TagRec: Towards A Standardized Tag Recommender Benchmarking Framework

Dominik Kowald  
Know-Center  
Graz University of Technology  
Graz, Austria  
dkowald@know-center.at

Emanuel Lacic  
KTI  
Graz University of Technology  
Graz, Austria  
elacic@know-center.at

Christoph Trattner  
Know-Center  
Graz University of Technology  
Graz, Austria  
ctrattner@know-center.at

ABSTRACT

In this paper, we introduce TagRec, a standardized tag recommender benchmarking framework implemented in Java. The purpose of TagRec is to provide researchers with a framework that supports all steps of the development process of a new tag recommendation algorithm in a reproducible way, including methods for data pre-processing, data modeling, data analysis and recommender evaluation against state-of-the-art baseline approaches. We show the performance of the algorithms implemented in TagRec in terms of prediction quality and runtime using an evaluation of a real-world folksonomy dataset. Furthermore, TagRec contains two novel tag recommendation approaches based on models derived from human cognition and human memory theories.

Categories and Subject Descriptors
H.2.8 [Database Management]: Database Applications—Data mining; H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval—Information filtering

Keywords
personalized tag recommendations; recommender framework; recommender evaluation; Java

1. INTRODUCTION

In recent years social tagging has become an important instrument of Web 2.0, which allows users to collaboratively annotate and search content. In order to support this process, current research has attempted to improve the performance and quality of tag recommendations. However, although various tag recommender approaches and experiments exist, most of them use different data pre-processing methods and evaluation protocols, making it difficult for researchers to reproduce these experiments and to compare these approaches with other algorithms.

To tackle this issue, we developed TagRec, a standardized tag recommender benchmarking framework that provides researchers with methods for data pre-processing, data modeling, data analysis and recommender evaluation against state-of-the-art baseline approaches. The purpose of TagRec is not only to increase the reproducibility in the tag recommender research but also to decrease the workload of developers who implement or test a new algorithm for tag recommendations.

2. SYSTEM OVERVIEW

TagRec was fully implemented in Java apart from the FM and PITF algorithms that were provided as a C++ framework by the University of Konstanz. TagRec is open-source and can be downloaded via Github.

Figure 1 shows the system architecture of TagRec, which consists of four main components:

- **Data pre-processing.** TagRec offers various methods for data pre-processing: (1) parsing and processing of social tagging datasets, such as CiteULike, BibSonomy, Delicious, LastFm, MovieLens and Flickr, into the system’s data format; (2) p-core pruning; (3) training/test set splitting (e.g., leave-one-out, time-based or 80/20 splits) [3] and (4) creating Latent Dirichlet Allocation [6] topics for category-based algorithms, such as 3Layers [4, 10].

- **Data model.** The data model of TagRec is generated from simple .csv files that contain the bookmarks (i.e., the combination of user-id, resource-id, timestamp and assigned tags) in a folksonomy. Furthermore, the data model is fully object-oriented and provides distinct classes and powerful methods for modeling and analyzing the relationship and interactions between users, resources and tags (e.g., the number of times a specific tag has been assigned to a target resource or the time since the last usage of a specific tag in the tag assignments of a target user).

- **Recommendation algorithms.** This component is the main part of TagRec and contains the implementations of the various algorithms shown in Table 1. Along with the state-of-the-art approaches
Table 1: Tag recommender algorithms implemented in TagRec.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Name</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>MP</td>
<td>Most popular tags</td>
<td>Jäschke et al. [3]</td>
</tr>
<tr>
<td>MPu</td>
<td>Most popular tags by user</td>
<td>Jäschke et al. [3]</td>
</tr>
<tr>
<td>MPu,r</td>
<td>Most popular tags by resource</td>
<td>Jäschke et al. [3]</td>
</tr>
<tr>
<td>MP</td>
<td>Mixture of MPu and MPu,r</td>
<td>Jäschke et al. [3]</td>
</tr>
<tr>
<td>CFu</td>
<td>User-based Collaborative Filtering</td>
<td>Marinho et al. [8]</td>
</tr>
<tr>
<td>CF,FM</td>
<td>Resource-based Collaborative Filtering</td>
<td>Marinho et al. [8]</td>
</tr>
<tr>
<td>CFu,r</td>
<td>Mixture of CFu and CFu,r</td>
<td>Marinho et al. [8]</td>
</tr>
<tr>
<td>APR</td>
<td>Adapted PageRank</td>
<td>Jäschke et al. [3]</td>
</tr>
<tr>
<td>FR</td>
<td>FolkRank</td>
<td>Jäschke et al. [3]</td>
</tr>
<tr>
<td>FM</td>
<td>Factorization Machines</td>
<td>Rendle et al. [9]</td>
</tr>
<tr>
<td>PITF</td>
<td>Pairwise Interaction Tensor Factorization</td>
<td>Rendle et al. [9]</td>
</tr>
<tr>
<td>LDA</td>
<td>Latent Dirichlet Allocation</td>
<td>Krestel et al. [6]</td>
</tr>
<tr>
<td>LDA&amp;LM</td>
<td>Mixture of LDA and MPu,r</td>
<td>Krestel et al. [6]</td>
</tr>
<tr>
<td>3L</td>
<td>3Layers</td>
<td>Seilitinger et al. [10]</td>
</tr>
<tr>
<td>3LT</td>
<td>Time-dependent 3L</td>
<td>Kowald et al. [4]</td>
</tr>
<tr>
<td>GIRP</td>
<td>Temporal Tag Usage Patterns</td>
<td>Zhang et al. [11]</td>
</tr>
<tr>
<td>GIRPTM</td>
<td>Mixture of GIRP and MPu</td>
<td>Zhang et al. [11]</td>
</tr>
<tr>
<td>BLL</td>
<td>Base Level Learning Equation</td>
<td>Kowald et al. [5]</td>
</tr>
<tr>
<td>BLL+C</td>
<td>Mixture of BLL and MPu</td>
<td>Kowald et al. [5]</td>
</tr>
</tbody>
</table>

3. RESULTS

To show the functionalities of TagRec, we evaluated and compared a selection of the implemented algorithms in terms of recommender quality and runtime using a real-world folksonomy dataset gathered from the image sharing portal Flickr. The dataset contained 9,590 users, 864,679 resources, 127,599 tags and 3,552,540 tag assignments and was split into a training and test set using the leave-one-out pre-processing method of TagRec (i.e., the latest bookmark for each user was used for testing and the rest for training). To quantify the prediction quality of the approaches, the set of well-known Information Retrieval metrics available in TagRec (R@k, P@k, F1@k, MRR, MAP, nDCG and UC) was used (see also [5]).

The first plot in Figure 2 shows the recommender quality of the various approaches in the form of recall/precision plots for k = 1 - 10 recommended tags. The results show that all algorithms, except for the simple MP approach, perform reasonably well on the dataset and that the two newly developed approaches based on human cognition (3L and 3LT) and human memory (BLL and BLL+C) theories perform best.

The runtime comparison is shown in the second plot in Figure 2, which indicates the full time required for providing tag recommendations for all user-resource pairs in the Flickr test set. Clearly, the BLL+C and 3LT approaches, which performed best in the recommendation quality experiment, also provided a reasonable runtime in contrast to the more complex algorithms, such as LDA, APR, FR, FM and PITF.

4. CONCLUSIONS & FUTURE WORK

In this work we presented TagRec, a standardized tag recommender benchmarking framework that provides researchers with methods for data pre-processing, data modeling, data analysis and recommender evaluation in a reproducible way. TagRec was fully implemented in Java and contained a rich set of state-of-the-art tag recommender algorithms along with two newly developed and published tag recommendation mechanisms based on models derived from human cognition (3L and 3LT) and human memory (BLL and BLL+C) theories.

In the future we plan to expand the framework by using more algorithms for tag recommendations and, especially, by content-based methods [1] since at the moment TagRec focuses on folksonomy-based approaches. Furthermore, we would like to adapt the implemented algorithms and evaluation procedures in order to also provide resource and user recommendations.

Acknowledgments: This work is supported by the Know-Center and the EU funded project Learning Layers (7th Framework Program, Grant Agreement 318209). The Know-Center is funded within the Austrian COMET Program - Competence Centers for Excellent Technologies - under the auspices of the Austrian Ministry of Transport, Innovation and Technology, the Austrian Ministry of Economics and Labor and by the State of Styria. COMET is managed by the Austrian Research Promotion Agency (FFG).
5. REFERENCES


