Telling Experts from Spammers
Expertise Ranking in Folksonomies

Michael G. Noll
Christoph Meinel

(Albert) Ching-Man Au Yeung
Nicholas Gibbins
Nigel Shadbolt

Hasso Plattner Institute
Uni Southampton
Opening joke
(under construction)
About Michael


- Bi-national Ph.D. candidate in Computer Science at the Hasso Plattner Institute in Potsdam, Germany, and the University of Luxembourg
- Working as external doctoral student at the satellite operator SES ASTRA (Luxembourg) in the industrial R&D project “Safer Internet”
- Thesis title:
  Understanding and Leveraging the Social Web for Information Retrieval
About Albert

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

- PhD candidate in Computer Science at the University of Southampton
- Previously obtained BEng (Information Engineering) and MPhil (Computer Science) from the Chinese University of Hong Kong
- Thesis title: *From User Behaviours to Collective Semantics*
  Study how implicit semantics can be harvested from social interactions on the Web, focusing on collaborative tagging as a prominent example
Why do we work together?

- Our common interest: ranking users according to their expertise
- In Web IR, we want to identify the experts so that we can get the best resources to satisfy our information needs
- Expertise/Trustworthiness of users is a kind of implicit quality of users that can be determined by analyzing collective user behavior
- It’s FUN! 😊
Introduction
Folksonomies and Collaborative Tagging

- Large and still increasing popularity in the WWW today

- Idea: Freely annotating resources with keywords aka “tags”
- Result: bottom-up “categorization” by end users, aka “folksonomy”
- Used for organizing resources, sharing, self-promotion, …
- Additional effect: **new means** of resource/information retrieval and discovery
Background

Web page

User bookmark

Example: Web page bookmarked by **754 users**, first bookmark from **09/2007**
Motivation

Two related goals for our work on expertise in folksonomies:

1. **Identifying and promoting experts for a given topic**
   Weighting user input, giving (better) recommendations, identify trendsetters for marketing/advertising/product promotion, etc.

   \[ \textit{Topic} := \textit{conjunction or disjunction of one or more tags} \]

2. **Demoting spammers**
   Reduce impact of spam and junk input thereby improving system quality, performance, operation
Models
What makes an expert an expert?

Postulation of two assumptions of expertise for resource discovery, grounded on literature from computer science (that’s you) and psychology.

1. Mutual reinforcement of user expertise and document quality
   Expert users tend to have many high quality documents, and high quality documents are tagged by users of high expertise.

2. Discoverers vs. followers
   Expert users are discoverers – they tend to be the first to bookmark and tag high quality documents, thereby bringing them to the attention of the user community. Think: researchers in academia.
Model of expert users

user network (social graph)

user network (social graph)

user

tags

timestamp

document network (Web graph)

document network (Web graph)

doc

doc

doc

doc

Context of social bookmarking / collaborative tagging
Model of expert users

user network
(social graph)

user

tags

timestamp

Our Focus

document network
(Web graph)

Telling Experts from Spammers | Michael G. Noll & Ching-man Au Yeung | SIGIR 2009
Model of expert users

Bookmarking history of a Web page

Web page

Timeline

Users
Model of expert users

Bookmarking history of a Web page

Discoverers

Followers

“john.smith”

December 2009
January
February
March
April
May
Credit score function $C(t)$ → earlier discovery = more credit

Model of expert users
SPEAR Algorithm
Proposed algorithm: SPEAR

SPEAR – SPamming-resistant Expertise Analysis and Ranking

- Based on the HITS (Hypertext Induced Topic Search) algorithm
  - Hubs: pages that points to good pages
  - Authorities: pages that are pointed to by good pages

- Expertise and Quality (SPEAR) similar to Hub and Authority (HITS)
  - Users are hubs – we find useful pages through them
  - Pages are authorities – provide relevant information

- Difference: only users can point (link) to pages but not vice versa
Proposed algorithm: SPEAR

HITS / WWW

SPEAR / Folksonomy
Proposed algorithm: SPEAR

Input

Number of users \( M \)
Number of pages \( N \)
Set of taggings \( R_{\text{tag}} = \{ (\text{user}, \text{page}, \text{tag}, \text{timestamp}) \mid \text{tag} = \text{tag} \} \)
Credit score function \( C() \)
Number of iterations \( k \)

Output:

 Ranked list \( L \) of users by expertise in topic \( \text{tag} \)

Algorithm:

Set \( E \) to be the vector \((1, 1, \ldots, 1) \in Q^M \)
Set \( Q \) to be the vector \((1, 1, \ldots, 1) \in Q^N \)
\( A \leftarrow \text{Generate}_{-}\text{Adjacency}_{-}\text{Matrix}(R_{\text{tag}}, C) \)
\( \text{for } i = 1 \text{ to } k \text{ do} \)
\( \quad E \leftarrow Q \times A^T \)
\( \quad Q \leftarrow E \times A \)
\( \quad \text{Normalize } E \)
\( \quad \text{Normalize } Q \)
\( \text{endfor} \)
\( L \leftarrow \text{Sort users by their expertise score in } E \)
return \( L \)
Proposed algorithm: SPEAR

Folksonomy (simplified)

Adjacency matrix, credits applied

<table>
<thead>
<tr>
<th></th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
</tr>
</thead>
<tbody>
<tr>
<td>U1</td>
<td>1.4</td>
<td>1.7</td>
<td>0.0</td>
</tr>
<tr>
<td>U2</td>
<td>1.0</td>
<td>1.4</td>
<td>0.0</td>
</tr>
<tr>
<td>U3</td>
<td>0.0</td>
<td>1.0</td>
<td>1.4</td>
</tr>
<tr>
<td>U4</td>
<td>0.0</td>
<td>0.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Ranked list of users by expertise

<table>
<thead>
<tr>
<th></th>
<th>Rank</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>U1</td>
<td>1</td>
<td>0.422</td>
</tr>
<tr>
<td>U2</td>
<td>2</td>
<td>0.328</td>
</tr>
<tr>
<td>U3</td>
<td>3</td>
<td>0.212</td>
</tr>
<tr>
<td>U4</td>
<td>4</td>
<td>0.038</td>
</tr>
</tbody>
</table>
Evaluation
Experimental Setup

- Problem: lack of a proper ground truth for expertise
- “Who is the best researcher in this room?” 😊
- Workaround: Inserting simulated users into real-world data from Delicious.com and check where they end up after ranking

- Real-world data set from Delicious.com comprising 50 tags with
  - 515,000 real users (and real spammers)
  - 71,300 real Web pages
  - 2,190,000 real social bookmarks
Evaluation

Experimental Setup

- **Probabilistic simulation**, simulated users generated with four parameters
  - **P1**: Number of user’s bookmarks – active or inactive user?
  - **P2**: Newness – fraction of Web pages not already in data set
  - **P3**: Time preference – discoverer or follower?
  - **P4**: Document preference – high quality or low quality?
Experimental Setup

- Simulation of 6 different user types
  Profiles (parameter values) based on recent studies + characteristics of our real-world data sets

- Experts
  - **Geek** – lots of high quality documents, discoverer
  - **Veteran** – high quality documents, discoverer
  - **Newcomer** – high quality documents, follower

- Spammers
  - **Flooder** – lots of random documents, follower
  - **Promoter** – some documents (most are his own), discoverer
  - **Trojan** – some documents, follower [next-gen spammer]
Performance baselines

- **FREQ(UENCY)**
  “Most popular” approach – simple frequency count, looks only at quantity. Seems to be the dominant algorithm in use in practice.

- **HITS**
  Algorithm on which SPEAR is based. Uses mutual reinforcement but does not analyze temporal dimension of user activity.

- In comparison: **SPEAR**
  Uses mutual reinforcement and exploits trusted temporal data for implementing the discoverer-follower scheme.
Experimental Results
Experts
Experts

- Only SPEAR distinguished geeks, veterans and newcomers
- FREQ and HITS clumped all expert-type users together
A closer look
Experts: “Ideal” result

Overlaps expected due to probabilistic simulation setup
Evaluation

Experimental Results – Promoting Experts

- SPEAR differentiated all expert types better than its competitors
- SPEAR kept expected order of “geeks > veterans > newcomers”
- SPEAR was less dependent on user activity (quality before quantity)
Qualitative analysis: manual examination of Top 10 experts for three tags “photography”, “semanticweb”, “javascript ∩ programming”

- No spammers found (…phew…)
- These users seemed to be more involved or “serious” about their Delicious usage, e.g. provided optional profile information such as real name, links to their Flickr photos or microblog on Twitter
- Their number of bookmarks: from 100’s to 10,000’s
- “semanticweb”: Semantic Web researcher among the experts
- “javascript ∩ programming”: Top 2 experts were professional software developers
Spammers
Evaluation

Spammers

- SPEAR consistently outperformed FREQ and HITS
- SPEAR was the only algorithm to handle trojans (tricky spammers)
A closer look
Spammers: “Ideal” result

Trojans expected to score higher because they mimic regular users for most of the time
Evaluation

Experimental Results – Demoting Spammers

- SPEAR demoted all spammer types significantly more than its competitors
- Only SPEAR demoted all trojans from the TOP 100 ranks
- FREQ completely failed to demote spammers
Evaluation

Qualitative analysis: manual examination of Top 50 users for the heavily spammed tag “mortgage” (without inserting simulated users)

- Ranked users by their number of bookmarks = FREQ strategy
- 30 out of 50 were (real) spammers, either flooders or promoters
- Compared to FREQ, both SPEAR and HITS were able to remove these spammers from the Top 50
- SPEAR demoted spammers significantly more than HITS
Discussion

SPEAR...

- demoted all spammer types while still ranking experts on top
- was much less vulnerable to spammers with its reduced dependence on the activeness of the users: quality >> quantity
- *increased* difficulty for spammers to game a collaborative tagging system
Preliminary study:
SPEAR and PageRank
SPEAR and PageRank

- Second SPEAR outcome: **document quality score**
- Relationship to other document quality / popularity measures?

Questions

- “Correlation between SPEAR (folksonomy) and PageRank (Web graph)?”
- “Are documents ranked high by SPEAR also ranked high by PageRank?”
SPEAR and PageRank

Experiments

- We built three data sets and compared PageRank distributions:
  - $ALL = \text{all documents from our n real-world data sets}$
  - $SPEAR-\text{TOP} = \text{joint set of SPEAR Top 100 docs of all n data sets}$
  - $SPEAR-\text{BOTTOM} = \text{joint set of SPEAR Bottom 100 docs …}$
Documents ranked higher (lower) by SPEAR tend to have higher (lower) PageRanks

Mean Pearson-$r$ correlation coefficient averaged over all data sets: $r = +0.324$
Still, SPEAR generally behaves quite different from PageRank!

- Best PR0 document [SPEAR #194] > best PR10 document [SPEAR #250]:
  PR0 document redirects to a PR8 document (homepage of TV show “The View”)
Summary
Conclusions

- Described a model of expertise in folksonomies for resource discovery
- Proposed an expertise ranking algorithm that is resistant to spammers
- Demonstrated how simulation techniques can be used for evaluation

Future Work

- Quality score of Web pages deserve more investigation
- Transfer to new problem domains, e.g. blogosphere or music
- Follow-up with user & item recommendation, trend detection
Backup Slides
Motivation

Resource retrieval ~ information retrieval

- Two types of resource discovery in collaborative tagging:
  1. Following the **tags**: subscribe or search tags to find relevant resources
  2. Following the **users**: subscribe to user feeds and receive notifications

- Following **expert users** provides more benefits
  - Should know the best resources with respect to a given topic
  - Should be quick in discovering and identifying new resources
Example: The user network of Joshua Schachter, founder of Delicious.com

Joshua “follows” these users and their activity