Library Hi Tech

Recommending research articles using citation data

Andre Vellino

Recommending research articles using citation data

Andre Vellino
University of Ottawa, Ottawa, Canada

Received 7 June 2015
Revised 5 September 2015
Accepted 10 September 2015

Abstract

Purpose – The purpose of this paper is to present an empirical comparison between the recommendations generated by a citation-based recommender for research articles in a digital library with those produced by a user-based recommender (ExLibris "bX").

Design/methodology/approach – For these computer experiments 9,453 articles were randomly selected from among 6.6 M articles in a digital library as starting points for generating recommendations. The same seed articles were used to generate recommendations in both recommender systems and the resulting recommendations were compared according to the "semantic distance" between the seed articles and the recommended ones, the coverage of the recommendations and the spread in publication dates between the seed and the resulting recommendations.

Findings – Out of the 9,453 test runs, the recommendation coverage was 30 per cent for the user-based recommender vs 24 per cent for the citation-based one. Only 12 per cent of seed articles produced recommendations with both recommenders and none of the recommended articles were the same. Both recommenders yielded recommendations with about the same semantic distance between the seed article and the recommended articles. The average differences between the publication dates of the recommended articles and the seed articles is dramatically greater for the citation-based recommender (+7.6 years) compared with the forward-looking user-based recommender.

Originality/value – This paper reports on the only known empirical comparison between the Ex Librix "bX" recommendation system and a citation-based collaborative recommendation system. It extends prior preliminary findings with a larger data set and with an analysis of the publication dates of recommendations for each system.

Keywords Digital libraries, Library services, Computer applications, Citation analysis, Collaborative filtering, Recommender systems

Paper type Research paper

1. Introduction

Despite the relatively slow acceptance of recommender systems technology in library settings (Wakeling *et al.*, 2012), libraries and digital libraries especially, continue to be an important application domain for recommender systems. The rapid growth of open access scholarly research and an increasingly long tail of rarely cited articles, puts an onus on intelligent discovery tools to suggest literature that is not necessarily keyword related to the items that the user has already found with search terms and yet remains topically relevant while offering some degree of serendipity.

The decision to either develop an in-house recommender system or to purchase a recommender system provided by a commercial third party is complex and depends not only on the choice of underlying recommender technology but also on its effectiveness for the user community, the quality of the user interface and the trustworthiness of the recommendations. This was the situation facing the Canada Institute for Scientific and Technical Information (CISTI) (now the National Research Council's National Science Library) in 2012 and the initial motivation for this study: to empirically compare the recommendations produced by two recommender systems employing different data sources and algorithms.

Library Hi Tech
Vol. 33 No. 4, 2015
pp. 597-609

0737-8831
DOI 10.1108/LHT-06-2015-0063

On the surface, research articles bear a sufficient similarity to other kinds of items, such as books, DVDs, and music that the problem of recommending new items of possible interest to users should be akin to if not the same as it is in e-commerce portals. Indeed a number of digital library recommenders have been deployed and studied since the year 2000, many of them employing a variety of types of data sources from which to generate recommendations: users' behaviour data from download logs, circulation data, text content in the articles, user-bookmarks and tags in bibliographic reference managers, and the network of article citations. These data can all be used to cluster users according to some measure of similarity: user-user or item-item similarity based on usage patterns or content-based similarity or some hybrid of the two (Burke, 2002).

Collaborative filtering (Su and Khoshgoftaar, 2009) is the most commonly used method for predicting user preferences and interests based on the collective data of a user community's past usage behaviour. To recommend an item using collaborative filtering, items must have some kind of "preference rating", obtained either explicitly from the user or implicitly from an analysis of usage patterns (clickstream, downloads, etc.) (Pohl *et al.*, 2007), or from an analysis of citation data (McNee *et al.*, 2002).

The effectiveness of a recommender depends as much on the data sources and on the specific algorithm used for measuring similarity as it does on the end-user's task (Herlocker *et al.*, 2004). In particular, it is known that user-based collaborative filtering does not produce useful recommendations if the usage data are sparse (Su and Khoshgoftaar, 2009), a particularly acute problem for items in digital libraries. However, for sparse matrices where the number of articles dominates the number of users it is known that user-based collaborative filtering recommenders have better accuracy than item-based ones (Bogers and van den Bosch, 2008).

Two main strategies have been employed to address the data sparsity problem: harvesting usage data from a large number of distributed usage logs and taking advantage of the citation network of articles as a proxy for preferences. The present study reports on empirical comparisons between the properties of two recommenders that employ these two strategies: the "bX" recommender by ExLibris, which applies collaborative filtering to distributed usage data, and the Sarkanto recommender, which applies collaborative filtering to article citations. This study extends initial results reported in Vellino (2010, 2013) by sampling a larger collection of documents, a greater number of recommendations, and includes, in addition, an analysis of publication date distributions.

The rest of the paper is organized as follows. Section 2 provides a brief survey of recommendation methods in digital libraries and Section 3 presents the motivation and background for this study. Section 4 discusses the methodology. Section 5 concludes the paper and discusses further research directions.

2. Related work

Recommendation systems for digital libraries have been developed using all three of the principal methods used in recommender systems research: user-based collaborative filtering, content-based filtering, and hybrids of the two. However, collaborative filtering remains the dominant method since it is especially useful when the items to be recommended have few or no content-based features. Webster *et al.* (2004) point out that since many traditional library resources, such as catalogues, contain only metadata about the items in a collection (i.e. there is no full text to index), content analysis techniques have limited usefulness. In such situations, collaborative filtering can help induce links between library objects for which there are no syntactic clues for relatedness.

For instance, BibTip (Mönnich and Spiering, 2008; Franke *et al.*, 2008), developed at Karlsruhe University, is an instance of a user-based collaborative filtering recommender for items in a library catalog that employs OPAC usage data and contains neither citation information nor full-text content. Users' session behaviours are mined to extract user-item preferences and in this respect, BibTip most closely resembles conventional commercial product recommenders, except for the absence of explicit ratings.

Similarly, the recommender system that is deployed on the bibliographic social reference manager site CiteULike (Bogers and van den Bosch, 2008) operates like BibTip except that the data source is the site users' collections of article bookmarks. Like OPAC usage data, bookmarked articles are implicit data and do not indicate the user's specific ratings for the articles.

In contrast with collaborative filtering, content-based recommenders need to measure item-item similarity according to some feature extracted from the items or from metadata (Balabanovíc and Shoham, 1997). For text items in a library, this could be the feature vectors obtained from the text, or, for items with no text content (e.g. scanned images), salient features could be provided by metadata such as bibliographic categories, authors, title, abstract, etc.

One well-known feature of content-based recommendations is that they rarely stray semantically from the content clusters of previously rated items. One approach used to overcome this overspecialization of recommendations is either to introduce randomness in the recommendation and to filter out items that are too similar or to complement them with collaborative filtering systems, which provide a source of naturally occurring serendipity from user behaviour (Zhang *et al.*, 2002).

The open-source repository platform DSpace (Elliott *et al.*, 2008) incorporates a content-based recommender that generates recommendations based on a user-selected set of examples that circumscribe the "research context" of the user. Recommendations from DSpace are generated by applying a Jaccard similarity coefficient on the metadata about the articles.

The first collaborative filtering recommender designed specifically for research papers was developed by the University of Minnesota's GroupLens team and later deployed by the University of Minnesota Library (McNee *et al.*, 2002). This system employed the strategy of using citations as a substitute for item ratings to address the problem of usage data sparsity. This is the same strategy that was implemented in the Sarkanto recommender studied in this paper. Its successor, TechLens+ (Torres *et al.*, 2004) adopted a hybrid (citation-based and content-based) approach to improve on its predecessors' precision. The authors also studied users' perceptions of paper recommendations generated by such a hybrid recommender.

Hybrid approaches take various forms, which are clearly summarized by Burke (2002) and Adomavicius *et al.* (2005). One approach uses content-based methods for developing user models and clustering users according to a content-based similarity measure in order to make collaborative recommendations. This enables recommendations to be made either by matching the item's content with the user's profile or by using other users' profiles (Shahabi *et al.*, 2001).

For instance, experiments with the Recommendz system (Garden and Dudek, 2006) have shown that usage data may be usefully combined with full-text information and semantic metadata to provide recommendations. Alternatively, the results of two separate recommenders may be either averaged or given a fair vote depending on the context.

Hybrids of item-based and user-based collaborative filtering systems also exist. Wang *et al.* (2006) describes item-item, user-item, and user-user collaborative filtering in combination with content-based methods both to cluster items and to cluster users. These experiments show that hybrid methods go some way towards alleviating the data sparsity problem and also provide higher quality recommendations.

3. Motivation and background

The initial motivation for this study was to examine empirically how the commercial "bX" recommender behaves compared with the citation-based recommender (Sarkanto) custom-built for CISTI. The objective was to compare and contrast some of the recommendation result-sets and behavioural characteristics of "bX" and Sarkanto with respect to prediction coverage (the percentage of the items for which the system is able to generate a recommendation) and serendipity (the extent to which the recommended items are unexpected to users).

Both these characteristics are hard to compare across recommender systems. Evaluating serendipity as a function of users' expectations would have required a user study beyond the scope of this research. As a proxy for this, I chose instead to consider a measure of "semantic distance" between the journals in which the recommended articles were published. Thus if one recommender produced recommendations mostly from semantically similar journals, it could be viewed as generating suggestions that were less serendipitous. If, on the other hand, recommendations came from a variety of semantically different journals, the recommender would be more serendipitous.

3.1 Article collection

The articles that formed the basis for both the previous and the present experiments were extracted from a collection of approximately 6.6 M articles held by CISTI. The majority of these articles are in the fields of Science, Technology and Medicine and date between 1995 and 2009. They were published in approximately 2,400 journals and conference proceedings in a variety of fields including Medicine (671 journals or 16 per cent of the total), Agricultural and Biological Sciences (9 per cent), Engineering and Technology (8 per cent), Biochemistry, Genetics, and Molecular Biology (6 per cent), Chemistry (5 per cent), and Computer Science (5 per cent).

The full article collection is rich in reference metadata but only 1.8 M articles among the 6.6 M contain references to articles within the collection. In that subset of 1.8 M articles, the average number of references to other articles in the 6.6 M collection was six per article. Although the full text of these articles was available for text analysis, this information was not used because many of the recommendations generated by "bX" were not also articles that existed in this collection and hence could not be text mined without an additional step to harvest the full text of the articles from a third party.

3.2 Data sparsity

If the sparsity of a user-item matrix is measured as the number of links between users and items (either ratings, or the occurrence of a download or citation) divided by the total number of possible links between users and items, then the sparsity of data used for typical collaborative filtering tasks, such as recommending movies with the Netflix data set (Bennett *et al.*, 2007), is about 1 per cent.

In a digital library these ratios are orders of magnitude smaller than in recommenders for commercial merchandise – on the order of tens of thousands of users per month for a collection of tens of millions of items. For example, the sparsity of the matrix of (scholars) to items (articles) in a substantial bibliographic portal such as Mendeley is 2.66×10^{-5}, almost three orders of magnitude smaller than Netflix. In the data provided by Mendeley in response to the DataTEL challenge (Jack *et al.*, 2010), out of the 3,652,286 unique articles, 3,055,546 (83.6 per cent) were referenced by only one user and 378,114 were referenced by only two users. Three or more users referenced less than 6 per cent of the referenced articles and the most frequently referenced article was referenced 19,450 times.

These extremely small user-item ratios are clearly insufficient for collaborative filtering to produce reliable recommendations. Hence the need for supplementary data, either from the aggregation of distributed usage logs or from citation data.

3.3 "bX"

"bX" is a commercially available web service from ExLibris (2009) that recommends research articles using data obtained from OpenURL logs of users' co-downloads. The recommender's design is based on research on the large-scale usage of scholarly resources that permits the harvesting of inter-institutional aggregation of log data (Bollen and van de Sompel, 2006). The quantity of data obtained with this method is sufficiently voluminous to make it possible to apply collaborative filtering effectively: as OpenURL resolver logs grow over time, "bX" recommendations reflect users' aggregate behaviour with increasing precision and accuracy (Herlocker *et al.*, 2004).

3.4 Sarkanto

Another strategy for addressing the data sparsity problem, first used by TechLens+ (Torres *et al.*, 2004), and re-implemented in Sarkanto is to take advantage of bibliographic citations in the articles as a proxy for user ratings. The idea is to consider an article as a "user" and the articles that it cites to be the article's "preferences" (or boolean ratings). Sarkanto is a user-based collaborative filtering recommender that implements *k*-nearest neighbour and cosine correlation in the Taste framework (now Mahout, 2009).

Since the document collection used for this experiment was static, the list of recommendations generated for each article (i.e. "user") was pre-computed rather than dynamically computed. However, recent advances in sparse matrix ordering and partitioning (Küçüktunç *et al.*, 2013) make it possible to generate real-time recommendations efficiently from large, sparse citation networks.

An obvious limitation of this approach is that bibliographic references, while an indicator of relevance, are not necessarily an indication of favourable relevance in the mind of the author. Findings in Case and Higgins (2000) showed that authors were motivated to cite a work for a variety of reasons, including the fact that citing it might promote the authority of their own work or that the cited work deserved criticism. As early as Garfield (1965) identified 15 such reasons for citing a work. Today, the Citation Typing Ontology (Peroni and Shotton, 2012) provides a rich machine-readable taxonomy for the characterization of bibliographic citations with almost 90 semantic relations such as "agrees with", "corrects", "supports", and "uses conclusions from".

While the application of such semantically rich annotations of citation intent would no doubt remove this resolve the relevance-ambiguity of references, there are, as yet, very few applications of the Citation Typing Ontology to present-day publications, let alone publications from the past. However, automated methods for identifying the influential references in an article might at least be useful for eliminating irrelevant references (Zhu *et al.*, 2015).

4. Methodology and quality comparison of Sarkanto and "bX"

A typical method for assessing the effectiveness of a recommender algorithm is leave-one-out cross-validation (Herlocker *et al.*, 2004). For a recommender of scholarly articles that uses citation-based "ratings" such as Sarkanto, a sample of test articles is selected and, for each article in that set, one reference is removed and the recommender is tested for whether it predicts the removed reference. If the removed reference ranks highest in the list of recommendations, it belongs to the Top-1 recommendations, if it ranks in the first five recommendations it belongs to the Top-5, etc.

Results from a previous study that compared citation-based recommendation with a method based on modelling cognitive memory (a model of human memory performance for cognitive tasks) show that the accuracy rate of Sarkanto for Top-10 predictions, using leave-one-out cross-validation on a subset of the collection used in the current study, is close to 20 per cent (Rutledge-Taylor *et al.*, 2008). Given the limitations of the significantly reduced citation data in that study, this is a respectable score.

However, the measures often used to evaluate the efficacy of algorithms such as Top-N or Mean Absolute Error (Herlocker *et al.*, 2004) are not applicable in this situation, principally because the absence of publically available OpenURL log data does not enable the tester to determine which are the gold-standard recommendations. Hence, a meaningful comparative evaluation of the quality of recommendations generated by each recommender could only be provided by a human-subject expert that inspects the results and assesses the relevance of each recommendation (Gunawardana and Shani, 2009).

Therefore, instead of using any of the above measures, I compared the recommendations generated by each of these strategies for other characteristics: coverage, diversity, complementarity, and publication date.

4.1 A priori comparison

The Sarkanto and "bX" recommenders each have a priori strengths and weaknesses in their respective approaches. For instance, while "bX" can take advantage of a voluminous amount of globally distributed usage data, this data may not reflect, even in the aggregate, the interests of specialists in any given field. Usage data from OpenURL logs is indiscriminate between expert researchers and undergraduate university students. In addition, a dependence on usage information makes such a recommender unable to address the recommendation needs of users interested in the end of long tail of sparsely researched areas. One consequence of this is that the publication dates of "bX" recommendations should typically be skewed towards the present.

On the other hand a recommender that uses bibliographic citations instead of usage data suffers from other limitations. One is that citations are static and citation-based recommendations do not reflect current usage trends. In addition, there is a lag period of about two years between the publication of an article for which there begins to develop co-downloading information and it being cited in other publications (Pohl *et al.*, 2007). Hence one would expect at least that much of a difference in the publication dates of recommendations. Finally, as noted earlier, an article's references are not necessarily a signal of endorsement by the author, although there is no reason to believe that co-download information is any more of an endorsement signal.

4.2 Experimental comparison

The present experiments are very similar to the experiments reported in Vellino (2010). They compares the semantic diversity of recommendations, the number of recommended articles the extent to which recommendations from these different sources overlap and the publication date characteristics from each source. Table I shows a sample of recommendations generated by both Sarkanto and "bX".

Citation-based recommendations	User-based recommendations
(1) Computing extreme avalanches (2004) *Cold Regions Science and Technology* 39 161-180	(1) Snow avalanche hazard modelling of large areas using shallow water numerical methods and CIS Gruber U., (2007-10-01) *Environmental modelling & Software* 22 1472-1481
(2) Error in a USGS 30-metre digital elevation model and its impact on terrain modelling (2000) *Journal of Hydrology* 233 154-173	(2) Characteristics and mitigation of the snow avalanche hazard in Kaghan Valley, Pakistan Himalaya De Scally F., (1994) *Natural Hazards* 9 197-213
(3) Dry granular flow modelling including erosion and deposition (2003) *Surveys in Geophysics* 24 569-585	(3) Cartographic modelling of snow avalanche path location within Glacier National Park, Montana Walsh S., (1990-05-01) *Photogrammetric Engineering and Remote Sensing* 56 615-621
(4) An alternative form for the statistical distribution of extreme avalanche runout distances (2005) *Cold Regions Science and Technology* 42 185-193	(4) The influence of tree and branch fracture, overturning and debris entrainment on snow avalanche flow Bartelt P., (2001) *Annals of Glaciology* 32 209-216
(5) Calculating internal avalanche velocities from correlation with error analysis (2003) *Surveys in Geophysics* 24 499-524	(5) Effects of release conditions uncertainty on avalanche hazard mapping Barbolini M., (2002) *Natural Hazards* 25 225-244
(6) Experimental devices to determine snow avalanche basal friction and velocity profiles (2004) *Cold Regions Science and Technology* 38 17-30	(6) Avalanche climatology of the western USA, with an emphasis on Alta, Utah Mock C., (1992) *The Professional Geographer* 44 307-318
(7) Optimization the basis of code making and reliability verification (2000) *Structural Safety* 22 27-60	(7) Altered streamflow and sediment entrainment in the Gunnison Gorge Elliott J.G., (1997) *Journal of the American Water Resources Association* 33 1041-1054
(8) On full-scale avalanche measurements at the Ryggfonn test site, Norway (2007) *Cold Regions Science and Technology* 49 39-53	(8) On probability analysis in snow avalanche hazard zoning Harbitz C., (2001) *Annals of Glaciology* 32 290-298
(9) Dense snow avalanche modelling: flow, erosion, deposition, and obstacle effects (2004) *Cold Regions Science and Technology* 39 193-204	(9) Mammoth Mountain, California Weaver T., (2008) *Skiing* 5 61-66
(10) Error propagation of DEM-based surface derivatives (2005) *Computers and Geosciences* 31 1015-1027	(10) Regionalization and reconstruction of snow water equivalent in the upper Colorado River basin Timilsena J., (2008) *Journal of Hydrology* 352 94-106

Source: The recommendations generated from the seed article – Snow avalanche hazard modelling of large areas using shallow water numerical methods and GIS – U. Gruber, P. Bartelt (2007) *Environmental Modelling and Software* 22:1472

Table I. Sample article recommendation using citation and usage data

The semantic distance between the seed article and the recommendations generated by both "bX" and Sarkanto was measured by examining the "journal diversity" between the recommended articles and the original article. It would have been preferable to use the semantic distance between the seed article and each of the recommended articles based on their full-text content. However, only the full text in our corpus was available for mining and only some of the full text was available for only some of the recommended articles generated by Sarkanto. Even fewer were available in full text for those generated by "bX".

Hence I chose to use an aggregate measure from the data that underlies the semantic map produced by Newton *et al.* (2009) and from which the semantic distances among 2,365 journals are calculated. For the purposes of this map, a "journal" is considered to be the concatenation of the full text of all the available articles in that journal. Each journal is represented by a coloured dot and each colour on this map corresponds to the publisher-generated main subject category to which the journal belongs. This map is reproduced in Figure 1.

The semantic distances between each journal in this map were computed using Widdows' Semantic Vectors method (Widdows and Ferraro, 2008) on the full text of a collection of 5.7 M articles, a curated subset of the collection described above. The average distance between a randomly selected pair of such journals is 0.79 where 1.0 is the distance between a journal and itself and 0.0 is the distance between two journals that have no terms in common. As a whole, this collection is relatively homogeneous in its subject matter. Thus, in this collection, the two most similar journals are at a distance of 0.998 and the two most dissimilar are at a distance of 0.2724.

From the collection of 1.8 M test articles that contain references to other articles in the collection, 9,453 articles were randomly selected as seed articles for generating

Note: Different colours indicate journals in different fields
Source: Newton *et al.* (2009)

Figure 1.
Journal semantic distance map in Newton, Callahan, and Dumontier

recommendations. For the citation-based recommender a "seed article" amounts to a collection of references that substitutes for the "user profile" from which a recommendation is generated. For the user-based recommender "bX", the metadata from the seed article (title, authors, publication date, etc.) are used to construct an OpenURL request a list of recommendations from the "bX" recommender web service. The distribution of subject areas and publication venues for the sampled seed articles was roughly the same as that of the collection as a whole.

For each seed article selected at random from the text collection, we compared the recommendations generated by "bX" and Sarkanto and counted:

(1) the number of recommended articles;

(2) the semantic distance between the seed article and the recommended articles;

(3) the number of times that "bX" and Sarkanto both recommended articles from a given seed article;

(4) for each instance where both "bX" and Sarkanto produced a set of recommendations from the same seed, which one of "bX" or Sarkanto had greater journal diversity; and

(5) the average distance, in years, between the publication date of the seed article and the publication dates of the recommended articles.

Note that the variety of journals that can be recommended in the "bX" system is significantly greater than the range available in Sarkanto, given the limited number of publishers (about 50) in the article collection used by Sarkanto.

4.3 Results

Out of the 9,453 test runs, 2,873 generated one or more recommendations using "bX" compared with 2,263 for Sarkanto (i.e. the recommendation coverage was 30 vs 24 per cent in Table II). Sarkanto recommended an average of 9.7 articles per seed article vs 8.4 for "bX", which was configured to generate as many as possible.

The number of seed articles that produced recommendations with both Sarkanto and "bX" was only 12 per cent meaning that most of the time either one or the other recommender would produce a result, indicating a high degree of complementarity between them. Furthermore, within this 12 per cent of articles for which both recommenders produced a result, none of the recommended articles were the same, as illustrated in Table I. The results of this part of the experiment do not differ significantly from those reported in Vellino (2010).

The results for semantic diversity, however, differ significantly and surprisingly from the earlier study. In this experiment both "bX" and Sarkanto yielded about the same semantic distance measure between the seed article and the recommended articles, namely, 0.948 (for "bX") and 0.956 (for Sarkanto). This is a significant discrepancy from previous results that can be explained by the relatively greater random sample size in this experiment.

	Seeds	Productive seeds	Sarkanto	"bX"	Both
Number	9,453	3,998	2,263	2,873	1,138
%	100	42	24	30	12

Table II. Summary table of citation and usage-based coverage

It is useful to have a baseline for these similarity measures. Given that the subject matter of the collection as a whole is relatively homogeneous and that a randomly selected article from the collection is likely (by this measure) to have a relatively high degree of similarity to the seed article, we also measured the semantic distance between the same seed articles and articles that were randomly generated from the collection to be 0.80. Thus, both "bX" and Sarkanto produce recommendations that have significantly greater semantic similarity to the seed article than to articles chosen at random.

The new result in this study centres on publication dates. We measured the average differences between the publication dates of the recommended articles and the seed articles. Both recommenders are able to and generally do recommend articles that were published either before or after the seed article's publication date. There is, however, an inherent bias in OpenURL usage-log data towards more recent content. Thus the average age of Sakanto recommendations was +7.6 years (prior) to the seed article's publication date whereas for "bX" the average age is −0.6 years, i.e. forward looking, on average. As a baseline reference, for randomly selected articles, the average age distance from the seed article is +6.3 years. Citation-based recommenders are therefore heavily weighted towards recommending older articles.

5. Discussion and future work

Initial experiments with these two recommenders suggested that they produced recommendations with significantly different journal-to-journal semantic diversity, with the citation-based recommender offering greater serendipity. A larger sampling indicates otherwise.

Certainly, both recommenders are topical. Inspection of the side-by-side recommendation lists generated by the 12 per cent of seed articles that generate results for both "bX" and Sarkanto clearly shows this. For instance, if a source article was about "avalanche modelling", as in Table I, the recommended articles tended to also be about snow or computer modelling. Yet both recommenders are also complementary in coverage, if only in the span of date ranges that they each cover. Thus, using citations as a method for dealing with the data sparsity problem is not only an alternative to harvesting large amounts of usage data. It also serves to generate different kinds of recommendations than those from usage data. One depends on the domain of authors' relevance judgments and the other on readers' relevance judgments. The user-based method recommends "articles that other users also downloaded" whereas the citation-based one recommends "articles whose citation patterns are similar to this one".

These differences between data sources play a significant role in the trust that the end-user places on the results. If you exclude survey articles, an author's co-citations are an indication of topic relevance whereas users' co-download correlations may be caused by other factors such as search-engine results. Cited articles are also usually read before they are cited, whereas the choice to download an article might be based primarily on its title or date of publication.

The complementarity in coverage between these two methods suggests that it might be useful to combine them to form a hybrid recommender. There are, however, several unresolved issues with hybridizing these methods, not least of which is how to compare their rankings. Furthermore, end users of recommenders need to understand the sources of data that are used to generate the recommendations if they are going to trust them. Explanations for such properties as topic diversity (when it arises), and publication date biases would help users choose which kind of recommender data sources are most relevant for their information retrieval tasks. Thus, it would be preferable for a system that combined usage-based and citation-based recommendations not to hybridize them but to offer them as complementary alternatives that are accompanied by different explanations for how they were generated.

The significant disparity between the publication dates of recommendations generated by the two methods would benefit from further analysis. While the differences between how citation data and how co-download data are generated may explain some of the differences, there may also be another citation-delay effect at play. According to Hajra and Sen (2006) the age distribution of references made to a paper obeys a power law decay while the age distribution of references made by a paper has an exponential decay. Such a model of the aging characteristics of the citation network for this collection is needed to explain why citation-based recommendations are so much older than even the average distance between a given article and the rest of the collection.

References

Adomavicius, G., Sankaranarayanan, R., Sen, S. and Tuzhilin, A. (2005), "Incorporating contextual information in recommender systems using a multidimensional approach", *ACM Trans. Inf. Syst.*, Vol. 23 No. 1, pp. 103-145.

Balabanovíc, M. and Shoham, Y. (1997), "Fab: content-based, collaborative recommendation", *CACM*, Vol. 40 No. 3, pp. 66-72.

Bennett, J., Elkan, C., Liu, B., Smyth, P. and Tikk, D. (2007), "KDD cup and workshop 2007", *SIGKDD Explor. Newsl.*, Vol. 9 No. 2, pp. 51-52.

Bogers, T. and van den Bosch, A. (2008), "Recommending scientific articles using citeulike", *RecSys '08: Proceedings of the 2008 ACM Conference on Recommender Systems*, *New York, NY*, pp. 287-290.

Bollen, J. and van de Sompel, H. (2006), "An architecture for the aggregation and analysis of scholarly usage data", *Proceedings of the Third International Conference on Digital Information Management, JCDL '06, ACM*, *New York, NY*, pp. 87-92.

Burke, R. (2002), "Hybrid recommender systems: survey and experiments", *User Modeling and User-Adapted Interaction*, Vol. 12 No. 4, pp. 331-370.

Case, D.O. and Higgins, G.M. (2000), "How can we investigate citation behavior? A study of reasons for citing literature in communication", *J. Am. Soc. Inf. Sci.*, Vol. 51 No. 7, pp. 635-645.

Elliott, D., Rutherford, J. and Erickson, J. (2008), "A recommender system for the dspace open repository platform", HP Labs Technical Reports No. HPL-2008-21, HP Laboratories, Bristol, April 8, available at: www.hpl.hp.com/techreports/2008/HPL-2008-21.pdf?jumpid=reg_R1002_USEN

ExLibris (2009), "bX Usage-Based Services transform your discovery experience!", available at: www.exlibrisgroup.com/category/bXUsageBasedServices/ (accessed 15 October 2015).

Franke, M., Geyer-Schulz, A. and Neumann, A.W. (2008), "Recommender services in scientific digital libraries", in Tsihrintzis, G.A. and Jain, L.C. (Eds), *Multimedia Services in Intelligent Environments*, Springer, Berlin, Heidelberg, pp. 377-417.

Garden, M. and Dudek, G. (2006), "Mixed collaborative and content-based filtering with user-contributed semantic features", *Proceedings of the 21st AAAI National Conference on Artificial Intelligence (AAAI'06)*, Vol. 2, *Boston, MA*, pp. 1307-1312.

Garfield, E. (1965), "Can citation indexing be automated?", in Stevens, M.E., *et al.* (Eds), *Statistical Association Methods for Mechanized Documentation, Symposium Proceedings*, National Bureau of Standards, Miscellaneous Publication 269, Washington, DC, pp. 188-192.

Gunawardana, A. and Shani, G. (2009), "A survey of accuracy evaluation metrics of recommendation tasks", *Journal of Machine Learning Research*, Vol. 10, December, pp. 2935-2962.

Hajra, K.B. and Sen, P. (2006), "Modelling aging characteristics in citation networks", *Physica A: Statistical Mechanics and its Applications*, Vol. 368 No. 2, pp. 575-582.

Herlocker, J., Konstan, J., Terveen, L. and Riedl, J. (2004), "Evaluating collaborative filtering recommender systems", *ACM Trans. Inf. Syst.*, Vol. 22 No. 1, pp. 5-53.

Jack, K., Hammerton, J., Harvey, D., Hoyt, J.J., Reichelt, J. and Henning, V. (2010), "Mendeley's reply to the datatel challenge", *Procedia Computer Science*, Vol. 1 No. 2, pp. 1-3.

Küçüktunç, O., Kaya, K., Saule, E. and Catalyürek, U.V. (2013), "Fast recommendation on bibliographic networks with sparse-matrix ordering and partitioning", *Social Network Analysis and Mining*, Vol. 3 No. 4, pp. 1097-1111.

Mahout (2009), "Machine learning libraries", availabe at: mahout.apache.org/ (accessed 15 October 2015).

McNee, S., Albert, I., Cosley, D., Gopalkrishnan, P., Lam, S., Konstan, J. and Riedl, J. (2002), "On the recommending of citations for research papers", *Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work*, pp. 116-125.

Mönnich, M. and Spiering, M. (2008), "Adding value to the library catalog by implementing a recommendation system", *D-Lib Magazine*, Vol. 14 Nos 5/6, available at: www.dlib.org/dlib/may08/monnich/05monnich.html

Newton, G., Callahan, A. and Dumontier, M. (2009), "Semantic journal mapping for search visualization in a large scale article digital library", Second workshop on Very Large Digital Libraries, ECDL, New York, NY.

Peroni, S. and Shotton, D. (2012), "FaBiO and CiTO: ontologies for describing bibliographic resources and citations", *Web Semantics: Science, Services and Agents on the World Wide Web*, Vol. 17, December, pp. 33-43.

Pohl, S., Radlinski, F. and Joachims, T. (2007), "Recommending related papers based on digital library access records", *The Joint Conference on Digital Libraries (JCDL'07), Vancouver, British Columbia, 18-23 June*.

Rutledge-Taylor, M.F., Vellino, A. and West, R.L. (2008), "A holographic associative memory recommender system", *Proceedings of the Third International Conference on Digital Information Management, London*, pp. 87-92.

Shahabi, C., Banaei-Kashani, F., Chen, Y.S. and McLeod, D. (2001), "Yoda: an accurate and scalable web-based recommendation system", in Batini, C., Giunchiglia, F., Giorgini, P. and Mecella, M. (Eds), *Cooperative Information Systems*, Springer, Berlin, Heidelberg, pp. 418-432.

Su, X. and Khoshgoftaar, T.M. (2009), "A survey of collaborative filtering techniques", *Advances in Artificial Intelligence 2009*, Vol. 4 No. 2, pp. 1687-7470.

Torres, R., McNee, S., Abel, M., Konstan, J. and Riedl, J. (2004), "Enhancing digital libraries with techlens+", *Proceedings of the 2004 Joint ACM/IEEE Conference on Digital Libraries*, pp. 228-236.

Vellino, A. (2010), "A comparison between usage-based and citation-based methods for recommending scholarly research articles", *Proceedings of the 73rd American Society for Information Science and Technology Annual Meeting, Pittsburgh, PA, 24-27 October*.

Vellino, A. (2013), "Usage-based vs citation-based methods for recommending scholarly research articles", ArXiv eprints, available at: http://arxiv.org/abs/1303.7149 (accessed 15 January 2014).

Wakeling, S., Clough, P., Sen, B. and Silipigni Connaway, L. (2012), "Readers who borrowed this also borrowed […]: recommender systems in UK libraries", *Library Hi Tech*, Vol. 30 No. 1, pp. 134-150.

Wang, J., de Vries, A. and Reinders, M. (2006), "Unifying user-based and item- based collaborative filtering approaches by similarity fusion", *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 501-508.

Webster, J., Jung, S. and Herlocker, J. (2004), "Collaborative filtering: a new approach to searching digital libraries", *New Review of Information Networking*, Vol. 10 No. 2, pp. 177-191.

Widdows, D. and Ferraro, K. (2008), "Semantic vectors: a scalable open source package and online technology management application", *Proceedings of the Sixth International Language Resources and Evaluation, (LREC 2008)*, pp. 1183-1190.

Zhang, Y., Callan, J. and Minka, T. (2002), "Novelty and redundancy detection in adaptive filtering", *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM Press*, *New York, NY*, pp. 81-88.

Zhu, X., Turney, P., Lemire, D. and Vellino, A. (2015), "Measuring academic influence: not all citations are equal", *Journal of the Association for Information Science and Technology*, Vol. 66 No. 2, pp. 408-427. doi: 10.1002/asi.23179.

Corresponding author
Dr Andre Vellino can be contacted at: avellino@uottawa.ca