World Wide Web
https://doi.org/10.1007/s11280-018-0545-6

Embedding and predicting the event at early stage

Zhiwei Liu[1,2] · Yang Yang[2] · Zi Huang[2] · Fumin Shen[2] · Dongxiang Zhang[2] · Heng Tao Shen[2]

Received: 9 October 2017 / Revised: 2 March 2018 / Accepted: 12 March 2018

Abstract Social media has become one of the most credible sources for delivering messages, breaking news, as well as events. Predicting the future dynamics of an event at a very early stage is significantly valuable, e.g, helping company anticipate marketing trends before the event becomes mature. However, this prediction is non-trivial because a) social events always stay with "noise" under the same topic and b) the information obtained at its early stage is too sparse and limited to support an accurate prediction. In order to overcome these two problems, in this paper, we design an event early embedding model (EEEM) that can 1) extract social events from noise, 2) find the previous similar events, and 3) predict future dynamics of a new event with very limited information. Specifically, a denoising

This article belongs to the Topical Collection: *Special Issue on Geo-Social Computing*
Guest Editors: Guandong Xu, Wen-Chih Peng, Hongzhi Yin, Zi (Helen) Huang

✉ Yang Yang
dlyyang@gmail.com

Zhiwei Liu
jim96liu@gmail.com

Zi Huang
huang@itee.uq.edu.au

Fumin Shen
fumin.shen@gmail.com

Dongxiang Zhang
zhangdo@uestc.edu.cn

Heng Tao Shen
shenhengtao@hotmail.com

1 Guizhou Provincial Key Laboratory of Public Big Data, GuiZhou University, Guizhou, Guiyang, 550025, China

2 Center for Future Media and School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China

Published online: 22 March 2018

approach is derived from the knowledge of signal analysis to eliminate social noise and extract events. Moreover, we propose a novel predicting scheme based on locally linear embedding algorithm to construct the volume of a new event from its k nearest neighbors. Compared to previous work only fitting the historical volume dynamics to make a prediction, our predictive model is based on both the volume information and content information of events. Extensive experiments conducted on a large-scale dataset of Twitter data demonstrate the capacity of our model on extract events and the promising performance of prediction by considering both volume information as well as content information. Compared with predicting with only the content or the volume feature, we find the best performance of considering they both with our proposed fusion method.

Keywords Social events · Volume dynamics · Content information · Early prediction

1 Introduction

Recent years have witnessed the tremendous power of social media reshaping the ways of generating, distributing and consuming information [12, 39], such as breaking news, topics and events [23]. Numerous research endeavors have been dedicated to characterizing social messages and events. For instance, on Twitter,[1] tweets are attached with timestamps, which can assist detecting the information flow [43] and depicting the growth and decay of certain events [9, 14, 15]. Besides, provided tremendous similar tweets with the same timestamps, several research endeavors have been made to characterize the dynamics of hashtags [16, 28], memes [2, 18] and events volume [29], predict the temporal dynamics of information [8, 10], or perform both of them [11, 24, 25].

Events delivered in social media environment are of potentially high influence to the public, such as presidential election, earthquake and terrorist attack maximizing the value of the new events. It is essential to timely identify and characterize them. In this paper, we study the problem of predicting about time-series volume of events. Different from majority of existing work [2, 24, 29] that build mathematic models to fit certain types of mature events, we utilise the limited information carried by the new events to foresee their future volume dynamics at a very early stage.

To facilitate the understanding of the organization of social data, we first clarify several concepts, including topic, event, volume and social noise. If a set of messages are related to some common subject, we define the set of messages as a **topic**. Generally, each topic comprises various underlying constituent parts which lie in different periods. Each constituent part reveals some important aspects of the topic. Hence, we define the constituent part as **event**. We illustrate the relationship between topic and event in Figure 1.

The number of messages in a topic (event) is always dynamically changing over time. To draw it clearly, we define **volume** illustrating the total number of messages in a predefined time window. **Topic volume** is the number of messages in a topic. Particularly, in Twitter, topic volume denotes the number of tweets with the same hashtags published in a certain time window (e.g., daily and hourly). Similarly, the searching interest volume is another type of topic volume provided in Google Trends.[2] **Event volume** is the number of messages

[1] https://twitter.com/

[2] https://www.google.com/trends/

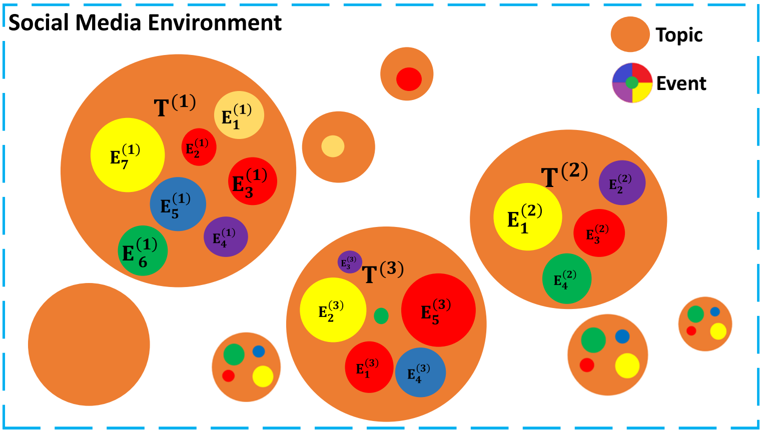


Figure 1 The organization of social data. An individual topic may contain several events. The orange color is stand for the topic. Some topic, e.g. the down-left topic, are of no event

in a constituent part of a topic. For instance, Figure 2 shows the topic volume dynamics of *Apple* in Google Trends. As seen, this topic is composed of several events [46], such as *Swift for iOS*, *Apple Special Event*, and each event has its own daily volume throughout the whole searching period.

The volume of an event rises from the emergence of the event and decays to zero when the event ends. However, events are not always readily intelligible within a topic in that there also exists social noise. In our context, we define **social noise** as the ever-lasting irrelevant part to the events (analogous to white noise in signal processing). Social noise is an independent variable which is unrelated to the event. For example, in Figure 2, we can observe obvious ascending trends of the searching volume of *Apple* when vital events

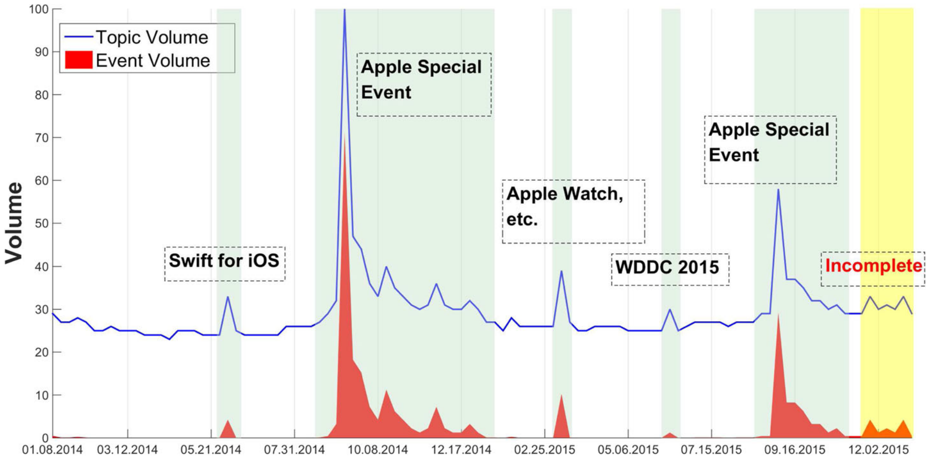


Figure 2 Search Interest for *Apple*. The blue line is the weekly topic volume of Apple, which is the hybrid of 5 events and random social noise volume. After denoising with our method, we can extract the 5 events volume shown in the red area and locate the start and end timestamp of them the same that in the real world. The maximum volume is set to 100 in this time duration

happen, such as *Apple Special Event*. However, there also exists regular volume of messages even no particular events happen, because searching request for information irrelevant to the events, such as fruit apple, keeps coming to Google. Such regular volume is defined as social noise volume in this paper. To notice, event volume and social noise volume may embody different properties. While the former grows and decays fast, the later one varies little under most circumstances. In this sense, events are more similar to subtopics under one topic [32] and social noise is more analogous to the ordinary public attention irrelevant to current events. In the predicting process, we focus more on the event part of a topic while ignoring the noise part. But in real life, events always exist with social noise in a topic at the same time. Thus, it is a challenge to extra events from a topic. To handle the problem of social noise, some of the existing research work regards social noise as the long-tail effect of social events, and thus builds response models to extract the true event volume [29]. Also, wavelet transform is introduced in extracting events in *Twitter* [31]. However, for this paper, in order to eliminate the influence to the great extent, we intend to eliminate social noise such that the event volume can be totally separated from topic volume.

To make a prediction of event volume, most existing work only consider the volume feature thus using the time-series analysis methods. However, the content information is also useful because events with the same content information, which implies people's attitude, often have similar dynamics. Thus, we predict the future dynamics of events at early stage with both volume and content feature [42]. Our predicting method is based on locally linear embedding algorithm [22]. For a new coming event, we try to find its neighbors and construct its future dynamics from previous events volume. To the best of our knowledge, it is the first attempts to predict the social event volume dynamics with only early stage information. The major contributions of our model are summarised as follows:

- **Early Prediction:** We propose a novel event early embedding model to predict event volume trend for the future dynamic given limited data at a very early stage.
- **Multi-Feature Fusion:** We construct a new event using both the volume feature and content feature of the k nearest neighbors of the event, where a novel similarity function is designed specifically for dealing with the incompatibility of different features.
- **Novel Evaluation Metric:** We define a novel divergence function evaluating the difference of our prediction and ground truth, which is a better evaluation metric than traditional methods.

The rest of the paper is organized as follows. We review the related work in Section 2. In Section 3, we present some preliminaries and the details of the proposed model. Experiments are reported and analyzed in Section 4, followed by the conclusion in Section 5.

2 Related work

There exist methods that try to model events and thus make a prediction from different angels. However, few of them can make a prediction at a very early stage. Moreover, most of the predictive models only predict the popularity of the events while ignoring the future trends. Different from previous work, we build a event predicting model which can not only extract the events from topics [30] but also make a prediction of social event trends at a very early stage. In this section, we review the related work on detecting, modeling, and predicting social events.

2.1 Social event detection

The methods of event detection in social media can be classified into *content-view* and *signal-view*. The former detects events by clustering the content information in corpus, while the latter extracts events by analyzing the time-series signals of topic volume.

From *content-view*, events can be detected by clustering documents [35] based on the semantics distance between documents [33, 37]. The goal of this method is to identify stories in several continuous news streams that pertain to new or previously unidentified events. Both hierarchical and incremental non-hierarchical clustering algorithms can be applied to explore events by using of content information. Additionally, with the booming and easy access of media resources, events can be detected by retrieving the information carried by online photos [13, 36, 41]. Event detection is treated as an unsupervised clustering problem by using the user-provided textual description. Also, supervised event detection is solved by learning a multi-feature similarity metrics [3, 19, 47].

From *signal-view*, event detection is a problem related to time-series analysis [38, 40], either in time domain or in frequency domain. By modeling the events volume as a convolution of the events importance and response function, events (volume) can be extracted from topic (volume) by applying the deconvolution of topic volume with predefined response funtion [29]. Except convolution, wavelet transformation can be utilized in extracting events volume from topic volume [31]. In this paper, we introduce a social noise volume as a important constituent part of topic volume. As the main task of our model is to make a prediction, we design a simple yet effective algorithm based on signal power theory to do event detection.

2.2 Social event model

In order to find the underline principal and predict the future dynamics of social events, different kinds of event model are proposed either from a topological view or from a statistical volume view.

Recent advances in topology-based event models mainly focus on two concepts: networking and event cascade. In a social network, nodes represent ordinary users and edges denote the relationships among users (e.g., follower and followee). Network or graph structure is an intuitive and straightforward organization for characterizing social influence and/or information diffusion [6, 7]. For instance, if a node have larger out degrees, then it would have higher influence because its outgoing messages can reach more followees [21]. For event analysis, cascade structure is as well an effective model, which concentrates more on information cascade, weakening the concept of networking [44]. In a cascade structure, if a user retweets a post, his/her followers would probably to retweet this post again. While networking emphasizes on the information propagation in the whole social network, mostly decided by the influence (i.e., in and out degrees) of nodes [21], cascade pays more attention to the retweeting mechanism of messages corresponding to time series, helping to predicting future popularity.

Given the collective volume of events, we can model the temporal dynamics of social events. In [18], a meme-tracking model was proposed to generate temporal curves of different memes with only two ingredients. To make a further step, in [2], various mathematical models were evaluated to determine an accurate characterization of the growth and fade of meme dynamics. To build an event volume model, an important phenomenon worth noticing is the long tail effect [1], which is common in real life. For instance, when a trending event emerges, its volume grows quickly at the first few days, then the public calm down and the volume gradually goes down. However, the subsequent discussion may still last for a long period. Moreover, under one topic, the tail of previous events volume can exert impact on the current event. To overcome this problem and extract the genuine volume of current event, it was claimed [29] that event volume is the response of social media to the external stimuli. The event volume was defined as the convolution of the events importance and social response function, and the true event volume can be split from the observed event volume by deconvolution. In this paper, we propose a simple yet effective approach to extract event volume by regarding the tail volume as social noise irrelevant to the current event.

2.3 Social event prediction

After building the event model, one of the popular applications of event model is to make a prediction. Prediction in social event focuses on predicting the user activity level in social networks [45], future information diffusion in social networks [8] or future popularity of tweets [17, 44]. User activity is another type of events corresponding with the connection of social networks. As a result of the sparsity of historical data, the relationship of users is necessary to be considered. Thus the user activity level in social network can be predicted by introducing a personalized and socially regularized time-decay model. And the activity level prediction are formulated as a binary classification problem Another type of prediction aims to predict the temporal dynamics of diffusion in social networks. Information diffusion is a process where a event is spread between different users. And in social network, users connect with each other by relationships. Thus, the prediction for information diffusion relies much on a graph connection between users. With a straightforward assumption that the users' interconnection causes the dynamics of the spreading process of a social events, the propagation of events can be predicted by modeling the social dimension and semantic dimension [8]. Additionally, the prediction of events popularity is also an important part of research work. The popularity of a tweet or a news can be treated as a popularity of a event, though different from our definition of event. Foresee the success of a event can help companies to prepare in advance. Previous work claim that modeling the collective behavior of users of a social media site allows the prediction of popularity of items from the users' early reaction. Thus they build a stochastic model of user behavior to make a prediction of how popular one item (event) will become [17]. After that, another generative model of predicting the final popularity of tweet is proposed. The model is based on the theory of self-exciting point process. And it can be used for estimating the spreading rate of given information (event), determining whether the propagation is growing up or dying outm, and predictin the final popularity of the event [44].

Different from previous work, in this work, we study how to predict the future volume of a new event based on limited information (e.g., 24 hours). Our prediction is not just a computation of final popularity rather we want to predict the numerical value at every time point so that the dynamic trends of events can be viewed clearly. We propose to connect the new events and historical events by exploring both volume feature and content feature, then reconstruct the volume dynamics of new events by event early embedding algorithm.

3 Preliminary and problem definition

In this section, we present some preliminary information of this work, including notations and problem definition.

3.1 Preliminary

Given a set of n tweets corresponding to a certain topic $\mathbf{T}$, denoted as $\mathcal{T} = \{\mathbf{t}_i\}_{i=1}^{n}$. The i-th tweet $\mathbf{t}_i$ is represented as a triplet (tp_i, c_i, ts_i), where tp_i is the topic-word $\mathbf{t}_i$ belonging to, c_i indicates $\mathbf{t}_i$'s content and ts_i represents the timestamp. We further define $V = (V(t))|_{t=1}^{l}$ and $C = (C(t))|_{t=1}^{l}$ as the volume sequence and content sequence, respectively. Here, $V(t)$ is the number of the tweets in the topic $\mathbf{T}$ during the t-th time interval (e.g., 1 hour), $C(t)$ is the corresponding collective contents, and l is the length of $\mathbf{T}$'s life cycle.

Suppose $\mathbf{T}$ comprises of m events, denoted as $\{\mathbf{E}_j\}|_{j=1}^{m}$, where $\mathbf{E}_j = (V_j, C_j, s_j, q_j)$, $V_j = (V_j(t))|_{t=1}^{l}$ is the sequence of tweet volume of $\mathbf{E}_j$ and C_j is the collective content of $\mathbf{E}_j$. Let s_j and q_j be the start time and end time of $\mathbf{E}_j$, respectively. Here we have $1 \leq s_j < q_j \leq l$. At any time before the starting point or after the ending point, the event volume is 0. By defining the volume of social noise in the topic as $\xi = (\xi(t))|_{t=1}^{l}$, we model the topic volume as below:

$$V(t) = \sum_{j=1}^{m} V_j(t) + \xi(t), \ t = 1, 2, \ldots, l. \tag{1}$$

3.2 Problem definition

The tremendous explosion of newly-emerging social events have posed great challenges yet opportunities to social media analysis. It is observable that these social events are likely to follow similar patterns of variation trends, such as detecting the shape clusters of temporal dynamics [34]. In this work, we mainly target at predicting the future volume of a given event at very early stage (e.g., the first day of the new event). The basic idea is that the similar events have the same dynamics both in early stage and in the future dynamics. The task is extremely non-trivial due to the obstacles of data sparsity and social noise. To this end, we first construct a corpus of historical events by extracting event information from topic information:

Problem 1 *Given the information of the topic* $\mathbf{T}$*, denoted as* $\mathbf{P} = (V, C)$*, extract all the events* $\mathbf{E}_j = (V_j, C_j, s_j, q_j)$*, including the volume* V_j*, the content* C_j*, the start time* s_j *and the end time* q_j*.*

Then, from the event information we can find both the early stage information and the whole stage information, such that we can explore the patterns of the variation, which is formulated as follows:

Problem 2 *Without loss of generality, suppose we have* N *social events extracted from some topics, denoted as* $\{\mathbf{E}_j\}|_{j=1}^{N}$*. Given very limited information (e.g.,* Te *hours) of a new event, denoted as* $\mathbf{Ee}^{(q)} = (\mathbf{E}^{(q)}(t))|_{t=1}^{Te}$*, predict* $\mathbf{E}^{(q)}$*'s future volume dynamics* $V^{(q)}(t), \ t > Te$*.*

To notice, the new event $\mathbf{E}^{(q)}$ only has the early information $\mathbf{Ee}^{(q)}$. Thus we need match the early stage information with the early stage information of previous events, which will be introduced in the rest part. And we elaborate the specific method and mathematical deduction in the next section.

4 Event early embedding model

In this section, we elaborate the proposed **E**vent **E**arly **E**mbedding **M**odel (EEEM). Our model has two important parts. The first part is the collection of event corpus so than the new events can be matched with some previous events. The second part is prediction part, where the future volume dynamics of a event can be predicted by applying our event early embedding algorithm.

In order to collect the event corpus, we design an algorithm for denoising the topic volume and extract events. The algorithm is based on the signal power theory and the basic assumption is that the noise is irrelavant to the current event. Then we briefly introduce Locally Linear Embedding (LLE) algorithm, which is the core idea of our predictive algorithm, before we describe the event early embedding algorithm for predicting (reconstructing) the future dynamics of new events.

4.1 Social denoising and event extraction

It is observed that within a topic there may exist certain levels of irrelevant content w.r.t. the events, such as spamming, advertisement, regular search logs. Recall from (1) that the volume of a topic is comprised of the volumes of all its events and the social noise. Besides, the volume of social noise volume for different topics may have different distributions. As illustrated in Figure 2, compared to the fast fluctuation of the volume of social events, the volume of noise only has slight variation over time, which hints us to make the rational assumption that the volume of social noise is an time-invariant constant. Thus, without loss of generality, the topic volume model is simplified as

$$V(t) = \sum_{j=1}^{m} V_j(t) + \xi. \tag{2}$$

Note that we simply use ξ to represent the constant volume of social noise. Performing integral with time-average for an infinite interval, we arrive at

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T V(t)dt = \lim_{T\to\infty} \frac{1}{T}\int_0^T \sum_{j=1}^{m} V_j(t)dt + \xi. \tag{3}$$

From the preliminary definition, each event $\mathbf{E}_j$ only has volume greater than zero between the start time s_j and the end time q_j, which means every event only lasts for a period and then decays to zero. It is observable that the sharp variation in topic volume normally corresponds to the emergence of events, which is similar to the impulse in a signal. Hence, we may conclude that most power of topic volume is from social noise. Another assumption is that under the same topic, different events do not overlap with each other in view of occurring time. In other words, if $j > h$, then $s_j \geq q_h$. Thus, as $T \to \infty$, we have

$$\begin{aligned} & \lim_{T\to\infty} \tfrac{1}{T}\textstyle\int_0^T \sum_{j=1}^{m} V_j(t)dt \\ = & \lim_{T\to\infty} \tfrac{1}{T}\textstyle\sum_{j=1}^{m} \int_{s_j}^{q_j} V_j(t)dt \\ = & \quad 0. \end{aligned} \tag{4}$$

Substituting (4) into (3), we can obtain the volume of social noise by calculating the time-average integral for infinite interval, which means observing for a long enough period:

$$\xi = \lim_{T\to\infty} \frac{1}{T}\int_0^T V(t)dt. \tag{5}$$

Now, we can extract the volume of social events by subtracting the volume of social noise from the volume of topic:

$$\sum_{j=1}^{m} V_j(t) = V(t) - \xi. \tag{6}$$

However, because of the slight fluctuation of noise, we might encounter unreasonable results of event volume, e.g., the error of negative number. Therefore, to prevent such situations, we transform the resultant event volume with a smoothing function $f(V(t) - \xi)$, which is defined as a half sigmoid linear function

$$f(x) = \begin{cases} x + 0.5, & \text{if } x \geq 0; \\ \frac{1}{1+e^{-x}}, & \text{if } x < 0. \end{cases} \tag{7}$$

Now, we can detect events as *the longest sub-sequences of consecutive non-zero members* in the sequence of the denoised topic volume $\{V(t) - \xi\}|_{j=1}^{l}$. The start point and end points of an sub-sequence respectively corresponds to the start time and end time of the found event. Figure 2 presents an illustration of the denoising results on “Apple”, and the red area parts is the extracted events.

4.2 Locally linear embedding algorithm

LLE is a kind of unsupervised learning algorithm, which is mostly used for nonlinear dimensionality reduction [22]. When we want to find the low-dimensional representation of a high-dimensional data, we can map the data into a low-dimensional space without change the underlying manifold. Suppose the time interval of the event volume is same so that each event volume can be written as a data point which consist of N real-valued vectors $\mathbf{V}_i$ with dimensionality D. Each data point is expected to lie on or close to a locally linear patch or the manifold. The reconstruction error can be measured by the cost function

$$J(W) = \sum_i \|\mathbf{V}_i - \sum_j W_{ij}\mathbf{V}_j\|^2 \tag{8}$$

which can be minimized to compute the reconstructed coefficient. For the constraints that first, each data is reconstructed only from its neighbors and second, the linear combination coefficients sum to one, the optimal weights of can be found by solving a least squared problem. After finding the reconstruction weights in high-dimensional space, the data point in the low-dimensional space is also expected to be reconstructed by its neighbors with the same weights so that the local geometry in the original data space can be invariant. Thus, LLE compute the coordinates $\mathbf{Y}_i$ as low-dimensional representation for the data point $\mathbf{V}_i$ by minimizing the embedding cost function

$$\Phi(Y) = \sum_i \|\mathbf{Y}_i - \sum_j W_{ij}\mathbf{Y}_j\|. \tag{9}$$

The embedding cost in (9) defines a quadratic form in the coordinates $\mathbf{Y}_i$, which can be minimized by solving a sparse $N \times N$ eigenvalue problem. The nonzero eigenvector corresponding with the lower eigenvalues provide an ordered set of orthogonal coordinates centered on the origin. Figure 3 is a showcase of mapping the events volume points into a two dimension space by applying the LLE algorithm.

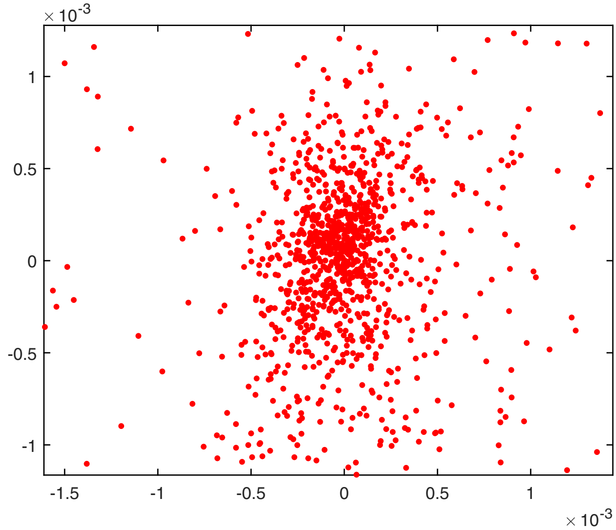


Figure 3 Figure of mapping the events volume into the two dimension space according to LLE algorithm. We randomly choose 1000 events to be projected into this space. And the original dimension is 48 corresponding with time interval 1 hour

4.3 Event dynamics prediction

This part is the core part of EEEM, targeting at solving Problem 2 in Section 2. We try to predict the new event volume future dynamics given only the early stage information. The basic idea is that we want to reconstruct the new events volume dynamics by a linear combination of its neighbors. First, we use the volume and content features of event to find its neighbors, and the fusion of two types of feature is based on the similarity of event. Second, we use the early volume dynamics of events to learn the weights of combination, which is based on LLE algorithm. Finally, we use the full dynamics rather than the early dynamics to reconstruct the future dynamics of the new events.

First we will use the early information to find the neighbors. However, previous events $\{\mathbf{E}_j\}|_{j=1}^{N}$ consist of the whole information of the event from the beginning to the end. Thus, in case of confusion, we separate the event early information denoting as $\{\mathbf{Ee}_j = (V_{ej}, C_{ej}, s_j, s_j + T_e)\}$, in which V_{ej} is the event early volume, C_{ej} is the event early content, and T_e is the early time duration. For example, when the early time duration T_{ej} is 24 hours, the event early volume V_{ej} is the first 24 values of the whole volume V_j, given the time interval is 1 hour, the same for C_{ej}.

Given the new event's early information $\mathbf{Ee}^{(q)}$, we represent the two types of early stage features vectors as $\mathbf{x}_v^{(q)}$ and $\mathbf{x}_c^{(q)}$. We may use early fusion [27] to concatenate the volume feature and the content feature, and then perform $k$nn search to find the neighbors. Nonetheless, the problem of incompatibility of different feature types may lead to unsatisfactory performance. Hence, we propose the following similarity-level [26] merge to facilitate $k$nn search:

$$S^{(e)}(\mathbf{Ee}^{(q)}, \mathbf{Ee}_j) = S^{(v)}(\mathbf{x}_{vj}^{(q)}, \mathbf{x}_v) \cdot S^{(c)}(\mathbf{x}_c^{(q)}, \mathbf{x}_{cj}), \tag{10}$$

where $S^{(e)}(\cdot)$ is the similarity factor of the new event and previous event, which is a product of $S^{(v)}(\cdot)$ and $S^{(c)}(\cdot)$, the volume similarity and content similarity respectively. $S^{(v)}(\mathbf{x}_v^{(q)}, \mathbf{x}_v)$ and $S^{(c)}(\mathbf{x}_c^{(q)}, \mathbf{x}_c)$ are defined as

$$\begin{cases} S^{(v)}(\mathbf{x}_v^{(q)}, \mathbf{x}_v) &= \frac{\|\mathbf{x}_v^{(q)} - \mathbf{x}_v\|}{\mathbf{max}(\|\mathbf{x}_v^{(q)} - \mathbf{x}_v\|)}, \\ S^{(c)}(\mathbf{x}_c^{(q)}, \mathbf{x}_c) &= \frac{\mathbf{x}_c^\top \mathbf{x}_c^{(q)}}{\|\mathbf{x}_c\| \|\mathbf{x}_c^{(q)}\|}, \end{cases} \tag{11}$$

where $\|\cdot\|$ is the ℓ_2 norm and $\mathbf{max}(\cdot)$ find the maximum value. Note that in order to make $S^{(v)}(\mathbf{x}_v^{(q)}, \mathbf{x}_v)$ and $S^{(c)}(\mathbf{x}_c^{(q)}, \mathbf{x}_c)$ comparable, we project volume feature and content feature to the scale of [0,1] by considering a maximum as denominator and cosine similarity respectively.

From the similarity of new event and previous events, we can find k neighbors which are k most similar events to the new event both in volume and in content. To find the reconstruction coefficient vector $\mathbf{w}$, like LLE algorithm in (8) except for the right regularization term, we try to minimize the following early reconstruction error:

$$\varepsilon(\mathbf{w}) = \frac{1}{2T_e} \left[\sum_{t=1}^{T_e} |V_e^{(q)}(t) - \sum_{j=1}^{k} \mathbf{w}_j V_{ej}(t)|^2 + \gamma ||\mathbf{w}||^2 \right] \tag{12}$$

where V_{ej} is the early volume of corresponding neighbor to $\mathbf{w}_j$ and γ is the regularization factor. The weight $\mathbf{w}_j$ summarizes the contribution of the jth event at early stage. To solve this problem, different with LLE algorithm solving a linear system, we implement gradient descent to find the optimizing weights $\mathbf{w}$.

Finally, we need to construct the future volume dynamics of the new event so that we can make a prediction. In particular, given a new event at its early stage, the predictive volume dynamics $V^{(q)}(t)$ of the new event is

$$V^{(q)}(t) = \sum_{j=1}^{k} \mathbf{w}_j V_j(t), \tag{13}$$

where $V^{(q)}$ is the predictive volume of the new event $\mathbf{E}^{(q)}$, $V_j(t)$ is volume of the neighbors corresponding with the weight $\mathbf{w}_i$. The k nearest neighbors are also the same that we utilize for computing the early reconstruction error in (12). The underlying principle of the model (13) is that the early volume dynamics of a event is an early embedding of the future dynamics thus we learn the weights from the early stage and the future dynamics of new event can be predicted by reconstruct the dynamics from the neighbors.

5 Experimental evaluation

In this section, we evaluate the proposed EEEM for predicting the future volumes of social events at very early stage on a large Twitter dataset.

5.1 Data

For evaluation, we employed the Twitter dataset published by [4]. The dataset contains 10,681,232 tweets posted from 2013-08-01 to 2013-11-30. We regarded the trending hashtags (e.g., #iPad) as topics, which results in 18,399 topics. We notice that most of the topics contain less than 50 tweets, which are probably outliers or noises. Hence, we sorted all the topics in descending order of the topic volume and kept the top 5,000 topics as our experimental data.

For each topic, we organized the corresponding tweets into a sequence of sub-groups of tweets at hourly intervals. Each element (sub-group) in the sequence contains an hour of tweets belonging to this topic. We utilized Stanford CoreNLP toolbox [20] to pre-process the tweet data, such as extracting basic tokens and generating lemmas from all the tokens. The words in Twitter is so casual that there exist 555081 different words. However, most of the words appear only once or twice, thus for easy computation and elimination of the noise-words, we neglect the word which appears less than 5 times among the whole corpus. Hence, the corpus only has 91095 different words. The words talked during one hour in a topic is the hourly topic content. And the number of tweets related to a topic in an hour is the hourly topic volume value. Then we will implement the social denoising method to get the events from topics.

5.2 Social noise reduction and event extraction

Recall that inspired by the knowledge in signal analysis, we develop a social denoising approach based on two rational assumptions of constant volume of social noise and infinite time duration interval of observation. We understand that it is impossible to satisfy the constraint $T \rightarrow \infty$, which requires an infinite time of observation. Nevertheless, as our experiments show, on average, each topic only comprises about 3 events and the average life-cycle of each event is 4 days. Every event needs an infinite time of observation. But on average, every event only lasts for 4 days. Therefore, we may reasonably argue that 122 days are sufficiently long for underpinning the assumption of infinite time interval observation and implementing our proposed denoising approach.

In order to gain complete information about event data for achieving accurate prediction of event volume, we also omit those events either happening before the very beginning of our data (incomplete early feature) or finishing after the very end of the data (incomplete information of event volume). Similar to the illustration in Figure 2, where the last possible event at the tail of the time series is actually incomplete, thus simply ignored. And the event volume $V_j(t)$ is the red area under the blue line of the topic volume in this figure. We can also locate the start time points s_j and the end points q_j of the event $\mathbf{E}_j$ after denoising. Moreover, we can also get the event content $C_j(t)$ from the topic content by extracting the words from the start time to the end time in topic content.

5.3 Event volume dynamics prediction

For the experiment, we selected the events lasting more than 48 hours so that the event volume dynamics is long enough, which gives us 16707 events in total. To make it more related to real-world scenario, our training events should be some historical and already happened events. And the test events should be some un-happened events. We sorted these events in ascending order of their start time, and chose the top 16507 events to form the historical event corpus and the rest latest 200 events as new events samples.

5.3.1 Evaluation metric

To find the best prediction samples and parameters, we need to use a common evaluation metric during the whole experiment. However, according to the error value calculated by traditional evaluation method, such as the mean squared error (MSE). We found that good prediction usually have a high error value, which can lead to bad evaluation. Thus, we first develop a criteria for evaluating the accuracy of volume prediction. Here we define a *Divergence* $\mathcal{D}(V^*, V^g)$ to characterize the difference between our predicted sequence V^* and the true volume sequence V^g of the given event:

$$\mathcal{D}(V^*, V^g) = \frac{\mathrm{Dist}(V^*, V^g)}{\mathrm{Sim}(V^*, V^g)}, \tag{14}$$

where $\mathrm{Dist}(\cdot,\cdot)$ and $\mathrm{Sim}(\cdot,\cdot)$ are defined as

$$\begin{cases} \mathrm{Dist}(V^*, V^g) &= \sum_{t=1}^{l} \frac{\left(V^*(t) - V^g(t)\right)^2}{\sqrt{V^g(t)+1}}, \\ \mathrm{Sim}(V^*, V^g) &= \frac{\mathbf{V}^* \cdot \mathbf{V}^g}{\|\mathbf{V}^*\| \|\mathbf{V}^g\|}, \end{cases} \tag{15}$$

where $\mathbf{V}^*$ and $\mathbf{V}^g$ are the vector versions[3] of V^* and V^g, respectively.

While $\mathrm{Dist}(\cdot,\cdot)$ measures the absolute difference between the two volume dynamics, $\mathrm{Sim}(\cdot,\cdot)$ is the cosine similarity, which guarantees that even if the absolute distance is somehow far, but we still regard it as a better prediction since the prediction has a similar variant shape of dynamics to the real dynamics. We illustrate the difference of using divergence, distance and Mean Squared Error (MSE) as evaluation metrics in Figure 4. The superiority of Divergence compared to the traditional MSE (Mean Squared Error) and distance measurement is in that it not only measures the absolute difference but also the distributional divergence. As we can see from Figure 4, we have the intuitive conclusion that (a) and (c) should be the better cases of prediction than (c) and (d), respectively, which is also consistent with the indication of divergence $\mathcal{D}$. However, if we simply use MSE or distance as measurement, we have the inverse results, i.e., (b) (Dist=98 and MSE=153) is better than (a) (Dist=120 and MSE=703) and (c) (MSE=110) fits worse than (d) does (MSE=69). Therefore, in the subsequent experiments, we always use divergence as the evaluation metric.

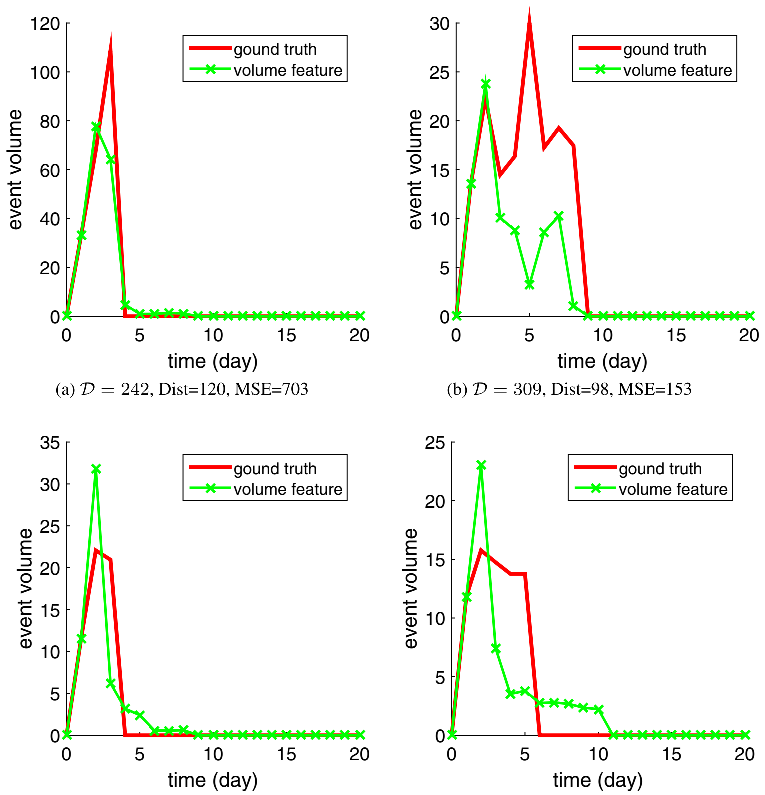
5.3.2 Optimization of parameter

In this part, we optimizing our parameter based on the evaluation result. Given a test event, we use (13) to predict the volume and then exploit (14) to compute the divergence value. We utilize the sum of the logarithm of the divergence values of all the testing samples as evaluation measurement, denoted as $\mathbf{M}$, as shown in (16):

$$\mathbf{M} = \sum log(\mathcal{D}). \tag{16}$$

The reason why we take a logarithm of the divergence is that a few very bad prediction samples, which have very high divergence value, may sharply affect the evaluation of the model if we sum the divergence directly. Our goal in this part is to find the best parameters that can make the measurement $\mathbf{M}$ low and meanwhile comparing the fusion results with methods that only considering one of the volume feature and content feature. The parameters in our model are the number of neighbors k and the regularization factor γ. However, the regularization term is just a penalty term for too big values of weights in case of overfitting, so

[3] In practice, we find that $\mathbf{V}^*$ and $\mathbf{V}^g$ may have different lengthes. To make them comparable, we simply expand the shorter one with value 0 to meet the length of the longer one.



(a) $\mathcal{D} = 242$, Dist=120, MSE=703

(b) $\mathcal{D} = 309$, Dist=98, MSE=153

(c) $\mathcal{D} = 97$, Dist=45, MSE=110

(d) $\mathcal{D} = 139$, Dist=33, MSE=69

Figure 4 Illustrative results of the predicted event volume v.s. the ground-truth volume. We choose 4 representative cases to demonstrate the superiority of using divergence as evaluation metric compared to Mean Squared Error (MSE) and distance

it has little effect on the whole model. Thus we set it fixed with value 0.1. Hence, the most important work is to find the hyperparameter k. In other words, we determine how many nearest neighbors (i.e., k) should be chosen to reconstruct a new event for volume prediction, and in this step, we set the $\gamma = 0.1$. Additionally, in this part we set the early time duration Te as 24 hours. We tune k in the range of $\{1, 2, \ldots, 50\}$. For each k, we learn an optimized $\mathbf{w}$ by implementing gradient descent for solving the optimization equation (12). The experimental results are shown in Figure 5.

As we can see, firstly, the fusion feature of volume feature and content feature is a better measurement than considering volume feature or content feature individually, as the measurement value $\mathbf{M}$ of fusion feature is always lower than the other two along with the increase of k from 1 to 50. Secondly, the measurement $\mathbf{M}$ keeps descending for all of these three types of features. The measurement of content feature and volume feature is poor at first and bump down quickly while the fusion feature reaching the optimal quickly and change slightly. When given only few neighbors, the information is too sparse for only using the content feature or volume feature, while the fusion feature learns a better information from two types of features thus has a better measurement. At the other hand, given too many neighbors, all of these three types of features improve little on the measurement thus we

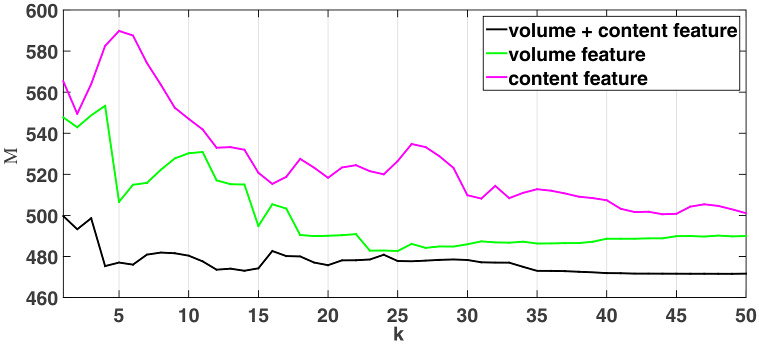


Figure 5 Measurement (sum of logarithm of divergence) of k w.r.t. three types of feature. The bottome line is the fusion feature measurement, which is the best measurement than volume feature and content feature

recommend use 35 neighbors for the optimizing number of neighbors. Hence, in the rest of the experiment, the number of neighbors is set as 35 for comparison.

5.3.3 Early duration comparison

As illustrated in (12), the reconstruction process can be implemented during any early duration Te. However, different feature types has different performance during different early duration. As the shortest events only have 48 hours lifecycle, we choose three stages of time early duration, which are 12 hours, 24 hours and 36 hours, representing half-day, one-day and one-and-a-half day early duration prediction. Figure 6 shows the result of different measurement for different types features when given different early duration Te.

From the result in Figure 6, we find that no matter what early duration, the fusion feature is always the better measurement. And when $Te = 12$, the worst measurement is the volume feature, and the fusion feature and content feature have the equal measurement. When $Te = 24$, the volume feature measurement drops more than the content feature and thus is a better measurement than content feature. Things worth noticing is that when the early duration is 36 hours, the measurement drops sharply to a very low value for all of three types of features. Here is a brief analysis for this result. During 12 hours or event 24 hours, most

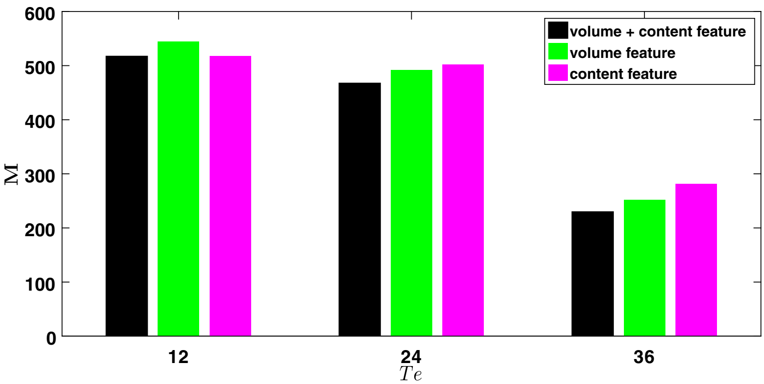


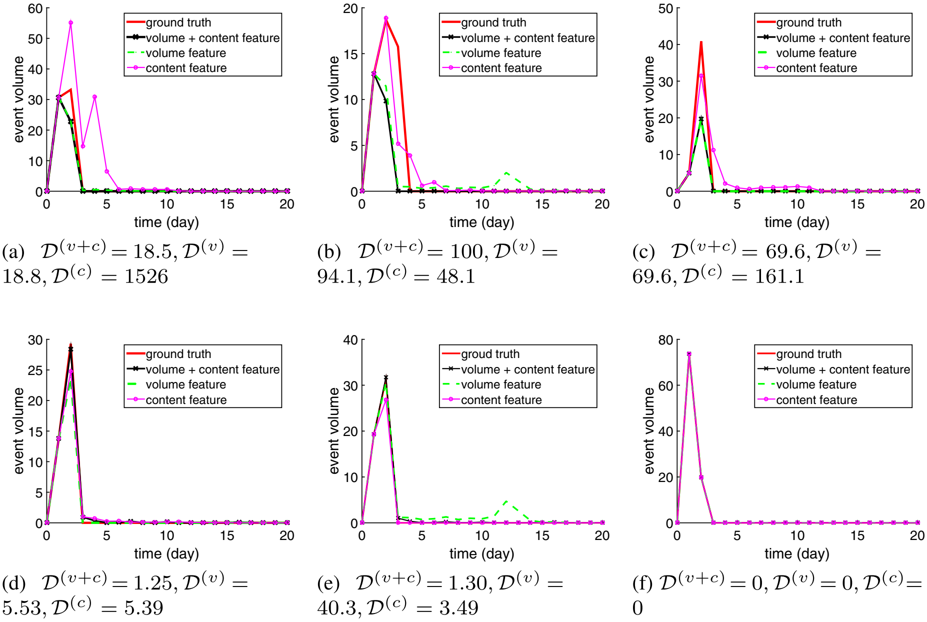
Figure 6 Measurement of different features w.r.t. different early duration Te, which is 12, 24 and 36

of the events have the same dynamics, each events grows up and the value changes continuously. However, after 24 hours, there exists a differentiation time point when some events climb up slowly, some events bump up quickly, whereas others decay gradually. Thus the early information is enough for distinguishing the correct corresponding neighbors of a new event.

5.3.4 *Individual prediction study*

In this part, we study some individual cases from which we can have an intuition of our prediction result. From good prediction measurement, the number of neighbors is set to 35 as before and the early duration Te is set to 36. In Figure 7, We show several illustrative predictions of using the fusion feature, volume feature, content feature. The divergence of each individual prediction is denoted as $\mathcal{D}^{(v+c)}$, $\mathcal{D}^{(v)}$ and $\mathcal{D}^{(c)}$, respectively. The values of each divergence are listed in the captions.

There are some interesting observations. First, prediction result with the content feature is an unstable results, which means the result can be very embarrassing. For example, the sub-figure (a) in Figure 7 is a poor prediction example of content feature while both the fusion feature and volume feature predict well. On the other hand, the sub-figure (b) is a good prediction by content feature, since the $\mathcal{D}^{(v+c)}$ and $\mathcal{D}^{(v)}$ is around 100 whereas $\mathcal{D}^{(c)} = 48.1$. It seems that the prediction by content feature are not that strict to the volume dynamics since content feature focus so much on the words similarity that neglecting the dynamics of events volume. Moreover, from the sub-figures (d) and (e) we find that the fusion feature can improve the prediction results under most circumstance. Finally, an amazing result is the sub-figure (f), where all of three types of feature get zero divergence. The reason for zeros divergence may because this type of dynamics of event volume is so common that all of these features can find the best neighbors and thus make a perfect prediction.



(a) $\mathcal{D}^{(v+c)} = 18.5, \mathcal{D}^{(v)} = 18.8, \mathcal{D}^{(c)} = 1526$

(b) $\mathcal{D}^{(v+c)} = 100, \mathcal{D}^{(v)} = 94.1, \mathcal{D}^{(c)} = 48.1$

(c) $\mathcal{D}^{(v+c)} = 69.6, \mathcal{D}^{(v)} = 69.6, \mathcal{D}^{(c)} = 161.1$

(d) $\mathcal{D}^{(v+c)} = 1.25, \mathcal{D}^{(v)} = 5.53, \mathcal{D}^{(c)} = 5.39$

(e) $\mathcal{D}^{(v+c)} = 1.30, \mathcal{D}^{(v)} = 40.3, \mathcal{D}^{(c)} = 3.49$

(f) $\mathcal{D}^{(v+c)} = 0, \mathcal{D}^{(v)} = 0, \mathcal{D}^{(c)} = 0$

Figure 7 Fusion of feature v.s. Single feature for predicting event volume

6 Conclusion

In this work, we studied the problem of predicting event volume with limited early information. In the context of social media, we formally defined the concepts of topic, event, volume and social noise. We devised a simple yet effective approach to eliminate irrelevant volume from the original topic volume and extract event volume. Furthermore, we view the future dynamics as a high dimensional embedding generating from the early low dimensional event dynamics. However, to overcome the sparsity in the low dynamics space, we proposed a novel prediction model, termed event early embedding model (EEEM), to reconstruct a new event from its k neighbors based on both volume and content features. Additionally, in order to evaluate and optimize the model accurately, we provide a novel evaluation metric that can not only calculate the volume distance but also the dynamics shape between prediction and groud truth. Extensive experiments on a large-scale Twitter dataset demonstrated the effectiveness in removing social noise, the necessity of exploiting content information of events, and the superiority of the proposed model in predicting future volume dynamics of events. And from the experiment, we find that the different early duration have different impact in making a prediction and th individual cases study gives us an intuition of our EEEM model.

In future, we may model event by fully exploring content information using sophisticated NLP techniques, such as word embedding and deep learning. Besides, we could also design a generative model which not merely searches the best matching events but also predicts the event dynamics incrementally. We could use different algorithm in different early stage so that the early prediction is more accurate. Finally, we may even exploring the pictures information or even the video [5] which may provide us more information in early stage.

Acknowledgements This work was supported in part by the National Natural Science Foundation of China under Project 61572108, Project 61632007 and Project 61502081.

References

1. Barabasi, A.L.: The origin of bursts and heavy tails in human dynamics. Nature **435**(7039), 207–211 (2005)
2. Bauckhage, C., Kersting, K., Hadiji, F.: Mathematical models of fads explain the temporal dynamics of internet memes. In: Proceedings of the Seventh International Conference on Weblogs and Social Media, ICWSM 2013, Cambridge, Massachusetts, USA, July 8-11, 2013. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/view/6022 (2013)
3. Becker, H., Naaman, M., Gravano, L.: Learning similarity metrics for event identification in social media. In: Proceedings of the Third International Conference on Web Search and Web Data Mining, WSDM 2010, New York, NY, USA, February 4-6, 2010, pp. 291–300. https://doi.org/10.1145/1718487.1718524. http://doi.acm.org/10.1145/1718487.1718524 (2010)
4. Cai, H., Tang, Z., Yang, Y., Huang, Z.: Eventeye: Monitoring evolving events from tweet streams. In: Proceedings of the ACM International Conference on Multimedia, MM '14, Orlando, FL, USA, November 03 - 07, 2014, pp. 747–748. https://doi.org/10.1145/2647868.2654871. http://doi.acm.org/10.1145/2647868.2654871 (2014)
5. Gao, L., Guo, Z., Zhang, H., Xu, X., Shen, H.T.: Video captioning with attention-based LSTM and semantic consistency. IEEE Trans. Multimed. **19**(9), 2045–2055 (2017)
6. Gomez-Rodriguez, M., Leskovec, J., Krause, A.: Inferring networks of diffusion and influence. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, July 25-28, 2010, pp. 1019–1028. https://doi.org/10.1145/1835804.1835933. http://doi.acm.org/10.1145/1835804.1835933 (2010)
7. Gomez-Rodriguez, M., Balduzzi, D., Schölkopf, B.: Uncovering the temporal dynamics of diffusion networks. In Proceedings of the 28th International Conference on Machine Learning, ICML 2011, Bellevue, Washington, USA, June 28 - July 2, 2011, pp. 561–568 (2011)

8. Guille, A., Hacid, H.: A predictive model for the temporal dynamics of information diffusion in online social networks. In: Proceedings of the 21st World Wide Web Conference, WWW 2012, Lyon, France, April 16-20, 2012 (Companion Volume), pp. 1145–1152. https://doi.org/10.1145/2187980.2188254. http://doi.acm.org/10.1145/2187980.2188254 (2012)
9. Guo, L., Zhang, D., Li, G., Tan, K., Bao, Z.: Location-aware pub/sub system: When continuous moving queries meet dynamic event streams. In: SIGMOD, pp. 843–857 (2015)
10. Gupta, M., Gao, J., Zhai, C., Han, J.: Predicting future popularity trend of events in microblogging platforms. Proc. Amer. Soc. Inf. Sci. Technol. **49**(1), 1–10 (2012)
11. Hong, R., Li, L., Cai, J., Tao, D., Wang, M., Tian, Q.: Coherent semantic-visual indexing for large-scale image retrieval in the cloud. IEEE Trans. Image Process. **26**(9), 4128–4138 (2017)
12. Hu, M., Yang, Y., Shen, F., Zhang, L., Shen, H.T., Xuelong, L.: Robust web image annotation via exploring multi-facet and structural knowledge. IEEE Trans. Image Process. **26**(10), 4871–4884 (2017)
13. Hu, M., Yang, Y., Shen, F., Xie, N., Shen, H.T.: Hashing with angular reconstructive embeddings. IEEE Trans. Image Process. **27**(2), 545–555 (2018)
14. Kamath, K.Y., Caverlee, J.: Discovering trending phrases on information streams. In: Proceedings of the 20th ACM Conference on Information and Knowledge Management, CIKM 2011, Glasgow, United Kingdom, October 24-28, 2011, pp. 2245–2248. https://doi.org/10.1145/2063576.2063937. http://doi.acm.org/10.1145/2063576.2063937 (2011)
15. Kanhabua, N., Nejdl, W.: Understanding the diversity of tweets in the time of outbreaks. In: 22nd International World Wide Web Conference, WWW '13, Rio de Janeiro, Brazil, May 13-17, 2013, Companion Volume, pp. 1335–1342. http://dl.acm.org/citation.cfm?id=2488172 (2013)
16. Lehmann, J., Gonçalves, B., Ramasco, J.J., Cattuto, C.: Dynamical classes of collective attention in twitter. In: Proceedings of the 21st World Wide Web Conference 2012, WWW 2012, Lyon, France, April 16-20, 2012, pp. 251–260. https://doi.org/10.1145/2187836.2187871. http://doi.acm.org/10.1145/2187836.2187871 (2012)
17. Lerman, K., Hogg, T.: Using a model of social dynamics to predict popularity of news. In: Proceedings of the 19th International Conference on World Wide Web, WWW 2010, Raleigh, North Carolina, USA, April 26-30, 2010, pp. 621–630. https://doi.org/10.1145/1772690.1772754. http://doi.acm.org/10.1145/1772690.1772754 (2010)
18. Leskovec, J., Backstrom, L., Kleinberg, J.M.: Meme-tracking and the dynamics of the news cycle. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, June 28 - July 1, 2009, pp. 497–506. https://doi.org/10.1145/1557019.1557077. http://doi.acm.org/10.1145/1557019.1557077 (2009)
19. Luo, Y., Yang, Y., Shen, F., Huang, Z., Zhou, P., Shen, H.T.: Robust discrete code modeling for supervised hashing. Pattern Recogn. **75**, 128–135 (2018)
20. Manning, C.D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S.J., McClosky, D.: The Stanford CoreNLP natural language processing toolkit. In: Association for Computational Linguistics (ACL) System Demonstrations, pp. 55–60. http://www.aclweb.org/anthology/P/P14/P14-5010 (2014)
21. Rong, Y., Cheng, H., Mo, Z.: Why it happened: Identifying and modeling the reasons of the happening of social events. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, NSW, Australia, August 10-13, 2015, pp. 1015–1024. https://doi.org/10.1145/2783258.2783305. http://doi.acm.org/10.1145/2783258.2783305 (2015)
22. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. Science **290**(5500), 2323–2326 (2000). https://doi.org/10.1126/science.290.5500.2323. http://science.sciencemag.org/content/290/5500/2323
23. Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake shakes twitter users: real-time event detection by social sensors. In: Proceedings of the 19th International Conference on World Wide Web, WWW 2010, Raleigh, North Carolina, USA, April 26-30, 2010, pp. 851–860. https://doi.org/10.1145/1772690.1772777. http://doi.acm.org/10.1145/1772690.1772777 (2010)
24. Shen, H., Wang, D., Song, C., Barabási, A.: Modeling and predicting popularity dynamics via reinforced poisson processes. In: Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27 -31, 2014, Québec City, Québec, Canada, pp. 291–297. http://www.aaai.org/ocs/index.php/AAAI/AAAI14/paper/view/8370 (2014)
25. Shen, F., Yang, Y., Liu, L., Liu, W., Dacheng Tao, H.T.S.: Asymmetric binary coding for image search. IEEE Transactions on Multimedia. https://doi.org/10.1109/TMM.2017.2699863 (2017)
26. Shen, F., Xu, Y., Liu, L., Yang, Y., Huang, Z., Shen, H.T.: Unsupervised deep hashing with similarity-adaptive and discrete optimization. IEEE Transactions on Pattern Analysis and Machine Intelligence. https://doi.org/10.1109/TPAMI.2018.2789887 (2018)
27. Song, X., Nie, L., Zhang, L., Akbari, M., Chua, T.: Multiple social network learning and its application in volunteerism tendency prediction. In: Proceedings of the 38th International ACM SIGIR Conference on

Research and Development in Information Retrieval, Santiago, Chile, August 9-13, 2015, pp. 213–222. https://doi.org/10.1145/2766462.2767726 (2015)

28. Song, J., Gao, L., Nie, F., Shen, H.T., Yan, Y., Sebe, N.: Optimized graph learning using partial tags and multiple features for image and video annotation. IEEE Trans. Image Process. **25**(11), 4999–5011 (2016)
29. Tsytsarau, M., Palpanas, T., Castellanos, M.: Dynamics of news events and social media reaction. In: The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14, New York, NY, USA - August 24 - 27, 2014, pp. 901–910. https://doi.org/10.1145/2623330.2623670 (2014)
30. Wang, B., Yang, Y., Xu, X., Hanjalic, A., Shen, H.T.: Adversarial cross-modal retrieval. In: ACM Multimedia, pp. 154–162 (2017)
31. Weng, J., Lee, B.: Event detection in twitter. In: Proceedings of the Fifth International Conference on Weblogs and Social Media, Barcelona, Catalonia, Spain, July 17-21, 2011. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2767 (2011)
32. Xing, C., Wang, Y., Liu, J., Huang, Y., Ma, W.: Hashtag-based sub-event discovery using mutually generative LDA in twitter. In: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA, pp. 2666–2672. http://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/12012 (2016)
33. Yang, Y., Pierce, T., Carbonell, J.G.: A study of retrospective and on-line event detection. In: SIGIR '98: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, August 24-28 1998, Melbourne, Australia, pp. 28–36. https://doi.org/10.1145/290941.290953. http://doi.acm.org/10.1145/290941.290953 (1998)
34. Yang, J., Leskovec, J.: Patterns of temporal variation in online media. In: Proceedings of the Forth International Conference on Web Search and Web Data Mining, WSDM 2011, Hong Kong, China, February 9-12, 2011, pp. 177–186. https://doi.org/10.1145/1935826.1935863. http://doi.acm.org/10.1145/1935826.1935863 (2011)
35. Yang, Y., Ma, Z., Yang, Y., Nie, F., Shen, H.T.: Multitask spectral clustering by exploring intertask correlation. IEEE Trans. Cybern. **45**(5), 1083–1094 (2015)
36. Yang, Y., Shen, F., Shen, H.T., Li, H., Li, X.: Robust discrete spectral hashing for large-scale image semantic indexing. IEEE Trans. Big Data **1**(4), 162–171 (2015)
37. Yang, Y., Shen, F., Huang, Z., Shen, H.T., Li, X.: Discrete nonnegative spectral clustering. IEEE Trans. Knowl. Data Eng. **29**(9), 1834–1845 (2017)
38. Yin, H., Cui, B., Chen, L., Hu, Z., Huang, Z.: A temporal context-aware model for user behavior modeling in social media systems. In: International Conference on Management of Data, SIGMOD 2014, Snowbird, UT, USA, June 22-27, 2014, pp. 1543–1554. https://doi.org/10.1145/2588555.2593685. http://doi.acm.org/10.1145/2588555.2593685 (2014)
39. Yin, H., Cui, B., Chen, L., Hu, Z., Zhou, X.: Dynamic user modeling in social media systems. ACM Trans. Inf. Syst. **33**(3), 10:1–10:44 (2015). https://doi.org/10.1145/2699670. http://doi.acm.org/10.1145/2699670
40. Yin, H., Cui, B., Lu, H., Huang, Y., Yao, J.: A unified model for stable and temporal topic detection from social media data. In: 29th IEEE International Conference on Data Engineering, ICDE 2013, Brisbane, Australia, April 8-12, 2013, pp. 661–672. https://doi.org/10.1109/ICDE.2013.6544864. https://doi.org/10.1109/ICDE.2013.6544864 (2013)
41. Zaharieva, M., Zeppelzauer, M., del Fabro, M., Schopfhauser, D.: Social event mining in large photo collections. In: Proceedings of the 5th ACM on International Conference on Multimedia Retrieval, Shanghai, China, June 23-26, 2015, pp. 11–18. https://doi.org/10.1145/2671188.2749313. http://doi.acm.org/10.1145/2671188.2749313 (2015)
42. Zhang, D., Li, Y., Fan, J., Gao, L., Shen, F., Shen, H.T.: Processing long queries against short text: Top-k advertisement matching in news stream applications. ACM TOIS **35**(3), 28:1–28:27 (2017)
43. Zhao, Q., Mitra, P., Chen, B.: Temporal and information flow based event detection from social text streams. In: Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence, July 22-26, 2007, Vancouver, British Columbia, Canada, pp. 1501–1506. http://www.aaai.org/Library/AAAI/2007/aaai07-238.php (2007)
44. Zhao, Q., Erdogdu, M.A., He, H.Y., Rajaraman, A., Leskovec, J.: SEISMIC: A self-exciting point process model for predicting tweet popularity. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, NSW, Australia, August 10-13, 2015, pp. 1513–1522. https://doi.org/10.1145/2783258.2783401. http://doi.acm.org/10.1145/2783258.2783401 (2015)

45. Zhu, Y., Zhong, E., Pan, S.J., Wang, X., Zhou, M., Yang, Q.: Predicting user activity level in social networks. In: 22nd ACM International Conference on Information and Knowledge Management, CIKM'13, San Francisco, CA, USA, October 27 - November 1, 2013, pp. 159–168. https://doi.org/10.1145/2505515.2505518. http://doi.acm.org/10.1145/2505515.2505518 (2013)
46. Zhu, X., Li, X., Zhang, S.: Block-row sparse multiview multilabel learning for image classification. IEEE Trans. Cybern. **46**(2), 450–461 (2016)
47. Zhu, X., Li, X., Zhang, S., Ju, C., Wu, X.: Robust joint graph sparse coding for unsupervised spectral feature selection. IEEE Trans. Neural Netw. Learn. Syst. **28**(6), 1263–1275 (2017)