Improving Search Using Graphs and Implicit Feedback

Ján SUCHAL*

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

With the coming era of semantic web, large, structured and linked datasets are becoming common. Unfortunately, current search engines mostly see web only as a graph of pages linked together by hyperlinks, thus becoming insufficient for searching in such new, structured and multidimensional data. When dealing with multidimensional data, identifying relations and attributes that are important for users to achieve their searching goals becomes crucial. Furthermore every user, can have different priorities, different goals which can even change in time.

One of the goals of this work is the extension of existing graph algorithms for multidimensional data, where the usage of tensor algebra and multigraphs can be useful, in contrast with currently preferred matrix algebra. Such extension of graph algorithms would be able to increase relevance and quality of search, and even enable new quality of query formulations.

Evaluation of relevance and quality of search can be done gathering implicit feedback (e.g. quality can be measured just by monitoring user interactions with the system). Another goal of this work is the exploitation of gathered (implicit or explicit) feedback from users to not only evaluates the underlying system, but also to analyse users' behaviour thus opening possibilities for adaptation and personalization.

The main goal of this work is the usage of implicit feedback in search and recommendation engines dealing with large multidimensional data, to improve search and recommendation result quality, speed and scalability.

In particular we focus our work on four major topics:

- Advanced techniques for mining knowledge and feedback from standard server logs, such as viral recommendation detection and probabilistic source identification, negative feedback from positive access logs and time-based trend characteristics.
- Performance and scalability issues of recommendation algorithms for real world large applications, such as collaborative news article recommendations and graph-based ranking algorithms.

---

* Supervisor: Pavol Návrat, Institute of Informatics and Software Engineering

Spring 2010 PeWe Workshop, April 18, 2010, pp. 35–36.

- Dealing with uncertainty and unknown data especially in large sparse matrices and tensors with power-law distributions typical for real-world applications to search and recommendation engines.
- Comparing synthetic evaluation methods for search and recommendation algorithms with real implicit and explicit feedback-based methods.

Theoretical and practical results of our work in these areas include:

- Nearest neighbourhood collaborative filtering algorithm based on generic full text engine exploiting power-law distributions and yielding recommendations comparable to spreading activation graph-based model. Furthermore having linear scalability characteristics with respect to dataset size and easy parallelization to multiple machines/cores.
- Viral recommendation detection and probabilitstic source identification from standard server access logs using time-based referer analysis and exploiting power-law distribution of recommendations [1].
- Negative feedback mining from standard server access logs applicable to news-based portals based on probabilistic „seen-but-not-clicked“ heuristic.
- Recommendation and evaluation framework for major news portal www.sme.sk.
- Application of spreading activation based recommendations on social network of slovak companies register (www.foaf.sk) [2].

Our future work is focused on more comprehensive evaluation of quality and performance characteristics for various datasets and parameter sensitivity.

Acknowledgement. This work was partially supported by the Scientific Grant Agency of Slovak Republic, grant No. VG1/0508/09.

References

[1] Suchal, Ján: Reconstructing Viral Recommendation Networks from Server Logs. In: Student Research Conference 2009. 5st Student Research Conference in Informatics and Information Technologies Bratislava, April 29, 2009: Proceedings in Informatics and Information Technologies. - : STU v Bratislave FIIT, 2009. - ISBN 978-80-227-3052-5. - p. 218-223

[2] Suchal, Ján - Vojtek, Peter: Navigácia v sociálnej sieti obchodného registra SR. In: Datakon 2009, Proceedings of the Annual Database Conference. Srní, Czech Republic, October 10-13. - Prague: University of Economics, 2009. - ISBN 978-80-245-1568-7. - p. 145-151