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

Table of Contents

MACHINE
LEARNING
LAB

Twitter

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

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?

Useful information?

Can we automatically notify Google about these users who like the new UI?

Table of Contents

MACHINE
LEARNING
LAB

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*)

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

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

Totally unsupervised!

Example

<div align="center">

topic $=$ google

</div>

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!`

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!

Premise

- 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 comon acronyms and abbreviation
- remove english stop-words
- stemming

Features

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:

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

- $f_w^{\mathrm{now}} \to$ current (last 3 days) relative frequency
- $f_w^{\mathrm{hist}} \to$ historic (last 14 days) relative frequency
- select the words with highest b_w

MACHINE
LEARNING
LAB

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_{pop} in $\{\texttt{low}, \texttt{medium}, \texttt{high}\}$
- sentiment polarity as Q_{sent} in $\{\texttt{positive}, \texttt{neutral}, \texttt{negative}\}$

starting from the *event corpus* T^*

Classification: popularity

1. compare current (last 3 days) T^* relative size vs. past sizes
2. set:
   - $Q_{\mathrm{pop}} = \mathtt{high}$, if current size is the largest
   - $Q_{\mathrm{pop}} = \mathtt{medium}$, if in top 50%
   - $Q_{\mathrm{pop}} = \mathtt{low}$, 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 ($2000 \rightarrow 50$), 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 $= Q_{\mathrm{sent}}$

Table of Contents

MACHINE
LEARNING
LAB

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

Topics

Three popular brands:

Brand	Keywords	Tweets	T./day
Apple	`apple, iphone, ipad, ipod, ios, mac, macintosh, macbook`	53,680	272
Google	`google, android, chrome`	29,073	148
Microsoft	`microsoft, explorer, msn, windows, winxp, vista, xbox`	20,897	106

Result

Some detected, classified and summarized events:

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

Result

Some detected, classified and summarized events:

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

Validation:

$\longrightarrow$

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:

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

Validation:

Google

News

$\longrightarrow$

Apple Store braces 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 »

MACHINE
LEARNING
LAB

Result

Some detected, classified and summarized events:

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

Validation:

Google

News

$\longrightarrow$

... 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 »

MACHINE
LEARNING
LAB

Result

Some detected, classified and summarized events:

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

Validation:

Google

News

$\longrightarrow$

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 »

Result

Some detected, classified and summarized events:

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

Validation:

Google

News

$\longrightarrow$

The Hindu : Sci-Tech / Internet : **Microsoft Security...**
The Hindu - Sep 30, 2009
The release of free **Microsoft Security Essentials** can potentially heat up the growing anti-virus market. Photo: PVSivakumar Amy Barzdukas has one of the ...
Microsoft Security Essentials shakes up... Register
Microsoft Security Essentials 1.0... PC Magazine
Testing out **Microsoft Security Essential...** ITworld.com
ZDNet Asia - InformationWeek
all 54 news articles »

MACHINE
LEARNING
LAB

Thanks!