

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 be deemed important by the community propagates through retweets (WWW 2011)

Problem?

---> Large variations

---> Effective using

---> Important opinions

---> Estimating the importance

---> Retweet

Information can be 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 didnt 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/>

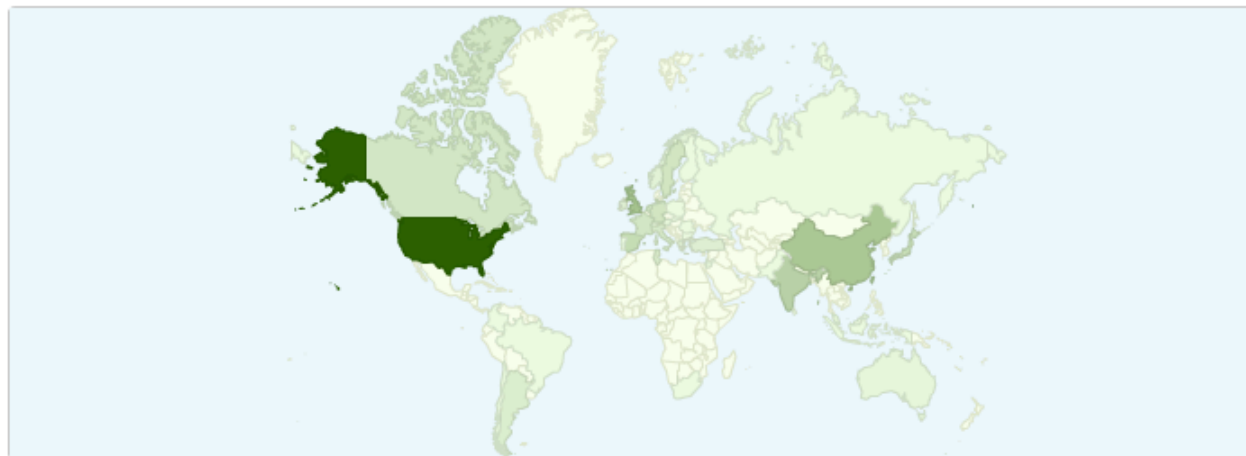
OpinionRetrieval_Twitter

Summary | Files | Reviews | Support | Wiki | Code | Tickets | Discussion

Brought to you by: zhunchenluo

[Home](#) / [Corpus.tar](#) [\(Change File\)](#)

Date Range: 2012-03-02 to 2013-09-20



DOWNLOADS

248

In the selected date range

TOP COUNTRY

United States

22% of downloaders

TOP OS

Windows

60% of downloaders

OS downloads as: Percent

Country	Linux	Macintosh	Unknown	Windows	Total
1. United States	4%	25%	4%	61%	57
2. United Kingdom	0%	20%	0%	40%	20
3. China	16%	5%	0%	42%	19
4. India	20%	20%	0%	60%	15
5. Spain	36%	18%	0%	45%	11

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

4時

Tragically not much has changed **RT** :“@SAlwadaei: **BBC** report on #Bahrain from 90s. History repeating itself. [youtube.com/watch?v=tApsza...](https://www.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

5月14日

Minimum price of alcohol in Scotland proposed at 50p per unit -
Scottish Health Secretary Nicola Sturgeon [bbc.in/Kkcyv0](https://www.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

	MAP		MAP
BM25	0.0997	TOR	0.1521
BM25+Retweetability	0.1077	TOR +Retweetability	0.1806
BM25+Opinionatedness	0.1146		
BM25+Textural Quality	0.1277	TOR +Textural Quality	0.1930
BM25+All	0.1317	TOR +All	0.1992

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/>