Collective User Behaviour and Tag Contextualisation in Folksonomies

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

Intelligence, Agents, Multimedia Group
School of Electronics and Computer Science
University of Southampton
Introduction

Delicious

BibSonomy

WI’08 – Collective User Behaviour and Tag Contextualisation in Folksonomies
Introduction

A Tag Cloud in Delicious

An example of collaborative tagging

WI’08 – Collective User Behaviour and Tag Contextualisation in Folksonomies
The Problem of Ambiguity

♦ A tag can be used to refer to different concepts

♦ Understanding the meaning of a tag is context-dependent

♦ Limits the effectiveness of collaborative tagging and causes problems when retrieving documents

♦ Example:
  * sf → *San Francisco*, or
  * sf → *Science Fiction*?
Background and Motivations
Contextualisation of Tags

- Putting tags in context by studying their relations with other tags, users and documents
- Limitation of large scale clustering of tags:

```
sf
  ↓
science
  ↓
bayarea
  ↓
california
  ↓
  ↓
travel
  ↓
  ↓
reading
  ↓
fiction
```
Contextualisation of Tags

- Instead, we look at a particular tag, construct networks based on the users’ behaviour.

An edge exists if some users assigned the tag to both A and B.

(Compare to the case in which they share some tags)
Preliminary Results

Network of bookmarks with edges weighted by number of users using the tag *sf* on the bookmarks.
(Data obtained from Delicious)

Related tags:
- bayarea
- california
- us
- san
- francisco

San Francisco related bookmarks

Science Fiction related bookmarks

Related tags:
- science
- fiction
- fantasy
- reading
- books
Preliminary Results

The same network with edges weighted less than 2 removed.
Preliminary Results

Summary

♦ Documents can be clustered by considering only the pattern of how the users use a particular tag

♦ Users tend to be consistent when using a tag (they tend not use it ambiguously)

♦ Modelling folksonomy data in this way gives insight into how the tags are actually used
### Preliminary Results

#### More Examples

<table>
<thead>
<tr>
<th>Tag</th>
<th>Context</th>
<th>Tags Extracted</th>
</tr>
</thead>
<tbody>
<tr>
<td>tube</td>
<td>YouTube</td>
<td>tube, youtube, video, funny, videos, fun, cool, music, feel.good</td>
</tr>
<tr>
<td></td>
<td>Vacuum Tubes</td>
<td>tube, audio, electronics, diy, amplifier, amp, tubes, music, elect</td>
</tr>
<tr>
<td></td>
<td>Underground</td>
<td>tube, london, underground, travel, transport, maps, map, uk</td>
</tr>
<tr>
<td>bridge</td>
<td>Design Pattern</td>
<td>bridge, programming, development, library, code, ruby, tools</td>
</tr>
<tr>
<td></td>
<td>Card Game</td>
<td>bridge, games, cards, game, imported, howto, conventions</td>
</tr>
<tr>
<td></td>
<td>Networking</td>
<td>bridge, networking, linux, network, howto, software, sysadmin</td>
</tr>
<tr>
<td></td>
<td>Architecture</td>
<td>bridge, bridges, structures, engineering, science, physics</td>
</tr>
<tr>
<td>wine</td>
<td>Linux Software</td>
<td>wine, linux, ubuntu, howto, windows, software, tutorial, emulation</td>
</tr>
<tr>
<td></td>
<td>Beverage</td>
<td>wine, food, shopping, drink, reference, vino, cooking, alcohol, blog</td>
</tr>
</tbody>
</table>
Method of Evaluation

Research Questions

1. Does the social context provide better information for understanding the semantics of tags?

2. Which method of modelling the data in a folksonomy best represent the associations between documents/users/tags?

3. How can we represent the different contexts extracted?
Different Kinds of Networks

- **User-based tag networks / document networks** (Social context considered)
- **Tag-based user networks / document networks** (Similar to the vector model in IR)
- **Tag similarity networks** (e.g. Calculate similarity of the contexts of two tags)
Method of Evaluation

Data Preparation

- We collected data of 50 tags (associated bookmarks, users and tags) from Delicious
- Recruited users to manually classify bookmarks into different contexts
- Compare the output of clustering process done on different networks
- Performance measures: precision, recall, coverage, etc.
Method of Evaluation

Other Possible Evaluation Methods

♦ Comparing with document clustering techniques e.g. Clusty

♦ Comparing the results with WordNet or other thesaurus
Applications

**Web Search Result Classification**

- We applied the result on classifying Web search results returned by search engines.

- New or domain-specific meanings of a keyword can be identified (e.g. tube, bridge).

- Paper: *A k-Nearest-Neighbour Method for Classifying Web Search Results with Data in Folksonomies* (Session 10-A-WI-4, 10 Dec, 10:45)
Applications

Submit query term → Delicious

Google

Get a list of Web documents

Example:
D1: Wikipedia - Science Fiction
D2: Wikipedia - San Francisco
D3: ...

Construct sets of keywords

Example:
D1: {scifi, fiction, books, ...}
D2: {california, bayarea, ...}
D3: {...}

Get documents associated with the query term

Example:
C1: {scifi, fiction, books, ...}
C2: {california, bayarea, ...}

Build Classifier

Search Result Classification

Classifying

Output Result

Example:
D1 ➔ C1
D2 ➔ C2
Conclusions

♦ Outlined our research on tag contextualisation based on collective user behaviour

♦ Social contexts should be considered when trying to understand how tags are used

♦ Results can be applied to applications such as enhancing Web search or recommendation systems
Thank You!

Albert Au Yeung
cmay06r@ecs.soton.ac.uk
http://users.ecs.soton.ac.uk/cmay06r/