

Department of Informatics
University of Fribourg (Switzerland)

SmartParticipation

A Fuzzy-Based Recommender System for Political Community-Building

Ph.D. Thesis

presented to the Faculty of Science of the University of Fribourg
(Switzerland) in consideration for the award of the academic grade of
Doctor scientiarum informaticarum

by

Luis Fernando Terán Tamayo

from

Quito, Ecuador

Thesis No. 1822

FRICOPY

28.01.2014

Accepted by the Faculty of Science of the University of Fribourg (Switzerland) upon the recommendation of:

- Prof. Ulrich Ultes-Nitsche: University of Fribourg, Switzerland, as Jury President.
- Prof. Andreas Meier: University of Fribourg, Switzerland, as Thesis Supervisor.
- Prof. Philippe Coudré-Mauroux: University of Fribourg, Switzerland, as Internal Expert.
- Prof. Andreas Ladner: University of Lausanne, Switzerland, as External Expert.

Fribourg, January 28th, 2014



Prof. Dr. Andreas Meier
Thesis Supervisor



Prof. Fritz Müller
Faculty Dean

I dedicate this thesis to my beloved wife and son.

Thanks, Pitu, for your endless support.

I love you so much.

Acknowledgements

First of all, I would like to thank Andreas Meier for his support during these years, not only as a supervisor but also as my mentor, for his academic and personal inputs. Thanks, Andreas, for encouraging and motivating me to reach this goal.

I would like to express my gratitude to Prof. Andreas Ladner for being always open to collaborate with me and support my work. I cannot forget to acknowledge the people from *smartvote*, especially to Jan Fivaz and Stefani Gerber for providing me with their advice and the necessary data to prove the concept presented in this thesis.

Special thanks goes to Philippe Coudré-Mauroux for his input, feedback, and for helping me with his knowledge to improve this work.

To all my colleagues at the IS group – Edy, Dany, Joël, Michael, Marcel, Aleksandar, Roland, Marco, and Darius – my gratitude for always being helpful and for all the great moments we had during the past few years.

Finally, I would like to thank my parents for being supportive of me during this time away from home. They are my inspiration; and, even though we were apart, they were always in my thoughts.

Luis Fernando Terán Tamayo
January 2014

Abstract

As the use of the Internet proliferates, so does the amount of information available. With the introduction of Web 2.0, which includes users as content generators, finding relevant information is even more complex. To tackle this problem of information overload, a number of different techniques have been introduced, including search engines, Semantic Web, and recommender systems, among others.

Social networks and communities have become an important environment for exchanging information about products, services, music, and movies, among other things. In an information and knowledge society, such technologies could also improve democratic processes, increase citizens' interest in political issues, enhance participation, and renew civic engagement. However, the difficulty of finding other citizens or parties that share common goals is still a barrier to overcome.

In this work, a fuzzy-based recommender system architecture for stimulating political participation and collaboration is proposed. The *SmartParticipation* project uses the database of *smartvote*, a well-known voting advice applications (VAAs) for local, cantonal, and national elections in Switzerland. The recommendation engine works with a modified fuzzy c-means algorithm (FCM) and the Sammon mapping technique, which is used for visualizing recommendations.

Additionally, an evaluation framework for *eParticipation* is presented, which allows one to analyze different projects and their development towards the enhancement of citizens' participation and empowerment. Initial results demonstrate the potential for building political communities and the stimulation of civic participation.

Kurzfassung

Mit zunehmendem Wachstum und Verbreitung des Internets steigt auch die Menge an vorhandenen Informationen stetig an. Besonders seit der Einführung des Web 2.0 und der damit verbundenen Etablierung von Nutzern als Informationsquelle, wurde die Suche nach der richtigen Information im Web komplexer. Eigens zur Meisterung dieser Problematik wurden Lösungen entwickelt, die in Form von Empfehlungssystemen, dem Semantic Web und Suchmaschinen auf dem Internet verfügbar sind.

Soziale Netzwerke und digitale Gemeinschaften sind zu wichtigen Umgebungen mutiert, in denen Informationen über Produkte, Dienstleistungen, Musik, Filme und andere Produkte ausgetauscht werden. In solch einer Informationsgesellschaft ist es dementsprechend naheliegend technische Hilfsmittel einzuführen und bereitzustellen, die Bürgern darin helfen ihre demokratischen Rechte besser wahrnehmen zu können. Dadurch wird der demokratische Prozess verbessert und das zivile Engagement erhöht. Lediglich die Auswahl der richtigen Kandidaten bei einer Wahl ist eine Hürde, die es noch zu überwinden gilt.

In dieser Arbeit wird ein Empfehlungssystem vorgeschlagen, welches auf Fuzziness basiert und dessen Ziel es ist den politischen Prozess und die Kandidatenwahl zu erleichtern. Das *SmartParticipation* Projekt nutzt Daten von *smartvote*, einem bekannten voting advice applications (VAAs) für lokale, kantonale und nationale Wahlen in der Schweiz. Das Empfehlungssystem arbeitet mit einem modifizierten c-means algorithmus (FCM) und der Sammon mapping zur visualisierung der Empfehlungen.

Zusätzlich wird ein Evaluations Framework für *eParticipation* vorgestellt. Dieses ermöglicht unterschiedliche Projekte und deren Beitrag zur Verbesserung der Partizipation und Stärkung des demokratischen Prozesses auszuwerten. Erste Resultate zeigen auf, das grosses Potenzial im Bereich der Bildung von digitalen, politischen Gemeinschaften besteht und der Stärkung der Partizipation von Bürgern im demokratischen Prozess.

Résumé

La popularisation d'Internet a graduellement accru la quantité de données disponibles. Avec l'apparition du Web 2.0, au sein duquel les utilisateurs sont également des créateurs de contenu, trouver des informations pertinentes est une tâche complexe. Dans le but d'aborder ce problème, un certain nombre de techniques ont été implémentées afin de faire face au risque de surcharge d'information. Parmi ces derniers on peut compter les moteurs de recherche, le Web sémantique, les systèmes recommandeurs.

Les réseaux sociaux et communautés sont devenus un environnement propice à l'échange d'informations au sujet de produits, de services, de musique, de films, etc. Dans une société de connaissance et d'information, de telles technologies sont également à même d'améliorer les processus démocratiques, d'éveiller l'intérêt des citoyens pour la politique, d'augmenter la participation, et de donner un nouveau souffle à l'engagement civique. Toutefois la difficulté à trouver des citoyens ou des partis qui partagent la même visée est toujours une barrière à franchir.

Dans cette étude, une architecture de système recommandeur basée sur une logique floue est proposée afin d'inciter à la participation politique et à la collaboration. Le projet *SmartParticipation* utilise la base de données *smartvote*, une voting advice applications (VAAs) répandue pour les élections locales, cantonales et nationales en Suisse. Le moteur de recherche recommandeur emploie un fuzzy c-means algorithme (FCM), ainsi que la technique de la mise en correspondance Sammon adoptée pour la visualisation des recommandations.

Un cadre d'évaluation pour la *eParticipation* est en outre présenté. Ce dernier permet d'analyser différents projets, de même que leur développement en fonction du renforcement de la participation des citoyens et de leur prise de pouvoir. Les premiers résultats ont démontré un potentiel notable pour la construction de communautés politiques ainsi que pour la dynamisation de la participation civique.

Declaration

I, the undersigned, hereby declare that this Ph.D. thesis entitled “SmartParticipation – A Fuzzy-Based Recommender System for Political Community-Building” has been compiled by me under the supervision of Prof. Dr. Andreas Meier, Information Systems Research Group, University of Fribourg (Switzerland).

This thesis has not been previously submitted for the award of any degree, diploma, associate-ship, fellowship or its equivalent to any other University or Institution.



Luis Fernando Terán Tamayo
January 2014

Contents

Contents	xv
List of Figures	xix
List of Tables	xxiii
I Motivation & Objectives	1
1. Introduction	3
1.1. Motivation	3
1.2. Objectives	5
1.3. Research Questions	6
1.4. Research Methods	7
1.5. Research Issues	8
1.6. Thesis Outline	9
1.7. Published Work	10
II Background	13
2. eGovernment	15
2.1. eGovernment Framework	15
2.1.1. Process Level I - Information and Communication	16

CONTENTS

2.1.2. Process Level II - Production	16
2.1.3. Process Level III - Participation	17
2.2. eCollaboration	17
2.3. eDemocracy	19
2.3.1. Process Steps for eVoting and eElections	19
2.4. eCommunity	21
2.5. Remarks on the eGovernment Framework	22
2.6. Further Readings	23
3. Recommender Systems	25
3.1. Introduction to RSs	25
3.1.1. Challenges and Problems of RSs	26
3.2. RSs for eCommerce	27
3.2.1. Collaborative Filtering	27
3.3. Recommender System Techniques	28
3.3.1. Memory-Based CF Algorithms.	29
3.3.2. Model-based CF Algorithms.	32
3.3.3. Dimensionality Reduction	35
3.4. Remarks on RSs	37
3.5. Further Readings	38
4. Fuzzy Logic	41
4.1. Fuzzy Logic and Fuzzy Sets Theory	41
4.2. Fuzzy Set Operations	43
4.3. Sharp vs. Fuzzy Clustering	47
4.3.1. C-Means Algorithm	47
4.3.2. Fuzzy C-Means Algorithm	49
4.4. Remarks on the FCM Algorithm	52
4.5. Further Readings	52
III Conceptual Framework	55
5. Fuzzy Recommender System	57

5.1. General Architecture	57
5.2. User Profile Generation	58
5.3. Recommendation Engine	61
5.3.1. Fuzzy Cluster Algorithm	61
5.4. Visualization of the FRS	62
5.4.1. Top-N Recommendations	64
5.4.2. Community-Building Process	66
5.4.3. Evaluation of Dimensionality Reduction Methods	67
5.4.4. Validation of Fuzzy Clustering Methods	88
5.5. Remarks about the FRS	98
5.6. Further Readings	99
6. SmartParticipation	101
6.1. eParticipation	101
6.1.1. Participation Levels	102
6.2. Project Description	106
6.3. eParticipation Maturity Models	107
6.3.1. eParticipation Maturity Model for eCollaboration	107
6.3.2. eParticipation Maturity Model for eDemocracy	109
6.3.3. eParticipation Maturity Model for eCommunity	110
6.4. Remarks on the SmartParticipation Project	111
6.5. Further Readings	112
IV Evaluation Framework	113
7. Evaluation Framework	115
7.1. eParticipation Evaluation Framework	115
7.1.1. Framework Description	116
7.2. Evaluation of Participation Levels	119
7.2.1. Voting Advice Applications	128
7.3. Evaluation of VAAs	129
7.4. Remarks on the Evaluation Framework	132
7.5. Further Readings	133

CONTENTS

V	Implementation	135
8.	Architecture and Implementation	137
8.1.	<i>SmartParticipation</i> Architecture Overview	137
8.2.	Web Interface	139
8.3.	User Profile	143
8.3.1.	Static Profile	143
8.3.2.	Dynamic Profile	144
8.4.	FRS Implementation	147
8.5.	Further Readings	150
VI	Conclusions	151
9.	Discussion and Conclusions	153
9.1.	Discussion	153
9.2.	Outlook	156
9.3.	Conclusions	158
A.	List of VAAs	165
B.	Evaluation of the <i>smartvote</i> project	167
C.	ICT Tools for eParticipation	169
D.	Mockup for SmartParticipation	171
E.	<i>smartvote</i> Match Points Computation	173
	Nomenclature	177
	References	185
	Internet Sources	195
	Curriculum Vitae	203

List of Figures

2.1. eGovernment Framework of the University of Fribourg. Adapted from Meier [2012]	16
2.2. Strategies of Organizational Development. Adapted from Meier [2012]	19
2.3. <i>eVoting</i> and <i>eElections</i> as Part of a Process Chain. Adapted from Meier [2012]	21
3.1. The CF Process. Adapted from Sarwar <i>et al.</i> [2001]	28
3.2. Neighborhood Formation	30
3.3. Isolation of Co-rated Items. Adapted from Sarwar <i>et al.</i> [2001] . .	33
3.4. Item-based CF Algorithm. Adapted from Sarwar <i>et al.</i> [2001] . . .	34
3.5. Neighbor Formation in a Bi-dimensional Space. Adapted from Sarwar <i>et al.</i> [2000b]	37
4.1. Examples of Sharp & Fuzzy Sets	42
4.2. Set Students	44
4.3. Convex and Non-convex Sets	45
4.4. Support, α -cut, and Kernel of a Fuzzy Set	45
4.5. Union and Intersection of Two Sets	46
4.6. C-means Algorithm Execution	49
4.7. FCM Instance	51
5.1. Fuzzy-based Recommender System Architecture	58

LIST OF FIGURES

5.2. Fuzzy Profile (FP)	59
5.3. Fuzzy Profile Component (fpc)	60
5.4. Illustration of Dimensionality Reduction	63
5.5. Fuzzy Cluster Analysis Graphical Interface	64
5.6. Top-N Recommendation Graphical Interface	65
5.7. Similarity Percentage Illustration	65
5.8. Political Communities Graphical Interface	67
5.9. Taxonomy of Dimensionality Reduction Techniques. Adapted from Van der Maaten <i>et al.</i> [2009]	68
5.10. Dimensionality Reduction by PCA for 3 and 15 Political Parties .	70
5.11. Dimensionality Reduction by Sammon Mapping for 3 and 15 Po- litical Parties	72
5.12. Dimensionality Reduction by t-SNE for 3 and 15 Political Parties	75
5.13. Evaluation of t-SNE for Different Initial Conditions	77
5.14. Comparison of Geodesic Distance and Euclidean Distance. Adapted from Lee & Verleysen [2007]	78
5.15. Dimensionality Reduction by Isomap for 3 and 15 Political Parties	80
5.16. Evaluation of Isomap for different local minima	81
5.17. Dimensionality Reduction by LLE for 3 and 15 Political Parties .	83
5.18. Evaluation of LLE for Different Local Minima	84
5.19. Evaluation of Five Dimensionality Reduction Methods	86
5.20. Cluster Shapes in \mathbf{R}^2 . Adapted from Balasko <i>et al.</i> [2005]	88
5.21. Fuzzy Clustering by the FCM for 3 and 13 Political Parties	91
5.22. Fuzzy Clustering by the FGK for 3 and 13 Political Parties	94
5.23. Fuzzy Clustering by the FGK for 3 and 13 Political Parties	96
6.1. eInforming	104
6.2. eConsulting	104
6.3. eDiscussion	105
6.4. eParticipation	105
6.5. eEmpowerment	105
6.6. The <i>SmartParticipation</i> Project	106
6.7. eParticipation Maturity Model for eCollaboration	108

LIST OF FIGURES

6.8. eParticipation Maturity Model for eDemocracy	110
6.9. eParticipation Maturity Model for eCommunity	111
7.1. eParticipation Evaluation Output Example	117
7.2. Evaluation Framework for eDemocracy	118
7.3. Web Evolution	119
7.4. Semantic Web Stack	123
7.5. Media Richness	124
7.6. Components Evaluation	130
7.7. VAAs Evaluation	131
8.1. Architecture of SmartParticipation Platform	138
8.2. Overview of Drupal Modules. Adapted from Tomlinson & VanDyke [2010]	139
8.3. Static Profile GUI	143
8.4. Block 1 - General Information	144
8.5. Block 2 - Additional Sources	145
8.6. Block 3 - Rating	145
8.7. Block 4 - Comments	146
8.8. Block 5 - List of Comments	146
8.9. The Process of Compilation and Distribution to End Users	147
8.10. FRS GUI	148
B.1. Evaluation of the <i>smartvote</i> project	167
D.1. Discussion Topics GUI Mockup	171

LIST OF FIGURES

List of Tables

1.1. Design-Science Research Guidelines. Adapted from Hevner <i>et al.</i> [2004]	7
4.1. Fuzzy Partition of the Data Set	52
5.1. Parameters Used for the Evaluation	85
5.2. Properties of Techniques for Dimensionality Reduction. Adapted from Van der Maaten <i>et al.</i> [2009]	87
5.3. Validation of Fuzzy Cluster Methods - 3 Political Parties	97
5.4. Validation of Fuzzy Cluster Methods - 15 Political Parties	97
6.1. eParticipation Levels from Grönlund [2009]	103
7.1. eParticipation Evaluation Frameworks	116
7.2. Characteristics of Media that Determine Richness of Information. Adapted from Daft & Lengel [1984]	123
7.3. Types of Collaboration Technologies	124
7.4. Web Evolution Levels	126
7.5. Media Richness Levels	128
B.1. Evaluation of the <i>smartvote</i> Project	168
C.1. ICT Tools for eParticipation	169
E.1. Budget Questions	174

LIST OF TABLES

E.2. Standard Questions	174
-----------------------------------	-----

List of Algorithms

4.1. C-Means Algorithm	48
4.2. Fuzzy C-Means Algorithm	51
5.1. Modified Fuzzy C-means Algorithm	63
5.2. Sammon Mapping Algorithm	72
5.3. t-Distributed Stochastic Neighbor Embedding Algorithm	76
5.4. Isomap Algorithm	79
5.5. Locally Linear Embedding Algorithm	82
5.6. Fuzzy Gustafson-Kessel Algorithm	93
5.7. Fuzzy Gath-Geva Algorithm	95

LIST OF ALGORITHMS

Part I

Motivation & Objectives

Chapter 1

Introduction

This chapter gives the reader a general overview and scope of this Ph.D. thesis, which uses a fuzzy-based recommender system approach for stimulating citizens' participation in the field of Information Systems (IS). Additionally, a summary of the technical contributions and the list of the papers previously published by the author of this dissertation are provided. The chapter is structured as follows: First, Sect. 1.1 provides the motivation of this work; then, Sect. 1.2 delimits the main objectives. Section 1.3 provides the research questions that will be answered during the development of the thesis; Sect. 1.4 delineates the research method framework; Sect. 1.5 enumerates the research issues that are addressed in this thesis. Finally, Sects. 1.6 and 1.7 provide the outline of this thesis and describe the published contributions that are part of this work.

1.1 Motivation

Nowadays, the use of the Internet has a specific purpose: the search for information. Unfortunately, the amount of information available on the Internet has grown unexpectedly and exponentially (Internet World Stats [2012]). This rapid growth has generated a question: How do we find relevant information? To resolve the problem of information overload on the World Wide Web (WWW), a number of technologies are used, such as search engines, recommender systems (RSs) and the Semantic Web. These technologies are briefly described as follows:

- **Search Engines.** These are technologies based on “search” robots (e.g., Google, Yahoo!, and Bing, among others), which automatically and periodically scan the Web in search of a new uniform resource locator (URL), which is a sequence of characters that complies with a standard format and a model that is used to name resources on the Internet for location or identification, indexing, and storage in databases.

When a user searches for information on the Web through a search engine, he must interact with the system by entering keywords. The system then will deliver, according to certain algorithms (these algorithms are not public and are only developed to meet certain general criteria), the Web addresses of sites that have something to do with what the user is looking for. These search engines are becoming more powerful, and because of the large amount of information the Web contains, you can be confident that the results obtained by the search engines are really the most relevant (Baeza-Yates & Ribeiro-Neto [2011]).

- **Recommender Systems (RSs).** The term “recommender system” refers to software tools and techniques that help reduce the problems of information overload, providing suggestions for items that might interest the users. RSs are based on collaborative filtering (CF) methods. These techniques are most used in *eCommerce* (some examples of sites that use RSs are amazon.com, and ebay.com, among others) to suggest items using different decision-making processes (Ricci *et al.* [2011]).
- **Semantic Web.** Also known as the “Web of Data,” the Semantic Web is based on the aggregation of semantic and ontological metadata of information contained on the Web. It is not a separate Web but an extension of the current one, and it gives a well-defined meaning, description of content, and relationship of the data with the aim of achieving a “Smart Web.” Examples of the standards used by Semantic Web are Extensible Markup Language (XML), Resource Description Framework (RDF) and Web Ontology Language (OWL) (Hitzler *et al.* [2009], Fensel *et al.* [2011]).

In the *eGovernment* sector, the amount of available information that is consumed by citizens (government plans, interviews, candidate profiles, political par-

ties, and discussion forums, among others) is also increasing. Political instability is not a recent problem, but a historical one. A key issue is that part of the population has not had access to education, among other basic goods and services. This social structure has made the “lack of memory” of citizens the best way to keep politicians in power. Thus, the social and political processes have based proposals on solving immediate problems and not a strategic development plan. On the other hand, constantly providing information about political proposals, offers and fulfillment thereof by the various political actors becomes a social obligation to make the political process truly democratic; and, thus, contribute to the so-called “Open Government” (Lathrop & Ruma [2010]).

Democratic processes are becoming a significant issue for citizens when they face election processes that require them to select their representatives from a large list of candidates since, in many cases, the candidates are relatively unknown to their constituents. Moreover, the need of citizens to create their own communities that can lead to projects, the creation of new political movements, and proposals for referendums, among others, can have a big impact on the inclusion of citizens and the enhancement of participation. Additionally to the problems mentioned above, the user must to be taken into account together with the limitations of the analysis of data, management of technology, and digital divide, among others. For that reason, providing a tool that can be used and understood by everyone is highly relevant.

1.2 Objectives

To address the problems listed above, this Ph.D. thesis proposes a Web-based platform called *SmartParticipation*, which intends to provide citizens with a simple, innovative, and independent alternative to enhance participation by using a fuzzy-based approach to provide recommendations and focusing on three participation areas: *eCollaboration*, *eDemocracy*, and *eCommunity*.

In the area of *eCollaboration*, the platform can be used by governments or private sectors to find citizens interested in taking part in various projects based on their profiles. In the area of *eDemocracy*, the platform can be used to monitor, evaluate and provide relevant information on different political actors. In the

area of *eCommunity*, the platform could provide tools for creating virtual communities. Additionally, the *SmartParticipation* platform can contribute toward enhancing the so-called *public memory*, which leads to better political control (Meier [2012], p. 160). The information presented by *SmartParticipation* should be displayed in a user-friendly interface and be easy to understand, aiming at a sector of society that might not be familiar with the latest technologies. This application offers the possibility for citizens to participate in national issues by opening channels of discussion and debate through the use of information and communication technologies (ICT), Web 2.0 and Semantic Web. To meet these objectives, the system must be designed taking into account different disciplines such as RSs, *eParticipation*, Semantic Web, human computer interaction (HCI), and data visualization.

1.3 Research Questions

In this section, a number of research questions are presented that provide an overview of the goals and scope of this Ph.D. thesis. In the conceptual and implementation phases, the following research questions shall be investigated:

1. What are the differences between classic RSs used in *eBusiness* and those used in *eGovernment*?
2. How can RSs be used to improve citizens' participation?
3. How can fuzzy logic be applied in RSs?
4. What are the advantages and disadvantages of using a fuzzy-based recommender system?
5. What type of architecture shall be chosen for the development of a fuzzy-based recommender system?
6. How should one conceptually evaluate the system developed and the implementation?

1.4 Research Methods

This Ph.D. thesis uses the guidelines for design science in IS research proposed by Hevner *et al.* [2004] and summarized in Table 1.1, which consists of seven guidelines to assist researchers, reviewers, editors, and readers to understand the requirements for effective design-science research.

Table 1.1: Design-Science Research Guidelines. Adapted from Hevner *et al.* [2004]

Guideline	Description
1st. Design as an Artifact	Design-science research must produce a viable artifact in the form of a construct, a model, a method, or an instantiation.
2nd. Problem Relevance	The objective of design-science research is to develop technology-based solutions to important and relevant business problems.
3rd. Design Evaluation	The utility, quality, and efficacy of a design artifact must be rigorously demonstrated via well-executed evaluation methods.
4th. Research Contributions	Effective design-science research must provide clear and verifiable contributions in the areas of the design artifact, design foundations, and/or design methodologies.
5th. Research Rigor	Design-science research relies upon the application of rigorous methods in both the construction and evaluation of the design artifact.
6th. Design as a Search Process	The search for an effective artifact requires utilizing available means to reach desired ends while satisfying laws in the problem environment.
7th. Communication of Research	Design-science research must be presented effectively both to technology-oriented as well as management-oriented audiences.

Design science is a problem-solving-based information technology research methodology that offers specific guidelines for evaluation and iteration within research projects. It focuses on the development and performance of artifacts with the explicit intention of improving the functional performance of the artifact. Design science research is typically applied to categories of artifacts including algorithms, human/computer interfaces, and design methodologies, among others. Since the main objective of this Ph.D. thesis is to develop a Web application, the design science approach gives the necessary framework for the implementation and development of this thesis. Design-science research requires the creation of an innovative artifact (1st guideline); and, in the case of this thesis, the artifact to

Introduction

develop is a Web application. The artifact belongs to a specified problem domain (2nd guideline); in the case of this thesis, the domain is *eParticipation*. Since the artifact is purposeful, it must be useful in resolving the specified problem. Hence, thorough evaluation of the artifact is crucial (3rd guideline).

Novelty is similarly crucial since the artifact must be innovative, solving a heretofore unsolved problem or solving a known problem in a more effective or efficient manner (4th guideline). In this way, design-science research is differentiated from the practice of design. The artifact itself must be rigorously defined, formally represented, coherent, and internally consistent (5th guideline). The process by which it is created, and often the artifact itself, incorporates or enables a search process whereby a problem space is constructed and a mechanism posed or enacted to find an effective solution (6th guideline).

Finally, the results of the design-science research must be communicated effectively (7th guideline), both to a technical audience (researchers who will extend them and practitioners who will implement them) and to a managerial audience (researchers who will study them in context and practitioners who will decide if they should be implemented within their organizations).

1.5 Research Issues

This Ph.D. thesis is divided in three phases that combine academic research, implementation of prototypes, and evaluation. The first phase of this thesis will be focused on literature research in the domains of RSs, fuzzy logic, fuzzy classification, fuzzy clustering, HCI, data visualization and open government. It will give first indications regarding which concepts can be used, modified, or need to be newly developed. This will help to demarcate the scope of the thesis.

The second part of this work is the implementation of prototypes of the desired system based on previous research that can meet the objectives of both user and designer in terms of usability and data analysis, among others. Finally, the evaluation of the system will take place using the prototypes developed and various techniques of clustering and data visualization. In order to evaluate the designed user interface (UI), a heuristic approach is used.

1.6 Thesis Outline

This thesis is organized in five parts, each part containing different chapters. The first part provides the background; the second part, the conceptual background; the third part, the evaluation framework; the fourth part, implementation; and the fifth part, concluding remarks and outlook. In this section, a brief description and the content of each chapter is presented.

Part II - Background

- **Chapter 2 - eGovernment.** In this chapter, the eGovernment framework used in this Ph.D. thesis is described. It introduces the concepts of *eCollaboration*, *eDemocracy*, *eCommunity*, and *eParticipation*, which make up the goals of this work.
- **Chapter 3 - Recommender Systems.** In this chapter, a brief introduction of RSs, challenges, and problems of RSs, the most-used techniques in RSs based on CF technologies, and different CF methods are presented.
- **Chapter 4 - Fuzzy Logic.** This chapter gives a brief introduction to fuzzy logic and fuzzy sets theory. Additionally, the main fuzzy set operations are described. This chapter gives a brief introduction of sharp and fuzzy clustering. It describes the main advantages of using the fuzzy methods used in the recommendation approach compared to the classical sharp clustering methods.

Part III - Conceptual Framework

- **Chapter 5 - Fuzzy Recommender System.** In this chapter, the architecture of the recommender system used by *SmartParticipation*, the prototype developed, and the different types of GUI are presented. The results presented in this section are used with the data provided by the smartvote [2012b] system.
- **Chapter 6 - SmartParticipation.** This chapter shows the growth of research in the area of eParticipation. Moreover, it introduces the five

Introduction

participation levels that are used by the *SmartParticipation* project. Additionally, it describes three maturity models for *eCollaboration*, *eDemocracy*, and *eCommunity*.

Part IV - Evaluation Framework

- **Chapter 7 - Evaluation Framework.** In this chapter, a quantitative framework for evaluating *eParticipation* projects is presented. Additionally, it presents the evaluation of a number of various VAAs, which are used in the area of *eDemocracy*.

Part V - Implementation

- **Chapter 8 - Architecture and Implementation.** This chapter presents the implementation of the *SmartParticipation* project and the different solutions implemented.

Part VI - Conclusions

- **Chapter 9 - Discussion and Conclusions.** In this chapter, concluding remarks and suggestions for future research are presented.

1.7 Published Work

In this section, a summary of the technical contributions and the list of the papers published by the author of this dissertation are provided. These resources are all related to the motivation or a part of this thesis.

- **Terán & Drobnjak [2013].** In this work, the authors describe five levels of participation, which takes into account the advantages of the social Web or Web 2.0, together with a quantitative approach for the evaluation of *eParticipation* projects. Each participation level is evaluated independently, taking into account three main components: Web evolution, media richness, and communication channels. This paper presents the evaluation of a number of existing VAAs. The results provide an overview of the main

- features implemented by each project, their strengths and weaknesses, and the participation levels reached.
- **Terán [2012]**. In this work, a new clustering visualization for the creation of Political Communities based on issues is presented. It also includes a new definition of the so-called fuzzy profiles that allows visualization of recommended citizens and candidates close to the preferences of users based taking as reference the various issues that are proposed by the *smartvote* project.
 - **Terán *et al.* [2012]**. This work compares the FRS introduced in the previous work using an existing VAA (*smartvote* [2012a]) that is used in this Ph.D. thesis for analysis and evaluation of results. The *smartvote* system provides recommendations to voters about candidates who are close to their political preferences and tendencies for local, cantonal, and national elections in Switzerland.
 - **Terán [2011]**. This paper introduces the creation of political communities using a variation of the FRS introduced in previous works and some information about the user's profile. Initial results demonstrate the potential for building political communities and the stimulation of civic participation.
 - **Terán & Meier [2011]**. In this paper, the *SmpartParticipation* platform is introduced. The goal of this project is to offer the possibility for citizens to participate in national issues by opening channels of discussion and debate through the use of ICT and the use of Web 2.0.
 - **Terán & Meier [2010]**. This paper presents the architecture of RSs for *eElections* using fuzzy clustering methods, is presented. The objective is to assist voters in making decisions by providing information about candidates close to the voters' preferences and tendencies. The use of RSs in *eGovernment* aims to increase participation of citizens in democratic processes through the use of information and communication technologies. In this paper, a first prototype of the FRS, which is used mainly in this thesis, is presented.

Introduction

Part II

Background

Chapter 2

eGovernment

The European Commission refers to *Electronic Government (eGovernment)* as the use of information technologies to improve the interaction between public administrations, citizens, and the private sector. In this chapter, the reader will be presented with the different levels that are part of the eGovernment framework, which is used in this work, in order to understand to what extent *Electronic Participation (eParticipation)* is important to this work and what can be covered by the *SmartParticipation* project. The chapter is structured as follows: First, Sect. 2.1 gives a brief introduction about the framework used in this Ph.D thesis. Then, Sect. 2.2 describes the different tools that can be used to enhance a collaborative environment. Section 2.3 gives a brief introduction of *Electronic Democracy (eDemocracy)* and the process steps for *Electronic Voting (eVoting)* and *Electronic Elections (eElections)*. In Sect. 2.4, some alternative communication and Web-based tools for community formation are presented. Section 2.5 presents some remarks about the eGovernment framework that is used by the *SmartParticipation* system. Finally, further readings are presented in Sect. 2.6.

2.1 eGovernment Framework

Three types of relationships are defined for *eGovernment*: Administration to Citizens (A2C), Administration to Business (A2B), and Administration to Administration (A2A). The eGovernment framework of the University of Fribourg

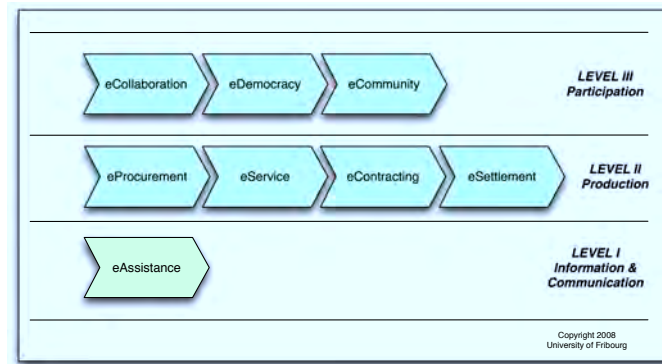


Figure 2.1: eGovernment Framework of the University of Fribourg. Adapted from Meier [2012]

is adopted from the work of Meier [2012], which is used by the *SmartParticipation* project and is a process-oriented maturity model with three primary levels. It is illustrated in Fig. 2.1.

2.1.1 Process Level I - Information and Communication

Electronic Assistance (eAssistance) corresponds to the lowest level, which provides information and communication facilities for *eGovernment*. It focuses on the design of communal Web portals; respectively, more extensive *eGovernment* portals, and the use of Web 2.0 technologies. For barrier-free Web access, compliance with the Web Content Accessibility Guidelines (WCAG) proposed by the World Wide Web Consortium (W3C) is imperative.

2.1.2 Process Level II - Production

The middle process level consists of administrative public services required for options A2A, A2C, and A2B. These services include management services for *Electronic Procurement (eProcurement)*; traditional services, such as taxation, residents' registration, identification acquisition, *Electronic Health (eHealth)*, and public education known as *Electronic Service (eService)*; contracts based on digitally signed electronic documents are managed by *Electronic Contracting (eContracting)*; and fulfillment elements, such as electronic shipment, electronic payment, and the assurance of data security and safety, are handled by *Electronic Settlement (eSettlement)*.

2.1.3 Process Level III - Participation

Notably, the top level emphasizes civic collaboration and participation, requisites of a progressive and responsible knowledge society. Virtual forms of collaboration (*eCollaboration*), including the utilization of social and Semantic Web technologies, enable the knowledge society to develop further. In addition to *eVoting*, communication platforms, including processes to build new communities and political networks (*eCommunity*), as well as stimulate participation and knowledge exchange between citizens.

In evaluating public services for *eGovernment*, most countries are focusing on information exchange and support for administrative processes in public affairs. In other words, governmental authorities have implemented process elements of the two lower levels (I and II), as shown in the eGovernment framework in Fig. 2.1.

2.2 eCollaboration

In the work of Meier [2012], important procedures and systems for computer-aided collaboration are described. The different tools that can be used to enhance collaborative environment in a Web-based information systems in *eGovernment* are briefly described in this section.

Document Management. This section analyzes the components of a Web-based information system and deals with document administration.

Content Management. Content management aims at planning and coordinating all activities for the supply and use of content.

Wiki Tools. Wiki tools were introduced first by Ward Cunningham (Leuf & Cunningham [2001]) and aim to facilitate users' ability to edit entries quickly and easily.

Use of Weblogs. A Weblog, or short blog, is a frequently updated digital journal whose entries are displayed in a chronologically descending way. The editor (blogger) of a Weblog is either an individual person (private blog) or a

group of people (corporate blog). A Weblog can be a textual or multimedia journal, or its content can be a linked collection dedicated to different matters and topics. Normally, the readers of a Weblog comment on the content.

Collaborative Working Environment. Groupware systems are technically mature products for information exchange, workflow control, collaboration, and data management.

Virtual Organizations and Forms of Cooperation. Depending on the complexity of the services, the administrative units must rethink their organization. Organizations and administrative units are undergoing changes due to market alterations and social developments. *Electronic Collaboration (eCollaboration)* can also be considered a type of crowdsourcing, which is a distributed problem-solving and production process that involves outsourcing tasks to a network of people, also known as a crowd. This method can be used to accomplish tasks. The constituent characteristics of virtual organizations are:

- **Voluntary cooperation of several independent network partners:** The fusion of organizations to a virtual organizational network is voluntary and requires that the individual group members and the management body trust each other.
- **Common goal:** Every virtual organization formulates a common organizational goal and comes to an agreement regarding task sharing and collaboration.
- **Bundling of core competences:** Virtual organizations try to obtain the required expertise through their network partners.
- **Utilization of information and communication technology:** Virtual organizations consequently use the possibility of electronic communication and the electronic exchange of service; for example, by running a collaborative portal. Such a platform is needed for information, communication, and handling of projects in the virtual organization.

Figure 2.2 illustrates four different options of organizations, depending on the complexity of the service and the dimension of the economical and social change.

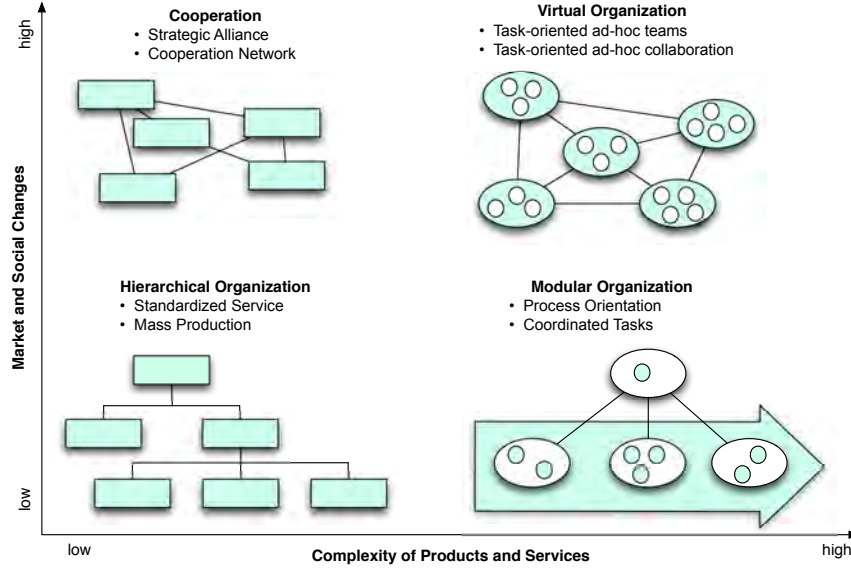


Figure 2.2: Strategies of Organizational Development. Adapted from Meier [2012]

2.3 eDemocracy

The term *eDemocracy* refers to the use of information and communication technologies to enable citizens to exercise their rights and fulfill their obligations in the information and knowledge society in a time- and place-independent manner. In his work, Meier [2012] mentions the importance of citizen participation in *eDemocracy* (e.g., *eElections* and *eVoting*). Meier defines the term *eDiscussion* as a stage in which citizens know more about the candidates or the subject during the voting process. *eDiscussion* uses information and communication technologies, such as discussion forums, decision-making aids, and subscription services, to aid voters (users) in making decisions. According to Meier, the next stage of *eDemocracy*, following *eVoting* and *eElections*, is *ePosting*, which facilitates the publication of results and gives voters (users) the chance to open up discussion channels about *eVoting* and *eElections*.

2.3.1 Process Steps for eVoting and eElections

Electronic votes and elections differ from traditional voting and election procedures mainly in their subsequent and post-processing phases (refer to Fig. 2.3) if

the advantages of electronic exchange relations are exploited. Through changed and expanded information and discussion of politics in the process steps of *eDiscussion* and *ePosting*, it is hoped that citizens will become more involved with political issues and engage in further community-building. Figure 2.3 distinguishes the following process steps:

eDiscussion. Prior to a vote or election, citizens can enhance their own opinion-forming process by requesting not only information, but also opinions and evaluations from discussion forums. Furthermore, subscription services allow the citizens to draw on documents or bases of decision-making and learn about changes in and extensions of topical issues.

eVoting. Within the timespan established by the authorities, citizens can fill out their electronic ballot and submit it. Before that, they identify themselves and register with a governmental institution; the subsequent vote, however, is made anonymously. The governmental institution can add an optional survey questionnaire to the ballot in order to, for example, get citizens' feedback on questions of procedure and implementation.

eElections. The published *spiderweb* profiles of political candidates and additional information on their abilities and skills make it easier for voters to fill out the electronic ballot (during the *eDiscussion* step). Again, citizens must register by means of an election and checking card and request a valid ballot from the governmental institution before voting electronically. Requesting that voters answer additional optional questions may be beneficial.

ePosting. The publication of voting and election results on the *eGovernment* portal for the associated governmental institution is directed at citizens but can also be studied and used by organizations and the press. To that end, suitable visualization and analysis tools can be offered so that electoral and voting behavior, and voting and election results, may be analyzed and discussed. Public blogs make it possible to comment on electronic votes and elections even after election day, enabling citizens to explore the relevant topics more deeply. Apart from

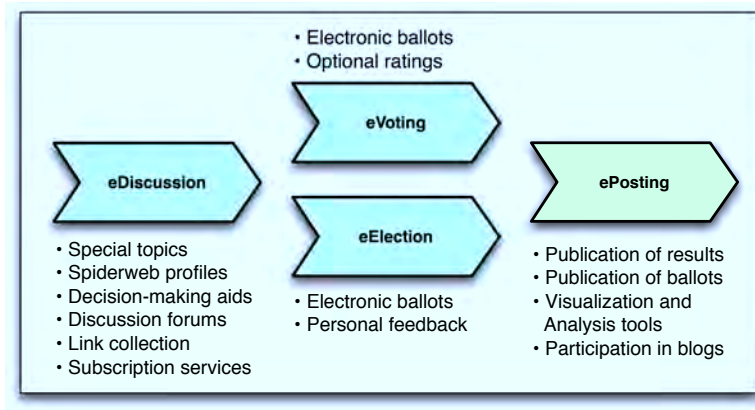


Figure 2.3: *eVoting* and *eElections* as Part of a Process Chain. Adapted from Meier [2012]

actual voting and election results, voting and election cards and un-ambiguous identification numbers should be published. By these means, all citizens will be able to verify whether their vote has actually been registered and processed. This method is more transparent than traditional votes and elections and will thus help to win the citizens' trust in *eVoting* and *eElections*. The description of the process steps for *eDiscussion*, *eVoting*, *eElections*, and *ePosting* shows that the use of electronic information and exchange relations increases citizens' involvement and stimulates public discussion.

2.4 eCommunity

The Internet is developing into an environment in which citizens display themselves, meet with others, exchange information and services, promote common projects, and overcome linguistic and cultural boundaries. This section presents some alternative communication and Web-based tools for *Electronic Community (eCommunity)*, as demonstrated by Meier [2012]. Computers and communication channels not only serve collaborators in the administration, enabling them to handle their workload, but also make encounters and communities possible. In the same way that street cafes, markets, and exhibitions serve as points of encounter in real life, besides home and the workplace, the computer network of networks develops into a virtual location. Topic-specific, cultural, or scientific meeting points on the Internet engender a new kind of community-building.

Computer networks are populated by citizens and avatars¹; Internet, or cyberspace, can enhance one's living environment. As in real living environments, infrastructures are developed for virtual spaces; platforms for exchange are supplied and services offered. In addition, rules of conduct and protective measures are implemented with the aim of safeguarding privacy and preventing misuses. Among the communities created on the Internet, two kinds are the most relevant:

- **Communities of interest.** These comprise citizens who share interest in a common thing or hobby.
- **Communities of practice.** These comprise groups of citizens who participate in a project together for a governmental institution, investing time and knowledge.

Both kinds of communities can benefit from information and communication systems. Community support systems serve as meeting places for members, as well as a place for them to exchange know-how and master tasks or challenges. Thus, Web-based platforms not only facilitate the development of communities, but also make it possible for people to meet other community members and utilize the community's collective know-how.

2.5 Remarks on the eGovernment Framework

The European Union has been pushing different measurement plans that point to enhancement of the development of *eGovernment*. Consequently, 12 areas of services for citizens and eight areas for services for the business sector are proposed. These services can be placed in the first and second levels of the eGovernment framework used in this Ph.D. thesis: Information & Communication and Production (Sects. 2.1.1 and 2.1.2).

The *eGovernment* framework proposed by the University of Fribourg is a maturity model that also aims to include in its third level (refer to Sect. 2.1.3) three processes that can lead to a better participation. The approach proposed in this Ph.D. thesis for advising citizens on elections (*eDemocracy*) and the creation of

¹In information society, avatars are images representing people who adopt a fictional identity on the Internet.

political communities (*eCommunity*) is based on the generation of a fuzzy cluster and uses a modified version of the fuzzy c-means algorithm (FCM) that is described in more detail in Sect. 5.3.1. Future work will also include *eCollaboration* in the *SmartParticipation* platform.

2.6 Further Readings

- **Meier [2012]**. “The reference book reviews and presents systematically the use of Internet in administration and politics. A process-oriented layer model defines the options of exchange and participation for all claim groups covering these topics: *eAssistance*, *eProcurement*, *eService*, *eContracting*, *eSettlement*, *eCollaboration*, *eDemocracy*, and *eCommunity*” (abs.).
- **Sloane [2011]**. “Open innovation and crowdsourcing are among the hottest topics in strategy and management today. The concept of capturing ideas in a hub of collaboration, together with the outsourcing of tasks to a large group of people or community is a revolution that is rapidly changing business culture” (abs.).
- **XRDS Magazine [2011]**. “This issue of XRDS is about how computer science can be used in service of democracy. There are so many projects under way—both inside and outside of government—aimed at improving governance with information technology, that it would be impossible to capture even a narrow segment of these in a single issue” (abs.).
- **Lathrop & Ruma [2010]**. “In a world where web services can make real-time data accessible to anyone, how can the government leverage this openness to improve its operations and increase citizen participation and awareness? Through a collection of essays and case studies, leading visionaries and practitioners both inside and outside of government share their ideas on how to achieve and direct this emerging world of online collaboration, transparency, and participation” (abs.).
- **Al-Hakim [2006]**. “Interest in e-government, both in industry and in academia, has grown rapidly over the past decade, and continues to grow.

Global E-Government: Theory, Applications and Benchmarking is written by experts from academia and industry, examining the practices of e-government in developing and developed countries, presenting recent theoretical research in e-government, and providing a platform to benchmark the best practices in implementing e-government programs” (abs.).

Chapter 3

Recommender Systems

Recommender systems (RSs) are computer-based techniques that attempt to present information about products that are likely to be of interest to a user. These techniques are mainly used in *Electronic Commerce (eCommerce)* in order to provide suggestions on items that a customer is, presumably, going to like. Nevertheless, there are other applications that make use of RSs, such as social networks and community-building processes, among others. In this chapter, the reader will be presented with different recommender system techniques, which are discussed in the academic literature and are part of the background information in this work. The *SmartParticipation* platform uses a fuzzy-based recommender system engine. This chapter is structured as follows: First, Sect. 3.1 gives a brief introduction of RSs, their challenges, and problems. Then, Sect. 3.2 describes the most-used techniques for RSs in *eCommerce*. Section 3.3 presents the most-used techniques in RSs, which are based on collaborative filtering (CF) technologies. Section 3.4 presents some remarks about the recommender system techniques presented in this chapter and the approach that is used by the *SmartParticipation* system. Finally, further readings are presented in Sect. 3.5.

3.1 Introduction to RSs

A recommender system is a specific type of information filtering technique that tries to present users with information about items (movies, music, books, news,

web pages, among others) in which they are interested. The term “item” is used to denote what the system recommends to users. To achieve this goal, the user profile is contrasted with the characteristics of the items. These features may come from the item content (content-based approach) or the user’s social environment (CF). The use of these systems is becoming increasingly popular in the Internet because they are very useful to evaluate and filter the vast amount of information available on the Web in order to assist users in their search processes and retrieval.

RSs have been highly used and play an important role in different Internet sites that offer products and services in social networks, such as Amazon, YouTube, Netflix, Yahoo!, TripAdvisor, Facebook, and Twitter, among others. Many different companies are developing RSs techniques as an added value to the services they provide to their subscribers. The use of RSs for *eGovernment* may reduce information overload, which could help to improve democratic processes. In this Ph.D. thesis, the recommendation engine developed provides citizens with a bi-dimensional, political/issue-based landscape to better understand their proximity to politicians or issues.

3.1.1 Challenges and Problems of RSs

A recommender system specifies two basic entities, which include the user (i.e., citizen, candidates) and the item (i.e., product, services). The main problems of RSs, according to Vozalis & Margaritis [2003], include the following:

1. **Quality of Recommendations:** The information received from a recommender system must be reliable; for that reason, RSs should minimize the number of false positive results (i.e., the products that the customer does not like).
2. **Sparsity:** A recommendation system is related to the number of recommendations made by customers. The sparsity problem of RSs emerges when the number of rated items is small compared to the total number of items, which leads to weak recommendations since the RSs are based on similarities between individuals.

3. **Scalability:** Increasing the number of users and products elevates the cost in terms of computations in RSs.
4. **Loss of Neighbor Transitivity:** The correlations between users cannot be expressed unless they have purchased and rated common items.
5. **Synonymy:** RSs generally cannot link products with different names that belong to the same category.
6. **First-rater Problem:** A product cannot be recommended unless another customer has previously rated it.
7. **Unusual User Problem:** This problem refers to users who cannot define their opinion about a product. This causes inconsistent recommendations.

In their work, Vozalis & Margaritis [2003] describe the challenges and problems of RSs. The most relevant are the *sparsity* and *first-rater* problems.

3.2 RSs for eCommerce

According to Yager [2003], RSs used in *eCommerce* are targeted marketing methods that rely on past experiences to increase the sales of products for *eCommerce* and specify two basic entities: the user (customer) and the item (product). The main goal of this type of recommender system used in *eCommerce* is basically to increase the sales of products. In the work of Vozalis & Margaritis [2003] and Sarwar *et al.* [2001], the more widely used techniques in RSs are based on CF methods. The goal of CF is to recommend or to predict the benefit of a specific product based on previously ranked items by the user and the opinions of other users with similar likings. CF methods include those that are memory-based (i.e., user-based) and model-based (i.e., item-based).

3.2.1 Collaborative Filtering

The goal of CF algorithms is to recommend a product or predict the benefit of a specific product based on previously ranked items by the user and the opinions of other users with similar likings. In a CF scenario, there is a list of users

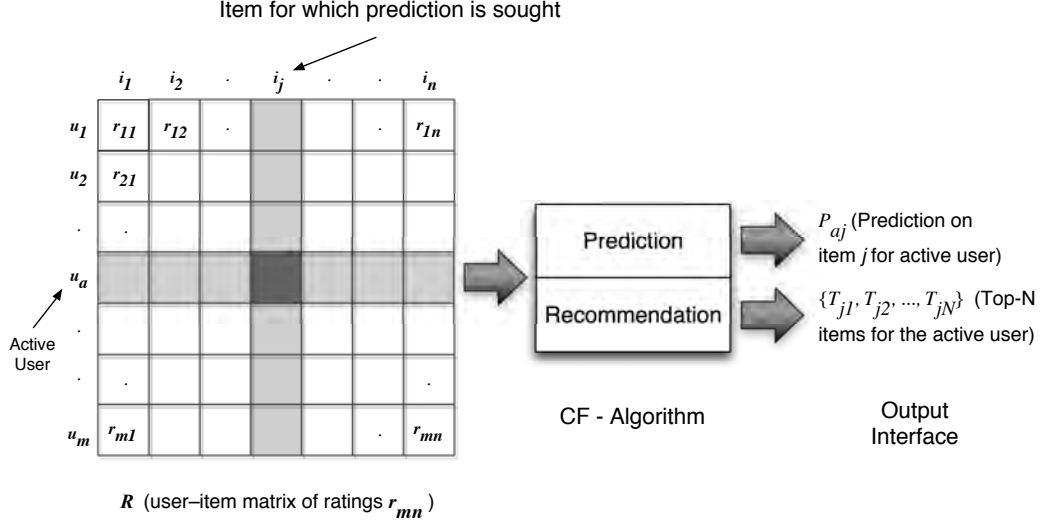


Figure 3.1: The CF Process. Adapted from Sarwar *et al.* [2001]

$U = \{u_1, u_2, \dots, u_m\}$ and a list of items $I = \{i_1, i_2, \dots, i_n\}$. Each user has a list of items, I_{ui} , of which the user has expressed his opinions (rating score). The active user, expressed as $u_a \in U$, is the user to whom the benefit of a specific product is recommended. In the matrix \vec{R} (user-item matrix), which has $m \times n$ dimensions, each entry, r_{ij} , represents the ratings of user u_i about an specific item, i_j . This rating is a numerical value varying from 0 to a maximum value (0 value means that the user has not ranked the product yet). Figure 3.1 shows a diagram of a CF process.

3.3 Recommender System Techniques

In the work of Ekstrand *et al.* [2011], six different collaborative algorithms to predict user preferences have been identified. These techniques are baseline predictors, user-user CF, item-item filtering, dimensionality reductions, probabilistic methods, and hybrid recommenders. The strategies presented in this work differ in their structure, but they keep in common the use of a preference-ranked list of items. In this section, the most-used techniques in RSs, which are based on CF technologies according to Guo & Lu [2007] and Sarwar *et al.* [2001], are presented. These two techniques are *memory-based (user-based)*, and *model-based*

(*item-based*). These techniques use the calculation of similarities between individuals (memory-based) or items (model-based). Additionally, the use of dimensionality reduction for RSs is presented, due to the kernel of the FRS proposed in this Ph.D. thesis. The use of dimensionality reduction, together with an evaluation of different algorithms, is described in Chap. 5.

3.3.1 Memory-Based CF Algorithms.

These techniques are based on the computation of “neighborhood formation” using the user–item matrix \vec{R} , which contains the ratings of items by users. Users are not required to provide their opinion on all items, which may cause the previously mentioned problem of sparsity. The most common techniques used to reduce the effect of sparsity consist of default voting, preprocessing using averages, the use of filterbots, and the use of dimensionality reduction techniques.

In order to compute the similarity between users u_i and u_k from \vec{R} , two methods are mentioned in this section: *Cosine/Vector Similarity* and *Pearson Correlation Similarity* (introduced by Pearson [1901]). Nevertheless, these are not the only similarity measures; a more detailed description can be found in Amatriain *et al.* [2011].

Cosine/Vector Similarity. To compute the proximity between two users, u_i and u_k , by calculating the similarity between their vectors, as the cosine of the angle formed between them, Eq. (3.1) is used as follows:

$$sim_{ik} = cos_{ik} = \frac{\sum_{j=1}^l r_{ij}r_{kj}}{\sqrt{\sum_{j=1}^l r_{ij}^2}\sqrt{\sum_{j=1}^l r_{kj}^2}} \quad (3.1)$$

The summations are computed over the l items that both users u_i and u_k have ranked.

Pearson Correlation Similarity. The Pearson correlation was initially introduced in the context of the GroupLens project by Resnick *et al.* [1994]. To find the Pearson correlation similarity between two users, u_i and u_k , Eq. (3.2) is used.

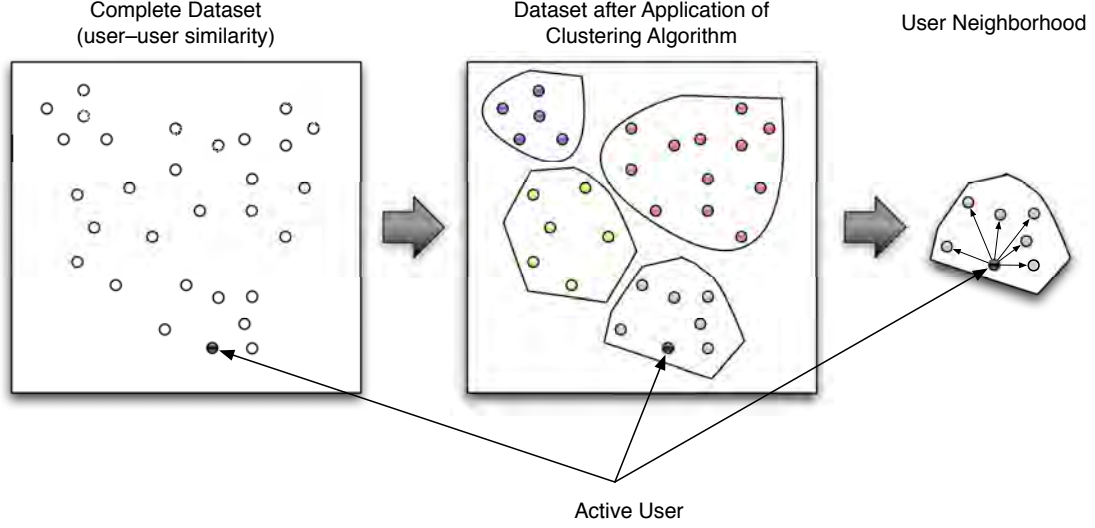


Figure 3.2: Neighborhood Formation

$$sim_{ik} = corr_{ik} = \frac{\sum_{j=1}^l (r_{ij} - \bar{r}_i)(r_{kj} - \bar{r}_k)}{\sqrt{\sum_{j=1}^l (r_{ij} - \bar{r}_i)^2 \sum_{j=1}^l (r_{kj} - \bar{r}_k)^2}} \quad (3.2)$$

\bar{r}_i and \bar{r}_k are the average rating value, which is computed using Eqs. (3.3) and (3.4) as follows:

$$\bar{r}_i = \frac{\sum_{j=1}^l r_{ij}}{l} \quad (3.3)$$

$$\bar{r}_k = \frac{\sum_{j=1}^l r_{kj}}{l} \quad (3.4)$$

The summations are computed over the l items that both users u_i and u_k have ranked. Using the *Pearson Correlation Similarity* or *Cosine/Vector Similarity*, the similarity matrix \vec{S} ($m \times m$), which includes the similarity values between users, can be generated. After the generation of \vec{S} , neighborhood formation can be computed based on several schemas. Figure 3.2 illustrates the process of neighborhood formation. According to Vozalis & Margaritis [2003], the most commonly used techniques for neighborhood formation are:

- **Centered-based schema.** Creates a neighborhood of size l for the active user (u_a) by simply selecting from the similarity matrix, \vec{S} ; and, more

specifically, from the row of matrix \vec{S} , which corresponds to the active user and those users who have the l highest similarity values with the active user.

- **Aggregate neighborhood schema.** Creates a neighborhood of users, not by finding the users who are closest to the active user (u_a), but by collecting the users who are closest to the center of the current neighborhood.

Once the neighborhood of users has been created, the final step in the recommendation process is to generate either a *prediction* or *top-n recommendation*, which are explained below:

Prediction. Prediction, P_{aj} , is a numerical value that expresses the possible opinion of the active user (u_a) about an item, i_j , regarding which he has expressed his opinion by computing the sum of the ratings given by the user on the items similar to i_j . It can be computed using Eq. (3.5):

$$P_{aj} = \bar{r}_a + \frac{\sum_{i=1}^v (r_{ij} - \bar{r}_i) \times sim_{ai}}{\sum_{j=1}^l |sim_{ai}|} \quad (3.5)$$

where v is the number of similar neighbors of u_a ; r_{ij} is the rating of item i_j for neighbor i ; \bar{r}_a and \bar{r}_i are the average rating value over all rated items for the users who belongs to the neighborhood; and sim_{ai} is the similarity between the users described above.

Top-N Recommendation. A list of N items that the active user (u_a) will like the most. Two techniques are the most commonly used in order to find the top-N recommendation according to Vozalis & Margaritis [2003] and Sarwar *et al.* [2000a]:

- **Most-Frequent Item Recommendations.** The system looks at the N neighborhoods of the active user (u_a) and performs a frequency count of the items that each neighbor user has purchased or rated. Once all neighboring users have been taken into account and the total counts for their rated items have been calculated, the system excludes the items rated by the

active user (u_a), sorts the remaining items according to their frequency counts, and returns the N most frequent items.

- **Association Rule-based Recommendation.** Based on the traditional data mining, association rule-based introduced by Lee & Stolfo [1998]. It does not make use of the entire population of customers to generate rules and considers only l neighbors for rule generation.

3.3.2 Model-based CF Algorithms.

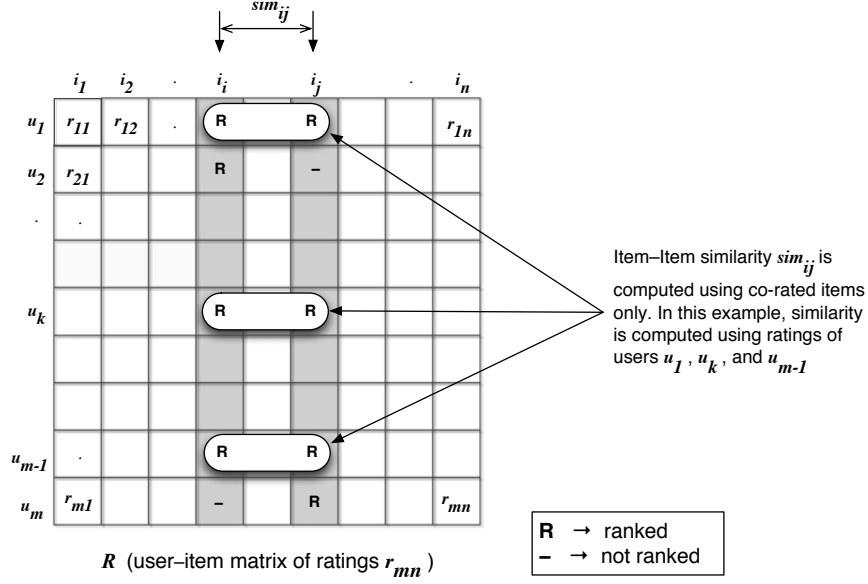
The model-based (i.e., item-based) CF algorithm uses the set of items that the active user (u_a) has ranked, computes the similarities between these items and a target item i_j , and to selects the n most similar items. The first step in the computation of similarities between items i_i and i_j , using the item-based recommendation algorithm, is to isolate the users who have rated both items. Figure 3.3 illustrates the use of isolation of co-rated items. It shows that the similarity is computed by taking into account those items that have been ranked by all users. The main techniques used to compute the similarity (sim_{ij}), according to Sarwar *et al.* [2001], are *Cosine/Vector Similarity*, *Pearson Correlation Similarity*, and *Weighted Sum*, which are briefly presented as follows.

Cosine/Vector Similarity. Equation (3.6) is used to compute the proximity between two items, i_i , and i_j , by calculating the similarity between their vectors, as the cosine of the angle formed between them.

$$sim_{ij} = cos_{ij} = \frac{\sum_{k=1}^l r_{ik}r_{jk}}{\sqrt{\sum_{k=1}^l r_{ik}^2} \sqrt{\sum_{k=1}^l r_{jk}^2}} \quad (3.6)$$

where r_{ik} and r_{jk} are the ratings of items i_i and i_j that have been given by user u_k . Obviously, the summations are computed over the l users who have been isolated (refer to Fig. 3.3), and who have expressed their opinions of both items.

Pearson Correlation Similarity. Equation (3.7) is used to find the *Pearson Correlation Similarity* between items i_i and i_j .


 Figure 3.3: Isolation of Co-rated Items. Adapted from Sarwar *et al.* [2001]

$$sim_{ij} = corr_{ij} = \frac{\sum_{k=1}^l (r_{ik} - \bar{r}_i)(r_{jk} - \bar{r}_j)}{\sqrt{\sum_{k=1}^l (r_{ik} - \bar{r}_i)^2 \sum_{k=1}^l (r_{jk} - \bar{r}_j)^2}} \quad (3.7)$$

where \bar{r}_i and \bar{r}_j are the average rating values, which are computed using Eqs. (3.8) and (3.9) respectively. The summations are computed over the l users that have been isolated (refer to Fig. 3.3).

$$\bar{r}_i = \frac{\sum_{k=1}^l r_{ik}}{l} \quad (3.8)$$

$$\bar{r}_j = \frac{\sum_{k=1}^l r_{jk}}{l} \quad (3.9)$$

Adjusted Cosine Similarity. A fundamental difference between *user-based* and *item-based* techniques is that, in the case of *item-based* techniques, the similarity is computed along the columns. This technique has a drawback, which is that the differences in rating scale between different users are not taken into account. The adjusted cosine similarity, presented by Sarwar *et al.* [2001], tackles this drawback by subtracting the corresponding user average from each co-rated

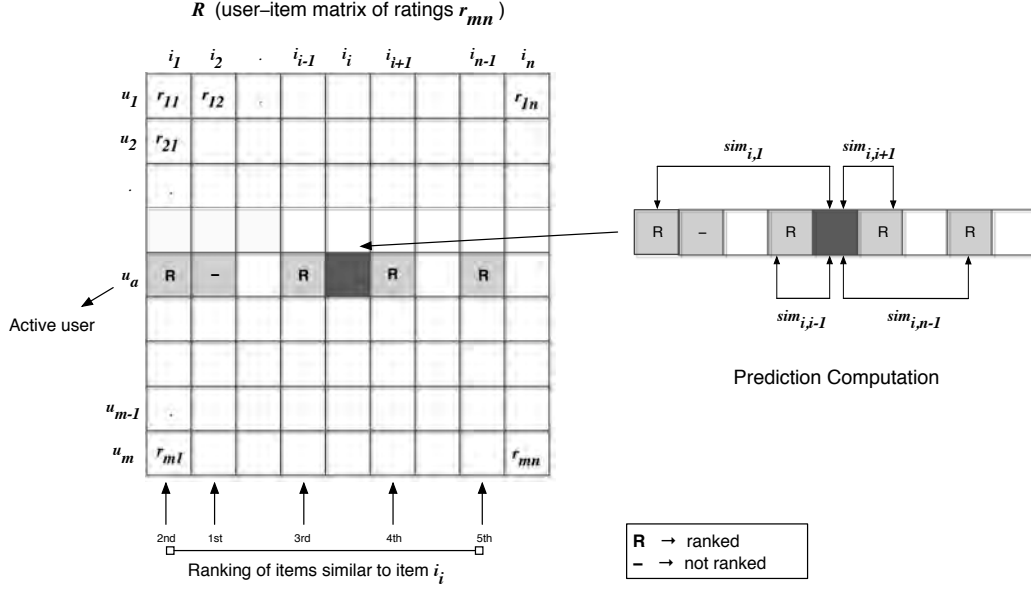


Figure 3.4: Item-based CF Algorithm. Adapted from Sarwar *et al.* [2001]

pair. To find the Pearson Correlation Similarity between items i_i and i_j , Eq. (3.10) is used.

$$sim_{ij} = adjcos_{ij} = \frac{\sum_{k=1}^l (r_{ik} - \bar{r}_i)(r_{jk} - \bar{r}_j)}{\sqrt{\sum_{k=1}^l (r_{ik} - \bar{r}_i)^2 \sum_{k=1}^l (r_{jk} - \bar{r}_j)^2}} \quad (3.10)$$

where r_{ik} and r_{jk} are the ratings that items i_i and i_j have received from user u_k , while \bar{r}_i is the average of the k -th user's ratings. The summations are computed over the l users that have been isolated (refer to Fig. 3.3).

After the computation of the similarities of items, the predictions of the n items that are the most similar to target item i_j are generated from its neighborhood N of items. According to Vozalis & Margaritis [2003], the most commonly used technique in order to find the prediction is the *Weighted Sum* (Sarwar *et al.* [2001]), which is described as follows.

Weighted Sum. Generates a prediction of item i_j for an active user (u_a) by computing the sum of ratings given by the active user on the items similar to i_j (belonging to the neighborhood of i_j). Each rating is weighted by the corresponding similarity, sim_{ik} , between items i_j and i_k . Using the notion shown in

Fig. 3.4 and Eq. (3.11), the prediction Pr_{aj} is computed. The summations are computed over n items in the neighborhood N .

$$Pr_{aj} = \frac{\sum_{k=1}^n sim_{jk} * r_{ak}}{\sum_{k=1}^n |sim_{ak}|} \quad (3.11)$$

Example 3.1 *Figure 3.4 is used to illustrate the computation of prediction using the item-based CF. It shows that the prediction is computed from the similarity among each of the ranked items by the active user and the objective item.*

3.3.3 Dimensionality Reduction

Traditional CF methods view the user–item domain as a vector space in a very high-dimensional space. A first approach to using dimensionality reduction techniques is described in the work of Sarwar *et al.* [2000b]. The approach presented in this work attempts to solve the sparsity problem (refer to Sect. 3.1.1) by integrating semi-intelligent filtering methods, which use singular value decomposition (SVD) algorithms to produce top-N recommendations. SVD is also used in information retrieval in order to find a document expressed as a term-document matrix, in which the cells represents the number of times each term is presented in a specific document.

Singular Value Decomposition

SVD is a matrix factorization method of a matrix, \mathbf{R} , divided into three matrices as defined in Eq. (3.12).

$$\mathbf{R} = \mathbf{U} \cdot \mathbf{S} \cdot \mathbf{V}^\top \quad (3.12)$$

where \mathbf{U} and \mathbf{V} are two orthogonal matrices of size $m \times r$ and $n \times r$, respectively. The rank of \mathbf{R} is represented by the values r . \mathbf{S} is a diagonal matrix of size $r \times r$, which has all singular values of matrix \mathbf{R} in its diagonal, and stored in descending order. \mathbf{V}^\top denotes the conjugated transposition of the $n \times n$ unitary matrix, \mathbf{V} . It is possible to reduce \mathbf{R} in terms of the Frobenius norm. Additionally, it is possible to reduce \mathbf{S} ($n \times n$) to have the k largest diagonal values and obtain S_k ,

Recommender Systems

where $k < r$. By reducing the matrices \mathbf{U} and \mathbf{V} in the same way, a reconstructed matrix, \mathbf{R}_k , which is close to \mathbf{R} , is generated. \mathbf{R}_k minimizes the Frobenius norm $\|\mathbf{R} - \mathbf{R}_k\|$.

$$\mathbf{R}_k = \mathbf{U}_k \cdot \mathbf{S}_k \cdot \mathbf{V}_k^\top \quad (3.13)$$

Example 3.2 *To illustrate how SVD is used for dimensionality reduction, the following matrices are used $\mathbf{R} = \mathbf{U} \cdot \mathbf{S} \cdot \mathbf{V}^\top$ correspondingly:*

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix} \cdot \begin{bmatrix} 9.64 & 0 \\ 0 & 5.29 \end{bmatrix} \cdot \begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix}$$

It is clear that \mathbf{S} is a 2×2 matrix with singular values arranged in descending order. The smallest singular value of \mathbf{S} is 5.29. If we extract the smallest singular value and the corresponding column and row from \mathbf{U} and \mathbf{V} , we have $\mathbf{R}_k \sim \mathbf{R}$ such that $\mathbf{R}_k = \mathbf{U}_k \cdot \mathbf{S}_k \cdot \mathbf{V}_k^\top$:

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0.18 \\ 0.36 \\ 0.18 \\ 0.90 \\ 0 \\ 0 \\ 0 \end{bmatrix} \cdot [9.64] \cdot \begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \end{bmatrix}$$

Finally, from the expression above, we can derive $\mathbf{R} \sim \mathbf{R}_k$ as follows:

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

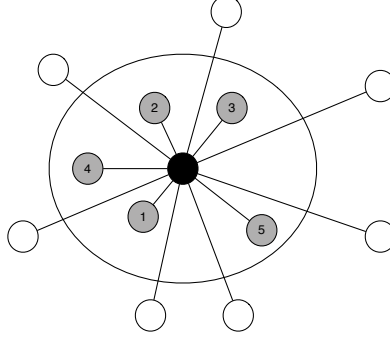


Figure 3.5: Neighbor Formation in a Bi-dimensional Space. Adapted from Sarwar *et al.* [2000b]

The use of dimensionality reduction in RSs allows a representation of original customer-product space and then the computation of the neighbor formation in the low-dimensional space. An example of a bi-dimensional representation of neighborhood formation ($k = 5$) is presented in Fig. 3.5.

Dimensionality reduction in RSs is used mainly to reduce the complexity and noise that could be presented in the data. At the same time, noise reduction allows a less sparse data set. Besides SVD, other matrix factorization methods are used for dimensionality reduction (e.g., the Eigentaste algorithm, introduced by Goldberg *et al.* [2001] uses Principal Component Analysis (PCA)). In Sect. 5.4.3 different dimensionality reduction methods are described and evaluated, which includes the PCA.

3.4 Remarks on RSs

According to Yager [2003], recommender systems, which are used in *eCommerce*, can be classified as “targeted marketing” since they use information that is based on the actions or past experiences of users. The accuracy of the recommendation of this type of method depends directly on users’ participation. In targeted marketing, the main objective of the recommendation is to increase the margin of sales by recommending products that the users are likely to find appealing.

This type of recommender system is only suitable for repeating items mentioned in Guo & Lu [2007], where the user can purchase an item even though some of them have been sold. Additionally, Yager [2003] makes a distinction

between RSs and targeted marketing. Yager considers a recommender system a “participatory” system in which the user intentionally provides information about his preferences. In a targeted marketing effort, the recommendation is based on extensional information, which is nothing but information predicated upon one’s actions or past experiences with respect to specific objects. The definition of RSs, as introduced by Yager, is used in this Ph.D. thesis. In addition, it is assumed that users are willing to provide information about their preferences. It is important to mention that participants (candidates and/or citizens) of RSs used for *eElections* in which events such as election processes occur only once cannot be considered unique since their presence at such events and their way of thinking can vary over time. The same argument is used for *eCommunity*. People’s political orientations can evolve or change over time. Therefore, using sharp clustering does not make sense even in the case of politicians who are usually members of a specific political party. According to their personal profiles, they may stand at different distances from the centers of political parties.

The approach proposed in this Ph.D. thesis for advising citizens on elections and the creation of political communities is based on the generation of a fuzzy cluster and uses a modified version of the fuzzy c-means algorithm (FCM). This approach takes most of the ideas from RSs for *eCommerce* since it provides top-N recommendations, of candidates, citizens, and political issues that are close to the user’s preference (profile). It also incorporates visualization by using a bi-dimensional political landscape to facilitate the evaluation of results and recommendations. This is described in more detail in Chap. 5.

Given that this work focuses on RSs, which can contribute to improved democratic processes in *eGovernment*, Yager’s definition of RSs is used in this thesis with the assumption that, in *eGovernment* systems, users are willing to participate in the process of providing information about their preferences (Yager [2003]).

3.5 Further Readings

- **Yoo *et al.* [2013].** “Whether users are likely to accept the recommendations provided by a recommender system is of utmost importance to system

designers and the marketers who implement them. By conceptualizing the advice seeking and giving relationship as a fundamentally social process, important avenues for understanding the persuasiveness of RSs open up” (abs.).

- **Ricci *et al.* [2011].**

“Recommender Systems Handbook, an edited volume, is a multi-disciplinary effort that involves world-wide experts from diverse fields, such as artificial intelligence, human computer interaction, information technology, data mining, statistics, adaptive user interfaces, decision support systems, marketing, and consumer behavior” (abs.).

- **Ekstrand *et al.* [2011].** “Collaborative Filtering Recommender Systems discusses a wide variety of the recommender choices available and their implications, providing both practitioners and researchers with an introduction to the important issues underlying recommenders and current best practices for addressing these issues” (abs.).

- **Jannach *et al.* [2010].**

“This book offers an overview of approaches to developing state-of-the-art RSs. The authors present current algorithmic approaches for generating personalized buying proposals, such as collaborative and content-based filtering, as well as more interactive and knowledge-based approaches. They also discuss how to measure the effectiveness of RSs and illustrate the methods with practical case studies” (abs.).

Chapter 4

Fuzzy Logic

Fuzzy logic is a multi-value logic that allows a better understanding of the result of a statement that is more approximate than precise in real life. In this chapter, the reader will be presented with the basic concepts of fuzzy logic and fuzzy sets, which are part of the core recommender system engine used by the *Smart-Participation* platform. This chapter is structured as follows: First, Sect. 4.1 gives a brief introduction of fuzzy logic and fuzzy set theory. Section 4.2 gives an overview of the basic operations of fuzzy sets. Then, Sect. 4.3 introduces the most-used techniques for sharp and fuzzy clustering. Section 4.4 presents some remarks about the fuzzy c-means algorithm (FCM) that is used by the *Smart-Participation* platform. Finally, further readings are presented in Sect. 4.5.

4.1 Fuzzy Logic and Fuzzy Sets Theory

In contrast with “sharp logic,” where the results of a statement are binary (“true or false” or “one or zero”), fuzzy logic admits a set of truth-values in the interval $[0, 1]$. Fuzzy logic is derived from fuzzy set theory, introduced by Zadeh [1965], where a fuzzy set is determined by a membership function with a range of values between 0 and 1. In his work, Zadeh [1965] gives a definition of fuzzy sets, which is shown in Definition 4.1.

Definition 4.1 *A fuzzy set is built from a reference set called a universe of discourse. The reference set is never fuzzy. Assuming $U = \{x_1, x_2, \dots, x_n\}$ as the uni-*

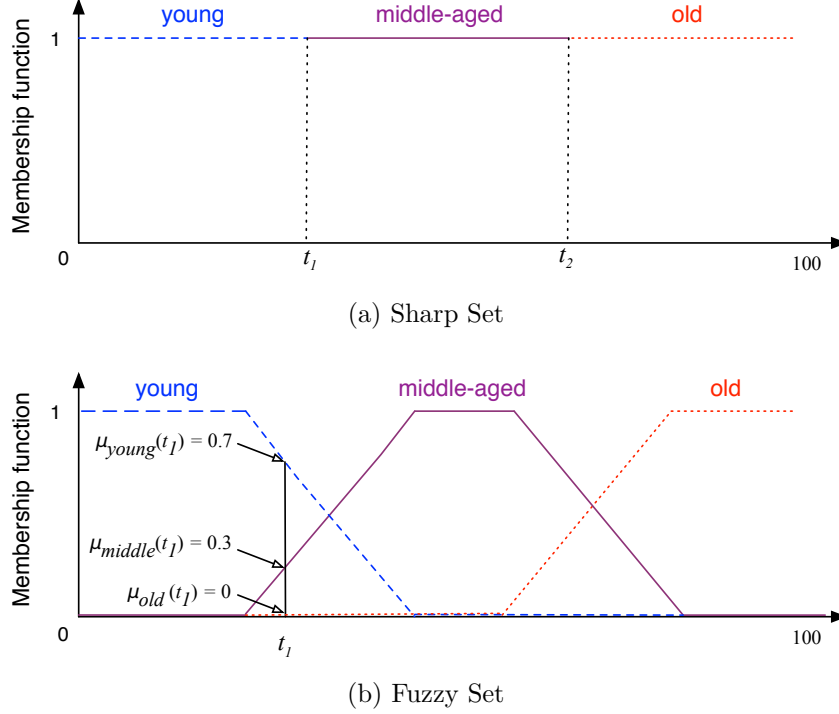


Figure 4.1: Examples of Sharp & Fuzzy Sets

verse of discourse, then a fuzzy set A in U ($A \subset U$) is defined as a set of ordered pairs: $\{(x_i, \mu_A(x_i))\}$ where $x_i \in U$, $\mu_A : U \rightarrow [0, 1]$ is the membership function of A , and $\mu_A(x) \in [0, 1]$ is the degree of membership of x in A .

Example 4.1 Figure 4.1 illustrates the difference between sharp sets and fuzzy sets, with the concepts “young,” “middle-aged,” and “old” as functions of the age of a citizen.

In Fig. 4.1a, it is clear that a person can move from young to middle and from middle-aged to old in a fraction of a second at t_1 and t_2 , respectively. Even though the sets young, middle-aged, and old are clearly delimited, they seem, therefore, unnatural as they do not match human perception due to their sharply fixed boundaries. On the other hand, Fig. 4.1b shows the membership functions $\mu_{\text{young}}(t)$, $\mu_{\text{middle}}(t)$, and $\mu_{\text{old}}(t)$ of age t . It is clear that, at age t_1 , a person belongs to three sets with different degrees of membership, which fits better into human perception.

Figure 4.1 introduces an essential concept in the fuzzy logic theory called a linguistic variable. A linguistic variable is a variable whose values are words or terms instead of numerical values, as described by Zadeh [1975a,b,c], and is presented in Definition 4.2. These terms (also linguistic or verbal terms) are represented by fuzzy sets.

Definition 4.2 *A linguistic variable is characterized by a quintuple:*

$$(X, T, U, G, M)$$

where X is the name of the variable, T is the set of terms of X , U is the universe of discourse, G is a syntactic rule for generating the name of the terms, and M is a semantic rule for associating each term with its meaning; i.e., a fuzzy set defined on U .

Example 4.2 *The linguistic variable represented in Fig. 4.1b is defined by the quintuple (X, T, U, G, M) where X is age and T is the set {young, middle-aged, old} generated by G and M specifies for each term a corresponding fuzzy set on the universe $U = [0, 100]$ given in years.*

As seen above, fuzzy logic is a form of many-valued logic that helps to better understand ambiguous concepts that cannot be sharply defined. In this section, an instance using the concepts young, middle-aged, and old as functions of the age of a person have been used to illustrate the advantages of fuzzy sets over classic sharp sets.

4.2 Fuzzy Set Operations

This section gives an overview of the basic operations of fuzzy sets introduced by Zadeh [1965], which are an extension of the classic set theory. In the literature, different authors have introduced additional operations that are applied in fuzzy set theory, which are described in more detail in Dubois & Prade [1980, 1985]; Yager [1980]; Zimmermann [2001]; Zimmermann & Zysno [1980].

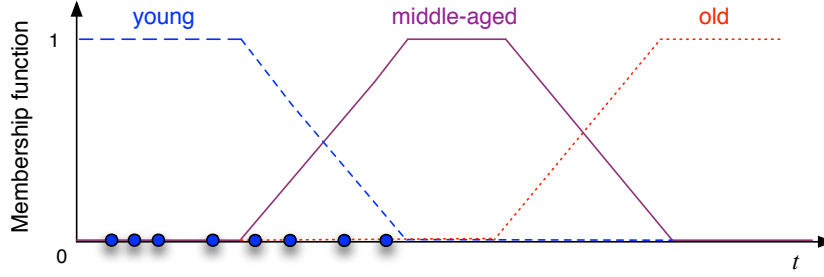


Figure 4.2: Set Students

Definition 4.3 The fuzzy set A is defined as empty if and only if:

$$A = \emptyset = \{x \mid x \in U, \mu_A(x) = 0\}$$

Example 4.3 Figure 4.2 is used to illustrate the definition of an empty fuzzy set. In this figure, people who belongs to a universe of students is presented and it is clear to conclude that: $\mu_{\text{young}} \neq \emptyset$, $\mu_{\text{middle-edge}} \neq \emptyset$ and $\mu_{\text{old}} = \emptyset$.

Definition 4.4 Two fuzzy sets, A and B , are defined as equal if and only if:

$$A = B = \{x \mid x \in U, \mu_A(x) = \mu_B(x)\}$$

Definition 4.5 The fuzzy set A is defined to be contained in another fuzzy set B if and only if:

$$A \subseteq B = \{x \mid x \in U, \mu_A(x) < \mu_B(x)\}$$

Definition 4.6 The complement of fuzzy set A , denoted by \bar{A} , is defined by:

$$\bar{A} = \{x \mid x \in U, \mu_{\bar{A}}(x) = 1 - \mu_A(x)\}$$

Definition 4.7 The fuzzy set A is defined as convex if and only if:

$$\mu_A(\lambda x + (1 - \lambda)y) \leq \min(\mu_A(x), \mu_A(y)), \forall x, y \in U, \forall \lambda \in [0, 1]$$

Example 4.4 Figure 4.3 is used to illustrate the definition of a convex and that of a non-convex fuzzy set.

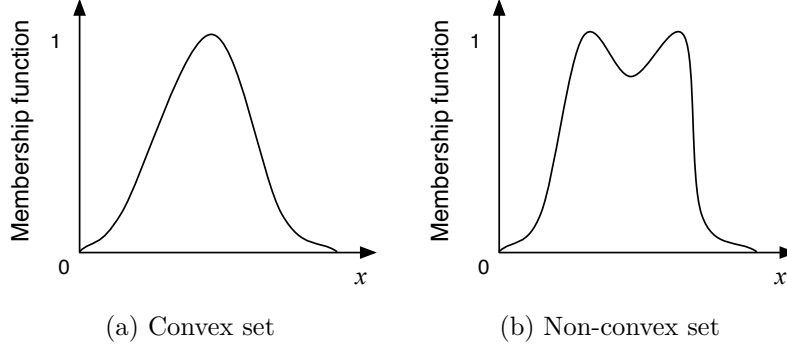


Figure 4.3: Convex and Non-convex Sets

Definition 4.8 The support of fuzzy set A is a crisp subset defined by:

$$Supp(A) = \{x \mid x \in U, \mu_A(x) > 0\}$$

Definition 4.9 The α -cut of fuzzy set A is a crisp subset defined by:

$$A_\alpha = \{x \mid x \in U, \mu_A(x) \leq \alpha, \alpha \in [0, 1]\}$$

Example 4.5 Figure 4.4 is used to illustrate the definition support, α -cut, and kernel of a fuzzy set. In this figure and for simplicity, only the linguistic term *middle-aged* has been taken from the set *age*.

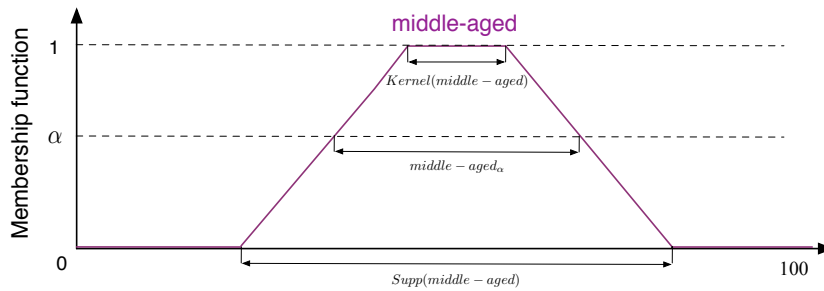
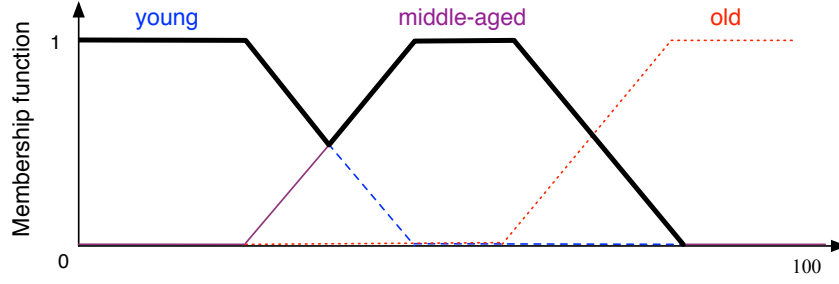


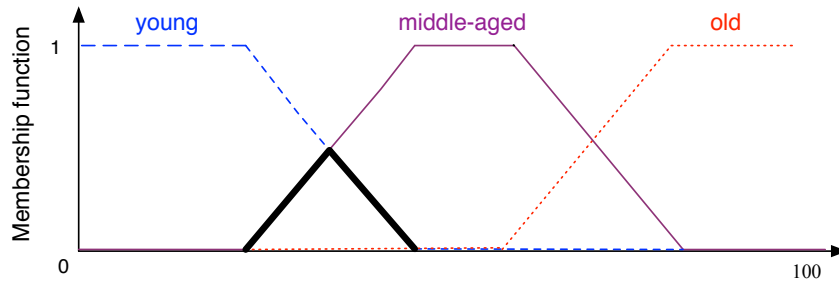
Figure 4.4: Support, α -cut, and Kernel of a Fuzzy Set

Definition 4.10 The kernel of fuzzy set A is a crisp subset defined by:

$$Kernel(A) = \{x \mid x \in U, \mu_A(x) = 1\}$$



(a) $young \cup middle - aged$



(b) $young \cap middle - aged$

Figure 4.5: Union and Intersection of Two Sets

Definition 4.11 *The union of two fuzzy sets, A and B , is another set defined by:*

$$A \cup B = \{x \mid x \in U, \mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x))\}$$

Definition 4.12 *The intersection of two fuzzy sets, A and B , is another set defined by:*

$$A \cap B = \{x \mid x \in U, \mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x))\}$$

Example 4.6 *Figure 4.5 is used to illustrate the definition of the union and intersection of two fuzzy sets.*

4.3 Sharp vs. Fuzzy Clustering

Clustering is an unsupervised learning task that aims to decompose a set of objects into “clusters” based on similarities, where the objects belonging to the same cluster are as similar as possible. Using sharp clustering, each element is associated with just one cluster. Two well known algorithms used to generate clusters are: c-means (sharp clustering) and FCM (fuzzy clustering). The FRS proposed in this Ph.D. thesis uses a modified version of the FCM that is explained in more detail in Sect. 4.3.2.

4.3.1 C-Means Algorithm

The c-means algorithm, also known as the k-means algorithm, was proposed by MacQueen [1967]. It is one of the simplest unsupervised learning algorithms that solve the well known clustering problem that aims to partition n observations into c clusters. Each observation belongs to one and only one cluster. This algorithm aims at minimizing an objective function; in this case, a squared error function defined in Eq. (4.1).

$$J_{cm} = \sum_{i=1}^c \sum_{j=1}^n u_{ij} \|x_j - y_i\|^2 \quad (4.1)$$

where c-means defines a given set of samples $X = \{x_1, \dots, x_n\}$, a set of clusters Γ_j ($i = 1, \dots, c$ and $2 \leq c < n$), and a $c \times n$ binary matrix $\vec{U} = [u_{ij}]$. The membership degree u_{ij} of an observation x_i in a cluster Γ_j is such that: $u_{ij} = \mu_{\Gamma_i}(x_j) \in \{0, 1\}$. The individual elements, $u_{ij} = \mu_{\Gamma_i}(x_j) \in \{0, 1\}$, indicate whether an element belongs to a cluster or not (e.g., $u_{ij} = 1$ if the element x_j is assigned to cluster Γ_i ; i.e., $x_j \in \Gamma_i$; and $u_{ij} = 0$ otherwise). Constraints in Eqs. (4.2) and (4.3) guarantee that clusters are not empty and that the sum of the membership for each x is equal to 1.

$$\sum_{j=1}^n u_{ij} > 0, \forall i \in \{1, \dots, c\} \quad (4.2)$$

$$\sum_{i=1}^c u_{ij} = 1, \forall j \in \{1, \dots, n\} \quad (4.3)$$

The c-mean algorithm is an iterative process that is defined as follows. First, the centers are placed at random (different locations cause different results). At this point each data point is assigned to its closest cluster using Eq. (4.4).

$$u_{ij} = \begin{cases} 1 & \text{if } i = \operatorname{argmin}_{l=1}^c \|x_j - y_l\|^2 \\ 0 & \text{otherwise} \end{cases} \quad (4.4)$$

then, the data partition matrix \vec{U} is held fixed and new cluster centers are computed using Eq. (4.5).

$$y_i = \frac{\sum_{j=1}^n u_{ij}^m x_j}{\sum_{j=1}^n u_{ij}^m} \quad (4.5)$$

Although it can be proved that the procedure will always terminate, the c-means algorithm does not necessarily find the most optimal configuration, corresponding to the global objective function minimum. The algorithm is significantly sensitive to the random nature of its initialization of cluster centers. In order to find the optimal solution, the c-means algorithm can be executed multiple times. A pseudocode of the c-means method is presented in Algorithm 4.1.

Algorithm 4.1 C-Means Algorithm

Input: c

Output: $\vec{U}^{(k+1)}, \vec{Y}^{(k+1)}$

- 1: Set iteration number: $k \leftarrow 0$
 - 2: Generate matrix of cluster centers: $\vec{Y}^{(k)} \leftarrow \text{random}$
 - 3: Compute $\vec{U}^{(k)} \leftarrow \vec{Y}^{(k)}$
 - 4: **repeat**
 - 5: Update $\vec{Y}^{(k+1)} \leftarrow \vec{U}^{(k)}$
 - 6: Update $\vec{U}^{(k+1)} \leftarrow \vec{Y}^{(k+1)}$
 - 7: **until** $\vec{U}^{(k+1)}$ and $\vec{Y}^{(k+1)}$ no longer move
 - 8: **return** $\vec{U}^{(k+1)}, \vec{Y}^{(k+1)}$
-

Example 4.7 Figure 4.6 illustrates the execution of the c-means algorithm. It presents the movement of two cluster centers, y_1 and y_2 , and shows that the centers moved from a starting point to their final position following the path described by the arrows. At the end of the run of the algorithm, the classification is identified by the final boundary.

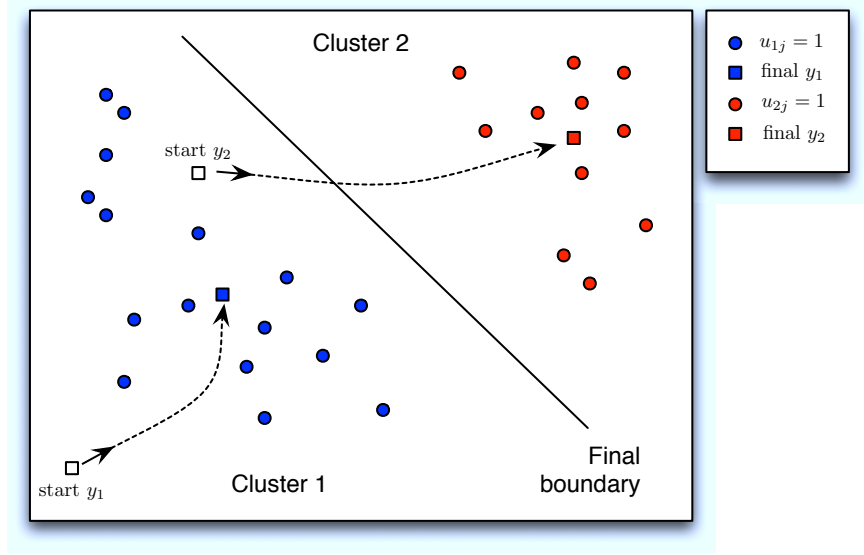


Figure 4.6: C-means Algorithm Execution

As mentioned above, the output of the final position of the centers depends on the initialization of the algorithm. To make sure that we are in the presence of a local optimum (a solution that is not exactly the best), several executions of the algorithm should be made.

4.3.2 Fuzzy C-Means Algorithm

The FCM originally developed by Dunn [1973a] and improved by Bezdek [1981] is an extension of the c-means algorithm that allows for gradual membership of data points in clusters with different degrees of membership according to the fuzzy set theory introduced by Zadeh [1965]. Thus, FCM defines a given set of samples: $X = \{x_1, \dots, x_n\}$, a set of clusters Γ_j ($i = 1, \dots, c$ and $2 \leq c < n$), and a $c \times n$ fuzzy partition matrix $\vec{U} = [u_{ij}]$. The membership degree u_{ij} of an observation x_i in a cluster Γ_j is such that: $u_{ij} = \mu_{\Gamma_i}(x_j) \in [0, 1]$. A probabilistic cluster partition defined by the constraints in Eqs. (4.6) and (4.7) guarantee that clusters are not empty and that the sum of the membership for each x is equal to 1.

$$\sum_{j=1}^n u_{ij} > 0, \forall i \in \{1, \dots, c\} \quad (4.6)$$

$$\sum_{i=1}^c u_{ij} = 1, \forall j \in \{1, \dots, n\} \quad (4.7)$$

Thus, the FCM algorithm is based on the minimization of the objective function shown in Eq. (4.8).

$$J_{fcm} = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m \|x_j - y_i\|^2 \quad (4.8)$$

where x_j is the j -th element of d -dimensional measured data, y_i is the d -dimensional center of cluster i , m is any real number greater than 1 (m determines the level of “fuzziness”; $m = 2$ is a typical value used), and $\|*\|$ is any norm expressing the similarity between any measured data and the center. In Eq. (4.8), $\vec{Y} = [y_i]$ is the matrix of cluster centers ($i = \{1, \dots, c\}$).

The membership function u_{ij} and the center of clusters y_i are computed by taking the derivative of the objective function J_m with respect to the parameters to optimize equal to zero. With the constraint in Eq. (4.7), Eqs. (4.9) and (4.10) can be derived as follows:

$$u_{ij} = \frac{1}{\sum_{l=1}^c \frac{\|x_j - y_i\|^{\frac{2}{m-1}}}{\|x_j - y_l\|^{\frac{2}{m-1}}}} \quad (4.9)$$

$$y_i = \frac{\sum_{j=1}^n u_{ij}^m x_j}{\sum_{j=1}^n u_{ij}^m} \quad (4.10)$$

The FCM algorithm is a two-step iterative process that is defined as follows. First, set the input variables c , m , and ϵ (ϵ is a termination criterion, normally $\epsilon \in [0, 1]$). Second, set an iteration number $k = 0$. Third, randomly generate a matrix of cluster centers $\vec{Y}^{(k)}$. Then, given the initial matrix $\vec{Y}^{(k)}$, compute the fuzzy partition matrix $\vec{U}^{(k)}$.

Finally, using a repeat-until loop, update $\vec{Y}^{(k+1)}$ using $\vec{U}^{(k)}$, and then update $\vec{U}^{(k+1)}$ using $\vec{Y}^{(k+1)}$. Repeat this process until the termination criterion is reached ($|\vec{U}^{(k+1)} - \vec{U}^{(k)}| \leq \epsilon$). To illustrate the use of FCM, a pseudocode is presented in Algorithm 4.2.

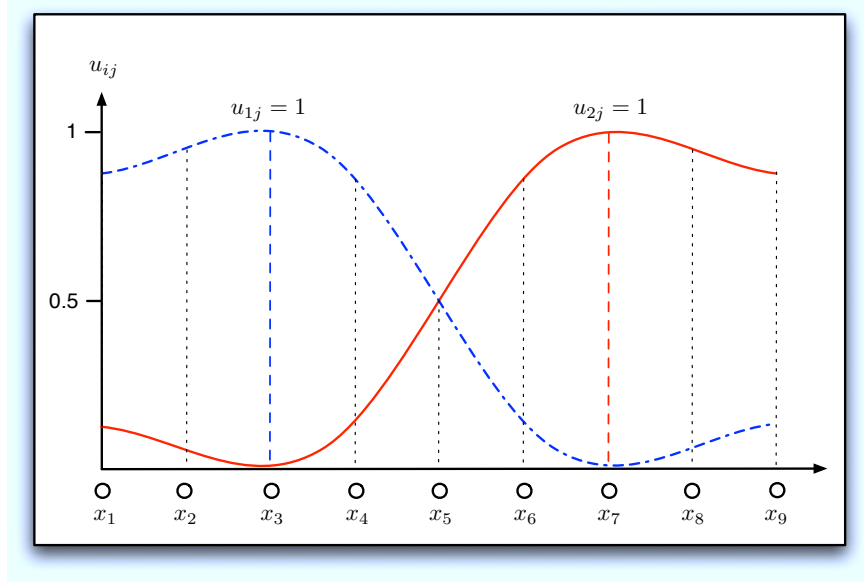


Figure 4.7: FCM Instance

Example 4.8 Figure 4.7 shows an example that illustrates how the FCM algorithm classifies in one dimension a set of nine elements. The x-axis shows the position of the data, and the y-axis is the membership function.

Algorithm 4.2 Fuzzy C-Means Algorithm

Input: c, m, ϵ

Output: $\vec{U}^{(k+1)}, \vec{Y}^{(k+1)}$

- 1: Set iteration number: $k \leftarrow 0$
 - 2: Generate matrix of cluster centers: $\vec{Y}^{(k)} \leftarrow \text{random}$
 - 3: Compute $\vec{U}^{(k)} \leftarrow \vec{Y}^{(k)}$
 - 4: **repeat**
 - 5: Update $\vec{Y}^{(k+1)} \leftarrow \vec{U}^{(k)}$
 - 6: Update $\vec{U}^{(k+1)} \leftarrow \vec{Y}^{(k+1)}$
 - 7: **until** $|\vec{U}^{(k+1)} - \vec{U}^{(k)}| \leq \epsilon$
 - 8: **return** $\vec{U}^{(k+1)}, \vec{Y}^{(k+1)}$
-

In Fig. 4.7, the highest point of each cluster also represents the final positions of the centers once the algorithm has finished its execution. The membership values of the data set are shown in Table 4.1; e.g., element x_5 is located between the two clusters, and the membership of this element is $u_{15} = 0.5$ and $u_{25} = 0.5$.

In the examples show the constraints on Eqs. (4.6) and (4.7) are guaranteed, since there are no empty clusters and the sum of the membership of each element is equal to 1.

Table 4.1: Fuzzy Partition of the Data Set

j -th element	u_{1j}	u_{2j}
1	0.87	0.13
2	0.93	0.7
3	1	0
4	0.81	0.19
5	0.5	0.5
6	0.19	0.81
7	1	0
8	0.7	0.93
9	0.13	0.87

4.4 Remarks on the FCM Algorithm

Although the FCM algorithm is used in different applications, it does have some weaknesses. First, a priori specification of the number of clusters is mandatory. To solve this problem, different techniques have been proposed (Dudoit & Fridlyand [2002]; Sugar & James [2003]; Tibshirani *et al.* [2001]). Secondly, the way to initialize the means was not specified. Given these issues, one popular way to start is to randomly choose c centers.

Finally, lower values of ϵ gives better result, but at the expense of more iterations. The approach proposed in this Ph.D. thesis for advising citizens on elections and the creation of political communities is based on the generation of a fuzzy cluster and uses a modified version of the FCM that is described in more detail in Sect. 5.3.1.

4.5 Further Readings

- **Lam *et al.* [2011].** “This book focuses on computational intelligence techniques and its applications—fast-growing and promising research topics

- that have drawn a great deal of attention from researchers over the years. It brings together many different aspects of the current research on intelligence technologies such as neural networks, support vector machines, fuzzy logic and evolutionary computation, and covers a wide range of applications from pattern recognition and system modeling, to intelligent control problems and biomedical applications” (abs.).
- **Miyamoto *et al.* [2010]**. “The main subject of this book is the FCM proposed by Dunn and Bezdek and their variations including recent studies. A main reason why we concentrate on fuzzy c-means is that most methodology and application studies in fuzzy clustering use fuzzy c-means, and hence fuzzy c-means should be considered to be a major technique of clustering in general, regardless whether one is interested in fuzzy methods or not” (abs.).
 - **Valente de Oliveira & Pedrycz [2007]**. “Fuzzy clustering is now a mature and vibrant area of research with highly innovative advanced applications. Encapsulating this through presenting a careful selection of research contributions, this book addresses timely and relevant concepts and methods, whilst identifying major challenges and recent developments in the area” (abs.).
 - **Nascimento [2005]**. “Development of models with explicit mechanisms for data generation from cluster structures is of major interest in order to provide a theoretical framework for cluster structures found in data. Especially appealing in this regard are the so-called typological structures in which observed entities relate in various degrees to one or several prototypes” (abs.).

Part III

Conceptual Framework

Chapter 5

Fuzzy Recommender System

The *SmartParticipation* project is a Web platform that uses a fuzzy-based approach for the creation of political/thematic groups, assuming that citizens and candidates cannot be considered unique items. In this chapter, the reader will be presented with the architecture and main features of the fuzzy recommender system's engine. Additionally, an evaluation of different dimensionality reduction and fuzzy clustering methods is presented. This chapter is structured as follows: First, Sect. 5.1 gives a brief overview of the systems architecture. Then, Sect. 5.2 describes the method used to generate the so-called fuzzy profiles. Section 5.3 presents the different components of the recommender system's engine used by the *SmartParticipation* system. Afterwards, Sect. 5.4 describes the different visualization methods of the recommendation approach proposed and includes an evaluation of different dimensionality and clustering algorithms. Section 5.5 provides some remarks about the recommender system techniques presented in this chapter and the approach that is used by the *SmartParticipation* system. Finally, further readings are presented in Sect. 5.6.

5.1 General Architecture

The recommendation procedure is divided into three steps. In the first step, the voters (users) and candidates must create their profiles using a fuzzy interface, which is a convenient tool used to determine the level of agreement, disagree-

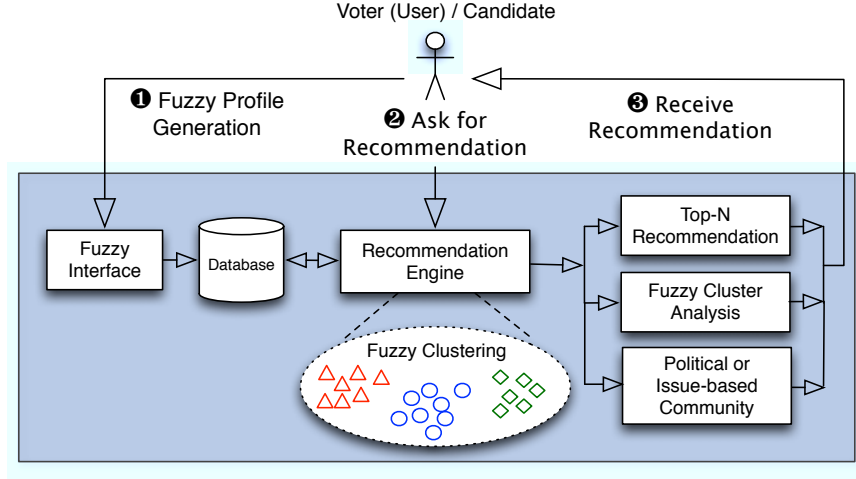


Figure 5.1: Fuzzy-based Recommender System Architecture

ment, and relevance for each specific question. The fuzzy profiles are stored in a database. In the second step, once all necessary profiles have been created, the user selects the recommendation target and the type of output (top-N recommendation, fuzzy cluster analysis, or political community). In the final step, once the recommendation engine has computed all the information, the user receives the recommendation in the pre-established format. The architecture of the fuzzy recommendation approach is presented in Fig. 5.1, and each element is presented in more detail in the following sections.

5.2 User Profile Generation

In order to provide a recommendation, voters (users) and candidates must generate a profile that describes their preferences using a fuzzy interface to complete a questionnaire regarding political issues (each question has different possible responses). The fuzzy interface is a convenient tool used to determine the level of agreement/disagreement and relevance for a specific question.

Unlike other similar tools, this interface provides a higher number of possibilities for each citizen/candidate to answer the questions. The interface is designed to be as intuitive and convenient as possible for users. In addition to the fuzzy interface, the system contains a profile representation, a so-called fuzzy profile

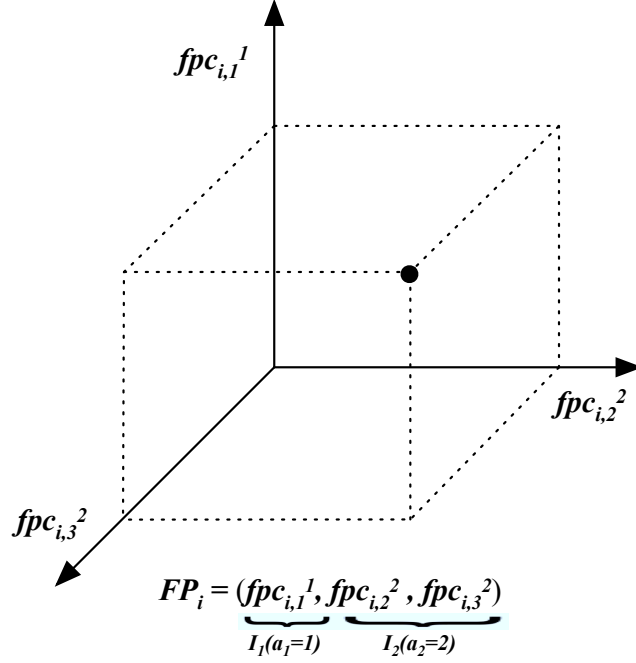


Figure 5.2: Fuzzy Profile (FP)

(FP), which is defined as a point in a multi-dimensional Euclidean space with a total of n elements (dimensions) and is defined as follows:

$$\begin{aligned}
 FP_i = & \left(\underbrace{fpc_{i,1}^1, fpc_{i,2}^1, \dots, fpc_{i,a_1}^1}_{I_1(a_1)}, \underbrace{fpc_{i,a_1+1}^2, \dots, fpc_{i,a_1+a_2}^2}_{I_2(a_2)}, \dots \right. \\
 & \left. \dots, \underbrace{fpc_{i,a_1+a_2+\dots+a_{k-1}+1}^k, \dots, fpc_{i,a_1+a_2+\dots+a_{k-1}+a_k}^k}_{I_k(a_k)} \right)
 \end{aligned}$$

where FP_i is the FP of i -th user. The fpc_{ij}^k is the j -th fuzzy profile component (fpc), and I_l represents a specific issue (with a total number of issues = k). Each issue I_l has a set of a positive number of questions defined by a_l , and the total of questions of all issues is equal to n , such that: $\sum_{l=1}^k a_l = n$.

Example 5.1 Figure 5.2 shows an instance of FP for the i -th user. It has three fpc's and two issues (I_1 and I_2). The I_1 has only one component ($a_1 = 1$), while I_2 has two components ($a_2 = 2$).

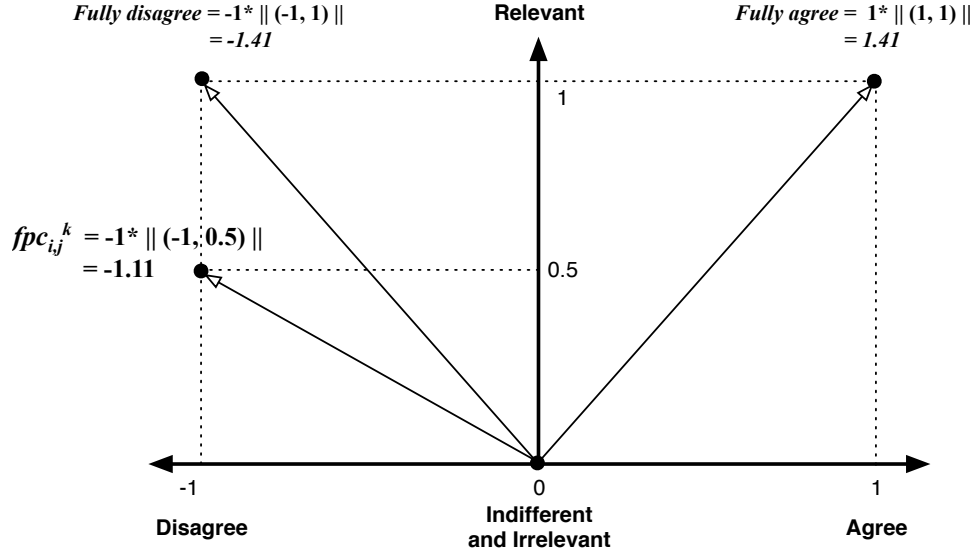


Figure 5.3: Fuzzy Profile Component (fpc)

The fpc has two components: agreement ($agr_{ij}^k \rightarrow [-1, 1]$) and relevance ($rel_{ij}^k \rightarrow [0, 1]$). Each fpc corresponds to the answer of a question that belongs to a specific issue I_l and is defined as follows:

$$fpc_{ij}^k = \alpha(agr_{ij}^k) * ||(agr_{ij}^k, rel_{ij}^k)||$$

and

$$\alpha(agr_{ij}^k) = \begin{cases} 1 & \text{if } agr_{ij}^k \in [0, 1] \\ -1 & \text{if } agr_{ij}^k \in [-1, 0) \end{cases}$$

where agr_{ij}^k is the j -th component of agreement on the k -th issue and rel_{ij}^k is the j -th component of relevance on the k -th issue for the i -th user.

Example 5.2 Figure 5.3 shows an instance of fpc_{ij}^k . It represents the j -th component of the k -th issue of the i -th user, and it is equal to -1.11 .

Additionally, Fig. 5.3 shows the references that are used for specifying full agreement, full disagreement, and indifference/irrelevance. They are used for building communities based on issues, which are discussed in more detail in Sect. 5.4.2.

5.3 Recommendation Engine

The recommendation engine is based on the generation of fuzzy clusters using a modified version of the FCM, which is described in more detail in the following section. Once the profiles are generated, the next step is to ask for a recommendation. At this point, the user selects a particular event and the type of recommendation (top-N recommendation, fuzzy clustering analysis, or political/issue-based community). The request is sent to the recommendation engine, which processes the query. The approach proposed in this Ph.D. thesis focuses on the Sammon mapping method for the visualization of clustering results, which preserves inter-pattern distances. Nevertheless, in Sect. 5.4.3, a non-exhaustive evaluation of dimensionality reduction methods is presented, and the results are discussed.

5.3.1 Fuzzy Cluster Algorithm

Once the profiles are mapped to a low-dimensional space with the Sammon mapping technique, the recommendation engine generates fuzzy clusters by using a modified FCM algorithm, which requires two main inputs: the number of clusters, and a matrix of cluster centers. For this reason, prior knowledge of the dataset is required. The modified FCM algorithm is a two-step iterative process that is defined as follows. First, set the input variables: c (number of clusters is equal to number of political parties or number of issues), m (level of fuzziness), ϵ (termination criterion, normally $\epsilon \in [0, 1]$), and t (type of clustering which can be either “Political Party” or “Issue-based”).

- **First Case.** If *type* is equal to “Political Parties,” the recommendation engine considers the number of clusters to be equal to the number of political parties. In the second step, the algorithm sets an iteration number, $k = 0$. The second input required by the FCM algorithm is the matrix of initial centers, which is generated at random. Consequently, the algorithm may converge to a local minima given the random nature of the algorithm. In order to avoid this problem, the modified version of the FCM algorithm initializes the matrix of cluster centers $\vec{Y}^{(k)}$, taking the mean average of answers from all candidates in the same political party (P_i).

The initialization process is based on two assumptions. First, cluster formation relies on the existence of political parties. Second, members of political parties have similar ideology as mentioned in the definition of political parties from the ACE Project [2012]. Thus, given the initial matrix $\vec{Y}^{(k)}$, compute the fuzzy partition matrix $\vec{U}^{(k)}$. Finally, using a repeat-until loop, update $\vec{Y}^{(k+1)}$ using $\vec{U}^{(k)}$ and then update $\vec{U}^{(k+1)}$ using $\vec{Y}^{(k+1)}$. Repeat this process until the termination criterion is reached ($|\vec{U}^{(k+1)} - \vec{U}^{(k)}| \leq \epsilon$). The termination criterion can also be a predefined number of iterations, or a condition that updates the centers, only if the number of candidates in a political party are the majority in the cluster; otherwise, the center does not update.

- **Second Case.** If *type* is equal to “issue-based,” the algorithm sets the center of all issues for both: full agreement and full disagreement (refer to Fig. 5.3). After that, and given the initial matrix $\vec{Y}^{(k)}$, compute the fuzzy partition matrix $\vec{U}^{(k)}$. The reason it takes only one iteration is that the centers are not expected to move. The modified FCM algorithm is presented in Algorithm 5.1.

The outputs of the modified FCM algorithm are a fuzzy partition matrix $\vec{U}^{(k+1)}$ that contains the membership degree of voters (users) and candidates/citizens with respect to each cluster, and a matrix of cluster centers $\vec{Y}^{(k+1)}$.

5.4 Visualization of the FRS

To provide a graphical representation of results that users can easily analyze, the recommendation engine transforms the high-dimensional space of profiles to a lower-dimensional space (bi-dimensional), which reduces the complexity of data analysis. In the following section, an analysis of five dimensionality reduction methods is presented. Figure 5.4 provides an illustration of dimensionality reduction from a three-dimensional space to a bi-dimensional space. The prototype developed in this work displays, in a bi-dimensional map, the locations of the voter (user) and the candidates (labeled by political parties), the clusters

Algorithm 5.1 Modified Fuzzy C-means Algorithm

Input: c, m, ϵ, t

Output: $\vec{U}^{(k+1)}, \vec{Y}^{(k+1)}$

```

1: if  $t = \text{Political} - \text{Party}$  then
2:   Set iteration number:  $k \leftarrow 0$ 
3:   for  $i = 1$  to  $c$  do
4:      $y_i \leftarrow$  mean average of answers from the  $i$ -th Political Party ( $P_i$ )
5:   end for
6:   Compute  $\vec{U}^{(k)} \leftarrow \vec{Y}^{(k)}$ 
7:   repeat
8:     Update  $\vec{Y}^{(k+1)} \leftarrow \vec{U}^{(k)}$ 
9:     Update  $\vec{U}^{(k+1)} \leftarrow \vec{Y}^{(k+1)}$ 
10:  until  $|\vec{U}^{(k+1)} - \vec{U}^{(k)}| \leq \epsilon$ 
11: else if  $t = \text{Issue} - \text{Based}$  then
12:   for  $i = 1$  to  $c$  do
13:      $y_i \leftarrow$  full agreement on issue  $I_i$ 
14:      $y_{i+1} \leftarrow$  full disagreement on issue  $I_i$ 
15:   end for
16:   Compute  $\vec{U}^{(k)} \leftarrow \vec{Y}^{(k)}$ 
17: end if
18: return  $\vec{U}^{(k+1)}, \vec{Y}^{(k+1)}$ 

```

that are generated according to each political party, and the percentage of the closeness of the voter (user) to each cluster.

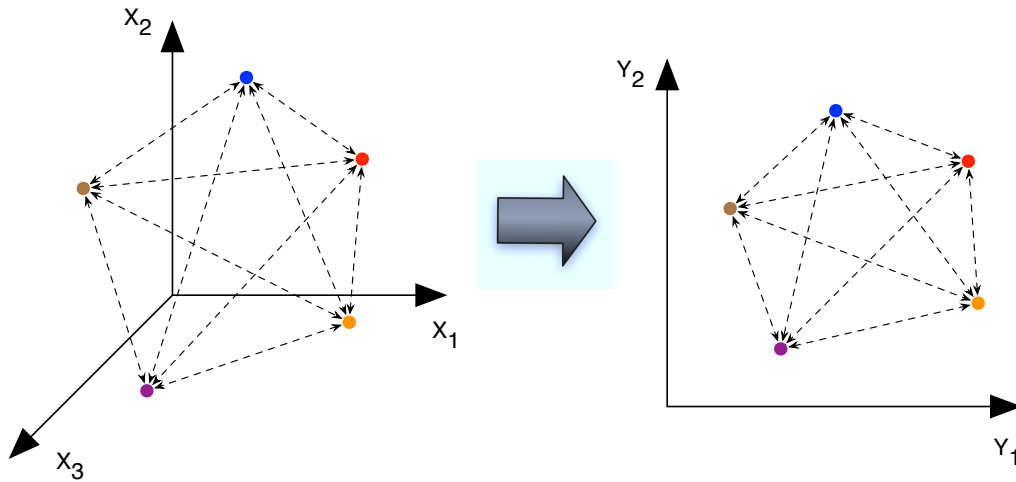


Figure 5.4: Illustration of Dimensionality Reduction

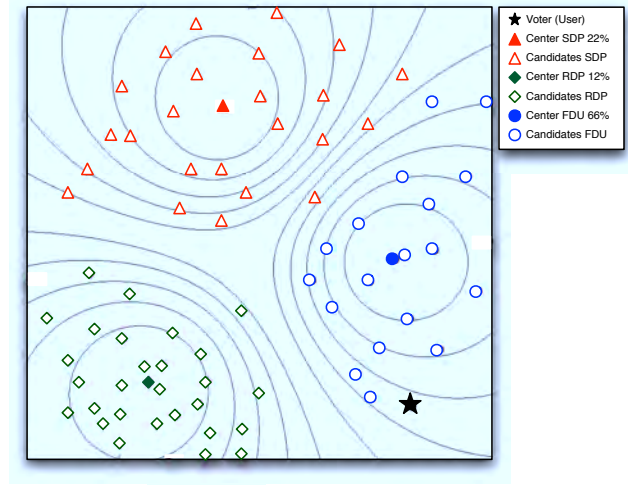


Figure 5.5: Fuzzy Cluster Analysis Graphical Interface

Figure 5.5 displays the formation of clusters with a clear concentration of candidates from the same political party. It shows that the closest political party with respect to the voter (user) is the Federal Democratic Union (66%), followed by the Social Democratic Party (22%) and the Radical Democratic Party (12%).

5.4.1 Top-N Recommendations

The top-N candidates/citizens similar to the preferences of the voter (user) v are generated by using the bi-dimensional profiles. The distances of all candidates, with respect to voter (user) v , are computed, and the N closest candidates are displayed. The similarity percentage ($S_{vc_i}(\%)$) of a voter (user) v and the i -th candidate (c_i) is computed using the most distant candidate or citizen (d_{max}) as a reference. Equation (5.1) presents the computation of similarity percentage.

$$S_{vc_i}(\%) = 100 - \left(\frac{100 * d_{vc_i}}{d_{max}} \right) \quad (5.1)$$

where d_{vc_i} is the distance between voter (user) v and the i -th candidate/citizen. The prototype developed in this Ph.D. thesis displays the location of a voter (user) and candidates (labeled by political parties), the clusters generated according to each political party (with a percentage of closeness of the voter (user) to each cluster), and the N closest candidates labeled with the percentage of proximity

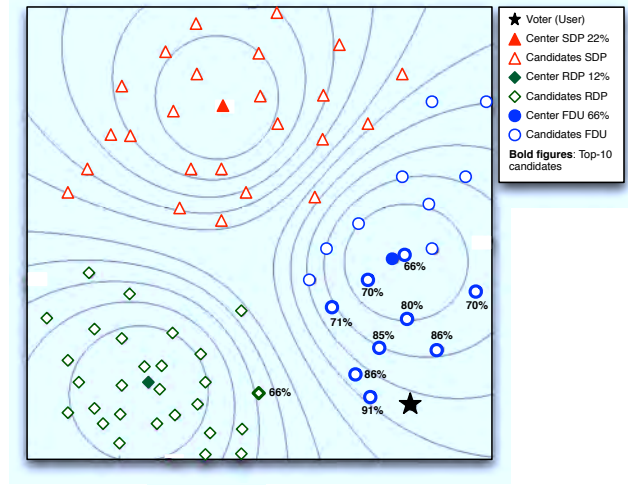


Figure 5.6: Top-N Recommendation Graphical Interface

to the voter (user). Figure 5.6 shows the results with the same dataset used in Fig. 5.5. It shows the formation of clusters by political party and the top 10 candidates close to the voter (user), together with the similarity percentages.

Example 5.3 Figure 5.7 illustrates the computation of top- N recommendations, taking three candidates (c_1 , c_2 , and c_3). It is clear that $d_{max} = d_{vc_3}$, so the similarity percentage between v and c_1 is given as follows:

$$S_{vc_1}(\%) = 100 - \left(\frac{100 * d_{vc_1}}{d_{max}} \right) = 100 - \left(\frac{100 * 1}{3} \right) = 66.66\%$$

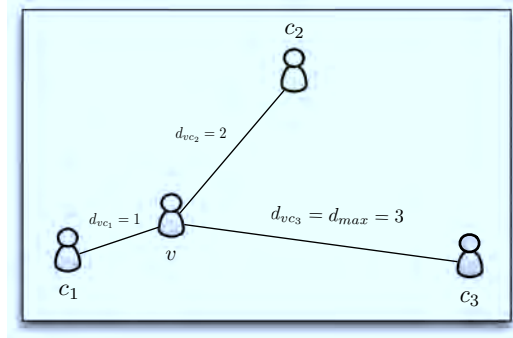


Figure 5.7: Similarity Percentage Illustration

5.4.2 Community-Building Process

The recommendation engine presented in this work can also be used during eDiscussion and ePosting, which is presented in the work of Meier [2012], with the creation of virtual communities, allowing citizens to interact through specific media and potentially crossing geographical and political boundaries in order to pursue mutual interests or goals. The use of the user-friendly, bi-dimensional interfaces can help voters (users) to establish which citizens are the most similar according to their preferences and tendencies (profiles). To create political communities, the recommendation engine uses the datasets of citizens together with the datasets of candidates. The prototype developed has two types of references for the centers of clusters: political parties and issue-based.

First Case: In the case of using political parties as references, the recommendation engine transforms the high-dimensional profiles into a bi-dimensional space. Secondly, in order to compute the fuzzy clusters, only the bi-dimensional profiles of candidates and the voter (user) looking for the recommendation are used. Once the fuzzy clusters are computed, the datasets are merged (voter, citizens, and candidates) and displayed in a bi-dimensional map. In Fig. 5.8a, not only the candidates but also the citizens involved in the system are included in the bi-dimensional map. The citizens are represented by black squares, and for this experiment, the 20 closest citizens are represented by filled black squares.

Second Case: In the case of using issues as references, the recommendation engine transforms the high-dimensional profiles into a bi-dimensional space. Secondly, in order to compute the fuzzy clusters, only the bi-dimensional references of issues are taken for both full agreement and full disagreement. Once the fuzzy clusters are computed, the datasets are displayed (voter and citizens) in a bi-dimensional map. Figure 5.8b shows how close the user is with respect to issues 3 and 5 (the issues are taken from the *smartvote* dataset). It shows that the selected user is 45% to the full agreement on issue 3 and 38% to full disagreement on issue 5. The citizens are represented by black squares, and for this experiment the 20 closest citizens are represented by filled black squares.

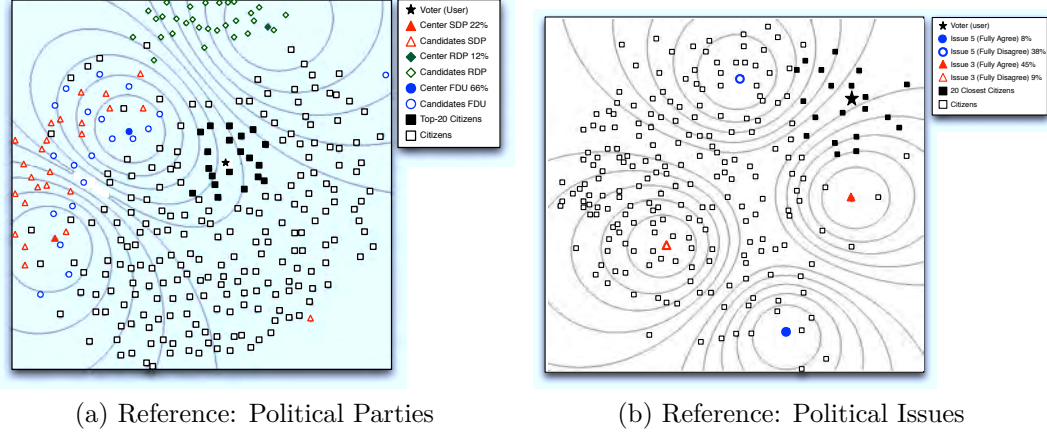


Figure 5.8: Political Communities Graphical Interface

5.4.3 Evaluation of Dimensionality Reduction Methods

Dimensionality reduction is being used in different research areas, such as image processing, multivariable data analysis, machine learning, and data mining, among others. The understanding of high-dimensional data requires the extraction of information from it. The FRS is based on dimensionality reduction to provide users with a visualization of political landscapes. For this reason, an evaluation of different methods is presented in this section. In the work of Van der Maaten *et al.* [2009], a taxonomy of different techniques for dimensionality reduction is presented. A subdivision is made into convex and non-convex techniques, and is presented in Fig. 5.14.

Evaluation Measures for Dimensionality Reduction

In the academic literature, several measures for evaluating non-linear dimensionality methods have been proposed. In this section, a methodology presented in the work of Venna & Kaski [2001, 2006] is used to evaluate five dimensionality reduction algorithms. It considers two measures: “trustworthiness” and “continuity.” Trustworthiness is defined in Eq. (5.2).

$$T(k) = 1 - \frac{2}{nk(2n - 3k - 1)} \sum_{i=1}^n \sum_{j \in U_i^{(k)}} r(i, j) - k \quad (5.2)$$

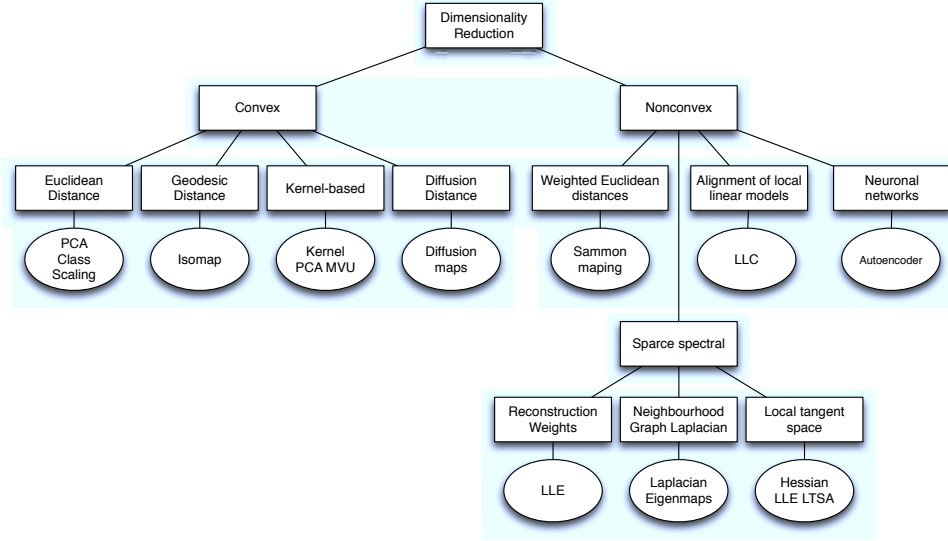


Figure 5.9: Taxonomy of Dimensionality Reduction Techniques. Adapted from Van der Maaten *et al.* [2009]

where $r(i, j)$ is the rank of low-dimensional data point j with respect to the pairwise distances between the low-dimensional set of points. The variable $U_i^{(k)}$ represents the set of k closest neighbors in the low-dimensional space, but not in the high-dimensional space. The second measure for evaluation of dimensionality reduction is continuity. It is defined in Eq. (5.3).

$$C(k) = 1 - \frac{2}{nk(2n - 3k - 1)} \sum_{i=1}^n \sum_{j \in V_i^{(k)}} \hat{r}(i, j) - k \quad (5.3)$$

where $\hat{r}(i, j)$ is the rank of high-dimensional data point j with respect to the pairwise distances between the high-dimensional set of points. The variable $V_i^{(k)}$ represents the set of k closest neighbors in the high-dimensional space, but not in the low-dimensional space.

Dimensionality Reduction Methods

The FRS proposed in this Ph.D thesis makes use of dimensionality reduction to better understand the inter-distance relations in a recommendation. A non-exhaustive analysis of dimensionality reduction methods is presented. In this section, five dimensionality reduction methods are analyzed and compared: (1) Prin-

Principal Component Analysis (PCA; Pearson [1901]), (2) Sammon mapping (Sammon [1969]), (3) t-Distributed Stochastic Neighbor Embedding (t-SNE; van der Maaten & Hinton [2008]), (4) Isomap (Tenenbaum *et al.* [2000]), and (5) Locally Linear Embedding (LLE; Roweis & Saul [2000]). To visualize the results provided by each fuzzy clustering method, two datasets provided by the *smartvote* project (smartvote [2012b]) are used. The first dataset corresponds to candidates of three political parties (83 candidates in total), and the second dataset corresponds to the candidates of 15 political parties (257 candidates in total). Both datasets are composed of profiles with 73 dimensions (questions).

Principal Component Analysis (PCA). A technique introduced by Pearson [1901], involves a mathematical procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called principal components. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. The main objectives of PCA are to identify new meaningful underlying variables and to discover or reduce the dimensionality of the data set.

The mathematical background lies in eigenanalysis. The eigenvector associated with the largest eigenvalue has the same direction as the first principal component. The eigenvector associated with the second largest eigenvalue determines the direction of the second principal component. In this thesis, the second objective of PCA is used. In that case, the covariance matrix of the data set, which is also called the “data dispersion matrix,” is defined as in Eq. (5.4) as follows:

$$F = \frac{1}{N}(x_k - \bar{x}_k)(x_k - \bar{x}_k)^T \quad (5.4)$$

where \bar{x}_k is the mean of the data set. N is equal to the number of objects in the data set. The PCA technique is based on the projection of correlated high-dimensional data onto a hyper-plane. This mapping uses only the first few q non-zero eigenvalues, and the corresponding eigenvectors of the covariant matrix are defined in Eq. (5.5) as:

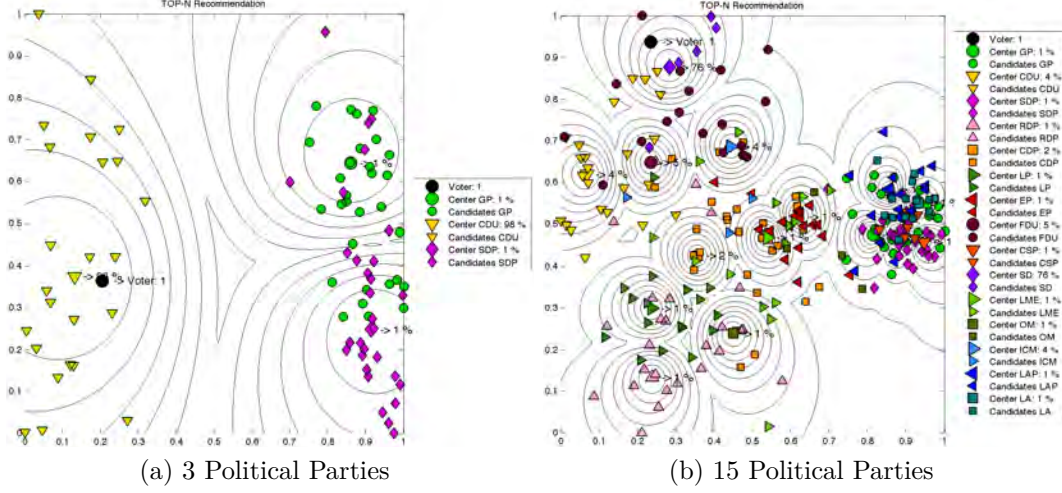


Figure 5.10: Dimensionality Reduction by PCA for 3 and 15 Political Parties

$$F_i = U_i \Lambda_i U_i^T \quad (5.5)$$

The covariant matrix F_i is decomposed to the matrix Λ_i , which includes the eigenvalues $\lambda_{i,j}$ of F_i in its diagonal in decreasing order, and to the matrix U_i , which includes the eigenvectors that correspond to the eigenvalues in its columns. The vector $y_{i,k} = W_i^{-1}(x_k) = W_i^T(x_k)$ is a q -dimensional reduced representation of the observed vector x_k , where the weight matrix W_i contains the q principal orthonormal axes in its column as shown in Eq. (5.6) as follows:

$$W_i = U_{i,q} \Lambda_{i,q}^{\frac{1}{2}} \quad (5.6)$$

Example 5.4 Figure 5.10 shows the output of the FRS using the PCA method for dimensionality reduction. Two datasets provided by the smartvote project (smartvote [2012b]) are used in this example. Figure 5.10a corresponds to the visualization of candidates of three political parties (83 candidates in total), and Fig. 5.10b corresponds to the visualization of candidates of 15 political parties (257 candidates in total). Both datasets are composed of profiles with 73 dimensions (questions).

Figure 5.10 shows that, due to the nonparametric nature of the PCA, only one output is presented for each dataset (3 and 15 political parties). Section 5.4.3 presents an evaluation of different dimensionality reduction methods.

Sammon Mapping. Sammon mapping is a well-known technique that transforms a high-dimensional space (n-dimensions) into a space with lower dimensionality (q-dimensions), finding N points in the lower dimensional space. Denoting the distances between two different points x_i and x_j ($i \neq j$) in the original space as d_{ij} , and the distance between points y_i and y_j in the mapped space as d'_{ij} , the mapping then becomes a problem of minimizing Sammon's stress E , as defined in Eq. (5.7):

$$E = \frac{1}{\lambda} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{(d_{ij} - d'_{ij})^2}{d_{ij}} \quad (5.7)$$

where $\lambda = \sum_{i=1}^{N-1} \sum_{j=i+1}^N d_{ij}$. In order to minimize E , the Sammon method applies a steepest descent technique in which the new y_{i_l} at iteration $t + 1$ is given by

$$y_{i_l}(t+1) = y_{i_l}(t) - \alpha \left[\frac{\frac{\partial E(t)}{\partial y_{i_l}(t)}}{\frac{\partial^2 E(t)}{\partial^2 y_{i_l}(t)}} \right] \quad (5.8)$$

where $y_{i_l}(t)$ is the l -th coordinate of point y_i in the mapped space and α is a constant that has been computed empirically to be $\alpha \approx 0.3$ or 0.4 . The partial derivatives in Eq. (5.8) are given by:

$$\frac{\partial E(t)}{\partial y_{i_l}(t)} = -\frac{2}{\lambda} \sum_{k=1, k \neq i}^N \left[\frac{d_{ki} - d'_{ki}}{d_{ki} d'_{ki}} \right] (y_{i_l} - y_{k_l})$$

$$\frac{\partial^2 E(t)}{\partial^2 y_{i_l}(t)} = -\frac{2}{\lambda} \sum_{k=1, k \neq i}^N \frac{1}{d_{ki} d'_{ki}} \cdot \left[(d_{ki} - d'_{ki}) - \left(\frac{(y_{i_l} - y_{k_l})^2}{d'_{ki}} \right) \left(1 + \frac{d_{ki} - d'_{ki}}{d_{ki}} \right) \right]$$

A pseudocode of the Sammon mapping method for dimensionality reduction is presented in Algorithm 5.2.

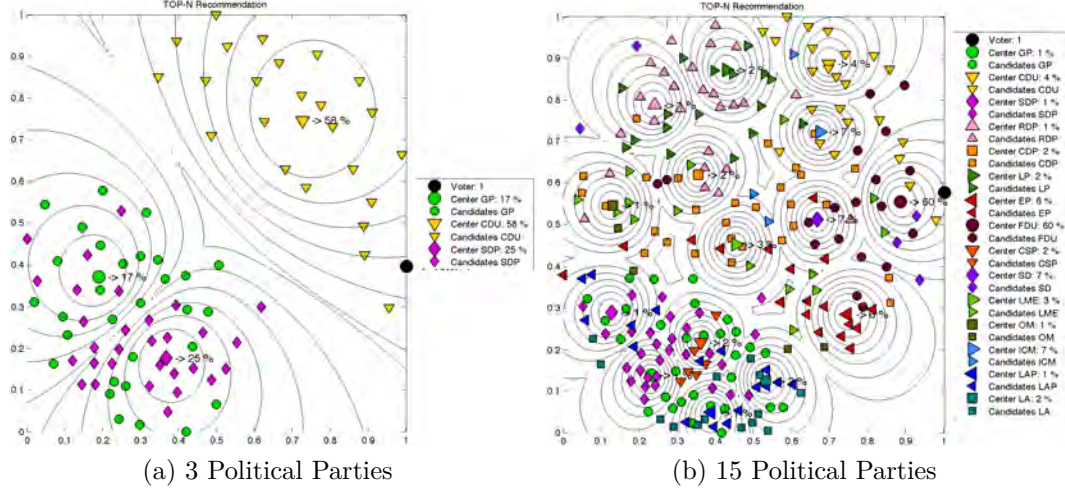


Figure 5.11: Dimensionality Reduction by Sammon Mapping for 3 and 15 Political Parties

Algorithm 5.2 Sammon Mapping Algorithm

Input: data set $X = \{x_1, x_1, \dots, x_n\}$, k-nearest neighbors k , lower dimension d

Output: low-dimensional data representation $Y = \{y_1, y_1, \dots, y_n\}$

- 1: compute all pairwise distances $d_y(i, j)$ in the d -dimensional data space.
 - 2: initialize the p -dimensional coordinates of all points $y(i)$, either randomly or with the p principal components (PCA or MDS)
 - 3: **repeat**
 - 4: compute the right-hand side of Eq. (5.8) for all points of $y(i)$
 - 5: update the co-ordinates of all points of $y(i)$
 - 6: **until** return until the value of the stress function no longer decreases
 - 7: **return** $Y^{(T)} = \{y_1, y_1, \dots, y_n\}$
-

Example 5.5 Figure 5.11 shows the output of the FRS using the Sammon mapping method for dimensionality reduction. Two datasets provided by the smartvote project (smartvote [2012b]) are used in this example. Figure 5.11a corresponds to the visualization of candidates of three political parties (83 candidates in total), and Figure 5.11b corresponds to the visualization of candidates of 15 political parties (257 candidates in total). Both datasets are composed of profiles with 73 dimensions (questions).

Three problems that must be taken into account when the Sammon mapping technique is used are as follows. First, the prototypes of clusters are usually not

known a priori. They are generally calculated while partitioning of the data. These prototypes can be vectors that are dimensionality equal to the examined data points, but they can also be defined as geometrical objects (i.e., linear or non-linear sub-spaces or functions). Sammon mapping is a projection method that is based on the preservation of the Euclidean inter-point distance norm, so it can only be used by clustering algorithms that are calculated with this type of distance norm. As mentioned in Sect. 5.2, FPs are defined to be a multi-dimensional Euclidean spaces, which fulfills the required condition of the Sammon mapping technique. Secondly, the Sammon mapping algorithm forces to find, in a high n -dimensional space, N points in a lower q -dimensional subspace, such that these inter-point distances correspond to the distances measured in the n -dimensional space. This causes a computationally expensive algorithm, since every iteration step requires the computation of $N(N - 1)/2$ distances.

Finally, this gradient-descent method has the possibility of reaching a local minimum in the error surface, while searching for the minimum of E , so experiments with different random initializations are necessary. In order to avoid this problem, the initialization is estimated using the PCA technique introduced by Pearson [1901], which maps the data points into a lower dimensional space. As mentioned previously, Sammon mapping has a disadvantage due to the random selection of the initial projection. In order to avoid this issue, PCA has been selected instead. This initial condition, and the nonparametric nature of Sammon mapping, provide a unique output for each dataset (3 and 15 political parties). For these reasons, no further evaluation is made for Sammon mapping. In Sect. 5.4.3, an evaluation of different dimensionality reduction methods is presented.

t-Distributed Stochastic Neighbor Embedding (t-SNE). t-SNE is a technique for dimensionality reduction introduced in the work of van der Maaten & Hinton [2008]. It is capable of capturing local structure, as well as revealing the presence of clusters, from the high-dimensional space to a low-dimensional space. t-SNE is an extended version of the Stochastic Neighbor Embedding (SNE) method present in the work of Hinton & Roweis [2002], and aims to tackle the problems of SNE, which are cost function optimization and the so-called crowd-

ing problem. In this section, a brief description of the algorithm used by t-SNE is presented. The cost function used by t-SNE uses a symmetric version of SNE with simpler gradients, and a Student-t distribution rather than Gaussian to compute similarity in the low-dimensional space. SNE converts high-dimensional Euclidean distances among data points into a representation of conditional probabilities $p_{j|i}$ of data point x_j to data point x_i , that x_i would pick x_j as a neighbor if neighbors were picked in relation to their probability density, and σ_i , which is the variance of the Gaussian that is centered on data point x_i . The conditional probability is given in Eq. (5.9).

$$p_{j|i} = \frac{\exp(\frac{\|x_i - x_j\|}{2\sigma_i^2})}{\sum_{k \neq i} \exp(\frac{\|x_i - x_k\|}{2\sigma_i^2})} \quad (5.9)$$

In the low-dimensional space, t-SNE converts distances into probabilities with a Gaussian distribution. The result mapping in the low-dimensional space, t-SNE can use a different distribution allowing moderate distance in the high-dimensional space to be modeled by a larger distance in the mapped space, eliminating unattractive mapped forces representing dissimilar data points. t-SNE employs a Student t-distribution with one degree of freedom (same as Cauchy distribution) as a distribution in the low-dimensional space. The joint probability is defined in Eq. (5.10).

$$p_{j|i} = \frac{\exp(\frac{\|x_i - x_j\|}{2\sigma_i^2})}{\sum_{k \neq i} \exp(\frac{\|x_i - x_k\|}{2\sigma_i^2})} \quad (5.10)$$

Finally, a gradient of the Kullback-Leibler divergence between P and the Student-t based joint distribution Q is given in Eq. (5.11).

$$\frac{\delta C}{\delta y_i} = 4 \sum_j (p_{ij} - q_{ij})(y_i - y_j)(1 + \|y_i - y_j\|^2)^{-1} \quad (5.11)$$

A pseudocode of the t-SNE method for dimensionality reduction is presented in Algorithm 5.3. In the examples presented in this section, the following parameters are used: initial dimensionality reduction using PCA to 30 dimensions to speed up the computation of pairwise distances, and perplexity $perp = 30$.

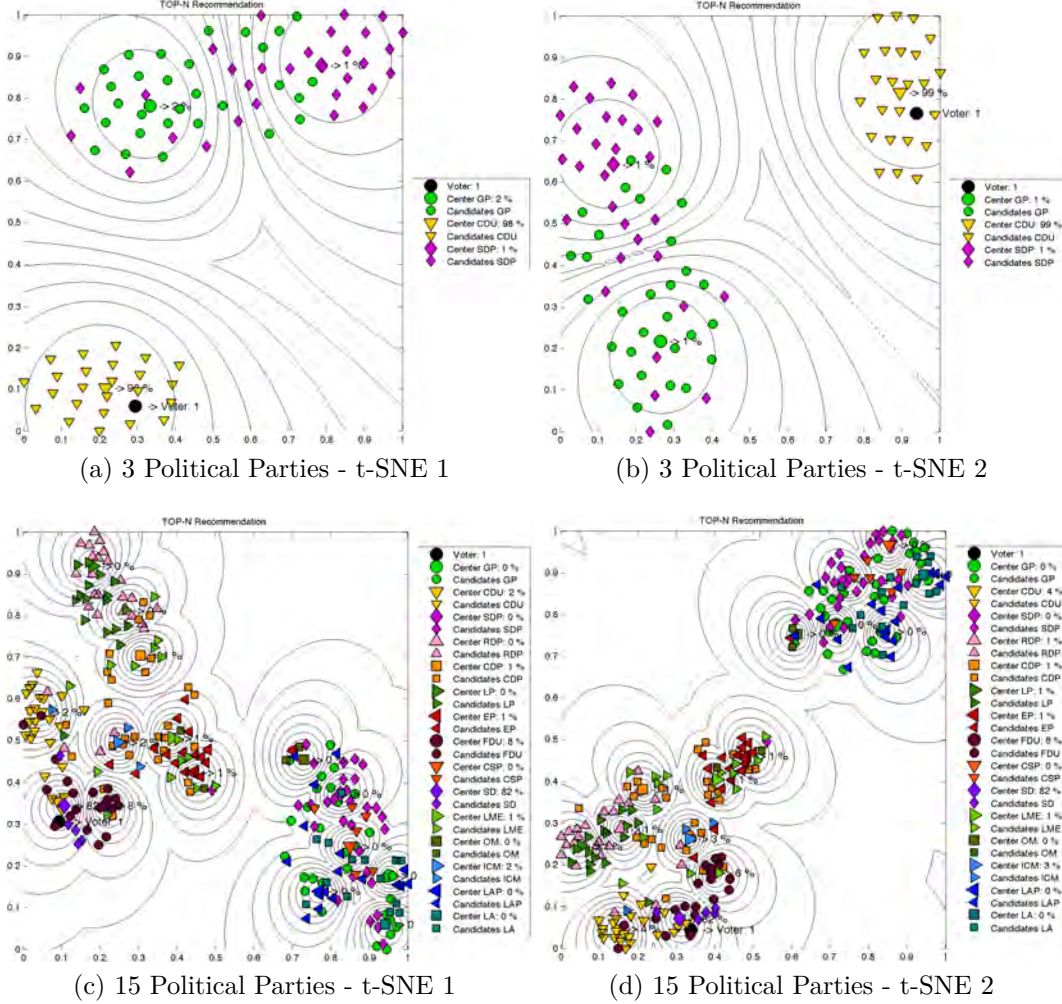


Figure 5.12: Dimensionality Reduction by t-SNE for 3 and 15 Political Parties

Example 5.6 Figure 5.12 shows the output of the FRS using the t-SNE method for dimensionality reduction. Two datasets provided by the smartvote project (smartvote [2012b]) are used in this example. Figures 5.12a and 5.12b correspond to the visualization of candidates of three political parties (83 candidates in total), and Figs. 5.12c and 5.12d correspond to the visualization of candidates of 15 political parties (257 candidates in total). Both datasets are composed of profiles with 73 dimensions (questions).

Figure 5.12 presents the execution of t-SNE with two different datasets. Figures 5.12a and 5.12b present two different executions of the algorithm taking

Algorithm 5.3 t-Distributed Stochastic Neighbor Embedding Algorithm

Input: data set $X = \{x_1, x_1, \dots, x_n\}$, perplexity $perp$, optimization parameters: number of iterations T , learning rate η , momentum $\alpha(t)$

Output: low-dimensional data representation $Y^{(T)} = \{y_1, y_1, \dots, y_n\}$

1: compute pairwise affinities $p_{j|i}$ with perplexity $perp$ (using Eq. (5.9))

2: set $p_{ij} = \frac{p_{j|i} + p_{i|j}}{2n}$

3: sample initial solution $Y^{(0)} = \{y_1, y_1, \dots, y_n\}$ from $N(0, 10^{-4}I)$

4: **for** $i = 1$ to T **do**

5: compute low-dimensional affinities q_{ij} (using Eq. (5.10))

6: compute gradient $\frac{\delta C}{\delta Y}$ (using Eq. (5.11))

7: set $Y^{(t)} = Y^{(t-1)} + \eta \frac{\delta C}{\delta Y} + \alpha(t)(Y^{(t-1)} - Y^{(t-2)})$

8: **end for**

9: **return** $Y^{(T)} = \{y_1, y_1, \dots, y_n\}$

three political parties. It shows that due to the random initial conditions of the algorithm (line 3, Algorithm 5.3) the results presented are different. The same results are presented in Figs. 5.12c and 5.12d, which include 15 political parties. In order to perform an evaluation of t-SNE and to compare it with different dimensionality reduction methods (refer to Sect. 5.4.3), different executions of the algorithm are used and are presented in Fig. 5.13, which shows that different initial conditions produce different performances in terms of trustworthiness and continuity. The evaluation of multiple executions of t-SNE, with two datasets, are presented and commented on. Figures 5.13a and 5.13b present the execution of t-SNE for 3 political parties and show that the second execution of t-SNE has a better performance for a higher number of neighbors. On the other hand, Figs. 5.13c and 5.13d show that different executions of t-SNE present similar performance for a higher number of political parties (bigger dataset).

Isomap. Isomap is a method of dimensionality reduction presented in the work of Tenenbaum *et al.* [2000] and is considered the simplest nonlinear dimensionality reduction. It uses graph distances as an approximation of geodesic distance. In order to illustrate the concept of geodesic distances, Fig. 5.14 shows an adapted version presented in the work of Lee & Verleysen [2007]. The intuition is that the manifold presented in the bi-dimensional space is unrolled and made into a straight line. In contrast with Euclidean distance, geodesic distances are measures

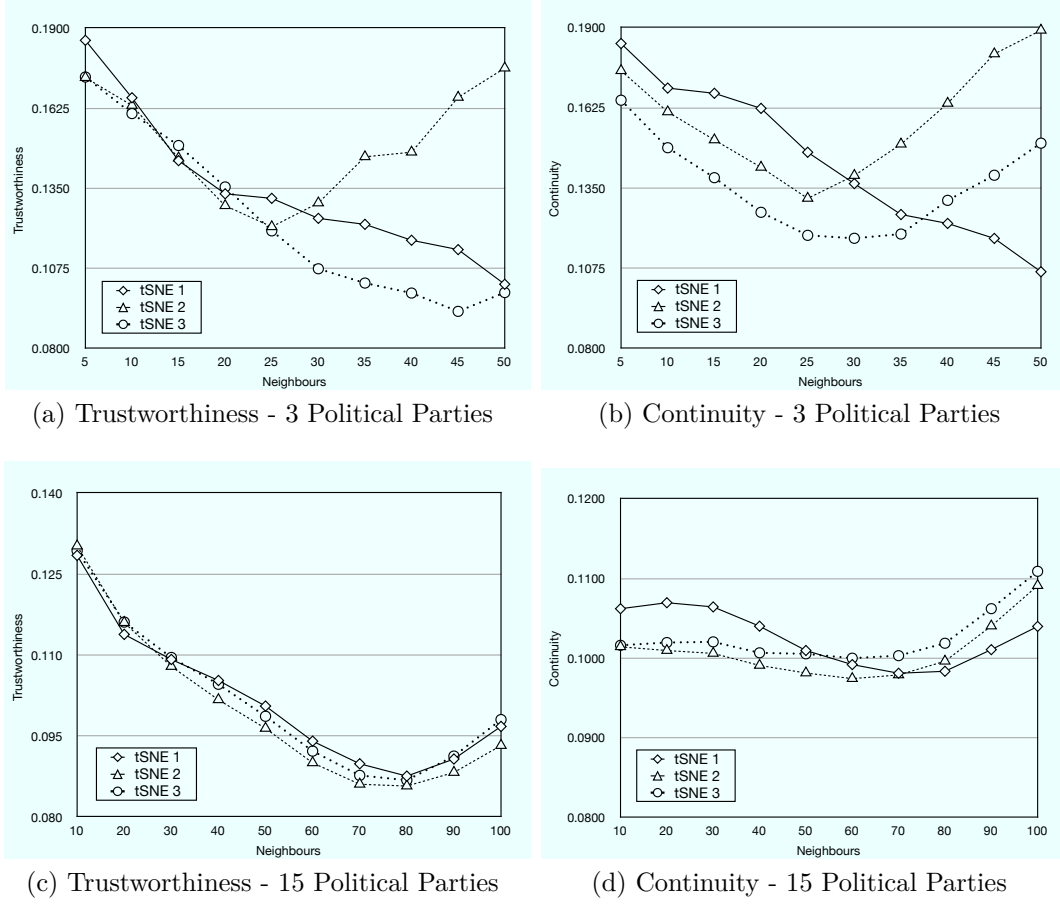


Figure 5.13: Evaluation of t-SNE for Different Initial Conditions

along the manifold. Isomap attempts to preserve pairwise geodesic distances between data points. The geodesic distances are computed by constructing a neighbor graph G , with each data point connected to the k nearest neighbors x_{i_j} with $j = 1, 2, \dots, k$ or within a fixed radius ϵ . Isomap replaces the Euclidean distances with the graph distances computing a \mathbf{D} matrix. The shortest paths are computed using Dijkstra's (Dijkstra [1959]) or Floyd's algorithm (Floyd [1962]). Isomap then applies multidimensional scaling (MDS; Cox & Cox [2008]) to the geodesic distances to find a low-dimensional mapping and is briefly described as follows. Given a $n \times n$ matrix \mathbf{D} of distances, MDS attempts to find n data points y_1, \dots, y_n in d dimensions such that, if \hat{d}_{ij} is the Euclidean distance between y_j and y_i , then $\hat{\mathbf{D}}$ is similar to \mathbf{D} . Metric MDS minimizes the following function:

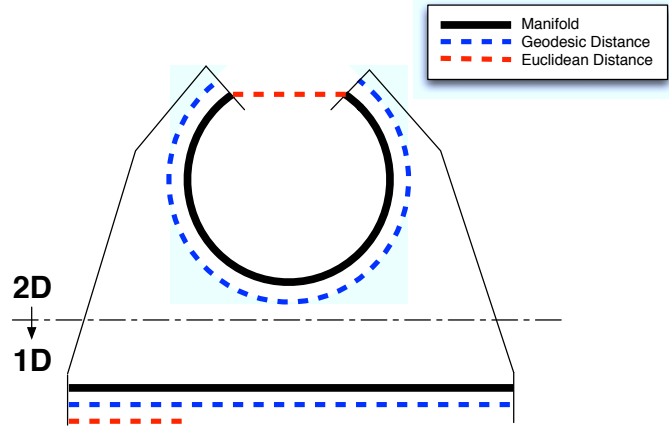


Figure 5.14: Comparison of Geodesic Distance and Euclidean Distance. Adapted from Lee & Verleysen [2007]

$$\min_Y \sum_{i=1}^n \sum_{j=1}^n (d_{ij}^{(X)} - d_{ij}^{(Y)})^2 \quad (5.12)$$

where $d_{ij}^{(X)} = \|x_i - x_j\|^2$ and $d_{ij}^{(Y)} = \|y_i - y_j\|^2$. Then, matrix $\mathbf{D}^{(X)}$ can be interpreted as:

$$\mathbf{X}^T \mathbf{X} = -\frac{1}{2} \mathbf{H} \mathbf{D}^{(X)} \mathbf{H} \quad (5.13)$$

where $\mathbf{H} = \mathbf{I} - \frac{1}{n} \mathbf{e} \mathbf{e}^T$ and \mathbf{e} is a column vector of all 1s. Equation (5.12) is reduced to:

$$\min_Y \sum_{i=1}^n \sum_{j=1}^n (x_i^T x_j - y_i^T y_j)^2 \quad (5.14)$$

Finally, the output of MDS is $Y = \Lambda^{1/2} \mathbf{V}^T$, where V is the eigenvectors of $\mathbf{X}^T \mathbf{X}$ corresponding to the d eigenvalues. Λ is the top d eigenvalues of $\mathbf{X}^T \mathbf{X}$. A pseudocode of the Isomap method is presented in Algorithm 5.4.

In the examples presented in this section, two values of k (closest neighbors) are used for each dataset. The first dataset (83 candidates) was tested with $k = 60$ and $k = 80$, and the second dataset (257 candidates) was tested with $k = 60$ and $k = 240$.

Algorithm 5.4 Isomap Algorithm

Input: data set $X = \{x_1, x_1, \dots, x_n\}$, k-nearest neighbors k

Output: low-dimensional data representation $Y = \{y_1, y_1, \dots, y_n\}$

- 1: build a graph with the k -rule or ϵ -rule
 - 2: weight the graph by labeling each edge with its Euclidean length.
 - 3: compute geodesic pairwise distances using and storing them in matrix \mathbf{D}
 - 4: convert matrix \mathbf{D} into a Gram matrix \mathbf{S} by double centering
 - 5: with Gram matrix \mathbf{S} , compute spectral composition $\mathbf{S} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$
 - 6: a p-dimensional representation of Y is obtained by computing the product
 $\hat{\mathbf{X}} = \mathbf{I}_{P \times N} \mathbf{\Lambda}^{1/2} \mathbf{U}^T$
 - 7: **return** $Y = \{y_1, y_1, \dots, y_n\}$
-

Example 5.7 *Figure 5.16 shows the output of the FRS using the Isomap method for dimensionality reduction. Two datasets provided by the smartvote project (smartvote [2012b]) are used in this example. Figures 5.15a and 5.15b corresponds to the visualization of all candidates of three political parties (83 candidates in total), and Figs. 5.15c and 5.15d corresponds to all candidates of 15 political parties (257 candidates in total). Both datasets are composed of profiles with 73 dimensions (questions).*

Figure 5.16 presents the execution of Isomap with two different datasets. Figures 5.12a and 5.12b present two different executions of the algorithm taking three political parties using $k = 60$ and $k = 80$, respectively. The same results are presented in Figs. 5.12c and 5.12d, which include 15 political parties using $k = 60$ and $k = 240$, respectively. In order to perform an evaluation of Isomap and compare it with different dimensionality reduction methods (refer to Sect. 5.4.3), three different values of k are used. Each execution presented different performances in terms of trustworthiness and continuity. Figures 5.16a and 5.16b present the execution of Isomap for 3 political parties. It is clear that both trustworthiness and continuity perform better for $k = 60$. Similar results are presented in Figs. 5.16c and 5.16d with $k = 60$ for a higher number of political parties (bigger dataset).

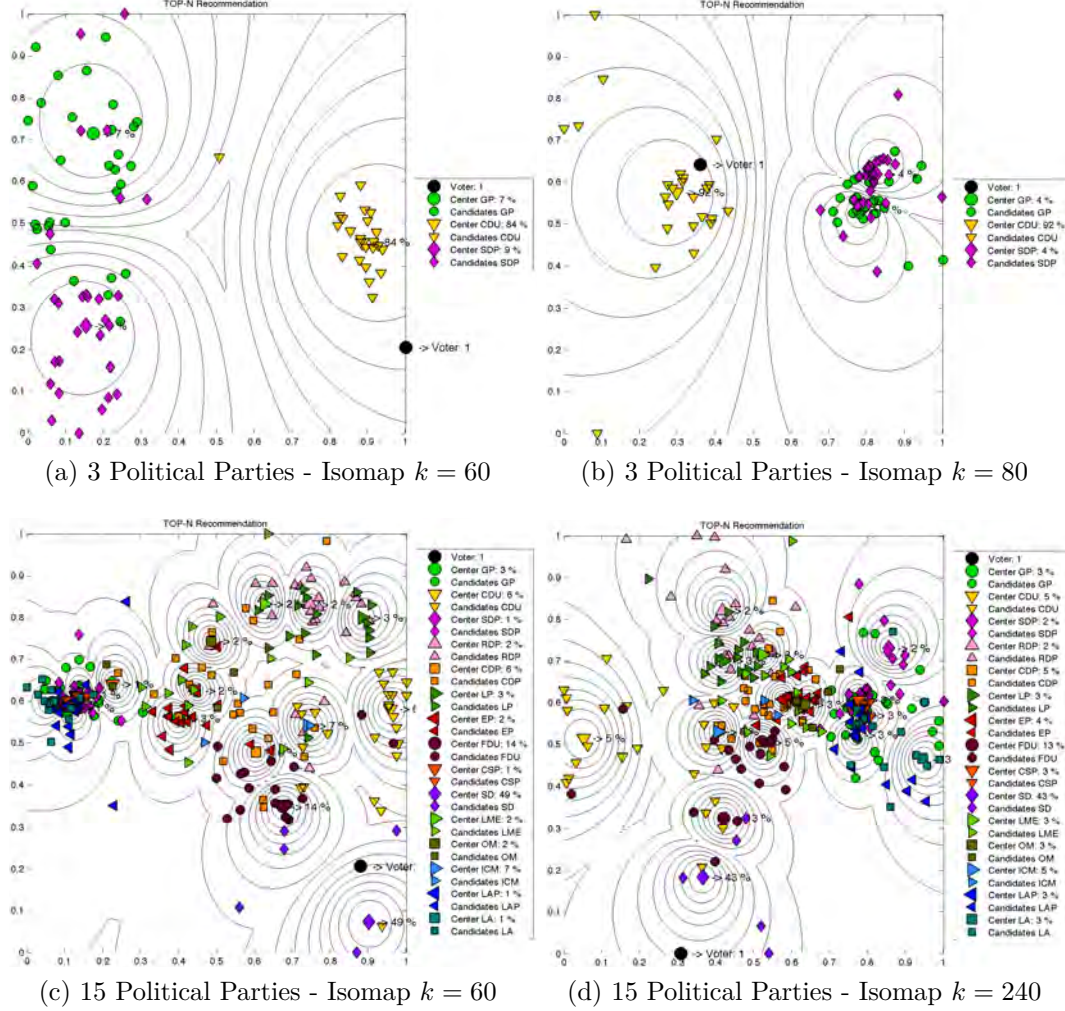


Figure 5.15: Dimensionality Reduction by Isomap for 3 and 15 Political Parties

Locally Linear Embedding (LLE). This method was introduced in the work of Roweis & Saul [2000], and it is based on the visualization of the manifold as a collection of overlapping coordinates. The same was applied with Isomap, which uses neighborhood formation in the initial phase. Let $N(i)$ be the set of k nearest neighbors of x_i . Initially, the algorithm attempts to characterize linear patches representing x_i as weighted convex combinations of closest neighbors. The weights are chosen to minimize the squared error for each i . It is presented in Eq. (5.15)

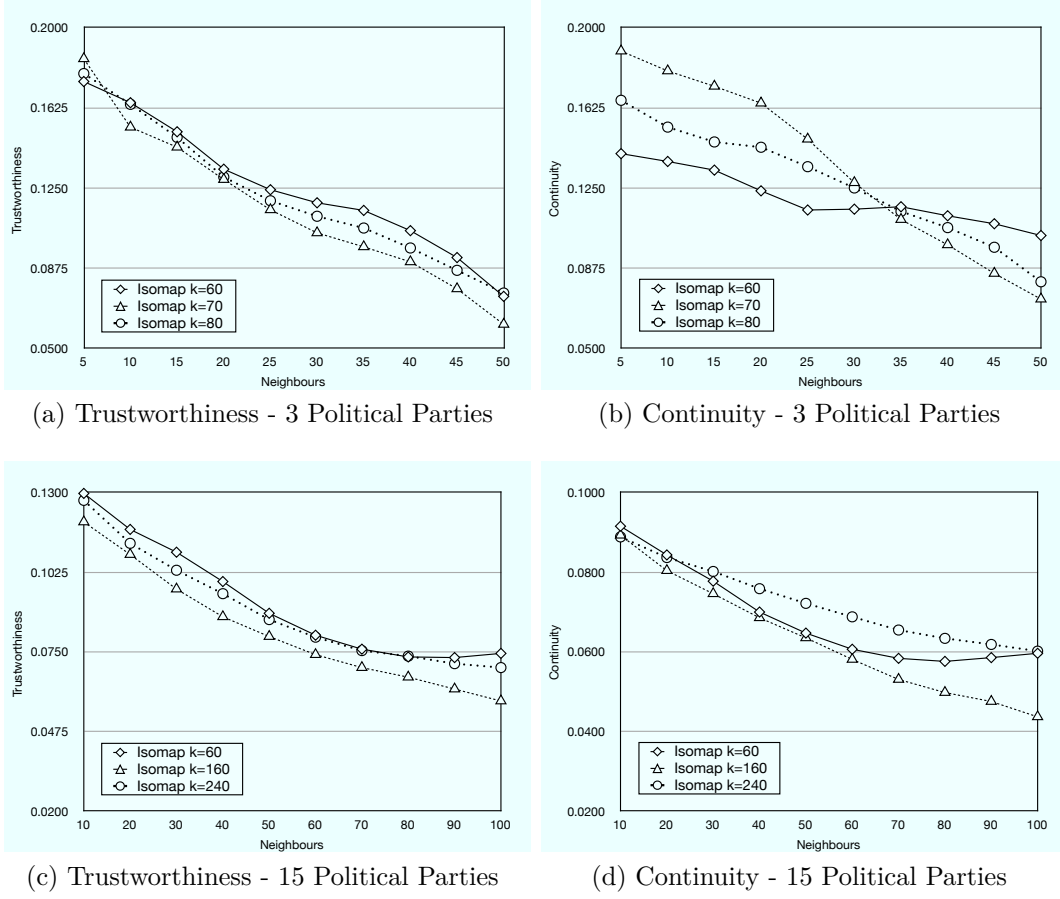


Figure 5.16: Evaluation of Isomap for different local minima

$$\|x_i - \sum_{j \in N(i)} W_{ij} x_j\| \quad (5.15)$$

where each row on the weighted matrix must sum to one and $W_{ij} = 0$ if $j \notin N(i)$, a closest solution using the concept of Lagrange multipliers. Particularly, the weights for each x_i are given in Eq. (5.16)

$$\tilde{W}_i = \frac{\sum_k C_{jk}^{-1}}{\sum_{lm} C_{lm}^{-1}} \quad (5.16)$$

where C is the local covariance matrix with $C_{jk} \stackrel{\text{def}}{=} (x_i - \eta_j)^T (x_i - \eta_k)$ and η_j, η_k are neighbors of x_i . Additionally, $W_i \stackrel{\text{def}}{=} \tilde{W}_i$ characterizes the local geometry of the

Fuzzy Recommender System

manifold around x_i . The second step of the algorithm is to find a configuration in the low-dimensional space, whose geometry is characterized by W and where the dimension d must be known a priori. Finding the d -dimensional representation Y corresponds to minimize the cost function, which is presented in Eq. (5.17)

$$\phi(Y) = \sum_i \|y_i - \sum_{j=1}^k w_{ij} y_{i_j}\|^2 \quad (5.17)$$

where $\|y^{(k)}\|^2 = 1$ for $\forall k$, and $y^{(k)}$ represents the k -th column of the solution matrix Y . The coordinates of the low-dimensional space y_i are found by computing the eigenvectors corresponding to the smallest nonzero d eigenvalues of the product:

$$(\mathbf{I} - \mathbf{W})^T (\mathbf{I} - \mathbf{W}) \quad (5.18)$$

where W is a sparse $n \times n$; $W_{ij} = \tilde{W}_{il}$ if x_j is the l -th neighbor of x_i , and is 0 if $j \notin N(i)$. In Eq. (5.18), \mathbf{I} is the $n \times n$ identity matrix. A pseudocode of LLE method for dimensionality reduction is presented in Algorithm 5.5.

Algorithm 5.5 Locally Linear Embedding Algorithm

Input: data set $X = \{x_1, x_1, \dots, x_n\}$, k -nearest neighbors k , lower dimension d

Output: low-dimensional data representation $Y = \{y_1, y_1, \dots, y_n\}$

1: each data point x_i , computes reconstruction weights for each x_i with:

$$W_i \stackrel{\text{def}}{=} \frac{\sum_k C_{jk}^{-1}}{\sum_{lm} C_{lm}^{-1}}$$

2: Let U be the matrix whose columns are the eigenvectors of $(\mathbf{I} - \mathbf{W})^T (\mathbf{I} - \mathbf{W})$ with nonzero accompanying eigenvalues.

3: **return** $Y \stackrel{\text{def}}{=} [U]_{n \times d}$

In the examples presented in this section, two values of k (closest neighbors) are used for each dataset. The first dataset (83 candidates) was tested with $k = 60$ and $k = 80$, and the second dataset (257 candidates) was tested with $k = 60$ and $k = 240$.

Example 5.8 Figure 5.18 shows the output of the FRS using the LLE method for dimensionality reduction. Two datasets provided by the smartvote project (smartvote [2012b]) are used in this example. Figures 5.17a and 5.17b correspond

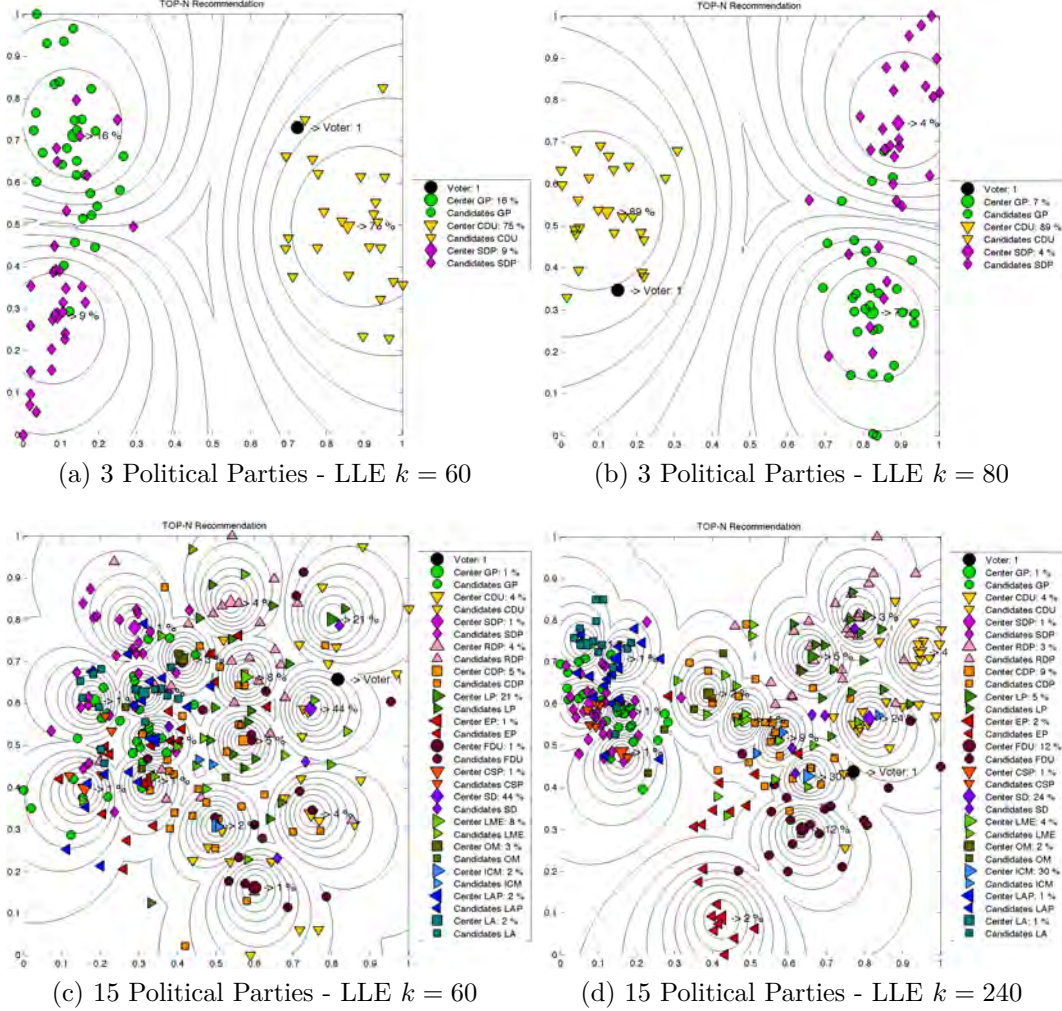


Figure 5.17: Dimensionality Reduction by LLE for 3 and 15 Political Parties

to the visualization of all candidates of three political parties (83 candidates in total), and Figs. 5.17c and 5.17d correspond to all candidates of 15 political parties (257 candidates in total). Both datasets are composed of profiles with 73 dimensions (questions).

Figure 5.18 presents the execution of LLE with two different datasets. Figures 5.17a and 5.17b present two different executions of the algorithm taking three political parties and two different values of k ($k = 60$ and $k = 80$), respectively. The same results are presented in Figs. 5.17c and 5.17d, which include 15 political

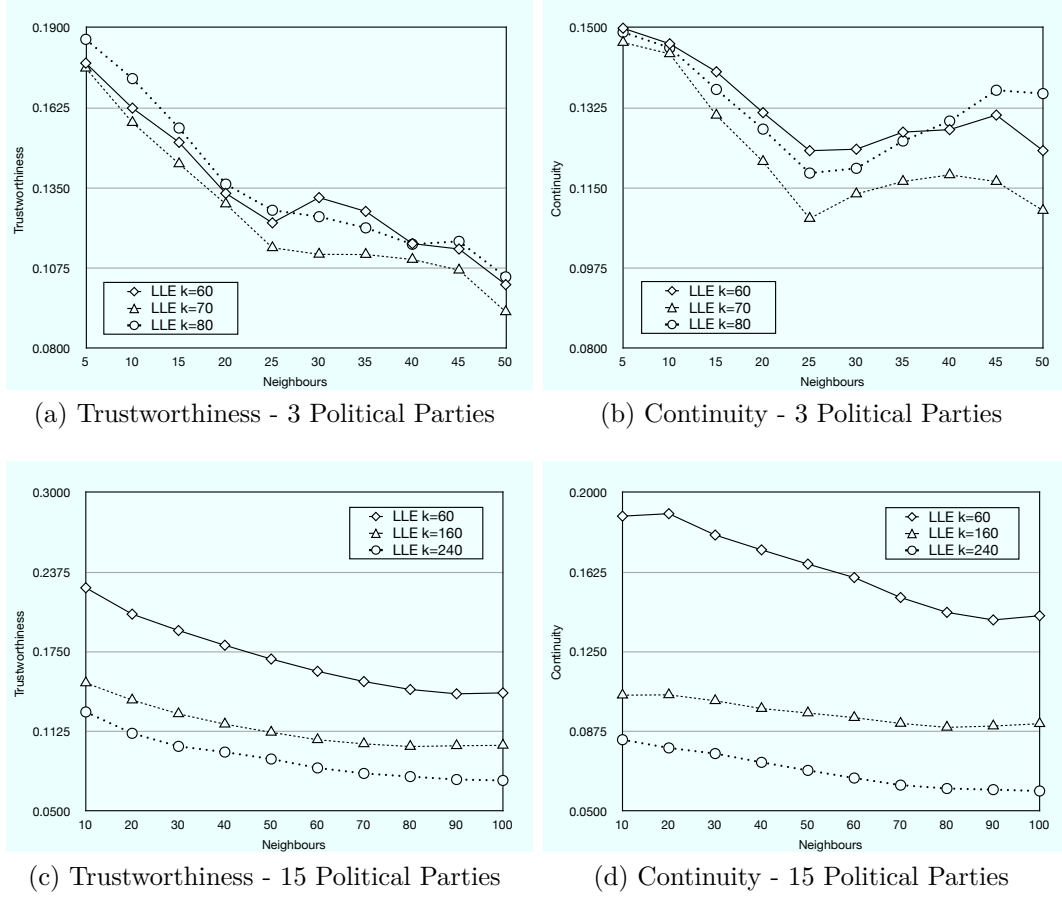


Figure 5.18: Evaluation of LLE for Different Local Minima

parties with $k = 60$ and $k = 240$, respectively. In order to perform an evaluation of LLE and compare it to different dimensionality reduction methods (refer to Sect. 5.4.3), three different values of k are used. Each execution presented different performances in terms of trustworthiness and continuity. Figures 5.18a and 5.18b present the execution of LLE for 3 political parties. It is clear that both trustworthiness and continuity perform better for $k = 60$. Similar results are presented in Figs. 5.18c and 5.18d with $k = 60$ for a higher number of political parties (bigger dataset).

Evaluation of Results

In this section, the evaluation of five dimensionality reduction methods, described in the previous section, is presented. Before presenting the results of the evaluation, it is important to describe the setup of the experiments for each algorithm. Table 5.1 summarizes the different parameters used for the evaluation. The results of the evaluation are presented in Fig. 5.19 and use the datasets provided by the *smartvote* project (smartvote [2012b]). To better understand the results presented, the analysis has been divided into two datasets, one with 3 political parties and the second with 15 political parties.

Table 5.1: Parameters Used for the Evaluation

Method	Parameters
PCA	none
Sammon	Initialization: PCA, max iterations: 1000, $\epsilon = 1e^{-10}$
t-SNE	Initialization: PCA, perplexity: 30
LLE	Dataset with 3 Political Parties, $k = 60$. Dataset with 15 Political Parties, $k = 60$
Isomap	Dataset with 3 Political Parties, $k = 60$. Dataset with 15 Political Parties, $k = 60$

Dataset with three political parties. Figures 5.19a and 5.19b present the evaluation of trustworthiness and continuity using a dataset with 3 political parties. In the case of trustworthiness, the performance of all the algorithms is similar for a small number of neighbors (5 neighbors). Nevertheless, for the higher number of neighbors, there is a considerable improvement with respect to the best (t-SNE) and worst (Sammon) cases of approximately 11.5%. For the case of continuity, there is a clear improvement of performance for the best (t-SNE) and worst (Sammon) cases while increasing the number of neighbors. The performance increased from approximately 5% (5 neighbors) to approximately 12.7% (50 neighbors).

Dataset with 15 political parties. Figures 5.20c and 5.19d present the evaluation of trustworthiness and continuity using a dataset with 15 political parties. The results presented show that, for both trustworthiness and continuity, there is a clear improvement with the best (LLE) and worst (PCA) cases while increasing

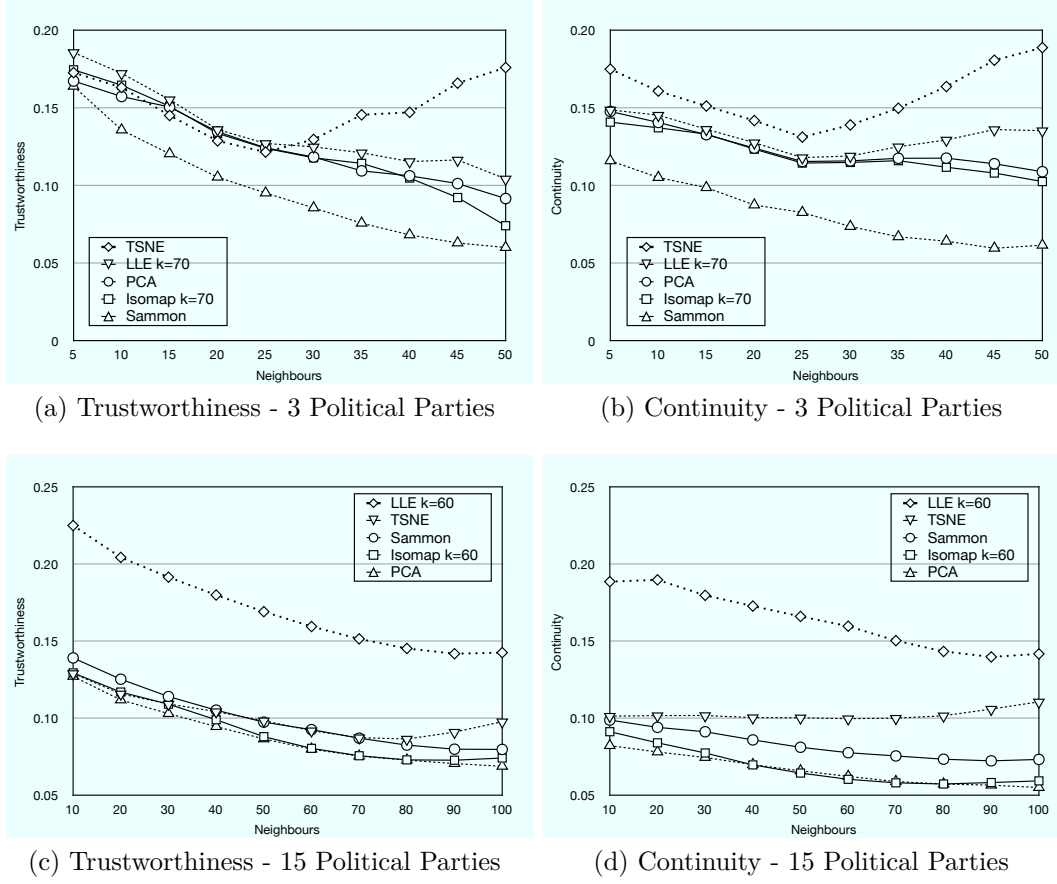


Figure 5.19: Evaluation of Five Dimensionality Reduction Methods

the number of neighbors. The performance increased from approximately 9% (5 neighbors) to approximately 7.3% (50 neighbors) in the case of trustworthiness, and from approximately 10.6% (5 neighbors) to approximately 8.6% (50 neighbors) in the case of continuity.

Analysis of Results

In the previous section, a non-exhaustive quantitative evaluation of dimensionality reduction methods is presented. Nevertheless, selecting the algorithm does not depend only on the performance but also on the number of parameters to be adjusted. It is important to mention that different algorithms have different performances depending on the dataset, number of dimensions, and number of

data points. In the work of Van der Maaten *et al.* [2009], 12 techniques were tested with artificial and natural datasets. The authors showed that, for artificial datasets, techniques based on neighborhood graphs such as Isomap, MVU, LLE, Laplacian Eigenmaps, Hessian LLE, and LTSA presented better performances. The techniques that do not employ neighborhood graphs such as PCA, Sammon mapping, and autoencoders, presented better performances. In the case of the FRS presented in this work, the selection of the dimensionality reduction method considers two factors: stability of the algorithm, and computational and memory performance. Table 5.2 summarizes the three main properties of each algorithm evaluated.

Table 5.2: Properties of Techniques for Dimensionality Reduction. Adapted from Van der Maaten *et al.* [2009]

Method	Parameters	Computational	Memory
PCA	none	$O(D^3)$	$O(D^2)$
Sammon	none	$O(in^2)$	$O(n^2)$
t-SNE	perplexity: <i>perp</i>	$O(n^2)$	$O(n^2)$
LLE	neighbors: <i>k</i>	$O(pn^2)$	$O(pn^2)$
Isomap	neighbors: <i>k</i>	$O(n^3)$	$O(n^2)$

Table 5.2 shows that the PCA and Sammon techniques can both be considered non-parametric. The other methods studied can be considered parametric, and their performance depends on the selection of these parameters. In the case of t-SNE, for different executions of the algorithm that produce different outputs, the interpretation is that they can fall in a local optimum (a solution that is not exactly the best), but only an approximation. They often require a careful adjustment of several hyperparameters (e.g., *k* neighbors) to fasten the convergence and avoid the local minima and converge to a global minima. Taking into account that the FRS works using Euclidean distances for simplicity proposes, the FRS uses the Sammon mapping technique in order to provide the visualization of dimensionality reduction. Additionally, to guarantee the presence of a local optimum, an initialization with the PCA is taken into consideration.

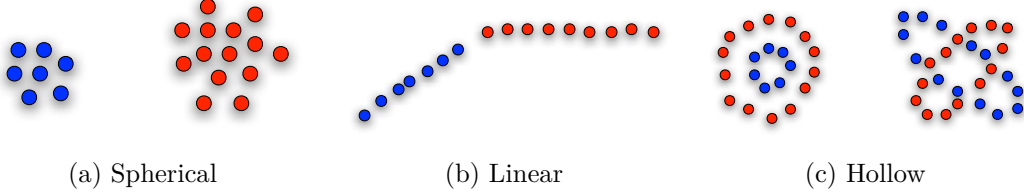


Figure 5.20: Cluster Shapes in \mathbf{R}^2 . Adapted from Balasko *et al.* [2005]

5.4.4 Validation of Fuzzy Clustering Methods

Balasko *et al.* [2005] refers to fuzzy cluster validation as the problem of whether a given fuzzy partition fits a given data point. Different fuzzy clustering methods are proposed in academic literature, aiming to find the best fit for a fixed number of clusters and cluster shapes. The results provided by different fuzzy clustering methods depend on the number of cluster selected, the initialization method selected and cluster shapes. Figure 5.20 shows four examples of different cluster shapes. As mentioned in previous sections, fuzzy clustering is part of the kernel used by the FRS. For that reason, different validation methods, which are proposed in the academic literature, are presented. Additionally, three different fuzzy clustering methods are validated with the methods presented.

Validation Methods for Clustering

In this section, seven cluster validation methods described in the work of Balasko *et al.* [2005] are presented as follows:

Partition Coefficient (PC). Method introduced by Bezdek [1981] that measures the amount of “overlapping” between clusters and is defined in Eq. (5.21).

$$PC(c) = \frac{1}{N} \sum_{i=1}^c \sum_{j=1}^N (u_{ij})^2 \quad (5.19)$$

where u_{ij} is the membership of the j -th cluster i . The highest value of PC represents a highest overlapping between clusters.

Classification Entropy (CE). It is similar to PC and measures the fuzziness of the cluster partition only. The highest value of CE represents the highest level of fuzziness. It is defined in Eq. (5.20).

$$CE(c) = \frac{1}{N} \sum_{i=1}^c \sum_{j=1}^N u_{ij} \log(u_{ij}) \quad (5.20)$$

Partition Index (SC). Method introduced by Bensaid *et al.* [1996] as the ratio of the sum of compactness and separation of clusters. The lowest value of SC indicates a better partition. It is defined in Eq. (5.21)

$$SC(c) = \sum_{i=1}^c \frac{\sum_{j=1}^N (u_{ij})^m \|x_j - v_i\|^2}{N_i \sum_{k=1}^c \|v_k - v_i\|} \quad (5.21)$$

Separation Index (S). Method introduced by Bensaid *et al.* [1996] that measures the separation index using a minimum-distance for partition. The lowest value of SC indicates a better separation of clusters. It is defined in Eq. (5.22).

$$S(c) = \frac{\sum_{i=1}^c \sum_{j=1}^N (u_{ij})^2 \|x_j - v_i\|^2}{N \min_{i,j} \|v_k - v_i\|^2} \quad (5.22)$$

Xie and Beni's Index (XB). Method introduced by Xie & Beni [1991] that aims to quantify the ratio of the total variation within clusters. The lowest value of XB represents the highest separation between clusters. It is defined in Eq. (5.23)

$$XB(c) = \frac{\sum_{i=1}^c \sum_{j=1}^N (u_{ij})^m \|x_j - v_i\|^2}{N \min_{i,j} \|x_j - v_i\|^2} \quad (5.23)$$

XB is used to find the optimal number of clusters with the minimum value of the index.

Dunn’s Index (DI). Method introduced by Dunn [1973b] that is used in the identification of “compact and well separated clusters.” The highest value of DI indicates better compactness and separation of clusters. It is defined in Eq. (5.24).

$$DI(c) = \min_{i \in c} \left\{ \min_{j \in c, i \neq j} \left\{ \frac{\min_{x \in C_i, y \in C_j} d(x, y)}{\max_{k \in c} \{ \max_{x, y \in C} d(x, y) \}} \right\} \right\} \quad (5.24)$$

Alternative Dunn’s Index (ADI). It aims to simplify the calculations when the dissimilarity between two clusters ($\min_{x \in C_i, y \in C_j} d(x, y)$) is rated by Eq. (5.25)

$$d(x, y) \geq |d(y, v_j) - d(x, v_j)| \quad (5.25)$$

where v_j is the cluster center of the j -th cluster. Highest value of ADI indicates better compactness and separation of clusters.

$$ADI(c) = \min_{i \in c} \left\{ \min_{j \in c, i \neq j} \left\{ \frac{\min_{x \in C_i, y \in C_j} |d(y, v_j) - d(x, v_j)|}{\max_{k \in c} \{ \max_{x, y \in C} d(x, y) \}} \right\} \right\} \quad (5.26)$$

Fuzzy Clustering Methods

The FRS in this Ph.D. thesis makes use of fuzzy clustering to provide recommendations using the center of clusters as a reference for both political parties and issues in a bi-dimensional landscape. A non-exhaustive analysis of fuzzy clustering methods is presented. In this section, three methods are analyzed and compared: (1) Fuzzy C-means Algorithm (FCM; Pearson [1901]), (2) Fuzzy Gustafson-Kessel Algorithm (FGK; Gustafson & Kessel [1978]), and (3) Fuzzy Gath-Geva Algorithm (FGG; Bezdek & Dunn [1975]). To visualize the results provided by each fuzzy clustering method, two datasets provided by the *smartvote* project (smartvote [2012b]) are used. The first dataset corresponds to candidates of three political parties (80 candidates in total), and the second dataset corresponds to the candidates of 15 political parties (213 candidates in total). Both datasets are composed of profiles with 73 dimensions (questions).

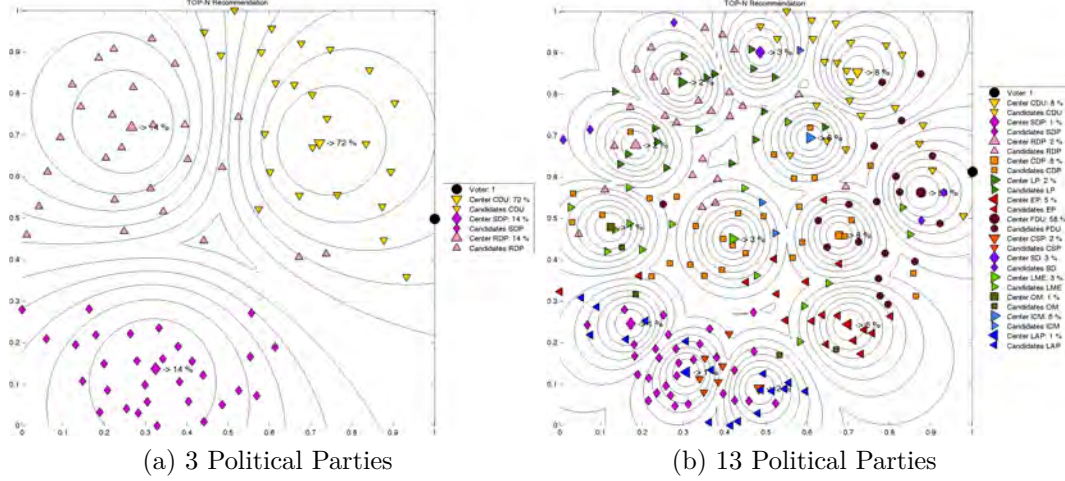


Figure 5.21: Fuzzy Clustering by the FCM for 3 and 13 Political Parties

Fuzzy C-means Algorithm (FCM). In this section, two outputs of the modified FCM, introduced in Sect. 5.3.1, are presented as in Fig. 5.21.

Example 5.9 Figure 5.21 shows the output of the FRS using the FCM method for clustering. Two datasets provided by the smartvote project (smartvote [2012b]) are used in this example. Figure 5.21a corresponds to the visualization of candidates of three political parties (81 candidates), and Fig. 5.21b corresponds to the visualization of candidates of 13 political parties (213 candidates in total). Both datasets are composed of profiles with 73 dimensions (questions).

Figure 5.21 presents the execution of the FCM with two different datasets. Figures 5.21a and 5.21b show that the clusters have similar shapes and can be considered as concentric clusters. The reason is that the FCM uses the standard Euclidean distance norm for the calculation of the partition matrix.

Fuzzy Gustafson-Kessel Algorithm (FGK). It extends the FCM by an adaptive distance norm to detect clusters of different shapes, and was introduced in the work of Gustafson & Kessel [1978]. Each cluster has a norm A_i with the following inner-product presented in Eq. (5.27).

$$D_{ikA}^2 = (\mathbf{x}_k - \mathbf{v}_i)^T A_i (\mathbf{x}_k - \mathbf{v}_i), 1 \leq i \leq c, 1 \leq k \leq N \quad (5.27)$$

where matrices A_i optimize the c-means, allowing clusters to adapt the distance norm to the topological structure of the data. Furthermore, the objective function of the FGK is defined in Eq. (5.28).

$$J(\mathbf{X}, \mathbf{U}, \mathbf{V}, \mathbf{A}) = \sum_{i=1}^c \sum_{k=1}^N (u_{ik})^m D_{ikA_i}^2 \quad (5.28)$$

Varying the determinant of A_i to optimize the cluster shape while its volume remains constant, the following expression is used:

$$||A_i|| = \rho_i, \rho > 0, \quad (5.29)$$

where ρ_i is fixed for each cluster. With the use of a Lagrange multiplier, A_i is derived and shown in Eq. (5.30).

$$A_i = [\rho_i \det(\mathbf{F}_i)]^{1/n} \mathbf{F}_i^{-1} \quad (5.30)$$

where \mathbf{F}_i is the fuzzy covariance matrix of the i -th cluster and is defined in Eq. (5.31)

$$\mathbf{F}_i = \frac{\sum_{k=1}^N (u_{ik})^m (\mathbf{x}_k - \mathbf{v}_i)(\mathbf{x}_k - \mathbf{v}_i)^T}{\sum_{k=1}^N (u_{ik})^m} \quad (5.31)$$

The validation conducted in this section uses an improved version of the FGK introduced by Babuka *et al.* [2002]. A pseudocode of the improved FGK method for clustering is presented in Algorithm 5.6.

Example 5.10 *Figure 5.22 shows the output of the FRS using the FGK method for clustering. Two datasets provided by the smartvote project (smartvote [2012b]) are used in this example. Figure 5.22a corresponds to the visualization of candidates of three political parties (81 candidates), and Fig. 5.22b corresponds to the visualization of candidates of 13 political parties (213 candidates in total). Both datasets are composed of profiles with 73 dimensions (questions).*

Figure 5.22 presents the execution of the FCM with two different datasets. Figures 5.22a and 5.22b show that the clusters have different shapes. The reason is that the FGK forces each cluster to have its own norm with the induction of matrix \mathbf{A}_i , so it adapts the distance norm to the local topological structure of

Algorithm 5.6 Fuzzy Gustafson-Kessel Algorithm

Input: c, ϵ, ρ

Output: $\vec{U}^{(l)}, \vec{V}^{(l)}$

```

1: Set iteration number:  $l \leftarrow 0$ 
2: for  $i = 1$  to  $c$  do
3:    $\mathbf{v}_i^l \leftarrow$  mean average of answers from the  $i$ -th political party ( $P_i$ )
4: end for
5: repeat
6:   Calculate cluster centers
    $\mathbf{v}_i^{(l)} = \frac{\sum_{k=1}^N (u_{ik}^{(l-1)})^m \mathbf{x}_k}{\sum_{k=1}^N (u_{ik}^{(l-1)})^m}, 1 \leq i \leq c$ 
7:   Compute the covariance matrices:
    $\mathbf{F}_i^{(l)} = \frac{\sum_{k=1}^N (u_{ik}^{(l-1)})^m (\mathbf{x}_k - \mathbf{v}_i^{(l)}) (\mathbf{x}_k - \mathbf{v}_i^{(l)})^T}{\sum_{k=1}^N (u_{ik}^{(l-1)})^m}$ 
   Add a scaled identity matrix:
    $\mathbf{F}_i := (1 - \gamma) \mathbf{F}_i + \gamma (\mathbf{F}_0)^{-1} / n \mathbf{I}$ 
   Extract eigenvalues  $\gamma_{ij}$  and eigenvectors  $\phi_{ij}$ , find  $\gamma_{i,\max} = \max_j \gamma_{ij}$  and set:
    $\gamma_{i,\max} = \gamma_{ij} / \beta, \forall j$  for which  $\gamma_{i,\max} / \gamma_{ij} \geq \beta$ 
   Reconstruct  $\mathbf{F}_i$  by:
    $\mathbf{F}_i = [\phi_{i,1} \dots \phi_{i,n}] \text{diag}(\gamma_{i,1} \dots \gamma_{i,n}) [\phi_{i,1} \dots \phi_{i,n}]^{-1}$ 
8:   Compute the distances:
    $D_{ikA_i}^2(\mathbf{x}_k, \mathbf{v}_i) = (\mathbf{x}_k - \mathbf{v}_i^{(l)})^T [(\rho_i \det(\mathbf{F}_i))^{1/n} \mathbf{F}_i^{-1}] (\mathbf{x}_k - \mathbf{v}_i^{(l)})$ 
9:   Update the partition matrix:
    $u_{ik}^{(l)} = \frac{1}{\sum_{j=1}^c (D_{ikA_i}(\mathbf{x}_k, \mathbf{v}_i) / D_{jk}(\mathbf{x}_k, \mathbf{v}_j))^{2/(m-1)}}, 1 \leq i \leq c, 1 \leq k \leq N$ 
10: until  $|\vec{U}^{(l)} - \vec{U}^{(l-1)}| \leq \epsilon$ 
11: return  $\vec{U}^{(l)}, \vec{V}^{(l)}$ 

```

the data points. Additionally, the FGK uses the Mahalanobis distance instead of the Euclidean distance. This includes another input parameter ρ_i . It is set equal to 1 if there is no a priori knowledge of it for each cluster.

Fuzzy Gath-Geva Algorithm (FGG). It is based on the computation of the fuzzy maximum likelihood estimator (FMLE) proposed by Bezdek & Dunn [1975]. It uses a distance norm based on FMLE that is shown in Eq. (5.32).

$$D_{ik}(\mathbf{x}_k, \mathbf{v}_i) = \frac{\sqrt{\det \mathbf{F}_{wi}}}{\alpha_i} \exp \left(\frac{1}{2} (\mathbf{x}_k - \mathbf{v}_i^{(l)})^T \mathbf{F}_{wi}^{-1} (\mathbf{x}_k - \mathbf{v}_i^{(l)}) \right) \quad (5.32)$$

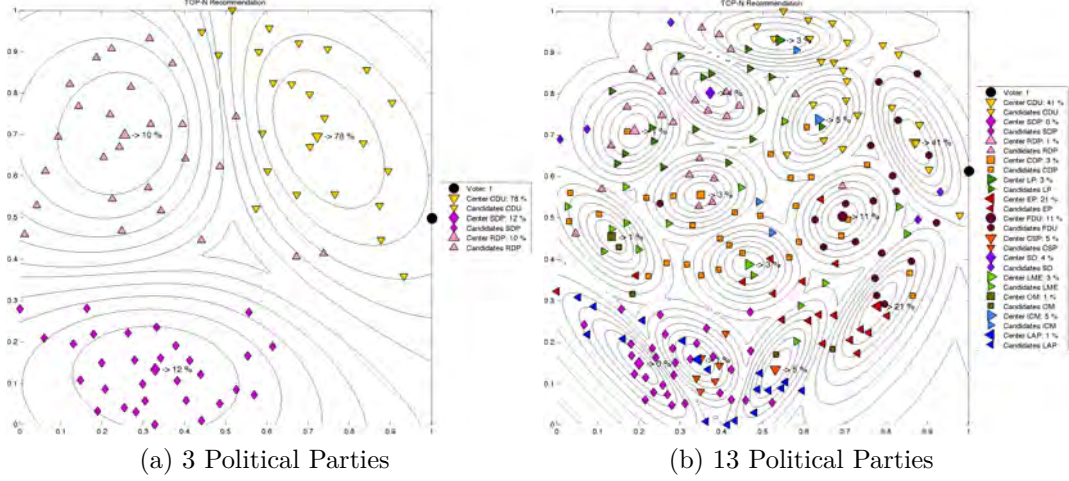


Figure 5.22: Fuzzy Clustering by the FGK for 3 and 13 Political Parties

The fuzzy covariance matrix of the i -th cluster is given by Eq. (5.33).

$$\mathbf{F}_{wi} = \frac{\sum_{k=1}^N (u_{ik})^w (\mathbf{x}_k - \mathbf{v}_i)(\mathbf{x}_k - \mathbf{v}_i)^T}{\sum_{k=1}^N (u_{ik})^w}, 1 \leq i \leq c \quad (5.33)$$

where $w = 2$, so the partition becomes fuzzy to compensate the exponential term of the distance norm. The prior probability of selecting cluster i is given in Eq. (5.34).

$$\alpha_i = \frac{1}{N} \sum_{k=1}^N u_{ik} \quad (5.34)$$

The membership degree u_{ik} is the posterior probability of selecting the i -th cluster given the data point x_k . A pseudocode of the FGG method for clustering is presented in Algorithm 5.7.

Example 5.11 Figure 5.23 shows the output of the FRS using the FGG method for clustering. Two datasets provided by the smartvote project (smartvote [2012b]) are used in this example. Figure 5.23a corresponds to the visualization of candidates of three political parties (81 candidates), and Fig. 5.23b corresponds to the visualization of candidates of 13 political parties (213 candidates in total). Both datasets are composed of profiles with 73 dimensions (questions).

Algorithm 5.7 Fuzzy Gath-Geva Algorithm

Input: c, m, ϵ

Output: $\vec{U}^{(l)}, \vec{V}^{(l)}$

```

1: Set iteration number:  $l \leftarrow 0$ 
2: for  $i = 1$  to  $c$  do
3:    $\mathbf{v}_i^l \leftarrow$  mean average of answers from the  $i$ -th political party ( $P_i$ )
4: end for
5: repeat
6:   Calculate cluster centers
    $\mathbf{v}_i^{(l)} = \frac{\sum_{k=1}^N (u_{ik}^{(l-1)})^w \mathbf{x}_k}{\sum_{k=1}^N (u_{ik}^{(l-1)})^w}, 1 \leq i \leq c$ 
7:   Compute the distance measure  $D_{ik}^2$ .
   The distance of the prototype is calculated with the fuzzy covariance matrices of clusters as follows:
    $\mathbf{F}_i^{(l)} = \frac{\sum_{k=1}^N (u_{ik}^{(l-1)})^w (\mathbf{x}_k - \mathbf{v}_i^{(l)}) (\mathbf{x}_k - \mathbf{v}_i^{(l)})^T}{\sum_{k=1}^N (u_{ik}^{(l-1)})^w}$ 
   The distance function is chosen as:
    $D_{ik}^2(\mathbf{x}_k, \mathbf{v}_i) = \frac{(2\pi)^{n/2} \sqrt{\det(\mathbf{F}_i)}}{\alpha_i} \exp \left( \frac{1}{2} (\mathbf{x}_k - \mathbf{v}_i^{(l)})^T \mathbf{F}_i^{-1} (\mathbf{x}_k - \mathbf{v}_i^{(l)}) \right)$ 
   with the a priori probability
    $\alpha_i = \frac{1}{N} \sum_{k=1}^N u_{ik}$ 
8:   Update the partition matrix:
    $u_{ik}^{(l)} = \frac{1}{\sum_{j=1}^c (D_{ik}(\mathbf{x}_k, \mathbf{v}_i) / D_{jk}(\mathbf{x}_k, \mathbf{v}_j))^{2/(m-1)}}, 1 \leq i \leq c, 1 \leq k \leq N$ 
9: until  $|\vec{U}^{(l)} - \vec{U}^{(l-1)}| \leq \epsilon$ 
10: return  $\vec{U}^{(l)}, \vec{V}^{(l)}$ 

```

Figure 5.23 presents the execution of the FCM with two different datasets. Figures 5.23a and 5.23b show that the clusters have different shapes. The outputs presented in this section using the FGK can be compared to c-means algorithm (refer to Sect. 4.3.1). In Fig. 5.23a one can see that voter 1 has a membership degree of 100% to the CDU, 0% to SDP, and 0% to RDP. Similar results can be observed in Fig. 5.23a; the only difference is that voter 1 has a membership degree of 99% to FDU and 1% to CDU. For the other 11 political parties, voter 1 has a membership degree of 0%. Another problem that is visualized in Fig. 5.23b is that the FGG can overflow at large values of c . This is problem is also mentioned in the work of Balasko *et al.* [2005].

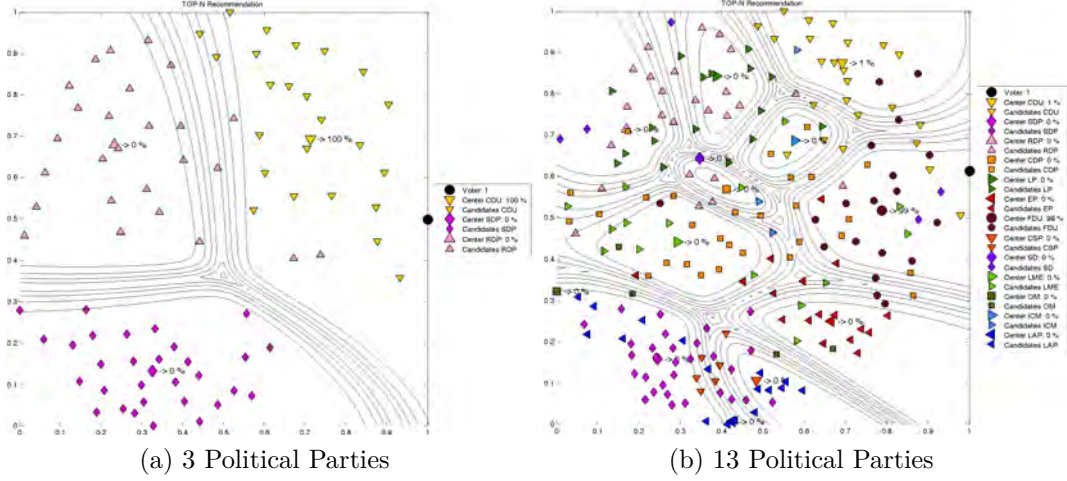


Figure 5.23: Fuzzy Clustering by the FGK for 3 and 13 Political Parties

Evaluation of Results

In this section, the validation of three fuzzy clustering methods described in the previous section is presented. Seven different methods for validation are used (refer to Sect. 5.4.4). To better understand the results presented, the analysis has been divided into two datasets, one with 3 political parties and the second with 13 political parties.

Dataset with 3 political parties. Table 5.3 presents the validation with methods described in Sect. 5.4.4 and use a dataset with 3 political parties. The results show that the FCM algorithm has a better performance using the validation methods CE and DI. The FGK algorithm shows better performance when using the validation methods SC and S. Finally, the FGK algorithm shows better performance when using the validation methods PC, XB, and ADI. The values shown in bold represent the highest performance.

Dataset with 15 political parties. Table 5.4 presents the validation with methods described in Sect. 5.4.4 and uses a dataset with 13 political parties. The first observation that has to be made is that the performance of different validation methods decreases when the number of clusters increases. The results presented in Table 5.4 show that the FCM algorithm has a better performance with the

Method	PC ¹	CE ²	SC ³	S ⁴	XB ⁵	DI ⁶	ADI ⁶
FCM	0.7069	0.5372	1.1041	0.0160	4.2152	0.1327	0.0014
FGK	0.7455	0.4719	1.0416	0.0154	3.5832	0.0905	0.0051
FGG	0.9614	0.0707	2.1258	0.0311	1.6861	0.0202	0.1118

¹ The highest value represents a highest overlapping between clusters.

² The highest value represents a highest level of fuzzyness.

³ The lowest value represents a better partition.

⁴ The lowest value represents a highest separation between clusters.

⁵ The lowest value represents a highest level of compactness and well separated clusters.

⁶ Highest value represents a highest level of compactness and well separated clusters.

Table 5.3: Validation of Fuzzy Cluster Methods - 3 Political Parties

validation methods CE, SC, XB, and DI. The FGK algorithm presented a better performance when using the validation method S. Finally, the FGG algorithm performed better using the validation methods PC and ADI. The values shown in bold represent the highest performance.

Method	PC ¹	CE ²	SC ³	S ⁴	XB ⁵	DI ⁶	ADI ⁶
FCM	0.4287	1.4122	0.7680	0.0064	1.8545	0.0652	0.0004
FGK	0.4703	1.2803	0.7795	0.0059	2.3409	0.0521	0.0015
FGG	0.8504	NaN ⁷	NaN ⁷	NaN ⁷	NaN ⁷	0.0514	0.0438

¹ The highest value represents a highest overlapping between clusters.

² The highest value represents a highest level of fuzziness.

³ The lowest value represents a better partition.

⁴ The lowest value represents a highest separation between clusters.

⁵ The lowest value represents a highest level of compactness and well separated clusters.

⁶ Highest value represents a highest level of compactness and well separated clusters.

⁷ Not a number.

Table 5.4: Validation of Fuzzy Cluster Methods - 15 Political Parties

Table 5.4 shows that FGG presents NaN values when using CE, SC, S, and XB. It can be explained due to the hardly detectable connection to the data structure and the high level of overlapping of cluster centers when using the FGG algorithm. In the work of Balasko *et al.* [2005], the authors mention that the most useful validation methods when comparing different clustering algorithms are SC and S with the same value of c .

Analysis of Results

In the previous section, a non-exhaustive quantitative evaluation of fuzzy clustering methods is presented. Nevertheless, making a selection of the algorithm does not depend only on the performance but also on the number of parameters to be adjusted. It is important to mention that different algorithms have different performances depending on the structure of the dataset and the number of

clusters. In their work, Balasko *et al.* [2005] mentioned that, besides of having different methods for the validation of fuzzy clustering, none of these methods can be considered as perfect. Based on the results presented in previous sections, some conclusions can be made.

Firstly, the FCM algorithm has a good performance and a lower complexity compared to the FGK and FGG algorithms. Both the FCM and FGK algorithms are more suitable for spherical clusters (refer to Fig. 5.20a) and the FGK algorithm is more suitable for linear clusters (refer to Fig. 5.20b).

Secondly, the FGK algorithm presents an advantage compared to the FCM algorithm since each cluster has the property to adapt the distance norm to the local topological structure. A disadvantages of this algorithm, is that, it requires a priori knowledge of ρ for each cluster. In the examples shown in previous sections, a constant value of $\rho = 1$ was used, so the algorithm finds clusters with approximately equal volume.

Finally, the FGG algorithm has shown better performance in terms of validation in the case of small number of clusters, Nevertheless, it can be compared to c-means algorithm (refer to Sect. 4.3.1). In Fig. 5.23a, one can see that voter 1 has a membership degree of 100% to the CDU, 0% to SDP, and 0% to RDP. Similar results can be observed in Fig. 5.23a; the only difference is that voter 1 has a membership degree of 99% to FDU and 1% to CDU. For the other 11 political parties, voter 1 has a membership degree of 0%.

For the reasons mentioned above, the author recommends the use of the FCM and the FGK for the implementation of the recommender system approach proposed in this Ph.D. thesis, which uses the datasets provided by the *smartvote* project. It is important to mention that the results provided by different algorithms directly depend on the datasets selected, number of clusters, and the particular shapes of the clusters, among other reasons.

5.5 Remarks about the FRS

In this chapter, a recommender system architecture for *eParticipation* has been proposed. The Web-based recommendation engine can be used to visualize differentiated clusters of politicians as well as of citizens. It, therefore, supports

collaboration, eElection processes for candidates, building processes for political communities that share common objectives, and civic participation. The recommender system proposed in this work can be used for *eCollaboration*, *eDemocracy*, and *eCommunity*. Based on a fuzzy clustering approach, it computes similarities between citizens and politicians in a multi-dimensional space. The Sammon mapping technique allows for a better understanding and evaluation of the relationships among citizens and/or politicians using a bi-dimensional graphical interface. The recommender system approach presented in this work differs from collaborative filtering methods in that the latter are based on past experiences. It is also suitable in the one-and-only scenario, in which events such as election processes occur only once, and their participants (candidates and/or citizens) cannot be considered unique, since their presence at such events and their way of thinking can vary over time.

5.6 Further Readings

- **Steele & Iliinsky [2010]**. “Visualization is the graphic presentation of data—portrayals meant to reveal complex information at a glance. Think of the familiar map of the New York City subway system, or a diagram of the human brain. Successful visualizations are beautiful not only for their aesthetic design, but also for elegant layers of detail that efficiently generate insight and new understanding” (abs.).
- **Mazza [2009]**. “Information Visualization is a relatively young field that is acquiring more and more consensus in both academic and industrial environments. Information Visualization explores the use of computer-supported interactive graphical representations to explain data and amplify cognition. It provides a means to communicate ideas or facts about the data, to validate hypotheses, and facilitates the discovery of new facts via exploration” (abs.).
- **Gorban [2007]**. “In 1901 Karl Pearson invented PCA. Since then, PCA serves as a prototype for many other tools of data analysis, visualization and dimension reduction: Independent Component Analysis (ICA), Mul-

tidimensional Scaling (MDS), Nonlinear PCA (NLPCA), Self Organizing Maps (SOM), etc” (abs.).

- **Lee & Verleysen [2007]**. “This book describes established and advanced methods for reducing the dimensionality of numerical databases. Each description starts from intuitive ideas, develops the necessary mathematical details, and ends by outlining the algorithmic implementation. The text provides a lucid summary of facts and concepts relating to well-known methods as well as recent developments in nonlinear dimensionality reduction” (abs.).
- **Balasko *et al.* [2005]**. “The Fuzzy Clustering and Data Analysis Toolbox is a collection of Matlab functions. Its propose is to divide a given data set into subsets (called clusters), hard and fuzzy partitioning mean, that these transitions between the subsets are crisp or gradual” (abs.).

Chapter 6

SmartParticipation

The *SmartParticipation* project is Web-based platform, which intends to enhance participation of citizens through the use of information and communication technologies. In this chapter, the reader will be presented with a five level participation model, which includes the concepts of Web 2.0 in order to provide community-building processes and discussion between citizens and authorities. The levels defined in this chapter are also used for evaluation of different participation platforms in Chap. 7. Additionally, this chapter presents a maturity model, which includes implementation instances of the fuzzy recommender system (FRS) presented in the previous chapter. This chapter is structured as follows: First, Sect. 6.1 discusses the growth of research on *eParticipation* and presents the definitions of participation levels, which will be used for evaluation in Chap. 7. Section 6.2 gives a brief introduction and scope of the *SmartParticipation* project. Then, Sect. 6.3 introduces three maturity models for *eCollaboration*, *eDemocracy*, and *eCommunity*. Section 6.4 presents some remarks about *SmartParticipation*. Finally, further readings are presented in Sect. 6.5.

6.1 eParticipation

The need of citizens and other stakeholders for a free democratic debate, and the right to be involved in the decision-making process, has been emphasized by many democratic theorists. The use of ICT has opened new channels for free

discussion of political issues, and day by day it is moving away from traditional media like TV, radio, mail, and newspapers. *eParticipation* is an emerging and growing research area that aims to increase citizens' participation in order to promote a fair and efficient society and government support by using the latest technology developments. In the last years, *eParticipation* has been addressed more often by the academic world. For example, Sanford & Rose [2007] provide a formative analysis of this emerging research area.

The authors identified 99 academic articles that are considered highly relevant to *eParticipation*. In the latest years, a number of dedicated conferences, workshops, and projects related to the field have appeared, such as the International Conference on *eParticipation* (ePart [2012]), IPIP e-government Conference (EGOV [2012]), International Conference on Electronic Government and the Information Systems (EGOVIS [2012]), International Conference on Information Society (i-Society [2012]), and MOMENTUM project the (MOMENTUM [2012]). Panopoulou *et al.* [2010] present an analysis and evaluation of different *eParticipation* initiatives in the European Union. In their work, the authors identified 255 initiatives from 23 European countries, 230 of which were contacted and evaluated via a survey. In this section, the author points out the increasing interest in the academic sector regarding *eParticipation*. Many different governmental, non-governmental, and research-oriented projects with the potential to support participation are readily available or in development.

6.1.1 Participation Levels

In academic literature, different methods for describing the level of participation have been proposed. For instance, the work of Grönlund [2009] describes different methods for determining the level of participation, which are summarized in Table 6.1. The table also shows a brief description and the levels of participation proposed by different authors.

Table 6.1: eParticipation Levels by Grönlund [2009]

Authors	Focus	Brief Description of Work	Particip. Levels
Arnstein [1969]	Participation	This model is designed to define stages of citizen influence over policy. The model is based on a direct democracy model.	1. Citizen Control 2. Delegated Power 3. Partnership 4. Placation 5. Consultation 6. Informing 7. Therapy 8. Manipulation
IAP2 [2012]	Participation	This model is designed to define stages of citizen influence over policy. The model is based on a direct democracy model.	1. Empower 2. Collaborate 3. Involve 4. Consult 5. Inform
Tambouris <i>et al.</i> [2007]	ICT use	This model is an attempt to produce a framework for assessing not only <i>eParticipation</i> projects but also <i>eParticipation</i> tools.	1. eEmpowerment 2. eCollaborating 3. eInvolving 4. eConsulting 5. eInforming
OECD [2001]	Improving democratic decision-making	This model is designed to improve representative democracy by introducing participation with citizens. It is open to different models of democracy.	1. Active Participation 2. Consultation 3. Information
Lukensmeyer & Torres [2006]	Improving democratic decision-making	This model is designed to improve representative democracy. It has four levels of participation.	1. Collaboration 2. Engagement 3. Consultation 4. Communication
Macintosh & Whyte [2008]	Improving citizen engagement	This model does not detail steps concerning either participation or democracy, but rather takes a project approach.	1. eEmpowering 2. eEngaging 3. eEnabling

The evaluation framework proposed in this Ph.D. thesis is inspired by the work of Tambouris *et al.* [2007] and includes the concepts of Web 2.0 in order to provide community-building processes and discussion between citizens and authorities. The model proposed in this work consists of five levels: *eInforming*, *eConsulting*,

SmartParticipation

eDiscussion, *eParticipation*, and *eEmpowerment*. Each of these levels is described in more detail as follows:

Level I - eInforming. This is the lowest level and uses a unidirectional (top-down) information channel to provide citizens with relevant information about different policies and projects. At this stage, citizens are only informed by the government; no interaction, participation, or decision is present.

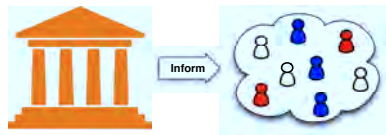


Figure 6.1: eInforming

Level II - eConsulting. This level of involvement uses a bi-directional information channel and gives the authorities the ability to collect feedback from citizens. At this stage, citizens are consulted by the government and minimal interaction is present. Nevertheless, neither participation nor decision is present.

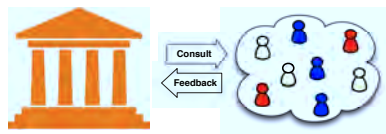


Figure 6.2: eConsulting

Level III - eDiscussion. This level of involvement uses a bi-directional information channel and provides citizens and government with the ability to establish discussion channels and create virtual communities by building citizen communication centers. Public project ideas and plans can be discussed and commented on, taking advantage of specialized groups (communities) in order to promote the opinion-forming process. At this stage, citizens are able to establish communication channels. Nevertheless, neither participation nor decision is yet present.

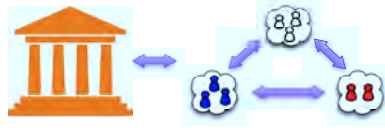


Figure 6.3: eDiscussion

Level IV - eParticipation. This level of involvement uses a bi-directional information channel and provides citizens with the ability to collaborate on public projects and developing bases for decision-making. At this stage, citizens are able to establish much bigger communication channels, which include more capabilities such as working collaboratively to enhance participation. The first steps toward empowerment are taken here.

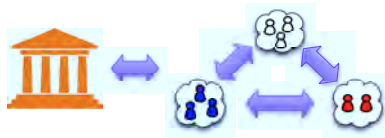


Figure 6.4: eParticipation

Level V - eEmpowerment. This level of involvement uses a bi-directional information channel and places the final decision in the hands of the citizens, thus implementing what they have decided. At this stage, citizens are empowered, as the communication channels are much bigger and include new and better capabilities towards empowerment.



Figure 6.5: eEmpowerment

It is important to point out that, from the processes in Fig. 6.1 to Fig. 6.5, the comparative size of the government is reduced gradually. This visual representation was created intentionally. The interpretation behind this idea is that citizens are empowered as they get closer to the highest level of participation.

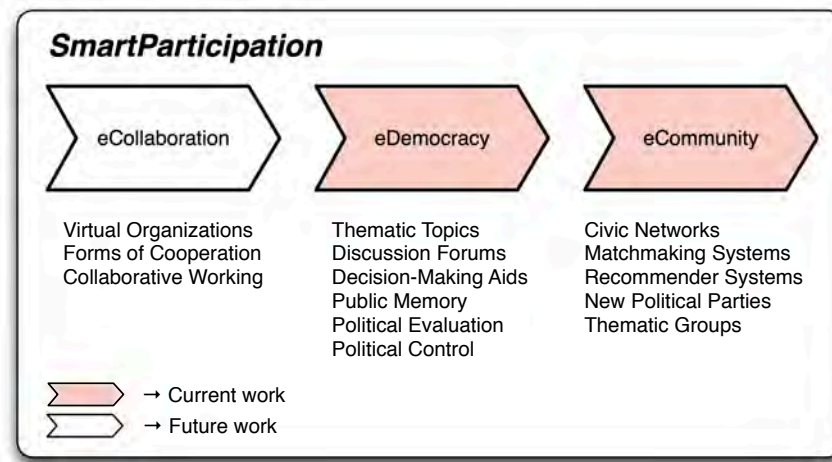


Figure 6.6: The *SmartParticipation* Project

Figure 6.5 shows that, at the *eEmpowerment* level, the final decisions are now placed with the citizens.

6.2 Project Description

The *SmartParticipation* project intends to provide citizens with a simple and innovative alternative using a fuzzy-based recommender system (see Chap. 5) to enhance participation of citizens in three main areas, *eCollaboration*, *eDemocracy*, and *eCommunity*, which are part of the eGovernment Framework used in this Ph.D. thesis (refer to Sect. 2.1). The project is based on the participation levels presented in previous sections in order to create maturity models taking into account three main components: Web evolution, media richness, and communication channels. More details about the use of these three components and an evaluation framework for *eParticipation* projects are described in Chap. 7. The current recommendation engine, described in this Ph.D. thesis, has been designed for *eDemocracy* and *eCommunity*. Further work will also cover *eCollaboration*. The current areas of *SmartParticipation* and the different ICT tools are presented in Fig. 6.6.

6.3 eParticipation Maturity Models

In this section, a brief description of a maturity model for *eParticipation* is presented. It uses the FRS (refer to Chap. 5) and the five levels of participation presented in previous sections. The FRS provides visualization tools for community creation processes and recommendations for citizens and governments looking for specific profiles or topics of interest. The participation levels determine the level of empowerment reached by different participation projects. The three participation components that are the focus of the *SmartParticipation* project (i.e., *eCollaboration*, *eDemocracy*, and *eCommunity*) are used to illustrate the use of the maturity model.

6.3.1 eParticipation Maturity Model for eCollaboration

Collaboration can be defined as a process by which people work together on intellectual, academic, or practical objectives. On the other hand, *eCollaboration* made this process possible by means of electronic technologies, using tools such as e-mail, instant messaging, application sharing, videoconferencing, collaborative workspace, and document management. The term *eCollaboration* is also related to collaborative working environments (CWE) and makes reference to other terms, such as online collaboration, online communities of practice, open source community, groupware, and open innovation principles.

CWE allows collaborators to communicate anytime and anyplace. People from different areas (e.g., buildings, states, countries, or continents) can exchange information, collaborate on shared documents and ideas, study together, or propose new projects. The *SmartParticipation* project presented in this work extends the idea of CWE by taking into account the five participation levels to enhance citizens' participation and to empower citizens in the decision-making process. Each level is briefly defined as follows:

eInforming. *eInforming* is often used by governments as a channel or tool to distribute and share information about new projects. At this level, citizens could get specialized information about different projects depending on their profiles and know-how.

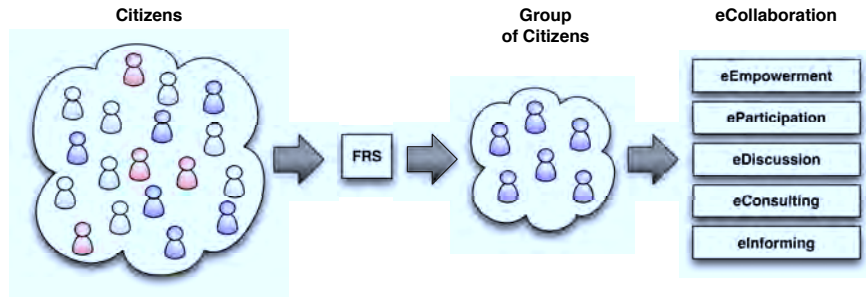


Figure 6.7: eParticipation Maturity Model for eCollaboration

eConsulting. This level of involvement uses a bi-directional information channel and gives authorities the ability to collect feedback from citizens about various projects.

eDiscussion. This level of involvement uses a bi-directional information channel and provides citizens and government with the ability to establish discussion channels and create virtual communities by building citizen communication centers. Public project ideas and plans can be discussed and commented on, taking advantage of specialized groups (communities) in order to promote the opinion-forming process.

eParticipation. Based on their know-how, citizens can form communities in order to take part in different initiatives and projects. The creation of groupware among citizens allows for interaction and participation through social media, potentially crossing geographical and political boundaries.

eEmpowerment. The highest level of participation could provide citizens the opportunity to express their will on various initiatives and projects. At this point, the decision of a specific initiative or project is placed in the hands of a group of citizens.

Figure 6.7 shows the schema of the maturity model for *eCollaboration* used in the *SmartParticipation* project.

6.3.2 eParticipation Maturity Model for eDemocracy

In academic literature, RSs that are used in *eDemocracy* are known as voting advice applications (VAAs) or decision-making aids. Such applications are used in the stage defined in the work of Meier [2012] as *eDiscussion* (refer to Sect. 2.3.1). The *SmartParticipation* project presented in this work extends the idea of FRS used mainly in *eDiscussion* to be used in the third level of the eGovernment framework at the University of Fribourg (refer to Fig. 2.1) to enhance and stimulate citizens' participation.

eInforming. *eInforming* is often used by governments as a channel or tool to distribute and share information with citizens. This one-sided approach, however, leaves room for improvement, meaning that citizens should get involved more in the process of supplying information. By balancing the amount of information supplied by different sources, citizens would have a larger pool of opinions to choose from, and, therefore, more freedom of choice.

eConsulting. Politicians, as well as citizens, should be given the opportunity to fill out questionnaires regarding their political preferences and receive, in return, customized recommendations for candidates and political parties. Such recommendation systems should not only include the names of candidates and parties, but also a short description of why these choices are being recommended by the application.

eDiscussion. While a majority of RSs are good at *eInforming*, only a few have gone a step further and implemented *eDiscussion* features, which give citizens and politicians the opportunity to discuss different political programs and issues. A more widespread implementation of *eDiscussion* would be beneficial for all parties, as it would allow more personalized questions to be raised and answered. An additional positive effect is the reduction of distance between citizens and politicians, which renders politicians more tangible.

eParticipation. Based on their political preferences, citizens can form communities in order to launch initiatives, programs, or even the creation of new

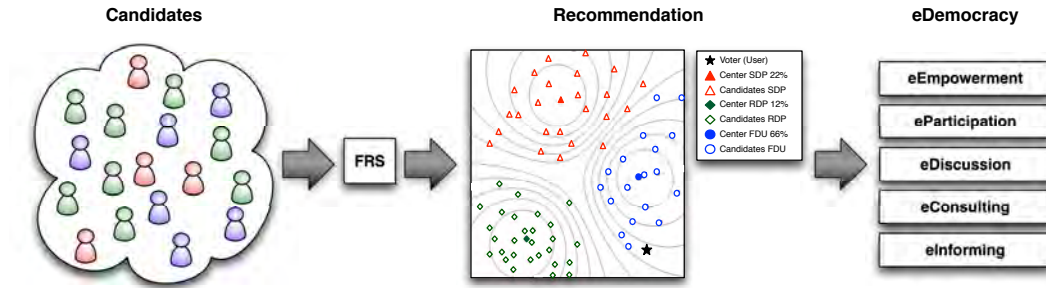


Figure 6.8: eParticipation Maturity Model for eDemocracy

political organizations (e.g., political parties, and NGOs).

eEmpowerment. The highest level of participation can provide citizens with the opportunity to express their will on various initiatives based on better informing, consulting, discussion, participation, and the recommendations provided by the VAAs. It could also be used in eElection or eVoting processes in the same platform. Citizens also use a *Political Controlling* process with *Public Memory* described in the work of Meier [2012]. This maturity model of VAAs can be used to evaluate whether candidates really act the way they claim they will. Additionally, citizens who are voting discuss the election results in order to influence the success of implemented solutions.

Figure 6.8 shows the schema of the maturity model for *eDemocracy* used in the *SmartParticipation* project.

6.3.3 eParticipation Maturity Model for eCommunity

The *SmartParticipation* project allows citizens to create virtual communities based on their profiles, such as new political parties, thematic groups, and civic networks, and participate in national issues by opening channels of discussion and debate through the use of ICTs and Web 2.0.

eInforming. At this level, citizens are informed on various communities of interest or communities of practice based on thematic profiles.

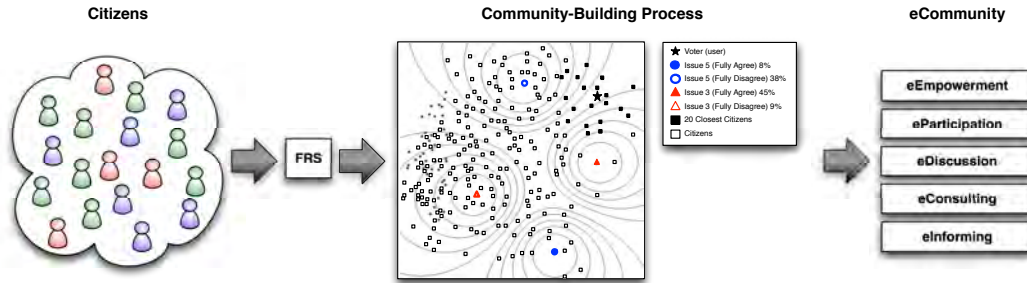


Figure 6.9: eParticipation Maturity Model for eCommunity

eConsulting. Citizens are given the opportunity to fill out questionnaires about their preferences and receive, in return, customized recommendations of various communities that are close to their preferences.

eDiscussion. This level of involvement uses a bi-directional information channel and provides citizens and government with the ability to establish discussion channels and create virtual communities by building citizen communication centers.

eParticipation. The creation of thematic communities and social networks among citizens allows for interaction and participation through social media, potentially crossing geographical and political boundaries. Contacting people with similar profiles, building exchange platforms, and stimulating participation will enrich the information and knowledge-based society in the future.

eEmpowerment. The highest level of participation can provide citizens with the opportunity to express their will on various initiatives proposed by civic networks.

Figure 6.9 shows the schema of the maturity model for *eCommunity* used in the *SmartParticipation* project.

6.4 Remarks on the SmartParticipation Project

In this chapter, five levels of participation based on the work of Tambouris *et al.* [2007] have been proposed. These participation levels includes Web 2.0 concepts

and emphasize community-building processes to enhance participation. Each of the participation levels can be evaluated independently. The creation of political communities and social networks among citizens allows for interaction and participation through social media, potentially crossing geographical and political boundaries.

Contacting people with similar political profiles, building exchange platforms, and stimulating participation will enrich the information and knowledge-based society in the future. The *SmartParticipation* project can be used to evaluate whether candidates really act the way they claim they will. The FRS can display their location in the bi-dimensional map as candidates and shows their changes in position during their political engagement as elected officials, allowing voters to easily understand politicians' behavior.

6.5 Further Readings

- **Tambouris *et al.* [2011].** “This book constitutes the refereed proceedings of the Third International Conference on Electronic Participation, ePart 2011, held in Delft, The Netherlands, in August/September 2011” (abs.).
- **Tambouris *et al.* [2010].** “This volume constitutes the refereed proceedings of the Second International Conference on Electronic Participation, ePart 2010, held in Lausanne, Switzerland, in August/September, 2010” (abs.).
- **Shark & Toporkoff [2009].** “To help local governments navigate the path through *eGovernment* implementation towards enhanced citizen engagement, Public Technology Institute (PTI) and ITEMS International introduce beyond *eGovernment*” (abs.).
- **Tambouris & Macintosh [2009].** “This book constitutes the refereed proceedings of the First International Conference on eParticipation, ePart 2009 held in Linz, Austria in August/September 2009” (abs.).

Part IV

Evaluation Framework

Evaluation Framework

In the last few years, eParticipation has been addressed more often by the academic world. Similarly, different approaches for evaluation have been proposed. In this chapter, the reader is presented with a quantitative method for the evaluation of different *eParticipation* projects. A number of voting advice applications (VAAs), used in *eDemocracy*, are evaluated using the proposed framework. The chapter is structured as follows: First, Sect. 7.1 briefly describes the framework used for evaluation. Then, Sect. 7.2 presents a quantitative evaluation of the different participation levels introduced in Sect. 6.1.1. Additionally, Sect. 7.3 shows the evaluation of different VAAs. Section 7.4 presents some remarks about the evaluation framework introduced. Finally, further readings are presented in Sect. 7.5.

7.1 eParticipation Evaluation Framework

In the first part, a number of existing frameworks proposed by different authors together with a brief description of the framework proposed are presented. Then, different participation levels proposed in the academic literature and a detailed description of the levels of participation that are part of the framework described in this section are discussed. Finally, a number of ICT tools for *eParticipation* are identified, and their levels of participation and participation areas are also described.

7.1.1 Framework Description

In academic literature, different frameworks for the evaluation of *eParticipation* projects have been proposed. In the work of Panopoulou *et al.* [2010], six frameworks have been identified. They are summarized in Table 7.1, which also presents a brief description of a number of frameworks for evaluation of *eParticipation*.

Table 7.1: eParticipation Evaluation Frameworks

Authors	Brief Description of Research
Smith <i>et al.</i> [2011]	This paper presents a framework for evaluating <i>eParticipation</i> , distinguishing between internal project components and external moderators and between the front and back regions of <i>eParticipation</i> from a governance perspective.
Macintosh & Whyte [2008]	The paper seeks to demonstrate the use of a range of perspectives and methods to evaluate <i>eParticipation</i> initiatives.
Macintosh [2004]	In this work, the authors present a characterization framework for <i>eParticipation</i> .
Kalampokis <i>et al.</i> [2008]	In this paper, the authors make an attempt to model the domain of <i>eParticipation</i> using a set of Unified Modeling Language (UML) packages and class diagrams.
Tambouris <i>et al.</i> [2007]	In this paper, the authors present a framework for assessing <i>eParticipation</i> projects and tools.
Rowe & Frewer [2000]	In this work, the authors present a framework for evaluation of public participation.

General Overview. In order to evaluate *eParticipation* projects, a simple framework is proposed that includes two main dimensions. The first dimension specifies the different participation areas. In this work, three main participation areas defined by the eGovernment framework (i.e., *eCommunity*, *eCollaboration*, and *eDemocracy*) are used. Nevertheless, the model proposed can be implemented in other participation areas.

The second dimension used by this framework takes into account the participation levels, which include the following: *eInforming*, *eConsulting*, *eDiscussion*, *eParticipation* and *eEmpowerment*. These levels were presented in a previous chapter (refer to Sect. 6.1.1). In the work of Tambouris *et al.* [2007], 20 different participation areas have been identified; e.g., community informatics, community-building, CWE, and citizenship education, among others.

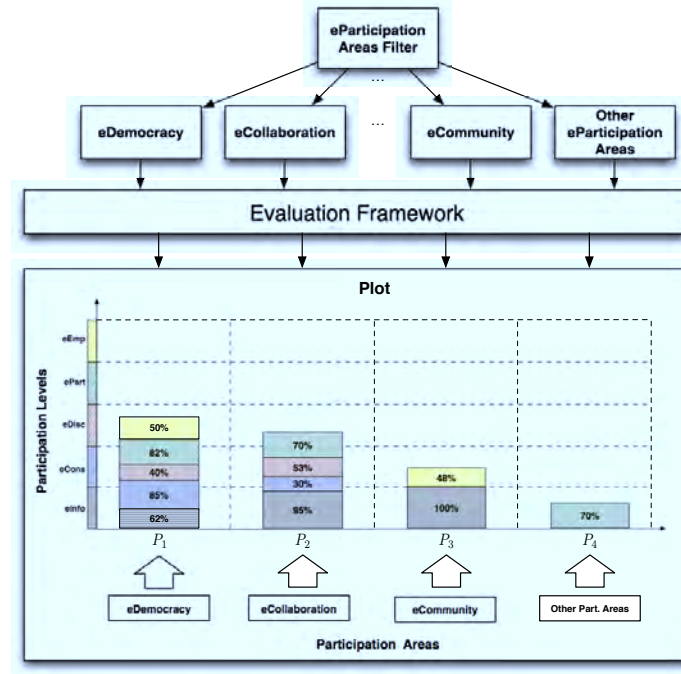


Figure 7.1: eParticipation Evaluation Output Example

Example 7.1 In order to visualize the evaluation provided by the framework, Fig. 7.1 illustrates an instance output of four eParticipation projects (P_1 to P_4) that belong to different participation areas.

In the previous example, each of these projects belongs to different participation areas. The evaluation shows the performance of each level by percentage. This example shows the following results: project P_1 has all participation levels (*eInforming*, *eConsulting*, *eDiscussion*, *eParticipation*, and *eEmpowerment*); project P_2 has four participation levels (*eInforming*, *eConsulting*, *eDiscussion*, and *eParticipation*); project P_3 has two levels (*eInforming* and *eEmpowerment*); and project P_4 has only one level (*eParticipation*).

The evaluation framework presented allows one to identify the strengths and weaknesses for the various levels. Each participation area and participation level are independent from on each other. The figure also shows that projects can focus on various participation levels. In the example shown, P_4 has a focus only on the *eParticipation* level. The quantitative approach used by the evaluation framework is presented in more detail in the following sections.

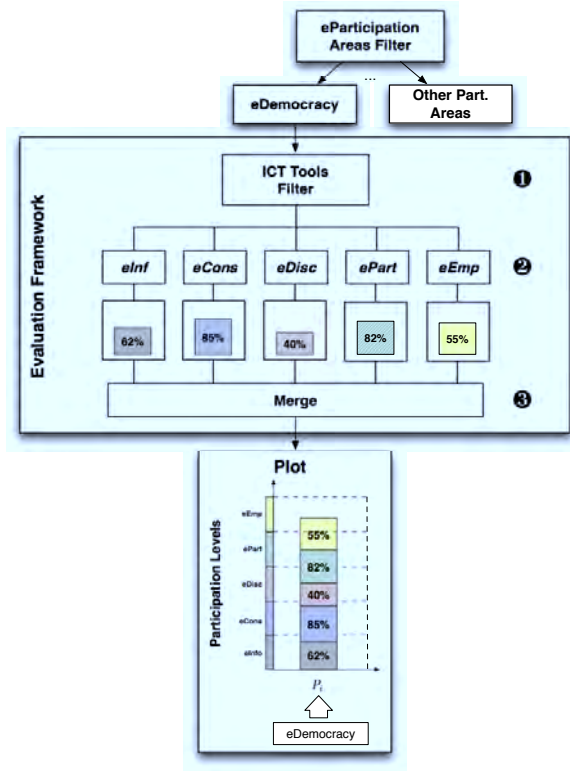


Figure 7.2: Evaluation Framework for eDemocracy

Framework Description. In this work, the author presents the evaluation of a number of VAAs, which are defined by Meier [2012] to be part of *eDemocracy*. Nevertheless, the model proposed can be implemented in other participation areas. Section 7.3 shows in more detail the evaluation of VAAs. The framework proposed in this work uses the participation levels described in Sect. 6.1.1, which are *eInforming*, *eConsulting*, *eDiscussion*, *eParticipation*, and *eEmpowerment*. The evaluation of each *eParticipation* areas includes three steps. In the first step, all the ICT tools are identified and filtered into each of the five participation levels. In the second step, a quantitative method is used to evaluate all the ICT tools identified; this quantitative method is described in detail in Sect. 7.2. Finally, in the third step, the results of the evaluation are merged and displayed.

Example 7.2 Figure 7.2 shows an example of the evaluation of an *eParticipation* project for *eDemocracy*.

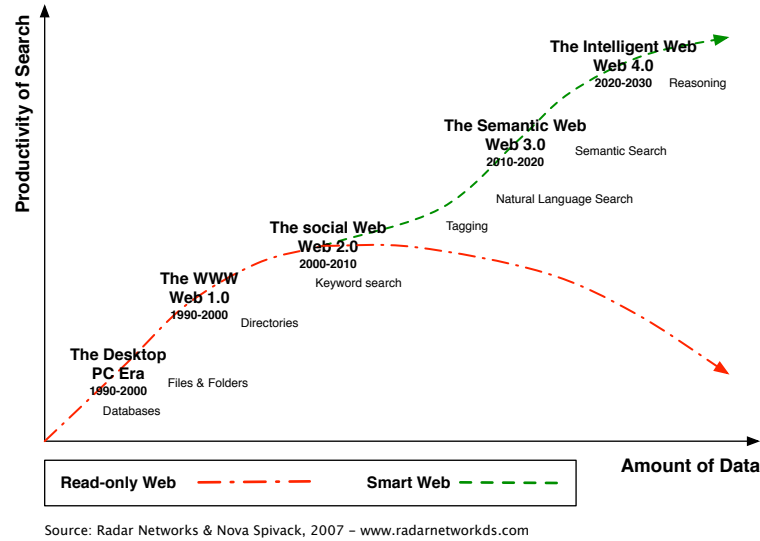


Figure 7.3: Web Evolution

7.2 Evaluation of Participation Levels

In order to evaluate the different participation levels described in Sect. 6.1.1, the framework proposed includes three components: Web evolution, media richness, and communication channels. Each of the dimensions are described in more detail below.

Web Evolution. Finding information on the World Wide Web is not an easy task. One of the main problems is the exponential growth of data available on the Internet, which can be considered as an almost infinite, non-structured, and evolving network. In academic literature, and in order to describe the evolution of the Web, two main descriptors have been used: Web X.0 (e.g., Murugesan [2009a]) and Web X.Y (e.g., Weber & Rech [2009]). The latter provides a higher granularity of each “version” of the Web. In this work, the descriptor Web X.0 is used for simplicity. It is described graphically in Fig. 7.3 and includes a brief description of the technologies used and the ratio of amount of data vs. the productivity of searches. It shows that, to increase the productivity of searches due to the increase in data available, higher standards of Web development have to be implemented. In this work, no extra argumentation will be made to support

the need for a higher standards of Web development. The evaluation framework uses the evolution of the Web as a main feature. Figure 7.3 identifies four stages of the evolution of the Web. A so-called PC era, where data was managed on local PCs and databases, took place before the Web era. The first stage of the Web era is Web 1.0 (World Wide Web), a system of hypertext documents accessed via the Internet. It is also known as the Information-centric Web, Web of cognition, and read-only Web. At this stage, the content was managed privately, and only the administrator could modify the content. The following technologies related to Web 1.0 have been identified:

- **Core Web Protocols:**

- HyperText Markup Language (HTML): markup language for displaying Web pages.
- Hypertext Transfer Protocol (HTTP): an application protocol.
- Uniform Resource Identifier (URI): a string of characters used to identify a name or a resource.

- **Server-side Scripting:**

- Active Server Pages (ASP): server-side script engine for dynamically generated Web pages.
- PHP: server-side scripting language designed to produce dynamic Web pages.
- JavaServer Pages (JSP): a technology that helps to create dynamically generated Web pages.
- Computer-generated Imagery (CGI): application to create or contribute to images.
- Perl: a high-level, general-purpose, interpreted, dynamic programming language.

- **Client-side Scripting:**

- JavaScript: prototype-based scripting language.

- Visual Basic Scripting Edition (VBScript): active scripting language.
- Flash: multimedia platform used to add animation, video, and interactivity to Web sites.
- **Downloadable Components:** ActiveX / Java.
 - ActiveX: framework for defining reusable software components.
 - Java: set of several computer software products and specifications.

The second stage is Web 2.0 (Social Web). It is also known under different names, such as Wisdom Web, People-centric Web, Participative Web, and Read/Write Web. It includes content sharing, social networks, democratization of information, and participation-oriented tools. At this level, not only are the administrators able to manage the content, but the consumers of the content are considered the source with the ability to create, update, and erase content. The following technologies related to Web 2.0 have been identified:

- **Blogs:** discussion or informational sites published on the Web.
- **Social Networks:** online services that focuses on building social interactions.
- **Wikis:** sites that allow their users to add, modify, or delete their content via a Web browser.
- **Communication Tools:** e.g., email, online chat, instant messaging.
- **Mashup:** site, or Web application, that uses and combines data, presentation, or functionality from two or more sources to create new services.
- **Really Simple Syndication (RSS):** family of Web feed formats used to publish frequently updated works.
- **Folksonomies:** system of classification derived from collaboratively creating and managing tags.
- **Tag:** non-hierarchical keyword or term assigned to a piece of information.

The third state, Web 3.0 (Semantic Web), is also known as a content-oriented Web, semantic-based Web, Web of cooperation, and context-sensitive Web. It

Evaluation Framework

promotes common data formats by encouraging the inclusion of semantic content and aims at converting unstructured and semi-structured Web documents into the so-called Web of data. Web 3.0 refers to the formats and technologies that enable it, are specified as W3C standards, and include the following:

- **Resource Description Framework (RDF):** general method for describing information.
- **RDF Schema (RDFS):** set of classes with certain properties that uses the RDF extensible knowledge representation language.
- **Simple Knowledge Organization System (SKOS):** family of formal languages designed for representation of thesauri, classification schemes, taxonomies, subject-heading systems, or any other type of structured controlled vocabulary.
- **SPARQL Protocol:** RDF Query Language.
- **Notation3 (N3):** designed with human-readability in mind.
- **N-Triples:** format for storing and transmitting data.
- **Turtle:** Terse RDF Triple Language.
- **Web Ontology Language (OWL):** family of knowledge representation languages.

According to the World Wide Web Consortium (W3C) [2012], “The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries”; thus, Web 3.0 is leading us towards the so-called Intelligent Web. Figure 7.4 presents the Semantic Web Stack to illustrate the architecture of the Semantic Web. The last state has been defined as Web 4.0 (Intelligent Web). It is also known as the agent-centric Web. At this state, the services provided are meant to be autonomous, proactive, content-exploring, self-learning, and collaborative, and content agents will include maturity technologies for semantics, reasoning, and artificial intelligence (AI). Despite the Web having not even reached maturity in Web 3.0, Web 4.0 has been proposed in the evaluation of future applications.

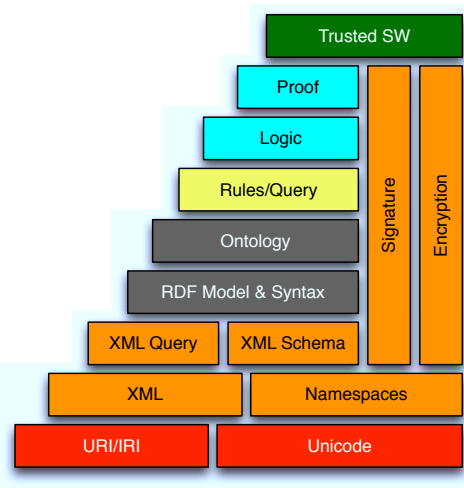


Figure 7.4: Semantic Web Stack

Media Richness. The second feature used to evaluate *eParticipation* projects is based on content richness. The model proposed in this work uses the Media Richness Theory proposed by Daft & Lengel [1984], which was primarily used to describe and evaluate communication media within organizations. A brief summary of the hierarchy of the Media Richness Theory is presented in Table 7.2.

Table 7.2: Characteristics of Media that Determine Richness of Information. Adapted from Daft & Lengel [1984]

Information Richness	Medium	Feedback	Channel
1. Highest	1. Face-to-Face	1. Immediate	1. Visual, Audio
2. High	2. Telephone	2. Fast	2. Audio
3. Moderate	3. Written, Personal	3. Slow	3. Limited Visual
4. Low	4. Written, Formal	4. Very Slow	4. Limited Visual
5. Lowest	5. Numeric, Formal	5. Very Slow	5. Limited Visual

In order to evaluate the content provided by the different *eParticipation* projects, a hierarchical model deducted from the complexity of media and Media Richness Theory is presented in Fig. 7.5. It starts at static media types (text and image), then moves to dynamic media types (audio and video), and, finally, to interactive media types (e.g., interactive television, interactive narrative, interactive advertizing, video games, social media, virtual reality, etc.). This is a five-level hierarchical model based on the supported media by various platforms on various ICT tools.

Evaluation Framework

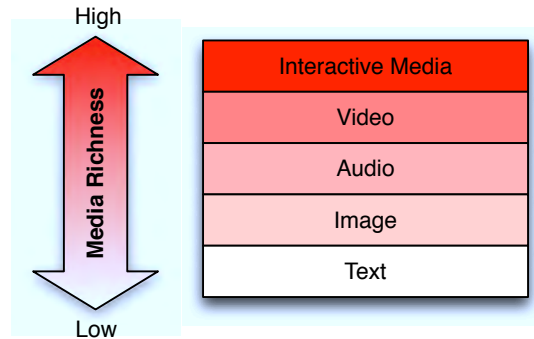


Figure 7.5: Media Richness

Communication Channels. The third feature used to evaluate *eParticipation* projects is based on two types of communication channels: asynchronous (different place/different time) and synchronous (different place/same time). The model proposed considers that synchronous channels facilitate collaborative processes. Table 7.3, adapted from Andriessen [2003], provides a non-exhaustive search of ICT tools used for various types of services.

Table 7.3: Types of Collaboration Technologies

	Support for Asynchronous Communica- tions.	Support for Synchronous Communica- tions.
Communication Systems	1. Fax 2. Email 3. Voice mail 4. Video mail	1. Telephone/mobile phones 2. Audio systems (e.g., micro, speaker) 3. Audio systems (e.g., camera, projector) 4. Chat systems
Information Sharing Systems	1. Document sharing systems 2. Message boards	1. Tele-consultant systems 2. Co-browsers ¹
Cooperation Systems	1. Document co-authoring (e.g., Wikipedia ²)	1. Shared CAD ³ 2. Whiteboards 3. Word processor 4. Spreadsheet (e.g., Open Office ⁴)

Continued on next page

¹Co-browsing: <http://cobrowsing.me>

²Wikipedia: <https://www.wikipedia.org>

³Shared CAD: <https://www.sharecad.org>

⁴Open Office: <https://www.openoffice.org>

Evaluation of Participation Levels

Table 7.3 – continued from previous page

	Support for Asynchronous Communica- tions.	Support for Synchronous Communica- tions.
Coordination Systems	1. Ground calendar 2. Shared planning (e.g., Zoho planner ¹) 3. Shared workflow management systems 4. Event manager 5. Subgroup spaces (e.g., Yahoo groups ² , Google groups ³)	1. Notification systems (e.g., Active Batch ⁴)
Social Encounter System	1. Social networking sites (e.g., MySpace ⁵ , Facebook ⁶ , Flickr ⁷)	1. Media spaces (e.g., IMVU ⁸) 2. Virtual reality (e.g., Second Life ⁹)

Quantitative Evaluation. The quantitative method presented in this section is a modified version of the framework proposed by Drobnjak [2012] and defined for evaluation of social network platforms. The objective of the evaluation method presented herein is to provide a quantitative analysis that takes into consideration the participation levels described in Chap. 6 and the components introduced in Sect. 7.2. One of the advantages of the model presented is the flexibility to add more components, which depend on the needs of the evaluation. In this work, the evaluation of three components is proposed, Web evolution, media richness, and communication channels. Equation (7.1) shows the quantitative evaluation of the i -th participation level.

$$PL_i = \frac{\sum_{j=1}^n \frac{\sqrt{we_j^2 + mr_j^2 + cc_j^2}}{\sqrt{n_c}}}{n} \quad (7.1)$$

where n_c corresponds to the number of components used for the evaluation. In this work, three components are defined.

¹Zoho Planner: <https://www.zoho.com>

²Yahoo Groups: <http://groups.yahoo.com>

³Google Groups: <https://groups.google.com>

⁴ActiveBatch: <http://www.advsyscon.com>

⁵MySpace: <http://www.myspace.com>

⁶Facebook: <http://www.facebook.com>

⁷Flickr: <http://www.flickr.com>

⁸IMVU: <http://www.imvu.com>

⁹Second Life: <http://secondlife.com>

Evaluation Framework

Web Evolution. The first component corresponds to the Web evolution of the j -th ICT tool(we_j), and it is defined as follows:

$$we_j = \sum_{k=1}^{level_{max}} wel_k = 1$$

where wel_k is the k -th level of Web evolution, and $level_{max} = 3$, which represents the highest level of Web development to be evaluated. In this section, only three levels are used: Web 1.0, Web 2.0, and Web 3.0. The values of wel_k have to fulfill the following constraint in the previous expression. Additionally, to guarantee that the presence of higher levels of Web evolution provide a comparative advantage to lower levels, the following expression is proposed to define the values of wel_k :

$$\sum_{k=1}^{level_{max}} wel_k = \sum_{k=1}^{level_{max}} k * \beta = 1, \text{ where: } \beta = \frac{1}{\sum_{k=1}^{level_{max}} k} = 1/6$$

Table 7.4 presents the values of wel_k for the k -th media richness level. It is clear that the highest levels have a better ranking compared with the lowest levels.

Table 7.4: Web Evolution Levels

<i>k</i> -th Level	Type of Media	Value
3	Web 3.0	$wel_3 = 3 * \beta = 3/6 = 1/2$
2	Web 2.0	$wel_2 = 2 * \beta = 2/6 = 1/3$
1	Web 1.0	$wel_1 = 1 * \beta = 1/6$

The values proposed in Table 7.4 are assumed by the author of this work to guarantee the importance of having higher standards of Web evolution implemented. The expression presented can be used for different goals and it can include more levels. An example of another Web evolution level is Web 4.0, which can be considered to have higher importance compared to previous levels.

Another example of Web evolution is the so-called Web 2.5, also known as the mobile Web. It was defined by Weber & Rech [2009] as a Web that is “always on” for users who carry mobile devices connected to the Internet. Some examples of this type of definition are the applications for mobile devices provided by

Facebook, Twitter, Google+, and LinkedIn, among others. The questions here is: how important is this for our evaluation.

Another question arises at this point, and it is the fact that different Web evolution levels include more than one standards. In the case of Web 3.0, its definition in the Semantic Web Stack (see Fig. 7.4) includes degrees of development that can also be used in this evaluation model.

For simplicity purposes, this work considers each Web evolution level as a group. The evaluation presented in Sect. 7.3 shows that, if a participation project uses any of the standards of the Semantic Web Stack, it will be considered part of the third level in the evaluation model.

Example 7.3 *To illustrate the evaluation using the Web evolution as an example, two sites (A and B) are used. Site A is built with HTML ($we = 1/6$), and site B includes RDF ($we = 1/6 + 1/2 = 0.66$).*

Media Richness. The second component for the evaluation method corresponds to media richness of the j -th ICT tool (mr_j), and it is defined as follows:

$$mr_j = \sum_{k=1}^{level_{max}} mrl_k = 1$$

where mrl_k is the k -th level of Media Richness, and $level_{max} = 5$, which represents the highest level of media richness to be evaluated. In this section, five levels are used: text, image, audio, video, and interactive video. The values of mrl_k have to fulfill the constraint previously expressed. Additionally, to guaranteed that the presence of higher levels of media richness provide a comparative advantage against lower levels, the following expression is proposed to define the values of mrl_k :

$$\sum_{k=1}^{level_{max}} mrl_k = \sum_{k=1}^{level_{max}} k * \beta = 1, \text{ where: } \beta = \frac{1}{\sum_{k=1}^{level_{max}} k} = 1/15$$

Table 7.5 presents the values of mrl_k for the k -th media richness level. It is clear to see that the highest levels have a better ranking compared with the lowest levels. The values proposed in Table 7.5 are assumed by the author of this work to guarantee the importance of having higher levels of media richness implemented.

Evaluation Framework

Like the Web evolution component, this evaluation method is flexible enough to use different values at different levels.

Table 7.5: Media Richness Levels

<i>k</i> -th Level	Type of Media	Value
5	Interactive Media	$mrl_5 = 5 * \beta = 5/15$
4	Video	$mrl_4 = 4 * \beta = 4/15$
3	Audio	$mrl_3 = 3 * \beta = 3/15$
2	Image	$mrl_2 = 2 * \beta = 2/15$
1	Text	$mrl_1 = 1 * \beta = 1/15$

Example 7.4 *To illustrate the use of media richness for evaluation, three sites that provide the ICT tool “chat room” are used. Site A provides text-only chat ($mr = 1/15$), site B provides audio-only chat ($mr = 3/15$), and site C provides text, audio, and video chat ($mr = 1/15 + 3/15 + 4/15 = 8/15$).*

Communication Channel. The third component corresponds to the communication channel of the j -th ICT tool (cc_j). It is a binary value defined as follows:

$$cc_j = \begin{cases} 1, & \text{if ICT tool is Synchronous} \\ 0, & \text{if ICT tool is Asynchronous} \end{cases}$$

The evaluation method presented gives the possibility of assessing, in different ways, the components proposed. In the case of communication channels, the method provides binary values when having either synchronous or asynchronous communication channels.

Example 7.5 *Two sites that provide customer support are used to illustrate the use of communication channels for evaluation. Site A provides email-based support ($cc = 0$), and site B has a chat room for support ($cc = 1$).*

7.2.1 Voting Advice Applications

The amount of data available on the Internet is growing rapidly, a phenomenon that affects not only our daily lives but politics and electoral campaigns as well.

For this reason, in recent years, the use of VAAs and different *eParticipation* projects have become popular. Thus, the advice given is of great political importance for opinion formation, decision-making, and voting behavior. VAAs are Web-based systems that provide voters with information about a political party or candidate who is closest to their preferences and political values. Voters are asked to create a political profile by filling out a questionnaire on different political issues. Then, the VAA compares their answers with the positions of parties or candidates in the system who have also completed the questionnaire.

Finally, voters are provided with a voting recommendation in the form of a list, ranking parties or candidates according to the degree of their issue congruence with the particular voter. In his work, Meier [2012] positioned VAAs as part of *eDemocracy* in a stage defined as *eDiscussion*; where, prior to a vote or election, citizens could enhance their own opinion-forming process by requesting not only information but also opinions and evaluations. VAAs are quite diverse; they vary in design as well as in the features they offer, but, in the end, they all share the same key functions. According to Ladner *et al.* [2010b], the first operational VAA was the Dutch project StemWijzer [2012]. It went online for the first time in 1998 and provided voting advice to 250,000 people. In 2006, this figure had exploded to 4.7 million people who received voting advice, which represented 40% of the Dutch electorate (Walgrave *et al.* [2008]). Fivaz & Felder [2009] also proffer clear evidence of the increasing popularity of VAAs, and Wagner & Ruusuvirta [2009] examine the recommendations given by 12 VAAs in seven European countries, describing problems of effectiveness at establishing party positions, which can lead to faulty recommendations.

7.3 Evaluation of VAAs

In this section, an evaluation of 21 VAAs (see Appendix A) is shown. The results are presented in four parts: the three main components to evaluate participation levels (Web evolution, media richness, and communication channels; see Fig. 7.6), and the participation level present in each project (see Fig. 7.7). The first step is to identify the ICT tools for each level of participation. Appendix C shows a number of ICT tools that have been identified and classified according to different

Evaluation Framework

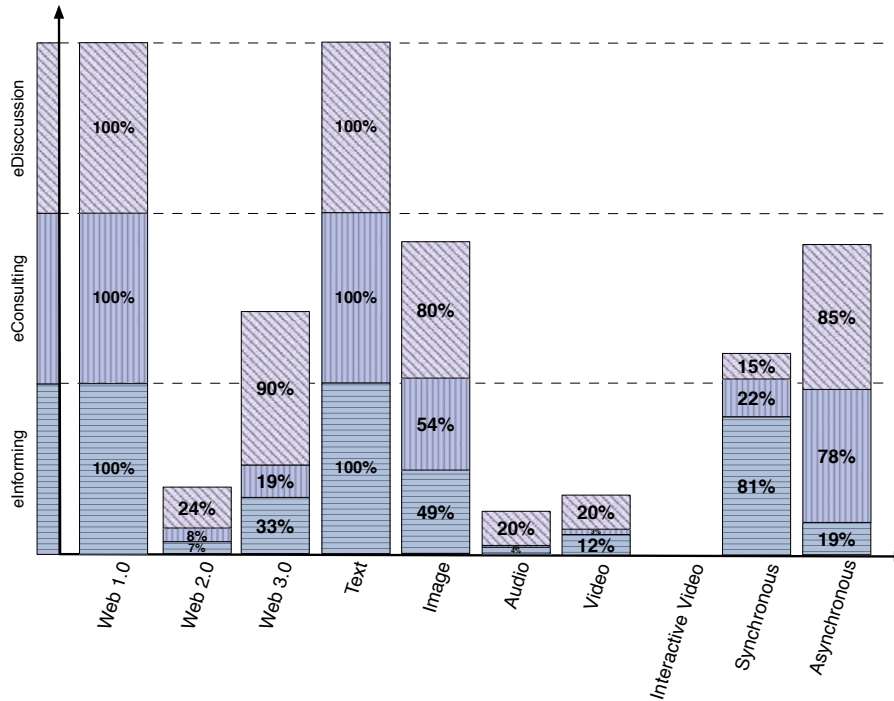


Figure 7.6: Components Evaluation

areas of *eParticipation*. Table C.1 provides a non-exhaustive list of ICT tools used for eParticipation, together with a brief description. For simplicity, the analysis of components takes the mean values of all VAAs used for evaluation. These results provide a general view of all projects evaluated. Nevertheless, to have a better understanding, one must make an individual evaluation. The analysis of participation levels presents the individual evaluation of all VAAs, and the following findings are presented below.

Web Evolution. The results presented show that there is a small percentage of development on Web 2.0, but apart from that, Web 3.0 has a bigger impact on Web evolution. It suggests that “Web developers” themselves need to appear in the search engine rankings. This effect is seen in all participation levels with an emphasis on *eDiscussion*.

Media Richness. The results provided show that both text and image are used by all participation levels as primary media. The use of video has a bigger

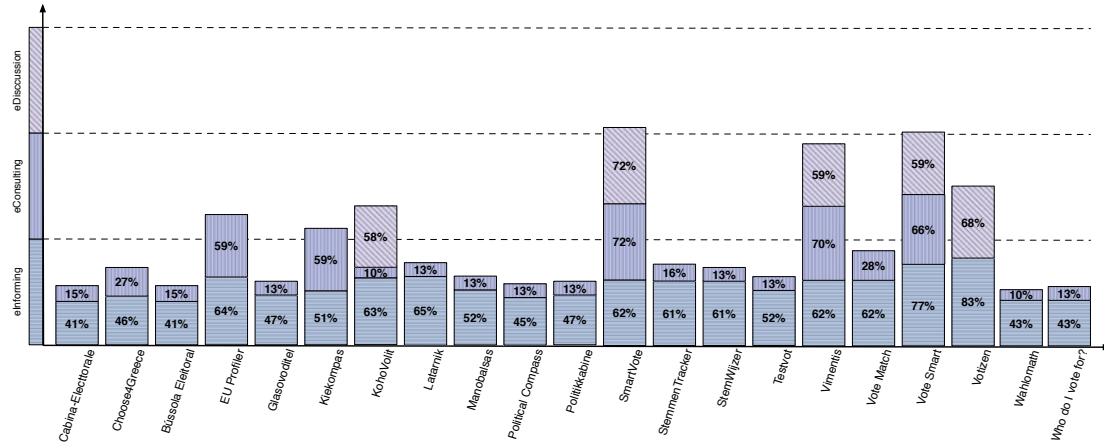


Figure 7.7: VAAs Evaluation

impact at the *eInforming* level than audio. Nevertheless, at higher levels, both are used in similar proportions. Interactive video was not considered at any level of participation.

Communication Channels. To understand the evaluation of communication channels (synchronous and asynchronous, shown in Fig. 7.6), the following services were defined as synchronous: presentation of information, contact including synchronous channels (e.g., phone), synchronous profile generation, synchronous community creation, and synchronous recommendations, among others. Additionally, the following services were defined as asynchronous: blogs, asynchronous profile generation, contact forms, asynchronous community creation, and asynchronous recommendations, among others. The results show that asynchronous services are used mainly for *eConsulting* and *eDiscussion*, and synchronous services are used mainly for *eInforming*. Figure 7.6 summarizes the evaluation of the three components mentioned above.

Participation Levels. According to the proposed framework, the highest participation level reached is *eDiscussion*. There are no ICT tools that can be considered as part of *eParticipation* or *eEmpowerment*. The first consideration to be made is how to filter each ICT tool for each participation level. The following unidirectional and informative ICT tools have been identified for *eInformation*:

Evaluation Framework

home pages, contacts, help, about us, campaigns, team, RSS feeds, newsletters, methodology, search engine, and FAQs, among others. The ICT tools that are used for *eConsulting* include political profile, recommendation of political parties, political landscapes, and recommendation of candidates.

Finally, the ICT tools that are used for *eDiscussion* are site blogs, blogs of politicians, community creation, discover friends that vote, discover races you can affect, and win votes for candidates. The evaluation shows that only 5 out of 21 VAAs have reached the level of *eDiscussion*. Figure 7.7 summarizes the evaluation of participation levels for all VAAs.

Example 7.6 *To better understand the evaluation method introduced in this chapter, an example of evaluation of the smartvote project is presented in Appendix B. The example shows the components that are characterized for the different participation levels. The final evaluation is computed using the mean value of all components that are part of the same participation level. Finally, Fig. B.1d shows that the smartvote project presented the following evaluation: *eInforming* (63%), *eConsulting* (78%), and *eDiscussion* (72%).*

7.4 Remarks on the Evaluation Framework

The framework proposed in this chapter presents a quantitative evaluation of different *eParticipation* projects based on three components: Web evolution, media richness, and communication channels. The framework can be extended depending on the objectives of the evaluation. In this chapter, the evaluation of 21 existing voting advice application projects was presented. The VAAs are considered by Meier [2012] to be part of *eDemocracy* in the eGovernment framework used in this Ph.D. thesis and developed at the University of Fribourg, Switzerland. The VAAs evaluated have reached only the first three levels of participation proposed (*eInforming*, *eConsulting*, and *eDiscussion*).

The results of the evaluation show that there is a lack of the use of the following technologies: Web 2.0, Web 3.0, audio, video, interactive video, and synchronous communication channels, which are considered by the authors to provide a competitive advantage compared with the other technologies used for the eval-

uation (Web 1.0, text, image, and asynchronous communications). Therefore, *eParticipation* is best performed on such platforms, granting an increased chance of reaching users from the targeted audience. At the same time, *eParticipation* remains highly efficient, meaning that the amount of work and time invested in transmitting information to other users from the targeted audience is kept at a minimum.

7.5 Further Readings

- **Fensel *et al.* [2011]**. “After years of mostly theoretical research, Semantic Web Technologies are now reaching out into application areas like bioinformatics, *eCommerce*, *eGovernment*, or Social Webs. Applications like genomic ontologies, semantic web services, automated catalogue alignment, ontology matching, or blogs and social networks are constantly increasing, often driven or at least backed up by companies like Google, Amazon, YouTube, Facebook, LinkedIn and others” (abs.).
- **Cedroni & Garzia [2010]**. “Voting Advice Applications (VAAs) have literally taken Europe by storm in the past decade, with millions of voters turning to these web-based tests at election time. VAAs help users casting a vote by comparing their policy preferences on major issues with the programmatic stands of political parties on such issues” (abs.).
- **Murugesan [2009b]**. “The Handbook of Research on Web 2.0, 3.0, and X.0: Technologies, Business, and Social Applications is a comprehensive reference source on next-generation Web technologies and their applications” (abs.).
- **Hitzler *et al.* [2009]**.
“The book concentrates on Semantic Web technologies standardized by the World Wide Web Consortium: RDF and SPARQL enable data exchange and querying, RDFS and OWL provide expressive ontology modeling, and RIF supports rule-based modeling” (abs.).

Part V

Implementation

Architecture and Implementation

The *SmartParticipation* platform aims to fulfill the five levels of participation presented in Chap. 6. In this chapter, the reader is presented with a general overview of the technologies, GUI, and features proposed for this platform. Additionally, a number of modules that can be used to fulfill the five participation levels are listed. The chapter is structured as follows: First, Sect. 8.1 presents the architecture used by *SmartParticipation*. Then, Sect. 8.2 briefly describes the Web framework used by the platform. Section 8.3 discusses the two types of profiles that utilize the architecture developed in this work. In Sect. 8.4, the implementation and features of the FRS are presented. Finally, further readings are presented in Sect. 8.5.

8.1 *SmartParticipation* Architecture Overview

In this section, the architecture of *SmartParticipation* is described in more detail. Two frameworks have been used to implement the platform, one for designing a dynamic Web interface (DWI) and the second for computing and visualizing recommendations by the FRS. The framework used for the development of the DWI is an open-source solution, and has been positioned as one of the most-used frameworks, according to the reports produced by Water&Stone [2011]; Water&Stone & CMS Wire [2009]. It includes a number of custom modules to develop the platform according to the different levels of participation proposed in this work

Architecture and Implementation

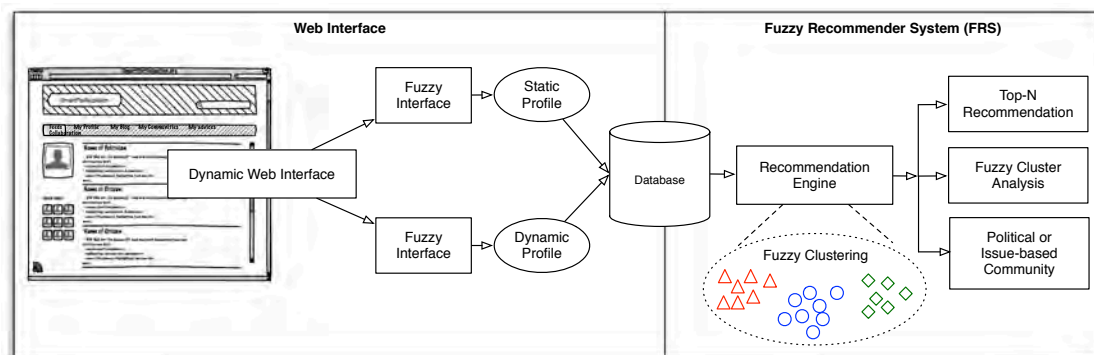


Figure 8.1: Architecture of SmartParticipation Platform

(refer to Sect. 6.1.1) and to fulfill the requirements proposed by the evaluation framework in Chap. 7.

Figure 8.1 shows a general overview of the architecture and is divided in two blocks. The right side of the architecture corresponds to the FRS and is implemented by using a MATLAB Compiler¹ that computes the dimensionality reduction, fuzzy clustering, and visualization of recommendations with a runtime engine called the MATLAB Compiler Runtime (MCR).² On the left side, a DWI, developed using the Drupal framework, is used to feed the FRS using two types of profiles: static and dynamic. A more detailed description of the two types of profiles is presented in Sect. 8.2.

Two types of user profiles are defined in this work: static and dynamic. The former is generated when a user subscribes to the system, and the latter is generated according to the activity of users. In contrast to the VAAs evaluated in the previous chapter, the inclusion of a dynamic profile gives users more options, and the enhancement of their profiles leads to better recommendations. The development of the FRS uses a separated framework, which requires that users install the application via MCR, which is clearly a disadvantage in terms of usability. Future work must include an FRS module developed exclusively for the Web framework that includes both computation and visualization of recommendations.

¹MATLAB Compiler: <http://www.mathworks.com/products/compiler/>

²MATLAB Compiler Runtime: <http://www.mathworks.com/products/compiler/mcr/>

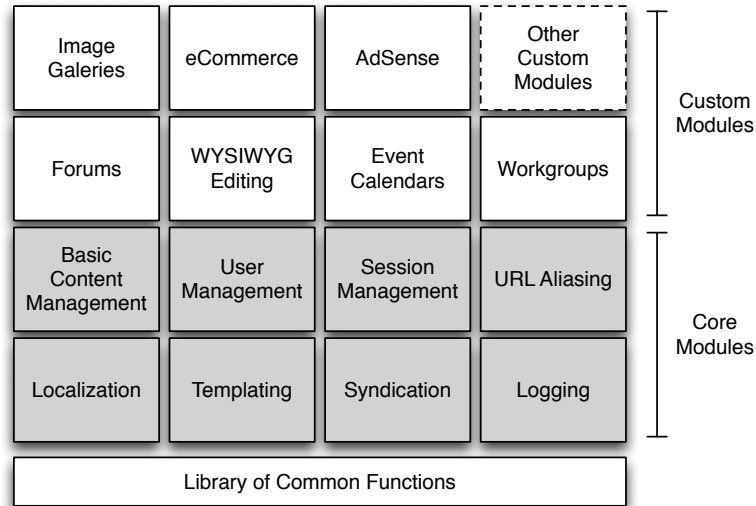


Figure 8.2: Overview of Drupal Modules. Adapted from Tomlinson & VanDyke [2010]

8.2 Web Interface

In order to implement the *SmartParticipation* platform, the first step is to select the appropriate framework that can fulfill the requirements proposed in the evaluation of *eParticipation* projects (see Chap. 7). This environment includes the use of Web. 2.0 and 3.0, synchronous communication channels, and media richness. For that reason, the Drupal content management systems (CMS),¹ an open-source software maintained and developed by a community of more than 630.000 users and developers. In the technical reports produced by Water&Stone [2011], the big three CMS, namely, WordPress, Joomla!, and Drupal, remain firmly in command of the market, according to the report presented by Water&Stone & CMS Wire [2009].

Drupal CMS is a framework with functionalities that can be added as modules, which can be enabled or disabled. It includes more than 22,000 modules developed for different purposes, such as social networking, Semantic Web, workgroups, user management, and forums. Figure 8.2 shows an overview of the main core and custom modules (not all modules are shown). To guarantee that the *SmartParticipation* platform fulfills the requirements of the evaluation frame-

¹Drupal: <https://drupal.org>

Architecture and Implementation

work (Web evolution, media richness, and communication channels) described in Chap. 7, a number of modules developed for Drupal have been used. A description of the main modules used by the *SmartParticipation* platform are presented as follows:

Web Evolution:

- ***Schema.org***. This module was developed by Corlosquet [2012] and enables the collections of schemas available at Schema.org.¹ The schemas are a set of “types,” each associated with a set of properties arranged in a hierarchy. Schema.org is recognized by major search engines such as Bing, Google, and Yahoo. This module allows RDF schemas for different fields, which are part of the content types developed.
- ***Taxonomy***. This module was developed by Drupal.org [2012], and it is one of the core modules used for categorization of articles using keywords or terms assigned to information known as tags.

Media Richness:

- ***Video Embed Field***. This module was developed by Caldwell [2012] and allows users to upload video media types and include descriptions.
- ***AudioField Module***. This module was developed by Zoubi & Bačelić [2012] and allows users to upload audio media types and include descriptions.

Communication Channels:

- ***DrupalChat***. This module was developed by Srivastava [2013] and allows visitors to chat with each other privately or in a public chatroom.
- ***Video Chat***. This module was developed by Trojan [2012] and adds a fully hosted free video chat to a Drupal website.

A number of modules that are used for the different participation levels proposed in this Ph.D. thesis (see Sect. 6.1.1) are described as follows:

¹Schema.org: <http://schema.org>

eInforming:

- ***Content Types.*** In Drupal, each information page is called a node, and each node belongs to a single content type. Each content type can have different features, templates, fields, and media types. Drupal comes with basic content types (e.g., article, basic page, forum) out of the box, as well as customized ones (e.g., poll, Web forms).
- ***CKEditor editor.*** This module was developed by Walc [2013] and allows Drupal to replace text area fields with the CKEditor,¹ which is a visual HTML editor.
- ***Views.*** This module was developed by Miles [2013]. It is one of the most-used modules in Drupal and includes a number of features such as sort data fields, filter content, contextual filtering, add relationships, creation of blocks, and creation of pages.

eConsulting:

- ***AJAX Poll.*** This modules was developed by Haug [2011] and allows users to vote on polls without reloading the page. It works with the normal poll module included with Drupal core.
- ***Advanced Poll.*** This module was developed by Kennedy [2012] and and provides multiple voting systems, decision-making tools, and management options. Voting systems implemented are basic polls, approval voting, Borda count, and instant-runoff voting.

eDiscussion:

- ***Organic Groups.*** This module was developed by Burstein [2013] and enables users to create, manage, subscribe to, and maintain groups.
- ***User Relationships.*** This module was developed by Karshakevich [2013] and enables administrators to create relationship types (e.g., friend, coworker). Relationship types can be set up to be one-way or mutual.

¹CKEditor: <http://ckeditor.com>

Architecture and Implementation

- ***Comment Easy Reply.*** This module was developed by Colella [2013] and allows users to directly reply to comments posted on a content type.

eParticipation:

- ***Wikitools.*** This module was developed by Gordon [2011] and provides a wiki-like environment and includes the following features: node creation, node search, redirect, and unique titles, among others.
- ***Advanced Forum.*** This module was developed by Troky & Webber [2013] and enhances the capabilities of Drupal's core forum module.
- ***Filedepot.*** This module was developed by Lang [2013] and is a full-featured document management module, which integrates file management support by role and user-based security.
- ***Calendar.*** This modules was developed by Agrawal [2013] and allows users to Add/Edit/View events in pop-up windows by clicking on a box in an event calendar view.

eEmpowerment:

- ***Voting API.*** This module was developed by Eaton [2013] for developers who want to use a standardized API and schema for storing, retrieving, and tabulating votes for Drupal content.
- ***Rate.*** This module was developed by Lawende *et al.* [2012] and provides flexible voting widgets for nodes and comments. It includes the following widgets: thumbs up/down, number up/down, five-star, emotion, yes/no, and slider.

The list of modules described above can be categorized as ICT tools for *eParticipation*. Appendix C provides a more detailed list of ICT tools developed for the different participation areas as part of this Ph.D. thesis (*eCollaboration*, *eDemocracy*, and *eCommunity*).

The screenshot displays a web interface for a static user profile. At the top, there is a horizontal navigation bar with eleven topic-based tabs: 'Welfare, family and health', 'Education and research', 'Migration and integration', 'Society, culture and ethics', 'Finances and taxes', 'Economy and work', 'Environment, transport and energy', 'State institutions and political rights', 'Justice, police and army', and 'Foreign policy and foreign trade'. The 'Welfare, family and health' tab is currently selected. Below this, on the left side, is a vertical list of seven questions, labeled 'Question 1' through 'Question 7'. 'Question 1' is highlighted. The main content area on the right shows the details for 'Question 1', which is: 'Do you support the raising of the pensionable age to 67 for men and women?'. Below the question text, there are two horizontal fuzzy sliders. The first slider is labeled 'Question 1' and has a numerical value of 46. The second slider is labeled 'Relevance Question 1' and has a numerical value of 29. Both sliders have a small square marker indicating the current value on a scale from 0 to 100.

Figure 8.3: Static Profile GUI

8.3 User Profile

The *SmartParticipation* platform has been designed to include two types of profiles: static and dynamic. Figure 8.1 shows that both profiles are generated using DWI, with the main difference between the two profiles being that the static profile is generated when the user signs up for the system and includes a fixed definition of preferences, whereas the dynamic profile is generated when the user participates in different blogs, discussion, comments, etc. A better description of the implementation of these profiles is presented in the next section.

8.3.1 Static Profile

The static profile is generated when the user subscribes to the system and includes a general overview of his/her preferences. In this Ph.D. thesis, the profile provided by smartvote [2012b] is used and is comprised of 11 topics, each containing different questions.

Figure 8.3 shows an example of a GUI that uses the questions and topics from *smartvote*. The profile generation includes a fuzzy interface to determine the level of agreement/disagreement and relevance for a specific question. This figure shows that, for the topic “welfare, family, and health,” seven questions are included. The first question is evaluated by using two fuzzy sliders, tendency and relevance. The profiles provided by *smartvote* are considered static because

they are proposed by the developers of system and do not consider user feedback (citizens and candidates).

8.3.2 Dynamic Profile

The dynamic profile aims to enhance users' participation and to improve profile generation. As discussed in Chap. 7, the profiles in this participation projects are generated only by answering predetermined questions about different political topics. Unlike VAAs, the *SmartParticipation* platform allows users to be content generators. The dynamic profile is generated on the basis of the activity of users using the DWI. In this work, an example of a dynamic profile is presented with the implementation of a content type called a *discussion topic*. Appendix D shows an instance of a mockup of the content type used to feed the dynamic user profile. It consists of five blocks: general information, additional sources, rating, comments, and list of comments. Each block has different fields defined with RDF using Schema.org.¹ The blocks are described in more detail as follows:

Block 1 - General Information. This block includes information about content types (e.g., discussion topic, blog, and article) and includes the following fields: title, tags, username, date of publication, body, and source. The fields tags are used by the fuzzy profile to define a specific issues (refer to Sect. 5.2). Figure 8.4 is a screenshot of block 1.

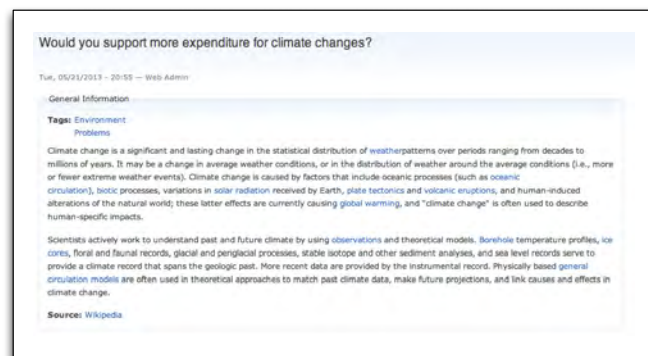


Figure 8.4: Block 1 - General Information

¹Schema.org: <http://schema.org>

Block 2 - Additional Sources. This block includes different media sources such as audio, video, and images. It contributes higher levels of media richness, which gives users a better understanding of various issues. Figure 8.5 is a screenshot of block 2.



Figure 8.5: Block 2 - Additional Sources

Block 3 - Rating. This block includes an rating tool that uses a fuzzy slider. It includes two inputs, tendency and relevance, and allows users to better describe their opinion about the topic being discussed. A user has to register with the system to be able to provide a rating, and the evaluation will be used by the FRS to better define the user's profile and preferences and for further recommendations. Figure 8.6 shows a screenshot of block 3.

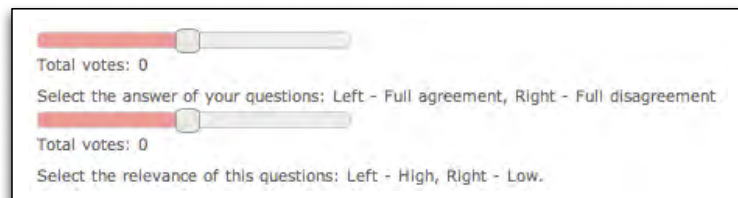
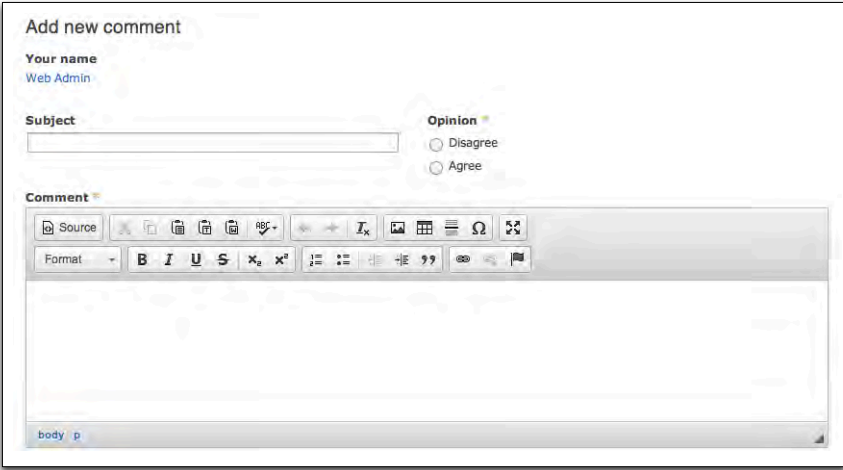


Figure 8.6: Block 3 - Rating

Block 4 - Comments. This block allows users to include comments on a specific topic. This field was developed using the CKEditor module described in

Architecture and Implementation

the previous section, which is a WYSIWYG tool used to include different types of media in the body of the comment. Additionally, a mandatory option field was included to allow a better presentation of opinions about the topic in discussion. Figure 8.7 is a screenshot of block 4.



The screenshot shows a web form titled "Add new comment". It contains the following elements:

- Your name:** A text input field with the value "Web Admin".
- Subject:** A text input field.
- Opinion:** Two radio buttons labeled "Disagree" and "Agree".
- Comment:** A large text area with a rich text editor toolbar. The toolbar includes buttons for Source, Bold (B), Italic (I), Underline (U), Strikethrough (ABC), Bulleted List, Numbered List, Indent, Outdent, Link, Unlink, and Insert.

Figure 8.7: Block 4 - Comments

Block 5 - List of Comments. This block is used to display the various comments submitted about the topic in discussion. This field was developed using the Views and Comment Easy Reply modules described in the previous section. This block has two main features and it separates the agree and disagree comments into two columns. Additionally, a reply feature puts comments in sequential order and lets users provide tags for multiple comments (two choices for default: #COMMENT_NUMBER and @COMMENT_AUTHOR). Figure 8.8 is a screenshot of block 5.



The screenshot shows a table-like layout with two columns: "Agree" and "Disagree".

Agree	Disagree
<p>Climate change is real</p> <p>Comment:</p> <p>Check this interesting facts about gobal warming.</p> <p>Source: National Geographic</p> <p>reply</p> <p>3 min 13 sec ago</p>	<p>Myths On Climate Change</p> <p>Comment:</p> <p>In contrast to your argument, chech the following facts & myths of climante change.</p> <p>Source: Prof. Robert M. Carter</p> <p>Link: Ten Facts & Ten Myths On Climate Change</p> <p>reply</p> <p>34 sec ago</p>

Figure 8.8: Block 5 - List of Comments

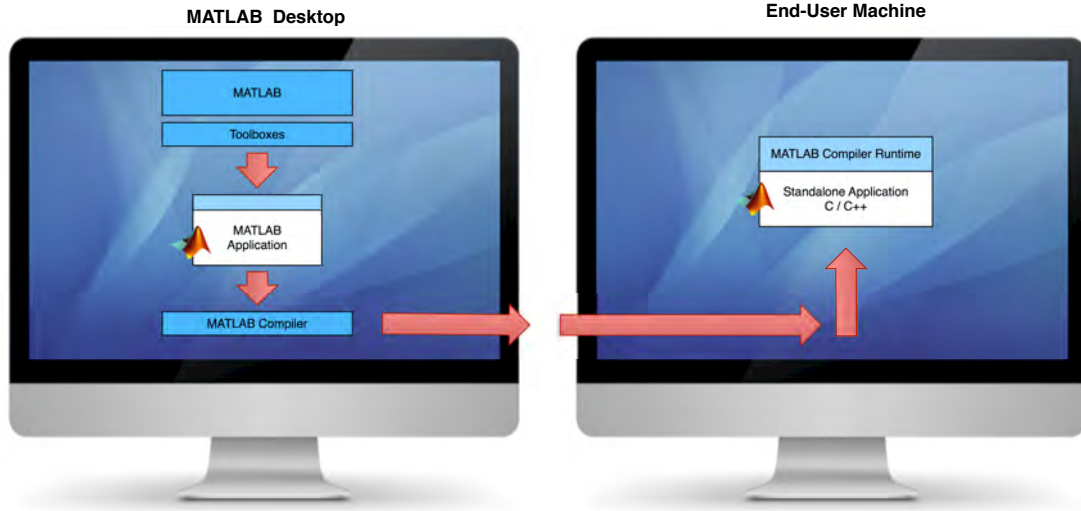


Figure 8.9: The Process of Compilation and Distribution to End Users

8.4 FRS Implementation

In this section, a description of the implementation and the main features implemented by FRS is presented. The computations of different dimensionality reduction methods, fuzzy clustering, and the GUI for visualization of recommendations have been implemented using the MATLAB Compiler, which generates standalone applications and C/C++ shared libraries from MATLAB code and distributes them to end users who do not have a MATLAB installed in their local computers. Figure 8.9 shows the compilation and distribution process of a MATLAB application. Besides the compiler package, MATLAB provides two additional packages: MATLAB Builder NE,¹ which lets one integrate these components into larger .NET, COM, and Web applications, and MATLAB Builder JA,² which enables one to create Java classes. The FRS developed in this work includes the following blocks: type of recommendation, type of relevance, dimensionality reduction method, political parties, topics, and recommendation GUI. Figure 8.10 shows the GUI of the FRS produced using the MCR. Each of the blocks developed by the FRS, which are shown in Fig. 8.10, and their main functionalities are briefly described below:

¹MATLAB Builder NE: <http://www.mathworks.com/products/netbuilder/>

²MATLAB Builder JA: <http://www.mathworks.com/products/javabuilder/>

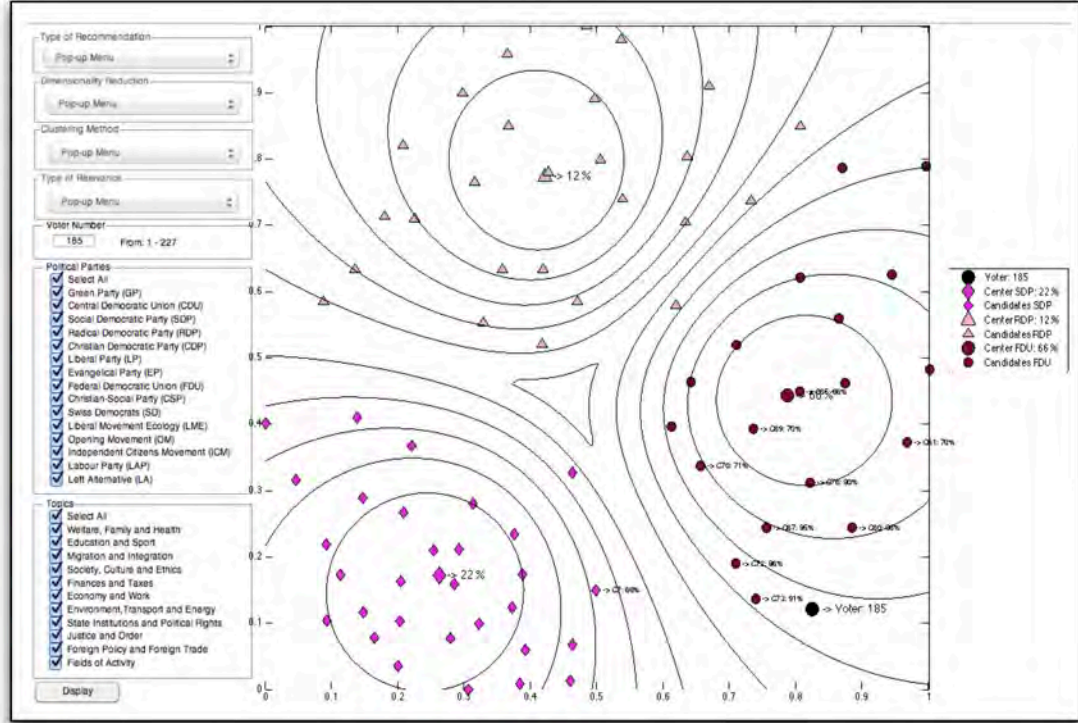


Figure 8.10: FRS GUI

Type of Recommendation. This field allows users to select the type of recommendation that will be displayed. It can be either fuzzy cluster analysis, top-N recommendation, political community, or issue-based community. More details about the options presented in this block are described in Sect. 5.4.

Types of Relevance. This field allows users to select the three types of relevance defined for the profile of candidates in the dataset provided by smartvote [2012b], which defines a measure of relevance for citizens only. For this reason, three methods have been implemented in this block: no-relevance, same-as-voter, and random. The no-relevance method extracts the relevance values from the profiles of all citizens. The same-as-voter method includes the same relevance of a specific voter to all candidates. Finally, the random method generates a random value of relevance for each candidate. These methods have been developed as research objectives. The option of no relevance is used in all the experiments presented in this work in order to avoid noise in candidates' profiles.

Dimensionality Reduction Method. This field allows users to select from among five reduction methods for visualization. It can be one of the following: (1) Principal Component Analysis (PCA; Pearson [1901]), (2) Sammon mapping (Sammon [1969]), (3) t-Distributed Stochastic Neighbor Embedding (t-SNE; van der Maaten & Hinton [2008]), (4) Isomap (Tenenbaum *et al.* [2000]), and (5) Locally Linear Embedding (LLE; Roweis & Saul [2000]). More details about the options presented in this block are described in Sect. 5.4.3. This field is used by the FRS proposed for evaluation and research objectives. It is not intended to be used in the production version of the *SmartParticipation* project.

Fuzzy Clustering Method. This field allows users to select three fuzzy clustering methods for visualization. It can be one of the following: (1) Fuzzy C-means Algorithm (FCM; Bezdek [1981]), (2) Fuzzy Gustafson-Kessel Algorithm (FGK; Gustafson & Kessel [1978]), and (3) Fuzzy Gath-Geva Algorithm (FGG; Bezdek & Dunn [1975]). This field is used by the FRS proposed for evaluation and research objectives. It is not intended to be used in the production version of the *SmartParticipation* project.

Political Parties. This field allows users to select from a list of political parties. This block can be used for three options of recommendations: fuzzy cluster analysis, top-N recommendation, and political community. More details about the options presented in this block are described in Sect. 5.4.

Topics. This field allows the user to select topics of interest to him/her. In the case of the FRS GUI, the topics selected are those selected by *smartvote*: (1) welfare, family, and health, (2) education and sport, (3) migration and integration, (4) society, culture, and ethics, (5) finance and taxes, (6) economy and work, (7) environment, transport, and energy, (8) state institutions and political rights, (9) justice and order, (10) foreign policy and foreign trade, and (11) fields of activity.

Recommendation GUI. This field provides users with a bi-dimensional landscape for both political parties and issues to better understand their proximity using the concept of membership degree and fuzzy clusters. Additionally, it pro-

vides a legend of the different political parties and issues with a percentage of proximity to each of the clusters generated.

8.5 Further Readings

- **Water&Stone [2011]**. “This report is about the brand strength and market share of 20 open source web content management systems. As such, it contains important information relevant to selecting a CMS, but it should not be read as a final judgment on the feature quality, stability, or a particular system’s suitability for any project” (abs.).
- **Nordin [2011]**. “This concise guide helps small teams and solo website designers understand how Drupal works by demonstrating the ways it outputs content. You’ll learn how to manage Drupal’s output, design around it, and then turn your design into a theme” (abs.).
- **Tomlinson & VanDyke [2010]**. “This book updates the most popular development reference for the newest major release of Drupal. With several new and completely-rewritten essential APIs and improvements in Drupal 7, this book will not only teach developers how to write modules ranging from simple to complex, but also how Drupal itself works” (abs.).
- **Siciliano [2008]**. “The book begins by looking at the main tools, in particular the Desktop, the Command and History Window, the Editor and the Help Browser. The selected number of functions, graphics objects, related properties and operators, considered fundamental in MATLAB, is a unique and remarkable feature of this book” (abs.).

Part VI

Conclusions

Discussion and Conclusions

In this chapter, the reader is presented with a description of the main contributions, outlook, and conclusions of this Ph.D. thesis. In the first section, the author presents a comparison of the fuzzy-based recommendation engine with the method used by *smartvote* for recommending candidates and political parties. The chapter is structured as follows: First, Sect. 9.1 presents a discussion of the fuzzy recommender system FRS compared to the *smartvote* project that points out the main contributions. Then, Sect. 9.2 gives the outlook and future work for the *SmartParticipation* project. Finally, Sect. 9.3 provides a the conclusions and a brief analysis of the research questions proposed in the introduction of this work.

9.1 Discussion

Wagner & Ruusuvirta [2009] examine the recommendations given by 12 voting advice applications (VAAs) in seven European countries. They describe the problems in effectiveness of establishing party positions, which can lead to faulty recommendations. Additionally, they show that *smartvote* includes a broader profile generation, which consists of 73 questions from 11 groups of topics, which is a clear advantage in terms of profile diversification. The second system that provides a broader user profile corresponds to Kieskompas [2012], which includes 36 questions. The rest of the VAAs analysed by Wagner & Ruusuvirta [2009] have an average of 26 questions.

Discussion and Conclusions

In order to evaluate the different existing VAAs, one must understand the algorithms used to provide recommendations. Unfortunately, few VAAs have made their algorithms public. This problem is also mentioned in the work of Wagner & Ruusuvirta [2009]. The authors assume that the VAAs follow the assumptions of the Downsian proximity model proposed by Downs [1957]. The algorithm used by *smartvote* is accessible and available online (refer to smartvote [2012a]). It gives points to candidates according to the distance between the responses of the candidates and those of the user. The concept used by *smartvote* has been the basis for other applications, such as Politikabine [2012], EU Profiler [2012], Koimipasva [2005], and Holyrood [2007]. In the work of Ladner *et al.* [2010a] other advantages of *smartvote* compared to different VAAs are presented and discussed.

Additionally, two characteristics of the profile generation proposed by *smartvote* must be taken into consideration for the analysis, and are mentioned as follows. First, *smartvote* provides two types of profiles, one for candidates and another for citizens. The candidates' profiles do not include the values of weight for each question; therefore, the questionnaire must be completed. On the other hand, citizens can include weights for each question and answer only those questions that they consider relevant. Second, the recommendations are based on a comparison of distances between the answers of candidates and those of citizens.

A simple calculation of distances is presented in Eq. (E.5) in Appendix E. In this equation, the value of weight provided by the citizens is used in the same way for both citizens and candidates. It can be considered an inclusion of noise, because it assumes candidates agree with each citizen's weight value for the recommendations since it is multiplied for the corresponding answers of citizens and candidates. A summary of the methodology used to provide the recommendations of candidates is given in Appendix E.

Contributions of SmartParticipation. This Ph.D. thesis focuses on the study of recommender systems (RSs) in *eGovernment*. In particular, the approach proposed corresponds to the fuzzy-based recommender systems used in *eParticipation*, which is suitable for the one-and-only item; here, past actions are not used for the recommendation.

The main contributions of this work, compared to the *smartvote* project, are listed as follows:

- ***Profile Generation.*** Unlike the *smartvote* system, the platform developed allows candidates to provide relevance to each question. This tool prevents the problem of adding noise to the profiles of candidates when providing a recommendation. Additionally, a fuzzy-based profile generation using sliders is included. The objective of this GUI is to provide users with a wider range of options to describe their tendencies and relevance for the issues they are concerned with.
- ***Dynamic Profiles.*** *SmartParticipation* provides new channels for discussion of various issues through the integration of dynamic profiles, which are generated on the basis of the activity of users using the DWI. Unlike VAAs, the platform developed allows users to become content generators, enhancing the participation and improving the profile generation. The recommendations would also improve, since they are not based only on static profiles as they are in most of the VAAs.
- ***Visualization of Political Parties.*** With the inclusion of fuzzy clustering methods, an alternative for the calculation of centers of political parties is presented. In general, VAAs compute the position of political parties by taking a mean average of the positions of all candidates of each political party. The same approach is used by *smartvote*. Unlike other VAAs, the centers of political parties are computed using a fuzzy clustering method, which moves the centers to where the mass of members is located.

The assumption for this approach is that the position of a political party should move along with the positions of its members. The author of this work considers that the status quo of a political party cannot be static in time; just as persons can evolve in the way they think, so can political parties.

- ***Issue-based Clustering.*** Another feature of the FRS developed for *SmartParticipation* is the provision of a bi-dimensional issue-based landscape. It allows users to better position themselves on different issues of

their interest. This recommendation approach allows communities of interest and, in the case of *eCollaboration*, helps users (candidates, government, companies, and citizens) look for citizens with specific skills.

- ***Higher Levels of Participation.*** Another objective of the platform developed is to improve citizens' participation. In Chap. 6, five levels of participation are identified and are used in this work as a reference framework. An advantage of using these guidelines is that, in contrast to other VAAs, *SmartParticipation* opens the possibility of becoming a more advanced discussion channel for political issues, not only during electoral campaigns (*eDemocracy*) but also for other ways of participation, such as *eCommunity* and *eParticipation*.

9.2 Outlook

RSs can be considered a multidisciplinary research topic that includes a wide range of areas, such as machine learning, data mining, information retrieval, human computer interaction (HCI), and data visualization, among others. In addition to the various solutions developed in this Ph.D. thesis, many ideas have been proposed and other questions remain open. These ideas and questions need to be analyzed in greater depth in future work. Some of the subjects that this work can be extended to include the following:

- ***Sentiment Analysis.*** The concept of sentiment analysis (or opinion mining) is defined by Feldman [2013] as the task of finding the opinions of authors about specific entities. The *eParticipation* platform developed includes a first approach of sentiment analysis with dynamic profiles. Nevertheless, more advanced tools for including sentiment analysis in profile generation, such as data mining, machine learning, and artificial intelligence, should be included.
- ***Context-Awareness in Recommendation.*** Contextual information could be used to improve the prediction accuracy of RSs. The use of context-awareness has been studied by a number of researchers to improve the quality of recommendations, as mentioned in the work of Adomavicius &

- Tuzhilin [2011]. Future work can include additional functionalities to support context-awareness for the *SmartParticipation* project.
- ***Self Control Platform.*** One of the problems in developing a discussion channel (e.g., forums, blocks) is how to promote communication and to avoid the censorship of administrators. A mechanism that includes “community censorship” can be implemented to guarantee the correct use of these channels. Data mining tools and a mechanism of community voting to control the misuse of comments and content could be developed. To guarantee that the voting community is not biased, the FRS can provide different points of view.
 - ***Fuzzy Voting.*** The *SmartParticipation* platform is intended to reach the highest level of participation, the so-called *eEmpowerment*, which places the final decision in the hands of the citizens. In addition to traditional crisp voting systems, an alternative, fuzzy-based, method could be tested. In the work of Côté-Real [2007], a comparison of crisp voting procedures with more general mechanisms of fuzzy preference aggregation is presented.
 - ***Evaluation of GUI.*** The platform developed in this Ph.D. thesis includes a number of GUIs, such as fuzzy clustering analysis, top-n recommendation, community-building, and dynamic profile. Future work could include an in-depth evaluation of the features provided by the platform, conducted by users. The work of Kortum [2008] describes the human factors involved in the design and implementation of non-traditional interfaces.
 - ***eElections & eVoting on SmartParticipation.*** VAAs can be considered as an additional source of information for the decision-making process in eElections and eVoting processes. In future work, *SmartParticipation* can provide the possibility of promoting candidates and/or political initiatives to enhance the political debate. The opportunity to link recommendations directly to polling stations can be addressed in future research, which will open new possibilities for research on legal, political and technical issues.
 - ***Open Government.*** The platform developed in this Ph.D. thesis presents the possibility of using the concepts of RSs used mainly in *Electronic Business (eBusiness)* for the community-building processes in *eParticipation*.

This type of solution could also contribute to so-called “Open Government.” An example of this type of project is the Swiss Open Government Data Pilot Portal (Swiss Confederation [2013]), introduced in september 2013, which contains 1625 datasets for evaluation and development of applications including visual representations. This type of platform is a great opportunity to include the features presented by the *SmartParticipation* project to promote citizens’ participation.

- ***Political Controlling and Public Memory.*** Besides all the features proposed by the *SmartParticipation* project, future work could include additional tools for improving political controlling. An example of this can be the analysis of voting from elected authorities. Their profiles can be analyzed before and after an election process using the answers provided to the system and the dynamic profile, which can give additional information for the analysis. Political programs can also be analyzed, including an evaluation of their performance.

9.3 Conclusions

RSs have been used mainly in *eCommerce* to evaluate and filter the vast amount of information available on the Web in order to assist users in their search processes and retrieval. These systems have been used to a great degree and play an important role for different Internet sites that offer products and services on social networks such as Amazon, YouTube, Netflix, Yahoo!, TripAdvisor, Facebook, and Twitter. Many different companies are developing recommender system techniques as an added value to the services they provide to their subscribers. The use of RSs for *eGovernment* is a new research area focused on reducing information overload, which could improve democratic processes and enhance participation. In the case of *eDemocracy*, a specific type of recommender system, also known as voting advice applications, has been used to provide recommendations to citizens about political parties and candidates facing *eVoting* and *eElections* processes. The proposed FRS addresses the research questions listed in Sect. 1.3 and is described as follows:

- ***What are the differences between classic RSs used in eBusiness with those used in eGovernment?***

RSs are computer-based techniques used to reduce information overload and to provide recommendations of products likely to interest a user. This technique is mainly used in *eCommerce* to suggest items that a customer is presumably going to buy.

The recommender system proposed in this work is a participatory system, in which users intentionally provide information about their preferences. In contrast, RSs used in *eCommerce*, also called targeted marketing methods, are based on extensional information, which refers to actions or past experiences with specific objects. The recommender system approach differs from collaborative filtering methods in that they are based on past experiences. It is also suitable in the one-and-only scenario in which events such as voting and election processes occur only once.

- ***How can RSs be used to improve citizens' participation?***

One of the main objectives of the recommender system approach proposed in this work is to increase citizens' participation, to provide more information to citizens about candidates, and to create political communities. The creation of political communities and social networks among citizens allows interaction and participation through social media, potentially crossing geographical and political boundaries. Contacting people with similar political profiles, building exchange platforms, and stimulating participation will enrich the information- and knowledge-based society in the future.

The recommender system approach proposed in this work can be used for *eCollaboration*, *eDemocracy*, and *eCommunity*, but could be extended to other participation areas.

- ***How can fuzzy logic be applied in RSs?***

Classical approaches of RSs are based on collaborative filtering methods. Nevertheless, Ekstrand *et al.* [2011] mention that the dimensionality reduction method is used more often. The kernel of the FRS is based on dimensionality reduction together with fuzzy clustering to provide recommen-

dations using a bi-dimensional political and issue-based landscape, which includes the percentage of similarity of the n -closest candidates/citizens. Therefore, relationships to the closest neighbor can be derived and analyzed by users.

- ***What are the advantages and disadvantages of using a fuzzy-based recommender system?***

The approach proposed in this Ph.D. thesis for advising citizens on elections and the creation of political communities is based on the visualization of political and issue-based landscapes, together with the generation of a fuzzy cluster. The main advantage of using fuzzy clustering is that it allows a gradual membership of data points to clusters with different degrees of membership according to the fuzzy set theory introduced by Zadeh [1965].

Nevertheless, the FRS approach has some weaknesses. First, a priori specification of the number of clusters is mandatory. Additionally, the output of the final position of the centers depends on how the algorithm is initialized. To make sure that we are in the presence of a local optimum (a solution that is not exactly the best), several executions of the algorithm should be made. Another problem of using fuzzy clustering methods is related to computational complexity. This issue has to be resolved when the systems begins to face large amounts of data, which may lead to performance and usability issues for users.

- ***What type of architecture shall be chosen for the development of a fuzzy-based recommender system?***

The architecture used to develop the *SmartParticipation* platform was divided in two blocks. One part is dedicated to generating the different profiles from users. This block uses a DWI supported by a CMS. The development of the DWI is very important, because it will feed the FRS. The second block corresponds to the FRS and is implemented using a MATLAB Compiler to execute the computation of the dimensionality reduction, fuzzy clustering, and visualization of recommendations with a runtime engine. This architecture allows a degree of flexibility for the development of additional features, algorithms, and distribution since, with the use of MATLAB Com-

pilers, end users do not need to have any type of license to install the runtime environment. A disadvantage of this architecture is the need for additional packages to execute the GUI. Future work should integrate solutions into the DWI.

- ***How can one conceptually evaluate the system developed and the implementation?***

One of the most challenging tasks faced by this work was the development of a framework that could be used to evaluate different *eParticipation* projects and be reduced to an analysis of VAAs only. For this reason, a five-level participation model is proposed in this work. The levels are *eInforming*, *eConsulting*, *eDiscussion*, *eParticipation*, and *eEmpowerment*. These participation levels are evaluated independently using a framework which includes three main components: web evolution, media richness, and communication channels.

The evaluation framework proposed is meant to be scalable, giving the opportunities to include more components for further development. For instance, for the evaluation of Web development, a so-called Web. 2.5, also known as the mobile Web, can be integrated. This version of the Web was defined by Weber & Rech [2009] as a Web that is “always on” for users who carry mobile devices connected to the Internet. Some examples of this type of definition are the applications for mobile devices provided by Facebook, Twitter, Google+, and LinkedIn, among others.

This type of applications could also improve the levels of participation providing what can be considered a synchronous channel of communication and participation.

“There are so many things that can be accomplished when a human being is connected to the rest of world.” Martin Cooper (BBC [2013]).

Discussion and Conclusions

Appendices

Appendix A

List of VAAs

List of VAAs used by the evaluation framework described in Chap. 7

- Bússola [2012]
- Choose 4 Greece [2012]
- Cabina Elettorale [2012]
- EU Profiler [2012]
- Glasovoditel [2012]
- Kieskompas [2012]
- KohoVolit [2012]
- Latarnik [2012]
- Manobalsas [2012]
- Political Compass [2012]
- Politikkabine [2012]
- smartvote [2012b]
- StemmenTracker [2012]
- StemWijzer [2012]
- Testvot [2012]
- Vimentis [2012]
- Vote Match [2012]
- VoteMatch Europe [2012]
- Vote Smart [2012]
- Wahlkabine [2012]
- Wahlomath [2012]
- Who Do I Vote For [2012]

List of VAAs

Appendix B

Evaluation of the *smartvote* project

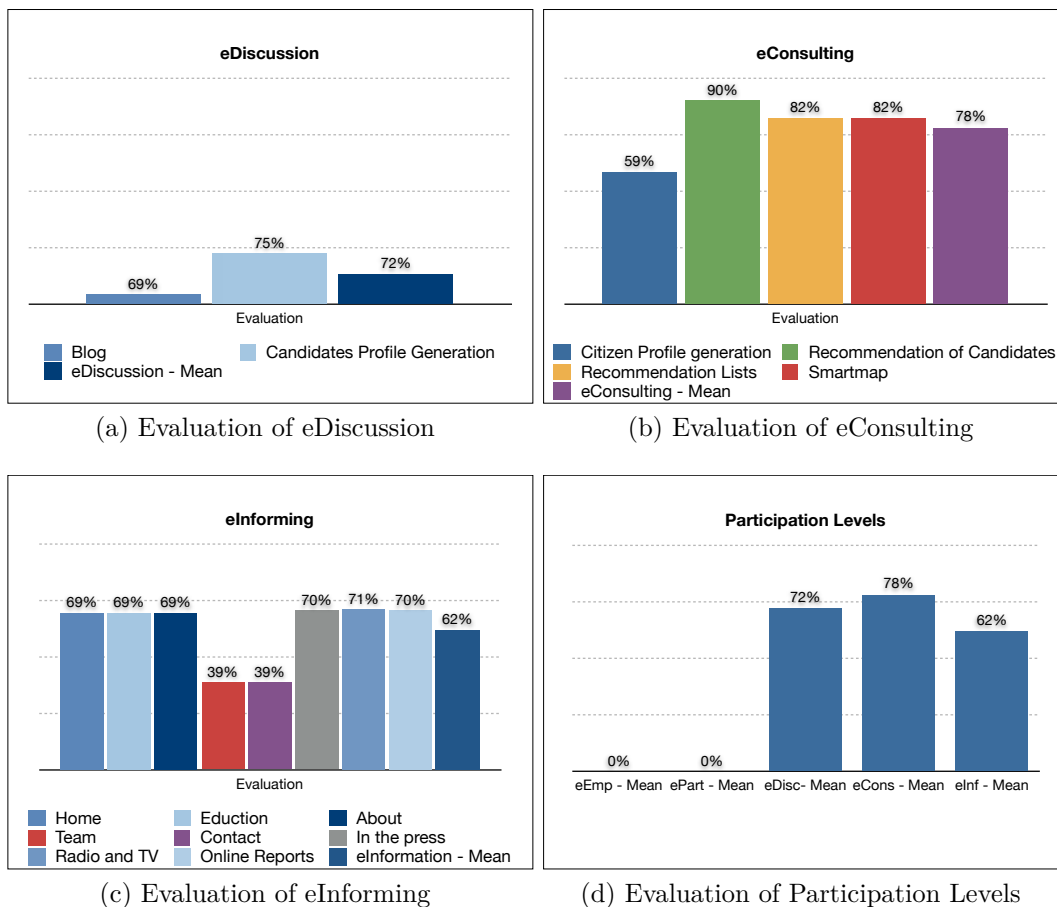


Figure B.1: Evaluation of the *smartvote* project

Evaluation of the *smartvote* project

Table B.1: Evaluation of the *smartvote* Project

Description	URL	WE ¹			MR ²				CC ³		Eval.
		Web 1.0	Web 2.0	Web 3.0	Text	Image	Audio	Video	Inter. Video	Sync.	
eEmpowerment											
none	none	✓	✓	✓	✓	✓	✓	✓	✓	✓	0
										Mean	0
eParticipation											
none	none	✓	✓	✓	✓	✓	✓	✓	✓	✓	0
										Mean	0
eDiscussion											
Blog	http://blog.smartvote.ch/	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.69
Candidates Profile Generation	https://smartvote.ch/candidate	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.75
										Mean	0.72
eConsulting											
Citizen Profile Generation	Depends on election selected	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.59
Recommendation of Candidates	Depends on election selected	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.90
Recommendation Lists	Depends on election selected	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.82
Smartmap	Depends on election selected	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.82
										Mean	0.78
eInforming											
Home	http://smartvote.ch/	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.69
Education	http://smartvote.ch/edu	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.69
About	https://smartvote.ch/about/idea	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.69
Team	https://smartvote.ch/about/team	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.39
Contact	https://smartvote.ch/about/team	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.39
In the Press	https://smartvote.ch/report/index?type=PRINT	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.70
Radio and TV	https://smartvote.ch/report/index?type=RADIO_AND_TV	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.71
Online Reports	https://smartvote.ch/report/index?type=WEB	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.70
										Mean	0.62

¹ Web Evolution.

² Media Richness.

³ Communication Channels.

Appendix C

ICT Tools for eParticipation

A number of ICT tools have been identified and classified according to different areas of *eParticipation*, which are defined in the eGovernment framework. Table C.1 provides a non-exhaustive list of ICT tools used for eParticipation, together with a brief description.

Table C.1: ICT Tools for eParticipation

ICT Tool	Description	Part. Levels					Areas	
		eInf	eCon	eDis	ePart	eEmp	eCol	eDem
Content Analysis Tools	These software tools allow analysis of communication content. An example of such a tool is the WebCAT project (WebCAT [2012]), which extracts meta-data and generates RDF descriptions for existing Web documents.	X					X	X
Natural Language User Interface (LUI)	A type of computer human interface where linguistic phenomena such as verbs, phrases, and clauses act as UI controls for creating, selecting, and modifying data in software applications. Examples include Ubiquity (Firefox), Wolfram Alpha, and Siri (Apple), among others.	X					X	X
Publication of Results	Software tools that allow citizens to see the results of eVoting and eElections. An example of such tools is the project Connect 2 Congress by Kinnaird [2012].	X						X
Public Memory & Political Controlling	Software tools that allow the digitalization of important works, documents, images, speeches, movies, television or radio recording, governmental programs and resolutions, citizens, initiatives, etc.	X						X
Search Engines	Also known as information retrieval systems, these are software tools designed to find information on a computer system.	X					X	X
Semantic Web	Aims to convert the current Web of unstructured documents into a "Web of data" in a common framework that allows data to be shared and reused across applications, enterprises, and community boundaries (W3C [2012]).	X					X	X
Subscription Services	Software tools that allow citizens to follow news in real time. Examples of such services are: RSS feeds, newsletters, SMS, mailing lists, podcasts, vodcasts, etc.	X					X	X
Consultation Platforms	Also known as eConsultation or ePanel, these are software tools used in consultation processes. An example of such a tool is the eConsultation project (eConsultation [2012]).		X				X	X
Online Survey Tools	They are software tools that allows to create and deploy web-based surveys to gather important feedback.		X				X	X

Continued on next page

ICT Tools for eParticipation

Table C.1 – continued from previous page

ICT Toolt	Description	Part. Levels					Areas		
		eInf	eCon	eDis	ePart	eEmp	eCol	eDem	eCom
Buddy Systems	Shows where colleagues or friends are currently located and how they can be reached electronically. It makes it possible for members to meet virtually or exchange experiences.			X					X
Civic Network Systems	Also known as community networks, these are electronic meeting points for citizens whose common ground is a shared place or living environment. Apart from discussion forums, these focus on projects that concern the community.			X					X
Content Management Systems (CMS)	CMSs are software tools that support the collection, management, and publishing of information, and provide procedures for managing the workflow in a collaborative environment.			X			X	X	X
Discussion Forums	Also known as Internet forums, these are online discussion platforms where people can hold conversations in the form of posted messages.			X			X	X	X
Matchmaking Systems	Matchmaking is about establishing social and economic exchange relations. These systems promote contacts and activities in a commonly used environment. Networks of acquaintances are utilized to make new contacts, based on an already existing mutual trust, and exchange information.			X			X	X	X
recommender systems (RSs)	These systems learn Internet users' preferences and make suggestions for further development.			X			X	X	X
Voting Advice Applications (VAAs)	Also known as Voting Aid Applications are Web-based applications that help citizens find a party or candidate that stands closest to their preferences.			X				X	
Web Logs	Software tools used as personal journals published on the Web and consisting of discrete entries (posts). They make it possible to share knowledge, discuss topics, and nurture relationships.			X			X	X	X
Web Virtual Meeting Places	Software tools that provide the development of virtual communities as supplement to conventional ones.			X			X		X
Collaborative Management Tools (Groupware)	Facilitate and manage group activities (e.g., electronic calendars, project management systems, workflow systems, knowledge management systems, enterprise bookmarking). Designed to help people involved in a common task achieve goals.				X		X		X
Collaborative Working Environments (CWE)	Supports people in their individual and cooperative work. They involve organizational, technical, and social issues. The following applications or services are considered elements of a CWE: email, instant messaging, application sharing, videoconferencing, collaborative workspace and document management, task- and workflow-management, wiki group, and blogging.				X		X		X
Computer-Supported Cooperative Work (CSCW)	Addresses how collaborative activities and their coordination can be supported by means of computer systems. CSCW focuses on the study of tools and techniques of groupware as well as their psychological, social, and organizational effects.				X		X		X
Deliberative Polling	Software tools intended to incorporate the principles of deliberative democracy to decision-making. An example of such a tool is the PICOLA project (PICOLA [2012]).					X	X	X	X
eVoting & eElections	Place- and time-independent voting and elections using ICT technologies. An example of such a tools is the eVoting project in Geneva (eVoting [2012]).					X		X	

Table C.1 illustrates that an *eParticipation* project can use different ICT tools, and that each tool can belong to different levels and participation areas.

Appendix D

Mockup for SmartParticipation

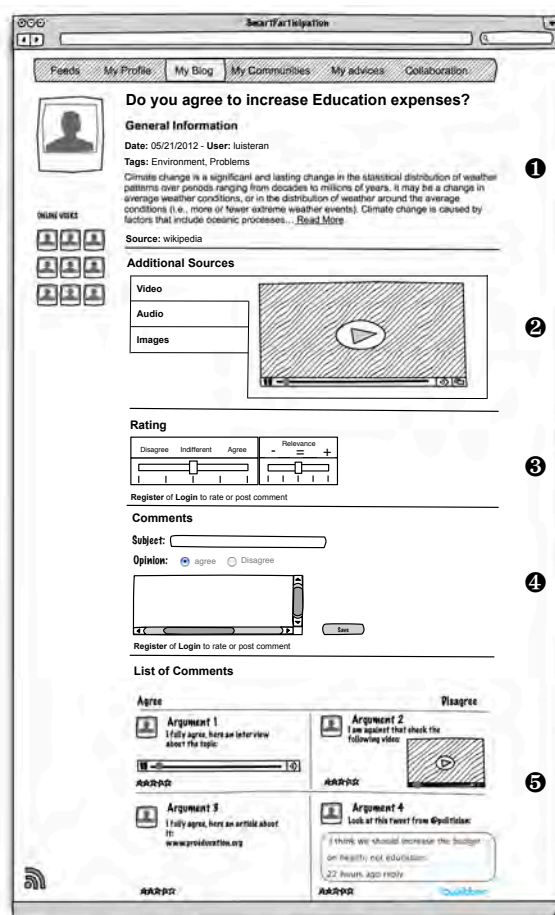


Figure D.1: Discussion Topics GUI Mockup

Appendix E

smartvote Match Points Computation

The latest version of matching points computation is presented in *smartvote* [2012a]. In order to generate the recommendations, *smartvote* computes the match points using the Euclidean distance between the position of candidates and voter according to Eq. (E.1).

$$dist(v, c) = \sqrt{\sum_{i=1}^n (v_i - c_i)^2} \quad (E.1)$$

where v_i is the position of voter v on question i , and c_i is the position of voter c on question i . Tables E.1 and E.2 show the values given for budget and standard questions respectively. The maximum possible distance is computed using Eq. (E.2).

$$mdist(v) = \sqrt{\sum_{i=1}^n (100)^2} = \sqrt{n * (100)^2} \quad (E.2)$$

This distance is normalized using the maximum distance according to Eq. (E.4).

$$dist_{nor}(v, c) = \frac{dist(v, c)}{mdist} \quad (E.3)$$

The matching value in terms of distance is computed by subtracting 1 from $dist_{norm}(v, c)$. For an easy interpretation of the results, the correlation value is given as a percentage.

$$match(v, c) = 100 * \left(1 - \frac{dist(v, c)}{mdist}\right) \quad (E.4)$$

***smartvote* Match Points Computation**

Table E.1: Budget Questions

Answer Options	Assigned Values
Clearly spend more (+ +)	100
Spend more (+)	75
Spend the same (=)	50
Spend less (−)	25
Spend significantly less (− −)	0

Table E.2: Standard Questions

Answer Options	Assigned Values
Yes	100
Probably yes	75
Probably no	25
No	0

Consideration of weights

The voters also have the option to assign an individual weight to each question. This consists of five options, where the following weighting factors are assigned:

- 3 → important
- 2 → somewhat important
- 1 → normal weight
- 0.5 → rather unimportant
- 0.33 → unimportant

Part of the calculation of the weighting choice recommendation is included as shown in Eq. (E.5):

$$dist_w(v, c) = \sqrt{\sum_{i=1}^n (w_i * (v_i - c_i))^2} \quad (\text{E.5})$$

The maximum possible distance including the weight is computed using Eq. (E.6),

$$mdist_w(v) = \sqrt{\sum_{i=1}^n (w_i * 100)^2} \quad (\text{E.6})$$

where W_i is the weight value assigned by voter v on question i .

Calculation of the lists-choice recommendation

In addition to a recommendation to vote for individual candidates, the system can also provide an internal list selection recommendation. This procedure is basically the same as for the recommendation for individual candidates. The only difference is that the answers from all candidates are aggregated into unified “response lists” for each question, before the matching computation. This is done by calculating the arithmetic mean of all candidates’ answers.

Nomenclature

Roman Symbols

A	Fuzzy set, page 42
\bar{A}	Complement of a fuzzy set A , page 44
ADI	Alternative Dunn's Index, page 90
$adjcos_{ij}$	Adjusted Cosine Similarity between items i_i and i_j , page 34
agr_{ij}^k	j -th agreement component of user i on the k -th issue, page 59
A_α	α -cut of a fuzzy set A , page 45
r_i	Average rating of user i , page 30
r_i	Average rating of item i , page 33
c	Number of clusters, page 50
CE	Classification Entropy, page 89
$corr_{ij}$	Pearson correlation similarity between items i and j , page 33
$corr_{ik}$	Pearson correlation similarity between user i and k , page 30
cos_{ik}	Cosine similarity between user i and k , page 29
sim_{ik}	Similarity between user i and k , page 29
cos_{ij}	Cosine similarity between items i and j , page 32
DI	Dunn's Index, page 90

***smartvote* Match Points Computation**

$dist(v, c)$	Distance between voter v and candidate c , page 173
$dist_{nor}(v, c)$	Normalized distance between voter v and candidate c , page 173
$dist_w(v, c)$	Distance between voter v and candidate c including weight values, page 174
$match(v, c)$	Matching value between voter v and candidate c , page 173
$mdist(v)$	Theoretical maximum possible match score including the weight values, page 174
$mdist(v)$	Theoretical maximum possible match score, page 173
F	Data dispersion matrix, page 69
\bar{x}_k	Mean of data set x_k , page 69
F_i	Covariance matrix, page 69
FP_i	Fuzzy Profile of user i , page 59
fpc_{ij}^k	j -th fuzzy profile component of user i on the k -th issue, page 59
I	Set of items, page 28
J_{fcm}	Objective function of FCM algorithm, page 50
i_j	Item j , page 28
J_{cm}	Objective function of c-means algorithm, page 47
A_α	Kernel of a fuzzy set A , page 46
m	Level of fuzziness, page 50
PC	Partition Coefficient, page 88
P_{aj}	Prediction of a possible opinion of active user u_a about an item i_j , page 31
\vec{R}	User-item matrix, page 29
rel_{ij}^k	j -th relevance component of user i on the k -th issue, page 59
\vec{S}	Similarity matrix, page 30
S	Separation Index, page 89

SC	Partition Index, page 89
$Supp(A)$	Support of a fuzzy set A , page 45
U	Set of users, page 28
U_i	Matrix of eigenvectors, page 69
u_a	Active user, page 28
I_{ui}	List of i items rated by user u , page 28
u_{ij}	Membership degree of an observation x_i in a cluster Γ_i , page 49
u_i	User i , page 28
U	Universe of discourse, page 42
\vec{U}	Fuzzy partition matrix, page 49
W_i	Weigh matrix, page 69
we	Web Evolution component, page 126
Pr_{aj}	Prediction of item i_j for an active user u_a , page 35
x_k	Observed vector in a <i>high-dimensional</i> space, page 69
XB	Xie and Beni's Index, page 89
x_i	i -th element of fuzzy set A , page 42
\vec{Y}	Vector of cluster centers, page 50
$y_{i,k}$	Vector of q -dimensional reduced representation o the observed vector x_k , page 69

Greek Symbols

ϵ	Termination criteria, page 50
Γ_i	Set of clusters, page 49
$\lambda_{i,j}$	Eigenvalues, page 69
Λ_i	Matrix of eigenvalues, page 69
μ_A	Membership function of fuzzy set A , page 42
$\mu_A(x_i)$	Degree of membership of x_i in a fuzzy set A , page 42

Acronyms

ADI Alternative Dunn's Index

AI Artificial Intelligence

ASP Active Server Pages

A2A Administration to Administration

A2C Administration to Citizens

A2B Administration to Business

CE Classification Entropy

CF Collaborative Filtering

CGI Computer-generated Imagery

CMS Content Management Systems

CMT Collaborative Management Tools

CSCW Computer-Supported Cooperative Work

CWE Collaborative Working Environments

DI Dunn's Index

DWI Dynamic Web Interface

eAssistance Electronic Assistance

eBusiness Electronic Business

eCommerce Electronic Commerce

eCommunity Electronic Community

eContracting Electronic Contracting

eConsulting Electronic Consulting

eCollaboration Electronic Collaboration

eDemocracy Electronic Democracy

eDiscussion Electronic Discussion

eElections Electronic Elections

eEmpowerment Electronic Empowerment

eGovernment Electronic Government

eHealth Electronic Health

eInforming Electronic Informing

eParticipation Electronic Participation

eProcurement Electronic Procurement

eService Electronic Service

eSettlement Electronic Settlement

eVoting Electronic Voting

FCM Fuzzy C-means Algorithm

FGG Fuzzy Gath-Geva Algorithm

FGK Fuzzy Gustafson-Kessel Algorithm

FMLE Fuzzy Maximum Likelihood Estimator

FRS Fuzzy Recommender System

FP Fuzzy Profile

fpc Fuzzy Profile Component

GUI Graphical User Interface

HCI Human Computer Interaction

HTML HyperText Markup Language

HTTP Hypertext Transfer Protocol

ICT Information and Communication Technologies

IS Information Systems

JSP JavaServer Pages

LLE Locally Linear Embedding
LUI Natural Language User Interface
MCR MATLAB Compiler Runtime
MDS Multidimensional Scaling
N3 Notation3
OWL Web Ontology Language
PC Partition Coefficient
PCA Principal Component Analysis
RDF Resource Description Framework
RDFS RDF Schema
RSs recommender systems
RSS Really Simple Syndication
S Separation Index
SC Partition Index
SKOS Simple Knowledge Organization System
SNE Stochastic Neighbor Embedding
SPARQL SPARQL Protocol An RDF Query Language
SVD Singular Value Decomposition
Turtle Terse RDF Triple Language
t-SNE t-Distributed Stochastic Neighbor Embedding
UI User Interface
UML Unified Modeling Language
URI Uniform Resource Identifier
URL Uniform Resource Locator
VAAAs Voting Advice Applications

VBScript Visual Basic Scripting Edition

WCAG Web Content Accessibility Guidelines

WWW World Wide Web

W3C World Wide Web Consortium

XHTML Extensible HyperText Markup Language

XML Extensible Markup Language

XB Xie and Beni's Index

References

- ADOMAVICIUS, G. & TUZHILIN, A. (2011). *Context-aware recommender systems*, chap. 7, 217–253. In Ricci *et al.* [2011], 1st edn. 156
- AL-HAKIM, L. (2006). *Global e-Government: Theory, Applications and Benchmarking*. IGI Global, 1st edn. 23
- AMATRIAIN, X., JAIMES, A., OLIVER, N. & PUJOL, J.M. (2011). Data mining methods for recommender systems. In *Recommender Systems Handbook*, 39–71, Springer. 29
- ANDRIESSEN, J.E. (2003). *Working with Groupware: Understanding and Evaluating Collaboration Technology*. Computer Supported Cooperative Work, Springer Verlag, 1st edn. 124
- ARNSTEIN, S.R. (1969). A Ladder Of Citizen Participation. *Journal of the American Institute of Planners*, **35**, 216–224. 103
- BABUKA, R., VAN DER VEEN, P. & KAYMAK, U. (2002). Improved covariance estimation for gaussian-kernel clustering. In *Fuzzy Systems, 2002. FUZZ-IEEE'02. Proceedings of the 2002 IEEE International Conference on*, vol. 2, 1081–1085, IEEE. 92
- BAEZA-YATES, R. & RIBEIRO-NETO, B. (2011). *Modern Information Retrieval: The Concepts and Technology behind Search*. ACM Press Books, 2nd edn. 4
- BALASKO, B., ABONYI, J. & FEIL, B. (2005). Fuzzy clustering and data analysis toolbox. *Department of Process Engineering, University of Veszprem, Veszprem*. xx, 88, 95, 97, 98, 100
- BENSAID, A.M., HALL, L.O., BEZDEK, J.C., CLARKE, L.P., SILBINGER, M.L., ARRINGTON, J.A. & MURTAGH, R.F. (1996). Validity-guided (re)

REFERENCES

- clustering with applications to image segmentation. *Fuzzy Systems, IEEE Transactions on*, **4**, 112–123. 89
- BEZDEK, J.C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. Advanced Applications in Pattern Recognition, Springer US. 49, 88, 149
- BEZDEK, J.C. & DUNN, J.C. (1975). Optimal fuzzy partitions: A heuristic for estimating the parameters in a mixture of normal distributions. *Computers, IEEE Transactions on*, **100**, 835–838. 90, 93, 149
- CEDRONI, L. & GARZIA, D., eds. (2010). *Voting Advice Applications in Europe. The State of the Art*. Scripta Web. 133, 188
- CÔRTE-REAL, P.P. (2007). Fuzzy voters, crisp votes. *International Game Theory Review*, **9**, 67–86. 157
- COX, M.A. & COX, T.F. (2008). Multidimensional scaling. In *Handbook of data visualization*, 315–347, Springer. 77
- DAFT, R.L. & LENGEL, R.H. (1984). Information Richness: A New Approach to Managerial Behaviour and Organizational Design. *Research in Organizational Behaviour*, **6**, 191–233. xxiii, 123
- DIJKSTRA, E.W. (1959). A note on two problems in connexion with graphs. *NUMERISCHE MATHEMATIK*, **1**, 269–271. 77
- DOWNS, A. (1957). An economic theory of democracy. 260–276. 154
- DROBNJAK, A. (2012). *Fuzzy Clustering in Social Networks*. Master’s thesis, University of Fribourg, Boulevard de Perolles 15, 1700 Fribourg. 125
- DUBOIS, D.J. & PRADE, H. (1980). *Fuzzy Sets and Systems: Theory and Applications*, vol. 144 of *Mathematics in Science and Engineering*. Academic Press. 43
- DUBOIS, D.J. & PRADE, H. (1985). A Review of Fuzzy Set Aggregation Connectives. *Information Sciences*, **36**, 85 – 121. 43
- DUDOIT, S. & FRIDLYAND, J. (2002). A Prediction-based Resampling Method for Estimating the Number of Clusters in a Dataset. *Genome Biology*, **3**, 1–21. 52
- DUNN, J.C. (1973a). A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *Journal of Cybernetics*, **3**, 37–57. 49

REFERENCES

- DUNN, J.C. (1973b). A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. 90
- EKSTRAND, M.D., RIEDL, J.T. & KONSTAN, J.A. (2011). Collaborative filtering recommender systems. *Now - The Essence of Knowledge*, **4**, 81–173. 28, 39, 159
- FELDMAN, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM*, **56**, 82–89. 156
- FENSEL, D., DOMINGUE, J. & HENDLER, J.A., eds. (2011). *Handbook of Semantic Web Technologies*, vol. 1. Springer, 1st edn. 4, 133
- FIVAZ, J. & FELDER, G. (2009). *Added Value of e-Democracy Tools in Advanced Democracies? The Voting Advice Application smartvote in Switzerland*, chap. 8, 109–122. In Shark & Toporkoff [2009]. 129
- FLOYD, R.W. (1962). Algorithm 97: Shortest path. *Commun. ACM*, **5**, 345–77
- GOLDBERG, K., ROEDER, T., GUPTA, D. & PERKINS, C. (2001). Eigentaste: A constant time collaborative filtering algorithm. *Information Retrieval*, **4**, 133–151. 37
- GORBAN, A.N. (2007). *Principal manifolds for data visualization and dimension reduction*, vol. 58. Springer. 99
- GRÖNLUND, Å. (2009). ICT Is Not Participation Is Not Democracy – eParticipation Development Models Revisited. In Tambouris & Macintosh [2009], 12–23. xxiii, 102, 103
- GUO, X. & LU, J. (2007). Intelligent E-government Services with Personalized Recommendation Techniques: Research Articles. *International Journal of Intelligent Systems*, **22**, 401–417. 28, 37
- GUSTAFSON, D.E. & KESSEL, W.C. (1978). Fuzzy clustering with a fuzzy covariance matrix. In *Decision and Control including the 17th Symposium on Adaptive Processes, 1978 IEEE Conference on*, vol. 17, 761–766, IEEE. 90, 91, 149
- HEVNER, A.R., MARCH, S.T., PARK, J. & RAM, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, **28**, 75–105. xxiii, 7
- HINTON, G. & ROWEIS, S. (2002). Stochastic neighbor embedding. *Advances in neural information processing systems*, **15**, 833–840. 73

REFERENCES

- HITZLER, P., RUDOLPH, S. & KRÖTZSCH, M. (2009). *Foundations of Semantic Web Technologies*. Chapman & Hall/CRC, 1st edn. 4, 133
- JANNACH, D., ZANKER, M., FELFERNIG, A. & FRIEDRICH, G. (2010). *Recommender Systems: An Introduction*. Cambridge University Press, 1st edn. 39
- KALAMPOKIS, E., TAMBOURIS, E. & TARABANIS, K. (2008). A domain model for eparticipation. In *Proceedings of the 2008 Third International Conference on Internet and Web Applications and Services, ICIW '08*, 25–30, IEEE Computer Society, Washington, DC, USA. 116
- KORTUM, P. (2008). *HCI beyond the GUI: Design for haptic, speech, olfactory, and other nontraditional interfaces*. Morgan Kaufmann. 157
- LADNER, A., FELDER, G. & FIVAZ, J. (2010a). *More than toys? A first assessment of voting advice applications in Switzerland*, chap. 5, 91–123. In Cedroni & Garzia [2010]. 154
- LADNER, A., FIVAZ, J. & PIANZOLA, J. (2010b). Impact of Voting Advice Applications on Voters' Decision-Making. In *Internet, Politics, Policy 2010: An Impact Assessment*, Oxford Internet Institute (OII, University of Oxford), Oxford, UK. 129
- LAM, H.K., LING, S.S.H. & NGUYEN, H.T., eds. (2011). *Computational Intelligence and Its Applications: Evolutionary Computation, Fuzzy Logic, Neural Network and Support Vector Machine Techniques*. Imperial College Press. 52
- LATHROP, D. & RUMA, L., eds. (2010). *Open Government. Collaboration, Transparency, and Participation in Practice*. O'Reilly Media, 1st edn. 5, 23
- LEE, J.A. & VERLEYSEN, M. (2007). *Nonlinear dimensionality reduction*. Springer. xx, 76, 78, 100
- LEE, W. & STOLFO, S.J. (1998). Data Mining Approaches for Intrusion Detection. In *Proceedings of the 7th conference on USENIX Security Symposium*, vol. 7 of *SSYM'98*, 6–6, USENIX Association, Berkeley, CA, USA. 32
- LEUF, B. & CUNNINGHAM, W. (2001). *The Wiki Way: Quick Collaboration on the Web*. Addison-Wesley Professional. 17
- LUKENSMEYER, C.J. & TORRES, L.H. (2006). *Public Deliberation: A Manager's Guide to Citizen Engagement*. Collaboration Series, IBM Center for the Business of Government, Washington, DC. 103

REFERENCES

- MACINTOSH, A. (2004). Characterizing e-participation in policy-making. In *Proceedings of the Proceedings of the 37th Annual Hawaii International Conference on System Sciences (HICSS'04) - Track 5 - Volume 5*, HICSS '04, 50117.1–, IEEE Computer Society, Washington, DC, USA. 116
- MACINTOSH, A. & WHYTE, A. (2008). Towards an Evaluation Framework for eParticipation. *Transforming Government: People, Process and Policy*, **2**, 16–30. 103, 116
- MACQUEEN, J.B. (1967). Some Methods for classification and Analysis of Multivariate Observations. In *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, 281–297, University of California Press, Berkeley, USA. 47
- MAZZA, R. (2009). *Introduction to Information Visualization*. Springer, 1st edn. 99
- MEIER, A. (2012). *eDemocracy & eGovernment - Maturity Levels of a Democratic Knowledge Society*. Springer, Berlin, Germany, 1st edn. xix, 6, 16, 17, 19, 21, 23, 66, 109, 110, 118, 129, 132
- MIYAMOTO, S., ICHIHASHI, H. & HONDA, K. (2010). *Algorithms for Fuzzy Clustering: Methods in c-Means Clustering with Applications (Studies in Fuzziness and Soft Computing)*. Springer, 2nd edn. 53
- MURUGESAN, S. (2009a). *Handbook of Research on Web 2.0, 3.0, and X.0: Technologies, Business, and Social Applications*, chap. 1, 1–11. In Murugesan [2009b], 1st edn. 119
- MURUGESAN, S. (2009b). *Handbook of Research on Web 2.0, 3.0, and X.0: Technologies, Business, and Social Applications*. Information Science Reference, 1st edn. 133, 189, 193
- NASCIMENTO, S. (2005). *Fuzzy Clustering Via Proportional Membership Model*. IOS Press, 1st edn. 53
- NORDIN, D. (2011). *Design and Prototyping for Drupal*. O'Reilly. 150
- OECD (2001). *Citizens as Partners: Information, Consultation and Public Participation in Policy-making*. OECD, Paris. 103
- PANOPOULOU, E., TAMBOURIS, E. & TARABANIS, K. (2010). eParticipation Initiatives in Europe: Learning from Practitioners. In Tambouris *et al.* [2010], 237. 102, 116

REFERENCES

- PEARSON, K. (1901). On Lines and Planes of Closest Fit to Systems of Points in Space. *Philosophical Magazine*, **2**, 559–572. 29, 69, 73, 90, 149
- RESNICK, P., IACOVOU, N., SUSHAK, M., BERGSTROM, P. & RIEDL, J. (1994). Grouplens: An open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*, CSCW '94, 175–186, ACM, New York, NY, USA. 29
- RICCI, F., ROKACH, L., SHAPIRA, B. & KANTOR, P.B., eds. (2011). *Recommender Systems Handbook*. Springer, 1st edn. 4, 39, 185
- ROWE, G. & FREWER, L.J. (2000). Public Participation Methods: A Framework for Evaluation. *Science, Technology & Human Values*, **25**, 3–29. 116
- ROWEIS, S.T. & SAUL, L.K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, **290**, 2323–2326. 69, 80, 149
- SAMMON, J.W. (1969). A Nonlinear Mapping for Data Structure Analysis. *IEEE Transactions on Computers*, **c-18**, 401–409. 69, 149
- SANFORD, C. & ROSE, J. (2007). Characterizing eparticipation. *International Journal of Information Management*, **27**, 406 – 421. 102
- SARWAR, B., KARYPIS, G., KONSTAN, J. & RIEDL, J. (2000a). Analysis of Recommendation Algorithms for e-Commerce. In *Proceedings of the 2nd ACM conference on Electronic commerce*, EC '00, 158–167, ACM, New York, NY, USA. 31
- SARWAR, B., KARYPIS, G., KONSTAN, J. & REIDL, J. (2001). Item-based Collaborative Filtering Recommendation Algorithms. In *Proceedings of the 10th International Conference on World Wide Web*, WWW '01, 285–295, ACM, Hong Kong. xix, 27, 28, 32, 33, 34
- SARWAR, B.M., KARYPIS, G., KONSTAN, J.A. & RIEDL, J.T. (2000b). Application of dimensionality reduction in recommender system – a case study. In *IN ACM WEBKDD WORKSHOP*. xix, 35, 37
- SHARK, A.R. & TOPORKOFF, S., eds. (2009). *Beyond eGovernment - Measuring Performance: A Global Perspective*. Public Technology Institute and ITEMS International, Washington, DC. 112, 187
- SICILIANO, A. (2008). *MATLAB: Data analysis and visualization*. World Scientific. 150

REFERENCES

- SLOANE, P. (2011). *A Guide to Open Innovation and Crowdsourcing: Advice From Leading Experts*. Kogan Page Publishers, 1st edn. 23
- SMITH, S., MACINTOSH, A. & MILLARD, J. (2011). A three-layered framework for evaluating e-participation. *International Journal of Electronic Governance*, **4**, 304–321. 116
- STEELE, J. & ILIINSKY, N., eds. (2010). *Beautiful Visualization: Looking at Data through the Eyes of Experts*. O'Reilly Media. 99
- SUGAR, C.A. & JAMES, G.M. (2003). Finding the Number of Clusters in a Data Set: An Information Theoretic Approach. *Journal of the American Statistical Association*, **98**, 750–763. 52
- SWISS CONFEDERATION (2013). Swiss Open Government Data Pilot Portal. [Online] Available at: <http://www.opendata.admin.ch/>, [Accessed 15-October-2013]. 158
- TAMBOURIS, E. & MACINTOSH, A., eds. (2009). *Electronic Participation: First International Conference, ePart 2009. Proceedings*, vol. 5694 of *LNCS*, Springer. 112, 187
- TAMBOURIS, E., LIOTAS, N. & TARABANIS, K. (2007). A Framework for Assessing eParticipation Projects and Tools. In *Proceedings of the 40th Annual Hawaii International Conference on System Sciences*, HICSS '07, 90–, IEEE Computer Society, Washington, DC, USA. 103, 111, 116
- TAMBOURIS, E., MACINTOSH, A. & GLASSEY, O., eds. (2010). *Electronic Participation: Second International Conference, ePart 2010. Proceedings*, vol. 6229 of *LNCS*, Springer. 112, 189
- TAMBOURIS, E., MACINTOSH, A. & DE BRUIJN, H., eds. (2011). *Electronic Participation: Third International Conference, ePart 2011. Proceedings*, vol. 6847 of *LNCS*, Springer. 112
- TENENBAUM, J.B., DE SILVA, V. & LANGFORD, J.C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, **290**, 2319–2323. 69, 76, 149
- TERÁN, L. (2011). A Fuzzy-Based Advisor for Elections and the Creation of Political Communities. In *Proceedings of International Conference on Information Society (i-Society)*, Computer Chapter, 180–185, IEEE, London, UK. 11

REFERENCES

- TERÁN, L. (2012). A Fuzzy-Based Platform for Stimulating Citizens' Participation. In *Proceedings of International Conference on E-Business and E-Government*, vol. 4 of *ICEE 2012*, 2353–2357, IEEE, Shanghai, China. 11
- TERÁN, L. & DROBNJAK, A. (2013). An Evaluation Framework for eParticipation: The VAAs Case Study. *World Academy of Science, Engineering and Technology, International Science Index* 73, 7, 315 – 324. 10
- TERÁN, L. & MEIER, A. (2010). A Fuzzy Recommender System for eElections. In K.N. Andresen, E. Francesconi, A. Grönlund & T.M. van Engers, eds., *Proceedings of International Conference on Electronic Government and the Information Systems Perspective (EGOVIS 2010)*, vol. 6267 of *LNCS*, 62–76, Springer, Bilbao, Spain. 11
- TERÁN, L. & MEIER, A. (2011). SmartParticipation – A Fuzzy-Based Platform for Stimulating Citizens' Participation. *International Journal for Infonomics (IJI)*, 4, 501–512. 11
- TERÁN, L., LADNER, A., FIVAZ, J. & GERBER, S. (2012). *Using a Fuzzy-Based Cluster Algorithm for Recommending Candidates in e-Elections*, chap. 6, 115–138. IGI Global. 11
- TIBSHIRANI, R., WALTHER, G. & HASTIE, T. (2001). Estimating the Number of Clusters in a Data Set Via the Gap Statistic. *Journal Of The Royal Statistical Society Series B*, 63, 411–423. 52
- TOMLINSON, T. & VANDYKE, J. (2010). *Pro Drupal 7 Development*. Apress. xxi, 139, 150
- VALENTE DE OLIVEIRA, J. & PEDRYCZ, W., eds. (2007). *Advances in Fuzzy Clustering and its Applications*. Wiley, 1st edn. 53
- VAN DER MAATEN, L. & HINTON, G. (2008). Visualizing high-dimensional data using t-sne. *Journal of Machine Learning Research*, 9, 2579–2605. 69, 73, 149
- VAN DER MAATEN, L., POSTMA, E. & VAN DEN HERIK, H. (2009). Dimensionality reduction: A comparative review. *Technical Report TiCC TR 2009-005*. xx, xxiii, 67, 68, 87
- VENNA, J. & KASKI, S. (2001). Neighborhood preservation in nonlinear projection methods: An experimental study. In *Proceedings of the International Conference on Artificial Neural Networks*, ICANN '01, 485–491, Springer-Verlag, London, UK, UK. 67

REFERENCES

- VENNA, J. & KASKI, S. (2006). Visualizing gene interaction graphs with local multidimensional scaling. In *ESANN*, 557–562. 67
- VOZALIS, E. & MARGARITIS, K.G. (2003). Analysis of Recommender Systems' Algorithms. In *Proceedings of the Sixth Hellenic-European Conference on Computer Mathematics and its Applications*, HERCMA, Athens, Greece. 26, 27, 30, 31, 34
- WAGNER, M. & RUUSUVIRTA, O. (2009). Faulty Recommendations? Party Positions in Online Voting Advice Applications. *SSRN eLibrary*, 39. 129, 153, 154
- WALGRAVE, S., VAN AELST, P. & NUYTEMANS, M. (2008). Do the Vote Test': The Electoral Effects of a Popular Vote Advice Application at the 2004 Belgian Elections. *International Journal of Political Science*, **43**, 50–70. 129
- WATER&STONE (2011). Open Source CMS Market Share. Tech. rep., Water&Stone. 137, 139, 150
- WATER&STONE & CMS WIRE (2009). Open Source CMS Market Share. Tech. rep., Water&Stone, CMS Wire. 137, 139
- WEBER, S. & RECH, J. (2009). *Handbook of Research on Web 2.0, 3.0, and X.0: Technologies, Business, and Social Applications*, chap. 2, 12–38. In Murugesan [2009b], 1st edn. 119, 126, 161
- XIE, X.L. & BENI, G. (1991). A validity measure for fuzzy clustering. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **13**, 841–847. 89
- XRDS MAGAZINE (2011). CS in Service of Democracy. *XRDS: Crossroads, The ACM Magazine for Students*, **18**. 23
- YAGER, R.R. (1980). On a General Class of Fuzzy Connectives. *Fuzzy Sets and Systems*, **4**, 235 – 242. 43
- YAGER, R.R. (2003). Fuzzy Logic Methods in Recommender Systems. *Fuzzy Sets and Systems*, **136**, 133–149. 27, 37, 38
- YOO, K.H., GRETZEL, U. & ZANKER, M. (2013). *Persuasive Recommender Systems: Conceptual Background and Implications*. SpringerBriefs in Electrical and Computer Engineering, Springer. 38
- ZADEH, L. (1975a). The Concept of a Linguistic Variable and its Application To Approximate Reasoning - Part I. *Information Sciences*, **8**, 199 – 249. 43

REFERENCES

- ZADEH, L. (1975b). The Concept of a Linguistic Variable and its Application To Approximate Reasoning - Part II. *Information Sciences*, **8**, 301 – 357. 43
- ZADEH, L. (1975c). The Concept of a Linguistic Variable and its Application To Approximate Reasoning - Part III. *Information Sciences*, **9**, 43 – 80. 43
- ZADEH, L.A. (1965). Fuzzy Sets. *Journal of Information and Control*, **8**, 338–353. 41, 43, 49, 160
- ZIMMERMANN, H.J. (2001). *Fuzzy Set Theory and its Applications*. Kluwer Academic Publishers, 4th edn. 43
- ZIMMERMANN, H.J. & ZYSNO, P. (1980). Latent Connectives in Human Decision Making. *Fuzzy Sets and Systems*, **4**, 37 – 51. 43

Internet Sources

ACE PROJECT (2012). Roles and Definition of Political Parties. [Online] Available at: <http://aceproject.org/ace-en/topics/pc/pca/pca01/pca01a>, [Accessed 07-March-2012].

AGRAWAL, V. (2013). Event Calendar | drupal.org. [Online] Available at: https://drupal.org/project/event_calendar, [Accessed 18-July-2013].

BBC (2013). Interview with mobile phone inventor Marty Cooper. [Online] Available at: <http://www.bbc.co.uk/news/technology-22020666>, [Accessed 31-July-2013].

BURSTEIN, A. (2013). Organic groups | drupal.org. [Online] Available at: <https://drupal.org/project/OG>, [Accessed 18-July-2013].

BÚSSOLA (2012). Bússola Eleitoral - Portugal. [Online] Available at: <http://www.bussolaeleitoral.pt>, [Accessed 06-March-2012].

CABINA ELETTORALE (2012). Elezioni Europee. [Online] Available at: <http://cabina-elettorale.it>, [Accessed 06-March-2012].

CALDWELL, J. (2012). Video Embed Field | drupal.org. [Online] Available at: https://drupal.org/project/video_embed_field, [Accessed 18-July-2013].

CHOOSE 4 GREECE (2012). Choose 4 Greece - Voting Advice Application. [Online] Available at: <http://www.choose4greece.org>, [Accessed 12-November-2012].

COLELLA, A. (2013). Comment Easy Reply | drupal.org. [Online] Available at: https://drupal.org/project/comment_easy_reply, [Accessed 18-July-2013].

INTERNET SOURCES

- CORLOSQUET, S. (2012). Schema.org | drupal.org. [Online] Available at: <https://drupal.org/project/schemaorg>, [Accessed 18-July-2013].
- DRUPAL.ORG (2012). Taxonomy | drupal.org. [Online] Available at: <https://drupal.org/node/1577546>, [Accessed 18-July-2013].
- EATON, J. (2013). Voting API | drupal.org. [Online] Available at: <https://drupal.org/project/votingapi>, [Accessed 18-July-2013].
- ECONSULTATION (2012). eConsultation Research Project. [Online] Available at: <http://www.e-consultation.org/>, [Accessed 30-May-2012].
- EGOV (2012). International conference on eParticipation (ePart). [Online] Available at: <http://www.egov-conference.org/>, [Accessed 29-May-2012].
- EGOVIS (2012). International Conference on Electronic Government and the Information Systems (EGOVIS). [Online] Available at: <http://www.dexa.org/egovisedem2012>, [Accessed 29-May-2012].
- EPART (2012). International conference on eParticipation (ePart). [Online] Available at: <http://www.epart-conference.org>, [Accessed 29-May-2012].
- EU PROFILER (2012). EU Profiler - Voting Advice Application. [Online] Available at: <http://www.euprofiler.eu>, [Accessed 06-March-2012].
- EVOTING (2012). eVoting Project - Geneva. [Online] Available at: <https://www.evotest.ch/ge>, [Accessed 04-June-2012].
- GLASOVODITEL (2012). Glasovoditel for European Parliament elections 2009. [Online] Available at: <http://glasovoditel.eu/media.html>, [Accessed 07-March-2012].
- GORDON, C. (2011). Wikitools | drupal.org. [Online] Available at: <https://drupal.org/project/WikiTools>, [Accessed 18-July-2013].
- HAUG, N. (2011). AJAX Poll | drupal.org. [Online] Available at: https://drupal.org/project/ajax_poll, [Accessed 18-July-2013].
- HOLYROOD (2007). Holyrood - Voting Advice Application. [Online] Available at: <http://www.holyrood2007.net/Scotlandnew/index.php>, [Accessed 06-May-2010].
- I-SOCIETY (2012). International Conference on Information Society (i-Society). [Online] Available at: <http://www.i-society.eu/>, [Accessed 29-May-2012].

INTERNET SOURCES

- IAP2 (2012). IAP2 Spectrum of Public Participation. [Online] Available at: <http://www.iap2.org/associations/4748/files/spectrum.pdf>, [Accessed 06-June-2012].
- INTERNET WORLD STATS (2012). World Internet Usage Statistics News and World Population Stats. [Online] Available at: <http://www.internetworldstats.com>, [Accessed 13-March-2012].
- KARSHAKEVICH, A. (2013). User Relationships | drupal.org. [Online] Available at: https://drupal.org/project/user_relationships, [Accessed 18-July-2013].
- KENNEDY, C. (2012). Advanced Poll | drupal.org. [Online] Available at: <https://drupal.org/project/advpoll>, [Accessed 18-July-2013].
- KIESKOMPAS (2012). Provincial Elections 2011. [Online] Available at: <http://provincie.kieskompas.nl>, [Accessed 07-March-2012].
- KINNAIRD, P. (2012). Connect 2 Congress - Visual Analytics for Unprecedented Civic Oversight. [Online] Available at: <http://www.connect2congress.com>, [Accessed 04-June-2012].
- KOHVOLIT (2012). Let you know who really represents you. [Online] Available at: <http://en.kohvolit.eu>, [Accessed 07-March-2012].
- KOIMIPASVA (2005). Koimipasva - Voting Advice Application. [Online] Available at: <http://www.koimipasva.bg>, [Accessed 06-May-2010].
- LANG, B. (2013). filedepot | drupal.org. [Online] Available at: <https://drupal.org/project/filedepot>, [Accessed 18-July-2013].
- LATARNIK (2012). Lamplighter Election 2011. [Online] Available at: <http://latarnik.nq.pl>, [Accessed 07-March-2012].
- LAWENDE, M., MOL, M. & KRAAIJ, P. (2012). Rate | drupal.org. [Online] Available at: <https://drupal.org/project/Rate>, [Accessed 18-July-2013].
- MANOBALSAS (2012). Mano Balsas LT - A rational system of voting online. [Online] Available at: <http://www.manobalsas.lt>, [Accessed 07-March-2012].
- MILES, E. (2013). Views | drupal.org. [Online] Available at: <https://drupal.org/project/views>, [Accessed 18-July-2013].
- MOMENTUM (2012). MOMENTUM Project. [Online] Available at: <http://www.ep-momentum.eu/>, [Accessed 29-May-2012].

INTERNET SOURCES

- PICOLA (2012). PICOLA Deliberative Poll. [Online] Available at: <http://caae.phil.cmu.edu/picola/index.html>, [Accessed 30-May-2012].
- POLITICAL COMPASS (2012). The Political Compass. [Online] Available at: <http://www.politicalcompass.org>, [Accessed 07-March-2012].
- POLITIKKABINE (2012). Politikkabine - Voting Advice Application. [Online] Available at: <http://www.politikkabine.at>, [Accessed 07-March-2012].
- SMARTVOTE (2012a). Method Description - Smart Choice Vote Recommendation. [Online] Available at: http://smartvote.ch/downloads/methodology_recommendation_de_CH.pdf, [Accessed 01-March-2012].
- SMARTVOTE (2012b). Smartvote - Voting Advice Application. [Online] Available at: <http://smartvote.ch>, [Accessed 29-February-2012].
- SRIVASTAVA, S. (2013). DrupalChat | drupal.org. [Online] Available at: <https://drupal.org/project/drupalchat>, [Accessed 18-July-2013].
- STEMMENTRACKER (2012). StemmenTracker - The Rate Tracker. [Online] Available at: <http://www.stemmentracker.nl/>, [Accessed 07-March-2012].
- STEMWIJZER (2012). StemWijzer - Voting Advice Application. [Online] Available at: <http://www.stemwijzer.nl>, [Accessed 07-March-2012].
- TESTVOT (2012). Testvot - Voting Advice Application. [Online] Available at: <http://www.testvot.eu>, [Accessed 07-March-2012].
- TROJAN, J. (2012). Video Chat | drupal.org. [Online] Available at: <https://drupal.org/project/videochat>, [Accessed 18-July-2013].
- TROKY, S. & WEBBER, D. (2013). Advanced Forum | drupal.org. [Online] Available at: https://drupal.org/project/advanced_forum, [Accessed 18-July-2013].
- VIMENTIS (2012). Vimentis elections St. Gallen - Find your perfect candidate! [Online] Available at: <http://www.vimentis.ch/wahlen>, [Accessed 07-March-2012].
- VOTE MATCH (2012). Vote Match - Voting Advice Application. [Online] Available at: <http://www.votematch.org.uk>, [Accessed 07-March-2012].
- VOTE SMART (2012). Project Vote Smart's VoteEasy TM. [Online] Available at: <http://www.votesmart.org/voteeasy>, [Accessed 08-March-2012].

INTERNET SOURCES

- VOTEMATCH EUROPE (2012). VoteMatch - Voting Advice Application. [Online] Available at: <http://votematch.eu>, [Accessed 07-March-2012].
- W3C (2012). Semantic Web - W3C. [Online] Available at: <http://www.w3.org/standards/semanticweb/>, [Accessed 05-June-2012].
- WAHLKABINE (2012). Wahlkabine - Voting Advice Application. [Online] Available at: <http://www.wahlkabine.at>, [Accessed 08-March-2012].
- WAHLOMATH (2012). Wahlomath.de - the rational choice help. [Online] Available at: <http://www.wahlomath.de>, [Accessed 08-March-2012].
- WALC, W. (2013). CKEditor - WYSIWYG HTML editor | drupal.org. [Online] Available at: <https://drupal.org/project/ckeditor>, [Accessed 18-July-2013].
- WEBCAT (2012). WebCat :: A Web Content Analysys Tool. [Online] Available at: <http://webcat.sourceforge.net/>, [Accessed 30-May-2012].
- WHO DO I VOTE FOR (2012). Who Do I Vote For? - Voting Advice Application. [Online] Available at: <http://www.whodoivotefor.co.uk/>, [Accessed 08-March-2012].
- WORLD WIDE WEB CONSORTIUM (W3C) (2012). W3C SEMANTIC WEB ACTIVITY. [Online] Available at: <http://www.w3.org/2001/sw/>, [Accessed 28-August-2012].
- ZOUBI, T. & BAČELIĆ, S. (2012). AudioField | drupal.org. [Online] Available at: <https://drupal.org/project/audiofield>, [Accessed 18-July-2013].

INTERNET SOURCES

INTERNET SOURCES

INTERNET SOURCES

Curriculum Vitae

Personal Information

Luis Fernando Terán Tamayo

University of Fribourg
Department of Informatics
Boulevard de Pérolles 90
1700 Fribourg, Switzerland

E-mail: luis.teran@unifr.ch
Web: <http://diuf.unifr.ch/main/is/members/luis-teran>

Research Interests

eGovernment, eParticipation, eCollaboration, eDemocracy, eElection, eVoting, eCommunities, ePassports, Recommender Systems, and Fuzzy Classification.

Education

University of Fribourg, Switzerland.

Ph.D. in Computer Science from the Information Systems Research Group, expected January 2014.

Swiss Federal Institute of Technology Lausanne, Switzerland.

M.Sc., School of Computer and Communication Sciences, February 2009.

Escuela Politécnica Nacional, Quito, Ecuador.

B.Sc., Faculty of Electrical Engineering, June 2004.

Honors and Awards

PROMISE Winter School 2012 - Information Retrieval meets Information Visualization

Best Poster Award - Information Visualization, Zinal, Switzerland, 2012.

Secretaria Nacional de Ciencia y Tecnologia (SENACYT) - actual SENESCYT

Fellowship - Programa de Becarios Fondos CEREPS, Quito, Ecuador, 2007.

Escuela Politécnica Nacional - Faculty of Electrical Engineering

Graduation with distinction Cum Laude, Quito, Ecuador, 2004.

Journal Articles

Terán L., Drobnjak A.: An Evaluation Framework for eParticipation: The VAAs Case Study. *World Academy of Science, Engineering and Technology Journal*, Issue 73, pp. 819–827 (2013).

Terán L., Meier A.: SmartParticipation – A Fuzzy-Based Platform for Stimulating Citizens' Participation. *International Journal for Infonomics (IJI)*, Volume 4, Issue 3/4, pp. 501–512 (2011).

Book Chapters

Meier A., **Terán L.**: Schwerpunkt Partizipation – Beteiligung als Zukunftsmodell.- Wahlhilfesystem für elektronische Wahlen unter Nutzung der unscharfen Logik. In *E-Government – Zwischen Partizipation und Kooperation*, Springer, chapter 4, pp. 73–87 (2012).

Terán L., Ladner A., Fivaz J., Gerber S.: Using a Fuzzy-based Cluster Algorithm for Recommending Elections. In *Fuzzy Methods for Customer Relationship Management and Marketing: Applications and Classification*, ICI Global, chapter 6, pp. 115–138 (2011).

Conference Papers

Terán L.: A Fuzzy-Based Platform for Stimulating Citizens' Participation. In *International Conference on E-Business and E-Government (ICEE 2012)*, IEEE, Volume 4, Shanghai, pp. 2353–2357, China (2012).

Terán L.: A Fuzzy-Based Advisor for Elections and the Creation of Political Communities. In *International Conference on Information Society (i-Society 2011)*, IEEE, London, UK, pp. 196–201 (2011).

Terán L., Meier A.: A Fuzzy Recommender System for eElection. In *EGOVIS 2010, International Conference on Electronic Government and the Information Systems Perspective*, LNCS 6267, Springer, Bilbao, Spain, pp. 62–76 (2010).

Terán L., Drygajlo A.: On Development of Inspection Systems for Biometric Passports Using Java. In *BioID_MultiComm 2009, Joint COST 2101 AND 2102 International Conference*, LNCS 5707, Springer, Madrid, Spain, pp. 268–275 (2009).

Magazine Articles

Drobnjak A., Meier A., **Terán L.**: Ein Reifegradmodell für Voting Advice Applications. In *eGovernment Review - Regionales E-Partizipation-Modell*, FH Kärnten, Edition No. 12, pp. 18–19 (2013).

Meier A., **Terán L.**: Empfehlungssystem für elektronische Wahlen. In *eGovernment Review - Regionales E-Partizipation-Modell*, FH Kärnten, Edition No. 7, pp. 16–17 (2011).

Internal Working Papers

Drobnjak A., Fasel D., Hugi P., Kaufmann M., Meier A., Portmann E., Schütze R., **Terán L.**, Wehrle M., Zumstein D.: Führungsinformationssysteme unter Nutzung der unscharfen Logik – Fallbeispiel coop@home. Information Systems Research Group. University of Fribourg (2011).

Academic & Teaching Activities

Universal Accessibility Workshop – Ecuadorian Technical Secretariat of Disabilities (SETEDIS), Quito, Ecuador.

Guest Lecturer

August 15th–16th, 2013

– *Title of lecture:* eGovernment Framework.

Jornadas de Ingeniería – Universidad de las Américas (UDLA), Quito, Ecuador.

Guest Lecturer

June 3rd–7th, 2013

– *Title of lecture:* Marco de Evaluación para Participación Electrónica.

International Seminar – eDemocracy & eGovernment (www.edem-egov.org), Quito, Ecuador.

Lecturer

April 8th–12th, 2013

– *Titles of lectures:* eService, eContracting, eDemocracy, eCommunity, SmartParticipation Project.

University of Fribourg, Switzerland.

Teaching Assistant and Guest Lecturer

September 2009–Present

Curriculum Vitae

- Teaching assistant and guest lecturer for Prof. Dr. Andreas Meier.
Master course: Electronic Government - Course HS-2013.
Titles of lectures: eSettlement, eCollaboration, eDemocracy, eCommunity, Data-warehouse exercise session.
- Teaching assistant and guest lecturer for Prof. Dr. Andreas Meier.
Master course: Electronic Government - Course HS-2012.
Titles of lectures: eSettlement, eCollaboration, eDemocracy, eCommunity, Data-warehouse exercise session.
- Guest lecturer at International Institute of Management of Technology, University of Fribourg.
Master course: Information Management & Decision Support – Module 3 – 2012.
Title of lecture: SmartParticipation project.
- Teaching assistant and guest lecturer for Prof. Dr. Andreas Meier.
Master course: Electronic Government - Course HS-2011.
Titles of lectures: eContracting, ePassports, eDemocracy, and SmartParticipation project.
- Teaching assistant and guest lecturer for Prof. Dr. Andreas Meier.
Master course: Electronic Government - Course HS-2010.
Titles of lectures: eContracting, ePassports, eDemocracy, and Fuzzy Recommender System project.

Teaching Assistant

September 2009–Present

- Teaching assistant for Prof. Dr. Andreas Meier.
Master course: Electronic Business - Course FS-2013.
- Teaching assistant for Prof. Dr. Andreas Meier.
Bachelor course: Databases for Bachelor Studies - HS-2012.
- Teaching assistant for Prof. Dr. Andreas Meier.
Bachelor course: Databases for Bachelor Studies - HS-2011.
- Teaching assistant for Prof. Dr. Andreas Meier.
Master course: Electronic Business - Course FS-2011.
- Teaching assistant for Prof. Dr. Andreas Meier.
Master course: Electronic Government - Seminar HS-2009.
- Teaching assistant for Prof. Dr. Andreas Meier.
Bachelor course: Databases for Bachelor Studies - HS-2009.

Swiss Federal Institute of Technology Lausanne, Lausanne, Switzerland.

Teaching Assistant

October 2010–February 2012

- Teaching assistant for Dr. Pearl Pu.
Master course: Human Computer Interaction.

Teaching Assistant and Guest Lecturer

June 2008–January 2009

- Teaching assistant and guest lecturer for Dr. Andrezej Drygajlo.
Master course: Biometrics.
Title of lecture: Implementation of an Inspection System for Biometric Passports based on ICAO Specifications.

Lucerne University of Applied Science and Arts, Lucerne, Switzerland.

Guest Lecturer

October 2010

- Guest lecturer at Lucerne University of Applied Science and Arts.
Master course: Business Information Analytics.
Title of lecture: Social Networking & Open Source.

Workshops

- Information Systems Research Group Workshop, Monto di Comino, Switzerland, May 2013.
- Information Systems Research Group Workshop, Lucerne, Switzerland, April 2011.
- Swiss e-Voting Workshop 2010, Fribourg, Switzerland, September 2010.
- Information Systems Research Group Workshop, Schwarzsee, Switzerland, April 2010.

Further Education

- 3rd ESWC Summer School 2013, Crete, Greece, September 2013.
- The Structure and Logic of Scientific English & Writing Titles and Abstracts, University of Geneva, Switzerland, February 2013.
- How does a Scientific Journal Work?, Doctoral Program in Computer Science, Conférence Universitaire de Suisse Occidentale (CUSO), University of Neuchâtel, Switzerland, October 2012.
- 2012 IFI Summer School, Department of Informatics, Business Intelligence Research Group, University of Zurich, Switzerland, June 2012.
- PROMISE Winter School 2012, Information Retrieval meets Information Visualization, Zinal, Switzerland, January 2012.
- CUSO Winter School in Computer Science, Managing and Engineering Complex Systems, Les Diablerets, Switzerland, January 2011.
- Human Computer Interaction, Master Course, Swiss Federal Institute of Technology Lausanne, Lausanne Switzerland, February–June 2010.

Project Experience

Swiss Federal Institute of Technology Lausanne, Lausanne, Switzerland.

Optional Project

September 2007–February 2008

“SWANS for TRANS: support for large scale realistic simulations of VANETs.” Supervisor: Prof. Dr. Jean-Pierre Hubaux; Assistants: Maxim Raya and Michal Piorkowski. Laboratory for Computer Communications and Applications.

Semester Project

March 2007–June 2007

“Experimenting with the IEEE 802.11 protocol in ad hoc mode.” Supervisor: Prof. Dr. Patrick Thiran; Assistant: Mathilde Durvy. Laboratory for Computer Communications and Applications.

Professional Experience

TELCONET, Quito, Ecuador

Logistics Department

December 2005–September 2006

- Head of Logistics Department.
- VIP Engineer in charge of key customers.

Curriculum Vitae

VIP Engineer

March 2005–December 2005

- VIP Engineer in charge of key customers.
- Network administration.

Senior Technician

March 2004–March 2005

- Installation and administration of microwave, fiber optic, SCPC and VSAT links.
 - Network administration.
-

Technical Skills

Extensive hardware and software experience in networking and information technology

MATLAB experience: linear algebra, Fourier transforms, nonlinear numerical methods, polynomials, statistics, visualization, communications, filter design, signal processing, and others.

Programming: C, C++, UNIX shell scripting, SQL, Java, Java Cards, and others.

Reporting Software: T_EX, L^AT_EX, B_IB_TE_X, Microsoft Office, and other common productivity packages for Windows, OS X, and Linux platforms.

Operating Systems: Microsoft Windows XP/2000, Apple OS X, Linux, and other UNIX variants.

Web Publishing: HTML, Javascript, XML, PHP, Drupal, Joomla, MODx and others.

Mathematical Expertise

Fuzzy Logic, Linear and Nonlinear Systems Theory, Probability, Random Variables, Stochastic Processes, and Game Theory.