

Detecting User Communities Based on Latent and Dynamic Interest on a News Portal

Marián HÖNSCH*

*Slovak University of Technology
Faculty of Informatics and Information Technologies
Ilkovičova 3, 842 16 Bratislava, Slovakia
marian.honsch@gmail.com*

One of the drawbacks of today's community-based collaborative recommender systems is that they group users based only on their aggregated similarity. Only those users are assigned to a community, whose profiles match completely the profile of the community. This prevents them from using the wisdom of crowd coming from users that match only parts of their interests. This drawback was also described in former research papers [1], [3]. We assume that recommendations coming from communities that address this issue can significantly improve the quality of recommendations. The demand for accurate recommender systems is very actual. Not only that users prefer personalized portals over non personalized ones, but internet companies can also raise their profits by having users spend more time on their web page thanks to the recommender and personalisation features.

Our target is to find communities that are defined by one particular interest. This community should include all users that share this interest. Depending on how many interests we have discovered in user's profile, she can belong to several communities.

We evaluated the whole approach on a news recommender system by recommending articles based on communities detected by our approach. We also performed an experiment to confirm our interest comparing strategy.

Interests are derived from the analysis of the domains corpus and are expressed as sets of words that have dense relatedness interconnection between each other. Next we cluster the interest based on our similarity metric and so detect communities. An interest of a user is deducted from the articles she has accessed before. Based on the time interval and frequency of accesses we can deduce long term and short term interests. We assume that keywords extracted from accessed articles determine her interests. Similar assumption can be found in [2]. Our interest extraction can be summarized in these steps:

1. Capture all articles accessed by a user,
2. Extract keywords from these articles,

* Supervisor: Michal Barla, Institute of Informatics and Software Engineering

3. Create keyword relatedness graph by connecting the keywords based on their relatedness (i.e. semantic relatedness),
4. Apply algorithms of finding virtual communities on the keyword relatedness graph.

Detected keyword groups are the base of user interests. With a similar approach on all articles we can detect interests in the whole corpus. We refer to users interests as local interests and to interests detected in corpus as global interest. Local interests have lesser cardinality than global interests and they are subsets of global interests.

Detecting communities

A community includes users who have one particular interest in common. To find neighbours for one of users interests we compare it to all other identified interests. The neighbours are then all interests that match to a certain threshold. We define two strategies during interest comparison:

- local interest to local interest
- local interest to global interest

The first approach detects all users that have a similar interest to the selected user interest and therefore it is always specific for one user, compared to the second approach which groups interests relevant to a specific global interest. The community size is cut-off by a threshold or top N rule.

A certain boost to both approaches is to consider the article category specified by the news editor (i.e. Sport, Politics) from which the keyword was extracted.

Our main contribution is the novel approach to user interest deduction and the detection of virtual communities based on these interests.

Acknowledgement. This work was partially supported by the Scientific Grant Agency of Slovak Republic, grant No. VG1/0675/11.

References

- [1] Braak, P.T., Abdullah, N., Xu, Y.: Improving the Performance of Collaborative Filtering Recommender Systems through User Profile Clustering. In *Web Intelligence & Intelligent Agents*, Washington, DC, USA, IEEE Computer Society, pp. 147–150, 2009.
- [2] Grineva, M., Grinev, M., Lizorkin, D.: Extracting key terms from noisy and multitheme documents. In *Proc. of the 18th Int. Conf. on World Wide Web*, New York, New York, USA, ACM Press, pp. 661–670, 2009.
- [3] Lops, P., Degemmis, M., Semeraro, G.: Improving Social Filtering Techniques Through WordNet-Based User Profiles. In *User Modeling 2007*. Vol. 4511, Lecture Notes in Computer Science, Springer, pp. 268–277, 2007.