Detecting Tip Spam in Location-based Social Networks

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

[1] *Federal University of Ouro Preto (UFOP) - Brazil*

[2] *Federal University of Minas Gerais (UFMG) - Brazil*

UFOP
Universidade Federal de Ouro Preto

LOCATION-BASED SOCIAL NETWORK (LBSN)

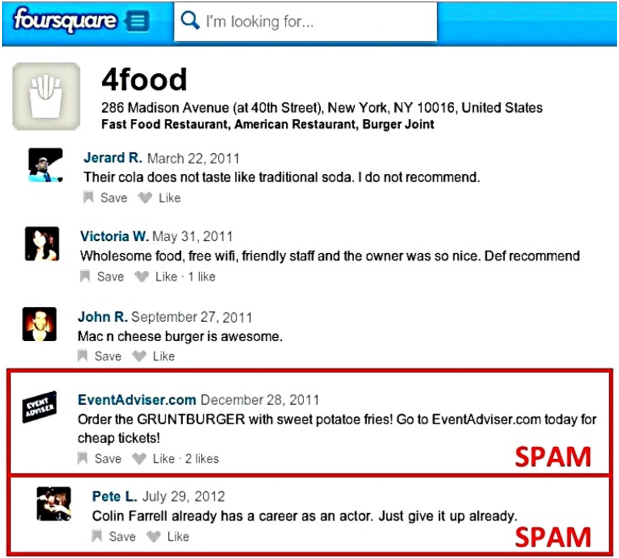
- What?
  - Systems that allow the users to share their geographic location with their social network
- How?
  - Check-in through a smartphone with GPS
  - Recommendations

UFOP Universidade Federal de Ouro Preto

MOTIVATION

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

TIP SPAM IN LBSN



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

STEP1. LABELED DATASET AND CRAWLED INFORMATION

- Apontador
  - Brazilian LBSN
  - Provided us tips labeled as spam or non-spam by their moderators

STEP1. 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.

UFOP
Universidade Federal de Ouro Preto

STEP1. 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

UFOP
Universidade Federal de Ouro Preto

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

STEP2. 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.

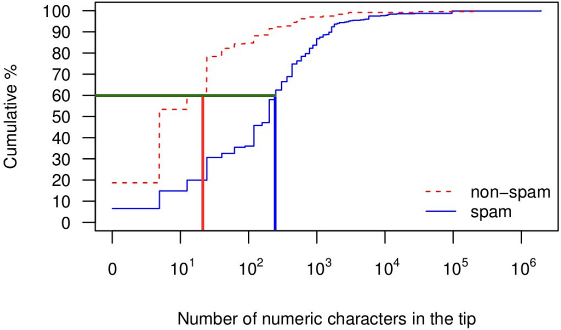
IMPORTANCE OF THE ATTRIBUTES

category	χ^2 ranking	description
place	**1**	Number of tips on the place
place	**2**	Place rating
content	**3**	Number of contact information on the text
content	**4**	Number of numeric characters
content	**5**	Number of phone numbers on the text
content	**6**	Number of email addresses on the text
content	**7**	Number of words
social	**8**	Number of followers (in-degree)
user	**9**	Maximum distance among all places reviewed by the user
user	**10**	Standard deviation of distance among all places reviewed by the user

It was important to investigate each attribute set

UFOP
Universidade Federal de Ouro Preto

DISTINGUISHING CLASSES OF TIPS



Spam tips tend to have phone numbers, and thus, more numerical characters

Universidade Federal de Ouro Preto

STEP3. 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**

Universidade Federal de Ouro Preto

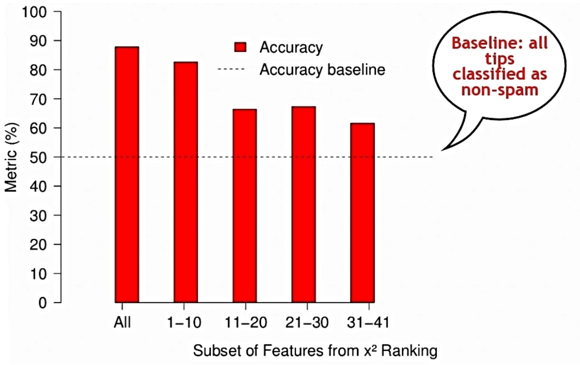
CLASSIFICATION RESULTS

		Predicted Label	
		Non-spam	**Spam**
True Label	**Non-spam**	**0.918**	0.082
	Spam	0.160	**0.840**

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

Universidade Federal de Ouro Preto

Reducing the attribute set



Different subsets of attributes can obtain competitive results

Universidade Federal de Ouro Preto

Conclusions and Future Work

- 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

UFOP Universidade Federal de Ouro Preto

Questions

helen.c.s.costa@gmail.com

UFOP
Universidade Federal de Ouro Preto