PROFILE REFINEMENT IN ONTOLOGY-BASED RECOMMANDER SYSTEMS FOR ECONOMICAL E-NEWS

Thomas HASSAN
Université de Bourgogne, LE2I, CNRS
thomas.hassan@u-bourgogne.fr
David WERNER
Université de Bourgogne, LE2I, CNRS
david.werner@u-bourgogne.fr
Aurélie BERTAUX
Université de Bourgogne, LE2I, CNRS
aurelie.bertaux@u-bourgogne.fr
Christophe CRUZ
Université de Bourgogne, LE2I, CNRS
christophe.cruz@u-bourgogne.fr

Abstract. *This paper is interested in a recommender system of economic news articles. More precisely, it focuses on automatic profile refinement of customers which is an important task over time by taken into account logs of the user concerning especially his/her actions, reading time, and domain specific knowledge. In our approach, ontologies are used to describe and automatically refine these profiles. This work focuses on one particular type of recommender systems which is content-based recommenders. The aim of these recommender systems is to build a user profile and to improve its precision over time. Several improvements that have been made to these recommender systems over the last decade are analyzed. We find that the improvements brought by the use of semantic knowledge are not negligible, therefore semantic web approaches should be more and more used in the future. Nevertheless improvements remain possible in this domain and further research could be interesting.*

Keywords: profile refinement, recommender system, ontology, economical e-news, Semantic web.
JEL classification: Innovation and Invention: Processes and Incentives O31

1. Introduction

The overload of news information is a particular case of information overload, which is a well-known problem, studied by Information Retrieval and Recommender Systems research fields. News recommender systems already exist such as [1], Athena [2], GroupLens [3] or News Dude [4]. A recommender system is used for the item ranking and a content-based approach is required to analyze the content of each article to structure information content. This kind of tool is used by stakeholders to ease the centralization and the intake of a large amount of information and to keep abreast with current market trends. Some systems use domain knowledge to improve the recommendation task [2], [1]. Content based recommender systems typically follow a two-step process: (i) the indexing of articles (also known as *classification*) and users also known as *profiling* (e.g. fig. 1 red square), and (ii) the comparison process which consists in comparing classification of articles and profiles of users. Our tool is specialized in the production and the distribution of press reviews about French regional economic activities (e.g. fig. 1).

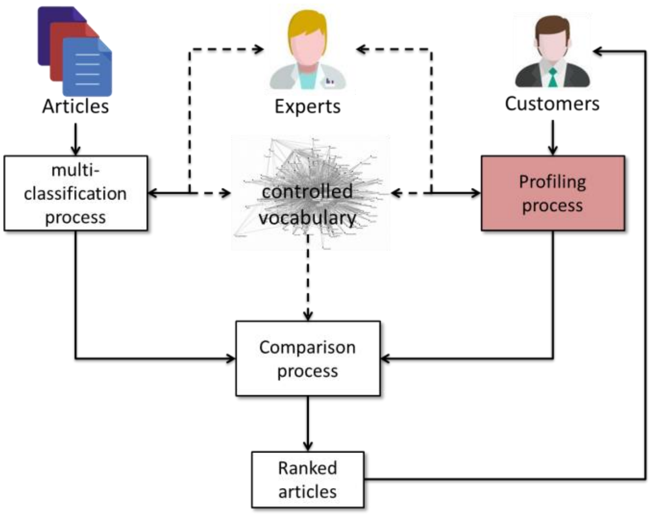

Figure 1. Our ontology-based recommender system

The paper focuses on automatic refinement of customers' profile which is an important task over time. This process takes into account the logs of the users concerning especially his/her actions, reading time, and domain specific knowledge. In our approach, ontologies are used to describe and automatically refine these profiles. Our recommender system takes advantage of the development of the semantic web. It brings new solutions to model the needs of the users by using a central controlled vocabulary using an ontology. The user profile represents the needs of the user, and the concept of profile refinement is used to describe the existing techniques which details this profile. Currently, the requirement of the company Actualis Sarl funder of this project is to move toward an automatic profile refinement in order to keep the profile as close as possible to the needs in economical information of the users. In the early content-based systems, the profile is described as a set of keywords. This technique has several flaws, the main ones being the synonymic and polysemic issues. The semantic web technologies brought new solutions for the representation of items and profiles, which are analyzed in section 2. Once items and profiles are described, we can try to improve the precision of the recommendations made. The first step consists in analyzing the user behavior within the application, whether they are explicit, i.e. feedbacks, or implicit, i.e. clicks, scrolling, etc. We analyze these methods in section 3. Section 4 presents the user action footprints for the profile refinement process created from sensors implemented in our solution. The last section concludes this paper.

2. Semantics in recommender systems

In this section we briefly describe several systems that consume knowledge which are lexical data and domain specific knowledge.

Lexical knowledge: The systems [7] and [10] use Wordnet and Multi-Wordnet which are the first tools that allowed the development of more complex recommenders. Wordnet gathers lexical information by grouping words as sets of synonyms called synsets [7]. SiteF[14] is one of the first news recommender that represents the information by their meaning. ITR[5] is a cultural products recommender, that uses Wordnet. The system represents the resources as bags of synsets. SEWeP[6] is a web request personalization system that uses Wordnet and a taxonomy of web pages categories. More complex systems, which combine these data with domain-specific, as well as encyclopedic ontologies are more efficient, when Wordnet is not sufficient to describe the interests of the users.

Domain specific knowledge: Informed Recommender [17] makes products recommendations based on a textual user feedback. An ontology is used to evaluate the quality of the feedback and the quality of the products. The system does not construct a user profile but it answers specific quality questions about the products. News hand[18] uses a set of ontologies based on the IPTC ontology (Ontology that deals with various themes like sport, education, politics). The profile and the resources are represented as a TF-IDF vector of the concepts of the ontology. Quickstep[19] is a scientific articles recommender that uses an ontology based on the ODP ontology. Concepts are associated to the several articles and the system retains the most relevant for the user.
The use of concepts that the user can understand makes possible the visualization of the profile. This point not only enhances the transparency of the system, but can also be used to get a user feedback on his own profile. In addition, general purpose ontologies such as Wikipedia, the ODP and Yahoo Web Directory have grown during the last decade. In our recommender system, the knowledge is domain dependent. General purpose ontologies are useful only to model lexical knowledge.

3. User Feedback

The aim of profile refinement is to enhance the relevance of the recommendations. If the profile is not built by an expert/vendor, it must be created automatically. The feedback provided by the interface of the application is the only source of information available in this case: explicit and implicit methods are distinguished. Several actions available to gather information about the user are categorized in [13].
Explicit methods: consist in asking information directly to the user in the interface. These methods are most of the time reliable, with few or no treatments needed. The establishment of an initial profile is made based on this information. An alternative is to create the profile manually which is actually done in the company by expert/vendors who analyze the needs of the customers during a short phone call.
Implicit methods: consist in the study of behavior of the user. As shown in figure 1, there are several actions available to evaluate the interest of the user to a resource. [19] studies the behavior of 75 students who are asked to surf on the web for 30 minutes, to analyze their clicks, reading time and scrolling of a page. After each page, the students have to rate it. The study shows that reading time combined with scrolling are correlated to the ratings. However, it doesn’t seem to have a correlation with the clicks, or scrolling alone. [16] studies reading times, saving and copies of news articles read by 8 people for 6 weeks. After reading an article the users have to rate it. No concrete correlation is found, but reading time seems to be higher for relevant documents. Also, the authors try to determine a threshold from which documents are considered as relevant.
Discussion: Most of the articles mentioned previously try to find correlations based on reading time and eventually scrolling. [13] states that there is a lack of research literature about other methods. Reading time is generally correlated to the relevance of the resources, but this correlation is highly dependent on the reading condition (test protocols), and more importantly on the task complexity [11][12]. Assuming it is possible to qualify these two characteristics in an application, it would be possible to determine a reading time threshold from which a document can be relevant. The use of scrolling, eyetracking and implicit methods in general are not sufficient, but these methods seem to improve the precision of the recommendations in conjunction with ratings[19][8][9]. [13] concludes by saying that if other actions are taken into account, they must be weighted, as saving or printing a document is probably more important than copying a part of it for example.

4. User action footprints for the profile refinement process

To capture the user behavior, the web application containing the economic articles includes some sensors. These sensors are of three types. The first one permits to capture the importance of an article due to the login of the user. When the user redirects an the information, his/her logs are updated and taken into account. The fig. 2 shows how the user can print or send by email a current piece of news. The second sensor is the time spend to read an article (e.g. fig. 3). In order to read the complete article, the user has to click on a button with start a timer. The click on the button is itself an important information. The last sensor is a news filter system that allows the regeneration of the web page regarding the most important information for the user (e.g. fig 4). In the example, the option boxes “BTP” and “Industrie manufacturière” are selected. These keywords are linked to the ontology and are the same used to define the profiles of the users and to index the news. The navigation behavior is stored in our recommender system to qualify the profile refinement, using the three kinds of footprints: actions, reading time, and domain specific knowledge. The profile refinement computation uses a query expansion principle already used in our comparison process [21].

Figure 2. Snapshot of a user action capture

Figure 3. Snapshot of a user reading-time capture

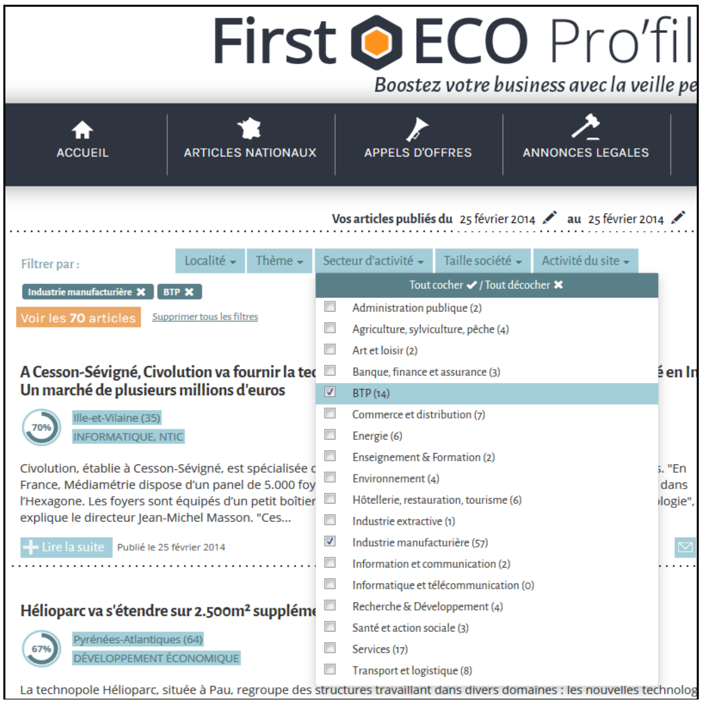

Figure 4. Snapshot of a user domain-specific knowledge capture

5. Conclusion and future work

Content-based recommender systems have been greatly impacted by the development of ontologies and semantic web in general. Research has yet to be done in the detection and analysis of the user behavior. It seems that technology slows advances for implicit methods, because of the difficulty to evaluate them and because they are highly dependent on the application. As for explicit methods, although they are easy to evaluate, they are really intrusive. An alternative would be to ask the user for a textual feedback that could be analyzed. We present three kinds of user interaction that are captured in the logs and taken into account in our system: actions of the user, his reading time, and domain specific knowledge. Due to a lack of space the refinement process is not detailed.

Acknowledgment
This project is founded by the company Actualis SARL and the French agency ANRT.

References

[1] S. E. Middleton, N. R. Shadbolt, and D. C. De Roure, ”Ontological user profiling in recommender systems”, Acm Trans Inf Syst, vol. 22, no 1, p. 54–88, janv. 2004.
[2] W. IJntema, F. Goossen, F. Frasincar, and F. Hogenboom, “Ontology-based news recommendation”, in Proceedings of the 2010 EDBT/ICDT Workshops, New York, NY, USA, 2010, p. 16:1–16:6.
[3] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, “GroupLens: an open architecture for collaborative filtering of netnews”, in Proceedings of the 1994 ACM conference on Computer supported cooperative work, NY, USA, 1994, p. 175–186.
[4] D. Billsus and M. J. Pazzani, “A personal news agent that talks, learns and explains”, in Proceedings of the third annual conference on Autonomous Agents, USA, 1999, p. 268–275.
[5] M. Degemmis, P. Lops, and G. Semeraro. “A content-collaborative recommender that exploits wordnet-based user profiles for neighborhood formation.”, User Modeling and User-Adapted Interaction, 17(3):217–255, 2007.

[6] M. Eirinaki, M. Vazirgiannis, and I. Varlamis. “Sewep: Using site semantics and a taxonomy to enhance the web personalization process.”, In ACM SIGKDD, KDD ’03, pages 99–108, New York, NY, USA, 2003. ACM.
[7] S. Gauch, M. Speretta, A, Chandramouli, and A. Micarelli, “User profiles for personalized information access.”, In The Adaptive Web, volume 4321 of Lecture Notes in Computer Science, pages 54–89. Springer Berlin Heidelberg, 2007.
[8] G. Jawaheer, M. Szomszor, and P. Kostkova. “Comparison of implicit and explicit feedback from an online music recommendation service.”, In Proceedings of the 1st International Workshop on Information Heterogeneity and Fusion in Recommender Systems, HetRec ’10, pages 47–51, New York, NY, USA, 2010. ACM.
[9] T. Joachims, L. Granka, B. Pan, H. Hembrooke, F. Radlinski, and G. Gay. “Evaluating the accuracy of implicit feedback from clicks and query reformulations in web search.”, ACM Trans. Inf. Syst., 25(2), April 2007.
[10] P. B. Kantor, F. Ricci, L. Rokach, and B Shapira. “Recommender systems handbook”, Springer, 2011.
[11] M. Kellar, C. Watters, J. Duffy, and M. Shepherd. “Effect of task on time spent reading as an implicit measure of interest.”, Proceedings of the American Society for Information Science and Technology, 41(1):168–175, 2004.
[12] D. Kelly and N. J. Belkin. “Reading time, scrolling and interaction: Exploring implicit sources of user preferences for relevance feedback.”, In Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR ’01, pages 408–409, New York, NY, USA, 2001. ACM.
[13] D. Kelly and J. Teevan. “Implicit feedback for inferring user preference: A bibliography.”, SIGIR Forum, 37(2):18–28, September 2003.
[14] H. R. Kim and P. K. Chan. “Learning implicit user interest hierarchy for context in personalization.”, In Proceedings of the 8th International Conference on Intelligent User Interfaces, IUI ’03, pages 101–108, New York, NY, USA, 2003. ACM.
[15] Y Koren. “Collaborative filtering with temporal dynamics.”, Commun. ACM, 53(4):89–97, April 2010.
[16] B. Magnini and C. Strapparava. “Improving user modelling with content-based techniques.”, In Mathias Bauer, PiotrJ. Gmytrasiewicz, and Julita Vassileva, editors, User Modeling 2001, volume 2109 of Lecture Notes in Computer Science, pages 74–83. Springer Berlin Heidelberg, 2001.
[17] M. Morita and Y. Shinoda. “Information filtering based on user behavior analysis and best match text retrieval.”, SIGIR ’94, pages 272–281, New York, NY, USA, 1994. Springer-Verlag New York, Inc.
[18] S. Aciar, D. Zhang, S. Simoff, and J. Debenham. “Informed recommender: Basing recommendations on consumer product reviews.”, Intelligent Systems, IEEE, 22(3):39–47, 2007.
[19] I. Cantador, A. Bellogan, and P. Castells. “News hand: A semantic web approach to recommending news.”, In Wolfgang Nejdl, Judy Kay, Pearl Pu, and Eelco Herder, editors, Adaptive Hypermedia and Adaptive Web-Based Systems, volume 5149 of Lecture Notes in Computer Science, pages 279–283. Springer Berlin Heidelberg, 2008.
[20] M. Claypool, P. Le, M. Wased, and D. Brown. “Implicit interest indicators.”, In Proceedings of the 6th International Conference on Intelligent User Interfaces, IUI ’01, pages 33–40, New York, NY, USA, 2001. ACM.
[21] D. Werner, and C. Cruz, “Precision Difference Management using a Common Sub-vector to Extend the Extended VSM Method”, ICCS 2013, Procedia Computer Science, Elsevier, Volume 18, 2013, Pages 1179–118.