Who Will Retweet Me? Finding Retweeters in Twitter
Zhunchen Luo (zhunchenluo@nudt.edu.cn) Miles Osborne (miles@inf.ed.ac.uk)
Jintao Tang (tangjintao@nudt.edu.cn) Ting Wang (tingwang@nudt.edu.cn)

Introduction

Problem: finding who (retweeters) will retweet a message posted on Twitter

Motivation:
• Deepening understanding of how information flows in Social Networks
• Being of interest to advertisers and media companies
• Treating finding retweeters as a ranking problem which retrieves followers retweeting a certain post
• Using a standard machine learning approach to learn a ranking function for followers that uses a range of features

Dataset: 500 tweets and 40574 followers (257 retweeters)

Features

Retweet History (RH)
• Num_fRu (Num_fMu): # follower retweeted (mentioned) author’s tweets before
• Num_uRf (Num_uRf): # author retweeted (mentioned) follower’s tweets before
• Ratio_retweet (Ratio_mention): % follower’s tweets are retweets (or contain mention ‘@’) before

Follower Status (FS)
• Posts: # tweets ever posted
• Followers: # followers
• Friend: # friends
• Listed: # has been listed
• Verified: ? verified

Follower Active Time (FAT)
• Timezone: ? author and follower are in the same timezone
• PostTimeConsis: % follower’s tweets posted before which were in the same hour interval to the tweet’s posting time

Follower Interests (FI)
• SimInterest: the cosine angle between the follower’s previous tweets and the tweet (based on vector space model)

Experiment

Baselines:
• Random: ranking followers randomly
• RPT: ranking followers by the number of times they retweeted the author’s previous tweets before

Leanring to rank model: SVM^Rank

Experimental result (Ten-fold cross-validation):

<table>
<thead>
<tr>
<th>Feature</th>
<th>MAP(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>2.17</td>
</tr>
<tr>
<td>PRT</td>
<td>6.93</td>
</tr>
<tr>
<td>RH</td>
<td>6.27</td>
</tr>
<tr>
<td>FS</td>
<td>3.66</td>
</tr>
<tr>
<td>FAT</td>
<td>2.91</td>
</tr>
<tr>
<td>FI</td>
<td>8.12</td>
</tr>
<tr>
<td>All</td>
<td>8.71</td>
</tr>
</tbody>
</table>

Examples

• We are having a bake sale today in the Student Union from 11-2! Come buy a midday snack from the Pretty Poodles!
  • RH model ranks the follower, who retweeted or mentioned the author of this tweet 30 times totally in previous tweets, in first place

• Excited to announce our debut London show. Full details here - http://t.co/P60Wc3Lj
  • FI model ranks the follower high who often posted about music and performance as shown in previous tweets

Conclusion

The retweet history, followers’ status and interest are effective information for finding retweeters in Twitter

The new task and our results will open the way for follow-up research better understanding how Social Media works.

Our Dataset URL: http://sourceforge.net/projects/retweeter/