Evaluating Tag Recommender Algorithms in Real-World Folksonomies: A Comparative Study

Dominik Kowald and Elisabeth Lex
dkowald@know-center.at, elisabeth.lex@tugraz.at

Goal

The goal of this study is to provide researchers and developers of tag-based recommender systems with an **extensive, transparent and reproducible evaluation of state-of-the-art tag recommender algorithms in real-world folksonomies**.

Contributions

- **Algorithms:** Not only classic algorithms (e.g., CF, FolkRank or PITF) but also novel time-based and cognitive-inspired approaches.
- **Datasets**: Six unfiltered folksonomies to demonstrate the performance of the algorithms in real-world settings.
- **Metrics:** Wide range of evaluation metrics measuring not only the accuracy and ranking but also the diversity, novelty and computational costs (runtime and memory) of the approaches.
- **Framework:** Our study was conducted using the open-source tag recommender evaluation framework *TagRec*.

Datasets

Dataset	|U|	|R|	|P|	|P| / |R|
Flickr	9,590	856,755	856,755	1.000
CiteULike	18,474	811,175	900,794	1.110
BibSonomy	10,179	683,478	772,108	1.129
Delicious	15,980	963,741	1,447,267	1.501
LastFM	1,892	12,522	71,062	5.674
MovieLens	4,009	7,601	55,484	7.299

Table 1: Statistics (number of users, resources, posts and narrowness degree) of the **real-world folksonomy datasets**.

Winning Algorithm

BLL$_{ac}$+MP$_r$, inspired by the activation equation of the cognitive architecture ACT-R:

$$A_i = B_i + \sum_j (W_j \cdot S_{j,i}) \quad (1)$$

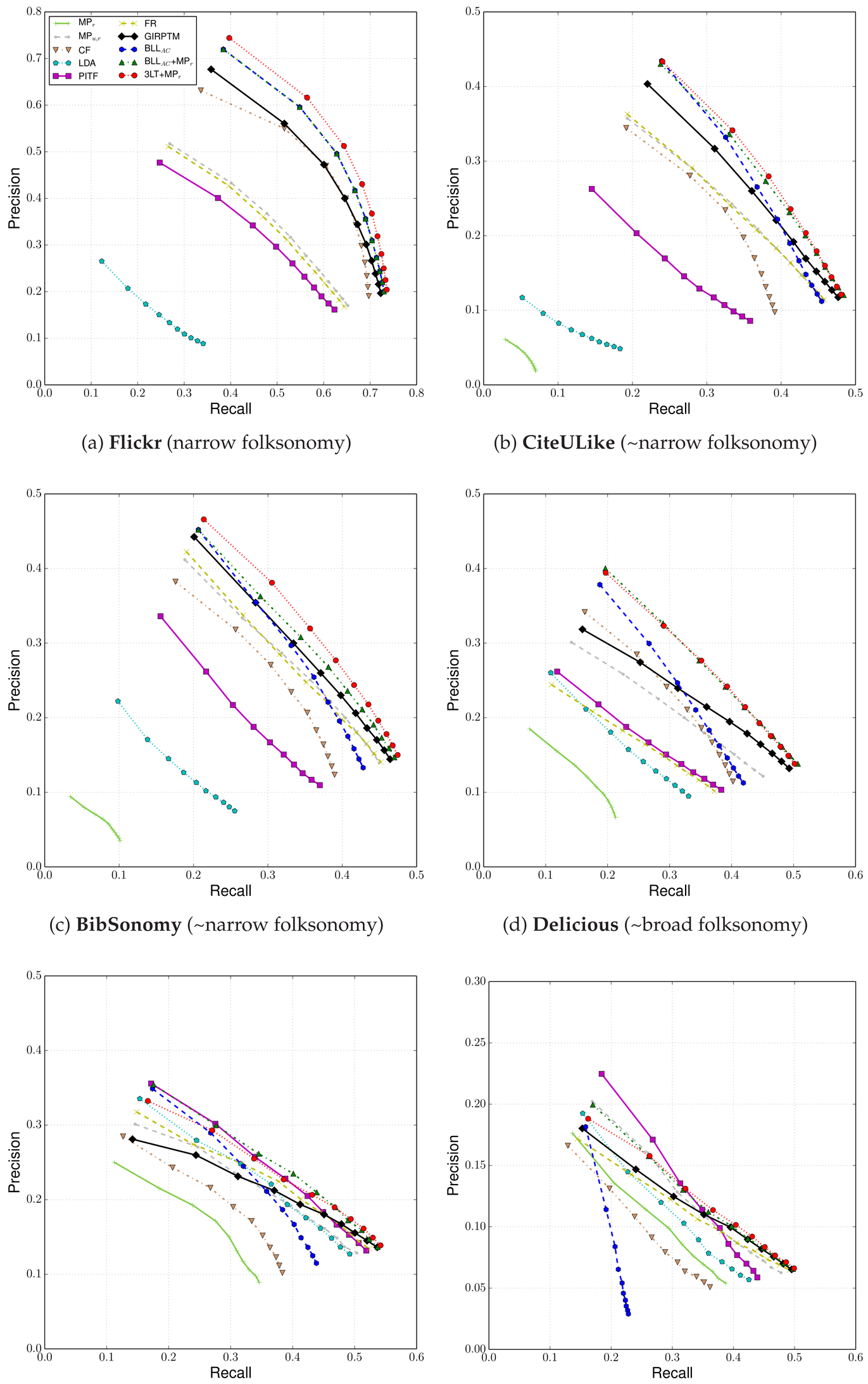
$$B_i = ln(\sum_{j=1}^{n} t_j^{-d}) \quad (2)$$

Evaluation Framework

[1] D. Kowald, E. Lacic, and C. Trattner. Tagrec: Towards a standardized tag recommender benchmarking framework. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media*, HT'14, NY, USA, 2014. ACM. **(best poster)**

https://github.com/learning-layers/TagRec/

Accuracy Results



(a) **Flickr** (narrow folksonomy)

(b) **CiteULike** (~narrow folksonomy)

(c) **BibSonomy** (~narrow folksonomy)

(d) **Delicious** (~broad folksonomy)

(e) **LastFM** (broad folksonomy)

(f) **MovieLens** (broad folksonomy)

Figure 1: **Recall / Precision plots** for k = 1 - 10 recommmended tags showing clear differences in algorithmic performance between narrow and broad folksonomies.

Summary

Algorithm	Name	Accuracy narrow	Accuracy broad	Diversity	Novelty	Runtime	Mem.
MP$_r$	Most popular tags by resource	-		-	-	++	+
MP$_{u,r}$	Most popular tags by user & resource					++	+
CF	User-based Collaborative Filtering			+			
LDA	Latent Dirichlet Allocation	-		++		-	-
PITF	Pairwise Interaction Tensor Factorization	-	+	++	+	-	++
FR	FolkRank		+	++			
GIRPTM	Temporal tag usage patterns and MP$_r$	+	+			++	+
BLL$_{ac}$	**BLL equation with context associations**	+		-	++	+	
BLL$_{ac}$+MP$_r$	**Mixture of BLL$_{ac}$ and MP$_r$**	++	++			+	
3LT+MP$_r$	**Time-dependent 3Layers and MP$_r$**	++	++			-	-

Table 2: Summary of the performance of the algorithms showing that providing helpful tag recommendations in real-world folksonomies **greatly depends on the given user needs**.