Intelligent browsing assistance for corporate knowledge portals*

Markus Zanker and **Sergiu Gordea**
University Klagenfurt, Austria
{markus,sergiu}@ifit.uni-klu.ac.at

Marius Silaghi
Florida Institute of Technology
marius.silaghi@fit.edu

Abstract

The management of knowledge, i.e. *knowing what is known and the ability to exploit it* is a burning issue for most organizations. Though knowledge management has a strong social perspective, information technology supports knowledge management strategies by providing tools that store codified knowledge and allow its retrieval. The value of knowledge management derives from the wide use of the stored knowledge, its annotation and refinement as well as its application in business practices. We support this issue of information browsing and retrieval by adapting and applying recommendation technology to reduce access barriers towards corporate knowledge portals. For this purpose we exploit domain heuristics to improve collaborative filtering techniques in order to cope with the problem of low numbers of users the knowledge portals have.

Introduction

Personalized recommendations are a powerful vehicle to reduce information overload and to help users to better orient themselves within Web-based information systems. There exist several different technological approaches for recommendation systems. Due to the commercial success of e-commerce sites like *amazon.com* the most well-known technique is collaborative or social filtering and variants thereof. It exploits similarities in the past behaviour of users and recommends those items to a user that his statistically nearest-neighbors liked. User behavior includes all actions that can be interpreted as a statement of preference or as a differentiating characteristic of a user, ranging from the explicit rating of items to simple clicks or views.

Literature documents a variety of application areas for recommender systems such as e-commerce, Webpages or even software development (Tsunoda *et al.* 2005). But so far there is only few published work on the application of recommendation technology on corporate knowledge portals and intranet environments. One of the reasons for that is the low performance of the algorithms due to the relatively small number of users. E-commerce environments typically address large communities of users, the same holds for websites that offer information such as news (Konstan *et al.* 1997). The user community of intranet environments and corporate knowledge portals is typically one size of magnitude smaller, therefore it is harder for recommendation algorithms to identify a properly sized group of users with similar preferences to make good predictions. (Middleton, Shadbolt, & de Roure 2004) describe the Foxtrot and Quikstep systems, that capture user preferences in research topics and help their users in quicker accessing interesting scientific literature documents. They employ domain knowledge in form of ontologies to learn more robust models of the user. Items that conform to the interest representation of the user are then recommended. Such a model-based approach is appropriate for domains, where users can be associated with a static representation (though this might be dynamically changed over time) of their interests and preferences. However, when interacting with a knowledge management portal we claim that a model-based approach suffers from a lack of serendipity, e.g. developers that only lookup java guides might not be proposed an XSLT-resource by Michael Kay, although this could be of interest in their current situation. Therefore, we propose the use of collaborative filtering techniques and domain heuristics, such as focusing on newer items or considering the context of the current session, to deliver reasonable recommendations to users of knowledge portals.

Recommendation technology

Adding recommendation services to knowledge portals intends to reduce access barriers for users and provide quicker retrieval opportunities next to standard browsing through folder categories. As already mentioned in the introduction, the application of collaborative filtering techniques to domains with few users (e.g. user numbers of knowledge portals typically range from 20 up to several hundreds of users) is not straightforward, as classic algorithms would suffer from sparse user data. Therefore, we first discuss characteristics of the domain that might allow the introduction of heuristic methods into the employed algorithms.

Characteristic 1 - continuous change of preferences: User preferences change over time. The music domain is an il-

*This work is carried out with financial support from the EU, the Austrian Federal Government and the State of Carinthia in the Interreg IIIA project Software Cluster South Tyrol - Carinthia.

lustrative example. Different trends (folk, rock, pop, house etc.) dominate the market for some period of time, but also special events influence users interest at short notice. This motivates us to introduce time frames to determine user preferences.

Characteristic 2 - different user groups: Users of knowledge portals can be differentiated according to their status and user type on the meta-level, e.g. system administrators, higher-level management staff and regular employees. When analyzing logs of different knowledge management systems we identified different user groups that distinguish themselves by a fundamentally different behaviour, e.g. administrators mainly add knowledge items to the system while regular employees are basically consumers.

Characteristic 3 - user intention: Users have a specific intention that might change between each session with the knowledge portal, that overrule their more stable general interests and preferences, i.e. within the context of a specific session or visit users are interested in a particular type of information. Therefore, we have to consider the current context of a user making a recommendation. We use for example the browsing behaviour into different folders as an indicator for his intentions and can therefore make context-dependent recommendations to him.

Characteristic 4 - novelty: Users of knowledge portals have a higher interest in new and currently added items. This demonstrates itself in the daily access rates of knowledge objects. The access typically peaks shortly after the object's creation date. In our experimental evaluation we observed that half of all accesses happened within the first month after creation.

We accommodate this domain characteristic weighting the degree of novelty into the *ranking function*:

$score_{ik} = h(t) * \frac{1}{N} * \sum_{j=1}^{N} sim_{ij} * r_{jk}$,
$h(t) = \frac{1}{1+\frac{1}{DIN_i}*t}$, $DIN_i = \frac{acc_i}{acct_i}$,

where r_{jkt} is the the user's j rating for item k, acc_i represents the number of distinct items accessed by user i, and $acct_i$ is the total number of item accesses of user i. The item's age is used in an attenuation function h(t) that reduces the weight of older items in the scoring function.

Collaborative filtering algorithms depend on the availability of item ratings, used to identify user's preferences. Therefore we use normalized *implicit ratings* computed as: $rate_{ik} = 1 - \alpha e^{-acc_{ik}}$, where $rate_{ik}$ is the rating of user i for item k, acc_{ik} is the number of times user i accessed item k, and α is the coefficient used to adjust the function's attenuation.

Evaluation

For evaluation we employ system logs of a productive knowledge management system, collected over a period of 10 months. The system contains 206 knowledge objects placed in 115 different folders (i.e. categories). The logs from the first 6 months were used to compute the user to user similarities, using Pearson correlation, while the log data from the last 4 months was used to evaluate the effectiveness of the recommendation algorithms. We evaluated the influence of introducing time and context based constraints in the recommendation process by analysing the effectiveness of 5 different recommender algorithms: top*n*, top*n*cf, *cftime*, *cfcontext* and *cfcontexttime*. The top*n* algorithm recommends the best rated *n* items, where *n* is the number of items accessed by the active user in the period of time considered for evaluation. Top*n*cf algorithm recommends the best ranked *n* items, rated by members of user neighborhood. The last three collaborative filtering (CF) algorithms use additional time and context constraints to reduce the number of elements in the list of recommended items without eliminating any of the candidate items. From the total of 57 users, we were able to identify 17(30%) with a consistent activity and user neighborhood for which we computed the recommendations. The evaluation of the algorithms was done with the help of precision, recall and F1 metrics (Herlocker *et al.* 2004). The top n algorithms have a poor performance showing that they are not designed to work under such restrictive conditions. The *cfcontext* presents a higher precision with the expense of a low recall, while the *cftimecontext* algorithm slightly improves the performance of *cftime* algorithm presenting the best performance with respect with all metrics (with an average of: F1=0.32, Precision=0.46, Recall=0.32), being also the algorithm with the lowest variance of the metrics computed over all predictions.

Conclusions

We presented our approach to improve the usability of intranet portals by using recommendation technology. We use a set of heuristics with respect to the use of knowledge portals in order to improve the effectiveness of the recommendation algorithms. Five different algorithm approaches were taken into account to evaluate the benefit of using time and context based heuristics. We can conclude that both time and context improved the effectiveness of the algorithms with respect to Precision, Recall and F1 metrics.
We introduced an enhanced scoring and rating functions that produced an advanced ranking of recommended items and overcomes problems of few users.

References

Herlocker, J.; Konstan, J.; Terveen, L.; and Riedl, J. 2004. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems* 22(1):5–53.

Konstan, J.; Miller, B.; Maltz, D.; Herlocker, J.; Gordon, L.; and Riedl, J. 1997. Grouplens: Applying collaborative filtering to usenet news. *Communications of the ACM* 3(40):77–87.

Middleton, S. E.; Shadbolt, N. R.; and de Roure, D. C. 2004. Ontological user profiling in recommender systems. *ACM Transactions on Information Systems* 22(1):54–88.

Tsunoda, M.; Kakimoto, T.; Ohsugi, N.; Monden, A.; and Matsumoto, K. 2005. Javawock: A Java Class Recommender System Based on Collaborative Filtering. In 17^{th} *International conference on Software Engineering and Knowledge Engineering (SEKE'05)*. ACM.