

Twitter NLP

Annie Louis

CIS 530 Lecture
05 Dec, 2011

Twitter: A LARGE source of public opinion



twitter

 **Bill Gates** 
@BillGates Seattle, WA
Sharing cool things I'm learning through my foundation work and other interests...
<http://www.thegatesnotes.com>

  Text follow **BillGates** to 40404 in the United States

Tweets Favorites Following Followers Lists ▾

 **BillGates** Bill Gates
Amazing: Microsoft Research using email spam fighting principles to help in HIV vaccine search: b-gat.es/uullav
#WAD11 #WorldAIDSDay
1 Dec

 **BillGates** Bill Gates
HappyBirthday to my dad! 86 & going full bore, you are an invaluable partner in our work at the @gatesfoundation. Thanks Dad for everything!
30 Nov

 **gatesfoundation** Gates Foundation  by BillGates
FACT: HIV treatment medicines have saved 2.5 million lives since 1995. #WorldAIDSDay #WAD11
28 Nov



twitter

Results for **bangkok floods**

Tweets · Top ▾ [Refine results »](#)

 **bedlamfury** Zashnain Zainal
Flood situation ends in 11 districts bit.ly/vcFxyo #Bangkok #floods #Thailand
13 minutes ago

 **PersonalThai** Personal Thailand
Governor new year gift no floods in Bangkok is.gd/rXuyD5 thaidflood
2 hours ago

 **DarlingOne** - P o M -
Thai floods crisis: Updated info for tourists | CNGo.com
cngo.com/bangkok/life/t...
2 hours ago

 **PairsonnalitesA** PAIRSonnalités | AS
| NEWSBLOG | Bangkok devastation: Bangkok devastation: The floods killed more than 600 people and caused...
dlvr.it/yvCmS | #THPR |

Very diverse/timely content



The image shows a screenshot of a Twitter profile for Bill Gates. The profile header includes the Twitter logo, a search bar, and the user's name 'Bill Gates' with a verified badge. Below the name is the handle '@BillGates', location 'Seattle, WA', and a bio: 'Sharing cool things I'm learning through my foundation work and other interests...'. A website link 'http://www.thegatesnotes.com' is also present. There is a 'Follow' button and a text follow option: 'Text follow BillGates to 40404 in the United States'. Below the profile are navigation tabs for 'Tweets', 'Favorites', 'Following', 'Followers', and 'Lists'. Three tweets are visible:

- BillGates** Bill Gates: Amazing: Microsoft Research using email spam fighting principles to help in HIV vaccine search: b-gat.es/uullav #WAD11 #WorldAIDSDay (1 Dec)
- BillGates** Bill Gates: HappyBirthday to my dad! 86 & going full bore, you are an invaluable partner in our work at the [@gatesfoundation](https://twitter.com/gatesfoundation). Thanks Dad for everything! (30 Nov)
- gatesfoundation** Gates Foundation  by BillGates: FACT: HIV treatment medicines have saved 2.5 million lives since 1995. #WorldAIDSDay #WAD11 (28 Nov)

- Different users: Celebrities, businesses, groups, general people
- Variety in tweet stream: different issues/topics/entities are discussed
- Timely & short: realtime updates

Today's lecture: Some twitter-based NLP tasks

1) Sentiment analysis

- Example application: predicting poll results
- Building a classifier -- what features work for tweets

2) Summarization

- Query results by themes
- Summaries explaining twitter trends
- Consumer opinion from business tweets

3) Event extraction

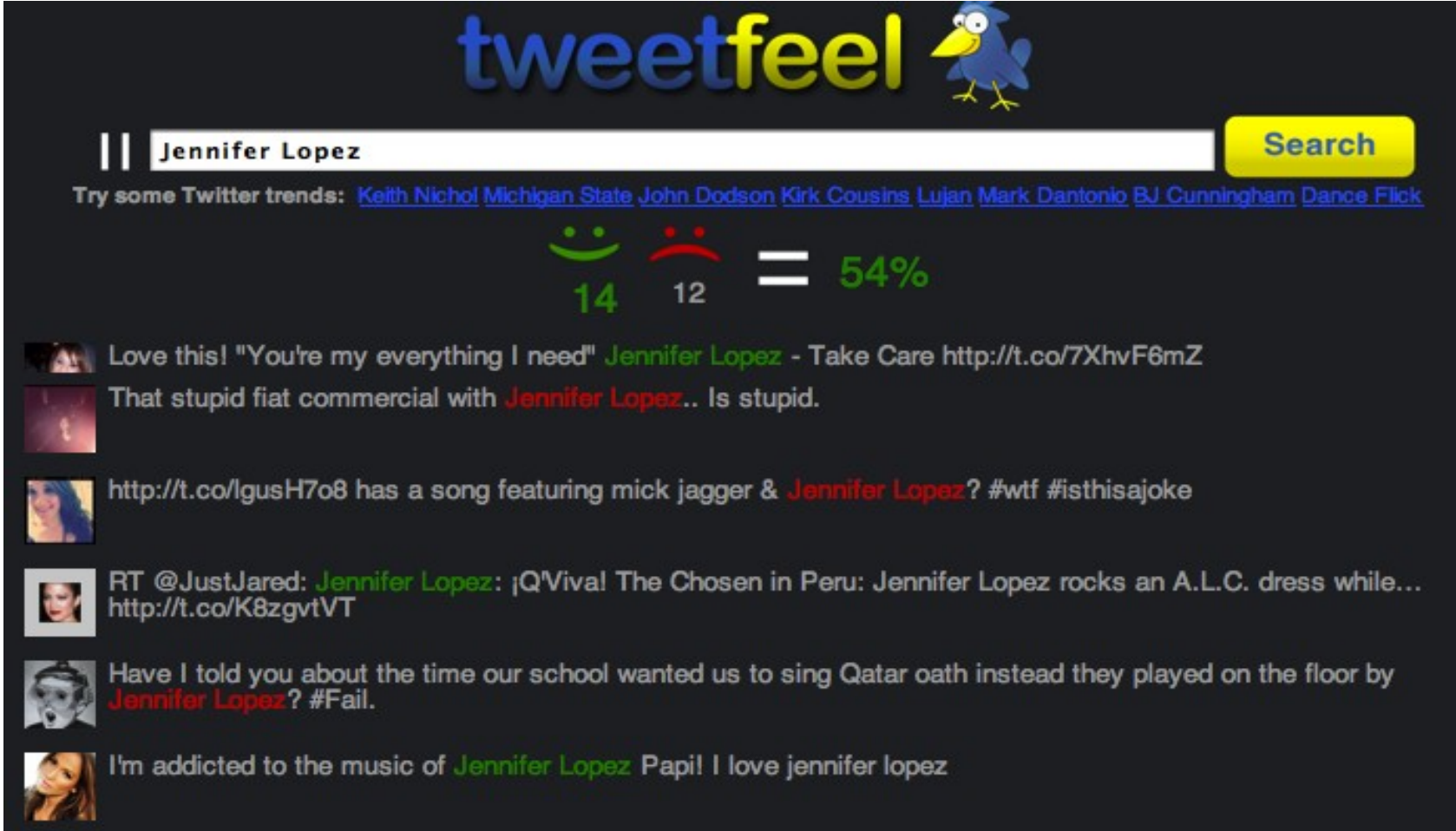
- Identifying trending topics

4) Understanding twitter communication

- How businesses use twitter
- User response to celebrity events

1. Sentiment Analysis on Twitter

Sentiment tagging for twitter search: A straightforward useful application



The screenshot shows the Tweetfeeling website interface. At the top, the logo "tweetfeeling" is displayed in blue and yellow, accompanied by a blue bird icon. Below the logo is a search bar containing the text "Jennifer Lopez" and a yellow "Search" button. Underneath the search bar, there is a section for "Try some Twitter trends:" with several blue links: [Keith Nichol](#), [Michigan State](#), [John Dodson](#), [Kirk Cousins](#), [Lujan Mark](#), [Dantonio BJ](#), [Cunningham](#), [Dance](#), and [Flick](#).

Below the trends section, a sentiment analysis summary is shown: a green smiley face icon with the number "14" below it, a red frowny face icon with the number "12" below it, an equals sign, and the percentage "54%".

The main content area displays a list of tweets with sentiment tagging. Each tweet is preceded by a small profile picture. The text of the tweets is as follows:

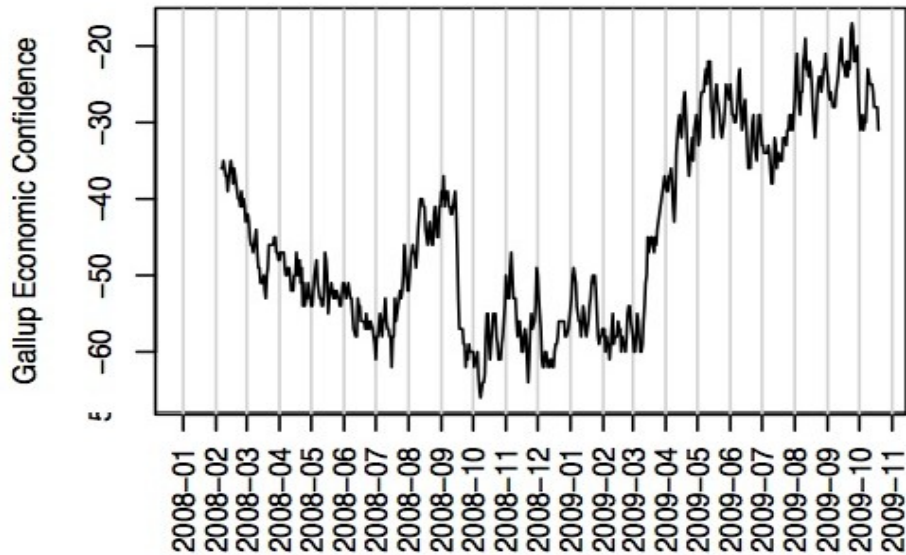
- Love this! "You're my everything I need" **Jennifer Lopez** - Take Care <http://t.co/7XhvF6mZ>
- That stupid fiat commercial with **Jennifer Lopez**.. Is stupid.
- <http://t.co/lgusH7o8> has a song featuring mick jagger & **Jennifer Lopez**? #wtf #isthisajoke
- RT @JustJared: **Jennifer Lopez**: ¡Q'Viva! The Chosen in Peru: Jennifer Lopez rocks an A.L.C. dress while... <http://t.co/K8zgvtVT>
- Have I told you about the time our school wanted us to sing Qatar oath instead they played on the floor by **Jennifer Lopez**? #Fail.
- I'm addicted to the music of **Jennifer Lopez** Papi! I love jennifer lopez

Volume of sentiment and aggregate across users: even more useful

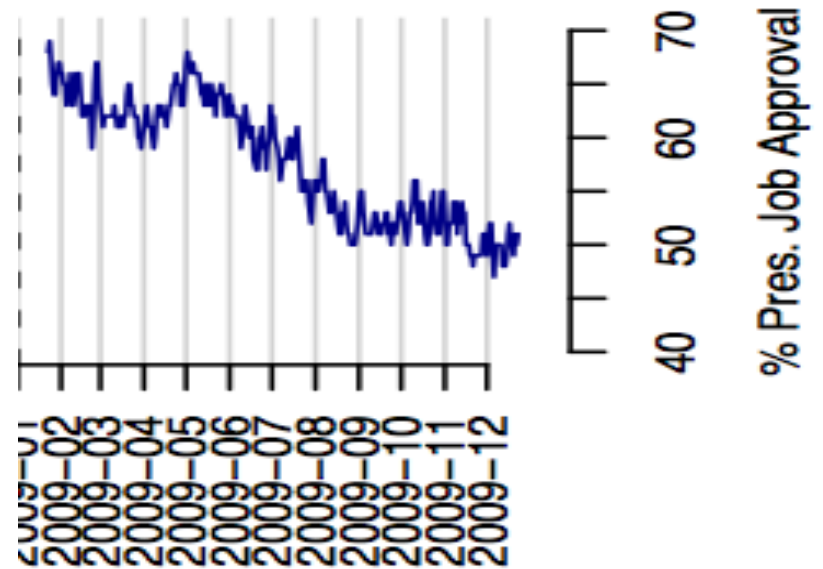
- Predicting consumer sentiment
- Predict stock prices
- Identifying events, disasters

Example application: Predicting results of opinion polls [O' Conner et al. 2010]

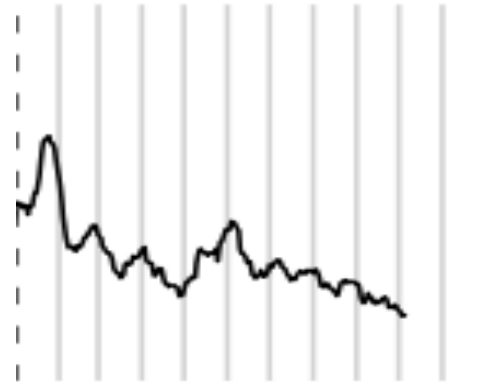
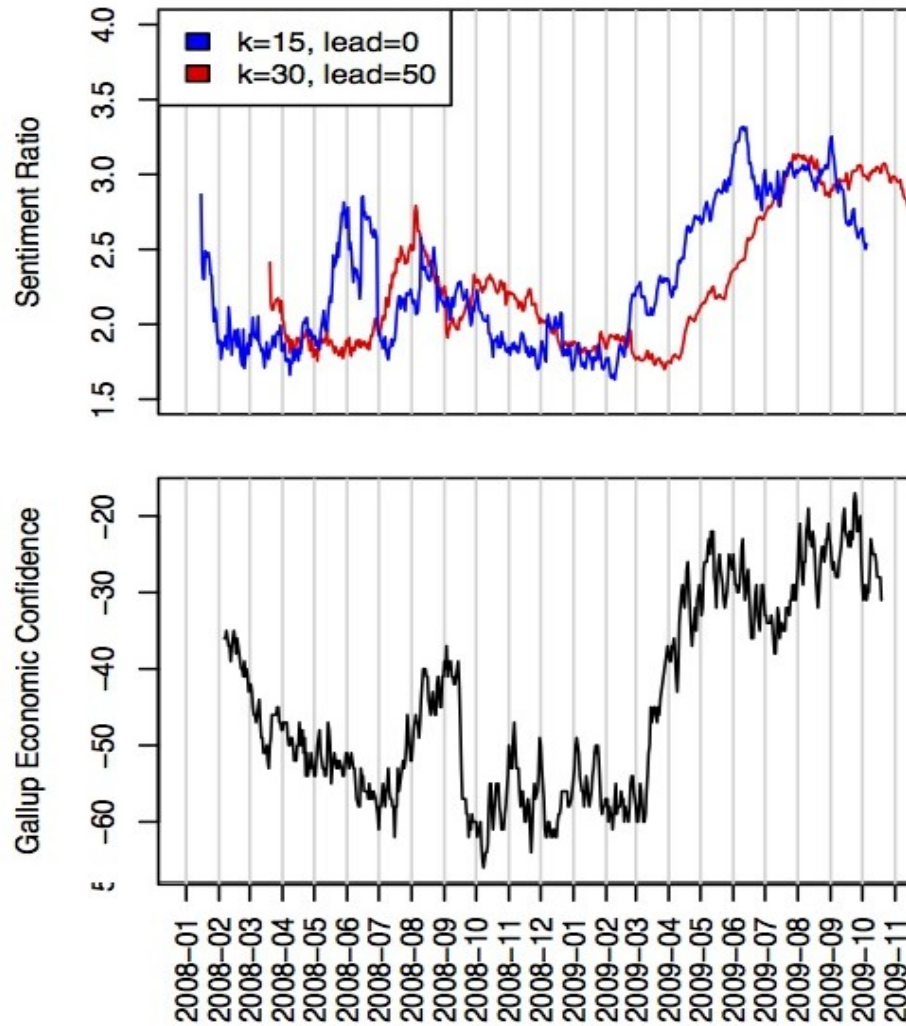
Consumer confidence poll results 2008 & 2009



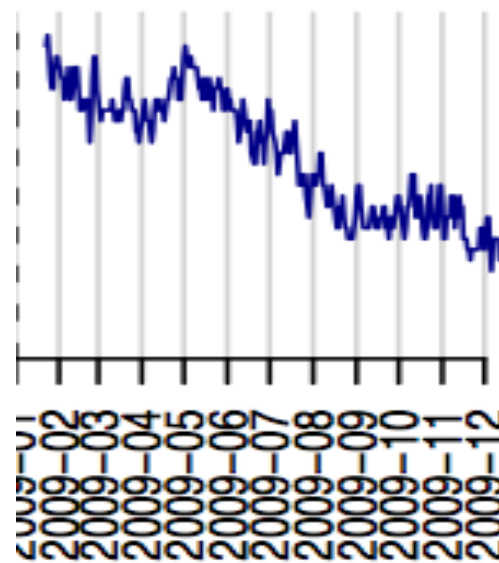
Presidential job approval for Obama in 2009



Example application: Predicting results of opinion polls [O' Conner et al. 2010]



Sentiment Ratio
on the topic
←



40 50 60 70
% Pres. Job Approval

Sentiment metric they used

- Tweet retrieval
 - Those matching 'job', 'economy' for consumer confidence
 - 'Obama' for presidential approval
- Sentiment Ratio
 - A message = positive if it has a positive word, negative if it has a negative word
 - $SR(x_t) = \#positive\ messages / \#negative\ messages$
- Smooth the ratio for over previous d days
 - $MA_t = 1/d * (x_{t-d+1} + x_{t-d+2} + \dots + x_t)$

Classifying the sentiment of business tweets [At MSR, 2011]

- Tweets which mention a company name



- 3 way
 - Positive/negative/neutral

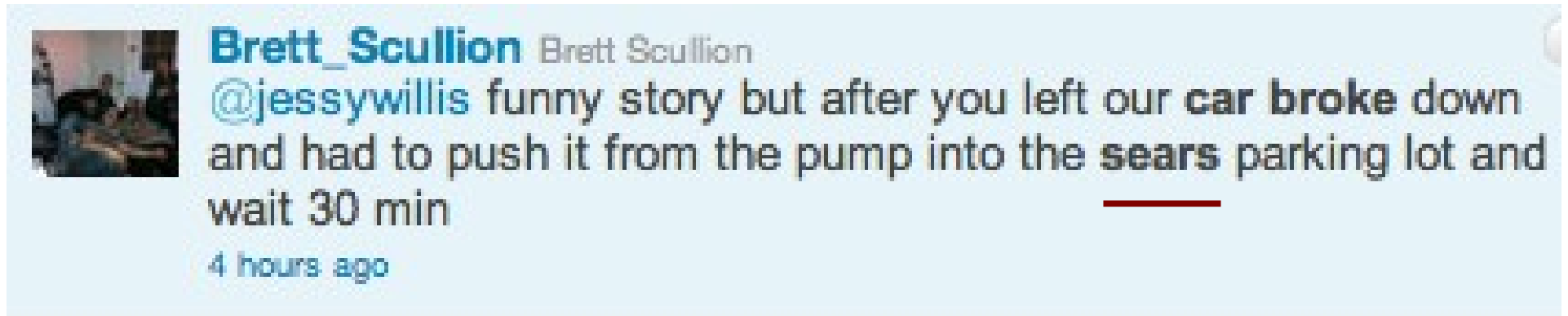
Word-based features

- Words
 - Unigrams, bigrams
- Sentiment lexicon
 - Counts of positive/negative words
 - Positive/negative word at the end of the tweet
 - #tag has a positive/negative word
 - Similar features with a slang dictionary: cool, lovin → positive, smh, grrr → negative

Realization features

- Part of speech and punctuation
 - Counts of different POS tags/ punctuations
 - Length of tweet
 - Presence of fully capitalized words
 - Multiple ! / ? in sequence

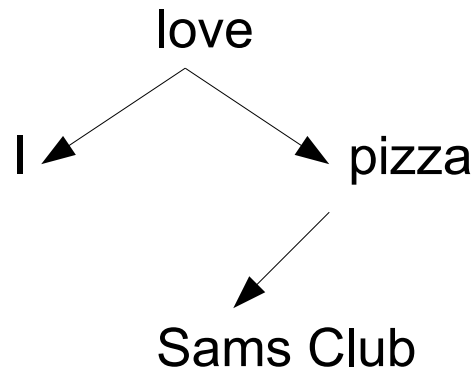
Target dependence is important



- When a tweet is not about the company
 - It is NEUTRAL regardless of any other sentiment in the tweet

Target-based features

- Company name near positive/negative words
 - In a window of size 2
- Dependency parse based features



- Company name linked to main verb
- Modifiers / siblings / parent / grand-parent of company name: POS tag and sentiment

Accuracy of the classifier

- Manually annotated 2470 tweets covering 18 companies
 - Positive: 23%, Negative: 28%, Neutral: 49%
- MaxEnt classifier
 - Majority baseline = 50%

Feature set	Accuracy
ngram	62.8
ngram + dictionary	64.6
ngram + dict + punct/pos	64.8
ngram + dict + punct/pos + target	82.1

2. Summarizing Tweets

Summarizing query results by themes – Tweet Motif



tweetmotif
discover twitter topics

(tweet this) (about) (tips)

What are people saying about...

#WolframAlpha

Search

related themes

over the last 50 minutes

#wikipedia	wolfram alpha
wolframalpha.com/	computational
fructose in onion	computational
with your input	knowledge
google killer	engine
wolfram	via @time
wolfram alpha	been alive
@time	bit.ly/sa6be
days old	got the number
meaning of life	wrong answer
i have tried	queries failed
on the day	eg
was born	i was born
engine	search engine
#fail	input

tweets by theme

“queries failed”

: not finding #WolframAlpha all I had hoped, data-driven queries failed every time (“ave # job applicants interviewed to fill position”) *peterjwolfgang*

So far #WolframAlpha has failed for the two queries that I have tried that weren't a play test #fail *marcad* show 1 similar tweet >

(drilldown 1)

“google killer”

#WolframAlpha is going to be huge, but it's not a google killer, and most people will never use it *briggs1*

@charlesf11 Yeah, it's bugging me that people are calling it a Google killer. They're not the same animal. #wolframalpha #wikipedia #google

mattbramanti

#WolframAlpha is no Google killer, but it's the smartest calculator ever.

PrestonStahley

Algorithm for identifying themes

- Extract phrases → unigrams, bigrams, trigrams with some filtering
- Rank by distinctiveness

$$\frac{P(\textit{phrase} \mid \textit{tweet result set})}{P(\textit{phrase} \mid \textit{general tweet corpus})}$$

General tweet corpus = results for common words like 'the', 'of'...

- Merge very similar topics

Twitter trend summarization [Liu et al. 2011]

Trends:

Jordy Nelson

Jason Garrett

Terrance NEwman

Stephens-Howling

Donald Driver

#GETTINCRAZY

Cardinals Win

#TurntUp

#cowgirls

Double D

Why is this name trending?



JPTrey13 Jordan Poydras

Jordy Nelson.. wow

14 minutes ago



FantasyShrink John Paulsen

What a catch by Jordy Nelson.

14 minutes ago

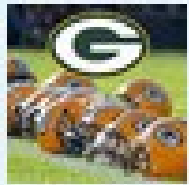


royal_flush2482 that dude Deshon

Jordy Nelson a beast

15 minutes ago

There are other more informative tweets on the topic



packers Green Bay Packers

WR **Jordy Nelson** catching passes right now at MetLife Stadium from Matt Flynn. yfrog.com/h7yv8syj

4 hours ago

- How to generate an automatic summary?

Gold standard summaries written by people



What the Trend™

Now Trending | [All Trends Today](#)

Why is Jordy Nelson popular right now?

0



Green Bay Packers quarterback Aaron Rodgers (12) gives a thumbs-up to wide receiver Jordy Nelson (87) after Nelson caught a 10-yard touchdown pass. (as of 10 months ago)

Define

Retweet

Flag as Inappropriate

Which text source would be the best input for this task?

- Original tweets on topic
 - Those matching the keyword given
- Normalized tweets
 - Better readability/ chance for summarization
- Linked web contents
 - More comprehensive, informative source on topic

Content quality of produced summaries

General Topics		R-1 F(%)
Input Source	Render	#Word
OrigTweets	Orig	30.21
	Norm	30.21
NormTweets	Norm	30.35
Web		25.07
Web + OrigTweets		30.44
Web + NormTweets		30.54

- Linked web text not found that useful yet
- Finding informative tweets is a good task

Readability of the summaries

	General	
	Tweet	Web
Gram.	3.13	3.42
NRedun.	3.93	4.64
Clarity	4.07	3.91
Focus	3.64	3.03

- Web text has better grammaticality and non-redundancy but lacks focus

Hashtag topics more difficult to summarize

- Human summaries are indicative and general. So hard to replicate.

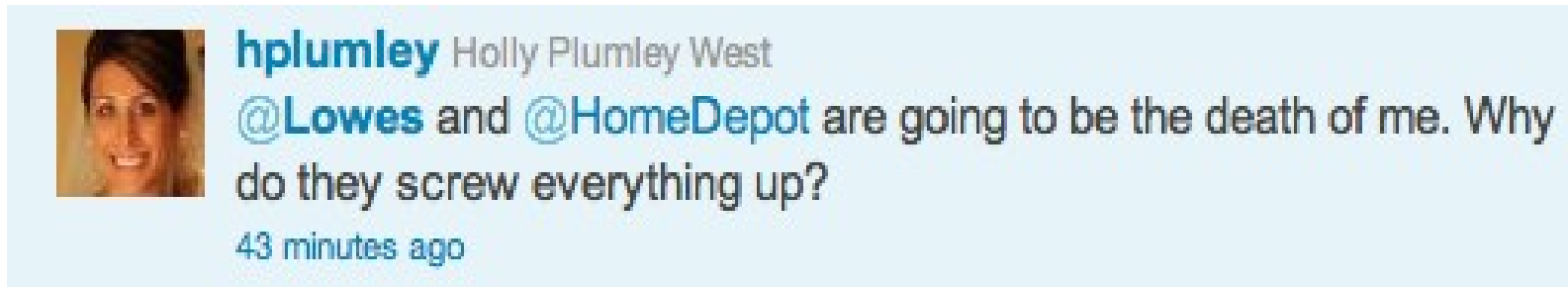
#2 #nevertrust

Updated definition needed. Old definition: People are tweeting why you shouldn't trust a person or a situation. ✓

- It would be nice to pick out a sample of interesting tweets or recommend to a user



Summarizing business-related tweets [MSR, 2011]



- Useful for company-analysts and consumers
- Input for summarization
 - Tweets matching a set of keywords for the company
 - Eg: Microsoft, Bill Gates, MS Office, Windows 7...
 - Obtained by crowd-sourcing

Types of business tweets

1. Postings of/related to news articles

- Wow! RT @nytimes: Sheila Bair, outgoing FDIC chair: "They should have let **Bear Stearns** fail." <http://.../>
- Glad we bailed out companies like **JPMorgan Chase** <http://.../>
#shameful

2. Comments on products/services

- Smh at **sams club** staff today, slow as eff!
- Thanks **JPMorgan Chase**...business accounts set up today.
#banking

Types of business tweets

3. About business aspects

- Sen. Rubio: we don't need new taxes, we need new taxpayers: Agreed, how about we start with GE and **Exxon/Mobil**.
- I feel like so many different companies are copying Axe's commercials. First DQ and now **Xfinity**. C'mon, originality plz!
- \$JPM - **JPMorgan Chase** Stock Analysis - volume down over 5%
- <http://www.dojinspace.com/stock-picks>

4. Non-specific comments

- I love **Sams Club**
- Uqhh Right Bout Now I Hatee **Xfinity** .,! Its So Stupid!!!!!!!

Types of business tweets

5. Posts from applications

- I'm at **JPMorgan Chase** in Lake Mary, FL <http://gowal.la/c/4zdNp>
- I just ousted Joshua B. as the mayor of Bank of America - Malvern on @foursquare! <http://4sq.com/bmAL0B>

6. Advertisements, jobs, deals

- Consultant- RI at **Bank of America** <http://.../> #programmer #dev
- FREE **Lowes** kid workshop TODAY – <http://.../>

7. No useful information about the company

- Walking around **Sams Club** Like a Boss xD
- Gonna put a fresh paint job on the walls of my room @ hm :) at Lowes trying na decide on a color

Nature of the summarization problem

- When collected for a short time period, often there is not much overlap in the tweets
 - Product reviews are archives and have much more recurring topics
- Companies are diverse
 - Goldman Sachs, Comcast, Walmart

Main idea: Learn possible business concepts in an offline process, once

- Some concepts are inherent to all businesses
 - Eg. people, assets, products, management, ...
- Automatically acquire them
 - From news articles, companies are heavily discussed in the news
 - Broad, semantic, good coverage
- Cluster tweets to these concepts
 - Rather than just word overlap

Step 1: Learn company-related words

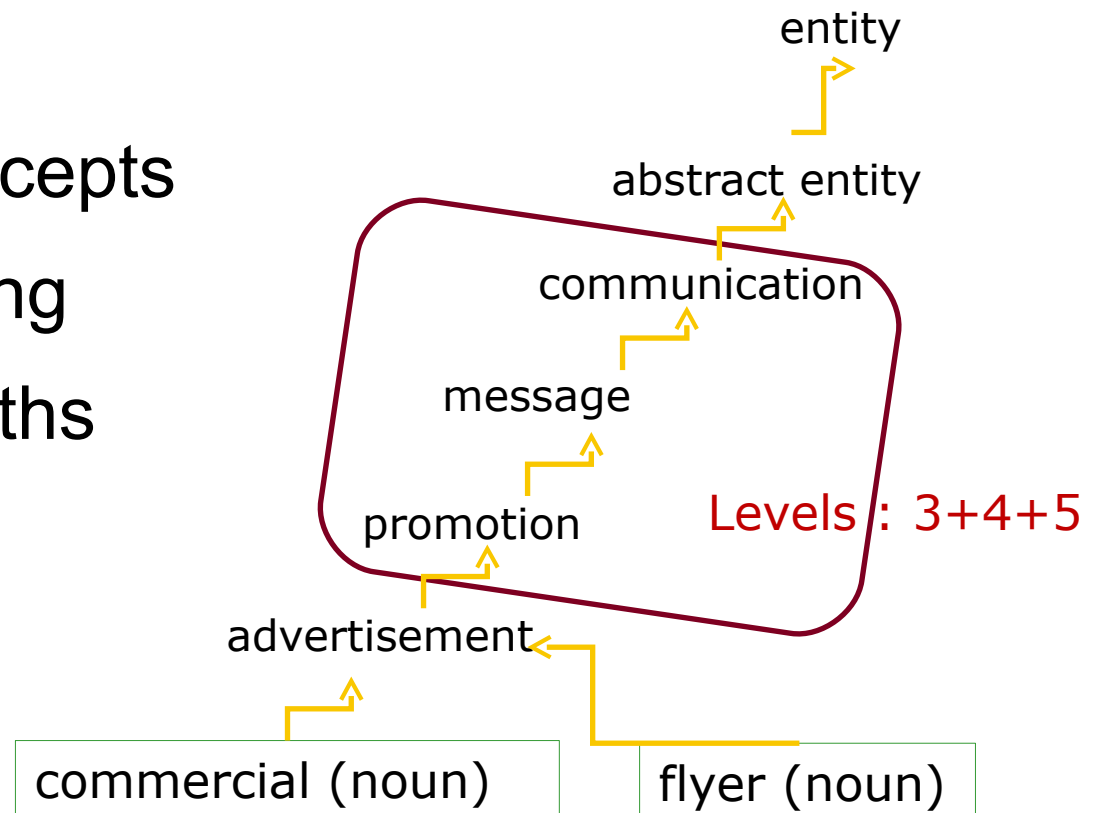
- Words related to businesses but not specific to any individual company
- From one year of New York Times news articles
 - About 100,000 articles
- Run a named entity tagger
 - Mark ORGANIZATION mentions in the business section articles as 'COMPANY'
 - Collect nouns in a window around 'COMPANY' mentions
 - Rank by mutual information with 'COMPANY'

Example 'business' words produced from the ranking

mergers
provider
stake
banking
acquisition
account
brand
appliance
carriers
liabilities
shortage
markets
analyst
food
shareholder

Step 2: Cluster the top business words into semantic concepts

- Increase coverage of company-related words
 - Get broader business-aspects
- Find more general concepts for top 2000 words using WordNet hypernym paths



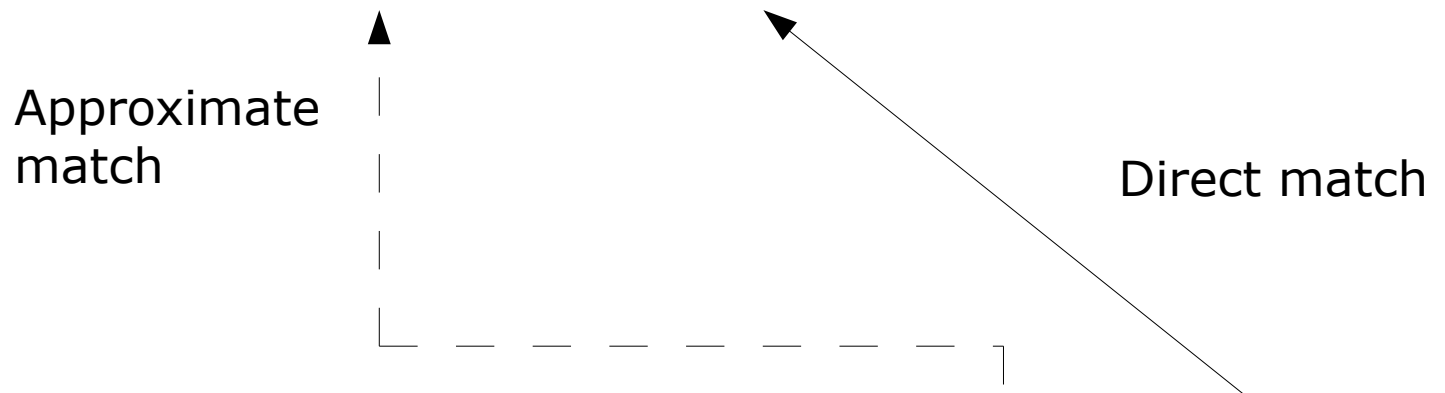
Example clusters of words

Words clustered	Covering Semantic Tag (WordNet)	Cluster name (manual)
merger, cancellation, banking, investment	human_activity	Activities
analyst, shareholder, executive, subscriber	animate_thing	People
manufacturer, firm, division, staff	organization	Organization
stake, equity, stocks, assets, share, sales	assets	Assets
semiconductor, shampoo, fiber, pesticide	material	Material/stuff
beer, candy, pizza, liquor, beverage	food stuff	Food

- 59 possible clusters
 - A cluster = semantic tag + prior company-words

Cluster the tweets to these pre-defined concepts

“PEOPLE” cluster = **animate_thing**: investor, customer, CEO, analyst...



@TPEconomy: Bank of America settlement with **investors** could cost thousands of distressed **borrowers** their homes

$$\text{score}(t, C) = |\text{direct matches}| * 0.8 + |\text{approx. matches}| * 0.2$$

$$\text{mapping}_{\text{words}}(C) = \bigcup_{t \in C} \{ \text{direct matches} \} \cup \{ \text{approx. matches} \}$$

Handle ties by assigning to the more popular cluster

- Once all the hypotheses are obtained
 - Each tweet maps to a max of 4-5 clusters
- Compute the most popular cluster list for this set of tweets
- When there is a tie in membership score, assign to the more popular of the clusters
 - Also helps assign similar tweets to the same cluster

Example tweet clusters: Bank of America

Assets:

Bank of America: I cannot get Bank of America to remove the PMI insurance from my **mortgage**. It has been 6 months

Stock Market Today – Bank of America (BAC) and JPMorgan (JPM) Biggest Dow Losers: Stock Market Today – Bank of ... <http://bit.ly/pzSBie>

Bank of America Merrill Lynch Previews **Earnings** From Time Warner Cable <http://bit.ly/oSgP0m>

Material/stuff:

Bank of America never sent me my bank **card** smh

I have to go to CityBank to cash a check, then pay the credit **card** for Bank of America, then deposit the rest into Chase ~Dad lmaoo #great

I Found My Bank Of America **Card** , An I Got Hella \$\$\$\$:)

People:

Bank Of America Appoints Terry Laughlin As Chief Risk **Officer** - RTT News <http://twlv.net/OKd6he>

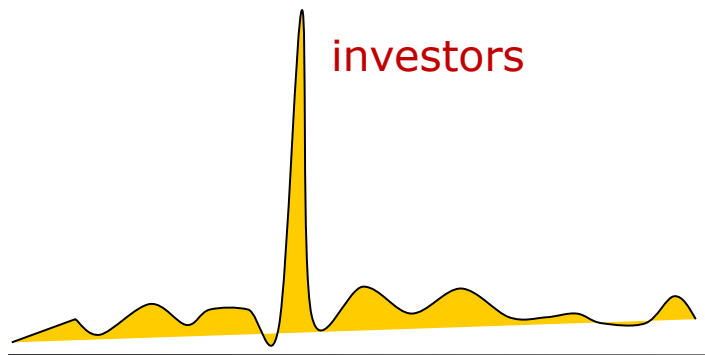
This **dude** is talking mighty loud about how he owes Bank of America \$2000

Customers hate Bank of America, repelled by excessive fees, impenetrable call centers, and foreclosure horror stories. But they love UPS

Cluster ranking

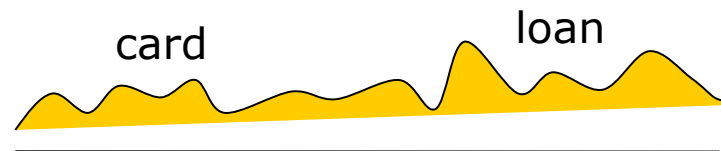
- Size was not informative enough
- Use bias in cluster content instead

More interesting cluster



Bank of America
"People"

Less interesting cluster



Bank of America
"Products"

Ranking metric

- Entropy of the mapping words for the cluster

$$E(C_j) = - \sum p(w_i) \log p(w_i) \quad w_i \text{ is a mapping word for cluster } C_j$$

- Incorporate sentiment into the probability value

$$\text{count}(w_i) = \sum \text{sentimentvalue}(S_k) \quad \text{where } S_k \text{ is a tweet } \in C_j$$

$$\text{sentimentvalue}(S_k) = |\text{positive confidence} - \text{negative conf}|$$

- Weight the entropy by cluster size

$$E_{adj}(C_j) = \left(1 - \frac{|C_j|}{\sum |C_k|}\right) E(C_j)$$

Cluster summarization

- Faceted
 - Map words ordered by probability forms the headline
 - For top 2 headline words, summarize the tweets that match those words
- 3 sentences for each cluster
 - 2 sentences containing first headword, 1 for the second
- Sentence ranking
 - Average probability of words
 - Probability uses the sentiment value just as in ranking

Example summary: Wells Fargo

(+57) People: customer, banker, employee, man

- Wells Fargo holding a daylong seminar to help customers having problems with mortgage <http://../>
- I swear Wachovia care more about customer service than anything
- Wells Fargo decided to exit reverse mortgages after federal officials insisted it foreclose on elderly customers <http://../>

(+14) Amount: money, cash, fund

- Wells Fargo act like they are mad about their little money
- Wells Fargo lost cash off my card. #smh I am sewing someone

Other approaches

Top positive/negative tweets

(+) I love [Wells Fargo](#). They let you customize your debit card!

(-) [Wells Fargo](#) has pissed me off one too many times. Time to move my money

Frequent tweets (motivated by retweets)

(+19) Banks financing Mexico drug gangs admitted in [Wells Fargo](#) deal

(+14) [Wells Fargo](#) to pay \$125 million in mortgage suit <http://t.co/>

Average probability ranking (no division of any sort)

- Wachovia banks to become [Wells Fargo](#)
- [Wells Fargo](#), Goldman Sachs and all other banks don't come close

Evaluating the summaries

- Summary length: 20 tweets
- 6 companies
 - Wells Fargo, Johnson & Johnson, LinkedIn, Northrop Grumman, Lowes, Nokia
- 14 judges, pairwise preferences
 - Which summary was more informative?
 - Which one was more interesting?
 - Would be more useful if you were an analyst?

Results

- Cluster summary paired with each baseline
 - 18 pairs for each question * 3 = 54
- Cluster preferred: 30 (55.5%)
 - Almost always preferred over 'no clustering' summary
 - But 'positive/negative' and 'frequency' summaries are also noted as useful
- Problems:
 - People expect fewer clusters in summary
 - Quality heavily depends on the tweet stream as well

3. Information Extraction from Tweets

First story detection [Petrovic et al. 2010]

- How to predict a trending topic?
- Sometimes, twitter precedes newswire

Approach

- For each tweet, (consider them coming as a stream)
 - Find most similar previous tweet
 - If similarity $>$ threshold, add to thread of matching tweet
 - Else create new thread
- Track which threads are growing very fast
- Needs techniques to overcome processing and space constraints

4. Understanding Twitter

The functions of business accounts

[Popescu & Jain, 2011]



Bank of America ✓

@BofA_Help

We are official Bank of America Twitter reps, here to help, listen & learn from our customers. Tweet with Bank of America reps 8-8 ET Mon-Fri.

<http://go.bofa.com/BofAHelp>

+ Follow

Text follow BofA_Help to 40404 in the United States

Tweets Favorites Following Followers Lists ▾



BofA_Help Bank of America

@ro1224 Please DM details of the issues or errors you are receiving. ^hl

3 Dec



BofA_Help Bank of America

@MsAvie758 If you'd like us to see if we can help please DM your name/zip/contact #. No acct #s. ^hl

3 Dec

- Several businesses have twitter accounts
- How are they using these accounts?
- By understanding the interactions, we can recommend content to both consumers and business owners on twitter

Types

Manual classification on a sample of 1000 tweets from 5245 verified business accounts on twitter

Engagement: all (0.04)	Examples
survey/poll (0.02)	Is an Alicia Keys written song good enough? decide: http://ow.ly/juuC
contest (0.01)	follow @miladyteam to win an ipod touch!
other (0.01)	Have a great 2010 !
Announcements (0.13)	Examples
events(0.07)	November 21st enjoy a Gingerbread House Workshop in Shakopee MN
deal/coupon/sale(0.03)	Deal of the Day: HP Photosmart C8180 All-in-One Printer / Scanner / Copier 199.99
product release (0.001)	New doubleTwist Windows release is out!..
schedule (0.001)	Closed Monday
other (0.088)	For charities: payment schedule over Christmas
job openings (0.01)	diesel mechanic (New Brunswick)
Brand awareness (0.157)	Best solutions for your scrapbooking needs!
Engagement: specific (0.22)	Examples
customer service (0.192)	@bhkahului Do you still have concerns ?
general discussion (0.02)	@milkyBarKed I'm a Pet Shop Boys fan
showcase/retweet (0.01)	RT @ralphryan: Personal responsibility Don't blame someone else (...)
other (0.01)	@popjustice what
Content +links (0.43)	Examples
news updates (0.37)	Commercial Vehicle Group Reports Second Quarter 2009 Results http://bit.ly/1tLydr
opinions/advice (0.006)	what an incredible goal for team USA woot
other 0.06	what is this

Event and deal identification

Announcements (0.13)	Examples
events(0.07)	November 21st enjoy a Gingerbread House Workshop in Shakopee MN
deal/coupon/sale(0.03)	Deal of the Day: HP Photosmart C8180 All-in-One Printer / Scanner / Copier 199.99

- Useful to separate these out for consumers

Results

- Fairly good results

Task	Gold standard	P	R	F-1
Deals	200+,200-	0.95	0.93	0.94
Deals	200+,6860-	0.96	0.76	0.85
Events	200+, 200-	0.96	0.97	0.96
Events	200+, 2860-	0.80	0.84	0.82

- Top features

Deals: top features	Events: top features
1. % tweet tokens from Deal lexicon	1. % tweet tokens from Event lexicon
2. % numeric tweet tokens	2. tweet length
3. % tweet tokens from Junk lexicon	3. % numeric tweet tokens
4. % tweet tokens which are URLs	4. % tweet tokens from Junk lexicon
5. tweet length	5. % tweet tokens which are pronouns

Business tweet topics

- Learn topics on 3 months of business tweets (LDA)
- Manually analyze top 25 topics from each month

Topic class	Num. topics	Example topics (manually labeled)
not clear	15	—
Calendar Occassion	11	New Year's, Valentine Day
Temporal Expressions	8	Hours, Week Days
Food and Drink	7	Chocolate, Ice cream, Beer, Wine
Sports	4	Superbowl, Spring training, March Madness
Weather	3	Snow, Rain
News	3	State of Union, Detroit/Ford news

Month	Topic	Example terms
01-2010	Haiti Earthquake	haiti,relief,donate,cross,red,text
01-2010	New Year's	year, resolution, healthy, calories
02-2010	Valentine Day	day, valentine's, romantic, dinner,reservations
02-2010	Superbowl	super, bowl, sunday, saints, colts, party
03-2010	Saint Patrick's Day	day, st, patrick's, patricks, irish, paddy's, parade
03-2010	Drink specials	happy, hour, half, earth, drinks

- Businesses respond a lot to news and social events
- Useful for recommendations

User response to celebrity events

[Popescu & Pennachiotti, 2011]

- What types of celebrities and events do people tend to discuss more in Twitter?
- What types of engagement with celebrity events do users exhibit?
- What comments do Twitter users have about celebrities?

Annotate buzzy events around celebrities

- 4 types of celebrities: Actors, Musicians, Politicians, Athletes
- Extract tweet streams, find buzzy snapshots
- 500 snapshots were manually annotated for topic discussed
 - Work, media, personal life, social activity

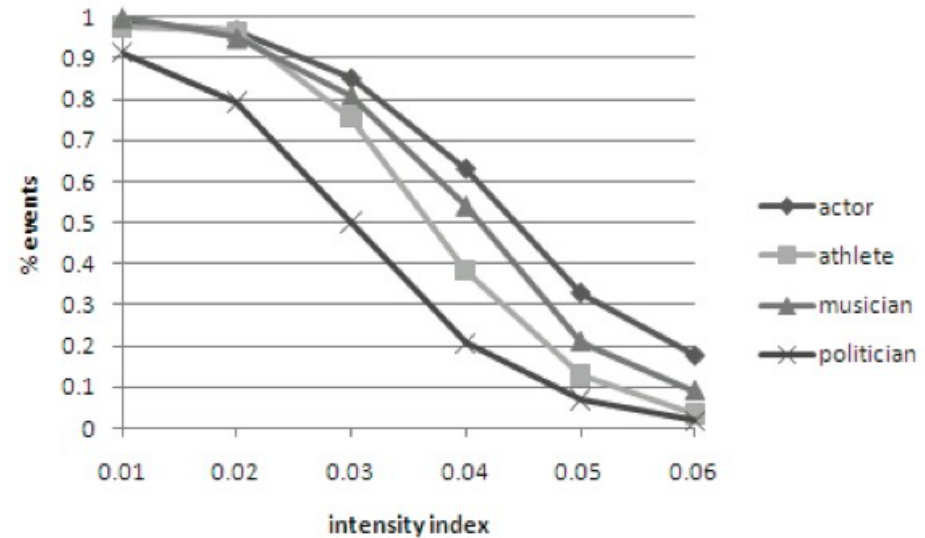
What topics are widely discussed for different entities?

	work	media	life	social	<i>total</i>
ACTOR	122	80	26	29	257
ATHLETE	79	1	2	8	90
MUSICIAN	61	11	17	14	103
POLITICIAN	35	17	4	3	59
<i>Total</i>	297	109	49	54	509

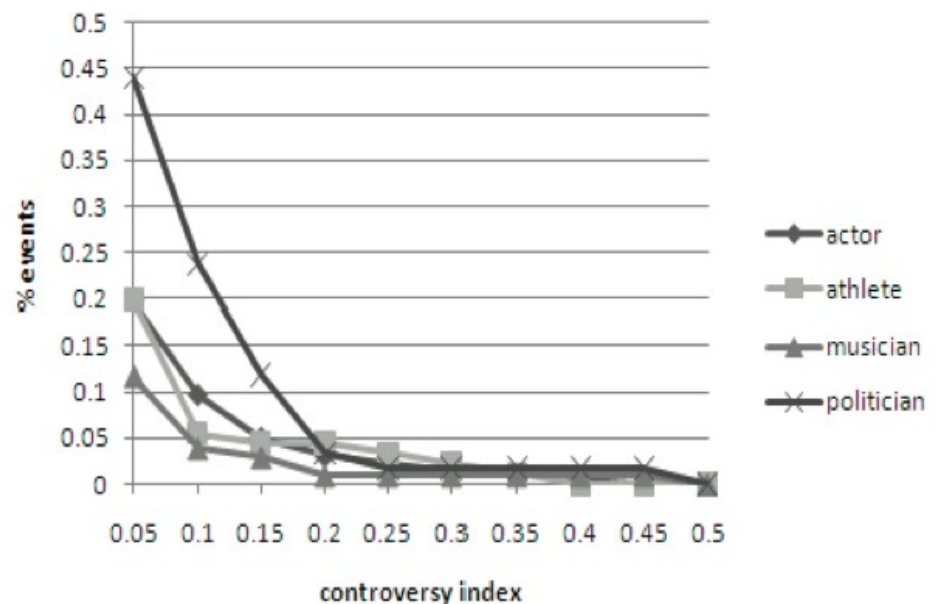
- Actors most popular, politicians least
- Most discussions are about work, very little gossip

Automatic characterization of discussions

- Intensity index: based on amount of punctuation and sentiment words



- Controversies: based on a manual dictionary



- Thank you!
- Questions?