Tag Meaning Disambiguation
through Analysis of Tripartite Structure of Folksonomies

Ching-man Au Yeung, Nicholas Gibbins, Nigel Shadbolt
Overview

- Background
- Motivations
- Tripartite structure of folksonomies
- Tag meaning disambiguation
- Experiments
- Conclusions and future work
• **Collaborative tagging systems** and **folksonomies**
Examples of collaborative tagging systems

- del.icio.us
- b.hatena.ne.jp

Understanding the Semantics of Ambiguous Tags in Folksonomies - C.M. Au Yeung, N. Gibbins, N. Shadbolt
**Background**

- **Advantages** [Adam 2004, Wu et al. 2006]
  - Freedom and flexibility
  - Quick adaptation to changes in vocabulary (e.g. ajax, youtube)
  - Convenience and serendipity

- **Disadvantages** [Adam 2004, Wu et al. 2006]
  - Ambiguity (e.g. apple, sf, opera)
  - Lack of format (e.g. how multiword tags are handled)
  - Existence of synonyms (e.g. semweb, semanticweb, semantic_web)
  - Lack of semantics
**Motivations**

- Many tags are ambiguous (possess multiple meanings)
- This affects the precision of retrieval and annotation of shared resources
- Current research works mainly focus on clustering of tags
- Few works deal with ambiguous tags, and in indirect ways only (e.g. [Wu et al. 2006])
Folksonomy (A hypergraph)

\[ F = \langle U, T, D, A \rangle; \ A \subseteq U \times T \times D \]

**Bipartite graph** \( UD_t \)

\[ UD_t = \langle U \cup D, E_{UD} \rangle \]

\[ E_{UD} = \{ \{u,d\} | \{u,t,d\} \in A \} \]

adjacency matrix multiplication

A weighted network of users

edge weight = # of users tagged the documents

A weighted network of documents

edge weight = # of documents tagged
A case study

- *sf* in *del.ici.ous* [Au Yeung et al. 2007]
Tag Meaning Disambiguation

• Basic ideas
  • Different clusters of nodes in the network correspond to different meanings of the tag
  • Different meanings of ambiguous tags can be obtained by partitioning the network into communities of nodes
  • The meanings can be understood by examining the most frequently used tags within a cluster

• Algorithms for discovering communities in a network
  • Modularity optimization by removing edges based on edge betweenness [Newman & Girvan 2004]
  • Modularity: a measure of the “goodness” of a partition of a network
  • Edge betweenness: a measure of how likely an edge is a bridge between two communities
Tag Meaning Disambiguation


describing the edge betweenness of edge \( e \) = number of shortest path running through \( e \)

(most likely to be a bridge between two communities in the network)
• **Proposed method**

1. Collect tagging data of the tag to be disambiguated (including documents with the tag, users and other tags involved)

2. Construct a *document network* out of the data

3. Apply the *community-discovering algorithm* to the network

4. For each community discovered, extract the *10 most frequently used tags* among those documents

5. The sets of tags should give different meanings of the tag being examined
Experiments

Disambiguation of the tag “sf”

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Tags</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>sf, scifi, fiction, books, sci-fi, writing, literature, science, sciencefiction, fantasy</td>
</tr>
<tr>
<td>2</td>
<td>sf, sanfrancisco, bayarea, san, francisco, california, travel, events, art, san_franisco</td>
</tr>
<tr>
<td>3</td>
<td>sf, sanfrancisco, design, bayarea, blog, food, todo, california, shopping, san</td>
</tr>
</tbody>
</table>
Experiments

Disambiguation of the tag “opera”

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Tags</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>opera, browser, web, software, javascript, browsers, tips, tools, internet, firefox</td>
</tr>
<tr>
<td>2</td>
<td>opera, shopping, imported, shop, design, store, home, inspiration, work, personal</td>
</tr>
<tr>
<td>3</td>
<td>opera, music, musique, classical, art, culture, musica, musica, classic, travel</td>
</tr>
</tbody>
</table>
Disambiguation of the tag “cambridge”
Disambiguation of the tag “tube”
Discussions

- Different meanings of a tag can be obtained from the result

- However, some problems exist:
  
  1. The meaning of a tag in some clusters is unclear
     (as in the results of *opera, cambridge*)

  2. Some clusters correspond to the same meaning of a tag
     (as in the result of *tube*)

  3. Some clusters correspond to the same meaning of a tag, but
     referring to such meaning in possibly different contexts
     (as in the result of *sf*)
Conclusions and future work

• **Conclusions**

  • The method is effective in clustering documents of different topic and discover the different meanings of a tag
  
  • Some post-processing may be required to clean up redundant or unclear clusters

• **Future work**

  • To improve the efficiency of the method by employing faster algorithms (e.g. [Clauset et al. 2004])
  
  • Investigate how the results can be refined to produce more useful disambiguation
  
  • Perform Larger scale of evaluation of the proposed method


