Classification of Twitter Follow Links Based on the Followers’ Intention

Hikaru Takemura  
Yahoo Japan Corporation

Atsushi Tanaka  
Mainichi Broadcasting System, Inc.

Keishi Tajima  
Kyoto University
Background

- **Twitter**: a micro-blogging service where users can post short messages (tweets).

- The most distinctive feature of Twitter is the mechanism of follow.
  - @A is a follower of @B
  - @B is a followee of @A
Background

- Twitter is used for various purposes.
  - to gather information as with RSS readers
  - to read celebrities’ chats as in blog services
  - to share discussions as in discussion forums
  - to have communication with friends as in SNSs
  - ... and more
Various types of follow intentions in Twitter

- gathering information
- a fan
- daily conversations

news media  famous singer  friend
Motivation

- Intentions behind follow links are important.
- e.g., user recommendation
Motivation

- Intentions behind follow links are important.
- e.g., user recommendation
Our Goal

- We proposed a scheme for classifying follow links based on the followers’ intention consisting of **the three axes**:
  - User-orientation
  - Content-orientation
  - Mutuality

- We develop a method of classifying links into **8 types** by combining these three axes.
Three Classification Axes

**User-orientation**: the follower is interested in the followee itself, and it cannot be replaced with another user with very similar tweets.

**Content-orientation**: the follower is interested in specific topic, and no reason to follow the followee if he stops tweeting about the topic.

**Mutuality**: the follower expect to have mutual communication.
Three Classification Axes

- User-orientation and Content-orientation are not exclusive.

I need business news.

Content-orientation

User-orientation

I like CNN. It has to be CNN, not BBC.
Three Classification Axes

content-orientation

information source

authoritative source

open discussion

closed discussion

spammer

celebrity

socializer

friend

user-orientation

mutuality
Three Classification Axes

content-orientation

information source

authoritative source

open discussion

closed discussion

spammer

celebrity

socializer

friend

user-orientation

mutuality
Three Classification Axes

content-orientation

information source

authoritative source

open discussion

closed discussion

spammer

celebrity

socializer

friend

user-orientation

mutuality
Three Classification Axes

content-orientation

information source

open discussion

authoritative source

closed discussion

spammer

celebrity

socializer

friend

user-orientation

mutuality
Three Classification Axes

- content-orientation
  - information source
  - authoritative source

- user-orientation
  - open discussion
  - closed discussion

- mutuality
  - spammer
  - celebrity
  - socializer
  - friend
Classification of Follow Links

- We use SVMs and Decision Trees for classification.

- For features, we use three types of properties.
  (a) properties of the follower
  (b) properties of the followee
  (c) properties of their relationship
Features for Classifiers

(a) follower
- # of followees, followers, reciprocal follows, lists
- reciprocal followee ratio, reciprocal follower ratio

(b) followee
- # of followees, followers, reciprocal follows, lists
- reciprocal followee ratio, reciprocal follower ratio
- proportion of information lists to all lists

(c) relationship
- reciprocity, # of common lists
- frequency of replies by “@”
- frequency of RT of followee’s tweets by follower
Features for Classifiers

(a) follower
- # of followees, followers, reciprocal follows, lists
- reciprocal followee ratio, reciprocal follower ratio

(b) followee
- # of followees, followers, reciprocal follows, lists
- reciprocal followee ratio, reciprocal follower ratio
- proportion of information lists to all lists

(c) relationship
- reciprocity, # of common lists
- frequency of replies by “@”
- frequency of RT of followee’s tweets by follower
Features for Classifiers

(a) follower
- # of followees, followers, reciprocal follows, lists
- reciprocal followee ratio, reciprocal follower ratio

(b) followee
- # of followees, followers, reciprocal follows, lists
- reciprocal followee ratio, reciprocal follower ratio
- proportion of information lists to all lists

(c) relationship
- reciprocity, # of common lists
- frequency of replies by “@”
- frequency of RT of followee’s tweets by follower
Features for Classifiers

reciprocal followee ratio, reciprocal follower ratio

\[
\text{reciprocal followee ratio} = \frac{\# \text{ of reciprocal follows}}{\# \text{ of one’s followee}}
\]

\[
\text{reciprocal follower ratio} = \frac{\# \text{ of reciprocal follows}}{\# \text{ of one’s follower}}
\]

These ratios identify users’ type.
- e.g. If both ratios are high, the user is probably an information source or a communication user.
Features for Classifiers

reciprocal followee ratio, reciprocal follower ratio

\[ \text{reciprocal follower ratio} = \frac{3}{5} \quad \text{reciprocal followee ratio} = \frac{3}{4} \]

- These ratios identify users’ type.
- e.g. If both ratios are high, the user is probably an information source or a communication user.
Features for Classifiers

reciprocal followee ratio, reciprocal follower ratio

- These ratios identify users’ type.
  - e.g. If both ratios are high, the user is probably an information source or a communication user.

\[
\text{reciprocal follower ratio} = \frac{3}{5} \quad \text{reciprocal followee ratio} = \frac{3}{4}
\]
Features for Classifiers

(a) follower
- # of followees, followers, reciprocal follows, lists
- reciprocal followee ratio, reciprocal follower ratio

(b) followee
- # of followees, followers, reciprocal follows, lists
- reciprocal followee ratio, reciprocal follower ratio
- proportion of information lists to all lists

(c) relationship
- reciprocity, # of common lists
- frequency of replies by “@”
- frequency of RT of followee’s tweets by follower
Features for Classifiers

(a) follower
- # of followees, followers, reciprocal follows, lists
- reciprocal followee ratio, reciprocal follower ratio

(b) followee
- # of followees, followers, reciprocal follows, lists
- reciprocal followee ratio, reciprocal follower ratio
- proportion of information lists to all lists

(c) relationship
- reciprocity, # of common lists
- frequency of replies by “@”
- frequency of RT of followee’s tweets by follower
Features for Classifiers

proportion of information lists to all lists

- **information lists**: lists used for grouping information sources of related topics

- **community lists**: lists used for grouping users belonging to a specific community

- If the followee is mainly included in information lists, the link is probably for gathering information.
- If the followee is mainly included in community lists, the link is probably for communication.
Features for Classifiers

proportion of information lists to all lists

■ for each list $l$,

$$ff(l) = \frac{1}{|\text{member}(l)|} \sum_{u \in \text{member}(l)} \frac{\# \text{ of } u\text{'s follower}}{\# \text{ of } u\text{'s followee}}$$

■ if $ff(l) > \delta$, $l$ is an information list
■ otherwise, $l$ is a community list
Features for Classifiers

(a) follower
- # of followees, followers, reciprocal follows, lists
- reciprocal followee ratio, reciprocal follower ratio

(b) followee
- # of followees, followers, reciprocal follows, lists
- reciprocal followee ratio, reciprocal follower ratio
- proportion of information lists to all lists

(c) relationship
- reciprocity, # of common lists
- frequency of replies by “@”
- frequency of RT of followee’s tweets by follower
Experimental Data Set

- We hired 44 Twitter users through crowdsourcing service in Japan (Lancers).

- For each Twitter user, we randomly chose 40 followees of the user, and ask the user to answer whether each follow link has user-orientation, content-orientation, and mutuality.
Experimental Data Set

- Content-oriented follows are more frequent than user-oriented ones.
- User-orientation and content-orientation have weak positive correlation.
- User-orientation and mutuality have weak positive correlation.
Experimental Data Set

- Content-oriented follows are more frequent than user-oriented ones.
- User-orientation and content-orientation have weak positive correlation.
- User-orientation and mutuality have weak positive correlation.

![Diagram with data points]

- **Content-orientation**: 1,120, 386, 553, 26, 379, 179
- **User-orientation**: 34, 48
- **Mutuality**:
Experimental Data Set

- Content-oriented follows are more frequent than user-oriented ones.
- User-orientation and content-orientation have weak positive correlation.
- User-orientation and mutuality have weak positive correlation.
Experimental Data Set

- Content-oriented follows are more frequent than user-oriented ones.
- User-orientation and content-orientation have weak positive correlation.
- User-orientation and mutuality have weak positive correlation.
Experimental Data Set

- Content-oriented follows are more frequent than user-oriented ones.
- User-orientation and content-orientation have weak positive correlation.
- User-orientation and mutuality have weak positive correlation.
Experimental Results

- classification of 1,253 follow links into 8 types
- SVMs and Decision Trees are used for classification.
- For each of them, we compared two methods:
  1. a single 8-class classifier
  2. 3 binary classifiers corresponding to the three axes
Experimental Results

- accuracy of classification into 8 classes

(a) follower (b) followee (c) relationship

<table>
<thead>
<tr>
<th>features</th>
<th>a+b</th>
<th>a+c</th>
<th>b+c</th>
<th>a+b+c</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 binary SVMs</td>
<td>36.95</td>
<td>50.20</td>
<td>34.08</td>
<td>42.70</td>
</tr>
<tr>
<td>3 binary Decision Trees</td>
<td>43.58</td>
<td>54.91</td>
<td>27.29</td>
<td>43.26</td>
</tr>
<tr>
<td>single 8-class SVM</td>
<td>46.93</td>
<td>43.26</td>
<td>37.51</td>
<td>50.28</td>
</tr>
<tr>
<td>single 8-class Decision Tree</td>
<td>49.16</td>
<td>55.87</td>
<td>29.93</td>
<td>50.12</td>
</tr>
</tbody>
</table>

- The type of follower is more important than followee.
- Decision Trees works better than SVMs.
- The difference between 3 binary classifiers and 8-class classifiers is not significant.
Experimental Results

- accuracy of Decision Trees without each feature

(a) follower (b) followee (c) relationship

<table>
<thead>
<tr>
<th>removed feature</th>
<th>3 binary</th>
<th>8-class</th>
</tr>
</thead>
<tbody>
<tr>
<td>with all features</td>
<td>43.26</td>
<td>50.12</td>
</tr>
<tr>
<td>(a) follower : reciprocal followee ratio</td>
<td>40.30</td>
<td>51.16</td>
</tr>
<tr>
<td>(c) relationship : frequency of @</td>
<td>40.94</td>
<td>48.76</td>
</tr>
</tbody>
</table>

- The accuracy becomes lowest when we remove
  - reciprocal followee ratio for 3 binary Decision Trees
  - frequency of @ for a 8-class Decision Trees

- But no single property can be a prominent discriminator.
Conclusion

- We classified Twitter follow links into 8 types along the three axes (user-orientation, content-orientation, and mutuality).

- We found the following facts.
  - Content-oriented follows are more frequent than user-oriented ones.
  - User-orientation and content-orientation have weak positive correlation.
  - User-orientation and mutuality have weak positive correlation.
  - Link types do not solely depend on the followees.
  - No single property can be a prominent discriminator.