Estimating Unbiased Sharer Reputation via Social Data Calibration

Jaewon Yang, Bee-Chung Chen and Deepak Agarwal
Goal: Whom to subscribe to

- “Subscription” feature
  - If user X subscribes to user Y, X receives the news articles shared by Y
  - Only about shared articles: Separate from getting connected to the user

- Topics for news: Marketing, Finance, …

- How can we recommend “whom to subscribe to” for each topic?
News articles in LinkedIn

• Users (& data) are segmented by topics
  – WLOG, ignore topics

• LinkedIn Today
  – News module for all users

• Network update stream
  – What the neighbors share
Problem definition

• Given data (for a given topic):
  – Clicks for articles from friends
  – Clicks for articles from all users (LinkedIn Today)

• Goal: Rank users for suggesting “whom to subscribe to”
Subscription to a user

• On the front page for the topic
• Any user in the topic (even if not connected to the recommended users) will see
Whom to recommend

• Good users to recommend: Users who share attractive articles for the subscription page
  – Attractive articles: Articles which would get high CTR when displayed on the subscription to all users
  – Subscription page is open to all users in the topic
Forming the objective

• We define Unbiased Sharer Reputation (USR) of a user:
  – Average CTR of articles that one shares if the articles would be displayed to all users in the topic

• We aim to estimate USR of all users
LinkedIn example

- LinkedIn Today (LI Today)
  - News module for all users in the topic
- Network update stream
- Unbiased sharer reputation:
  - Average CTR of the articles shared by the user if the articles would be on LI Today
Data: Responses in social network

- Shared articles get responses (clicks) from the neighbors in the social network
- Responses are available for all users
- However, the responses in the social network contain two kinds of bias to be used for USR
Bias in social responses

• Selection bias: From neighbors only

Grad student

Share ➔

Click!

Friends (Grad students)

• Response bias: Affected by social status

Advisor

Share ➔

Click!

Grad student
USR: Novel concept

• USR: Responses from all users
• Social responses: From friends in the social network
• Estimating USR is a new problem
  – Many methods to rank users based on social responses
    • Top users in LinkedIn with existing methods: CEO and Recruiters
Solution: LinkedIn Today data

- LinkedIn Today (LI Today) data: Unbiased user action
  - No selection bias
  - No response bias

- Assign high USR to the users who share high CTR articles at LI Today
  - Problem: For many users, LI Today data is not available

The Wall Street Journal
Problem definition

• Given data:
  – Social responses in the social network
  – Unbiased user action (LI Today CTR)
    • Only available for the articles displayed on LI Today
  – Which users share which articles

• Goal: Estimate unbiased sharer reputation (USR) of all users
Modeling challenges

• Social responses:
  – Available for all users and all shared items
  – Containing social bias

• Unbiased user action (LI Today):
  – Available only for a small subset of articles. (and subset of users)

• Our solution: Calibrate social responses with small unbiased user action
Model diagram

Social responses (complete, biased)

Unbiased user action (LI Today) (incomplete, unbiased)

Biased estimate

Feature based on the user profile

Regression

Unbiased estimate

Calibration by regression
We learn regression between biased estimate and unbiased estimate for USR.

For users with no unbiased user action (LI Today) available:
- Learn biased estimate
- Apply regression
Experiments

• Dataset: LinkedIn
  – May ~ Aug, 2012

• Unbiased user action: LinkedIn Today clicks

• Social responses: Network update stream
Baselines

• Two PageRank-based baselines (Similar to [Saez-Trumper KDD ‘12])
  – Build an influence graph from social response data
  – Run PageRank

• Two baselines from our model
  – Model using only social responses
  – Model using only unbiased user action
Experimental setup

- Rank users using data in May 2012
- Evaluate ranking in Jun ~ Aug
  - Evaluation: ground-truth (GT) USR
    - GT: the average CTR for articles on LinkedIn Today
- Evaluation metric:
  - Kendall’s tau
  - GT reputation of top K users (In the paper)
Results

<table>
<thead>
<tr>
<th>Model</th>
<th>PageRank1</th>
<th>PageRank2</th>
<th>Model (Unbiased)</th>
<th>Model (Social)</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>All users</td>
<td>-0.0176</td>
<td>-0.0215</td>
<td>0.1879</td>
<td>0.0211</td>
<td>0.2078</td>
</tr>
<tr>
<td>Cold-start</td>
<td>-0.0325</td>
<td>-0.0365</td>
<td>N/A</td>
<td>0.0329</td>
<td>0.1124</td>
</tr>
</tbody>
</table>

- Cold start case: Ranking users for whom unbiased user action is not available
- PageRank: Negative correlation
- Our model: 2.4X improvement in Cold-start
Generalization to other networks

• Key intuition: We want to estimate the (global) responses to a user from all users
  – Calibrate social responses with sparse global responses

• Example:
  – Product rating networks (e.g., Goodreads)
    • Global responses: Feedback from all users
    • Social responses: Feedback from the friends
Conclusion

• We investigate the problem of estimating unbiased sharer reputation (USR)
  – USR: Average CTR of the articles when articles are displayed to all users
  – Social responses need calibration
  – We calibrate using unbiased user action

• In the paper:
  – Only small unbiased data needed
  – Mathematical details about the model
THANK YOU!
Calibration by unbiased user action

- Article has latent popularity which generates user action.
- Sharer has unbiased reputation, which generates articles’ popularity.
- Markov Random Prior: Sharers who have same items have similar reputation.
- Unbiased reputation is a function of uncalibrated reputation.