

A Comparison Between Usage-based and Citation-based Methods for Recommending Scholarly Research Articles

André Vellino
CISTI, National Research Council
Ottawa, Ontario K1A 0R6
andre.vellino@nrc.ca

ABSTRACT

This study compares some of the behavioural characteristics of two recommender systems for scholarly articles in a digital library: a usage-based recommender and an experimental citation-based recommender. Experimental results show that article recommendations based only on usage data are slightly better at solving the perennial data-sparsity problem that plagues collaborative filtering recommenders in digital libraries. However, citation-based recommendations are more semantically diverse and have less in common with conventional search results than the usage-based method. However both of these methods are complementary since most of the time if one recommender produces a list of recommendations the other does not.

Keywords

Recommendation, collaborative filtering, citation analysis, comparison study.

INTRODUCTION

To the scholarly researcher existing search engines are of limited value for discovering relevant “related literature”. Most search engines are based either on full-text indexes – which limit the scope of search and discovery to the terms that are contained in the index – or on databases that index each bibliographic field (author, title, abstract, keyword). Even specialized search engines – such as GoPubMed – that include semantic search features are limited to domain-specific and vocabulary-restricted ontologies.

While these conventional IR methods serve an important function, they do not address the problem of finding research papers that may be relevant but pertain to a different discipline and which contain different terminology that the user does not know to use as search terms.

Most recommendation systems for Digital Libraries attempt to address this issue by suggesting research papers to the user that may be related but don't necessarily share any

terms in common. This is done by applying collaborative filtering methods that are based either on the prior behavior of users (co-downloads) or their explicit preference ratings.

RECOMMENDATION STRATEGIES

One of the central practical problems for recommenders of scholarly articles, stems from the fact that there are simply too few scholars (users) compared to the relatively large number of articles (items) to enable collaborative filtering algorithms to operate effectively. Synthese and 'bX' are instances of two different strategies to overcome this “data sparsity” problem.

'bX' is a commercial, usage-based recommendation service from ExLibris which recommends journal articles by mining SFX logs that record information about users' co-downloads. The recommender's design originates from research on the large scale usage of scholarly resources (Bollen & Sompel, 2006) that permits the harvesting of inter-institutional aggregation of usage data. With this data it is then possible to apply a collaborative filtering algorithm that predicts recommended articles in much the same manner as Netflix recommends movies or iTunes recommends songs.

On the other hand, TechLens+ (Torres et al., 2004) and Synthese (Vellino & Zeber, 2007) address the data sparsity problem by using bibliographic citations in the articles as a substitute for user ratings. The idea is to consider an article as a “user” and view articles that it cites to be the articles' “preferences”.

COMPARISONS

In this experiment, we compared the number of times each recommender produced a result, the “semantic diversity” of the recommendations and the extent to which each of these two approaches was complementary.

The semantic distance between an article and its associated recommendations was determined by measuring the “journal diversity” among the recommended articles relative to the journal in which the starting article was published. Journal diversity was determined from 2.7 million journal-to-journal semantic distances (Newton, Callahan & Dumontier, 2009) that were computed from a large full-text library of scientific, technical and medical articles using Widdows' Semantic Vectors method (Widdows & Ferraro, 2008).

This is the space reserved for copyright notices.

ASIST 2010, October 22–27, 2010, Pittsburgh, PA, USA.
© NRC Canada 2010

For this experiment, a statistically significant collection of 1886 test-articles were randomly selected from the collection of 7.6 million articles spanning 2,750 scientific, technical and medical journals that contain 4.3 million references. This collection was data-mined by Synthese to generate 1.9 million citation-based recommendations.

For each of the test-articles, 'bX' and Synthese attempted to generate recommendations and the results were compared for (i) the number of recommended articles, (ii) the semantic distance between the starting article and the recommended ones, (iii) the number of times that both 'bX' and Synthese recommended articles from a given starting article, and, (iv) for each instance where both 'bX' and Synthese produced recommendations, which one of 'bX' or Synthese had greater semantic diversity. A web demonstration of how these two recommenders compare is available at <http://lab.cisti-icist.nrc-cnrc.gc.ca/Sarkanto/>.

We note that despite the significant size of this collection, the range of journals from which articles can be recommended in 'bX' is still broader than the range available to Synthese.

RESULTS

Out of the 1886 randomly selected test-samples, 'bX' produced recommendations more frequently than Synthese (487 v. 372, or 26% vs. 20%) but the total number of recommended articles was about the same (51% vs. 49%).

Of the total number of instances where a recommendation was produced, the number of times that Synthese and 'bX' both generated results for a given starting article was only 13% (116 out of a total of 859). This means that, most of the time either recommender system generated a result, only one or the other recommender would produce it, indicating a high degree of complementarity.

Furthermore, within this 13% of articles for which both recommenders produced a result, none of the individual recommendations overlapped. Yet the semantic diversity of recommendations in Synthese was 60% compared with 34% in 'bX'. Sometimes (6%) both recommenders produced results with no semantic diversity (i.e. all the recommended articles came from the same journal as the starting article). Overall, the average semantic distance between the starting article and the recommendations was 0.906 for Synthese and 0.935 for 'bX'.

DISCUSSION

A manual inspection of the recommendation lists showed that both 'bX' and Synthese generated recommendations that were topical – if the source article was about “angioplasty” the recommended articles tended to also be about that or a related topic like “coronary artery disease”.

However, 'bX' recommendations tend to contain more occurrences of words that also occur in the title of the original article. This and the lesser diversity of 'bX' recommendations could be explained by the possibility that users' co-downloaded articles originated with a list that was

first generated by a search engine. This hypothesis could be verified by examining SFX logs in combination with the search terms that led to the invocation of the resolver.

The low rate of articles for which both recommenders produces results and the complete lack of intersection in the recommendations themselves reflects the fundamentally different nature of these two methods. The user-based method recommends “other articles that other users also downloaded” whereas the citation-based one recommends “other articles whose citation pattern is similar to this one”.

FUTURE WORK

One likely consequence of employing a usage-only method of recommending articles is that, as researchers download recommended articles, the SFX logs generated by such usage will begin to reinforce these same recommendations in the future, which may further reduce recommendation diversity (Fleder & Hosanagar, 2008).

It would be interesting to undertake a human-subject study that compares the usefulness of a usage-based system that converges to a fixed point over time with one that has already achieved a more diverse fixed point from relatively static citation patterns.

ACKNOWLEDGMENTS

I would like to thank Glen Newton for providing access to data about journal-similarities used in his research and to Heather Moulaison for comments on a previous draft.

REFERENCES

- Bollen, J., & Sompel, H. van de. (2006). An architecture for the aggregation and analysis of scholarly usage data. In *Proc. of the 6th ACM-IEEE joint conference on digital libraries* (p. 298-307). New York, NY: JCDL '06, ACM.
- Fleder, D., & Hosanagar, K. (2008). Blockbuster culture's next rise or fall: The effect of recommender systems on sales diversity (Tech. Rep.). *NET Institute Working Paper 07-10*.
- Newton, G., Callahan, A., & Dumontier, M. (2009). Semantic journal mapping for search visualization in a large scale article digital library. In *Second workshop on very large digital libraries, ECDL 2009*. New York, NY.
- Torres, R., McNee, S., Abel, M., Konstan, J., & Riedl, J. (2004). Enhancing digital libraries with TechLens+. *Proceedings of the 2004 Joint ACM/IEEE Conference on Digital Libraries*, 228-236.
- Vellino, A., & Zeber, D. (2007). A Hybrid, multidimensional recommender for journal articles in a scientific digital library. *Proceedings of the 2007 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, 111-114.
- Widdows, D., & Ferraro, K. (2008). Semantic vectors: A scalable open source package and online technology management application. In *Proceedings of the sixth international language resources and evaluation, LREC '08*. New York, NY: ECDL 2009, ACM.