Detecting Tip Spam in Location-based Social Networks

Helen Costa ¹  Fabricio Benevenuto ²  Luiz H. C. Merschmann ¹

¹Federal University of Ouro Preto (UFOP) - Brazil
²Federal University of Minas Gerais (UFMG) - Brazil
LOCATION-BASED SOCIAL NETWORK (LBSN)

▶ What?
  ▶ Systems that allow the users to share their geographic location with their social network

▶ How?
  ▶ Check-in through a smartphone with GPS
  ▶ Recommendations
MOTIVATION

- The most interesting function: post tips
- New form of spam
  - Unsolicited message
  - Spread advertise
  - Disseminate pornography
  - Fake tips
Tip Spam in LBSN

**4food**
286 Madison Avenue (at 40th Street), New York, NY 10016, United States
Fast Food Restaurant, American Restaurant, Burger Joint

* Jerard R. March 22, 2011
Their cola does not taste like traditional soda. I do not recommend.

* Victoria W. May 31, 2011
Wholesome food, free wifi, friendly staff and the owner was so nice. Def recommend

* John R. September 27, 2011
Mac n cheese burger is awesome.

* EventAdviser.com December 28, 2011
Order the GRUNTBURGER with sweet potatoe fries! Go to EventAdviser.com today for cheap tickets!

* Pete L. July 29, 2012
Colin Farrell already has a career as an actor. Just give it up already.
NEGATIVE IMPACT OF TIP SPAM

- Jeopardize the trust of users on the existing tips
- Waste human attention
Goal and methodology

- **Goal**: Detect tip spam in LBSN

- 3-step approach
  1. Obtain tips labeled as spam or non-spam and crawler additional information
  2. Identify attributes able to distinguish spam from non-spam tips
  3. Classification approach to detect tip spam
Step 1. Labeled Dataset and Crawled Information

- Apontador
  - Brazilian LBSN
  - Provided us tips labeled as spam or non-spam by their moderators
Step 1. Labeled dataset and crawled information

- Labeled dataset
  - Week period
  - 1,260 tips classified as spam and the 1,260 tips classified as non-spam
  - Each tip contains the following information
    - tip content
    - tip ID
    - user ID
    - place ID
    - etc.
Step 1. Labeled Dataset and Crawled Information

- Crawled information
  - Information not available in the provided data:
    - Geographical location of places
    - User information
    - Social graph
  - In total our labeled dataset contains
    - 2,520 tips
    - 1,984 unique users
    - 2,216 different places
    - Social graph with 137,464 users
VERIFYING LABELING ACCURACY

- Volunteers from our research group manually verify 100 randomly selected spam tips
- Volunteers classified 2 tips as non-spam and 98 as spam
  - 65 were local advertises
  - 29 pollution (i.e., unrelated or irrelevant text)
  - 4 were aggressive comments about the places
Step 2. Attributes

- **Content attributes** (total=16)
  - number of words
  - number of numeric characters
  - number of URLs
  - etc.

- **User attribute** (total=8)
  - number of tips posted by the user
  - average distance among all places reviewed by the user
  - etc.

- **Place attributes** (Total=5):
  - number of tips on the place
  - place rating
  - etc.

- **Social attribute** (Total=12)
  - number of followers
  - number of followees
  - clustering coefficient
  - etc.
# Importance of the Attributes

<table>
<thead>
<tr>
<th>category</th>
<th>$\chi^2$ ranking</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>place</td>
<td>1</td>
<td>Number of tips on the place</td>
</tr>
<tr>
<td>place</td>
<td>2</td>
<td>Place rating</td>
</tr>
<tr>
<td>content</td>
<td>3</td>
<td>Number of contact information on the text</td>
</tr>
<tr>
<td>content</td>
<td>4</td>
<td>Number of numeric characters</td>
</tr>
<tr>
<td>content</td>
<td>5</td>
<td>Number of phone numbers on the text</td>
</tr>
<tr>
<td>content</td>
<td>6</td>
<td>Number of email addresses on the text</td>
</tr>
<tr>
<td>content</td>
<td>7</td>
<td>Number of words</td>
</tr>
<tr>
<td>social</td>
<td>8</td>
<td>Number of followers (in-degree)</td>
</tr>
<tr>
<td>user</td>
<td>9</td>
<td>Maximum distance among all places reviewed by the user</td>
</tr>
<tr>
<td>user</td>
<td>10</td>
<td>Standard deviation of distance among all places reviewed by the user</td>
</tr>
</tbody>
</table>

It was important to investigate each attribute set
Distinguishing classes of tips

Spam tips tend to have phone numbers, and thus, more numerical characters.
Step 3. Classification approach

- *RandomForest* implemented in the Weka tool as classifier
- Use all 41 attributes
- 5-fold cross validation (repeated 10 times with different seeds)
- Results are average of 50 runs
## Classification Results

<table>
<thead>
<tr>
<th>True Label</th>
<th>Non-spam</th>
<th>Spam</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-spam</td>
<td>0.918</td>
<td>0.082</td>
</tr>
<tr>
<td>Spam</td>
<td>0.160</td>
<td>0.840</td>
</tr>
</tbody>
</table>

- 84% of the spam tips were correctly classified as spam
- For the non-spam tips, 92% were classified correctly
- 88% of accuracy
Reducing the Attribute Set

Different subsets of attributes can obtain competitive results
We propose a mechanism to detect tip spam in LBSN

- Apontador dataset and labeled collection
  - Publicly available (soon) at homepages.dcc.ufmg.br/~fabricio/
- Attribute identification
- Classification approach
  - Correctly identified majority of spam tips
  - Different subsets of attributes can obtain competitive results

Future work
- Identify subclasses of tip spam
QUESTIONS

helen.c.s.costa@gmail.com