

# 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

## 1 Scenario

## 2 Approach

- Preprocessing
- Event detection
- Event classification
- Event summarization

## 3 Experimental evaluation



# 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?

```

93525760,Fri Feb 24 13:42:25 +0000 2012,@iwa_chann お疲れさま V 確かに夜は寒い(ToI)私、汗がいて冷えた(笑)2000くらいじゃないのか
83078784,Fri Feb 24 13:42:21 +0000 2012,RT @WatsonRadcliffe: RT if you wanna #FridayLove :),76958485,Anna Toml
97715968,Fri Feb 24 13:42:25 +0000 2012,@Joeaffolter go for it,294619607,Samuel Kemp,en,,326,,,,,,,,,,,,,
93525761,Fri Feb 24 13:42:25 +0000 2012,fallen in love with @lilygandmaguire !xoxoxox,359003979,katiee gregory
89319169,Fri Feb 24 13:42:25 +0000 2012,@ZinClaudio para tailandia me exigian la de la fiebre amarilla pero xq
8683392,Fri Feb 24 13:42:25 +0000 2012,Ice cream remix fukkiinnngg helllllllllllll,157379381,Discarda,en,Wileho
27067904,Fri Feb 24 13:42:25 +0000 2012,RT @hawks19morifuku: お疲れ様です☆多今日も平日なのにたくさんのファンの皆さんにビックリで
01910272,Fri Feb 24 13:42:25 +0000 2012,http://t.co/Jp6x5idV работа в орехово-зуюво вакансии охранником,466704
93521664,Fri Feb 24 13:42:25 +0000 2012,"@ahardhaug Må gjerne bygge flere boliger i Oslo. Spørsmålet er om det
27067905,Fri Feb 24 13:42:25 +0000 2012,にゃんにゃん。にゃんに書こうかにゃん。シリーズもので、お誕生日書きたい?? 目の前に、丟能パロあっ
27067995,Fri Feb 24 13:42:25 +0000 2012,Very happy for the new #google maps ui - like it!,4147432073,Joshua Sm
93509377,Fri Feb 24 13:42:25 +0000 2012,RT @GuyKawasaki: Steampunk insects add mechanical edge to creep factor
93509378,Fri Feb 24 13:42:25 +0000 2012,"Hah, jaruuuuk. . RT @iitott: Sya mamanya apika, kq gk izin dlu ngjak
0286592,Fri Feb 24 13:42:25 +0000 2012,FREE Sari Caste http://t.co/cmiFixFK. #Fear #desperation #murder Manas
25889280,Fri Feb 24 13:42:23 +0000 2012,#FF Strudy: @kurtworkholding @Automation_AGV @Strategicpms @BastianSol
89315072,Fri Feb 24 13:42:25 +0000 2012,"RT @thedouglaas_: Ironia é alguém que assiste Big Brother falar: ""Va
25881088,Fri Feb 24 13:42:23 +0000 2012,Vanavond pelikaan with best friend x,164384561,chinouk,en,Via BlackBer
0290688,Fri Feb 24 13:42:26 +0000 2012,إخواني باقي بطايله ليوم الأحد الى حاب يحجز يراسلني وهدى صورتها, http://t
93513473,Fri Feb 24 13:42:25 +0000 2012,Zo maar even naar de kapper,214033047,jouke waaksma,en,,6688,,,,,,,,,,,,,
4489088,Fri Feb 24 13:42:25 +0000 2012,He Me With The Less Than Symbol && Number 3 #&t!;3,326932705,CoutureIm

```

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



# 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

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!



# 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  $\rightarrow$  a pre-compiled set of keywords
  - we focused on brands (google, apple, microsoft)
  - topic= apple  $\rightarrow$  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  $w$ :

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

- $f_w^{\text{now}}$  → current (last 3 days) relative frequency
- $f_w^{\text{hist}}$  → 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:
  - $Q_{\text{pop}} = \text{high}$ , if current size is the largest
  - $Q_{\text{pop}} = \text{medium}$ , if in top 50%
  - $Q_{\text{pop}} = \text{low}$ , 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  $\rightarrow$  50), using a linear model
- 2 set  $Q_{\text{sent}}$  to the sentiment of the largest portion  $T_{\text{sent}}^*$  of  $T^*$



# Summarization

Goal: select one tweet which represents the event

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

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



# Table of Contents

## 1 Scenario

## 2 Approach

- Preprocessing
- Event detection
- Event classification
- Event summarization

## 3 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





# 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! <a href="#">http://****</a>
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: <a href="#">http://****</a>
Google	10/3/12	direction map bike	high	neut.	RT @****: (BikeRadar) Cycling directions added to Google Maps in US <a href="#">http://****</a> #proccycling
Microsoft	09/10/1	security essentials microsoft	high	pos.	Microsoft Security Essentials is now available - free AntiVirus/Malware protection - <a href="#">http://****</a>



## 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! <a href="#">http://****</a>
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: <a href="#">http://****</a>
Google	10/3/12	direction map bike	high	neut.	RT @****: (BikeRadar) Cycling directions added to Google Maps in US <a href="#">http://****</a> #procyling
Microsoft	09/10/1	security essentials microsoft	high	pos.	Microsoft Security Essentials is now available - free AntiVirus/Malware protection - <a href="#">http://****</a>

Validation:



News



[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 of 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! <a href="http://****">http://****</a>
→ 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: <a href="http://****">http://****</a>
Google	10/3/12	direction map bike	high	neut.	RT @****: (BikeRadar) Cycling directions added to Google Maps in US <a href="http://****">http://****</a> #procyling
Microsoft	09/10/1	security essentials microsoft	high	pos.	Microsoft Security Essentials is now available - free AntiVirus/Malware protection - <a href="http://****">http://****</a>

Validation:



News



[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 »](#)



# 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! <a href="#">http://****</a>
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: <a href="#">http://****</a>
Google	10/3/12	direction map bike	high	neut.	RT @****: (BikeRadar) Cycling directions added to Google Maps in US <a href="#">http://****</a> #proccycling
Microsoft	09/10/1	security essentials microsoft	high	pos.	Microsoft Security Essentials is now available - free AntiVirus/Malware protection - <a href="#">http://****</a>

Validation:



News



[... 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:

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! <a href="#">http://****</a>
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: <a href="#">http://****</a>
→ Google	10/3/12	direction map bike	high	neut.	RT @****: (BikeRadar) Cycling directions added to Google Maps in US <a href="#">http://****</a> #proccycling
Microsoft	09/10/1	security essentials microsoft	high	pos.	Microsoft Security Essentials is now available - free AntiVirus/Malware protection - <a href="#">http://****</a>

Validation:



News



[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:



News



[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 »](#)



Thanks!

