Propagated Opinion Retrieval in Twitter

Zhunchen Luo, Jintao Tang and Ting Wang
Opinion Retrieval in Twitter

• Extension work: opinion retrieval in Twitter (ICWSM 2012)

• Twitter: an important source for people to collect opinions

• Previous Task: finding relevant tweets that express either a negative or positive opinion about some topic
Relevant Tweet (previous)

• Given a topic: “UK strike”

• Relevant tweet

  - Perhaps if the public sector workers on #strike today go Christmas shopping then at least it will give the high street / UK economy a boost!

• Irrelevant tweet

  - UK: BBC - Up to TWO Million Set to Strike http://t.co/wBrsgrKh #tcot #gop #ows
Problem?
Problem?

---→ Large variations
Problem?

---> Large variations

---> Effective using
Problem?

-->Large variations

-->Effective using

-->Important opinions
Problem?

---> Large variations

---> Effective using

---> Important opinions

---> Estimating the importance
Problem?

--- > Large variations

--- > Effective using

--- > Important opinions

--- > Estimating the importance

--- > Retweet
Problem?

--->Large variations

--->Effective using

--->Important opinions

--->Estimating the importance

--->Retweet

Information can deemed important by the community propagates through retweets (WWW 2011)
Problem?

--- > Large variations

--- > Effective using

--- > Important opinions

--- > Estimating the importance

--- > Retweet

Information can deemed important by the community propagates through retweets (WWW 2011)
Our Task

• Goal: finding propagated opinions -- tweets that express an opinion about some topics, but will be retweeted
Relevant Tweet (now)

• Given a topic: “Obama”

• Relevant tweet
  • RT@KG_NYK: The fact that Obama “lost” the debate b/c he didn’t call Romney’s lies out well enough is pretty harrowing commentary on surf

• Irrelevant tweet
  • MyNameisGurley AND I HATE OBAMA
Our Work

- A new ranking task aiming at finding opinionated tweets that will be propagated in the future

- Learning-to-rank for Twitter propagated opinion retrieval
  - Retweetability feature: whether a tweet in general will be retweeted
  - Opinionatedness feature: opinionatedness score of a tweet
  - Textural quality features: textural information of a tweet
Data

• 50 queries and 5000 judged tweets
• 3.4 relevant tweets per query

• https://sourceforge.net/projects/ortwitter/
Retweetability Feature

• Predicting the retweetability score of a tweet (ICWSM2011: “RT to win! Predicting Message Propagation in Twitter”)

• 30 millions training tweets

• Machine Learning: passive-aggressive algorithm

• Features: content; followers number, listed number, verified

• Accuracy: 95.99% (testing 100,000 tweets)
Opinionatedness Feature

• Estimating the opinionatedness score of a tweet (ICWSM2012: “Opinion Retrieval in Twitter”)
• Lexicon-based approach
• Automatically construct opinionated lexical for Twitter
Opinionated Tweets

• “Pseudo” Subjective Tweet (PST): a tweet of the form “RT @username” with text before the retweet

[Amber Lyon @AmberLyon
Tragically not much has changed RT: “@SAlwaddei: BBC report on #Bahrain from 90s. History repeating itself. youtube.com/watch?v=tApsza…”]

• “Pseudo” Objective Tweet (POT): If a tweet satisfies two criteria: (1) it contains links and (2) the user of this tweet posted many tweets before and has many followers

[BBC Breaking News @BBCBreaking
Minimum price of alcohol in Scotland proposed at 50p per unit - Scottish Health Secretary Nicola Sturgeon bbc.in/Kkcyv0]

• A term can be measured how dependent with PST set and POT set
Textural Quality Features

- Length
- Part of speech
- Fluency (language model)
Experiment

- Experimental Settings
  - SVM Rank
  - 10 fold cross-validation
  - Evaluation metric: Mean Average Precision (MAP)

- Baselines
  - BM25
  - TOR (ICWSM2012 Twitter opinion retrieval approach): BM25, URL, Mention, Statuses, Followers, Opinionatedness
## Result

<table>
<thead>
<tr>
<th>Model</th>
<th>MAP</th>
<th>TOR</th>
<th>MAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>BM25</td>
<td>0.0997</td>
<td>TOR</td>
<td>0.1521</td>
</tr>
<tr>
<td>BM25+Retweetability</td>
<td>0.1077</td>
<td>TOR +Retweetability</td>
<td>0.1806</td>
</tr>
<tr>
<td>BM25+Opinionatedness</td>
<td>0.1146</td>
<td>TOR +Textural Quality</td>
<td>0.1930</td>
</tr>
<tr>
<td>BM25+Textural Quality</td>
<td>0.1277</td>
<td>TOR +All</td>
<td>0.1992</td>
</tr>
<tr>
<td>BM25+All</td>
<td>0.1317</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Comparison with Humans

• Our approach for identifying the propagated opinion in Twitter can achieve human subject’s ability as well!!!

• 100 pairs of tweets (same topic + one relevant tweet + the other is irrelevant)

• Result (accuracy):
  • Person A: 75%
  • Person B: 69%
  • Our approach: 71% (persons: 72%)
Conclusion

- A new task aims at finding propagated opinions in Twitter.
- Features, such as the retweetability, opinionatedness and textural quality of a tweet, are effective for solving this problem.
- Our approach can achieve the human subjects' ability to identify the propagated opinions in Twitter.
Thank you for your attention!

Zhunchen Luo

zhunchenluo@nudt.edu.cn
https://sites.google.com/site/zhunchenluo/