

A PERSONALIZED FORUM ENVIRONMENT

Anna Stavrianou

Laboratoire ERIC, Université Lumière Lyon 2, Université de Lyon, Lyon, France

Magdalini Eirinaki

Computer Engineering Department, San Jose State University, San Jose, CA, U.S.A.

Keywords: Forum, Recommender systems, Social networks, Web personalization.

Abstract: Web2.0 has resulted in an increasing popularity of personalized recommender systems, especially in the context of social networking applications. Although there exist design approaches available for such systems, most of them make very explicit assumptions on the application domain as well as on the availability and data types to be used as input. In this position paper, we discuss the requirements and challenges of Forum Recommender Systems. Such systems aim at generating automatically posting recommendations for the different user profiles that deal with a forum. Despite the fact that these systems share characteristics with other social media, they have hardly been explored due to the particularities they present in terms of structure, context and user differences. Here, we discuss the particularities of Forum Recommender Systems and we propose a framework that enables the gathering of profile data and the generation of posting recommendations. The proposed framework can also be adjusted to other social networks.

1 INTRODUCTION

The blogosphere, the forums, the web newsgroups, the social network sites aggregate masses of user-generated and personalized data. Nowadays, data such as social network relationships (e.g., friendship) and respective ratings/opinions are employed to recommend items (Guy et al., 2009; Konstas et al., 2009; Massa and Avesani, 2007) or users (Kunegis et al., 2009; Leskovec et al., 2010; Varlamis et al., 2010; Weng et al., 2010). The recommended items may be news stories, blog posts, or communities and the recommended users may be bloggers, or likely-to-be-friends, depending on the context.

A big challenge of nowadays is the design and the evaluation of personalized systems and recommender applications in the context of social networking media. Although many approaches exist, most of them make specific assumptions on the application domain as well as on the availability and type of data to be used as input. However, questions such as, “How can we design a system so as to enable the gathering of profile data?”, “What characteristics do we have to take into account?”, “How can we evaluate such a system?”, have not been answered in a systematic way.

In this position paper, we propose a generic framework that addresses the aforementioned challenges, focusing on Forum Recommender Systems. Although, nowadays, the forum Web sites provide social networking functionality such as, who is fan of whom, whether a user is popular, etc., this social network (SN) information, although available, is hardly used for purposes other than statistical or purely informative. We focus on formulating the framework that forums should be based on and the information that should be gathered from forum web sites in order to facilitate the implementation of a Forum Recommender System.

We begin by defining the characteristics of a Forum Recommender System, then, in Section 3, we discuss the requirements based on the different user types and tasks. Section 4 presents profiling users and items, while in Section 5 challenges and open issues are discussed. Related work is in Section 6, and the conclusion in Section 7. We should note that since forums share common characteristics with other social media, such as, user connectivity, shared authorship, tagging, and reviewing/commenting, the proposed framework can be easily adjusted to other social networks as well.

2 FORUM RECOMMENDER SYSTEMS

Our objective is to define a framework that enables the generation of recommendations in forums. For this purpose we have to consider a set of design features which implicitly generate data that can be mined in order to create user profiles and/or input in the recommendation process. These are characteristics that are common to any social networking application whose actors author texts. Here, we focus on forums.

A forum can be represented by a graph whose nodes show the users, and the edges may represent various types of relationships. In the simple case, an undirected edge between two nodes implies that two users have posted on the same thread. Expanding to a directed graph introduces follow-up discussions or enriches the graph with social relationships, such as sharing of information. By applying social network analysis algorithms, we can mine interesting information regarding the position and importance of each user in the graph (e.g. by using degree/betweenness centrality, or variations of PageRank).

Authorship is a major component of forums. Users author postings and may also comment or augment other users' postings too. These postings which are on average much lengthier than any other social network's counterpart, are rich in terms of content, context, they may contain different subtopics, and also express specific sentiments/opinions.

Forums keep evolving and changing with time. Thus, postings that may be popular at a certain time period, may be displaced by other postings some minutes afterwards. Similarly, the location of the user in the network, her role or the degree of interaction of users may change at any time.

Based on these characteristics, we define a "Forum Recommender System" (FRS) based on the recommendation problem (Adomavicius and Tuzhilin, 2005):

Definition 1. *Let us consider a forum F having a set of postings P . A Forum Recommender System is a system that recommends a subset $P' \subset P$ of postings to a user usr such that the utility u of the user is the maximum:*

$$P'_{usr} = \operatorname{argmax}_{P' \subset P} u(P', usr) \quad (1)$$

The utility function represents the satisfaction of the user regarding how interesting a posting is.

The definition of an FRS points out the presence of different user roles and tasks (i.e. why they need

a recommendation) in a forum. Different users have different requirements and needs from a system that recommends postings.

3 FRS REQUIREMENTS

In this section we discuss the requirements of an FRS system per type of user.

3.1 End User

An end user is a user that browses and navigates the forum. She desires to understand the gist of what has been discussed and participate as well. Due to the abundance of information, it is evident that the system needs to generate recommendations of postings, in order to facilitate navigation.

A FRS accessed by an end-user should be able to comply with the following requirements:

Recommended items:

- The system should recommend a number of forum postings to a user.
- The number of the recommended postings should not exceed a certain threshold which could be either user-defined or user-friendly.

Access/Usability of recommendations:

- The recommendations should ideally be accompanied by the location of the posting in a forum. If the user is interested in the recommended posting she should be able to follow the thread, the replies to this posting or to what this posting replies to.
- The system should allow the user to browse the results quickly (e.g. show the beginning of the text content of each recommendation).
- The user should be able to access the recommended postings quickly (e.g. by clicking on them).
- The system should visualize the recommendations in a user-intuitive way.
- The system could recommend postings from a list of topic-related forums. Thus, the user would be encouraged to visit different forums on the same topic whose existence may have ignored.

These requirements apply also to the other user roles and tasks, since efficient browsing of forums is a requirement for any type of user.

3.2 Forum Administrator

A forum administrator is responsible for the content of the forum in terms of guiding and supervising its

flow. She wants to know how the discussion has evolved, which postings have caused a lot of reactions, in which parts of the discussion people argue, etc. In order to assist this user, the FRS should make available the selection of content-based criteria, as outlined in the following requirements:

- The system should recommend postings that contain certain (user-provided) keywords.
- The system should recommend postings that belong to specific topics.
- The system should recommend “controversial” postings, where people seem to have disputed over a point or a subject.
- The system should recommend postings that are considered “interesting”, “uninteresting”, etc. In this way, the rejection of hostile, insulting or spam postings becomes more efficient.

3.3 Analyst

Forum analysts are also interested in the content of the forum. The challenges such users face involve locating the relevant information, and subsequently accessing and analyzing the useful portions related to it. It is evident that for the analyst the most important aspect of forum discussions is the opinions of the forum participants. Thus, in addition to the aforementioned end-user requirements, this user category’s requirements include the following:

- The system should recommend postings which contain “positive” or “negative” comments for a certain (user-defined) product/subject.
- The system should recommend sets of postings where participants seem to agree or disagree over a specific (user-defined) subject.
- The system should recommend postings authored by forum “influencers” or experts.

4 CONSTRUCTING USER AND ITEM PROFILES

In this Section we focus on the construction of user and item profiles that characterize an FRS.

4.1 Profiling Non-participants

Users who have not yet participated in a forum are more challenging to profile, since we have little, or no explicit knowledge about them.

4.1.1 Implicit Preferences

An FRS should be designed in such a way that allows the gathering of implicit data. For instance, it may facilitate the tracking of the postings that the user clicks on, assuming that these are postings the user selects to read (Stavrianou, 2010).

Posting Content. By analyzing the content of the read posts, the FRS can infer the information the user is interested in. As a result, the recommendation list can be updated with postings similar in content or topic to the clicked ones.

Posting Navigation Depth. Navigating on the same thread by clicking on various posts, increases the degree of certainty about the user’s interest on the specific content/topic. This is in turn reflected in the recommendation process by re-ranking the results placing the most relevant ones higher on the list.

Duration. An FRS could also monitor the time the user spends on a posting, by logging the time difference between two clicks. Of course, the time lap can lead us to erroneous conclusions, since a user may not necessarily be dealing with the clicked posting before he clicks somewhere else. For this reason, the time lap should be used cautiously and always in combination with other behavioral user attitudes.

Uninterest. An FRS can even log the postings not chosen by the user. This information may reveal the topics or style of postings the user is not interested in so as to avoid their recommendation.

4.1.2 Explicit Preferences

Apart from the implicit ways to monitor a user, we can also use explicit ones in order to build her profile.

Profile Preferences. The FRS could allow the user to set certain criteria. The profile characteristics could be topic-related. For instance the user may select which thematic areas are of interest for her (e.g. World Economy and Technology) and/or which are not. The user may be able to update this profile. The profile can be also implicitly updated by the postings the user actually reads.

Moreover, depending on the user’s role and task, a user could choose between spam postings, postings with negative comments, etc. These preferences could be logged in the profile of the user together with their frequencies (i.e. how often they are demanded), or could be stored permanently and only changed when the user chooses to do so.

Posting Feedback/Ratings. The user can be given the possibility to tag the postings she reads as to whether they are interesting or not. In order to achieve this, we need to include a ranking system in the web site of the forum. In this case, the system should clarify to the user the meaning of tagging a news story or a product with a “thumbs up” or a “thumbs down” (or a more extensive rating, e.g. in the scale of 1-5 stars).

Recommendation Feedback/Ratings. Finally, once the system starts generating posting recommendations, a user may provide feedback on whether these recommendations are useful or not. This functionality is similar to the one of posting feedback and requires a similar mechanism to gather information.

4.2 Profiling Participants

A user who is already an author in the forum, aggregates in her profile the information of a non-participant. Additional information can also be gathered resulting from the user’s own *activity*.

Authored Content. The postings which are authored by the user give an indication of the topics that the user is interested in. Similarly, the content of the postings to which the user has replied (or not replied), can show the interests (or non-interests) of the user in the particular forum.

Social Network Relationships. Regarding the reply-activity, the behavior of an author can also be tracked in the context of the SN. If, for example, a participant tends to always reply to the same person, the recommendations can be updated to include postings sent by the particular person. These postings may be located in another thread of the forum not yet visited by the specific user.

Expertise/Influence. Identification of the user’s expertise or influence (Agarwal and Liu, 2008; Estévez et al., 2007; Kale et al., 2007; Zhang et al., 2007), could re-arrange the recommendations so that postings adjusted to the role of the participant are given priority to in the recommendation list. An expert, for example, maybe be more eager to answer to questions and a non-expert will be more interested to receive answers written by experts.

4.3 Profiling Postings

The item profiles of a FRS refer to the profiles of the different postings. These profiles can mainly contain semantic information such as keywords and opinion polarity information.

Keywords. The keywords extracted from a posting give information about its topic. The challenge in the case of forum postings is that the text can be so small that Text Mining techniques based on keywords (Mooney and Bunescu, 2006) cannot always work efficiently. Additionally, there is often a forum-specific vocabulary used.

Opinion Polarity. The opinion polarity of a posting could automatically be retrieved by Opinion Mining methods (Ding and Liu, 2007; Ghose et al., 2007; Hu and Liu, 2004; Turney and Littman, 2003). Knowing the opinion polarity enables their recommendation to users who desire to have an opinion-oriented overview, positive or negative, of what has been said about a product, an idea or a topic.

Timestamp Information. The “age” of a posting and its temporal distance between other postings in the same or other forums, together with the opinion mining can give an indication of the opinion flow or exchange between forum postings.

5 CHALLENGES, OPEN ISSUES

It is evident from the aforementioned analysis that the design and implementation of a Forum Recommender System raises many challenges. We discuss the most important of them in this section.

User Profiling. Although a user profile with some basic user characteristics (e.g. age, gender) can easily be kept into a database, a profile that contains information regarding what makes a message interesting for a user or not, is not easily catered for. Profiling information needs to be constantly updated, since the interests of a user change, as well as her expertise. Being interested in a topic today does not necessarily mean that a user desires to receive recommendations for the same topic all the time. In addition, a user that is considered expert for answering certain questions in a forum can be “outsmarted” by another user who joined recently the SN.

Influencers. Influential users play a key role in spreading information. A simple way to measure influence is by using the structure of the SN and apply measures such as, centrality or prestige (PageRank). Lately, other activity/profile parameters (e.g. number and frequency of posts, comments, etc.) have been studied in the context of blogs (Agarwal et al., 2008) and social networks (Kim and Han, 2009). Measuring influence in a forum presents a unique challenge; influence is strongly related to trust: how much do we trust the opinion of a user in a specific topic. Trust is

context specific and this discrimination is important in the FRS context: we might trust a user's opinion on world politics, but not on financial matters. Thus, the "influence" of a user is context-related. Another challenge that needs to be addressed is defining each user's "circle of trust" among the users of a forum.

Interest of a Posting. Defining the interest of a message per user or user-community is not easy. Is it the content or the author that makes a post interesting? What makes a posting more interesting than another? From the content point of view, the opinion presence, the way the arguments are presented as well as the type of arguments may make it interesting. On the other hand a post authored by an expert or an influencer has an increased weight of interest independent of the actual content. Modeling and measuring the interest of a posting has not yet been dealt with.

Opinion Mining. The postings of a forum may contain opinions about products, ideas, proposals, social and economic changes. The presence of opinion in forums can be used as a criterion for recommending interesting messages. Until now the Opinion Mining techniques deal with the identification of the polarity and its strength, but they do not, yet, consider the opinion flow between two posts or the opinion exchange between two people. The opinion-based information flow has an impact on the way the opinion changes and on the evolution of the discourse.

User Similarities. The most common technique used in recommender systems is collaborative filtering. The same algorithms can be applied in the context of FRS. The challenge, in this case, is how to model users and the similarity among them. A naive way would be to assume two users to be similar if they comment on the same posts. However, this approach excludes the notion of opinion, discussed previously. Two users are similar if they have the same opinion on the same (or similar) topics. Thus there are two parameters that need to be considered in the FRS context: content as well as opinion polarity.

Visualization. One design issue is how to present the recommendation list to the users. Navigating quickly to the chosen recommendations that correspond to the user criteria or interests becomes an issue, especially when the posts are long and the related forum contains hundreds of postings. Visualization techniques need to be studied in order to represent a forum together with the posts-related recommendation list, allowing at the same time the user to browse efficiently the posts given the recommendations.

Evaluation. Evaluating recommender systems is not easy (Herlocker et al., 2004). In the case of FRS,

the evaluation is an important issue since each forum has a different distribution of users and postings, different content and style of language used. Implicitly, we could monitor the behavior of users towards the recommendation list. If they actually click on the proposed postings, this could be an initial indication that they find them interesting, depending on how much time they do spend on them, or whether they actually follow the specific discussion thread. Otherwise, we could explicitly ask users to rate the postings. Evaluation techniques need to be studied carefully, since no benchmark exists for the time being.

6 RELATED WORK

A personalized forum environment may use methodologies from various domains. In (Kunegis et al., 2009; Leskovec et al., 2010), several algorithms are proposed for the recommendation problem, based on content similarity, social link information, and common items among users. The proposed models are only applicable to social networking applications and not other social media. In the case of blogs, ranking algorithms have been suggested that exploit explicit (Nakajima et al., 2005) and/or implicit (Kritikopoulos et al., 2006; Adar et al., 2004) hyperlinks between blogs. A similar effort is presented in (Weng et al., 2010), while a more generic model has been presented in (Varlamis et al., 2010).

Identifying influencers in a SN is often modeled as a combinatorial optimization problem: given a fixed number of nodes find the ones with maximum influence over the network (Domingos and Richardson, 2001). The proposed approaches (Estévez et al., 2007; Kimura et al., 2008; Kempe et al., 2003) are based on the link structure, and do not consider parameters, such as activity, rate of updates, and trust among users. Link analysis techniques (Song et al., 2007) and activity-related parameters have been used in order to identify influencers in blogs (Agarwal et al., 2008) and social networks (Kim and Han, 2009). Recently, trust has also been introduced in the context of recommender systems (Golbeck, 2006; Golbeck, 2005; Massa and Avesani, 2007), and trust propagation has been studied in the case of virtual communities (Guha et al., 2004; O'Donovan, 2009; Ziegler, 2009).

The need of the use of opinion mining techniques is evident due to the opinion that resides inside review sites, blogs and forums. The majority of approaches, such as (Hu and Liu, 2004; Ding and Liu, 2007), use a seed list of adjectives and they attempt to identify the relation between the words in a text and those of

the seed list. An original approach is proposed in (Ghose et al., 2007) where the opinion is inferred by observing the effect of user comments found in a reputation system on the prices of the products sold.

7 CONCLUSIONS

Forum Recommender Systems may use knowledge and techniques from various research fields such as the generation of recommendations in social networks, the presence of influence, the trust propagation. Although much work has been done in identifying and incorporating these notions in other types of social media, there does not exist an in-depth study of how they can be incorporated in the context of forums. This paper provides the researchers with a generic framework and outlines the main challenges and open areas that still need to be explored.

REFERENCES

- Adar, E., Zhang, L., Adamic, L., and Lukose, R. (2004). Implicit structure and the dynamics of blogspace. In *Workshop on the Blogging Ecosystem, WWW*.
- Adomavicius, G. and Tuzhilin, A. (2005). Towards the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6):734–749.
- Agarwal, N. and Liu, H. (2008). Blogosphere: research issues, tools, and applications. *SIGKDD Explor. Newsl.*, 10(1):18–31.
- Agarwal, N., Liu, H., Tang, L., and Yu, P. S. (2008). Identifying the influential bloggers in a community. In *WSDM*.
- Ding, X. and Liu, B. (2007). The utility of linguistic rules in opinion mining. In *SIGIR-07*.
- Domingos, P. and Richardson, M. (2001). Mining the network value of customers. In *SIGKDD*, pages 57–66.
- Estévez, P. A., Vera, P. A., and Saito, K. (2007). Selecting the most influential nodes in social networks. In *IJCNN*.
- Ghose, A., Ipeirotis, P., and Sundararajan, A. (2007). Opinion mining using econometrics: a case study on reputation systems. In *ACL*.
- Golbeck, J. (2005). Computing and applying trust in web-based social networks. *PhD Dissertation*.
- Golbeck, J. (2006). Trust on the world wide web: a survey. *Found. Trends Web Sci.*, 1(2):131–197.
- Guha, R., Kumar, R., Raghavan, P., and Tomkins, A. (2004). Propagation of trust and distrust. In *WWW*.
- Guy, I., Zwerdling, N., Carmel, D., Ronen, I., Uziel, E., Yogev, S., and Ofek-Koifman, S. (2009). Personalized recommendation of social software items based on social relations. In *RecSys*, pages 53–60.
- Herlocker, J., Konstan, J. A., Terveen, L., and Riedl, J. (2004). Evaluating collaborative filtering recommender systems. *ACM TOIS*.
- Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. In *SIGKDD*, pages 168–177.
- Kale, A., Kolari, P., Java, A., Finin, T., and Joshi, A. (2007). Modeling trust and influence in the blogosphere using link polarity. In *ICWSM 2007*.
- Kempe, D., Kleinberg, J., and Tardos, E. (2003). Maximizing the spread of influence through a social network. In *SIGKDD*, pages 137–146.
- Kim, E. and Han, S. S. (2009). An analytical way to find influencers on social networks and validate their effects in disseminating social games. In *ASONAM 2009*.
- Kimura, M., Yamakawa, K., Saito, K., and Motoda, H. (2008). Community analysis of influential nodes for information diffusion on a social network. In *IJCNN*.
- Konstas, I., Stathopoulos, V., and Jose, J. M. (2009). On social networks and collaborative recommendation. In *SIGIR*, pages 195–202.
- Kritikopoulos, A., Sideri, M., and Varlamis, I. (2006). Blogrank: ranking blogs based on connectivity and similarity features. In *AAA-IDEA'06*. ACM.
- Kunegis, J., Lommatzsch, A., and Bauckhage, C. (2009). The slashdot zoo: mining a social network with negative edges. In *WWW*, pages 741–750.
- Leskovec, J., Huttenlocher, D. P., and Kleinberg, J. M. (2010). Predicting positive and negative links in online social networks. In *WWW*, pages 641–650.
- Massa, P. and Avesani, P. (2007). Trust-aware recommender systems. In *RecSys*, pages 17–24.
- Mooney, R. and Bunesco, R. (2006). Mining knowledge from text using information extraction. *SIGKDD Explor. Newsl.*, 7(1):3–10.
- Nakajima, S., Tatemura, J., Hino, Y., Hara, Y., and Tanaka, K. (2005). Discovering important bloggers based on analyzing blog threads. In *Workshop on the Blogging Ecosystem: Aggregation, Analysis and Dynamics*.
- O'Donovan, J. (2009). Capturing trust in social web applications. In Golbeck, J., editor, *Computing with Social Trust*, chapter 9. Springer London.
- Song, X., Chi, Y., Hino, K., and Tseng, B. (2007). Identifying opinion leaders in the blogosphere. In *CIKM*.
- Stavrianou, A. (2010). Modeling and mining of web discussions. *PhD Dissertation*.
- Turney, P. and Littman, M. (2003). Measuring praise and criticism: inference of semantic orientation from association. *ACM TOIS*, 21(4):315–146.
- Varlamis, I., Eirinaki, M., and Louta, M. (2010). A study on social network metrics and their application in trust networks. In *ASONAM 2010*.
- Weng, J., Lim, E.-P., Jiang, J., and He, Q. (2010). Twitter-rank: Finding topic-sensitive influential twitterers. In *WSDM*.
- Zhang, J., Ackerman, M., and Adamic, L. (2007). Expertise networks in online communities: Structure and algorithms. In *WWW*, pages 221–230.
- Ziegler, C.-N. (2009). On propagating interpersonal trust in social networks. In *Computing with Social Trust*. Springer.