Survey of Social Network Analysis

Xiaosong Jia

04/10/2018

Content

- Automatic Detection and Classification of Social Events
- Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors
- Automatic sub-event detection in emergency management using social media
- Discover breaking events with popular hashtags in twitter
- TopicSketch: Real-Time Bursty Topic Detection from Twitter
- Real-Time Disease Surveillance Using Twitter Data: Demonstration on Flu and Cancer
- A unified model for stable and temporal topic detection from social media data.

Automatic Detection and Classification of Social Events

Apoorv Agarwal
Department of Computer Science
Columbia University
New York, U.S.A.
apoorv@cs.columbia.edu

Owen Rambow
CCLS
Columbia University
New York, U.S.A.
rambow@ccls.columbia.edu

2010 EMNLP(CCF B)

Owen Rambow

	总计	2013 年至今
引用	8507	4285
h 指数	44	27
i10 指数	148	93

Sentiment analysis of twitter data A Agarwal, B Xie, I Vovsha, O Rambow, R Passonneau ACL HLT 2011, 30	1079	2011
Arabic tokenization, part-of-speech tagging and morphological disambiguation in one fell swoop N Habash, O Rambow Proceedings of the 43rd Annual Meeting on Association for Computational ...	437 *	2005
MADA+ TOKAN: A toolkit for Arabic tokenization, diacritization, morphological disambiguation, POS tagging, stemming and lemmatization N Habash, O Rambow, R Roth Proceedings of the 2nd international conference on Arabic language resources ...	276	2009
A fast and portable realizer for text generation systems B Lavoie, O Rambow Proceedings of the fifth conference on Applied natural language processing ...	257 *	1997
MADAMIRA: A Fast, Comprehensive Tool for Morphological Analysis and Disambiguation of Arabic. A Pasha, M Al-Badrashiny, MT Diab, A El Kholy, R Eskander, N Habash, ... LREC 14, 1094-1101	219	2014

Definition

- We take a "social network" to be a network consisting of individual human beings and groups of human beings who are connected to each other by the virtue of participating in social events.
- social events are events that occur between people where at least one person is aware of the other and of the event taking place.
- broad types of social events
  ——Interaction event (INR): When both entities participating in an event are aware of each other and of the social event, we say they have an INR relation

  —— Observation event (OBS): When only one person is aware of the other and of the social event, we say they have an OBS relation.
- Of the type OBS: there are three subtypes:

  ——PPR requires that one entity can observe the other entity in real time

  ——PCR, where one entity observes the other through media (TV, radio, magazines etc.)

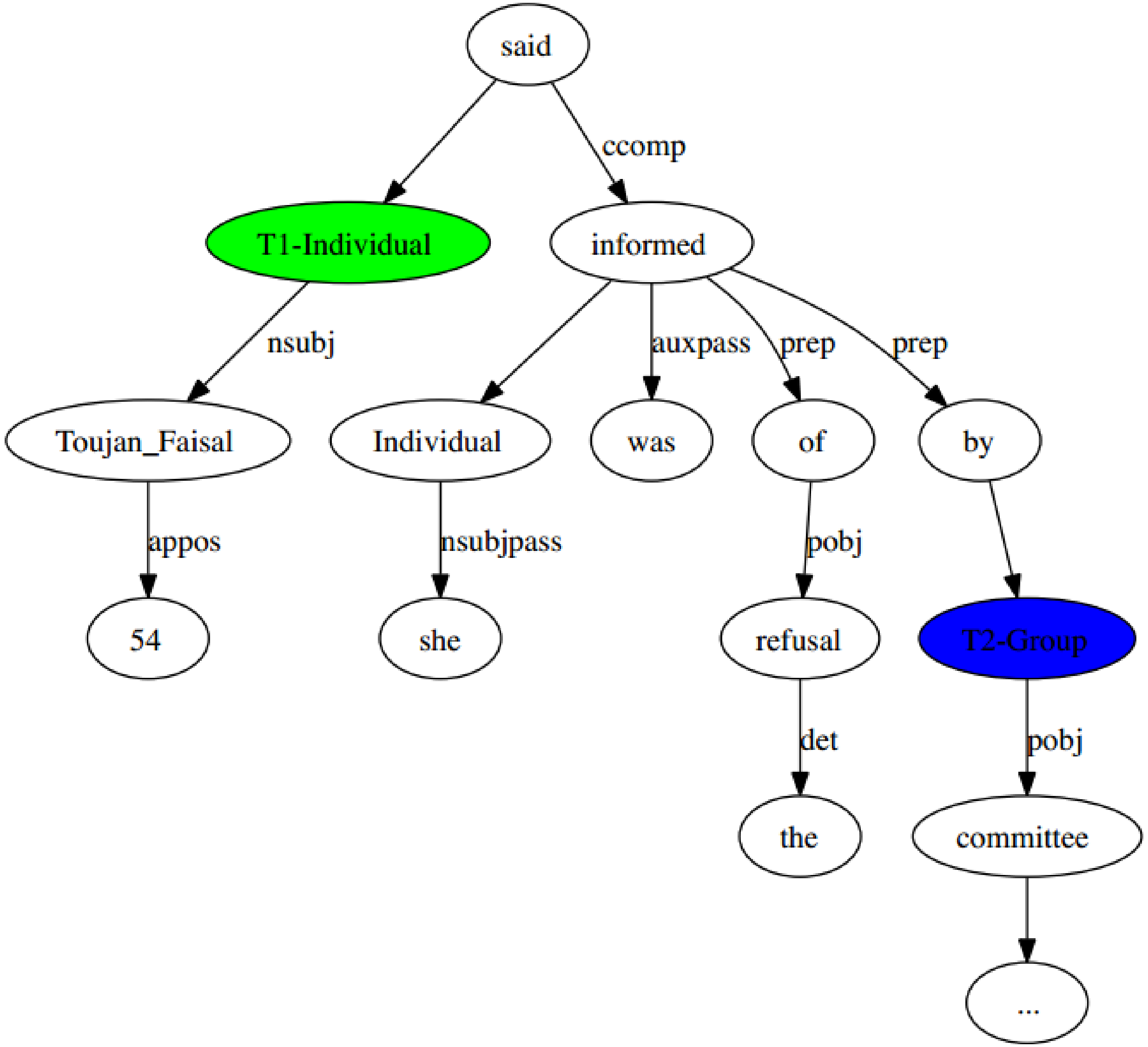
  ——Any other observation event that is not PPR or PCR is COG.

Motivation

- We are interested in modeling classes of events which are characterized by the cognitive states of participants–who is aware of whom. The predicate-argument structure of verbs can encode much of this information very efficiently, and classes of verbs express their predicate-argument structure in similar ways.
- For example, many verbs of communication can express their arguments using the same pattern: *John talked/spoke/lectured/ranted/testified to Mary about Percy*. Independently of the verb, **John** is in a COG relation with **Percy** and in an INR relation with **Mary**. All these verbs allow us to drop either or both of the prepositional phrases, without altering the interpretation of the remaining constituents. And even more strikingly, any verb that can be put in that position is likely to have this interpretation;

Method

- Linear learning machines are used for classification problems.
- The well-known kernel trick aids us in finding similarity between feature vectors in a high dimensional space without having to write down the expanded feature space.
- Phrase Structure Trees (PST)
- Dependency Words (DW) tree
- Grammatical Relation (GR) tree
- Grammatical Relation Word (GRW) tree
- Sequence Kernel of words (SK1)
- Sequence in GRW tree (SqGRW)
- We also use combinations of these structures
- We use the Partial Tree (PT) kernel
- We employ two well-known data sampling methods on the training data before creating a model for test data; random under-sampling and random over-sampling



Experiment

Baseline:

Kernel	P	R	F1
PET	70.28	21.46	32.38
GR	**87.79**	15.21	25.55
GRW	76.42	8.26	14.8
SqGRW	48.78	6.08	10.38
PET_GR	70.21	**27.76**	**38.89**
PET_GR_SqGRW	71.06	26.74	38.02
GR_SqGRW	82.0	24.47	36.12
GRW_SqGRW	68.19	17.01	25.06
GR_GRW_SqGRW	79.81	21.99	32.57

Under-sampled:

Kernel	P	R	F1
PET	28.89	77.06	41.96
GR	**35.68**	72.47	47.37
GRW	29.7	83.6	43.6
SqGRW	34.31	**84.15**	**48.61**
PET_GR	34.38	83.94	**48.52**
PET_GR_SqGRW	34.34	83.66	**48.52**
GR_SqGRW	33.45	81.73	47.27
GRW_SqGRW	32.87	**84.44**	47.11
GR_GRW_SqGRW	32.73	83.26	46.82

Over-sampled:

Kernel	P	R	F1
PET	50.9	57.21	53.62
GR	43.57	67.21	52.59
GRW	46.05	64.15	53.31
SqGRW	42.4	**72.75**	53.5
PET_GR	56.42	66.2	60.63
PET_GR_SqGRW	**57.28**	66.26	**61.11**
GR_SqGRW	44.35	71.17	54.52
GRW_SqGRW	44.77	68.79	54.12
GR_GRW_SqGRW	46.79	71.54	56.45

WWW 2010 • Full Paper April 26-30 • Raleigh • NC • USA

Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors

Takeshi Sakaki
The University of Tokyo
Yayoi 2-11-16, Bunkyo-ku
Tokyo, Japan
sakaki@biz-model.t.u-tokyo.ac.jp

Makoto Okazaki
The University of Tokyo
Yayoi 2-11-16, Bunkyo-ku
Tokyo, Japan
m_okazaki@biz-model.t.u-tokyo.ac.jp

Yutaka Matsuo
The University of Tokyo
Yayoi 2-11-16, Bunkyo-ku
Tokyo, Japan
matsuo@biz-model.t.u-tokyo.ac.jp

2010 WWW(CCF A)

Yutaka Matsuo

Konnichiwa! Thank you for visiting.
My name is Yutaka Matsuo. I am a project associate professor at the University of Tokyo (UT), working on Information Technology and Artificial Intelligence. I got my Ph.D degree from the University of Tokyo in 2002. From Oct. 2005 to Oc. 2007, I was a visiting scholar at CSLI, Stanford University.[Vita]

My current research interests are in web mining, social networks, and deep learning.
[Research topics] [Publication]

I belong to Department of Technology Management for Innovation (at Graduate School) and Program for Social Innovation (for undergraduates). My laboratory is weblab .

Please feel free to contact me!

	总计	2013 年至今
引用	8690	5276
h 指数	36	23
i10 指数	81	41

Earthquake shakes Twitter users: real-time event detection by social sensors T Sakaki, M Okazaki, Y Matsuo Proceedings of the 19th international conference on World wide web, 851-860	3310	2010
Keyword extraction from a single document using word co-occurrence statistical information Y Matsuo, M Ishizuka International Journal on Artificial Intelligence Tools 13 (01), 157-169	747	2004
Measuring semantic similarity between words using web search engines. D Bollegala, Y Matsuo, M Ishizuka www 7, 757-766	657	2007
POLYPHONET: an advanced social network extraction system from the web Y Matsuo, J Mori, M Hamasaki, T Nishimura, H Takeda, K Hasida, ... Web Semantics: Science, Services and Agents on the World Wide Web 5 (4), 262-278	407	2007
Tweet analysis for real-time event detection and earthquake reporting system development T Sakaki, M Okazaki, Y Matsuo IEEE Transactions on Knowledge and Data Engineering 25 (4), 919-931	256	2013
A web search engine-based approach to measure semantic similarity between words D Bollegala, Y Matsuo, M Ishizuka	174	2011

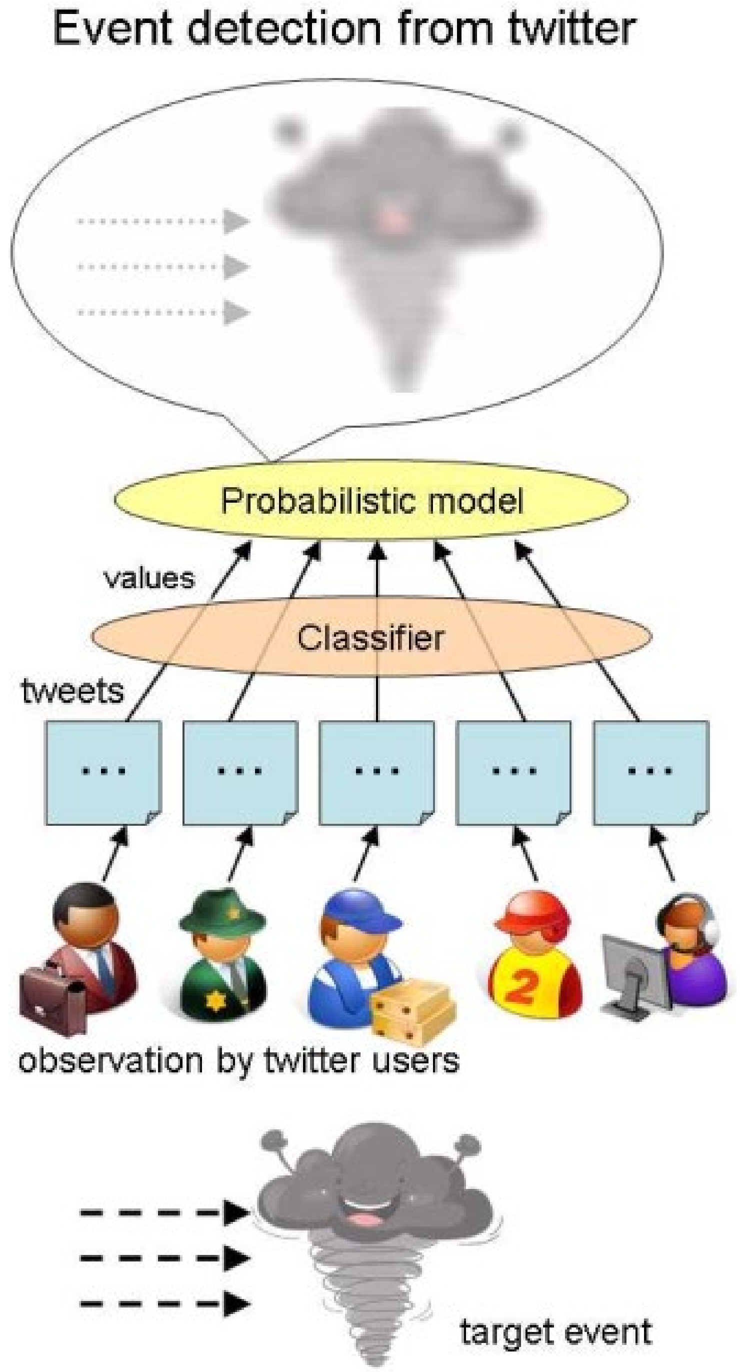
Motivation

- An *event* is an arbitrary classification of a space–time region.
  An event might have actively participating agents, passive
  factors, products, and a location in space/time.

- We target events such as earthquakes, typhoons, and traffic jams,
  which are visible through tweets. These events have several properties:
- i) they are of large scale (many users experience the event),
- ii) they particularly influence people's daily life (for that reason, they are
  induced to tweet about it),
- iii) they have both spatial and temporal regions (so that real-time
  location estimation is possible)

Thought

- To classify a tweet into a positive class or a negative class, we use a support vector machine.
- Each Twitter user is regarded as a sensor. A sensor detects a target event and makes a report probabilistically.
- Each tweet is associated with a time and location, which is a set of latitude and longitude.



Temporal Model

In the Twitter case, we can infer that if a user detects an event at time 0, assume that the probability of his posting a tweet from t to Δt is fixed as λ. Then, the time to make a tweet can be considered as an exponential distribution.

$$f(t;\lambda) = \lambda e^{-\lambda t} \text{ where } t > 0 \text{ and } \lambda > 0.$$

The false-positive ratio *pf* of a sensor is approximately 0.35

Sensors are assumed to be independent and identically
distributed (i.i.d.)

$$p_{occur}(t) = 1 - p_f^{n_0(1-e^{-\lambda(t+1)})/(1-e^{-\lambda})}$$

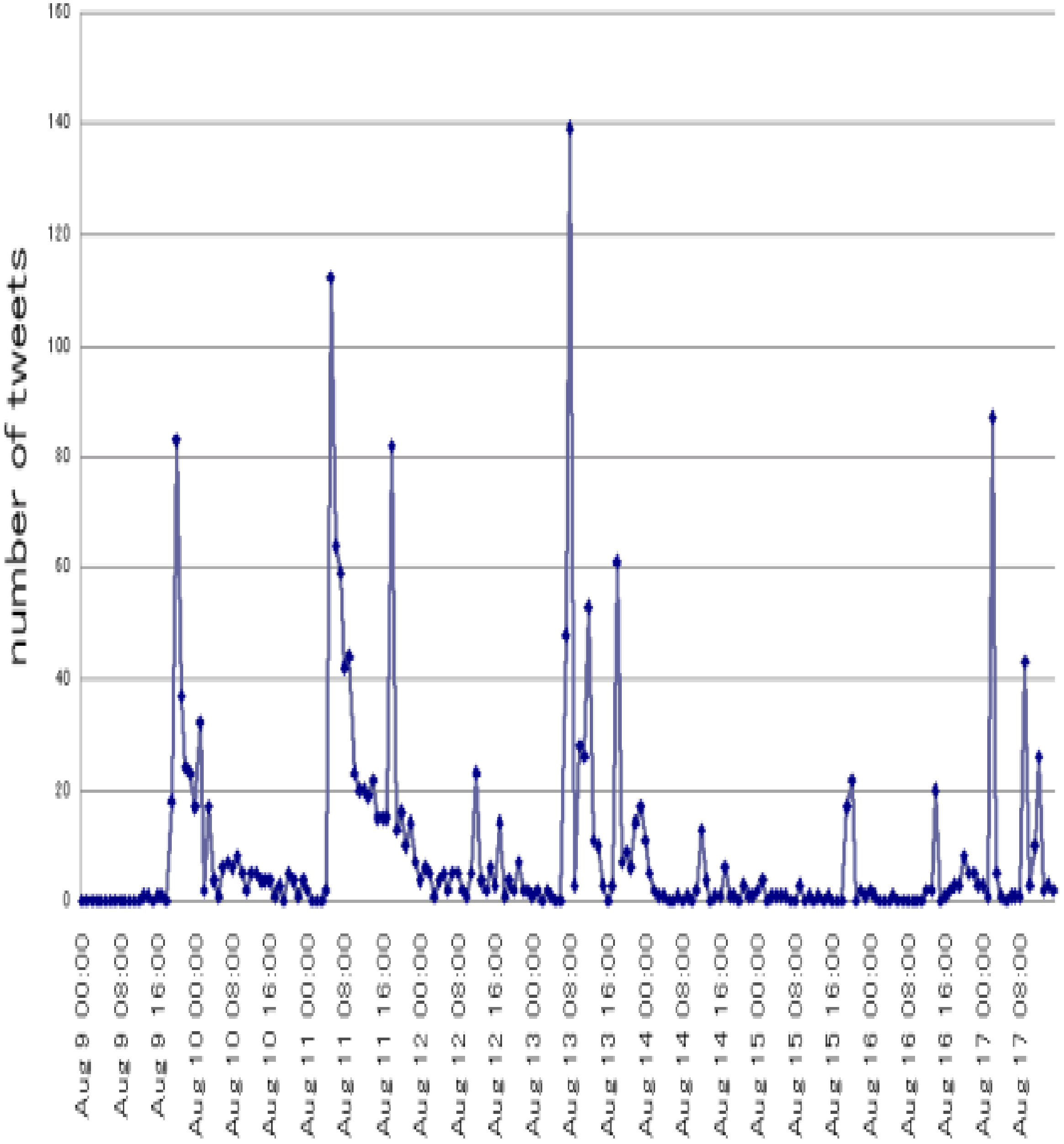


Figure 4: Number of tweets related to earthquakes.

Spatial Model

- From a Bayesian perspective, the tracking problem is to calculate, recursively, some degree of belief in the state *xt* at time *t*, given data *zt* up to time *t*.
- Kalman Filters
- Particle Filters

Experiment

Table 1: Performance of classification.

(i) *earthquake* query:

Features	Recall	Precision	*F*-value
A	87.50%	63.64%	73.69%
B	87.50%	38.89%	53.85%
C	50.00%	66.67%	57.14%
All	87.50 %	63.64%	73.69%

(ii) *shaking* query:

Features	Recall	Precision	*F*-value
A	66.67%	68.57%	67.61%
B	86.11%	57.41%	68.89%
C	52.78%	86.36%	68.20%
All	80.56 %	65.91%	72.50%

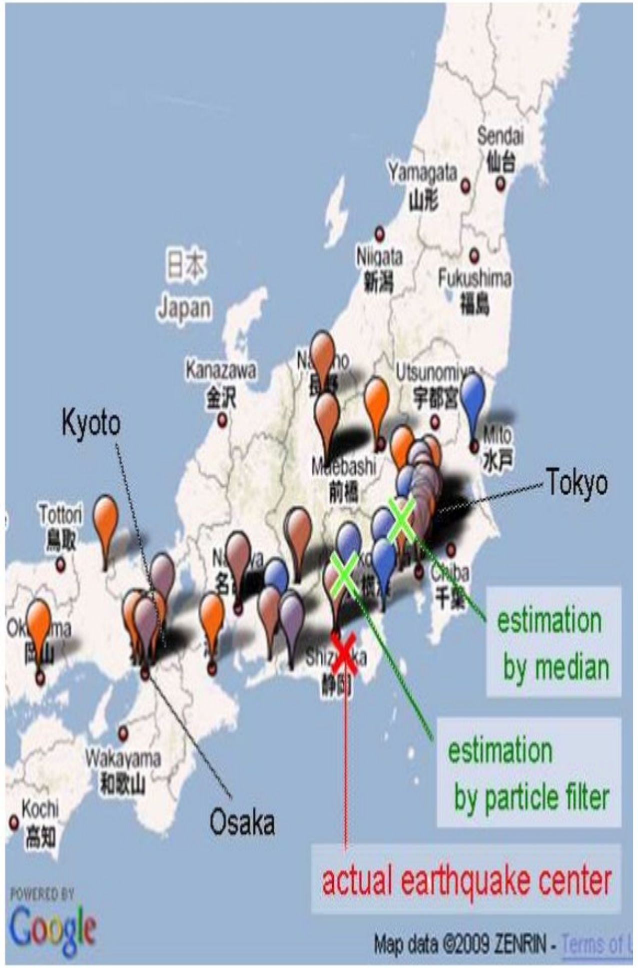


Figure 11: Screenshot of Toretter, an earthquake reporting system.

```
Dear Alice,

We have just detected an earthquake
around Chiba. Please take care.

Toretter Alert System
```

Figure 12: Sample alert e-mail.

WWW 2012 – SWDM'12 Workshop April 16–20, 2012, Lyon, France

Automatic Sub-Event Detection in Emergency Management Using Social Media *

Daniela Pohl
Institute of Information Technology
Klagenfurt University, Austria
daniela@itec.uni-klu.ac.at

Abdelhamid Bouchachia
Institute of Informatics Systems
Klagenfurt University, Austria
hamid@isys.uni-klu.ac.at

Hermann Hellwagner
Institute of Information Technology
Klagenfurt University, Austria
hellwagn@itec.uni-klu.ac.at

2012 WWW(CCF A)

Motivation

- Social media platforms (e.g., Flickr, YouTube, Twitter, Facebook) turn out to be a valuable technology for collecting data (e.g., continuous status update, context information) of different types (e.g., pictures, videos, text messages), making such technology very useful for emergency management.

Method

- First, the framework performs a *pre-selection* of the data from different repositories using user-supplied keywords
- Second, sub-event detection Subevents are events during a disaster which are separated from other events w.r.t. time or location.
- Third, labeling and the assessment of the resulting sub-events.

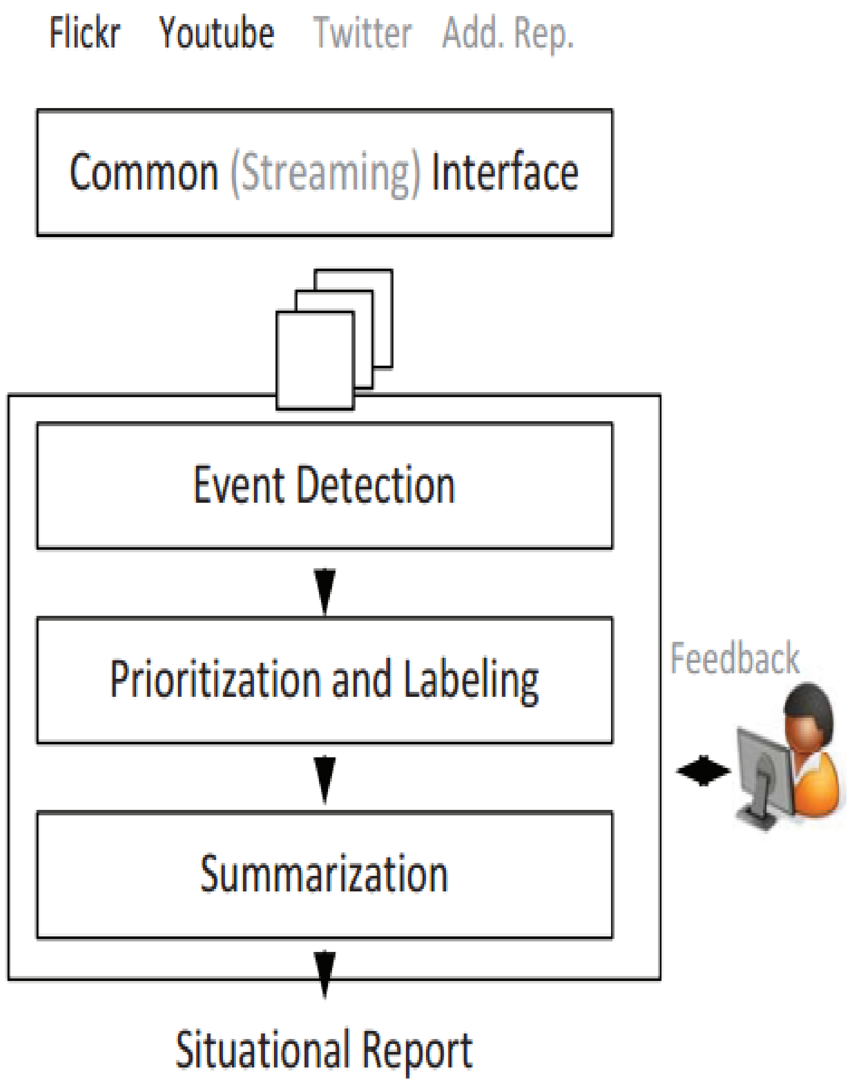


Figure 1: Multimedia (Metadata) Exploration Framework

Sub-event dection: A clustering approach

- It's based on a *Self Organizing Map*(SOM). SOM is a special case of a neural network without any hidden layer
- It maps input vectors into a lower-dimensional map.

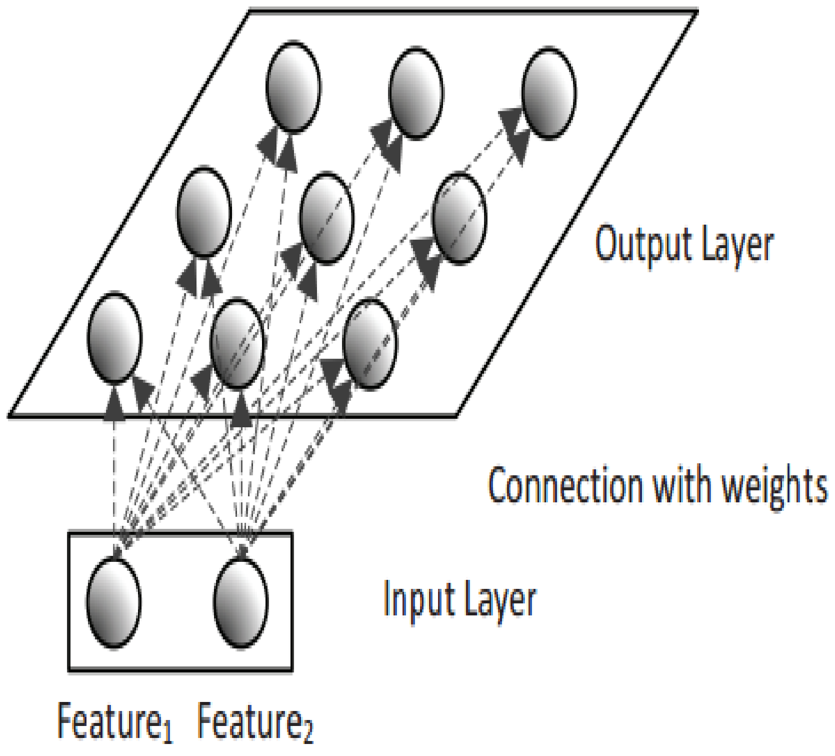


Figure 2: Self Organizing Map (SOM), with a 3x3 map resulting in 9 clusters (adopted from [6])

Experiment

Table 2: UK Riots 2011: Clustering results

Cluster (#hits)	Top 4 Words
Cluster 1 (151)	Polit*, Anarch*, ***Salford***, ***Manchester***
Cluster 2 (118)	***Birmingham***, UK, peopl*, burn*
Cluster 3 (104)	***London***, loot*, riot*, pol*
Cluster 4 (60)	***London***, ***Birmingham***, loot*, riot*
Cluster 5 (10)	Polit*, ***Manchester***, str*, ***Salford***
Cluster 6 (9)	Polit*, Anarch*, ***Salford***, ***Manchester***

Table 3: Oslo Bombing 2011: Clustering results

Cluster (#hits)	Top 4 Words
Cluster 1 (131)	terror, attack, ***shoot****, kil*
Cluster 2 (59)	governm*, ***Oslo***, expl*, bomb*
Cluster 3 (47)	injur*, ***car***, peopl*, kil*
Cluster 4 (16)	expl*, ***Oslo***, governm*, bomb*

Discover breaking events with popular hashtags in twitter

Discover Breaking Events with Popular Hashtags in Twitter*

Anqi Cui, Min Zhang, Yiqun Liu, Shaoping Ma, Kuo Zhang
State Key Laboratory of Intelligent Technology and Systems
Tsinghua National Laboratory for Information Science and Technology
Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China
cuianqi@gmail.com, {z-m, yiqunliu, msp}@tsinghua.edu.cn,
zhangkuo@sogou-inc.com

2012 CIKM(CCF B)

Definition

- **Hashtag Instability** is how unlikely the hashtag keeps a stable amount based on previous observation.
- *Twitter meme possibility is* to tell the difference between the memes and event topics
- For detecting automated agents and robots， we take the *authorship entropy* as a measurement on how concentrated the contributed authors are.

Method

- **Hashtag Instability**

$$\tilde{P}(x) \quad = \quad Pr(X > x \bigvee X < 2\mu - x)$$

$$Inst(x) = -\log \tilde{P}(x), Inst(H) = \frac{1}{n} \sum_{\tilde{P}(x)<p} Inst(x) \qquad (2)$$

- **Twitter Meme Possibility**

$$p_{\text{word}} = 1 - N/L \qquad TMP(hashtag) = p_{\text{word}} \cdot p_{\text{pos}}.$$

$$p_{\text{pos}} = \frac{|\{\text{tweets starting with } h\}|}{|\{\text{tweets containing } h\}|}$$

Method

- **Authorship Entropy**

$$Ent(hashtag) = -\sum_{i=1}^{k} \frac{c_i}{n} \cdot \log\left(\frac{c_i}{n}\right)$$

- **Categorization**

The hashtag instability, Twitter meme possibility and authorship entropy are three orthogonal dimensions which are independent from each other. Considering each feature a lower or a higher value, the hashtag space is divided into eight subspaces

Experiment

Table 4: Experiment Results (Precision, Recall and F-Measure) of Hashtag Categories

Dataset	*Tweets6*						*Tweets3*					
Algorithm	Popularity Pattern			Subspace			Popularity Pattern			Subspace		
Accuracy	17.8%			**40.0%**			31.5%			**38.0%**		
Breaking events	0.250	**0.231**	0.240	**0.333**	0.205	**0.254**	**0.192**	**0.192**	**0.192**	0.167	0.154	0.160
Twitter memes	0.000	0.000	0.000	**0.681**	**0.595**	**0.635**	**1.000**	0.060	0.113	0.725	**0.248**	**0.369**
Advertisements		–		**0.258**	**0.370**	**0.304**		–		**0.053**	**0.385**	**0.093**
Miscellaneous	**0.162**	**0.926**	**0.276**	0.125	0.148	0.136	**0.240**	**0.909**	**0.379**	0.220	0.205	0.212

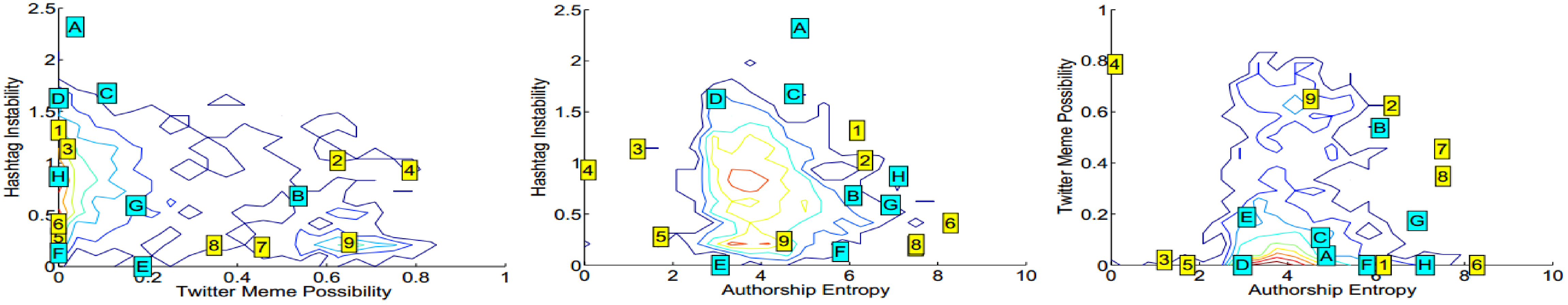


Figure 5: Contour of hashtag distributions. ***Tweets6*****: 1–#hcr, 2–#nowplaying, 3–#property, 4–#praytweets, 5–#abbeydawn, 6–#fb, 7–#musicmonday, 8–#followfriday, 9–#iaintafraidtosay.** ***Tweets3*****: A–#sopa, B–**

TopicSketch: Real-Time Bursty Topic Detection from Twitter

TopicSketch: Real-time Bursty Topic Detection from Twitter

Wei Xie, Feida Zhu, Jing Jiang, Ee-Peng Lim
Living Analytics Research Centre
Singapore Management University
{wei.xie.2012, fdzhu, jingjiang, eplim}@smu.edu.sg

Ke Wang
Simon Fraser University
*Singapore Management University**
wangk@cs.sfu.ca

2012 ICDM(CCF B)

Ee-Peng Lim

SMU Director, Living Analytics Research Centre

LARC
LIVING ANALYTICS RESEARCH CENTRE

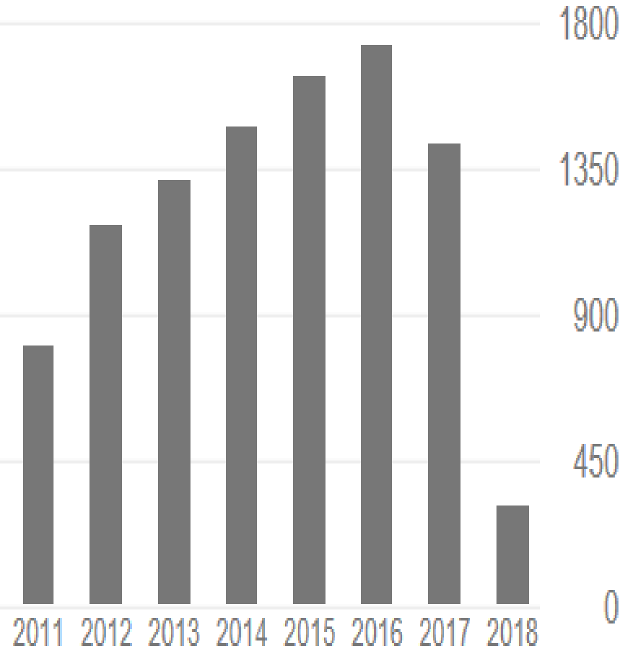
Professor of Information Systems

School of Information Systems
Singapore Management University
(a member of i-School Caucus and CiSAP)
80 Stamford Road, Singapore 178902 (direction)

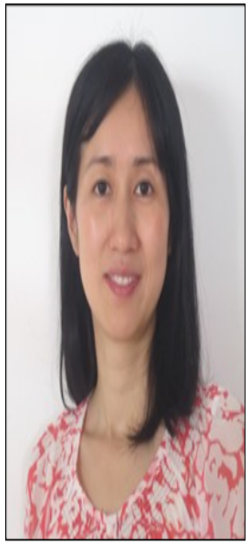
Tel: +65-6828-0781
Fax:+65-6828-0919
Email: eplim(at)smu.edu.sg

引用次数 查看全部

	总计	2013 年至今
引用	13810	7895
h 指数	51	39
i10 指数	202	111



标题	引用次数	年份
Twitterrank: finding topic-sensitive influential twitterers J Weng, EP Lim, J Jiang, Q He Proceedings of the third ACM international conference on Web search and data ...	1847	2010
Comparing twitter and traditional media using topic models WX Zhao, J Jiang, J Weng, J He, EP Lim, H Yan, X Li European Conference on Information Retrieval, 338-349	835	2011
Mobile commerce: promises, challenges, and research agenda K Siau, L Ee-Peng, Z Shen Journal of Database management 12 (3), 4	573	2001
Detecting product review spammers using rating behaviors EP Lim, VA Nguyen, N Jindal, B Liu, HW Lauw Proceedings of the 19th ACM international conference on Information and ...	511	2010
Hierarchical text classification and evaluation A Sun, EP Lim Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on ...	415	2001
Research issues in web data mining SK Madria, SS Bhowmick, WK Ng, EP Lim International Conference on Data Warehousing and Knowledge Discovery, 303-312	365	1999



Jing Jiang

Associate Professor

School of Information Systems
Singapore Management University
80 Stamford Road
Singapore 178902

Phone: (+65) 6828 0785
Email: jingjiang at smu dot edu dot sg

引用次数 查看全部

	总计	2013 年至今
引用	7562	5738
h 指数	33	29
i10 指数	58	51

标题	引用次数	年份
TwitterRank: Finding topic-sensitive influential Twitterers J Weng, EP Lim, J Jiang, Q He Proceedings of the Third ACM International Conference on Web Search and Data ...	1847	2010
Comparing Twitter and traditional media using topic models WX Zhao, J Jiang, J Weng, J He, EP Lim, H Yan, X Li The 33rd European Conference on Information Retrieval, 338-349	835	2011
Adaptive filters for continuous queries over distributed data streams C Olston, J Jiang, J Widom Proceedings of the 2003 ACM SIGMOD International Conference on Management of ...	571	2003
Instance weighting for domain adaptation in NLP J Jiang, CX Zhai Proceedings of the 45th Annual Meeting of the Association for Computational ...	552	2007
Jointly modeling aspects and opinions with a MaxEnt-LDA hybrid WX Zhao, J Jiang, H Yan, X Li Proceedings of the 2010 Conference on Empirical Methods in Natural Language ...	295	2010

Topic Detection

- Our idea of early detection is to monitor the acceleration of a topic

minimize

$$f = w_X \cdot e_X + w_Y \cdot e_Y \tag{4}$$

s.t.

$$\sum_{k=1}^{K} a_k(t) = \mathbb{S}''(t) \tag{5}$$

$$\sum_{i=1}^{N} p_{k,i} = 1, 1 \leq k \leq K \tag{6}$$

$$p_{k,i} \geq 0, 1 \leq k \leq K, 1 \leq i \leq N \tag{7}$$

where

$$e_X = \sum_{i=1}^{N} (\sum_{k=1}^{K} a_k(t) \cdot p_{k,i} - \mathbb{X}_i''(t))^2 \tag{8}$$

$$e_Y = \sum_{i=1}^{N} \sum_{j=1}^{N} (\sum_{k=1}^{K} a_k(t) \cdot p_{k,i} \cdot p_{k,j} - \mathbb{Y}_{i,j}''(t))^2 \tag{9}$$

Realtime Detection Techinique

- Dimension Reduction

   1. Hashing all the distinct words into *B* buckets.

   2. After hashing, what we obtain is the distribution over buckets

   3. we use count-min algorithm to estimate the probability of each word *i* as *min*1 ⩽ *h* ⩽ *H*{*p*(*k*, *hH*) (*i*)}, and return the words of high probability {*i*|*min*1 ⩽ *h* ⩽ *H*{*p*(*k*, *hH*) (*i*)} ⩾ *s*}, where *s* is a probability threshold, e.g., 0.02.

- Efficient Sketch Maintenance

$$\mathbb{S}'_{\Delta T}(t) = \begin{cases} \mathbb{S}'_{\Delta T}(t_{d_{i-1}}) \cdot e^{\frac{(t_{d_{i-1}} - t)}{\Delta T}}, & t \in (t_{d_{i-1}}, t_{d_i}) \\ \mathbb{S}'_{\Delta T}(t_{d_{i-1}}) \cdot e^{\frac{(t_{d_{i-1}} - t)}{\Delta T}} + \frac{1}{\Delta T}, & t = t_{d_i} \end{cases} \quad (15)$$

$$\mathbb{S}''_{\Delta T_1, \Delta T_2}(t) = \frac{\mathbb{S}'_{\Delta T_1}(t) - \mathbb{S}'_{\Delta T_2}(t)}{\Delta T_2 - \Delta T_1}$$

Topic Inference

while stop criterion is not satisfied:

  for $h = 1 \ldots H$ (in parallel)

    for $k = 1 \ldots K$

      fixing $\boldsymbol{a}$ and $\{\boldsymbol{p}_{k'}^{(h)}\}_{k' \neq k}$, use Newton-Raphson approach to find best $\boldsymbol{p}_k^{(h)}$ based on $\frac{\partial f}{\partial \boldsymbol{p}_k^{(h)}}$ and $\frac{\partial^2 f}{\partial \boldsymbol{p}_k^{(h)} \partial \boldsymbol{p}_k^{(h)T}}$

    endfor

  endfor

  fixing $\{\boldsymbol{p}_k^{(h)}\}_{k=1}^{K}$, use the Newton-Raphson approach to find the best $\boldsymbol{a}$ based on $\frac{\partial f}{\partial \boldsymbol{a}}$ and $\frac{\partial^2 f}{\partial \boldsymbol{a} \partial \boldsymbol{a}^T}$

endwhile

Experiment

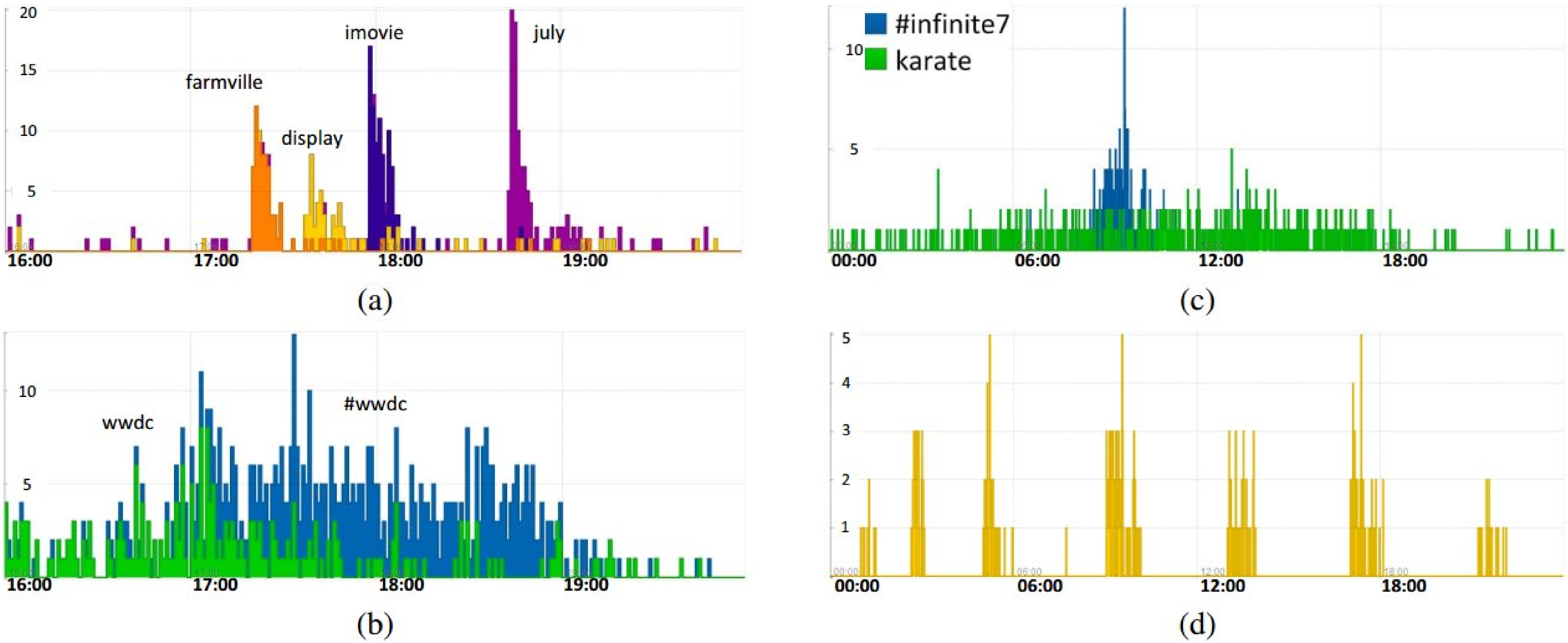


Figure 6. Case studies. (a)-(b) Apple WWDC 2010; (c) events *infinite7* and *karate*; (d) detected bursty topic created by spam.

Real-Time Disease Surveillance Using Twitter Data: Demonstration on Flu and Cancer

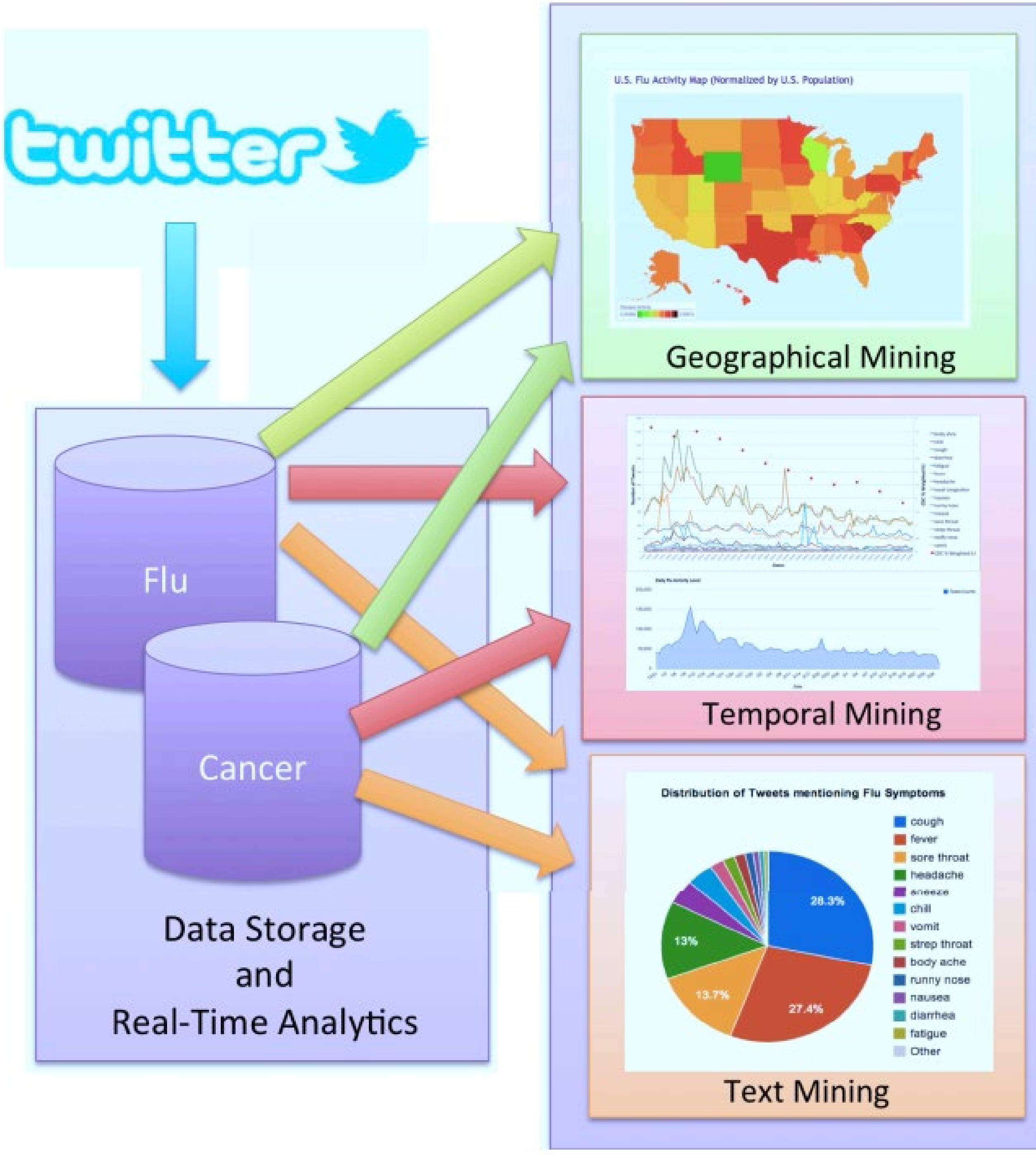
Real-Time Disease Surveillance Using Twitter Data: Demonstration on Flu and Cancer

Kathy Lee Ankit Agrawal Alok Choudhary
EECS Department
Northwestern University
Evanston, IL USA
{kml649, ankitag, choudhar}@eecs.northwestern.edu

2013 SIGKDD(CCF A)

Method

- The proposed system continuously downloads flu and cancer related twitter data using Twitter streaming API
- They apply spatial, temporal, and text models on this data to discover national flu and cancer activities and popularity of disease-related terms.
- The output of the three models is summarized as pie charts, time-series graphs, and US disease activity maps on our project website [1][2] in real time.



Details

- Geographical Analysis :

The dataset for geographic analysis is all users who mention ‘flu’ or ‘cancer’ and have a valid US state info (e.g., ‘Evanston, IL’, ‘somewhere in NY’) in their home location field.

- Temporal Analysis :

Disease Daily Activity Timeline The data for flu/cancer timeline is created by counting the number of tweets mentioning ‘flu’ or ‘cancer’ generated daily.

- Text Analysis :

We are interested in investigating the popularity of terms used in three categories: (1) disease types (2) symptoms (3) treatments

Experiment

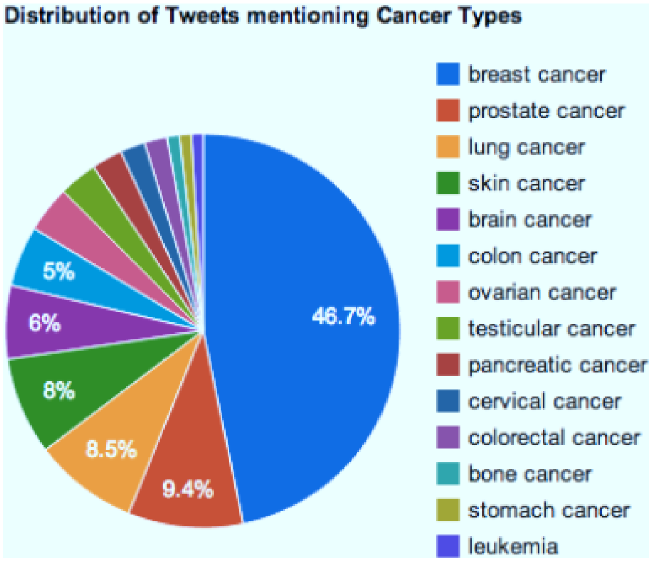
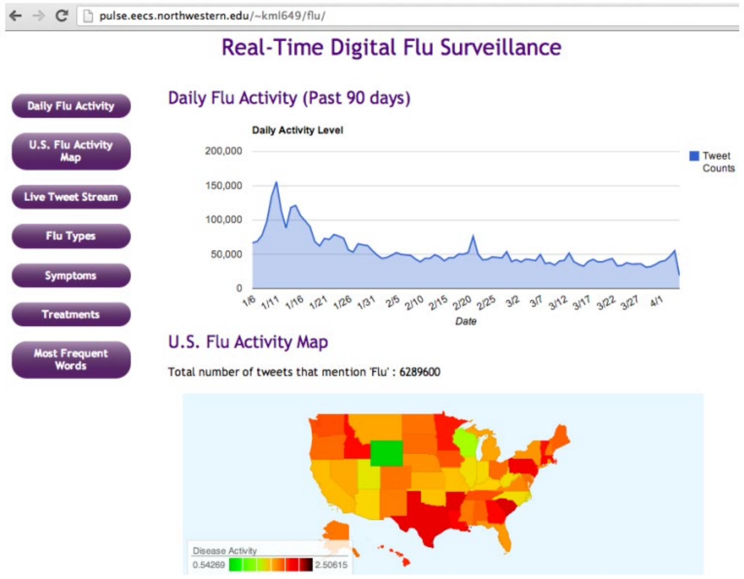
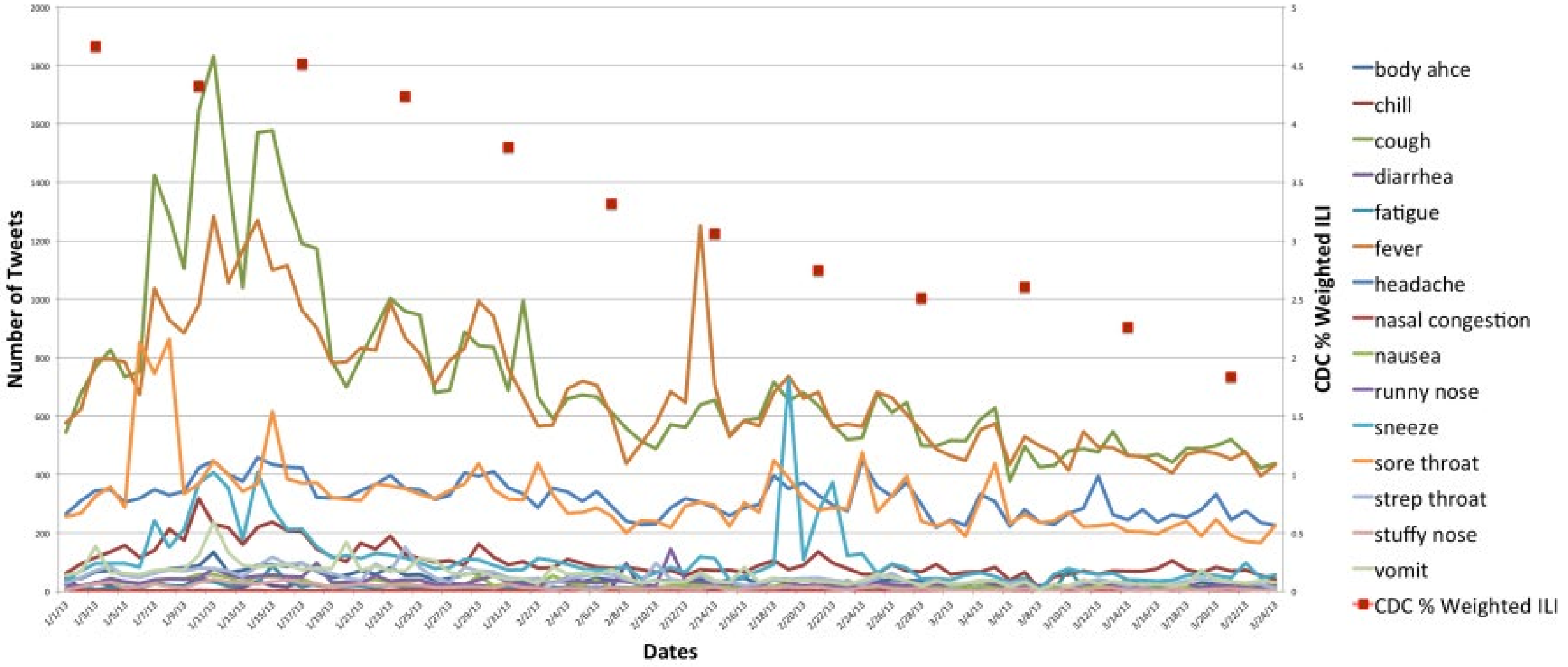


Figure 4: Cancer Types.





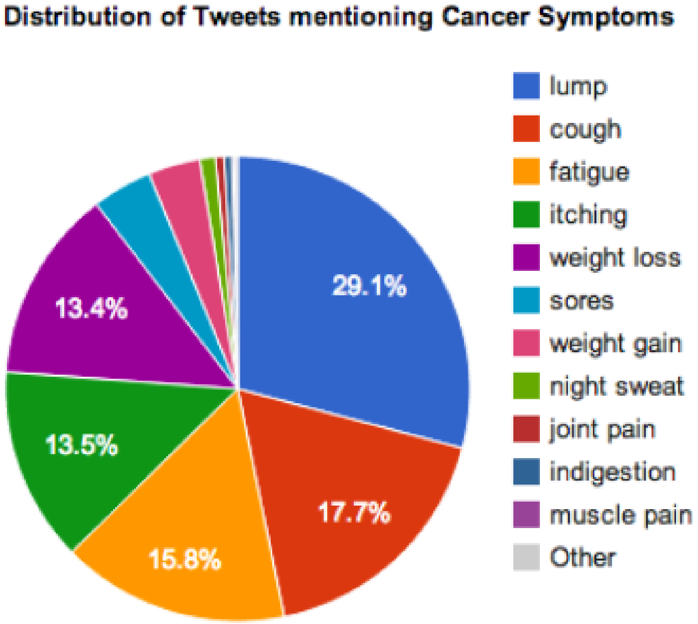


Figure 5: Cancer Symptoms.

A unified model for stable and temporal topic detection from social media data.

A Unified Model for Stable and Temporal Topic Detection from Social Media Data

Hongzhi Yin[†] Bin Cui[†] Hua Lu[‡] Yuxin Huang[†] Junjie Yao[†]

[†]*Department of Computer Science and Technology*

Key Laboratory of High Confidence Software Technologies, Peking University

[‡]*Department of Computer Science, Aalborg University*

[†]{bestzhi, bin.cui, huangyuxin, junjie.yao}@pku.edu.cn, [‡]luhua@cs.aau.dk

2013 ICDE(CCF A)

Motivation

- Given a document collection *C*, a user-time-keyword matrix *M*, and a social network *G*, our intension in this paper is to find interesting topics from *C* by exploiting the information captured in *M* and *G*.
- Task 1: Extracting Stable Topics. This task is to model and extract a set of stable topic models, $\Theta U = \{\theta i\}$, where $|\Theta U| = k1$ and k1 is a user specified parameter.
- Task 2: Detecting Temporal Topics. The task is to discover and detect a set of temporal topic models, $\Theta T = \{\theta j\}$, where $|\Theta T| = k2$ and k2 is a user specified parameter.

Method

$$p(w|u,t) = \lambda_U \sum_{\theta_i \in \Theta_U} p(\theta_i|u)p(w|\theta_i) + \lambda_T \sum_{\theta_j \in \Theta_T} p(\theta_j|t)p(w|\theta_j)$$

SYMBOL	DESCRIPTION
u, t, w	user, time stamp, keyword
U, T, W	set of users, time stamps and keywords
$M[u, t, w]$	frequency of w used by u within time stamp t
λ_U, λ_T	parameter controlling the branch selection
θ_i	stable topic indexed by i
θ_j	temporal topic indexed by j
Θ_U, Θ_T	stable and temporal topic set

$$L(C) = \sum_U \sum_T \sum_W M[u, t, w] \log p(w|u,t)$$

Enhancement of the model

- 1 $\mathcal{O}(\mathcal{C}, G) = L(\mathcal{C}) - \lambda R(\mathcal{C}, G)$ $R(\mathcal{C}, G) = \frac{1}{2} \sum_{(u,v) \in E} \pi(u, v) \sum_{\Theta_U} (p(\theta_i|u) - p(\theta_i|v))^2$
- 2 $\mathcal{O}(\mathcal{C}, T) = L(\mathcal{C}) - \xi R(\mathcal{C}, T$ $R(\mathcal{C}, T) = \sum_{t=1}^{|T|-1} \sum_{\Theta_T}^{k} (p(\theta_j|t) - p(\theta_j|t+1))^2$
- 3 Burst-Weighted Smoothing

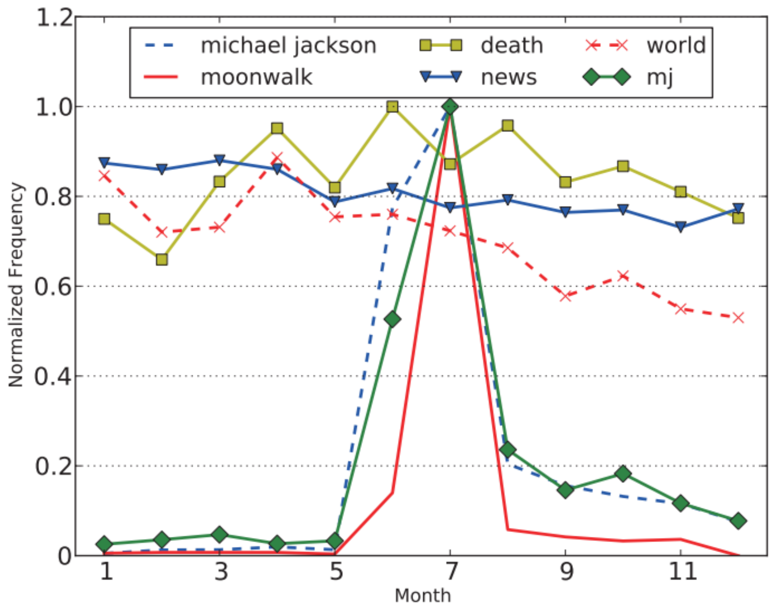


Fig. 2. Normalized Word Frequency Distribution on "Michael Jackson's Death" in 2009

Experiment

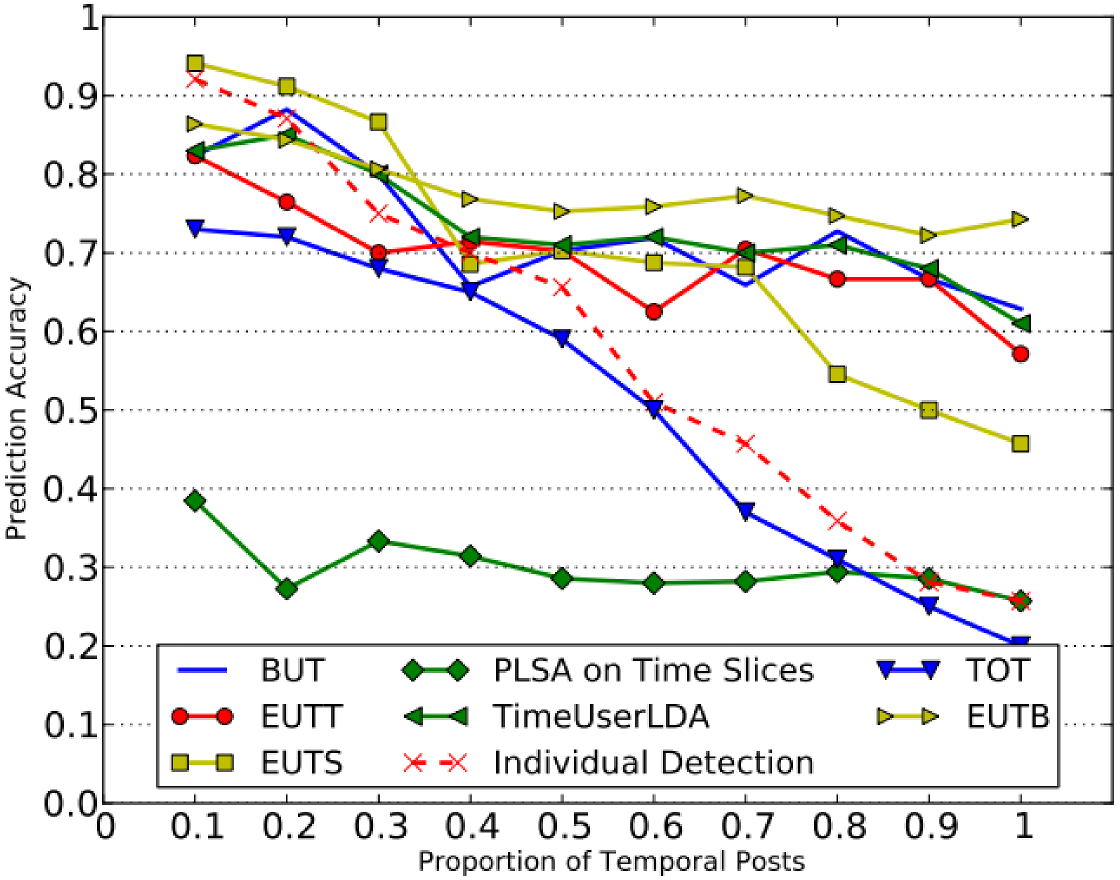


Fig. 3. Accuracy Curve of Time Stamp Prediction

Topic Detection Approach		N-Variance	Difference
BUT	stable topics	0.36	0.95
	temporal topics	1.31	
EUTS	stable topics	**0.21**	1.05
	temporal topics	1.26	
EUTT	stable topics	0.38	0.92
	temporal topics	1.30	
EUTB	stable topics	0.26	**1.34**
	temporal topics	**1.60**	
TOT	stable topics	0.39	0.11
	temporal topics	0.50	
Individual Detection	stable topics	0.38	0.61
	temporal topics	0.99	
TimeUserLDA	stable topics	0.39	0.93
	temporal topics	1.32	
Twitter-LDA	stable topics	0.38	0.58
	temporal topics	0.96	

	Excellent	Good	Poor
EUTB	42.5%	32.5%	25%
TOT	10%	40%	50%
Individual Detection	20%	37.5%	42.5%
TimeUserLDA	29.5%	38%	32.5%
Twitter-LDA	13.5%	39%	47.5%