

Personalized Recommendation of Interesting Texts

Michal KOMPAN*

Slovak University of Technology
Faculty of Informatics and Information Technologies
Ilkovičova 3, 842 16 Bratislava, Slovakia
kompan05@student.fiit.stuba.sk

The data amount on the web is serious problem for the common user. One of the most relevant sources of information over the web are news portals (nytimes.com, reuters.com, etc.). Most of users prefer large renowned news metaportals. They include thousands of daily added news from the whole world and there is no chance to access them in a fast and comfortable way for every user. The only way to help the user is to filter large amount of information and reduce it to an acceptable amount. There are several filtering systems in this domain nowadays [3], [4].

The main problem in the content-based filtering is effective and enough expressive representation of items (or articles). This is often done by means of text summarization [1] or keywords extraction [2]. These techniques are commonly used in English based systems and cannot be easily applied to other languages. Keywords extraction and summarization brings better results as the other methods but are more time consuming. These methods cannot represent non-text documents without modification.

Proposed representation compresses article information value to short vectors, which are used for fast similarity computation over the specific articles time-window. This vector represents article in an effective way, so there is no need to store whole articles. Proposed method expects pre-processed article as an input and produces vector representation usually no longer than 30 words. Then these vectors can be easily used for similarity computations or we can use them in special structures for recommendation e.g. binary trees [5].

Our method for content-based news recommendation uses this effective article representation. We use similar articles to create recommended content based on implicit user model. The method for recommendation is based on three basic steps – computing article similarity, creating user model and recommendation based on first two steps (Fig. 1).

* Supervisor: Mária Bielíková, Institute of Informatics and Software Engineering

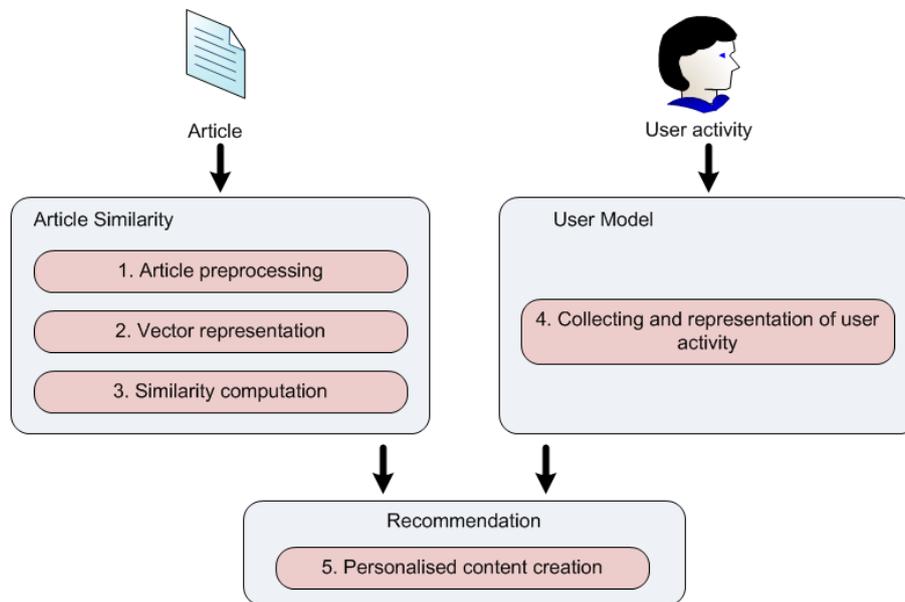


Fig. 1. Proposed news recommendation method.

In the article similarity step it is necessary to preprocess every article to reduce word space. Then is article represented in an effective vector representation, which is used in cosine similarity computation. As a result of article similarity step we obtain a list of similar articles for every article in the dataset. User model is created implicitly based on server logs by identification of visited and recommended article for unique cookie. Finally is the recommended content from both similar articles and user model created.

Acknowledgement. This work was partially supported by by the Cultural and Educational Grant Agency of the Slovak Republic, grant No. KEGA 345-032STU-4/2010.

References

- [1] Dakka, W., Gravano, L. 2007. Efficient summarization-aware search for online news articles. In Proc. of the 7th ACM/IEEE-CS Joint Conf. on Digital Libraries. JCDL '07. ACM, New York, NY, pp. 63-72.
- [2] Kou, H., Gardarin, G., 2009. Keywords Extraction, Document Similarity and Categorization, Tech.rep. No.2002/22, PRiSM Laboratory of Versailles Univ.
- [3] Wongchokprasitti, C., Brusilovsky, P., 2007. Newsme: A case study for adaptive news systems with open user model. In: Proc. of Int. Conf. on Auton. and Auton. Systems, 2007. ICAS07. pp. 69.
- [4] Yoneya, T., Mamitsuka, H., 2007. Pure: a pubmed article recom. system based on content-based filtering. In Proc. of Int. Conf. on Genome Inf. 18, pp. 267-276.
- [5] Zelenik, D., Bielikova, M., 2009. Dynamics in hierarchical classification of news. In Proc. of the 4th Work. on Intelligent and Knowledge oriented Technologies (WIKT 2009), pp. 83-87.