Brand-related Events Detection, Classification and Summarization on Twitter

Eric Medvet    Alberto Bartoli

DIA - University of Trieste

December 7th, 2012

http://machinelearning.inginf.units.it
1 Scenario

2 Approach
   • Preprocessing
   • Event detection
   • Event classification
   • Event summarization

3 Experimental evaluation
Twitter

A microblogging platform spreading people’s short texts (tweets)

- several millions of users
- hundreds of millions of tweets per day including
  - first- or second-hand news
  - chatter
  - opinions about facts, persons, products
  - …
Twitter

A microblogging platform spreading people’s short texts (tweets)

- several millions of users
- hundreds of millions of tweets per day including
  - first- or second-hand news
  - chatter
  - opinions about facts, persons, products
  - ...

Does it contain/can we extract useful information?
Can we automatically notify Google about these users who like the new UI?
# Table of Contents

1. Scenario

2. Approach
   - Preprocessing
   - Event detection
   - Event classification
   - Event summarization

3. Experimental evaluation
A method for, given a topic, from all today’s tweets related to the topic:

1. identifying one (or more) popular conversation themes (events)
2. assessing the sentiment polarity and popularity of the event
3. choosing one (or more) tweets which represent the event
Contribution in a nutshell

A method for, given a *topic*, from all today’s tweets related to the topic

1. identifying one (or more) popular conversation themes (*events*) detection

2. assessing the sentiment polarity and popularity of the event classification

3. choosing one (or more) tweets which represent the event summarization
A method for, given a *topic*, from all today’s tweets related to the topic

1. identifying one (or more) popular conversation themes (*events*) detection
2. assessing the sentiment polarity and popularity of the event classification
3. choosing one (or more) tweets which represent the event summarization

Totally *unsupervised*!
Example

**Approach**

- **Identify events**: new maps ui
- **Assess sentiment/popularity**: positive/high
- **Choose one representing tweet**: Very happy for the new #google maps ui - like it!
Example

**topic = google**

1. identify events: new maps ui
2. assess sentiment/popularity: positive/high
3. choose one representing tweet: Very happy for the new #google maps ui - like it!

Can provide useful information? Yes!
• topic → a pre-compiled set of keywords
  • we focused on brands (google, apple, microsoft)
  • topic = apple → keywords = {apple, iphone, ipad, mac, ...}

• “current” tweets about the topic can be obtained in real-time using
  the Twitter Streaming API
Preprocessing

Applied to all the tweets related to the topic, up to 14 days old

- convert to lowercase
- replace URLs with token T_URL
- replace emoticons with token T_POS_EMOT and T_NEG_EMOT
- replace numbers with token T_NUM
- expand common acronyms and abbreviation
- remove English stop-words
- stemming
Tweet $\rightarrow$ features

- compute unigrams (word counts in each tweet)
- drop all but most occurring 2000 unigrams
Detection

Event → a set of 3 words occurring with unusually high rate

Three steps:

1. find words with unusually high rate
2. build sets from words
3. select set with unusually high rate
Detection: set vs. word

Why sets instead of single words?

- avoid considering recurring words as event (#ff, #musicmonday)
- preserve coverage by associating a tweet with a 3-words-set when contains at least 2 words
Detection: unusually high rate

Burstiness index of word $w$:

$$b_w = f_w^{\text{now}} - 3f_w^{\text{hist}}$$

- $f_w^{\text{now}} \rightarrow$ current (last 3 days) relative frequency
- $f_w^{\text{hist}} \rightarrow$ historic (last 14 days) relative frequency

select the words with highest $b_w$
Detection: word $\rightarrow$ set

Given a word $w$:

1. compute *co-burstiness index* $b_{v \wedge w}$ of other words $v$ occurring with $w$
2. select the two words $v_1, v_2$ with the highest $b_{v \wedge w}$ to form set $W$
3. select the set $W^*$ with the highest *set burstiness index* $b_{W^*}$
Classification: goal

Determine, for each event
- popularity as $Q_{\text{pop}}$ in \{low, medium, high\}
- sentiment polarity as $Q_{\text{sent}}$ in \{positive, neutral, negative\}

starting from the event corpus $T^*$
Classification: popularity

1. compare current (last 3 days) $T^*$ relative size vs. past sizes
2. set:
   - $Q_{pop} = \text{high}$, if current size is the largest
   - $Q_{pop} = \text{medium}$, if in top 50%
   - $Q_{pop} = \text{low}$, otherwise
Classification: sentiment polarity

1. SVM classification of each tweet of $T^*$
   - SVM is trained on a static sentiment corpus of hand labeled tweets
   - feature selection (2000 → 50), using a linear model

2. set $Q_{sent}$ to the sentiment of the largest portion $T^*_{sent}$ of $T^*$
Summarization

Goal: select one tweet which represents the event

1. compute the centroid of $T_{sent}^*$ features
2. select the tweet nearest to the centroid

$T_{sent}^*$ is the corpus of tweets with sentiment polarity $= Q_{sent}$
Table of Contents

1 Scenario

2 Approach
   ● Preprocessing
   ● Event detection
   ● Event classification
   ● Event summarization

3 Experimental evaluation
Experimental evaluation

Data

Stream corpus (i.e., live tweets):

- more than 8,000,000 tweets of US users
- September, 2009–March, 2010
- about 41,000 tweets per day

Sentiment corpus:

- 1,611 tweets
- balanced among labels positive, neutral, negative
### Three popular brands:

<table>
<thead>
<tr>
<th>Brand</th>
<th>Keywords</th>
<th>Tweets</th>
<th>T./day</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apple</td>
<td>apple, iphone, ipad, ipod, ios, mac, macintosh, macbook</td>
<td>53,680</td>
<td>272</td>
</tr>
<tr>
<td>Google</td>
<td>google, android, chrome</td>
<td>29,073</td>
<td>148</td>
</tr>
<tr>
<td>Microsoft</td>
<td>microsoft, explorer, msn, windows, winxp, vista, xbox</td>
<td>20,897</td>
<td>106</td>
</tr>
</tbody>
</table>
## Result

### Some detected, classified and summarized events:

<table>
<thead>
<tr>
<th>Brand</th>
<th>Date</th>
<th>Event keywords</th>
<th>$Q_{pop}$</th>
<th>$Q_{sent}$</th>
<th>Summarizing tweet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apple</td>
<td>09/10/21</td>
<td>imac apple mouse</td>
<td>high</td>
<td>pos.</td>
<td>the new Apple Magic Mouse looks glorious! http://****</td>
</tr>
<tr>
<td>Apple</td>
<td>10/3/14</td>
<td>pre ipad order</td>
<td>high</td>
<td>neut.</td>
<td>@**** did you pre order your iPad?</td>
</tr>
<tr>
<td>Google</td>
<td>09/11/20</td>
<td>event os chrome</td>
<td>high</td>
<td>neut.</td>
<td>Google Chrome OS on Mixx: http://****</td>
</tr>
<tr>
<td>Google</td>
<td>10/3/12</td>
<td>direction map bike</td>
<td>high</td>
<td>neut.</td>
<td>RT @<strong><strong>: (BikeRadar) Cycling directions added to Google Maps in US http://</strong></strong> #procycling</td>
</tr>
<tr>
<td>Microsoft</td>
<td>09/10/1</td>
<td>security essentials microsoft</td>
<td>high</td>
<td>pos.</td>
<td>Microsoft Security Essentials is now available - free AntiVirus/Malware protection - http://****</td>
</tr>
</tbody>
</table>
## Experimental evaluation

### Result

Some detected, classified and summarized events:

<table>
<thead>
<tr>
<th>Brand</th>
<th>Date</th>
<th>Event keywords</th>
<th>$Q_{pop}$</th>
<th>$Q_{sent}$</th>
<th>Summarizing tweet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apple</td>
<td>09/10/21</td>
<td>imac apple mouse</td>
<td>high</td>
<td>pos.</td>
<td>the new Apple Magic Mouse looks glorious! http://****</td>
</tr>
<tr>
<td>Apple</td>
<td>10/3/14</td>
<td>pre ipad order</td>
<td>high</td>
<td>neut.</td>
<td>@**** did you pre order your iPad?</td>
</tr>
<tr>
<td>Google</td>
<td>09/11/20</td>
<td>event os chrome</td>
<td>high</td>
<td>neut.</td>
<td>Google Chrome OS on Mixx: http://****</td>
</tr>
<tr>
<td>Google</td>
<td>10/3/12</td>
<td>direction map bike</td>
<td>high</td>
<td>neut.</td>
<td>RT @<strong><strong>: (BikeRadar) Cycling directions added to Google Maps in US http://</strong></strong> #procycling</td>
</tr>
<tr>
<td>Microsoft</td>
<td>09/10/1</td>
<td>security essentials microsoft</td>
<td>high</td>
<td>pos.</td>
<td>Microsoft Security Essentials is now available - free AntiVirus/Malware protection - http://****</td>
</tr>
</tbody>
</table>

### Validation:

First Look: ***Apple Magic Mouse - PCWorld***

PC World - Oct 21, 2009

Apple's new Magic Mouse has a creative new design that, upon first impression, you'll either love or hate. The Bluetooth Magic Mouse has no visible buttons. ...  

*Apple Magic Mouse wows without wires*, Christian Science Monitor  

Full Review: ***Apple Magic Mouse makes***... CNET  

***Apple's 'Magic Mouse' Swaps Buttons for***... PC Magazine  

CNET UK - myGLOSS  

all 59 news articles »
## Result

### Some detected, classified and summarized events:

<table>
<thead>
<tr>
<th>Brand</th>
<th>Date</th>
<th>Event keywords</th>
<th>$Q_{pop}$</th>
<th>$Q_{sent}$</th>
<th>Summarizing tweet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apple</td>
<td>09/10/21</td>
<td>imac apple mouse</td>
<td>high</td>
<td>pos.</td>
<td>the new Apple Magic Mouse looks glorious! http://****</td>
</tr>
<tr>
<td>Apple</td>
<td>10/3/14</td>
<td>pre ipad order</td>
<td>high</td>
<td>neut.</td>
<td>@**** did you pre order your iPad?</td>
</tr>
<tr>
<td>Google</td>
<td>09/11/20</td>
<td>event os chrome</td>
<td>high</td>
<td>neut.</td>
<td>Google Chrome OS on Mixx: http://****</td>
</tr>
<tr>
<td>Google</td>
<td>10/3/12</td>
<td>direction map bike</td>
<td>high</td>
<td>neut.</td>
<td>RT @<strong><strong>: (BikeRadar) Cycling directions added to Google Maps in US http://</strong></strong> #procycling</td>
</tr>
<tr>
<td>Microsoft</td>
<td>09/10/1</td>
<td>security essentials</td>
<td>high</td>
<td>pos.</td>
<td>Microsoft Security Essentials is now available - free AntiVirus/Malware protection - http://****</td>
</tr>
</tbody>
</table>

### Validation:

- **Apple Store bracces for iPad pre-order blitz** - CSMonitor.com
  Christian Science Monitor - Mar 12, 2010
  The online Apple Store went down for several hours early Friday morning, as Apple updated the site to accommodate the iPad pre-order rush. ...
  Apple iPad Pre-Order Pre-Sale! LALATE
  iPad Pre-Order Guesses Range from 15K... Mac Observer
  Apple iPad preorders are a go, limit... VentureBeat
  i4u.com - i4u.com
  all 83 news articles »
## Result

### Some detected, classified and summarized events:

<table>
<thead>
<tr>
<th>Brand</th>
<th>Date</th>
<th>Event keywords</th>
<th>$Q_{\text{pop}}$</th>
<th>$Q_{\text{sent}}$</th>
<th>Summarizing tweet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apple</td>
<td>09/10/21</td>
<td>imac apple mouse</td>
<td>high</td>
<td>pos.</td>
<td>the new Apple Magic Mouse looks glorious! http://****</td>
</tr>
<tr>
<td>Apple</td>
<td>10/3/14</td>
<td>pre ipad order</td>
<td>high</td>
<td>neut.</td>
<td>@**** did you pre order your iPad?</td>
</tr>
<tr>
<td>Google</td>
<td>09/11/20</td>
<td>event os chrome</td>
<td>high</td>
<td>neut.</td>
<td>Google Chrome OS on Mixx: http://****</td>
</tr>
<tr>
<td>Google</td>
<td>10/3/12</td>
<td>direction map bike</td>
<td>high</td>
<td>neut.</td>
<td>RT @<strong><strong>: (BikeRadar) Cycling directions added to Google Maps in US http://</strong></strong> #proycling</td>
</tr>
<tr>
<td>Microsoft</td>
<td>09/10/1</td>
<td>security essentials microsoft</td>
<td>high</td>
<td>pos.</td>
<td>Microsoft Security Essentials is now available - free AntiVirus/Malware protection - http://****</td>
</tr>
</tbody>
</table>

### Validation:

... You Need to Know About Google's Chrome OS | Lance...
PC Magazine - Nov 19, 2009
By Lance Ulanoff Before today, Google's Chrome Operating System was an exciting idea that offered far more questions than answers. That changed today. ...
PCWorld Google Chrome OS Unveiled... PCWorld
5 Expectations for Google's Chrome OS... PC World
Five Reasons Google Chrome OS will... PC World
Telegraph.co.uk - Techtree.com
all 114 news articles »
## Result

### Some detected, classified and summarized events:

<table>
<thead>
<tr>
<th>Brand</th>
<th>Date</th>
<th>Event keywords</th>
<th>(Q_{\text{pop}})</th>
<th>(Q_{\text{sent}})</th>
<th>Summarizing tweet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apple</td>
<td>09/10/21</td>
<td>imac apple mouse</td>
<td>high</td>
<td>pos.</td>
<td>the new Apple Magic Mouse looks glorious! http://****</td>
</tr>
<tr>
<td>Apple</td>
<td>10/3/14</td>
<td>pre ipad order</td>
<td>high</td>
<td>neut.</td>
<td>@**** did you pre order your iPad?</td>
</tr>
<tr>
<td>Google</td>
<td>09/11/20</td>
<td>event os chrome</td>
<td>high</td>
<td>neut.</td>
<td>Google Chrome OS on Mixx: http://****</td>
</tr>
<tr>
<td>Google</td>
<td>10/3/12</td>
<td>direction map bike</td>
<td>high</td>
<td>neut.</td>
<td>RT @<strong><strong>: (BikeRadar) Cycling directions added to Google Maps in US http://</strong></strong> #procycling</td>
</tr>
<tr>
<td>Microsoft</td>
<td>09/10/1</td>
<td>security essentials</td>
<td>high</td>
<td>pos.</td>
<td>Microsoft Security Essentials is now available - free AntiVirus/Malware protection - http://****</td>
</tr>
</tbody>
</table>

### Validation:

Google Maps Bike Directions Roll Out at National Bike Summit...
Tonic - Mar 11, 2010
Wired, PC World and Bikeradar.com are reporting that Google Maps now includes a bike path feature that gives directions specifically tailored to those...
Inside Google Maps' New Cycling Features... BikeRadar.com
Google Maps adds biking-route feature AZ Central.com
Help Us Review Google Maps for Bikes [... Wired News
CCTV - Techtree com
all 10 news articles »
### Experimental evaluation

## Result

### Some detected, classified and summarized events:

<table>
<thead>
<tr>
<th>Brand</th>
<th>Date</th>
<th>Event keywords</th>
<th>$Q_{pop}$</th>
<th>$Q_{sent}$</th>
<th>Summarizing tweet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apple</td>
<td>09/10/21</td>
<td>imac apple mouse</td>
<td>high</td>
<td>pos.</td>
<td>the new Apple Magic Mouse looks glorious! http://****</td>
</tr>
<tr>
<td>Apple</td>
<td>10/3/14</td>
<td>pre ipad order</td>
<td>high</td>
<td>neut.</td>
<td>@**** did you pre order your iPad?</td>
</tr>
<tr>
<td>Google</td>
<td>09/11/20</td>
<td>event os chrome</td>
<td>high</td>
<td>neut.</td>
<td>Google Chrome OS on Mixx: http://****</td>
</tr>
<tr>
<td>Google</td>
<td>10/3/12</td>
<td>direction map bike</td>
<td>high</td>
<td>neut.</td>
<td>RT @<strong><strong>: (BikeRadar) Cycling directions added to Google Maps in US http://</strong></strong> #procycling</td>
</tr>
<tr>
<td>→ Microsoft</td>
<td>09/10/1</td>
<td>security essentials microsoft</td>
<td>high</td>
<td>pos.</td>
<td>Microsoft Security Essentials is now available - free AntiVirus/Malware protection - http://****</td>
</tr>
</tbody>
</table>

### Validation:

![Google News](https://via.placeholder.com/150)
Thanks!