Nikos Manouselis · Hendrik Drachsler
Katrien Verbert · Olga C. Santos *Editors*

Recommender Systems for Technology Enhanced Learning

Research Trends and Applications

Springer

Recommender Systems for Technology Enhanced Learning

Nikos Manouselis • Hendrik Drachsler
Katrien Verbert • Olga C. Santos
Editors

Recommender Systems for Technology Enhanced Learning

Research Trends and Applications

Foreword by Joseph A. Konstan

Springer

Editors
Nikos Manouselis
Agro-Know
Athens, Greece

Katrien Verbert
Department of Computer Science
VUB & KU Leuven
Leuven, Belgium

Hendrik Drachsler
Welten Institute – Research Centre for Learning, Teaching and Technology
Faculty of Psychology and Educational Sciences
Open University of the Netherlands
Heerlen, The Netherlands

Olga C. Santos
aDeNu Research Group, Artificial Intelligence Department
Computer Science School, UNED
Madrid, Spain

ISBN 978-1-4939-0529-4 ISBN 978-1-4939-0530-0 (eBook)
DOI 10.1007/978-1-4939-0530-0
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2014935878

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Foreword

It was an inauspicious beginning in Barcelona in 2010. I had agreed to give talk to a workshop I hadn't heard of before on Recommender Systems for Technology Enhanced Learning. That morning was sunny and hot, and the city's usually efficient transit was on strike. I was advised that the easiest way to get to the workshop would be a long walk, so I set off for the workshop reflecting on the theme of my talk—that recommender systems had great potential in education, but that we weren't there yet. Arriving hot and tired, I re-told the story I'd been telling for almost 15 years—about how recommending products was relatively easy, and that it was a quick win for the technology. Product recommenders certainly have improved quality of life—making shopping and television watching easier. But for people seeking a deeper impact, they may fall short.

By contrast, education raised all sorts of challenges for recommender systems. But it also presented the potential for a deep win—for making a difference that would affect the quality of life for billions of people. The technical challenges are formidable. Education is fundamentally interdependent and sequential. A learning module or lesson that may be ideal for a student at one time may be completely useless too early or too late. So in a very real way, technology-enhanced learning should be a "grand challenge" for recommender systems researchers—but at that time, it mostly wasn't happening.

There were many reasons why. Making progress on educational recommenders presented at least three formidable obstacles to the typical recommender systems researcher. First, the researcher needed to gain understanding of education and learning research—any successful effort in education would require such an understanding. Second, the researcher would need real datasets—part of the challenge at the time was the lack of large datasets in general and of cases where there are more than one or two alternatives for given content modules specifically. And third, the researcher would need to learn how to conduct meaningful evaluation—this is no longer simply a question of which learning modules a student "prefers" but of what leads to actual learning, competence, and performance, not just on an immediate post-test basis, but later as the knowledge gets integrated. So while I was happy to lead the cheers for the whole area of RecSysTEL, and enjoyed seeing the work being done at the time, I left that day somewhat discouraged that this field would remain in the margins.

Three years later, how things have changed! Who knew that we'd have online courses with tens and hundreds of thousands of students? And who would have expected entire campuses (physical and virtual) committed to the idea of scientific exploration of personalised education? We are surely entering an era of new interest and new possibilities.

But what's most exciting is that we are entering that area through strength. As I look through the collection of articles in this book, I see a variety of advances that bring together the best ideas in recommender systems with important TEL applications. It is gratifying to see the expansion of available datasets that can allow researchers to explore ideas offline first, and even more gratifying to see the increased diversity of research approaches and questions—with issues ranging from trust to affect, and methods ranging from data analysis to field and experimental research.

So we are entering what may well become the golden age of RecSysTEL research, and this is a well-timed volume to help bring those new to the field up to speed.

Minneapolis, MN Joseph A. Konstan

Preface

Technology-enhanced learning (TEL) aims to design, develop, and test socio-technical innovations that will support and enhance learning practices of both individuals and organisations. It is an application domain that generally addresses all types of technology research and development aiming to support teaching and learning activities, and considers meta-cognitive and reflective skills such as self-management, self-motivation, and effective informal and self-regulated learning. It was in 2007 when our first efforts to create opportunities for researchers working on topics related to recommender systems for TEL found their way in workshops like the Workshop on Social Information Retrieval for Technology Enhanced Learning (SIRTEL), the Workshop on Context-Aware Recommendation for Learning, and the Workshop Towards User Modelling and Adaptive Systems for All (TUMAS-A).

Still, it was only in 2010 when a really rare opportunity rose: during the same week of September and at the same location (Barcelona, Spain), two very prestigious and very relevant events (the fourth ACM Conference on Recommender Systems and the fifth European Conference on Technology Enhanced Learning) took place, giving us the chance to bring the two communities together. And so we did, by organising a joint event called the *1st Workshop on Recommender Systems for Technology Enhanced Learning (RecSysTEL)*.

Since then, lots of things have happened to mainstream educational applications in recommender systems' research. The most important achievement is an initial pool of datasets that have been collected and can be used to compare the outcomes of different TEL Recommender Systems to create a body of knowledge about the effects of different algorithms on learners. Furthermore, running research projects like Open Discovery Space[1] and LinkedUp[2] aim to create a publicly accessible Linked Data cloud[3] that can be used as a reference dataset for RecSysTEL research. Along these infrastructure improvements various scientific events and publications

[1] www.opendiscoveryspace.eu/

[2] www.linkedup-project.eu/

[3] http://data.linkededucation.org/linkedup/catalog/

have been realised. The most relevant are the organisation of subsequent editions of the RecSysTEL workshop with bi-annual periodicity; authoring a review article for the Recommender Systems Handbook; expanding it to an introductory handbook on Recommender Systems for Learning; and contributing (as co-editors or as authors) to several relevant Special Issues in scientific journals and specialised books.

We thought that this is a good time to build upon this previous experience and to collect some state-of-the-art contributions to a volume that will give a fresh view of the status of this area. Our interest was to collect a representative sample of high-quality manuscripts that will illustrate some important research trends, identify key challenges and demonstrate some innovative applications. This volume is the result of an open call that helped us collect, peer-review, select and propose for publication 14 articles (out of 49 proposed works; 29 % acceptance rate) that give a very good picture of the current status of research in recommender systems for TEL. The first four chapters (Karampiperis et al.; Cenichel et al.; Dietze et al.; Bienkowski and Klo) deal with user and item data that can be used to support recommendation systems and scenarios. The next four (Hulpus et al.; Santos et al.; Schwind and Buder; Tang et al.) focus on innovative methods and techniques for recommendation purposes. And the last six (Fazeli et al.; Bielikova et al.; Nowakowski et al.; Fernandez et al.; Sie et al.; Petertonkoker et al.) present examples of educational platforms and tools where recommendations are incorporated.

The bibliography covered by this book is available in an open group created at the Mendeley research platform[4] and will continue to be enriched with additional references. We would like to encourage the reader to sign up for this group and to connect to the community of people working on these topics, gaining access to the collected bibliography but also contributing pointers to new relevant publications within this very fast developing domain.

We hope that you will enjoy reading this volume as much as we enjoyed editing it.

Athens, Greece	Nikos Manouselis
Heerlen, The Netherlands	Hendrik Drachsler
Leuven, Belgium	Katrien Verbert
Madrid, Spain	Olga C. Santos

[4] http://www.mendeley.com/groups/1969281/recommender-systems-for-learning/

Acknowledgements

We would like to thank all the people that are continuously inspiring and contributing to this application domain of recommender systems. We would particularly like to thank the people that are listed below (in alphabetical order), since they have contributed to the quality of this volume by reviewing the submissions received, providing valuable arguments that helped us select the most appropriate contributions as well as giving valuable feedback to the authors so that they improve their final articles. This list includes the people that have served as part of the RecSysTEL workshops Steering Committees and/or Program Committees as well as in the Review Committee for this volume.

Alexander Felfernig, Graz University of Technology (Austria)
Brandon Muramatsu, Massachusetts Institute of Technology (USA)
Carlos Delgado Kloos, University of Carlos III de Madrid (Spain)
Charalampos Karagiannidis, University of Thessaly (Greece)
Christina Schwind, Knowledge Media Research Center (Germany)
Christoph Rensing, Technische Universität Darmstad (Germany)
Cristian Cechinel, Fundação Universidade Federal do Pampa (Brazil)
Cristóbal Romero, University of Cordoba (Spain)
David Massart, ZettaDataNet, LLC (USA)
Davinia Hernández-Leo, Universitat Pompeu Fabra (Spain)
Denis Gillet, Swiss Federal Institute of Lausanne (Switzerland)
Denis Parra, University of Pittsburgh (USA)
Eelco Herder, L3S (Germany)
Elina Megalou, Computer Technology Institute and Press – Diophantus (Greece)
Erik Duval, Katholieke Universiteit Leuven (Belgium)
Felix Mödritscher, Vienna University of Economics and Business (Austria)
Fridolin Wild, Open University (UK)
Geert-Jan Houben, Technical University Delft (The Netherlands)
George Kyrgiazos, National Technical University of Athens (Greece)
Giannis Stoitsis, Agro-Know (Greece)

Hannes Ebner, Royal Institute of Technology (Sweden)
Hans-Christian Schmitz, Fraunhofer FIT (Germany)
Ivana Bosnic, University of Zagreb (Croatia)
Jad Najjar, Eummena (Belgium)
Jan Pawlowski, University of Jyväskylä (Finland)
Jesus G. Boticario, UNED (Spain)
Joel Duffin, Tatemae (USA)
Jon Dron, Athabasca University (Canada)
Joris Klerkx, Katholieke Universiteit Leuven (Belgium)
Julien Broisin, Université Paul Sabatier (France)
Katrin Borcea-Pfitzmann, Dresden University of Technology (Germany)
Leonardo Lezcano, University of Alcala (Spain)
Liliana Ardissono, Universita di Torino (Italy)
Maiga Chang, Athabasca University (Canada)
Martin Memmel, German Research Center for Artificial Intelligence (Germany)
Martin Wolpers, Fraunhofer FIT (Germany)
Miguel-Angel Sicilia, University of Alcala (Spain)
Mimi Recker, Utah State University (USA)
Nikolas Athanasiadis, Intrasoft Int. (Luxembourg)
Paul Libbrecht, Karlsruhe University of Education and Martin Luther University of Halle (Germany)
Pedro J. Munoz Merino, University of Carlos III de Madrid (Spain)
Peter Brusilovsky, University of Pittsburgh (USA)
Peter Scott, Open University (UK)
Ralf Klamma, RWTH Aachen University (Germany)
Rick D. Hangartner, Strands (USA)
Riina Vuorikari, European Schoolnet (Belgium)
Rita Kuo, Knowledge Square, Inc. (Taiwan)
Rosta Farzan, Carnegie Mellon University (USA)
Salvador Sanchez-Alonso, University of Alcala de Henares (Spain)
Sandy El Helou, Swiss Federal Institute of Lausanne (Switzerland)
Sergey Sosnovsky, DFKI GmbH (Germany)
Sotiris Konstantinidis, University of Athens (Greece)
Stavros Demetriadis, Aristotle University of Thessaloniki (Greece)
Stefan Dietze, University Hanover, (Germany)
Stefanie Lindstaedt, Know-Center Graz (Austria)
Sten Govaerts, K.U.Leuven (Belgium)
Tiffany Tang, Kean University (USA)
Toby Dragon, Saarland University (Germany)
Tomislav Šmuc, Rudjer Bošković Institute (Croatia)
Tsukasa Hirashima, Hiroshima University (Japan)
Wolfgang Greller, Open University of the Netherlands (Netherlands)
Wolfgang Nejdl – L3S & Leibniz Universitat (Germany)
Wolfgang Reinhardt, University of Paderborn (Germany)
Xavier Ochoa, Escuela Superior Politecnica del Litoral (Ecuador)

The compilation of the work presented in this book has been carried out with European Commission and national funding support. More specifically, the involvement of Nikos Manouselis has been supported by the EU project Open Discovery Space (ODS)—297229 of the CIP PSP Programme (http://opendiscoveryspace.eu), the work of Hendrik Drachsler by the EU project LinkedUp—317620 of the FP7 Programme, and the work of Olga C. Santos by the EU project "European Unified Approach for Accessible Lifelong learning" (EU4ALL)—034778 of the IST FP6 (http://www.eu4all-project.eu/) and the national project "Multimodal approaches for Affective Modelling in Inclusive Personalized Educational scenarios in intelligent Contexts" (MAMIPEC) funded by the Spanish Ministry of Science and Innovation (TIN2011-29221-C03-01). Katrien Verbert is a post-doctoral fellow of the Research Foundation—Flanders (FWO). This publication reflects the views only of the authors, and the funding bodies cannot be held responsible for any use that may be made of the information contained therein.

Contents

Part I User and Item Data

Collaborative Filtering Recommendation of Educational Content in Social Environments Utilizing Sentiment Analysis Techniques 3
Pythagoras Karampiperis, Antonis Koukourikos, and Giannis Stoitsis

Towards Automated Evaluation of Learning Resources Inside Repositories 25
Cristian Cechinel, Sandro da Silva Camargo, Salvador Sánchez-Alonso, and Miguel-Ángel Sicilia

A Survey on Linked Data and the Social Web as Facilitators for TEL Recommender Systems 47
Stefan Dietze, Hendrik Drachsler, and Daniela Giordano

The Learning Registry: Applying Social Metadata for Learning Resource Recommendations 77
Marie Bienkowski and James Klo

Part II Innovative Methods and Techniques

A Framework for Personalised Learning-Plan Recommendations in Game-Based Learning 99
Ioana Hulpuş, Conor Hayes, and Manuel Oliveira Fradinho

An Approach for an Affective Educational Recommendation Model 123
Olga C. Santos, Jesus G. Boticario, and Ángeles Manjarrés-Riesco

The Case for Preference-Inconsistent Recommendations 145
Christina Schwind and Jürgen Buder

Further Thoughts on Context-Aware Paper Recommendations for Education 159
Tiffany Y. Tang, Pinata Winoto, and Gordon McCalla

Part III Platforms and Tools

Towards a Social Trust-Aware Recommender for Teachers 177
Soude Fazeli, Hendrik Drachsler, Francis Brouns, and Peter Sloep

ALEF: From Application to Platform for Adaptive Collaborative Learning 195
Mária Bieliková, Marián Šimko, Michal Barla, Jozef Tvarožek, Martin Labaj, Róbert Móro, Ivan Srba, and Jakub Ševcech

Two Recommending Strategies to Enhance Online Presence in Personal Learning Environments 227
Samuel Nowakowski, Ivana Ognjanović, Monique Grandbastien, Jelena Jovanovic, and Ramo Šendelj

Recommendations from Heterogeneous Sources in a Technology Enhanced Learning Ecosystem 251
Alejandro Fernández, Mojisola Erdt, Ivan Dackiewicz, and Christoph Rensing

COCOON CORE: CO-author REcommendations Based on Betweenness Centrality and Interest Similarity 267
Rory L.L. Sie, Bart Jan van Engelen, Marlies Bitter-Rijpkema, and Peter B. Sloep

Scientific Recommendations to Enhance Scholarly Awareness and Foster Collaboration 283
Jan Petertonkoker, Wolfgang Reinhardt, Junaid Surve, and Pragati Sureka

Part I
User and Item Data

Collaborative Filtering Recommendation of Educational Content in Social Environments Utilizing Sentiment Analysis Techniques

Pythagoras Karampiperis, Antonis Koukourikos, and Giannis Stoitsis

Abstract Collaborative filtering techniques are commonly used in social networking environments for proposing user connections or interesting shared resources. While metrics based on access patterns and user behaviour produce interesting results, they do not take into account qualitative information, i.e. the actual opinion of a user that used the resource and whether or not he would propose it for use to other users. This is of particular importance on educational repositories, where the users present significant deviations in goals, needs, interests and expertise level. In this paper, we examine the benefits from introducing sentiment analysis techniques on user-generated comments in order to examine the correlation of an explicit rating with the polarity of an associated text, to retrieve additional explicit information from user comments when a standard rating is missing and expand tried recommendation calculation with qualitative information based on the community's opinion before proposing the resource to another user.

Keywords Recommender systems • Educational repositories • Sentiment Analysis • Qualitative analysis

P. Karampiperis (✉) • A. Koukourikos
Software and Knowledge Engineering Laboratory, Institute of Informatics and Telecommunications, National Center for Scientific Research "Demokritos"Agia Paraskevi Attikis, P.O.Box 60228, 15310 Athens, Greece
e-mail: pythk@iit.demokritos.gr; kukurik@iit.demokritos.gr

G. Stoitsis
Agro-Know Technologies, Grammou 17 Str., Vrilissia Attikis, 15235 Athens, Greece

Universidad de Alcalá, Pza. San Diego, s/n, 28801 Alcalá de Henares, Madrid, Spain
e-mail: stoitsis@ieee.org

N. Manouselis et al. (eds.), *Recommender Systems for Technology Enhanced Learning: Research Trends and Applications*, DOI 10.1007/978-1-4939-0530-0_1,

Introduction

Recommender Systems are of particular importance within social environments, where users share access to a common set of resources. The variability of crucial user characteristics, like their background, their special interests, their degree of expertise, pose interesting issues in terms of proposing a resource that is interesting, useful and comprehensible to a particular user.

Collaborative filtering approaches based on explicitly given user ratings do not always reflect the differentiation between the various criteria that apply to a resource and the weight that the users give to each criterion. On the other hand, techniques that examine access patterns may suffer from the appearance of stigmergy phenomena. That is, the resources that are more popular or favourably regarded by the community at a given time tend to be favoured as recommendations to new users. The visibility of a resource, or even more elaborate features like the time spent in a resource, the amount of downloads etc. are not directly connected to its quality or suitability. Hence, the examination of access and use patterns can lead to poor recommendation that will be further propagated due to the users continuing to follow previously defined paths within the repository of available content.

The evolvements of Web 2.0, however, led to the provision of more explicit information from the user side. User comments, discussions and reviews can constitute valuable information for determining the quality, appeal and popularity of a resource.

In this context, we propose the exploitation of user generated comments on the resources of a repository of educational content in order to deal with the lack of explicit ratings and discover qualitative information related to a specific resource and the impressions it left to the users that accessed it. To this end, we applied sentiment analysis to comments on educational content and examined the accuracy of the results and the degree to which these comments reflect the perceived user satisfaction from the content. At this stage, a Collaborative Filtering Recommendation system was built, that is, content characteristics and features were not taken into account in the analysis.

The rest of the paper is structured as follows. We provide an overview of Collaborative Filtering approaches in "Collaborative Filtering Recommender Systems". Our quality-centric approach on Collaborative Filtering Recommendation is analysed in "Quality-Centric Recommender System Methodology". "Sentiment Analysis Techniques for Collaborative Filtering Recommender Systems" describes the Sentiment Analysis techniques that were implemented and examined for incorporation in a Recommender System. The experimental setup for determining the appropriateness of these Sentiment Analysis techniques and evaluating our Recommender System is described in "Experimental Setup", while the experimental results are presented in "Experimental Results." We conclude and define our next steps in "Conclusions and Future Work".

Collaborative Filtering Recommender Systems

Recommender systems aim to predict the preferences of an individual (user/customer) and provide suggestions of further resources or entities (other users of the same system, resources, products) that are likely to be of interest.

In broad terms, a recommender system can be defined formally as follows; Let U be the set of all users of a system and R the set of all items (resources) that are available within the system and accessible by the users. A utility function $f: U \times R \rightarrow S$, associates a score to user-resource pairs, which indicates the suitability of the specific resource to the specific user. As it is obvious, the common case for environments with that structure is that there do not exist scores for every pair in $U \times R$. To this end, the role of a recommender system is to "predict" the scores for the user-resource pairs that do not have a score readily available.

The main approaches for building a recommender system, i.e. defining the characteristics that are taken into account by the utility function employed, are the following:

- Content-based approaches; the utility function examines the similarity of new/unknown items with the ones already declared as likeable by the user and proposes the most similar to him/her.
- Collaborative filtering approaches; the recommendations provided to the user are based on the explicit usefulness declared by other users with similar tastes and activity with him/her.
- Hybrid approaches that combine the characteristics of the previous methods.

Collaborative recommender systems can generally be grouped into heuristic-based and model-based systems [1, 2]. In the first case, the score for a user-resource pair is calculated using the scores of other users for the examined resource. The main goals in heuristic-based approaches are to determine user similarity and the way that the degree of similarity is used to weigh the effect of a user's activity to another user's preferences. Various metrics have been examined for computing user similarity, like calculating the angle between the rating vectors of the users [3], computing the mean squared difference of users' ratings [4] and calculating the correlation coefficient between a given pair of users [5]. In the latter case, existing scores are used to construct a rating model, to which the predicted scores are expected to conform. Similarly, the aggregation of peer ratings to produce a predicted rating can be achieved in various ways, such as calculating the average of the ratings of similar users, using a weighted average where the weights are based on the degree of similarity etc. In the case of model-based approaches,

The usage of recommender systems is widely spread in e-commerce environments [6] but the general principle is applicable to multiple and diverse environments. In the case of TEL, multiple solutions have been proposed and examined [7, 8]. The proposed systems use a variety of methods and elements for producing recommendations. For example, RACOFI [9] takes into account user ratings and content associations, CoFind [10] applies folksonomies to better define context and purpose before producing a recommendation, while [11] exploits a multi-attribute rating of educational resources. Further techniques from different fields have been used in TEL Recommenders, like creating clusters of users based on their interests [12], ontology-based strategies [13] and methods that combine social and contextual information [14].

Due to the particularities of the domain, some of the most common algorithms for collaborative filtering have been shown to struggle in the setting of a learning object repository [15, 16]. Furthermore, the explosion of Social Networking environments in the context of Web2.0 has established new interesting issues and proposed solutions for the field [17–19] and urged the pre-dominance of collaborative filtering methods in many environments with such functionality. The incorporation of social networking functionality in educational repositories is continuously increasing. Platforms like MERLOT [20] and Organic. Edunet [21] offer to their users the ability to comment on the presented material, stating their opinions or remarks regarding various aspects of the available content.

Taking into account the previous two statements, the presented service tries to exploit the newly-introduced information from user comments and reviews and examine an alternative approach for producing recommendations of educational resources. As mentioned, the presented techniques are to be incorporated in a recommender system over a social platform that provides access to educational content. Linguistic techniques, such as sentiment analysis, can be of use for alleviating some of the drawbacks of traditional algorithms in terms of differentiating users belonging in different audiences (e.g. teachers from students) and bypassing the need for explicit ratings (via a star system).

Quality-Centric Recommender System Methodology

The proposed Collaborative Filtering methodology is based on the users' social connectivity and their rating activities within the relevant Social environment in order to compute the expected ratings for resources unknown to the user. It should be noted that the approach is domain-agnostic; hence additional information for the resources (metadata, categorization, keywords etc.) is not examined by the method. In this paper, the equations and scores rely on the widely used 5-scale rating system.

For the purposes of our research we consider a social service that incorporates the following functionalities:

- Membership: The service enables user registration and associates each user with a unique account.
- Organisation in Communities: Users can create communities (groups) within the social environment. The communities can be considered as subsets of the overall social network.
- Rating attribution: The users of the service can apply a rating to the resources available through the system, by assigning 1–5 stars to the resource.

- Comment submission: The users can comment on resources, essentially providing a review in natural language. The length of a comment can be arbitrary. For the purposes of the experiment, we consider comments written in English.

In this context, the activities of interest from a registered user are the ones of (a) assigning a rating to a resource and (b) commenting on a resource. The purpose of the recommendation service, therefore, is twofold; to generate a rating of the user from his/her actual activities (ratings and comments); and to generate a rating for resources for which there is no activity from the particular user. We consider the first case as an *explicit* rating for the user-resource pair, while in the second case, we consider the rating *implicit*. In the next paragraphs, we proceed to elaborate on the two types of ratings and their formal definition for the proposed recommender system.

Explicit Rating**:** The proposed system relies on the attribution of a rating to a resource by a user. This rating could be direct, via the aforementioned star system, or indirect, via the analysis of comments/discussion related to the specific resource. These ratings—if existent—are used to provide a *score* for a user-resource pair. The score is defined as:

$$Score(u,r) = \begin{cases} Rating(u,r), \left[\exists Rating(u,r)\right] AND \left[\nexists Comment(u,r)\right] \\ Sentiment(u,r), \left[\nexists Rating(u,r)\right] AND \left[\exists Comment(u,r)\right] \\ \dfrac{Rating(u,r) + Sentiment(u,r)}{2}, otherwise \end{cases} \quad (1)$$

Where $Rating(u, r)$ is the explicit rating of the resource r by user u and $Sentiment(u, r)$ is the sentiment score assigned by the sentiment analysis that will be applied to user comments and is described in detail in "Sentiment Analysis Techniques for Collaborative Filtering Recommender Systems".

Implicit Rating using User Similarity**:** In the case that a user has not explicitly provided a rating for a specific resource, the recommender system provides a predicted rating, taking into account the user's similarity with other users who have actually provided a rating for this resource.

This similarity represents the trust of a certain user to the expressed opinion of other users of the social service. In the relevant literature, it is evident that this type of information can provide meaningful recommendations in the case where no explicit qualitative information has been provided by the user himself/herself [17, 18, 22].

The calculation of the predicted rating relies on the similarity and distribution of scores provided by the system's users. Specifically, the predicted score given to a resource r by user u is defined as:

$$Score(u,r) = \frac{1}{2|S|} \cdot \left(\sum_{i=1}^{|S|} \left(L(u,S_i) + P(u,S_i) \right) Score(S_i,r) \right) \quad (2)$$

In this function, S is the set of system users that have provided an explicit score for the resource r. The L metric is a modification of the trust-centred algorithm proposed by [23] and is defined as:

$$L(a,b) = \left(1 - \frac{1}{5 \cdot |C|} \sum_{i=1}^{|C|} \left| Score(a,i) - Score(b,i) \right| \right)$$

The set $C = R_a \cap R_B$ is the conjunction of the sets R_a and R_b of resources that bear explicit scores provided by users a and b respectively. So, in broad terms, L is a measure for the similarity in the ratings of users a and b. The score difference is normalized to the [0, 1] space, since the ratings on the examined dataset belong in the (0, 5] range.

P is the normalized Pearson correlation metric as applied to our system. Specifically,

$$P(a,b) = \begin{cases} \frac{1}{5}\overline{Score(a,Ra)}, & \left[Score(a,i) = \overline{Score(a,R_a)} \forall i \in C \right] \\ \frac{1}{5}\overline{Score(b,R_b)}, & \left[Score(b,i) = \overline{Score(b,R_b)} \forall i \in C \right] AND \\ & \left[\neg\left(Score(a,i) = \overline{Score(a,R_a)} \forall i \in C \right) \right] \\ \left(\frac{\sum_{i=1}^{|C|} \left(Score(a,i) - \overline{Score(a,R_a)} \right) \cdot \left(Score(b,i) - \overline{Score(b,R_b)} \right)}{\sqrt{\sum_{i=1}^{|C|} \left(Score(a,i) - \overline{Score(a,R_a)} \right)^2 \cdot \sum_{i=1}^{|C|} \left(Score(b,i) - \overline{Score(b,R_b)} \right)^2}} \right) & \end{cases}$$

The $Score(u,R_u)$ construct denotes the complete set of scores provided by a user u. Hence, the average of these scores is $\overline{Score(u,R_u)}$.

Both quantities participate in the definition of the proposed score, as L reflects the "trust" that the examined user can have in the opinions of others, while P computes the differentiations on their rating habits and adjusts the score accordingly.

***Community-driven Implicit Rating*:** In the case that a user has not explicitly provided a rating for a specific resource and, additionally, does not have any common activities with the other users of the system, i.e. $R_u \cap R_k = \varnothing \, \forall k \in U, k \neq u$, where U is the set of users known to the system, the recommendation module provides a rougher estimate for the scores to be proposed to the user by calculating the average of the scores provided by users belonging to the same communities with user u.

Formally, let $M(u,c)$ denote that user u is a member of community c. If $R_u \cap R_k = \varnothing \, \forall k \in U, k \neq u$ the estimated score of user u or a resource r is calculated by the following formula:

$$Score(u,r) = \begin{cases} \frac{5}{2}, T = \varnothing \\ \\ \frac{\sum_{i=1}^{|T|} Score(T_i, r)}{|T|} \end{cases} \tag{3}$$

$$T = \{ p \in U : \exists c \, M(p,c), M(a,c) \}$$

Sentiment Analysis Techniques for Collaborative Filtering Recommender Systems

Sentiment analysis regards extracting opinion from texts and classifying it into positive, negative or neutral valence [24]. Work on the field focuses on two general directions; lexical approaches and solutions using supervised machine learning techniques.

Lexical approaches rely on the creation of appropriate dictionaries. The terms present in the dictionary are tagged with respect to their polarity. Given an input text, the presence of dictionary terms is examined and the overall sentiment of the text is computed based on the existence of "positive" and "negative" terms within it. Despite its simplicity, the lexical approach has produced results significant better than "coin-toss" [25–27]. The way of constructing the lexica that are used for sentiment analysis is the subject of several works. In [27] and [28] the lexicons comprised solely adjective terms.

The usage of pivot words (like "good" and "bad") and their association with the target words is also a frequently met approach. In [29] and [30], the minimum path between each target word and the pivot terms in the WordNet hierarchy was calculated in order to determine the polarity of the term and its inclusion in the dictionary. In [26], the authors executed search queries with the conjunction of the pivot words and the target word given as input. The query that returned the most hits determined the polarity of the given word.

Machine learning techniques focus on the selection of feature vectors and the provision of tagged corpora to a classifier, which will be used for analysing untagged corpora. The most frequent routes for choosing the feature vectors are the inclusion of unigrams or n-grams, counting the number of positive/ negative words, the length of the document etc. The classifiers are usually implemented as a Naive Bayes classifiers or as Support Vector Machines [27, 31]. Their accuracy is dependent on the selection of the aforementioned feature vectors, ranging in the same space as the lexical approaches (63–82 %).

This section presents the sentiment analysis techniques that were examined and tested in order to identify the most suitable method for the case of the proposed Social Recommendation Service in terms of precision, recall and execution time. Sentiment analysis in the context of the Social Recommendation Service refers to the task of extracting the polarity of an opinionated text segment with respect to the quality of a certain resource. In this case, a number of different techniques could be applied (presented in the following subsections), with different performance and characteristics. The fact that we are dealing with user generated content drives us to take into account its unstructured nature and the potential unbalanced distribution it may present. This gives rise to the fact that our training set may be unbalanced and therefore learning may not be able to cope with such diversity in the number of instances per class. In this paper, we focus on lexical approaches for sentiment analysis, in order to avoid the consequences of erroneous training due to the distribution of the ratings in MERLOT (i.e. the positive ratings are much more than the negative ones). The following subsections discuss the techniques that we have implemented and tested for inclusion in the proposed recommender system.

Affective Term Frequency

This technique relies on the existence of sets of terms that bear a predetermined polarity. In most cases, there are two sets of terms; a set containing terms with positive polarity and a set containing terms with negative polarity. Let P be the set of terms bearing positive polarity and N the set of terms bearing negative polarity. Also, let $T = \{t_1, t_2, \cdots, t_n\}$, the set of distinct tokens in the text segment to be examined. We define a positive score for a token t as:

$$PosScore(t, P) = \begin{cases} 1, & t \in P \\ 0, & t \notin P \end{cases}$$

Similarly, the negative score for a token is defined as:

$$NegScore(t, N) = \begin{cases} 1, & t \in N \\ 0, & t \notin N \end{cases}$$

For the entire text, i.e. the complete set of tokens T, we define the positive score as:

$$PosScore(T, P) = \sum_{i=1}^{|T|} PosScore(t_i, P)$$

Similarly, the negative score for the entire text is defined as:

$$NegScore(T,N) = \sum_{i=1}^{|T|} NegScore(t_i, N)$$

We describe two distinct variations of this approach below.

Domain-agnostic Term Frequency

In this case, the sets of positive and negative terms are constant and known beforehand [27]. The algorithm discovers all the terms of a text segment that can be found in the positive set and all the terms that can be found in the negative set. For example, in the sentence "Don't you just love this camera? It's great!", the presence of sentiment-bearing terms determines the polarity of the overall statement.

Keeping in mind the previous definitions, the overall polarity of the text segment (normalized to the [−1, 1] space) is defined as:

$$Sentiment(T,P,N) = \frac{PosScore(T,P) - NegScore(T,N)}{\max\{PosScore(T,P), NegScore(T,N)\}} \quad (4)$$

Domain-Aware Term Frequency

In this variation of the term frequency approach, the sets of polarized terms are constructed from a corpus of already classified, in-domain text segments [31]. Every term found in segments characterized as positive is added to the positive set *P*. Similarly, every term found in segments characterized as negative is added to the negative set *N*. An example for showcasing the differentiation of term polarity with respect to the domain at hand is a text segment such as "This phone is amazing! It's so small and light", where the term "small" carries a positive valence, in contrast with the general notion that small is a negative attribute. In this method, we introduce the notion of neutral polarity, where the text segment was characterized as neither positive nor negative. Hence, the algorithm uses another set of terms *Neu*, in which the terms found in segments characterized as neutral are added. Similar to the cases of positive and negative sets, the neutral score of a token is equal to:

$$NeuScore(t, Neu) = \begin{cases} 1, & t \in Neu \\ 0, & t \notin Neu \end{cases}$$

And the neutral score for the entire text equals to:

$$NeuScore(T, Neu) = \sum_{i=1}^{|T|} NeuScore(t_i, Neu)$$

In this case, the function for determining the polarity of a text segment T is formulated as follows.

$$Sentiment(T,P,N,Neu) = \begin{cases} 0, & \begin{array}{l}[NeuScore(T,Neu) > PosScore(T,P)]\ AND \\ [NeuScore(T,Neu) > NegScore(T,N)]\end{array} \\ \dfrac{PosScore(T,P) - NegScore(T,N)}{\max\{PosScore(T,P), NegScore(T,N)\}}, & otherwise \end{cases}$$

Affective and Domain Terms Correlation

The main drawback of the previous approach is that it does take into account the fact that a certain text segment may express opinions not directly related to the entity we are actually interested in. For example, a comment on an educational object may contain praises for a similar work that is recommended as a possible reference. The following techniques try to address this problem by examining ways to associate the affective terms with specific tokens that refer to the entity for which we want to mine the writer's opinion.

Distance-Based Correlation

This method relies on the proximity of the domain and affective terms in order to determine which of the latter are more likely to determine the polarity of the text towards the entity of interest. Let $D = \{D_1, D_2, \cdots, D_n\}$ the set of terms/phrases that are used to define the entity of interest (e.g. "the paper", "the article", the title of the article etc.). For each element of D, we calculate the distance between the term and all the affective terms in the positive and negative sets, i.e. the number of words between the descriptive and the affective term. If there is not an affective term within the text segment, the distance is set to zero.

Dependency-Based Correlation

A more sophisticated approach for estimating the polarity of a text towards a specific entity is to examine if the terms associated with the latter are syntactically linked with one or more affective terms.

In this method, we split the input text into sentences and obtain the parse tree for each sentence by employing a shallow parser. For sentences containing a term or phrase that describes the entity of interest, we examine the dependencies of these terms from the parse tree. If the connected terms are also found in the positive or the negative sets, the PosScore and NegScore are respectively incremented by 1. Finally, we employ (4) to calculate the overall polarity of the text.

Experimental Setup

Experimental Corpus

The conducted experiments used the material available in the MERLOT platform. MERLOT is an online repository, providing open access to resources for learning and online teaching. It provides learning material of higher education aiming at promoting access to scientific data and as a result to their manipulation and exploitation by research communities.

MERLOT users can evaluate the available resources in two distinct ways. They can write comments on the resource, along with providing a rating in the 0–5 scale. We consider ratings of 0–2 as negative, ratings of 3 as neutral and ratings of 4–5 as positive. Additionally, MERLOT users can provide comments in a more formal manner, by submitting an “expert review”. Expert reviews follow a structured template. Reviewers can provide an overview of the examined content and evaluate it with respect to its (a) Content Quality; (b) Effectiveness as a teaching tool; and (c) Ease of Use for students and faculty. Figures 1, 2 and 3 depict a resource description, a user comment and an expert review respectively, as they are presented within MERLOT.

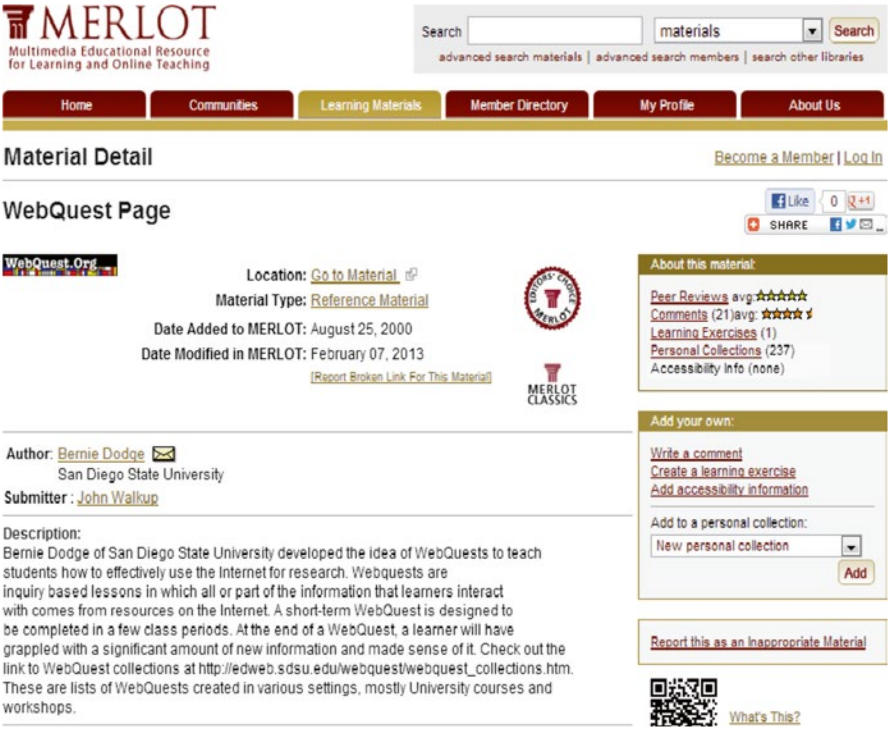

Fig. 1 A MERLOT resource as represented in the constructed XML file

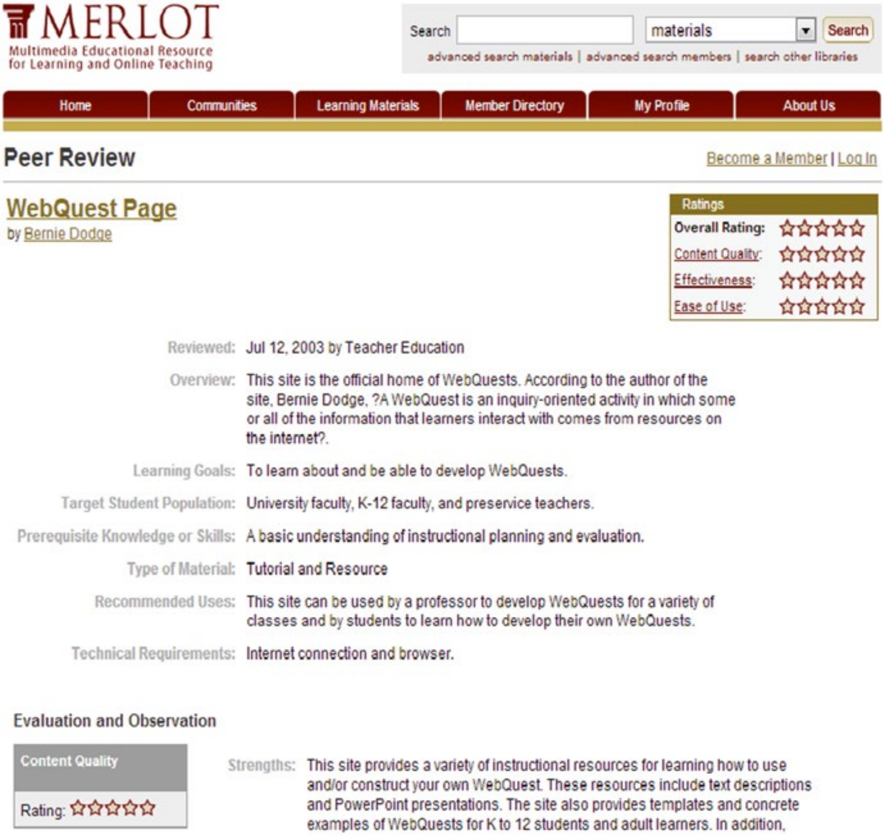

Fig. 2 Structure of the reviews element in a MERLOT resource description

Fig. 3 Structure of the comments element in a MERLOT resource description

In order to build our test corpus, the MERLOT user-generated content was retrieved via the following process: starting from a list of resources presented via the MERLOT web environment, we crawled the pages of each individual included object by following the appropriate hyperlinks in the object's page. For each resource, we retrieved the following elements:

- The title of the resource
- The description provided within MERLOT
- The keywords associated with the resource
- User comments provided for the resource
- Expert reviews submitted to MERLOT

For the user comments, we store the URL of the comment and the actual comment text. For the expert reviews associated with a resource, we store the following information:

- The URL of the review
- The ID code of the user that provided the review
- Information pertaining to the quality of the content, as expressed by the reviewer. This includes descriptions in natural language, of the strengths and concerns regarding the content quality.
- Information pertaining to the effectiveness of the examined resource, as a learning object. This includes descriptions, in free-form text, of the strengths and concerns regarding the effectiveness of the resource.
- Information pertaining to the ease of use of the resource, again indicating the strengths and weaknesses in free text.

All of the above information is organized and stored in an XML file, in order to facilitate the extraction of information with respect to the resources and the contributing users (via XPath querying). Each tag associated with a MERLOT resource encloses all the elements associated with it.

The dataset used for our experiments incorporates information for 6,720 MERLOT resources. There are 9,623 comments and 3,459 expert reviews in total. Hence, the average comment count is 1.43 comments per resource and 0.514 expert reviews per resource. The majority of the resources had 1 or 2 comments and no reviews. However, the maximum number of comments in the examined datasets was 23, while the maximum number of expert reviews was 4.

Sentiment Analysis Techniques

The described methods were tested in terms of precision, recall and execution time in order to reach to a decision for their suitability in the context of a recommendation service. As the recommendation methodologies have an execution overhead, we incorporate the execution time metric into the quality analysis of each implementation. This section provides a description of the dataset used for the experiments, the definition of the quality score for a sentiment analysis method and the results of the experiments on the given dataset.

Corpus Preparation

In order to build our test corpus, the MERLOT user-generated content was retrieved and underwent trivial linguistic processing (HTML cleaning, Stop-word removal, Lemmatization) before being fed to implementations of the aforementioned sentiment analysis methods.

Sentiment Analysis Quality Score

As mentioned, besides the precision and recall performance of a sentiment analysis method, we are especially concerned with the execution time of the respective module, as it will be incorporated into an already time-demanding recommendation process.

In this regard, we introduce a Quality Score (QS) metric for the implementation of each method. Let *P* denote the achieved precision from the application of the method on the MERLOT corpus, and *R* the achieved recall. Let also *T* denote the execution time (in milliseconds) for 1,000 iterations of the method (that is, the analysis of 1,000 specific textual segments) and *maxT*, *minT* the worst and best observed execution times, that is, the execution time of the slowest and fastest method respectively. The Quality Score for the method is defined as:

$$QS = \frac{2}{3} \cdot \left(\frac{P+R}{2} \right) + \frac{1}{3} \cdot \left(\frac{\max T - T}{\max T - \min T} \right)$$

Since $0 \leq P \leq 1$, $0 \leq R \leq 1$ and $T \leq maxT$, it is obvious that $0 \leq QS \leq 1$ and that a higher Quality Score indicates a more suitable method. The QS metric assigns a higher weight to the effectiveness of the method in comparison to its execution speed.

Quality-Centric Collaborative Filtering Recommender

Building the Training and Evaluation Sets

In this phase of the experiment, the retrieved corpus was divided in two subsets; the first subset was considered the initial input of the recommender system, that is, the records were considered as the information known to the service; the second subset was held as the evaluation corpus, that is, the information present was considered unknown and the results of the recommendation service are to be compared with