Reasonable Tag-Based Collaborative Filtering - for - Social Tagging Systems

WICOW2008
October 30, 2008

Reyn Nakamoto, kizasi Company, Inc.
Shinsuke Nakajima, Kyoto Sangyo University
Jun Miyazaki, Nara Institute of Science and Technology
Shinsuke Uemura, Nara Sangyo University
Hirokazu Kato, Nara Institute of Science and Technology
Youichi Inagaki, kizasi Company Inc.

Outline

• Background
• Reasonable Tag-Based Collaborative Filtering
• Live-Updating Website Recommendation System Using RCF
• Experiment & Results
• Conclusions and Future Work

2

Information Explosion

• Amount of information available on the Internet is growing exponentially
• Blogs, Reviews, Message Boards, etc.
• Finding relevant and credible information is increasingly difficult

Need for new, personalized information recommendation methods

3

Background

Information Explosion

• Amount of information available on the Internet is growing exponentially
• Blogs, Reviews, Message Boards, etc.
• Finding relevant and credible information is increasingly difficult

Need for new, personalized information recommendation methods

4

Collaborative Filtering (CF)

• Using people to collaborate and filter out and find the relevant information
• Personalized recommendation method based upon matching users’ preferences
• Effective with subjective information recommendation
  + Credible recommendations through group knowledge
  + Credible recommendations through depth of relationship
  - Does not consider why a user likes something

5

Social Tagging Systems

• Tags are natural language keywords that users attach to items
• Subject, Reason for Liking, Defining Characteristic, etc.
• Easier organization & retrieval by tag searching
• Discover new resources from other users

+ Credible recommendations through social filtering/group evaluation
  + Tags provide clues as to why a user liked an item

6
Reasonable Tag-Based Collaborative Filtering
Proposed Algorithm

Reasonable Tag-Based Collaborative Filtering (RCF)
• Personalized information recommendation algorithm which combines CF and Social Tagging Systems
• CF provides relevant and credible recommendations based upon matching users’ preferences
• Social Tagging provides:
  • Relevant, credible results through social filtering
  • Clues to the reasons why users liked some item
  • Based upon social bookmarking systems

RCF uses the power of the community and people to discover the relevant and credible information

Base System Overview
• Bookmark websites using tags
• Search for bookmarks by tags
• Discover other users’ bookmarks through tags

Social Website Bookmarking System
Features
• Bookmark websites using tags
• Search for bookmarks by tags
• Discover other users’ bookmarks through tags

Website Recommendations using RCF
Recommendation Process

User Perspective

1. User Similarity Calculation
   - Find users who share the same interest for same reason
2. Score Prediction Calculation
   - Find recommendation candidate websites from similar users that match the common interest
3. Current Topic Recommendation
   - Show the recommendation candidates that match the user’s current interest
1. User Similarity Calculation
Find users who share the same interest for same reason

1. Find users with common bookmarks with B

User B and C’s user similarity is high due to similar tags
User B and A’s user similarity is low due to dissimilar tags

2. Score Prediction Calculation
Find recommendation candidate websites from similar users that match the common interest

1. Get all similar users of user B and get all their bookmarks

"MacRumors" gets a high score prediction and becomes a recommendation candidate
"ATP Tennis" does not

3. Current Topic Recommendation
Show the recommendation candidates that match the user’s current interest

Website 3 gets recommended due to similar topic to current
Experiment and Results

Experiment Goals
- Goals
  - Determine feasibility of RCF with users
  - Test the effectiveness of RCF in providing relevant, credible website recommendations
  - Determine the most effective way of implicit recommendations
- Tested Methods
  - topic - Topic Vector Similarity
  - bookmark-count (tbc) - Total Bookmark Count
  - cf-topic - Standard CF with Topic Filtering
  - RCF - Proposed Method

Experiment Data
- Mined del.icio.us for three months
- 100,000 users
- 2.5 million webpages
- 3.6 million bookmarks
- 870,000 distinct tags

Experiment Procedure
1. Create User Profile
   - User bookmarks 20+ pages using tags
2. Evaluate Recommendations
   - User selects fifteen or more webpages and browse them.
   - Upon browsing a website, the system would randomly select one of the above described algorithms and implicitly generate up to six recommendations
   - User clicks on a recommendation if it is interesting
   - User rates websites as ‘Good’, ‘Fair’, or ‘Bad’
     - Relevant to user preferences
     - Related to their current interest
3. Test Users

Results

How Many Bookmarks are Necessary?
User’s Number of Bookmarks vs. Percentage of Users with Recommendation Candidates

- [Graph showing the relationship between the number of bookmarks and the percentage of users with recommendation candidates]
Are there enough recommendations?

<table>
<thead>
<tr>
<th></th>
<th>total # requests</th>
<th>total # recom.</th>
<th>recom/ request</th>
<th>% requests with recom.</th>
</tr>
</thead>
<tbody>
<tr>
<td>topic</td>
<td>173</td>
<td>645</td>
<td>3.73</td>
<td>62.1%</td>
</tr>
<tr>
<td>tbc</td>
<td>172</td>
<td>641</td>
<td>3.73</td>
<td>62.1%</td>
</tr>
<tr>
<td>cf-topic</td>
<td>169</td>
<td>628</td>
<td>3.72</td>
<td>61.9%</td>
</tr>
<tr>
<td>RCF</td>
<td>170</td>
<td>358</td>
<td>2.11</td>
<td>35.1%</td>
</tr>
</tbody>
</table>

How Effective are Recommendations?

<table>
<thead>
<tr>
<th></th>
<th>User Feedback Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>topic</td>
<td>0.521</td>
</tr>
<tr>
<td>tbc</td>
<td>0.449</td>
</tr>
<tr>
<td>cf-topic</td>
<td>0.536</td>
</tr>
<tr>
<td>RCF</td>
<td>0.594</td>
</tr>
</tbody>
</table>

Precision = (Num. of Good / Total Num. of Ratings)

Conclusions

- Proposed effective website recommendation based upon a social bookmarking systems
- Can recommend with reasonable effort
- RCF performs better than other commonly used search methods
- Possible to recommend subjective information like blogs, reviews, user generated content
- Provides an avenue toward providing relevant and credible recommendations

Future Work

- Implementation of Negative Feedback
- Incorporation of fallback method for websites not in the bookmarking database
- Improving interface to facilitate recommendation usage

Questions?