Learning Layers
Scaling up Technologies for Informal Learning in SME Clusters

The Influence of Frequency, Recency and Semantic Context on the Reuse of Tags in Social Tagging Systems

Hypertext 2016
Halifax, Nova Scotia, Canada

Dominik Kowald, Elisabeth Lex
Know-Center, Graz University of Technology (Austria)

http://Learning-Layers.eu – Scaling up Technologies for Informal Learning in SME Clusters – layers@learning-layers.eu
Social Tagging

- Social tagging is the process of collaboratively **annotating** content
- Essential instrument of Web 2.0 to **structure and search** Web content

**Issues**
- No rules for tags → can be **freely chosen**
- Hard for people to come up with a set of descriptive/relevant tags **by their own**
- People are **lazy** in applying tags
- **Language**: Synonyms, spelling errors, singular/plural ...

[http://blog.zubiaga.org/2009/02/what-are-social-tags/]
Solution: Tag Recommendations


http://Learning-Layers.eu — Scaling up Technologies for Informal Learning in SME Clusters – layers@learning-layers.eu
Previous Work: Cognitive-Inspired Tag Recommendations

• **Activation equation** of the cognitive architecture ACT-R [Anderson et al, 2004]

\[
A_i = B_i + \sum_j (W_j \cdot S_{j,i})
\]

• Activation of memory unit i (i.e., **tag**) =
  • Base-level activation (**general usefulness**: tag frequency and recency in the past via power function [Anderson et al., 1991])
  • Associative activation (**usefullness** in the current **semantic context**: similarity with resource tags)

• Evaluation results showed that this approach **outperforms other state-of-the-art methods** (e.g., [Kowald et al., 2014] @ WWW; [Trattner et al., 2016] @ Journal of Web Science)
Present Work: Factors that Influence Tag Reuse

• **RQ1**
  – How are the factors of frequency, recency and semantic context influencing a tag's probability of being reused in social tagging systems?
    → Empirical study

• **RQ2**
  – Can the factors of frequency, recency and semantic context be exploited to efficiently predict a user's tag reuse given a specific folksonomy type?
    → Prediction study
Datasets

- Six social tagging datasets from different domains and of different folksonomy types

| Dataset    | $|U|$  | $|R|$  | $|T|$  | $|P|$  | $|P|/|R|$ |
|------------|------|------|------|------|--------|
| Flickr     | 9,590| 856,755| 125,119| 856,755| 1.000  |
| CiteULike  | 18,474| 811,175| 273,883| 900,794| 1.110  |
| BibSonomy  | 10,179| 683,478| 201,254| 772,108| 1.129  |
| Delicious  | 15,980| 963,741| 184,012| 1,447,267| 1.501  |
| LastFM     | 1,892| 12,522| 9,748| 71,062| 5.674  |
| MovieLens  | 4,009| 7,601| 15,238| 55,484| 7.299  |

- Train / test sets
  - For each user → most recent post in test set
    - Rest is used for training (reflecting the past)
    - Test set defines tag reuse (reflecting the future)
The more frequently a tag was used in the past \((k > 0)\), the higher its reuse probability is.

The more recently a tag was used in the past \((k < 0)\), the higher its reuse probability is.

The more similar a tag is to tags in the current semantic context \((k > 0)\), the higher its reuse probability is.

→ All three factors are important
## Results (RQ2)

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Metric</th>
<th>Individual factors</th>
<th>Combination</th>
<th>Social</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Frequency</td>
<td>Recency</td>
<td>SemCon</td>
</tr>
<tr>
<td>Flickr</td>
<td>(F_1@5)</td>
<td>.371</td>
<td>.464</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>nDCG@10</td>
<td>.569</td>
<td>.702</td>
<td>-</td>
</tr>
<tr>
<td>CiteULike</td>
<td>(F_1@5)</td>
<td>.231</td>
<td>.236</td>
<td>.041</td>
</tr>
<tr>
<td></td>
<td>nDCG@10</td>
<td>.367</td>
<td>.385</td>
<td>.069</td>
</tr>
<tr>
<td>BibSonomy</td>
<td>(F_1@5)</td>
<td>.253</td>
<td>.252</td>
<td>.063</td>
</tr>
<tr>
<td></td>
<td>nDCG@10</td>
<td>.371</td>
<td>.368</td>
<td>.090</td>
</tr>
<tr>
<td>Delicious</td>
<td>(F_1@5)</td>
<td>.173</td>
<td>.179</td>
<td>.108</td>
</tr>
<tr>
<td></td>
<td>nDCG@10</td>
<td>.267</td>
<td>.287</td>
<td>.158</td>
</tr>
<tr>
<td>LastFM</td>
<td>(F_1@5)</td>
<td>.193</td>
<td>.189</td>
<td>.202</td>
</tr>
<tr>
<td></td>
<td>nDCG@10</td>
<td>.292</td>
<td>.293</td>
<td>.302</td>
</tr>
<tr>
<td>MovieLens</td>
<td>(F_1@5)</td>
<td>.077</td>
<td>.076</td>
<td>.077</td>
</tr>
<tr>
<td></td>
<td>nDCG@10</td>
<td>.177</td>
<td>.183</td>
<td>.176</td>
</tr>
</tbody>
</table>

### Folksonomy type

<table>
<thead>
<tr>
<th>Folksonomy type</th>
<th>Frequency</th>
<th>Recency</th>
<th>SemCon</th>
<th>Comb</th>
<th>Social</th>
</tr>
</thead>
<tbody>
<tr>
<td>Narrow</td>
<td>+/-</td>
<td>+</td>
<td>-</td>
<td>+/-</td>
<td>-</td>
</tr>
<tr>
<td>Mixed</td>
<td>+</td>
<td>+</td>
<td>+/-</td>
<td>+</td>
<td>+/-</td>
</tr>
<tr>
<td>Broad</td>
<td>+/-</td>
<td>+/-</td>
<td>+</td>
<td>+/-</td>
<td>+</td>
</tr>
</tbody>
</table>
Conclusion & Future Work

- **RQ1**: All three factors influence tag reuse
  - $k > 0$ for frequency and semantic context
  - $k < 0$ for recency

- **RQ2**: Prediction accuracy depends on folksonomy type
  - Recency is most important in the narrow case
  - Combination of the factors works best in the mixed case
  - Social influence become better, the broader the folksonomy is

- **Future Work**
  - Analyze social influence (tag „imitation“)
  - Extent semantic context (e.g., resource title or content)
  - Apply findings to hashtags (e.g., Twitter, Facebook, Instagram)
References

Thank you for your attention!
Questions?

TagRec: open-source recommender evaluation framework
https://github.com/learning-layers/TagRec/

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