Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence

User Group Oriented Temporal Dynamics Exploration

Zhiting Hu[1], **Junjie Yao**[2], **Bin Cui**[1]
[1]Department of Computer Science,
Key Lab of High Confidence Software Technologies (MOE), Peking University
[2]University of California, Santa Barbara
zhitinghu@gmail.com, bin.cui@pku.edu.cn,
jjyao@cs.ucsb.edu

Abstract

Temporal online content becomes the zeitgeist to reflect our interests and changes. Active users are essential participants and promoters behind it. Temporal dynamics becomes a viable way to investigate users. However, most current work only use global temporal trend and fail to distinguish such fine-grained patterns across groups. Different users have diverse interest and exhibit distinct behaviors, and temporal dynamics tend to be different.

This paper proposes GrosToT (Group Specific Topics-over-Time), a unified probabilistic model to infer latent user groups and temporal topics at the same time. It models group-specific temporal topic variation from social content. By leveraging the comprehensive group-specific temporal patterns, GrosToT significantly outperforms state-of-the-art dynamics modeling methods. Our proposed approach shows advantage not only in temporal dynamics but also group content modeling.

The dynamics over different groups vary, reflecting the groups' intention. GrosToT uncovers the interplay between group interest and temporal dynamics. Specifically, groups' attention to their medium-interested topics are *event-driven*, showing rich bursts; while its engagement in group's dominating topics are *interest-driven*, remaining stable over time.

Introduction

Social media, has become a pervasive and convenient platform for billions of users to share their thoughts and feelings every day. Prominent examples include social network and micro-blog service Facebook, Twitter and Weibo[1]. Active users are essential participants and promoters behind it. Users are not alone, and they form the interest groups explicitly/implicitly. As the basis for user participation and engagement, user groups become a critical factor in participating and promoting social media dynamics (Zhang, Wang, and Feng 2013; Danescu-Niculescu-Mizil et al. 2013).

Different users enjoy varying behavior patterns and propensities, leading to distinct temporal dynamic behaviors in response to popular topics or breaking events. One example of group dynamics diversity is illustrated in Figure 1. We present the temporal distributions of topic "Astronomy" in two user groups of a micro-blog site, interested in "Astronomy" and "Sports", respectively. We can clearly observe highly different patterns. The huge burst in group (b) coincides with the meteorite crash in Russia, Jan 15, 2013. In contrast, though group (a) exhibited increasing activities at that time, the degree is not remarkable.

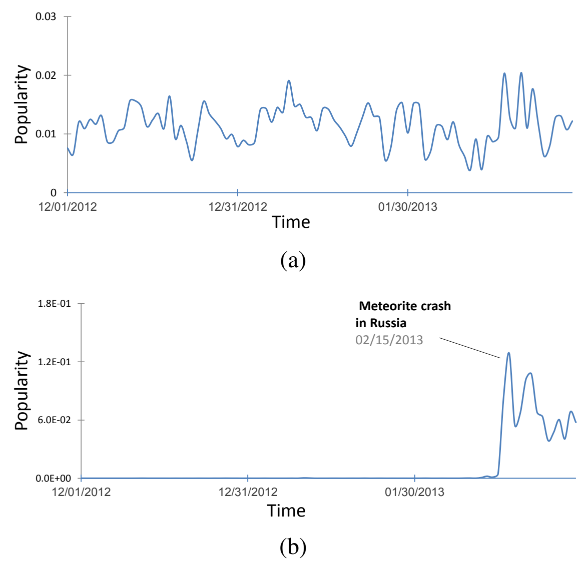

Figure 1: Temporal dynamics of topic "Astronomy" within two groups, namely, (a) the group mainly interested in "Astronomy", and (b) the group mainly interested in "Sports", from Dec 2012 to Feb 2013. The breaking event of Russian meteor explosion can be easily identified from (b).

Temporal dynamic patterns across different groups have practical implications for content extraction and group modeling. It opens up new insights into temporal dynamics mechanism, e.g., how users' interest correlates with their temporal behaviors; and how a piece of content attracts attentions from different users.

In this paper, we combine user group and social media dynamic analysis by exploring *group-specific topic temporal dynamics*. Specifically, we extract and analyze temporal variation patterns of topics within different user groups at the same time.

[1]facebook, twitter.com, weibo.com

Though a growing line of research has focused on the issues raised by the rich temporal variation of social media content (Matsubara et al. 2012; Yang and Leskovec 2011; Wang and McCallum 2006), they are limited in the aggregated dynamics of social media, and failed to distinguish such fine-grain patterns across groups. This problem can be even challenging due to volatile user behaviors. Groups can enjoy interest in multiple topics with various levels, while topics can exhibit diverse temporal patterns within different groups. Such interdependence requires simultaneous extraction of groups and topics, and accurately modeling the intricate correlation is also nontrivial.

This paper tackles the above challenges by presenting GrosToT (Group Specific Topics-over-Time), a unified model that uncovers latent groups and topics. It also extracts group-specific topic temporal dynamics. Basic idea of GrosToT is to leverage temporal content characters to capture groups' changing concern on various topics. Specifically, we associate each group with a distribution over topics, modeling its diverse interest. Each topic is also associated with a set of distributions over time, one for each group, flexibly capturing the different dynamic patterns within different groups. Though seeming challenging, we provide a well designed generative model structure to guarantee the performance.

We investigate the performance of GrosToT on a large-scale micro-blog dataset consisting of 14M posts generated by 0.52M users, spanning three months period, from Dec 2012 to Feb 2013. CrosTot shows significant improvement over state-of-the-art temporal modeling methods, not only in topic extraction but also the temporal prediction. We also find novel interplay phenomenons between user groups and temporal dynamics. Group's attention to their medium-interested topics are *event-driven*, exhibiting bursty engagement in the particular topic. In contrast, dominated topics of a group usually bear stable popularity for this topic along the time, showing an *interest-driven* difference. Take the example of online marketing: with the better temporal modeling under user group angle, ad campaigns can be designed and be launched specifically to different user groups and time slots.

To summarize, we make the following contributions in our work.

- We identify the problem of topic temporal variation across user groups. To the best of our knowledge, this angle has not been offered by previous research.
- We propose a unified probabilistic model, GrosTOT, which uncovers the topics and groups as well as captures the group-specific temporal dynamics of topics. This model achieves better performance over state-of-the-art methods.
- We present and analyze the interplay between group interest and temporal dynamics. This further demonstrates the potential of GrosToT in temporal extraction and group modeling.

The rest of this paper is organized as follows: §2 reviews related work; §3 formulates the problem and describes the proposed model; §4 presents our experimental results and §5 shows the analysis cases; and finally, §6 concludes this paper.

Related Work

There has been an increasing research interest of online temporal dynamics. Both content and structure based methods are proposed.

Topic modeling approaches are used to extract temporal content (Blei, Ng, and Jordan 2003; Wang and McCallum 2006; Blei and Lafferty 2006). In Latent Dirichlet Allocation(LDA), documents are modeled as a distribution over a shared set of topics, while topics themselves are distributions over words. Topics Over Time (TOT) (Wang and McCallum 2006) jointly models the text and time stamp by assuming that both words and time stamps are generated by latent topics. Specifically, time stamps are drawn from a Beta distribution, which only allows a unimodal temporal distribution over time for each topic. EUTB (Yin et al. 2013) distinguishes stable topics (e.g. topics on user interest) and bursty topics (e.g. topics on emergencies) in a unified PLSA-based model. It models the topic distribution of time periods by assuming that topics are generated by users or time periods. Another set of topic model based methods makes Markovian assumption on topic variation. They divide time into epoches, and assume that topics evolve based on their states in the previous epoch (Blei and Lafferty 2006; Zhang et al. 2010; Ren, Dunson, and Carin 2008; Hong et al. 2011). In contrast, our proposed model, GrosToT, utilizes multinomial distribution which is able to model multimodal variation and thus is more flexible for capturing topic temporal dynamics.

Some recent efforts have been made on analyzing temporal behaviors of users w.r.t. certain aspects. A meme-tracking approach is presented in (Leskovec, Backstrom, and Kleinberg 2009) to monitor dynamics of news cycle by identifying and tracking short textual phrases. The K-Spectral Centroid clustering algorithm (Yang and Leskovec 2011) finds out six classes of temporal variation pattern in online media. A flexible analytical model SPICKM (Matsubara et al. 2012) generalizes theoretical models for the rise and fall patterns of propagation. A framework is proposed in (Danescu-Niculescu-Mizil et al. 2013) to track linguistic change of two online communities and identify a determined two-stage lifecycle w.r.t. community members' susceptibility to linguistic evolution. Public opinion shift in response to political events is explored (Lin et al. 2013), by categorizing twitter users into two biased sets and performing sentiment analysis in these two groups respectively. These works focus on profiling user groups' reaction to some particular topics or events. Our model is much more general in the sense of automatically identifying groups and topics, as well as their interplay from social media stream.

All the above methods only explore aggregated dynamics, and ignore the fact that different user groups with diverse propensities and behavior patterns tend to have varying temporal dynamics. GrosToT distinguishes patterns of topic temporal dynamics across different groups, which allows fine-grained analysis of temporal dynamics.

Group Specific Topics-over-Time Model

Preliminaries

We introduce the notation used in GrosToT and formally define our problem. Without loss of generality, we use micro-blog as our problem setting. Tweets are messages posted by users. We consider a stream of D tweets $\mathcal{D} = \{d_1, d_2, \ldots, d_D\}$. Each tweet d_i is generated by a user $u_i \in \{1, 2, \ldots, U\}$ at timestamp $t_i \in \{1, 2, \ldots, T\}$. Here U and T denote the total number of users and time points, respectively. Each tweet d_i contains a set of words $\mathcal{W}_i = \{w_{i1}, w_{i2}, \ldots, w_{iN_i}\}$ from a given vocabulary. Here N_i denotes the number of words in d_i.

Groups and topics are both latent factors to be extracted, indexed by $g \in \{1, 2, \ldots, G\}$ and $k \in \{1, 2, \ldots, K\}$ respectively. Here we assume there are G groups and K topics in the social media.

For each topic k and group g, we define *group-specific topic temporal dynamics* as a multinomial distribution over time points η_{kg}. Our fundamental task is to uncover η_{kg} for each topic within each group from the input tweet stream. Later η_{kg} can be used for temporal analysis and group modeling.

Model Structure

Now we present the structure and generative process of proposed model, GrosToT (Group Specific Topics-over-Time). Tweets are usually written by users under certain social context. Therefore, the temporal characters of tweets effectively represent the group's real-time concern on various topics and events, as well as provide opportunities to accurately capture group-specific temporal topic dynamics.

Each topic k has a distribution over words ϕ_k. We denote the number of topics as K. Unlike long documents such as news articles which are modeled with a mixture of topics (Blei, Ng, and Jordan 2003), a tweet is very short (i.e. less than 140 characters) and thus is most likely to be about a single topic (Diao et al. 2012; Zhao et al. 2011). Each tweet is therefore suitable to be assigned a single latent topic responsible for generating its words. This facilitates the following model structure.

We assume there are G groups. Each group g is associated with a topic distribution θ_g to capture its interest in multiple topics with varying strengths. In real life, users are usually characterized by multiple group memberships (Xie, Kelley, and Szymanski 2013; Airoldi et al. 2008). We therefore associate each user u with a distribution over groups π_u.

As mentioned above, a tweet reflects the temporal context of a certain group. Hence each tweet is associated with a group indicating the group context it is related to, and its timestamp is then generated by the time distribution η_{kc} specific to the group and topic assigned to it.

Figure 2 shows the graphical representation of GrosToT. Consider a user u who writes a tweet d_i, she first selects the group context g_i according to group distribution π_u, and then chooses a topic z_i according the group's topic distribution θ_{g_i}. The words in the tweet are generated from the word distribution ϕ_{z_i} specific to that topic, and the timestamp is generated from the temporal distribution η_{kc} for that topic and group.

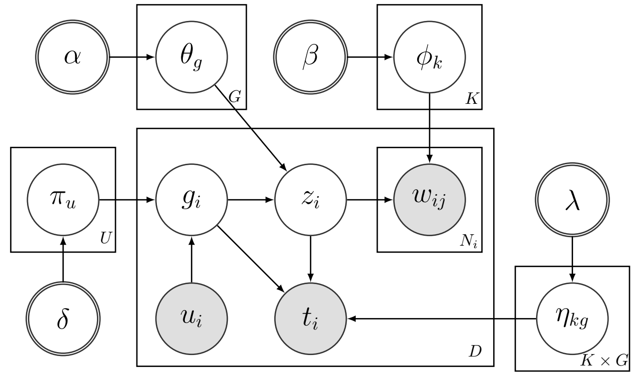

Figure 2: GrosToT model. A double circle indicates a hyperparameter; a single hollow circle indicates a latent variable; and a filled circle indicates an observed variable.

We therefore set up the following generative process for the tweet stream.

- For each group $g = 1, 2, \ldots, G$,
  - Draw the distribution over topics, $\theta_g|\alpha \sim$ Dirichlet(α).
- For each user $u = 1, 2, \ldots, U$
  - Draw the distribution over groups, $\pi_u|\delta \sim$ Dirichlet(δ).
- For each topic $k = 1, 2, \ldots, K$,
  - Draw the distribution over words, $\phi_k|\beta \sim$ Dirichlet(β).
  - For each group $g = 1, 2, \ldots, G$,
    - ∗ Draw the distribution over time stamps, $\eta_{kg}|\lambda \sim$ Dirichlet(λ).
- For each tweet $i = 1, 2, \ldots, D$,
  - Draw group indicator, $g_i|\pi_{u_i} \sim$ Multi(π_{u_i}).
  - Draw topic indicator, $z_i|\theta_{g_i} \sim$ Multi(θ_{g_i})
  - For each word $j = 1, 2, \ldots, N_i$,
    - ∗ Draw word, $w_{ij}|\phi_{z_i} \sim$ Multi(ϕ_{z_i}).
  - Draw time stamp, $t_i|\eta_{z_i g_i} \sim$ Multi($\eta_{z_i g_i}$).

Model Inference

Exact inference is intractable due to the coupling parameters in GrosToT model. We therefore use collapsed Gibbs Sampling (Griffiths and Steyvers 2004) to obtain samples of the hidden variable assignment, based on which the unknown parameters $\{\pi, \theta, \phi, \eta\}$ can be estimated. For simplicity we fix the hyperparameters to $\beta = \lambda = 0.01$, $\delta = 50/C$ and $\alpha = 50/K$.

The inference process is listed as follows.

We first sample the group indicator g_i for each post d_i:

$$P(g_i = g | z_i = k, t_i = t, \boldsymbol{g}_{-i}, \boldsymbol{z}_{-i}, \boldsymbol{t}_{-i}, .) \propto \frac{n_{i,g} + \delta}{n_{i,\cdot} + C\delta} \cdot \frac{n_{g,k} + \alpha}{n_{g,\cdot} + K\alpha} \cdot \frac{n_{gk,t} + \lambda}{n_{gk,\cdot} + T\lambda},$$

where $n_{i,c}$ is the number of tweets by user u_i assigned to group g; $n_{g,k}$ is the number of tweets assigned to group g and generated by topic k; $n_{ck,t}$ denotes the number of times that timestamp t is generated by group g and topic k; Marginal counts are represented with dots. All the counters mentioned above are calculated with the tweet d_i excluded.

We then sample the topic indicator z_i for each post d_i:

$$P(z_i = k | g_i = g, t_i = t, \boldsymbol{g}_{-i}, \boldsymbol{z}_{-i}, \boldsymbol{w}, \boldsymbol{t}_{-i}, .)$$
$$\propto \frac{n_{g,k} + \alpha}{n_{g,\cdot} + K\alpha} \cdot \frac{n_{gk,t} + \lambda}{n_{gk,\cdot} + T\lambda} \cdot \frac{\prod_{v=1}^{V} \prod_{q=0}^{n_{i,v}-1} (n_{k,v} + q + \beta)}{\prod_{q=0}^{n_{i,\cdot}-1} (n_{k,\cdot} + q + V\beta)},$$

where $n_{i,v}$ is the number of times word v occurs in the tweet d_i; $n_{k,v}$ is the number of times word v is assigned to topic k. Marginal counts are represented with dots. Note that $n_{k,v}$ and $n_{k,\cdot}$ is calculated with the tweet d_i excluded.

Extraction Performance

We conduct empirical study to evaluate GrosToT's modeling performance. Both topic extraction and temporal prediction tasks are investigated.

Experiment Setup

Data Set We use a large real dataset crawled from Sina Weibo[2], one of the most popular micro-blog platforms. We randomly sample users and get their streaming updates. To save the API quota, low frequent users are ignored since they have low contribution to groups' dynamics. After removing stop words and low quality posts, we obtain a dataset consisting of about 0.52M users, 14M posts and 112M words with a vocabulary of size 89K. The posts are distributed almost evenly in a three-month time period from Dec 2012 through Feb 2013. We use one day as the time window.

We randomly select 80% tweets as the training set while the remaining 20% as testing set.

Baselines We compare our proposed GrosToT with two state-of-the-art competitors, i.e., EUTB (Yin et al. 2013) and Topics over Time (TOT) (Wang and McCallum 2006). EUTB extracts temporal patterns by modeling topic distributions of time slots. Similar to GrosToT, TOT treats both words and time stamps as variables generated by latent topics. Both of these two approaches only uncover global trends of topics without distinguishing dynamic patterns across groups.

Topic Extraction

A fundamental requirement of temporal modeling is that it can extract intuitive and representative topics of groups or temporal events. We display one of the extracted topics in Figure 3. From the word cloud we see that the discussed topic is mainly about "Sports". It reflects not only general themes, e.g., "football" and "basketball", but also the seasonable focuses, e.g., "Real Madrid" and "Kobe Bryant".

We next proceed to a quantitative way to measure the topic extraction performance. Perplexity (Blei, Ng, and Jordan 2003) is a common metric in the topic modeling area,

[2]http://weibo.com

Figure 3: Word cloud of extracted topic "Sports".

measuring how well the words of test documents are represented by the word distribution of extracted topics. A lower perplexity value indicates better performance of the model.

For a test set of D posts, the perplexity score is measured by:

$$perplexity(\mathcal{D}_{test}) = \exp\left\{ - \frac{\sum_{i=1}^{D} \log p(\mathcal{W}_i)}{\sum_{i=1}^{D} N_i} \right\},$$

where $p(\mathcal{W}_i)$ is the probability of the words in the test document d_i. In the case of GrosToT, it is computed as:

$$p(\mathcal{W}_i) = \sum_{g} \pi_{u_i g} \sum_{k} \theta_{gk} \prod_{j} \phi_{k w_{ij}}.$$

Figure 4 shows the perplexity values for three competitors as the numbers of topics (i.e. K) and groups (i.e. G) change. Our proposed model consistently yields better performance, indicating higher quality of extracted topics. Suitable numbers of topics (i.e. 100) and groups (i.e. 50) give the best performance. The result justifies the advantage of leveraging group information of social users and extracting fine-grained representation of topic variation patterns.

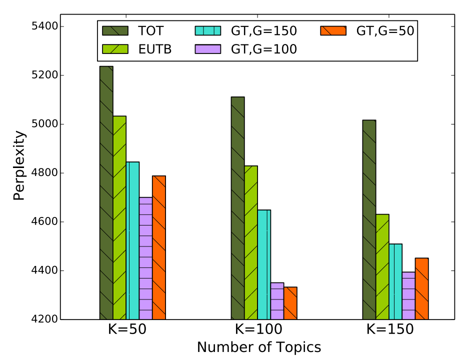

Figure 4: Topic extraction performance of different methods.

Temporal Dynamic Modeling

We then evaluate the model's capacity of capturing temporal dynamics by measuring the *time stamp likelihood* of a held-out test set, i.e., computing the likelihood of time stamps of previously unseen tweets based on their content. Specifically, GrosToT has the time stamp likelihood of a tweet d_i as:

$$\mathcal{L}_{d_i} = \sum_{g} \pi_{u_i g} \sum_{k} \theta_{gk} \eta_{kgt_i} \prod_{j} \phi_{k w_{d_i j}}.$$

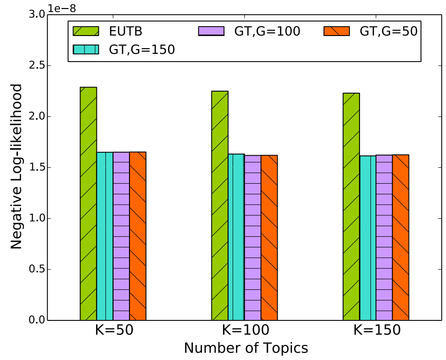

Figure 5: The time stamp likelihood of different methods.

A higher likelihood value indicates better generalization performance.

We compare GrosToT with EUTB which has been shown better performance over TOT (Yin et al. 2013). Figure 5 shows the negative log-likelihood values for the competitors with varying numbers of topics and groups. We find that GrosToT consistently outperforms the baselines. The reason is that our model incorporates diversity of different user groups' temporal patterns. In contrast, the baseline methods simply aggregate all temporal actions to infer topics' global trends. Such coarse-grained extraction fails to gain accurate representation of dynamics. It is also notable that the performance of these methods remains nearly stable as K and G vary.

Group and Dynamics Exploration

With the help of GrosToT, we analyze group characteristics and topic dynamics. For clarity we fix $K = 100$ and $G = 100$ in the following study. We present illustrative case studies to show the advantage of in-depth exploration. Besides, we also quantitatively report their correlations to reveal the interplay between groups and dynamics.

Group and Dynamics Cases

To identify breaking events related to a certain topic, e.g., "Sports", one would intuitively focus on the groups whose major interest lies in this topic. The following results, however, suggest that analyzing groups that are not regularly concerned with the particular topic may provide an easier way.

Figure 6 shows the topic distributions of two extracted groups, labeled as Group-A and Group-B, respectively. We find that the major theme inside each group has a clear topic focus. Members of Group-A mainly concentrate on topic "Sports". In contrast, though topic "Movie" dominates the activity of Group-B, the members also pay attentions to various other topics.

Figures 7 (a) and (b) demonstrate the temporal distributions of topic "Sports" within Group-A and Group-B respectively. From Figure 7 (b) we can easily identify four burst periods. For example, the burst at time stamp Jan 26, 2013 coincides with the Australian Open Final which the famous

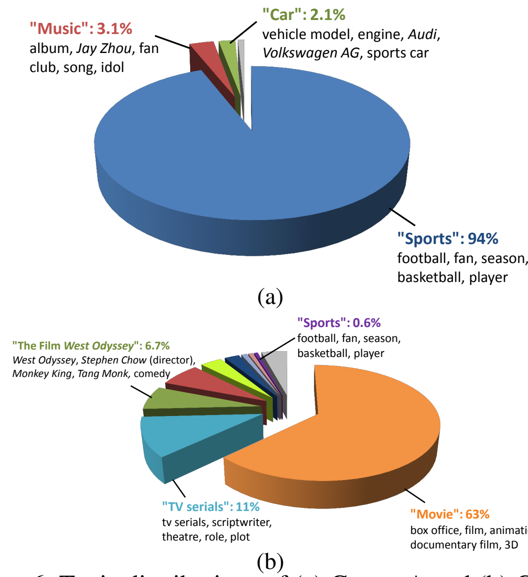

Figure 6: Topic distributions of (a) Group-A and (b) Group-B. For each topic, top 5 words w.r.t. its word distribution are listed; we manually label topic using a concise phrase; the percentage denotes the proportion of the topic in the group's topic distribution.

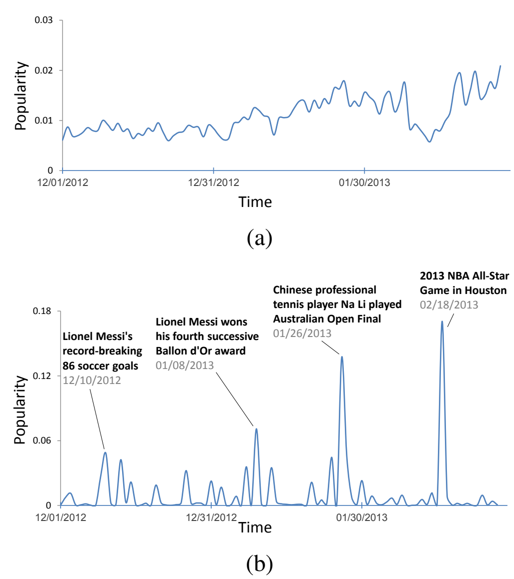

Figure 7: Temporal dynamics of topic "Sports" within (a) Group-A (b) Group-B. Group-A bears a probability of 94% on topic "Sports", while Group-B bears a probability of 0.6% on topic "Sports". Breaking events can be easily identified from (b).

Chinese professional tennis player *Na Li* attended. On the contrary, the timeline in Figure 7 (a) is much smoother and without clear bursts.

By checking the posts in each group, we find that Group-A does talk about the particular events detected by Figure 7 (b) when they happened. However, these bursty behaviors are concealed since members keep talking about sports in their daily life, rendering the events hard to be identified by the analysis of topic popularity variation. On the other hand, members of Group-B generate sports-related tweets only when significant sports events happened. Therefore a spike in its timeline typically corresponds to some bursty event in real life.

Group Behavior and Temporal Topics

We next focus on the interplay between group interest and fluctuation of topic popularity within groups. Such a new analysis can be valuable in understanding user's interest and behavior patterns in response to topics.

Group Interest and Fluctuation of Topic Popularity Group interest is modeled as a distribution over topics, i.e., θ. To quantitatively measure the fluctuation of topic popularity within groups, especially how intensive topics burst over time, we define the *burst degree* of topic temporal distribution by exploiting *sliding window*.

Specifically, we assume that the popularity of a stable topic in a sliding window can be modeled by a Gaussian distribution $G(\mu, \sigma^2)$. For a topic k within a group g, we then compare its real popularity at time t, i.e., η_{kgt} with sliding window containing its recent history periods $[t-n, t-1]$, and define topic burst degree at t as:

$$burst\_degree(k, g, t) = \frac{\eta_{kgt} - \mu_t}{\sigma_t},$$

where μ_t and σ_t are mean and standard deviation estimated by the popularity in the recent history window of t:

$$\mu_t = \frac{1}{n} \sum_{i=t-n}^{t-1} \eta_{kgt},$$

$$\sigma_t = \sqrt{\frac{1}{n} \sum_{i=t-n}^{t-1} (\eta_{kgt} - \mu_t)^2}.$$

We finally compute the sum over all the time stamps, i.e., $\sum_t burst\_degree(k, g, t)$ as the aggregate burst degree of the temporal distribution of topic k within group g.

Correlation between Groups and Temporal Topics Figure 8 displays the correlation between group interest and burst degree of topic temporal distributions. We can clearly conclude that topics tend to exhibit higher burst degree within those groups which bear topic probability between 0.001% to 10%; while the average burst degree is near zero if topic probability is smaller than 0.001% or larger than 10%.

This result is reasonable, and also revealed by the previous examples shown in Figure 7. For medium-interested topics (i.e. with probabilities between 0.001% to 10%), group users are not concerned with them in the daily life, but their attentions would still be greatly drawn when significant events happen. This *event-driven* attention leads to the bursty temporal patterns. In contrast, members in groups with extremely high preferences for the particular topic keep talking about it throughout the time period. This *interest-driven* behaviors result in stable temporal distributions without clear bursts. (Note that for topics with extremely low probabilities ($< 0.001\%$), the results would largely be too noisy regardless of the large data set, thus we only focus on the previous two categories of topics.)

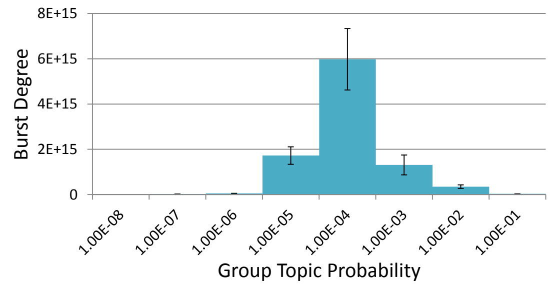

Figure 8: Correlation between group topic probability θ_{gk} and burst degree of group-specific topic temporal distribution η_{kg}. E.g., Topics within those groups which have topic probability between 1.0E-04 to 1.0E-03 have an average burst degree of 6.0E15.

The above exploration clearly reveals that the different behavior patterns of user groups response to varying kinds of topics. This research enables fine-grained insights into the characteristics of topic temporal patterns and at the same time provide practical guidance for multiple areas such as online marketing and content management.

Conclusion

In this paper, we have addressed the problem of group oriented temporal dynamics extraction. We discussed the drawbacks of current global temporal extraction methods and presented a unified probabilistic generative model, i.e., GrosToT (Group Specific Topics-over-Time). GrosToT can simultaneously uncover latent user groups and temporal topics and extract group-specific topic temporal variation.

We demonstrated that GrosToT shows advantage not only in temporal dynamics modeling but also group content exploration. GrosToT gains significant performance improvement over state-of-the-arts methods in modeling temporal dynamics. Novel findings in varying groups and rich dynamics were also provided.

This study shows that GrosToT can improve temporal topic extraction and user group modeling at the same time. The intuitive result opens interesting directions in the future work. Probably the most exciting one is the integration of theme spectrum of temporal evolution and group granularity for comprehensive temporal modeling.

Acknowledgments

This research is supported by the National Natural Science Foundation of China under Grant No. 61272155, and 973 program under No. 2014CB340405.

References

Airoldi, E. M.; Blei, D. M.; Fienberg, S. E.; and Xing, E. P. 2008. Mixed membership stochastic blockmodels. *JMLR* 9:1981–2014.

Blei, D. M., and Lafferty, J. D. 2006. Dynamic topic models. In *Proc. of ICML*, 113–120.

Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2003. Latent dirichlet allocation. *JMLR* 3:993–1022.

Danescu-Niculescu-Mizil, C.; West, R.; Jurafsky, D.; Leskovec, J.; and Potts, C. 2013. No country for old members: User lifecycle and linguistic change in online communities. In *Proc. of WWW*, 307–318.

Diao, Q.; Jiang, J.; Zhu, F.; and Lim, E.-P. 2012. Finding bursty topics from microblogs. In *Proc. of ACL*, 536–544.

Griffiths, T. L., and Steyvers, M. 2004. Finding scientific topics. *PNAS* 101(Suppl 1):5228–5235.

Hong, L.; Yin, D.; Guo, J.; and Davison, B. D. 2011. Tracking trends: incorporating term volume into temporal topic models. In *Proc. of KDD*, 484–492.

Leskovec, J.; Backstrom, L.; and Kleinberg, J. 2009. Memetracking and the dynamics of the news cycle. In *Proc. of KDD*, 497–506. ACM.

Lin, Y.-R.; Margolin, D.; Keegan, B.; and Lazer, D. 2013. Voices of victory: a computational focus group framework for tracking opinion shift in real time. In *Proc. of WWW*, 737–748.

Matsubara, Y.; Sakurai, Y.; Prakash, B. A.; Li, L.; and Faloutsos, C. 2012. Rise and fall patterns of information diffusion: model and implications. In *Proc. of KDD*, 6–14.

Ren, L.; Dunson, D. B.; and Carin, L. 2008. The dynamic hierarchical dirichlet process. In *Proc. of ICML*, 824–831.

Wang, X., and McCallum, A. 2006. Topics over time: a non-markov continuous-time model of topical trends. In *Proc. of KDD*, 424–433.

Xie, J.; Kelley, S.; and Szymanski, B. K. 2013. Overlapping community detection in networks: The state-of-the-art and comparative study. *ACM Comput. Surv.* 45(4):43:1–43:35.

Yang, J., and Leskovec, J. 2011. Patterns of temporal variation in online media. In *Proc. of WSDM*, 177–186.

Yin, H.; Cui, B.; Lu, H.; Huang, Y.; and Yao, J. 2013. A unified model for stable and temporal topic detection from social media data. In *Proc. of ICDE*, 661–672.

Zhang, J.; Song, Y.; Zhang, C.; and Liu, S. 2010. Evolutionary hierarchical dirichlet processes for multiple correlated time-varying corpora. In *Proc. of KDD*, 1079–1088.

Zhang, W.; Wang, J.; and Feng, W. 2013. Combining latent factor model with location features for event-based group recommendation. In *Proc. of KDD*, 910–918.

Zhao, W. X.; Jiang, J.; Weng, J.; He, J.; Lim, E.-P.; Yan, H.; and Li, X. 2011. Comparing twitter and traditional media using topic models. In *Proc. of ECIR*, 338–349.