Information Seeking and Responding Networks in Physical Gatherings: A Case Study of Academic Conferences in Twitter

Xidao Wen and Yu-Ru Lin
University of Pittsburgh
Challenges

• Participants may seek and disseminate information while rushing to keep up with the event.

• Participants often find themselves exposed to new places, new people, and new topics.
Research Questions

• RQ1: How do we meaningfully capture the distinctive information needs on Twitter during academic conferences?
Research Questions

• RQ2: Can communications of the information needs be inferred from users’ prior tweeting patterns and network positions?

• Who are the users that tend to post certain types of information needs?

• Can we identify potential responders to these information needs?
Data Collection

• 190k tweets posted in 66 CS conferences over five years (2009-2013).

• ~12k information-seeking tweets (tweets that contain at least one “?” [Paul 2011]).
Information Seeking Annotation & Labels

• We developed a category scheme ($Fleiss\ Kappa = 0.58$) of the information seeking tweets based on previous studies [Morris 2011 & Efron 2010].

• We obtain annotations of the type labels for ~2k information-seeking tweets through Amazon Mechanical Turk.

• We trained n-gram multi-class SVM classifiers ($accuracy = .721$) to automatically label the rest of the information-seeking corpus.
Getting set to head to LA for #siggraph2012 with a stable of #makerbot replicators. Anyone wanna join me there?

"Since we couldn't do either qual or quant research, we decided on mixed methods." huh?! #chi2012

Can't make it to #FAST13 this week? Consider attending via live stream instead: https://t.co/mNyA4IGe

I wish I was at #CHI2010. I wish I got to listen to Genevieve Bell. Does anyone know if I can listen to the keynote online?
RQ2A:
Who are the users that tend to post certain types of information needs?
Coordinate Events/Ask Favors
Express Opinions
Promote Information
Request Information
tweetCount
URL-Ratio
Mention-Ratio
RT-Ratio
is-Mentioned
is-Retweeted

Coordinate Events/Ask Favors
Express Opinions
Promote Information
Request Information
Logistic Regression Model 1

Logistic Regression Model 2

Logistic Regression Model 3

Logistic Regression Model 4

Coordinate Events/Ask Favors

Express Opinions

Promote Information

Request Information
2009 2010 2011 2012

Logistic Regression Model 1
Logistic Regression Model 2
Logistic Regression Model 3
Logistic Regression Model 4

2013?
Coordinate
Events/Ask
Favors
Express
Opinions
Promote
Information
Request
Information
Coordinators

Less likely to give information pointers (via posting URLs).

SIGKDD retweet network (Size of the nodes are proportional to the closeness centrality)
Requestors

Less likely to give information pointers or mention others before.

SIGKDD retweet network (Size of the nodes are proportional to the closeness centrality)
Promoters

More likely to give information pointers, be mentioned, or be retweeted.

SIGKDD retweet network (Size of the nodes are proportional to the out-degrees)
Opinionators

Less likely to give information pointers, but more likely to be mentioned before.

SIGKDD retweet network (Size of the nodes are proportional to the out-degrees)
RQ2B:
Can we identify potential responders to the information needs?
I have a question?
I have a question?

I have an answer.
I have a question?

I have an answer.
Link Prediction Task

Network Proximity

Common Neighbors (CN)
Katz Index (CN)
Jaccard Index (JI)
Preferential Attachment (PA)
Link Prediction Task

Content-based Similarity
Link Prediction Task

Content-based Similarity

- User Similarity
  - Similarity between users’ tweets in the past.
Link Prediction Task
Content-based Similarity

- Text Similarity
  - Similarity between question and users’ tweets in the past.
• Baseline
  • The time difference between the question's posted time and the potential respondent's last seen activity.

• Network Proximity

• Content Similarity
  • User Similarity, Text Similarity, LIWC Similarity

• Balanced dataset
<table>
<thead>
<tr>
<th>Feature Sets</th>
<th>RF</th>
<th>ADA</th>
<th>Bagging</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>0.73</td>
<td>0.774</td>
<td>0.77</td>
<td>0.669</td>
</tr>
<tr>
<td>Baseline+Network</td>
<td>0.824</td>
<td>0.774</td>
<td>0.834</td>
<td>0.689</td>
</tr>
<tr>
<td>Baseline+Content</td>
<td>0.784</td>
<td>0.783</td>
<td>0.81</td>
<td>0.688</td>
</tr>
<tr>
<td>Combined</td>
<td>0.843</td>
<td>0.783</td>
<td>0.859</td>
<td>0.714</td>
</tr>
</tbody>
</table>
Similarity between question and users past tweets is more important!
Summary

• We provided the first large-scale empirical study on the information seeking and responding networks in physical gatherings, using Twitter in Academic conferences as a case study.

• We profiled the four prototypical information seekers in the conferences.
  • Opinionators, Coordinators, Promoters, Requesters.

• With a set of similarity measures, we are able to predict the replying action to the questions with high AUC.
Potential Applications & Implications

• For the Individuals:
  • We can route the less noted questions to the probable respondents.

• For the communities:
  • Event organizers could consider constructing question leader list, so that people with various interests could selectively join the discussions more quickly.
  • Information Seeking and Responding dynamics relates to the sustainability of the community.
Thank you!
Any Questions?