Aalborg Universitet

AALBORG UNIVERSITY
DENMARK

Report on RecSys 2014 Workshop on New Trends in Content-Based Recommender Systems (CBRecSys 2014)

Bogers, Toine; Koolen, Marijn; Cantádor, Ivan

Published in:
SIGIR Forum

Publication date:
2015

Document Version
Publisher's PDF, also known as Version of record

Link to publication from Aalborg University

Citation for published version (APA):
Bogers, T., Koolen, M., & Cantádor, I. (2015). Report on RecSys 2014 Workshop on New Trends in Content-Based Recommender Systems (CBRecSys 2014). *SIGIR Forum*, *49*(1), 20-26. http://sigir.org/files/forum/2015J/p020.pdf

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy
If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Report on RecSys 2014 Workshop on New Trends in Content-Based Recommender Systems

Toine Bogers[1] **Marijn Koolen**[2] **Iván Cantador**[3]
toine@hum.aau.dk marijn.koolen@uva.nl ivan.cantador@uam.es
[1]Aalborg University Copenhagen, Denmark
[2]University of Amsterdam, Netherlands
[3]Universidad Autónoma de Madrid, Spain

Abstract

While content-based recommendation has been applied successfully in many different domains, it has not seen the same level of attention as collaborative filtering techniques have. However, there are many recommendation domains and applications where content and metadata play a key role, either *in addition to* or *instead of* ratings and implicit usage data. For some domains, such as movies, the relationship between content and usage data has seen thorough investigation already, but for many other domains, such as books, news, scientific articles, and Web pages we still do not know *if* and *how* these data sources should be combined to provided the best recommendation performance. The *CBRecSys 2014* workshop aimed to address this by providing a dedicated venue for papers dedicated to all aspects of content-based recommender systems.

1 Motivation and Goals

While content-based recommendation has been applied successfully in many different domains [9], it has not seen the same level of attention as collaborative filtering techniques have. In recent years, competitions like the Netflix Prize[1], CAMRA[2], and the Yahoo! Music KDD Cup 2011 [5] have spurred on advances in collaborative filtering and how to utilize ratings and usage data from a community of users. However, there are many recommendation domains and applications where item content and metadata play a key role, either *in addition to* or *instead of* ratings and implicit usage data. For some domains, such as movies, the relationship between content and usage data has seen thorough investigation already (e.g. [12]), but for many other domains, such as books, news, scientific articles, and Web pages we still do not know *if* and *how* these data sources should be combined to provided the best recommendation performance. On top of that, in certain contexts of search, content-based recommendation plays a role as well. Personalised IR

[1]http://www.netflixprize.com/
[2]http://www.dai-labor.de/camra2010/

is strongly related to recommendation, but recent research in book search also identified search scenarios that combine aspects of retrieval and recommendation [8].

The *CBRecSys 2014* workshop aimed to address this by providing a venue for papers dedicated to all aspects and new trends of content-based recommender systems. This would include both recommendation in domains where textual content is abundant (e.g. books, news, scientific articles, jobs, educational resources, and Web pages) as well as dedicated comparisons and combinations of content-based techniques with collaborative filtering approaches.

2 Workshop Focus & Format

Authors were encouraged to submit papers on a variety of topics related to Content-Based Recommendation (CBR), including:

- Developing novel recommendation approaches
- Exploiting user-generated content for recommendation
- Processing text reviews
- Mining contextual data from content
- Addressing limitations of recommender systems

In particular, papers submitted to the the workshop focused on the following topics. Several papers presented hybrid systems combining collaborative filtering and content-based recommendation, finding them complementary, with content-based recommendation components especially suitable for tackling the cold-start problem. Other papers investigated how different content features can be used for similarity measures, and explored ways to identify which features are the most relevant for a given context. Some papers presented approaches to mine textual reviews for inferring user preferences on specific attributes of items, essentially deriving more structured feature information from unstructured text. Finally, several papers looked at semantic frameworks and Linked Open Data to measure item similarity across different domains. A total of 16 papers were submitted, of which 10 were accepted for an oral presentation (acceptance rate of 63%). All papers were reviewed by at least 3 reviewers.

2.1 Keynote

The workshop started with a keynote presentation on semantics-aware recommender systems by Pasquale Lops (University of Bari Aldo Moro, Italy). Pasquale's keynote gave a great introduction to CBR and highlighted the main problems, such as limited content analysis and overspecialization. He proposed several research directions for overcoming these challenges, such as top-down semantic approaches based on the use of different open knowledge sources (ontologies, Wikipedia, DBpedia) as well as bottom-up semantic approaches based on the distributional hypothesis (which states that "words that occur in the same contexts tend to have similar meanings"). He also touched upon the potential for cross-language recommender systems and algorithms for generating more serendipitous recommendations, which parallels an active research topic in IR research.

2.2 Accepted Papers

Bauman and Tuzhilin [1] presented a new method of discovering relevant contextual information from the user-generated reviews. They compared word-based and LDA-based approaches on Yelp data and found that the methods provide excellent, complementary performance.

Chow et al. [2] proposed HybridRank, a novel hybrid algorithm for mobile game recommendation. Their algorithm is based on a personalised random walk approach which incorporates both content-based and user-based information and is successful in countering the cold-start problem.

De Clercq et al. [3] investigated on extracting and adding semantic features based on Linked Open Data to a content-based book recommender. Their frame-based approach is shown to outperform a basic bag-of-words approach.

De Pessemier et al. [11] presented a hybrid strategy combining collaborative filtering and content-based techniques for mobile shopping with the primary aim of preserving the customers privacy.

Dias et al. [4] proposed Improvise, a personalized music recommendation solution for daily activities, whose approach associates music content (acoustic features) with activities (context). Preliminary evaluation shows that Improvise is able to both effectively recommend songs for daily activities and adjust to individual users tastes.

Fernández-Tobías & Cantador [6] presented an alternative cross-domain recommendation model based on a novel extension of the SVD++ algorithm that enriches both user and item profiles with independent sets of tag factors. Evaluation on the books and movie domains show that it outperforms existing approaches, even in cold-start situations.

Herzog and Wörndl [7] propose a travel recommender system that supports travelers in creating composite trips consisting of multiple destinations or activities while still respecting their limitations in time and money. An expert study showed that their algorithm provided more satisfactory recommendations than a set of baseline algorithms.

Musto et al. [10] investigated the contribution of several features extracted from the Linked Open Data cloud to the accuracy of different recommendation algorithms. Their results showed the potential of Linked Open Data-enabled approaches to outperform existing state-of-the-art algorithms.

Sahebi and Walker [13] provide a generic framework for content-based cross-domain recommendations that includes an efficient method of feature augmentation to implement domain adaptation. They study an implementation of their framework based on logistic regression on a job recommendation dataset from LinkedIn and find their results show promising performance.

Sinha and Roy [14] investigate the potential of preference mapping for identifying the most relevant product attributes for a marketing campaign. They test their approach on user reviews mined from the Web and provide an effective visualization of their results.

2.3 Discussion Sessions

CBRecSys 2014 proved to be a popular workshop as it was attended by over 60 participants and attendance remained high throughout the day. The morning session was concluded by a general discussion on future challenges in content-based recommender systems. Challenges that were commonly mentioned by participants included dealing with (1) real-time scenarios such as twitter

and news streams, (2) shopping: both online and the connection between on- and offline shopping, and (3) content-based recommendation in non-text domains such as music and video.

The afternoon session concluded the workshop with a general discussion on the desirability of a competition on content-based recommendation similar to the RecSys Challenges, and which requirements are important for such a challenge to be relevant and successful. In addition to spacing such challenges evenly throughout the year, workshop participants agreed that it was essential that the domain should closely match realistic commercially interesting domains, users and items.

3 Book Recommendation Challenge

To facilitate exploration of the above mentioned topics, CBRecSys featured an in-workshop challenge on book recommendation. This challenge focused on recommending new, interesting books to LibraryThing users based on usage data (which books they had added to their collection), and content-based information about the books available in LibraryThing. The challenge is directly linked to the CLEF Social Book Search Lab[3] with the aim to bring the communities of IR and (content-based) recommender systems closer together, through shared data sets and scenarios to range from pure retrieval to pure recommendation. The rich textual nature of the task made the challenge an excellent venue to revisit questions about the benefits of content-based filtering vs. collaborative filtering, and metadata versus ratings information. Presentation of the results was scheduled during the workshop. Unfortunately, the book recommendation challenge competed for interest with the main RecSys 2014 challenge as well as ESCW 2014's recommendation challenge and received no submissions.

3.1 Challenge Dataset

For this challenge, a large dataset containing user profiles with book ratings and tags, and 2.8 million book descriptions with library metadata, user ratings, tags, and reviews from Amazon and LibraryThing was be made available.

The dataset for the book recommendation challenge was comprised of two parts: usage data and book metadata. The first part of the dataset for book recommendation was a log of usage data: who added which books to their collection at what point in time. In addition to this, ratings and tags assigned to books were also included in the usage dataset (where available). The user profiles in this data set contained a total number of 1,830,958 unique books added by 78,633 different LibraryThing users, anonymized through different privacy-preserving measures. This usage data served as the main data source for evaluating our challenge.

The second part of our challenge dataset for book recommendation was a collection of metadata records for 2.8 million books cataloged on LibraryThing. This collection was crawled from Amazon and LibraryThing by the University of Duisburg-Essen in early 2009. From Amazon, there was formal metadata like book title, author, publisher, publication year, library classification codes, Amazon categories, and similar product information, as well as user-generated content in the form of user ratings and reviews. From LibraryThing, there were user tags and user-provided metadata

[3]URL: `http://social-book-search.humanities.uva.nl/`

on awards, book characters and locations, and blurbs. This part of the challenge data had been used successfully for more retrieval-oriented challenges at the INEX/CLEF 2011–2015 Social Book Search tracks[4].

For more details about the data set and the evaluation process of the CBRecSys challenge, please consult the workshop website at `http://ir.ii.uam.es/cbrecsys2014/challenge.html`.

4 Future Workshop

The high number of participants clearly showed that there is a renewed interest in research on content-based recommendation. The CBRecSys workshop will therefore continue with a second edition at RecSys 2015 in Vienna, Austria, on September 16–20. The 2015 workshop website can be found at `http://humanities.uva.nl/~mkoolen1/CBRecSys15/`.

5 Acknowledgments

We would like to thank ACM and RecSys for hosting this workshop, the RecSys workshop chairs Lior Rokach and Badrul Sarwar. We would also like to thank the program committee: Linas Baltrunas, Alejandro Bellogn, Shlomo Berkovsky, Robin Burke, Pablo Castells, Federica Cena, Paolo Cremonesi, Marco De Gemmis, Ernesto W. De Luca, Tommaso Di Noia, Peter Dolog, Juan M. Fernández-Luna, Ignacio Fernández-Tobías, Frank Hopfgartner, Juan F. Huete, Marius Kaminskas, Alexandros Karatzoglou, Birger Larsen, Pasquale Lops, Alan Said, Markus Schedl, Giovanni Semeraro, Nava Tintarev, Marko Tkalčič, David Vallet, and Markus Zanker. Final thanks are due to the paper authors, the keynote speaker Pasquale Lops, and the participants for a great and lively workshop.

The workshop material (list of accepted papers, invited talk, and the workshop schedule) can be found on the CBRecSys 2014 workshop website at `http://ir.ii.uam.es/cbrecsys2014/`. The proceedings were published as a CEUR Workshop Proceedings volume, available at `http://ceur-ws.org/Vol-1245/`.

References

[1] K. Bauman and A. Tuzhilin. Discovering contextual information from user reviews for recommendation purposes. In *Proceedings of the 1st Workshop on New Trends in Content-based Recommender Systems co-located with the 8th ACM Conference on Recommender Systems, CBRecSys@RecSys 2014, Foster City, Silicon Valley, California, USA, October 6, 2014.*, pages 2–9, 2014.

[2] A. Chow, M. N. Foo, and G. Manai. Hybridrank: A hybrid content-based approach to mobile game recommendations. In *Proceedings of the 1st Workshop on New Trends in Content-based Recommender Systems co-located with the 8th ACM Conference on Recommender Systems,*

[4] `http://social-book-search.humanities.uva.nl/`

CBRecSys@RecSys 2014, Foster City, Silicon Valley, California, USA, October 6, 2014., pages 10–13, 2014.

[3] O. D. Clercq, M. Schuhmacher, S. P. Ponzetto, and V. Hoste. Exploiting framenet for content-based book recommendation. In *Proceedings of the 1st Workshop on New Trends in Content-based Recommender Systems co-located with the 8th ACM Conference on Recommender Systems, CBRecSys@RecSys 2014, Foster City, Silicon Valley, California, USA, October 6, 2014.*, pages 14–21, 2014.

[4] R. Dias, M. J. Fonseca, and R. Cunha. A user-centered music recommendation approach for daily activities. In *Proceedings of the 1st Workshop on New Trends in Content-based Recommender Systems co-located with the 8th ACM Conference on Recommender Systems, CBRecSys@RecSys 2014, Foster City, Silicon Valley, California, USA, October 6, 2014.*, pages 26–33, 2014.

[5] G. Dror, N. Koenigstein, Y. Koren, and M. Weimer. The Yahoo! Music Dataset and KDD-Cup '11. In *JMLR Workshop and Conference Proceedings*, volume 18 of *Proceedings of KDD Cup 2011*, pages 3–18, 2012.

[6] I. Fernández-Tobías and I. Cantador. Exploiting social tags in matrix factorization models for cross-domain collaborative filtering. In *Proceedings of the 1st Workshop on New Trends in Content-based Recommender Systems co-located with the 8th ACM Conference on Recommender Systems, CBRecSys@RecSys 2014, Foster City, Silicon Valley, California, USA, October 6, 2014.*, pages 34–41, 2014.

[7] D. Herzog and W. Wörndl. A travel recommender system for combining multiple travel regions to a composite trip. In *Proceedings of the 1st Workshop on New Trends in Content-based Recommender Systems co-located with the 8th ACM Conference on Recommender Systems, CBRecSys@RecSys 2014, Foster City, Silicon Valley, California, USA, October 6, 2014.*, pages 42–48, 2014.

[8] M. Koolen, T. Bogers, A. van den Bosch, and J. Kamps. Looking for books in social media: An analysis of complex search requests. In *Advances in Information Retrieval: 37th European Conference on IR Research (ECIR 2015)*, LNCS. Springer, 2015.

[9] P. Lops, M. de Gemmis, and G. Semeraro. Content-based Recommender Systems: State of the Art and Trends. In *Recommender Systems Handbook*, pages 73–105. 2011.

[10] C. Musto, P. Basile, P. Lops, M. de Gemmis, and G. Semeraro. Linked open data-enabled strategies for top-n recommendations. In *Proceedings of the 1st Workshop on New Trends in Content-based Recommender Systems co-located with the 8th ACM Conference on Recommender Systems, CBRecSys@RecSys 2014, Foster City, Silicon Valley, California, USA, October 6, 2014.*, pages 49–56, 2014.

[11] T. D. Pessemier, K. Vanhecke, and L. Martens. A hybrid strategy for privacy-preserving recommendations for mobile shopping. In *Proceedings of the 1st Workshop on New Trends*

in Content-based Recommender Systems co-located with the 8th ACM Conference on Recommender Systems, CBRecSys@RecSys 2014, Foster City, Silicon Valley, California, USA, October 6, 2014., pages 22–25, 2014.

[12] I. Pilászy and D. Tikk. Recommending New Movies: Even a Few Ratings Are More Valuable Than Metadata. In *RecSys '09: Proceedings of the Third ACM Conference on Recommender Systems*, pages 93–100. ACM, 2009.

[13] S. Sahebi and T. Walker. Content-based cross-domain recommendations using segmented models. In *Proceedings of the 1st Workshop on New Trends in Content-based Recommender Systems co-located with the 8th ACM Conference on Recommender Systems, CBRecSys@RecSys 2014, Foster City, Silicon Valley, California, USA, October 6, 2014.*, pages 57–64, 2014.

[14] M. Sinha and R. S. Roy. Preference mapping for automated recommendation of product attributes for designing marketing content. In *Proceedings of the 1st Workshop on New Trends in Content-based Recommender Systems co-located with the 8th ACM Conference on Recommender Systems, CBRecSys@RecSys 2014, Foster City, Silicon Valley, California, USA, October 6, 2014.*, pages 65–68, 2014.