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.

HT'14, September 1–4, 2014, Santiago, Chile.
ACM 978-1-4503-2954-5/14/09.
http://dx.doi.org/10.1145/2631775.2631781.

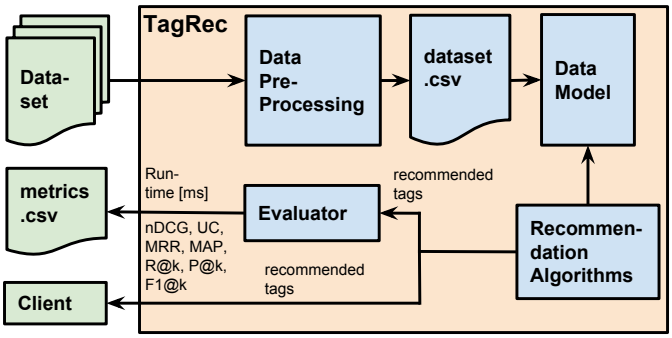


Figure 1: *TagRec* system architecture.

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[1].

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

[1]https://github.com/learning-layers/TagRec/

Algorithm	Name	Authors
MP	Most popular tags	Jäschke et al. [3]
MP_u	Most popular tags by user	Jäschke et al. [3]
MP_r	Most popular tags by resource	Jäschke et al. [3]
$MP_{u,r}$	Mixture of MP_u and MP_r	Jäschke et al. [3]
CF_u	User-based Collaborative Filtering	Marinho et al. [8]
CF_r	Resource-based Collaborative Filtering	Marinho et al. [8]
$CF_{u,r}$	Mixture of CF_u and CF_r	Marinho et al. [8]
APR	Adapted PageRank	Jäschke et al. [3]
FR	FolkRank	Jäschke et al. [3]
FM	Factorization Machines	Rendle et al. [9]
PITF	Pairwise Interaction Tensor Factorization	Rendle et al. [9]
LDA	Latent Dirichlet Allocation	Krestel et al. [6]
LDA&LM	Mixture of LDA and $MP_{u,r}$	Krestel et al. [6]
3L	3Layers	Seitlinger et al. [10]
3LT	Time-dependent 3L	Kowald et al. [4]
GIRP	Temporal Tag Usage Patterns	Zhang et al. [11]
GIRPTM	Mixture of GIRP and MP_r	Zhang et al. [11]
BLL	Base Level Learning Equation	Kowald et al. [5]
BLL+C	Mixture of BLL and MP_r	Kowald et al. [5]

Table 1: Tag recommender algorithms implemented in *TagRec*.

to folksonomy-based tag recommendations (e.g., Collaborative Filtering, FolkRank or Pairwise Interaction Tensor Factorization) [7] the engine contains two newly developed and recently published algorithms based on models derived from human cognition (3L and 3LT) and human memory (BLL and BLL+C) theories. All algorithms implement a common interface which, making it easy to develop and integrate new approaches. The predicted tags generated by the different algorithms can be forwarded either to the evaluation engine or directly to a client application.

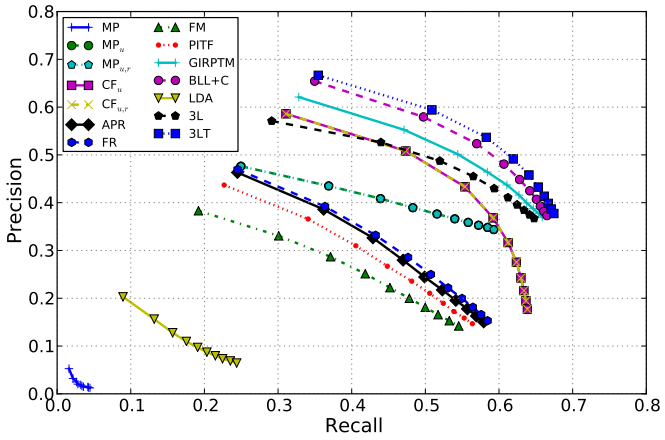
Evaluator. This component evaluates the algorithms based on a training/test set split of a dataset with respect to standard Information Retrieval (IR) metrics, such as Recall (R@k), Precision (P@), F1-score (F1@k), Mean Reciprocal Rank (MRR), Mean Average Precision (MAP), Normalized Discounted Cumulative Gain (nDCG) and User Coverage (UC) [2]. Moreover, the evaluation engine offers data post-processing to limit the evaluation to users with the minimum or maximum number of bookmarks.

3. RESULTS

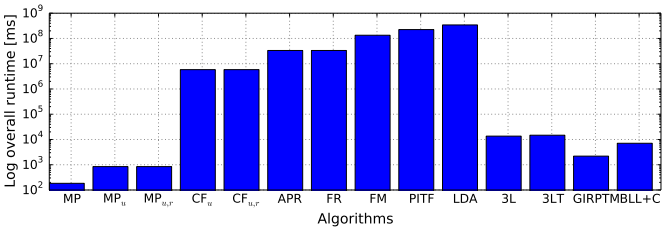
To show the functionalities of *TagRec*, we evaluated and compared a selection of the implemented algorithms in terms of recommender quality and runtime using a real-world folksonomy dataset gathered from the image sharing portal Flickr. The dataset contained 9,590 users, 864,679 resources, 127,599 tags 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 (3LT) and human memory (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.



(a) Recommender quality



(b) Recommender runtime

Figure 2: Evaluation results for the Flickr dataset showing the quality and overall runtime of the recommender algorithms.

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

[1] I. Cantador, A. Bellogín, and D. Vallet. Content-based recommendation in social tagging systems. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 237–240. ACM, 2010.

[2] R. Jäschke, F. Eisterlehner, A. Hotho, and G. Stumme. Testing and evaluating tag recommenders in a live system. In *Proceedings of the third ACM conference on Recommender systems*, pages 369–372. ACM, 2009.

[3] R. Jäschke, L. Marinho, A. Hotho, L. Schmidt-Thieme, and G. Stumme. Tag recommendations in social bookmarking systems. *Ai Communications*, 21(4):231–247, 2008.

[4] D. Kowald, P. Seitlinger, C. Trattner, and T. Ley. Forgetting the words but remembering the meaning: Modeling forgetting in a verbal and semantic tag recommender. *arXiv preprint arXiv:1402.0728*, 2014.

[5] D. Kowald, P. Seitlinger, C. Trattner, and T. Ley. Long time no see: The probability of reusing tags as a function of frequency and recency. In *Proc.*, WWW '14. ACM, 2014.

[6] R. Krestel and P. Fankhauser. Language models and topic models for personalizing tag recommendation. In *Proc.*, pages 82–89. IEEE, 2010.

[7] L. Marinho, A. Nanopoulos, L. Schmidt-Thieme, R. Jäschke, A. Hotho, G. Stumme, and P. Symeonidis. Social tagging recommender systems. In F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, editors, *Recommender Systems Handbook*, pages 615–644. Springer US, 2011.

[8] L. B. Marinho and L. Schmidt-Thieme. Collaborative tag recommendations. In *Data Analysis, Machine Learning and Applications*, pages 533–540. Springer, 2008.

[9] S. Rendle and L. Schmidt-Thieme. Pairwise interaction tensor factorization for personalized tag recommendation. In *Proc.*, WSDM '10. ACM, 2010.

[10] P. Seitlinger, D. Kowald, C. Trattner, and T. Ley. Recommending tags with a model of human categorization. In *Proc.*, CIKM '13, pages 2381–2386. ACM, 2013.

[11] L. Zhang, J. Tang, and M. Zhang. Integrating temporal usage pattern into personalized tag prediction. In *Web Technologies and Applications*, pages 354–365. Springer, 2012.