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 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 been thoroughly investigated 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 provide the best recommendation performance. The CBRecSys 2014 workshop aims to address this by providing a dedicated venue for papers dedicated to all aspects of content-based recommendation.

Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval—information Filtering; D.2.8 [Software Engineering]: Metrics—performance measures

General Terms

Algorithms, Experimentation, Human Factors, Theory

Keywords

recommender systems, content-based recommendation, text reviews, user-generated content, implicit feedback, semantics, context

1. MOTIVATION AND GOALS

While content-based recommendation has been applied successfully in many different domains [2], it has not seen the same level of attention as collaborative filtering techniques have. In recent years, competitions like the Netflix Prize¹, CAMRA², and the Yahoo! Music KDD Cup 2011 [1] have spurred on advances in collaborative filtering and how to utilize ratings and usage data. 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 been thoroughly investigated already (e.g. [3]), 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 provide the best recommendation performance.

The CBRecSys 2014 workshop aims to address this by providing a venue for papers dedicated to all aspects and new trends of content-based recommendation. 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. TOPICS OF INTEREST

Relevant topics of the workshop include:

- Developing novel recommendation approaches
  - Hybrid strategies combining content-based and collaborative filtering recommendations
  - Content-based approaches to cross-system and cross-domain recommendation
  - Latent factor models for content-based and hybrid recommendation

- Exploiting user-generated content for recommendation
  - Mining microblogging data in content-based recommender systems
  - Social tag-based recommender systems
  - Exploiting Semantic Web and Linked Open Data in content-based recommender systems

- Processing text reviews
  - Estimating (implicit) ratings associated with text reviews
  - Opinion mining and sentiment analysis of text reviews to support content-based recommendation
  - Extracting user personality traits and factors from text reviews for recommendation

¹http://www.netflixprize.com/
²http://www.dai-labor.de/camra2010/
- Mining contextual data from content
  - Extraction of contextual signals from textual content for recommendation
  - Incorporating the temporal dimension in content-based recommendation
  - Mood-based recommender systems
- Addressing limitations of recommender systems
  - Addressing the cold-start problem with content-based recommendation approaches
  - Increasing diversity of content-based recommendations
  - Providing novelty in content-based recommendations

In particular, papers submitted to the workshop have focused on the following topics. Several papers present 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 investigate how different content features can be used for similarity measures and explore ways to identify which features are the most relevant for a given context. Some papers present approaches to mine user reviews for inferring user preferences on specific attributes of items, essentially deriving more structured feature information from unstructured text. Finally, several papers look at semantic frameworks and Linked Open Data to measure item similarity across different domains.

3. CHALLENGE ON BOOK RECOMMENDATION

To facilitate exploration of the above mentioned topics, the workshop features an in-workshop challenge on book recommendation. This challenge focuses on recommending new, interesting books to LibraryThing users based on usage data (which books they have added to their collection) and content-based information about the books available in LibraryThing. The rich textual nature of the task makes the challenge an excellent venue to revisit questions about the benefits of content-based filtering vs. collaborative filtering, and metadata versus ratings information. At the workshop the evaluation results of the challenge were presented.

3.1 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 made available.

The dataset for the book recommendation challenge is comprised of two parts: usage data and book metadata. The first part of the dataset for book recommendation is 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 are also included in the usage dataset (where available). The user profiles in this data set contain a total number of 1,830,958 unique books added in the usage dataset (where available). The user profiles in this data set contain a set of users that are not included in the original training material to avoid over-fitting. To allow participants to train their systems on these new users as well, we also released a small amount of extra training material corresponding to the users near the end of the challenge.

4. WEBSITE AND PROCEEDINGS

The workshop material (list of accepted papers, invited talk, and the workshop schedule) can be found on the CBRRecSys 2014 workshop website at http://iri.ii.uam.es/cbrecsys2014/.

The proceedings are published as a CEUR Workshop Proceedings volume.

5. REFERENCES


https://inex.mmci.uni-saarland.de/tracks/books/