UNED NLP & IR Group
VII Jornadas MAVIR

Modelo basado en LDA para la Gestión de la Reputación de Compañías

Tamara Martín-Wanton

Universidad de Educación a Distancia
Departamento de Lenguajes y sistemas Informáticos

27/11/2012

Motivation / Scenario

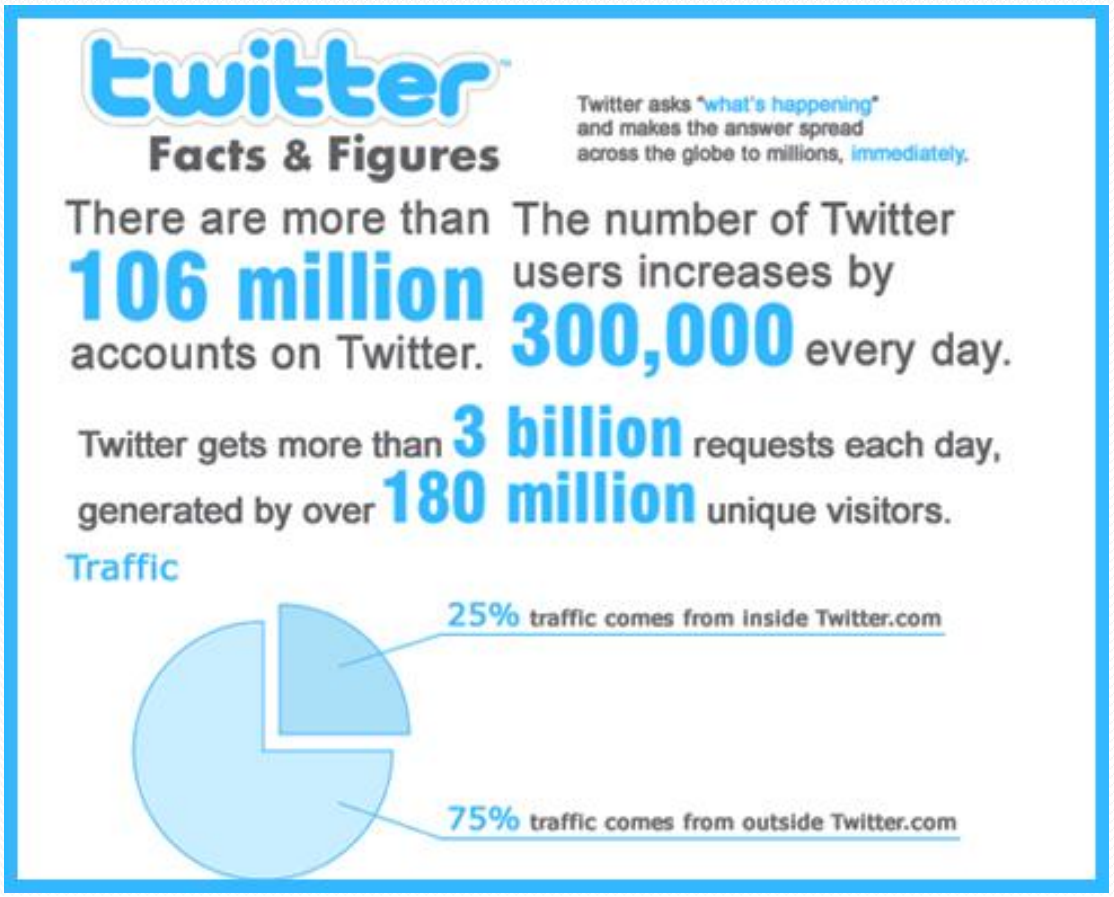
Online Reputation Management

Tasks:

- Which topic is being discussed??
- Which are the most disturbing/important topics??
- Who's talking about me?
- …

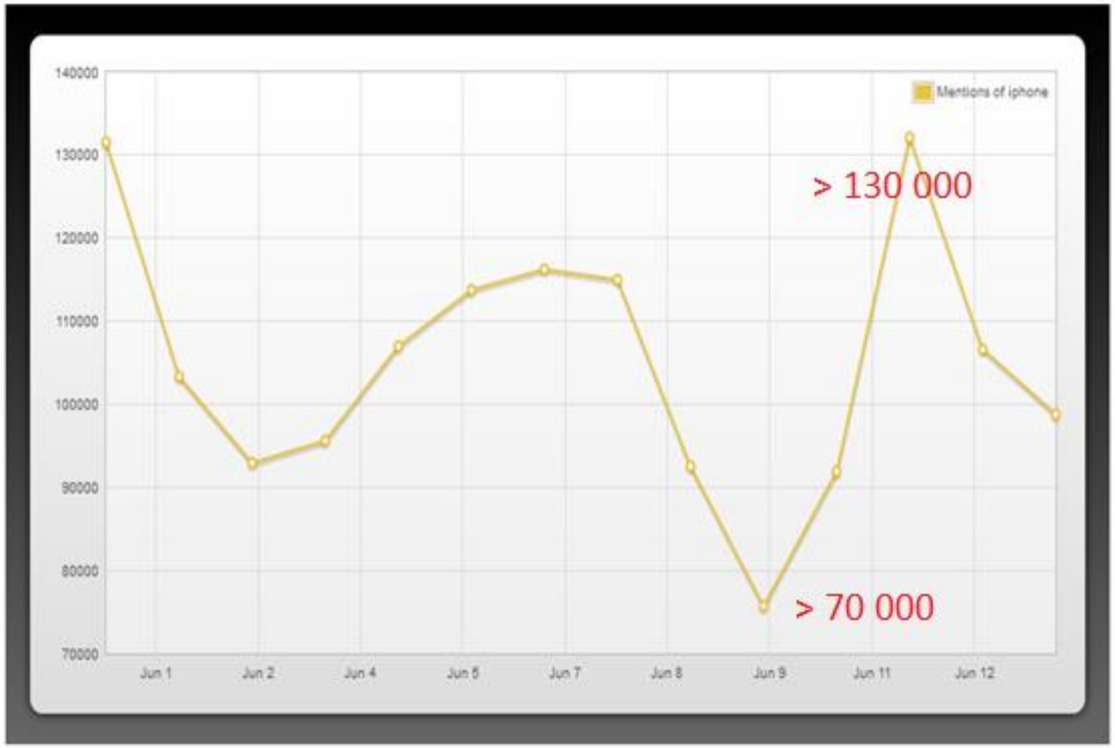
Online Reputation Management

- Source:



Online Reputation Management

A **BIG** problem for the analyst !!!!



Evaluation framework: RepLab 2012 monitoring task

Input

Output

topic 1

2

3

...

less important

Our proposal

- Clustering strategy:
  LDA-based Model

- Priority
  - Sentiment-based heuristic

Topic models for Twitter data??

- Do not make any assumptions about the ordering of words.
- Disregard grammar as well
- Each document is represented as a numerical vector that describes its distribution over the topics.
- Training a topic model is easy since it uses unsupervised learning

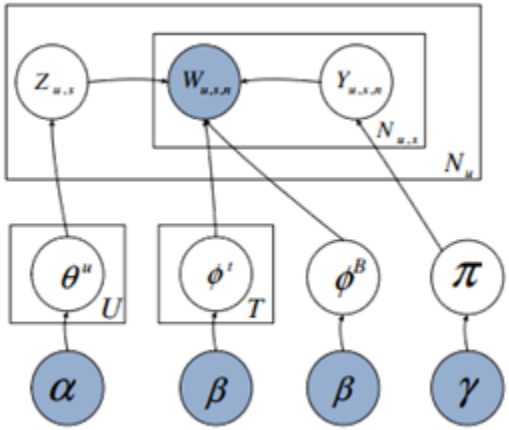
Probabilistic Graphical Models

LDA-based models used for Twitter data

- **Standard LDA [Kireyev2009, Celikyilmaz2010]**
  - Use standard LDA model
  - Problem: Tweets are too short. A single tweet is usually about a single topic.
- **Author-topic models [Weng2010, Puniyani2010]**
  - Consider as document all the tweets of an author
  - Problem: A single tweet is usually about a single topic.
- **Labeled-LDA [Ramage2010]**
  - Partially supervised learning model that use the hashtags to classify the tweets
  - Problem: the model may not include all topics
- **Twitter-LDA model [Zhao2011a, Zhao2011b]**
  - A single tweet is about a single topic
  - Potential improvement: incorporate metadata from twitter

Clustering strategy: TwitterLDA

- Approach based on a latent variable topic model: Latent Dirichlet Allocation (LDA)
- Variant of LDA proposed by (Zhao, 2011): TwitterLDA
  - Adapted to the characteristics of Twitter: **tweets are short** (140-character limit) and a **single tweet tends to be about a single topic**.
- Model assumptions
  - A set of topics T in Twitter (represented by a word distribution)
  - Each user has her topic interests modeled by a distribution over the topics
  - Each tweet (document) has only one topic



Clustering strategy: TwitterLDA

- Each entity has few tweets to be annotated

  Add more tweets as **background information**:

  - Background of the entity (5000 tweets)

    [provide additional information to cluster tweets that has the same topic]

  - Background from a different entity (15000 tweets)

    [differentiate between topics that do not refer to the entity]

Clustering strategy: TwitterLDA

- Input:

  TE : tweets about the entity to annotate

  TBE : tweets of the entity (not in TE)

  TBO: tweets from another entity

- Clustering:

  TwitterLDA (TE + TBE + TBO)

  K = 100

  Each tweet in (TE + TBE + TBO) is assigned a topic

- Output:

  TE clusters (notice that the number of final clusters is different of K)

Priority: Sentiment-based heuristic

- The priority of a topic depends on the sentiment expressed in the tweets that refer to it.
- A tweet-level sentiment analysis classifier (Carrillo de Albornoz, 2012)
  - Works at the concept level (WSD)
  - Uses emotions instead of terms to represent the text as a set of emotional meanings
  - Deals with negations and intensifiers
  - Uses this information to train a Machine Learning Algorithm
  - English - Spanish

Priority: Sentiment-based heuristic

- Priority of a topic:

$$Priority(T_i) = \begin{cases} 3 & \text{if} \quad |Negative(T_i)| = |T_i| \quad or \quad |Positive(T_i)| = |T_i| \\ 2 & \text{if} \quad |Negative(T_i)| \geq |Positive(T_i)| \quad and \quad |Negative(T_i)| \geq |Neutral(T_i)| \\ 2 & \text{if} \quad |Positive(T_i)| > |Negative(T_i)| \quad and \quad |Positive(T_i)| \geq |Neutral(T_i)| \\ 2 & \text{if} \quad |Positive(T_i)| + |Negative(T_i)| \geq |Neutral(T_i)| \\ 1 & \text{if} \quad |Neutral(T_i)| = |T_i| \\ 1 & \text{if} \quad |Neutral(T_i)| > |Positive(T_i)| + |Negative(T_i)| \\ 0 & \text{if} \quad |Neutral(T_i)| + |Positive(T_i)| + |Negative(T_i)| = 0 \end{cases}$$

RepLab at CLEF 2012

RepLab Monitoring Task:

- Early detection of alerts that may damage the reputation of a company
- Clustering + Ranking task
- Twitter
- Multilingual: English + Spanish
- Evaluation Metrics: *Reliability and Sensitivity*
  - *Precision and Recall over binary relations between documents.*
  - *In the context of clustering tasks, Reliability and Sensitivity are equivalent to Bcubed Precision and BCubed Recall, respectively.*

Results

Clustering relationships

System	CLUSTERING R (BCubed P)	S (BCubed R)	F(R,S)
baseline0%	0,40	1	**0,50**
baseline10%	0,50	0,70	0,49
baseline20%	0,89	0,32	0,42
UNED_3	0,72	0,32	0,40
baseline30%	0,95	0,26	0,35
baseline40%	0,97	0,23	0,34
baseline50%	0,97	0,22	0,33
baseline60%	0,97	0,21	0,32
baseline70%	0,98	0,20	0,31
cirgdisco_1	0,95	0,24	0,35
baseline80%	0,98	0,19	0,29
baseline90%	0,98	0,17	0,27
OPTAH_1	0,70	0,34	0,38
baseline100%	0,98	0,17	0,26
UNED_2	0,85	0,34	0,39
UNED_1	0,90	0,20	0,30

- **Baseline_0%:** assigns all the tweets to a single cluster (all-in-one system)
- With regards to F-Measure, obtains the highest score.
- Reaches **perfect recall**
- **Precision is relatively high** on entities with few topics (above 0.95 in 5/24 test cases)

Baseline X%: Agglomerative clustering. Stopping threshold = X% Jaccard word distance. Single Linkage.

Results

Clustering relationships

System	CLUSTERING R (BCubed P)	S (BCubed R)	F(R,S)
baseline0%	0,40	1	**0,50**
baseline10%	0,50	0,70	0,49
baseline20%	0,89	0,32	0,42
UNED_3	0,72	0,32	0,40
baseline30%	0,95	0,26	0,35
baseline40%	0,97	0,23	0,34
baseline50%	0,97	0,22	0,33
baseline60%	0,97	0,21	0,32
baseline70%	0,98	0,20	0,31
cirgdisco_1	0,95	0,24	0,35
baseline80%	0,98	0,19	0,29
baseline90%	0,98	0,17	0,27
OPTAH_1	0,70	0,34	0,38
baseline100%	0,98	0,17	0,26
UNED_2	0,85	0,34	0,39
UNED_1	0,90	0,20	0,30

- **Baseline_0%:** assigns all the tweets to a single cluster (all-in-one system)
- With regards to F-Measure, obtains the highest score.
- Reaches **perfect recall**
- **Precision is relatively high** on entities with few topics (above 0.95 in 5/24 test cases)

Results

Clustering relationships

System	CLUSTERING R (BCubed P)	S (BCubed R)	F(R,S)
baseline0%	0,40	1	**0,50**
baseline10%	0,50	0,70	0,49
baseline20%	0,89	0,32	0,42
UNED_3	0,72	0,32	0,40
baseline30%	0,95	0,26	0,35
baseline40%	0,97	0,23	0,34
baseline50%	0,97	0,22	0,33
baseline60%	0,97	0,21	0,32
baseline70%	0,98	0,20	0,31
cirgdisco_1	0,95	0,24	0,35
baseline80%	0,98	0,19	0,29
baseline90%	0,98	0,17	0,27
OPTAH_1	0,70	0,34	0,38
baseline100%	0,98	0,17	0,26
UNED_2	0,85	0,34	0,39
UNED_1	0,90	0,20	0,30

When most of the **tweets are not related to the entity** of interest (ex: Indra, ING, BP) all of the proposed systems obtain F-1 scores below 0.25

Treatment of ambiguity is needed

Results

Priority relationships

System	PRIORITY R	S	F(R,S)
baseline0%	0	0	0
baseline10%	0,35	0,16	0,16
baseline20%	0,32	0,33	**0,28**
UNED_3	0,25	0,30	0,26
baseline30%	0,32	0,28	0,27
baseline40%	0,31	0,31	0,27
baseline50%	0,31	0,31	0,27
baseline60%	0,30	0,30	0,27
baseline70%	0,30	0,30	0,27
cirgdisco_1	0,24	0,30	0,24
baseline80%	0,30	0,30	0,26
baseline90%	0,29	0,29	0,25
OPTAH_1	0,19	0,16	0,16
baseline100%	0,28	0,27	0,24
UNED_2	0	0	0
UNED_1	0	0	0

Baseline: assigns all non-single clusters to the same level, and single clusters are assigned to a secondary level.

Results

Priority relationships

System	PRIORITY R	S	F(R,S)
baseline0%	0	0	0
baseline10%	0,35	0,16	0,16
baseline20%	0,32	0,33	**0,28**
UNED_3	0,25	0,30	0,26
baseline30%	0,32	0,28	0,27
baseline40%	0,31	0,31	0,27
baseline50%	0,31	0,31	0,27
baseline60%	0,30	0,30	0,27
baseline70%	0,30	0,30	0,27
cirgdisco_1	0,24	0,30	0,24
baseline80%	0,30	0,30	0,26
baseline90%	0,29	0,29	0,25
OPTAH_1	0,19	0,16	0,16
baseline100%	0,28	0,27	0,24
UNED_2	0	0	0
UNED_1	0	0	0

Upper-bound Priority Sentiment Based Method:

Perfect clustering

Perfect polarity

Sentiment-based heuristic

R	S	$F(R,S)$
0.75	0.70	0.71

- **Suggesting that the overall sentiment of the topic is a helpful variable for the priority annotation**

Results

Monitoring task (clustering and priority)

System	CLUSTERING			PRIORITY			ALL		
	R (BCubed P)	S (BCubed R)	F(R,S)	R	S	F(R,S)	R	S	F(R,S)
baseline0%	0,40	1	**0,50**	0	0	0	0,40	0,43	**0,41**
baseline10%	0,50	0,70	0,49	0,35	0,16	0,16	0,34	0,33	0,33
baseline20%	0,89	0,32	0,42	0,32	0,33	**0,28**	0,38	0,26	0,30
UNED_3	0,72	0,32	0,40	0,25	0,30	0,26	0,32	0,26	0,29
baseline30%	0,95	0,26	0,35	0,32	0,28	0,27	0,37	0,23	0,27
baseline40%	0,97	0,23	0,34	0,31	0,31	0,27	0,35	0,21	0,26
baseline50%	0,97	0,22	0,33	0,31	0,31	0,27	0,35	0,21	0,26
baseline60%	0,97	0,21	0,32	0,30	0,30	0,27	0,34	0,20	0,25
baseline70%	0,98	0,20	0,31	0,30	0,30	0,27	0,33	0,20	0,25
cirgdisco_1	0,95	0,24	0,35	0,24	0,30	0,24	0,29	0,22	0,25
baseline80%	0,98	0,19	0,29	0,30	0,30	0,26	0,33	0,19	0,24
baseline90%	0,98	0,17	0,27	0,29	0,29	0,25	0,31	0,17	0,22
OPTAH_1	0,70	0,34	0,38	0,19	0,16	0,16	0,37	0,19	0,22
baseline100%	0,98	0,17	0,26	0,28	0,27	0,24	0,30	0,16	0,20
UNED_2	0,85	0,34	0,39	0	0	0	0,85	0,09	0,14
UNED_1	0,90	0,20	0,30	0	0	0	0,90	0,05	0,10

Not considering priority relationships significantly drops Sensitivity scores.

Conclusions

- A clustering + ranking approach based in LDA + sentiment-based ranking algorithm
- Clustering
  - High Reliability (BCubed Precision) scores -> There is a terminology that describes each topic
  - Low Sensitivity (Bcubed Recall) scores -> It's hard to find!
- Priority
  - Sentiment expressed in tweets of the same cluster -> useful indicator of the topic priority
  - Need to be combined with other signals: novelty, centrality, potential impact

UNED NLP & IR Group
VII Jornadas MAVIR

Modelo basado en LDA para la Gestión de la Reputación de Compañías

Tamara Martín-Wanton

Universidad de Educación a Distancia
Departamento de Lenguajes y sistemas Informáticos

27/11/2012