ECML PKDD Discovery Challenge 2009
Tag Recommendations for Social Bookmarking Systems

Folke Eisterlehner, Andreas Hotho, Robert Jäschke

Knowledge & Data Engineering Group
University of Kassel

sponsored by
Tag Recommendations for Social Bookmarking Systems

- Support users during the tagging process
- Recommend tags on the posting page
- http://www.kde.cs.uni-kassel.de/ws/dc09/
Last Year's Tag Recommendation Results

U_ (14925 posts)

all train tags known
1st
2nd
3rd
4th
5th
most popular tags
most popular tags by user

Recall

UR (1245 posts)

all train tags known
1st
2nd
3rd
4th
5th
most popular tags

Recall

_R (966 posts)

all train tags known
1st
2nd
3rd
4th
5th
most popular tags by resource

Recall

Number of recommended tags
Tasks

• Given a user and a resource, recommend up to five tags.

• Task 1: Content-Based Tag Recommendations
  ■ User, resource, or tags might be unknown

• Task 2: Graph-Based Recommendations
  ■ User, resource, and tags contained in training post-core

• Task 3: Online Tag Recommendations
  ■ Deliver live recommendations respecting timeouts, etc.

• 48h time for Tasks 1 & 2, Task 3 running for five weeks
Social bookmarking data from BibSonomy
http://www.bibsonomy.org/

• Training data
  ■ Released on March 25th, 2009
  ■ Posts from BibSonomy up to 31\textsuperscript{st} Dec. 2009
  ■ Complete snapshot: 3,617 users, 93,756 tags, 378,378 resources
  ■ Post-core at level 2: 1,185 users, 13,276 tags, 22,389 resources

• Test data
  ■ Released on July 6th, 2009
  ■ 6 months snapshot (2009-01-01 - 2009-06-30)
  ■ Task 1: 1,591 users, 34,045 tags, 40,729 resources, 43,002 posts
  ■ Task 2: 136 users, 862 tags, 667 resources, 778 posts
Dataset: Preprocessing, Identity, Post-core

- **Preprocessing**
  - Removed spam, posts from user *dblp*
  - Tag cleansing: only letters+numbers, common system tags removed

- **Identity of resources**
  - *Intrahash* + user name uniquely identifies a *post*
  - *Overlap* between resources by interhash

- **Post-core at level 2**
  - Each user, tag, resource appears in at least two posts
  - Iterative pruning process
  - Interhash to identify resources
Evaluation

- Test data = posts without tags
- Submitted data = recommended tags for the posts
- First 5 recommended tags compared against true tags from user (again, cleaning tags - only letters+numbers, no system tags)

- Precision and recall per post:

\[
\text{precision} = \frac{|\text{recommended} \cap \text{true}|}{|\text{recommended}|} \quad \text{recall} = \frac{|\text{recommended} \cap \text{true}|}{|\text{true}|}
\]

- Averaged over all posts

- Final criterion: F1-Measure = \( \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \)
Participants & Submissions

- 150 registered mailing list users (= access to training data)
- 21 result submissions for each of the Tasks 1 & 2
- 27 paper submissions - 24 accepted
## Results Task 1

<table>
<thead>
<tr>
<th>submission</th>
<th>f1m</th>
<th>team</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.18740</td>
<td><strong>Tag Sources for Recommendation in Collaborative Tagging Systems</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Marek Lipczak, Yeming Hu, Yael Kollet, and Evangelos Milios</td>
</tr>
<tr>
<td>2</td>
<td>0.18001</td>
<td><strong>Content- and Graph-based Tag Recommendation: Two Variations</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Johannes Mrosek, Stefan Bussmann, Hendrik Albers, Kai Posdziech,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Benedikt Hengefeld, Nils Opperman, Stefan Robert, and Gerrit Spira</td>
</tr>
<tr>
<td>3</td>
<td>0.17975</td>
<td><strong>A Weighting Scheme for Tag Recommendation in Social Bookmarking</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Systems</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sanghun Ju and Kyu-Baek Hwang</td>
</tr>
<tr>
<td>4</td>
<td>0.17661</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.15464</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.15412</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.14398</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.14199</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.14151</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.14018</td>
<td></td>
</tr>
</tbody>
</table>
Results Task 1

![Graph showing the relationship between the number of tags and a metric labeled "f1m".]
Results Task 2

<table>
<thead>
<tr>
<th>submission</th>
<th>f1m</th>
<th>team</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>79768</td>
<td><strong>Factor Models for Tag Recommendation in BibSonomy</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Steffen Rendle and Lars Schmidt-Thieme</td>
</tr>
<tr>
<td>2</td>
<td>80999</td>
<td><strong>Relational Classification for Personalized Tag Recommendation</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Leandro Balby Marinho, Christine Preisach, and Lars Schmidt-Thieme</td>
</tr>
<tr>
<td>3</td>
<td>30478</td>
<td><strong>Tag Sources for Recommendation in Collaborative Tagging Systems</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Marek Lipczak, Yeming Hu, Yael Kollet, and Evangelos Milios</td>
</tr>
<tr>
<td>4</td>
<td>76565</td>
<td><strong>A Collaborative Filtering Tag Recommendation System based on Graph</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yuan Zhang, Ning Zhang, and Jie Tang</td>
</tr>
<tr>
<td>5</td>
<td>82722</td>
<td>0.32134</td>
</tr>
<tr>
<td>6</td>
<td>61431</td>
<td>0.32124</td>
</tr>
<tr>
<td>7</td>
<td>45839</td>
<td>0.32039</td>
</tr>
<tr>
<td>8</td>
<td>56390</td>
<td>0.31396</td>
</tr>
<tr>
<td>9</td>
<td>16379</td>
<td>0.31368</td>
</tr>
<tr>
<td>10</td>
<td>11472</td>
<td>0.30751</td>
</tr>
</tbody>
</table>
Results Task 2
Lessons learnt

• Task 1 = mostly unknown data
  ■ Keep it simple
  ■ Gather tags from various sources (title, content, external services, different spellings)
  ■ Re-order tags for personalization
  ■ Clever selection and combination strategies win

• Task 2 = user, resource, tags known
  ■ Sophisticated methods are useable and useful
  ■ Factor models, classification, graph-based, CF, etc.
  ■ Simple methods work, too

Since Task 1 is the >90% case, use simple methods (they're quick, too!)
Thanks

• Conference organizers, in particular Tina Anzic

• Our sponsors

• YOU - for your vivid participance
Schedule

9:00 - 10:30
- Welcome & Introduction
- Tag Sources for Recommendation in Collaborative Tagging Systems - Marek Lipczak, Yeming Hu, Yael Kollet, and Evangelos Milios
- Factor Models for Tag Recommendation in BibSonomy - Steffen Rendle and Lars Schmidt-Thieme

10:30 - 11:00
- Coffee Break (+poster preparation)
11:00 - 12:30

- **Content- and Graph-based Tag Recommendation: Two Variations** - Johannes Mrosek, Stefan Bussmann, Hendrik Albers, Kai Posdziech, Benedikt Hengfeld, Nils Opperman, Stefan Robert, and Gerrit Spira [PDF]

- **Relational Classification for Personalized Tag Recommendation** - Leandro Balby Marinho, Christine Preisach, and Lars Schmidt-Thieme [PDF]

- **A Weighting Scheme for Tag Recommendation in Social Bookmarking Systems** - Sanghun Ju and Kyu-Baek Hwang [PDF]

12:30 - 14:00

- Lunch
Schedule

14:00 - 15:20
- A Collaborative Filtering Tag Recommendation System based on Graph - Yuan Zhang, Ning Zhang, and Jie Tang  PDF
- **Minute Madness**: Every poster presenter has one minute to convince everybody to look at his/her poster.
- Closing Session

15:20 - 15:40
- Coffee Break (+ poster preparation)

15:40 - 17:00
- Poster Session

17:20 - 18:00
- Conference Opening & Awards