

Burst Time Prediction in Cascades

Senzhang Wang^{*}, Zhao Yan^{*}, Xia Hu[†], Philip S. Yu[‡], Zhoujun Li^{*}

^{*} State Key Laboratory of Software Development Environment,
Beihang University, Beijing, 100191, China

[†] Arizona State University, Tempe, AZ 85287, USA

[‡] University of Illinois at Chicago, Chicago, IL 60607, USA
{szwang, yanzhao, lizj}@uic.edu, psyu@uic.edu, xia.hu@asu.edu

Abstract

Studying the bursty nature of cascades in social media is practically important in many applications such as product sales prediction, disaster relief, and stock market prediction. Although the cascade volume prediction has been extensively studied, how to predict when a burst will come remains an open problem. It is challenging to predict the time of the burst due to the “quick rise and fall” pattern and the diverse time spans of the cascades. To this end, this paper proposes a classification based approach for burst time prediction by utilizing and modeling rich knowledge in information diffusion. Particularly, we first propose a time window based approach to predict in which time window the burst will appear. This paves the way to transform the time prediction task to a classification problem. To address the challenge that the original time series data of the cascade popularity only are not sufficient for predicting cascades with diverse magnitudes and time spans, we explore rich information diffusion related knowledge and model them in a scale-independent manner. Extensive experiments on a Sina Weibo reposting dataset demonstrate the superior performance of the proposed approach in accurately predicting the burst time of posts.

Introduction

Burst, defined as “a brief period of intensive activity followed by long period of nothingness” (Barabási 2011), is a common phenomenon in human activities. The bursty nature of human behavior is observed and studied extensively in many domains, such as electronic communication (Barabási 2005), library visiting (Vazquez et al. 2006), stock trading (Zhu and Shasha 2003), as well as cascades spreading in social media (Matsubara et al. 2012).

With the bursty nature of the cascades and the challenge of information overload in social media, an interesting problem arises: *Can we predict the burst time during cascade spreading?* Predicting the burst time of cascades is of outstanding interest for many applications in various domains, such as product sales prediction (Gruhl et al. 2005; Chen et al. 2013), disaster relief (Sakaki, Okazaki, and Matsuo 2010), and stock market prediction (Pinsén 2012). Ya-

hoo! Finance reported that, Didier Sornette, a former physicist has developed a statistical model with the help of social media data to predict when a financial bubble will burst (Pinsén 2012). As said in the report: “the Sornette model is now predicting a stock market crash as early as next year.”

Existing related works on cascades prediction mainly focus on predicting their future volume, ranging from the future popularity prediction (Szabo and Huberman 2008; Hong, Dan, and Davison 2011) to the aggregate size prediction (Petrovic, Osborne, and Lavrenko 2011; Cheng et al. 2014). Recently, some efforts have also been devoted to modeling the burst patterns and using the burst patterns to cluster the cascades (Crane and Sornette 2008; Yang and Leskovec 2011; Matsubara et al. 2012). Another related area is burst detection (Kleinberg 2002; Zhu and Shasha 2003; Parikh and Sundaresan 2008). Burst detection focuses on detecting instead of predicting the burst. How to predict the burst time of cascades still remains an open problem.

Predicting the burst time of the cascades is a non-trivial task due to the following two major challenges. First, existing prediction approaches cannot be directly applied to predict burst time due to the “quick rise-and-fall” pattern of the bursts. For example, traditionally, regression is widely used for predicting and forecasting (Mill 1990; Li et al. 2011) by learning relationships among features based on historical data. With the “quick rise-and-fall” property (Matsubara et al. 2012) of bursts in cascades, the correlation between historical data and future data becomes difficult to be captured by regression-based methods. This motivates us to investigate: are there any other knowledge could be helpful and how to model them? The second challenge is the significantly distinct life spans of different cascades. The vast difference in magnitude and time span of the cascades makes extracting comparable features difficult, and results in building predictive models challenging. For example, given two cascades with the time spans of 10 days and 10 hours respectively, it is more meaningful to predict the formal one on the daily basis and the latter one on the hourly basis. Therefore, instead of simply modeling all the cascades in the original time scale without distinguishing their diverse time spans, a more general and time-independent model is necessary.

In this paper, we propose to formulate the burst time prediction task as a classification problem by time window based transformation. Instead of predicting the exact occur-

ring time of the burst, we predict in which time window the burst will appear. Since we conduct the prediction in the time window granularity, cascades with diverse time spans can be handled in a unified way. Motivated by previous studies on utilizing social theories to analyze and predict information diffusion in social media (Oh, Susarla, and Tan 2008; Hu et al. 2013; Wang et al. 2014a; Hu et al. 2009), we explore rich social knowledge available during cascades spreading such as knowledge on user profile and social relation to help this task. To utilize rich knowledge in a unified way and eliminate the difference of cascades in magnitude and time span, we model them in a scale-independent manner by deriving scale-independent features. We evaluate the proposed approach on a Sina Weibo reposting dataset that contains 300,000 posts and each post is reposted for 80 times on average. The results show the effectiveness of the proposed approach in accurately predicting the burst time of the posts.

Problem Statement

We will start with some definitions, and then will give two problems need to solve for the burst time prediction task. There is no clear definition of the “burst time” of a cascade, because it is hard to exactly say at what time a burst begins or in which time interval a burst exists. Alternatively, we consider the time of the *global spike* of the cascade defined as follows as its burst time we need to predict.

Definition 1 Global Spike. Suppose 1) the time span T^c of the cascade c can be equally divided into K time windows, that is $T^c = \{(n_1^c, 1), (n_2^c, 2), \dots, (n_K^c, K)\}$, where $n_j^c (j = 1, 2, \dots, K)$ is the number of reposts in the j th time window, and 2) the number of reposts n_k^c is a function of the time window k : $n_k^c = f_c(k)$. The global spike of c is the $f_c(k_{max})$ that satisfies $\forall 1 \leq k \leq K, f_c(k_{max}) \geq f_c(k)$, and k_{max} is the time window of the global spike.

The time series of the cascades may also have some local peaks where the values are larger than their neighbors. Such values can be considered as *local spikes* of the cascade.

Definition 2 Local Spike. Given the time window related function $n_k^c = f_c(k)$ of cascade c and the divided time windows $T^c = \{(n_1^c, 1), (n_2^c, 2), \dots, (n_K^c, K)\}$ in Definition 1, $f_c(k_{lmax})$ is a local spike of cascade c if the following condition is satisfied: $\forall -s \leq i \leq +s, f_c(k_{lmax}) \geq f_c(k_{lmax} + i)$, where s is a predefined threshold.

Previous studies showed that most cascades usually have one notable global spike with several less remarkable local spikes (Crane and Sornette 2008; Yang and Leskovec 2011; Matsubara et al. 2012). Therefore, it makes sense to use the time of global spike as the burst time of the cascade c .

To address the challenge that the time spans of cascades may differ significantly, we propose a time window based approach to eliminate the difference of time spans for various cascades. Specifically, we first divide the time spans of all the cascades into K time windows, and then try to predict the bursts of the cascades appearing in which future time window. Before stating the problem, we first define the μ_{th} future time window of a cascade as follow.

Definition 3 The μ_{th} future time window. Given constant K , μ and the cascade c with an observed spreading time interval $[t_0^c, t_{current}^c]$, where t_0^c is the starting time of the cascade c and $t_{current}^c$ is the current time, the μ_{th} future time window of c is defined as such a time interval $[t_{current}^c + \frac{\mu-1}{K} \times (t_{current}^c - t_0^c), t_{current}^c + \frac{\mu}{K} \times (t_{current}^c - t_0^c)]$.

Based on the above definitions, we next introduce how to address the burst time prediction task by answering the following two questions.

- I. Given a new cascade c with an early stage of observed diffusion process, how could we predict whether a burst will occur in its μ_{th} future time window?
- II. How could we further predict in which future time window the burst will appear?

Question I can be considered as a binary classification problem and solved by a general classification method, such as SVM or decision tree. If we can accurately answer Question I, Question II can be solved based on the solution of Question I. A straightforward approach is to consider Question II as a multi-classification problem. However, our later experiment results will show that this attempt usually cannot get desirable results due to the fact that the classification performance with different μ may be significantly distinct. In the next section, we will introduce an effective approach to answer question II by recursively solving question I.

CPB: Classification based Framework for Burst Time Prediction

In this section, we first describe how we transform the time prediction task to a classification problem. Then we introduce what knowledge we exploit and how to model them in a scale-independent manner to help the classification task.

The intuition is that though the magnitudes and time spans of the cascades may be significantly different, the shapes of their time series curves may be similar. Motivated by this, we propose the *time window based transformation* which equally divides the time spans of all the cascades into the same number of time windows. Then instead of predicting at which exact time point the burst will appear, we predict in which time window it occurs.

The general methodology would be to represent a cascade with a set of features extracted from rich information diffusion related knowledge, and then we use Classifiers to Predict the Burst will occur in which future time window (CPB). As an illustration, Figure 1 shows how we construct the classifier to predict whether a burst will occur in the 1_{st} future time window based on two cascades with significantly different time spans and popularities. The upper part illustrates how we extract the positive and negative samples from the raw times series data of the cascades. Horizontal axis is time and vertical axis is the reposting count. Given a cascade c with observed spreading process in the time interval $[t_0^c, t_{current}^c]$ (the green vertical line represents $t_{current}^c$), we first equally divide $[t_0^c, t_{current}^c]$ into K time windows. If a burst appears in the following future time window, the partial data between $[t_0^c, t_{current}^c]$ is considered as a positive sample; otherwise, it is a negative sample. We will introduce

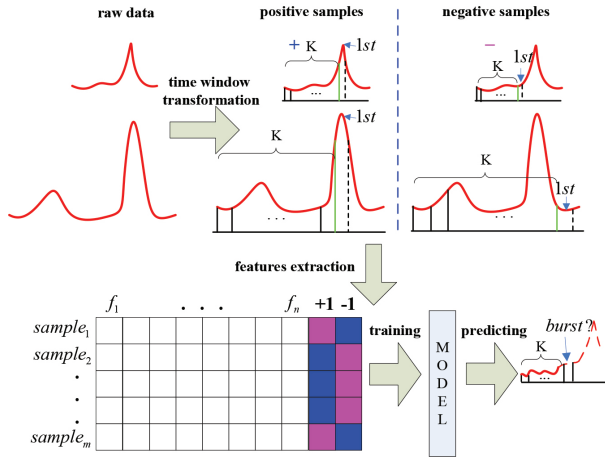


Figure 1: Illustration of the classifier construction to predict whether a burst will occur in the 1_{st} future time window.

how to extract these training samples in details later. Next we construct the classifier based on the extracted samples. Since the popularity information is insufficient, we will later elaborate what knowledge we will use and how to extract features from them. One can see that the proposed framework enables us to handle cascades with various time spans and popularities uniformly. Finally, when a new cascade comes, we use the trained model to predict whether a burst will appear in its next time window.

Time Window based Transformation to Construct Classifiers

To answer question I in Section 2, we first introduce how to extract training samples based on the time window based transformation and how to use these samples to construct classifiers. We next present how to answer question II by recursively solving question I.

Classifier Construction for Question I. For each cascade c , we construct classifiers for predicting whether the burst will appear in its 1_{st} , 2_{nd} , ..., μ_{th} future time windows respectively, and extract corresponding training samples. For brevity, we only introduce how to construct the 1_{st} future time window classifier as an example, and all the other classifiers can be constructed in the similar way.

To construct positive samples for the 1_{st} future time window classifier, we first identify the time of the global spike t_{max}^c for cascade c , and then we equally divide the time interval $[t_0^c, t_{max}^c]$ into $K + 1$ time windows $\{[t_0^c, t_1^c], (t_1^c, t_2^c] \dots (t_K^c, t_{max}^c]\}$. The time of the burst t_{max}^c can be considered to be in the last time window. If the current time is t_K^c and we can only observe the reposting data of c before t_K^c , the burst will occur in the next time window, namely the 1_{st} future time window. Therefore, cascade c with observed time interval $[t_0^c, t_K^c]$ can be considered as a positive sample of the 1_{st} future time window classifier. For the negative samples, we first randomly select a time point \bar{t}_K^c such that the burst will not appear in the next time window. To do this, we equally divide

the time interval $[t_0^c, t_{max}^c]$ into $K + l + 1$ time windows $\{[t_0^c, \bar{t}_1^c], \dots (\bar{t}_{K-l}^c, t_K^c], \dots, (\bar{t}_{K+l}^c, t_{max}^c]\}$, where l is a random positive or negative integer. A positive l means the burst does not occur before the K_{th} window; while a negative l means the burst occurs before the K_{th} window. Cascade c with observed time interval $[t_0^c, \bar{t}_K^c]$ can be considered as a negative sample of the 1_{st} future time window classifier. For a testing sample, assuming the start time is t_0 and the current time is $t_{current}$, we also divide $[t_0, t_{current}]$ into K time windows.

Answering Question II. To answer question II, we start with predicting whether the burst will appear in the 1_{st} future time window using the 1_{st} future time window classifier. If the answer is YES, the process stops and outputs the result; otherwise, we use the 2_{nd} future time window classifier to predict whether it will appear in the 2_{nd} future time window. The above process continues recursively until some classifier gives a positive prediction. If all the classifiers give the negative answer, we predict the burst appears in the last time window.

The reason why we conduct the prediction in this way is that, as shown in later experiment, the classification performance decreases with the increase of the parameter μ . Intuitively, bursts occurring in the near future time windows are easier to predict than those in farther future time windows. If two classifiers, for example the 1_{st} and 2_{nd} future time window classifiers both give positive predictions, we think the burst is more likely to appear in the 1_{st} future time window because the former classifier is more accurate.

Model Information Diffusion Related Knowledge in a Scale-Independent Manner

Besides the repost count, the cascades are also associated with a lot of other information, such as user profile and social relation. Oh et al.'s study showed that there are a number of mechanisms by which social influence is transmitted such as networked structure and conformity (Oh, Susarla, and Tan 2008). Cheng et al. also found that the network structure information are helpful to predict cascades (Cheng et al. 2014). Motivated by these works, we explore rich knowledge in information diffusion and roughly categorize them into four types: general time series based knowledge, fluctuation knowledge, user profile knowledge, and social relation knowledge. For each type of knowledge, we extract corresponding scale-independent features.

General Time Series based Knowledge The popularity of a cascade in each time window can be considered as general time series data (Wang et al. 2014b). We extract the following scale-independent features from the general time series based knowledge.

Average one-step increase rate (AIR_{+1}): Given the number of reposts n_k^c and n_{k+1}^c in the k_{th} and $(k + 1)_{th}$ time windows respectively, the average one-step increase rate between two successive windows is defined as

$$AIR_{+1}^c = \frac{1}{K-1} \sum_{i=1}^{K-1} \frac{n_{k+1}^c - n_k^c}{n_k^c} \quad (1)$$

Average two-step increase rate (AIR_{+2}): Similar to

AIR_{+1} , we extract the average increase rate between every other time window (n_k^c, t_k^c) and (n_{k+2}^c, t_{k+2}^c) as the AIR_{+2} .

Recent data may be more useful to predict the future trend of the cascade; hence we also extract some features that are only related to the latest data.

Average spreading speed in the latest l time windows (ASS_l). The average spreading speed in the latest l windows can be defined as

$$ASS_l^c = \frac{1}{l} \sum_{i=1}^l \frac{n_{K-i}^c}{t^c} \quad (2)$$

where t^c is the length of the time window in cascade c .

Average one-step increase rate in the latest l time windows (AIR_{+1}). The average one-step increase rate in the latest l time windows of cascade c is defined as

$$AIR_{+1}^c = \frac{1}{l-1} \sum_{i=0}^{l-1} \frac{n_{K-i}^c - n_{K-i-1}^c}{n_{K-i-1}^c} \quad (3)$$

Average two-step increase rate in the latest l time windows (AIR_{+2}). We extract two-step increase rate between every other time windows in the latest l time windows.

Fluctuation Knowledge The spreading process of cascades is rather dynamic and fluctuates over time (Petrovic, Osborne, and Lavrenko 2011; Myers and Leskovec 2014). An important reason causing the temporal dynamic is that users' behavior is highly related to the time. Myers et al. (Myers and Leskovec 2014) and Sakaki et al. (Sakaki, Okazaki, and Matsuo 2010) discovered that the fluctuation property of cascade spreading is helpful to predict the future popularity of photo reshare cascades and tweets. Therefore, we also explore the fluctuation knowledge of cascades and extract fluctuation features.

Hour (H). We use the hour of current time as a feature with 24 values from 0 to 23.

Day (D). The day of the week is selected as the second time related feature with 7 values from 0 to 6.

Number of local spikes N_l . Given a cascade c and the current time $t_{current}$, we identify all the local spikes before $t_{current}$ and use the number of local spikes as a feature.

Average normalized distance between two successive local spikes ($ADLL$). Assuming $k_{lmax,m}$ and $k_{lmax,(m+1)}$ are two time windows in which two successive local spikes occurs, the normalized distance between two successive local spikes is defined as $d^c(m, m+1) = k_{lmax,(m+1)} - k_{lmax,m}$. The average normalized distance between two successive spikes can be computed by

$$ADLL^c = \frac{1}{M} \sum_{i=1}^M d^c(i, i+1) \quad (4)$$

where m is the number of local spikes.

The normalized distance between the latest local spike and the current time (DLC). Assuming $k_{lmax,l}$ is the time window of the latest local spike and $k_{current}$ is the current time window, the normalized distance between the latest local spike and current time can be defined as

$$DLC^c = k_{current} - k_{lmax,l} \quad (5)$$

One-step consistency (F_{+1}^c). The one-step consistency between time windows k and $k+1$ is defined as

$$f_{+1}^c(k, k+1) = \begin{cases} 0 & \text{if } n_k^c \geq n_{k+1}^c \\ 1 & \text{if } n_k^c < n_{k+1}^c \end{cases} \quad (6)$$

The one-step consistency of c is the sum of the one-step consistency between all the successive two time windows

$$F_{+1}^c = \sum_{k=1}^{K-1} f_{+1}^c(k, k+1) \quad (7)$$

Two-step consistency (F_{+2}^c). We also extract the two-step consistency between every other time windows k and $k+2$.

User Profile Knowledge Different from traditional time series data, the cascades are triggered and driven by users. The posts originating from different users may have significantly different impact on the spreading of the cascades (Kupavskii et al. 2013). Hence we also use the user profile knowledge and categorize them into two types: profile based knowledge and authority based knowledge. The profile based knowledge includes **gender**, **location**, and **number of posts**. The authority based knowledge includes **number of followers**, **number of followees**, whether the user is a **verified** user, **PageRank score**, and **HITS score** of the user. The PageRank and HITS scores are computed based on the following relationship graph of all the users. For each type of knowledge, we first obtain the corresponding data in each time window, and then derive scale-independent features based on the time series data in all the time windows. Due to space limitation, we only take the gender as an example to illustrate how we extract scale-independent features from the gender knowledge.

In each time window (n_k^c, t_k^c) , assuming the numbers of male and female users reposting post c are $n_{k,m}^c$ and $n_{k,f}^c$ respectively, we compute the ratio $g_k^c = \frac{n_{k,m}^c}{n_{k,f}^c}$. By calculating the ratios in all the time windows, we obtain such a gender related time series $G^c = \{g_1^c, g_2^c, \dots, g_K^c\}$. Based on G^c , we can finally derive some scale-independent features such as average one-step increase rate, average two-step increase rate, the latest one-step gender ratio, and the latest two-step gender ratio. As the calculation of above derived features is similar to that of the general time series features, we omit the mathematical equations for brevity. For other user profile knowledge, we also extract scale-independent features in the similar way.

Social Relation Knowledge The social relations among users and the spreading paths of the cascades may also be helpful. Cheng et al. studied whether the cascade is spreading primarily within a community or across many to predict the future popularity of the cascade (Cheng et al. 2014). Ma et al. also studied that the total number of exposed users is an important feature in predicting Twitter hashtag popularity (Ma, Sun, and Cong 2013). To examine whether the social relation knowledge is helpful to predict burst time of cascades, we explore some structure knowledge based on users' social relations in cascade diffusion, and derive corresponding scale-independent features.

Table 1: Classification result for various learning algorithms.

Algorithm	F1-measure	AUC	Accuracy
Logistic Regression	0.826	0.895	82.8%
Multilayer Perceptron	0.899	0.915	90.2%
Adaboost	0.854	0.929	85.4%
Random Forest	0.892	0.904	89.2%
J48 Decision Tree	0.928	0.922	92.2%
Random Tree	0.904	0.934	92.4%
LibSVM	0.824	0.828	82.8%

We use five network structure measures to represent the social relation knowledge: **Wiener index** (Goel et al. 2013), **graph edit distance** (Papadimitriou, Dasdan, and Garcia-Molina 2008), **vertex and edge overlap** (VEO) (Papadimitriou, Dasdan, and Garcia-Molina 2008), **graph density** (Cheng et al. 2014), and **Entropy of degree distribution** (Kong et al. 2014). Wiener index is a measure of the structure vitality of a cascade. Graph edit distance and VEO are both used to measure the similarity of two graphs. Graph density is to measure how closely the nodes are connected in a graph. Entropy of degree distribution is a measure to describe the degree distribution of a graph.

We take the graph edit distance as an example to show how we model the social relation knowledge and extract features. All the other measures can be modeled in the similar way. For each time window k , we first extract a graph G_k^c based on the nodes involved in cascade c before the current time window and the following relationships among these nodes. Then we can obtain a set of graphs $\mathcal{G}^c = \{G_1^c, G_2^c, \dots, G_K^c\}$. Based on \mathcal{G}^c we can compute the graph edit distance $d_{k,k+1}$ between two successive graphs G_k^c and G_{k+1}^c . Then we can further obtain the time series data of the graph edit distance $D = \{d_{1,2}, d_{2,3}, \dots, d_{K-1,K}\}$. Using the time series data of D , we derive some scale-independent features similar to the general time series features.

Evaluations

We use the public available Sina Weibo reposting dataset¹ (Zhang et al. 2013) to evaluate the proposed approach. This dataset contains 1,776,950 users, 308,489,739 following relationships, 300,000 popular microblog diffusion episodes with the original microblog and all its reposts. On average each microblog has been reposted for about 80 times. First, we verify whether the proposed approach can learn an accurate classifier by examining the classification performance with various learning algorithms. Then we study the effect of parameters on the classification performance. Next we conduct the feature importance analysis to verify the importance of different knowledge in the studied task. Finally, we quantitatively evaluate how accurate CPB can predict the burst time of the cascades by comparing it against baselines.

Performance Analysis. We first exam the classification performance of various learning algorithms. We use 10-fold cross validation to evaluate on three metrics: **F1-measure**,

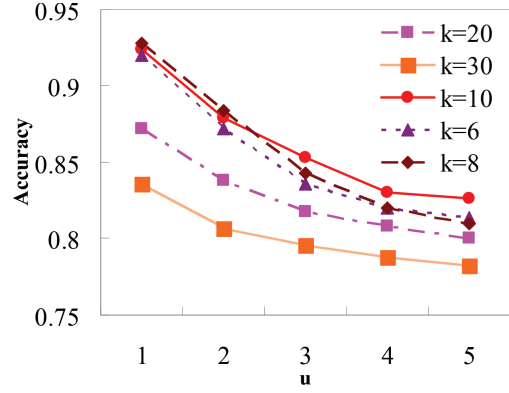


Figure 2: Classification accuracy with various K and μ .

AUC (Area Under ROC Curve) and **classification accuracy**.

In this experiment, we divide the time spans of all the cascades into 10 time windows and predict whether the burst occurs in the first future time window. The result is given in Table 1. As shown in the table, the classification accuracy of all the used classification techniques are over 80%. The performance of Multilayer Perceptron, Random Forest, and J48 Decision Tree are not significantly different: the accuracy is around 90%. It implies that when sufficient features are available, this prediction task is not much sensitive to the choice of the learning algorithms. In our experiment, Random Tree is shown to be the most accurate algorithm with a classification accuracy of 92.4% and AUC value of 0.934. Hence, in the following experiments we use Random Tree as our classification method.

Effect of Parameters. To study the effect of parameters K and μ on the classification performance, we conduct experiment with various K over different μ . Here K is the number of divided time windows and μ is the future time window in which we predict whether the burst occurs. Note that given a cascade c , the burst time window μ may change if K changes. The results are given in Figure 2. The x-axis is the future time window with μ from 1 to 5 and the y-axis shows the classification accuracy. We first set K to a set of relatively small numbers: $K = 6$, $K = 8$, and $K = 10$. Ones can observe a monotonically decrease trend in classification accuracy with the increase of the parameter μ . It means the burst occurring in a farther away future time window is harder to predict than that occurring in a time window which is closer to the current time. Ones also can see the classification performance does not show significant difference when the K values are relatively small.

To further verify whether larger K can impact the classification performance, we set K to two larger values: $K = 20$ and $K = 30$. Ones can see that the classification performance drops remarkably if K is set to a relatively large value. The result is not surprising as larger K means smaller time window and more fine-grained prediction. Smaller time window makes the difference between the data in two successive time windows smaller and thus harder to distinguish.

¹<http://arnetminer.org/Influencelocality#b2354>

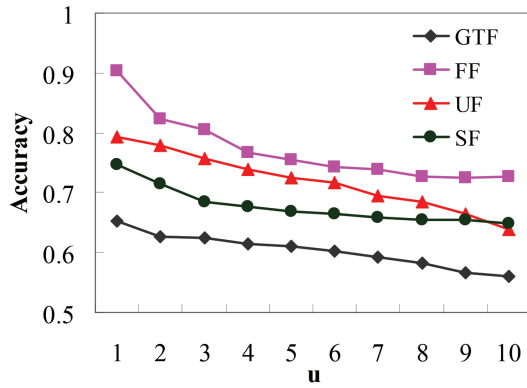


Figure 3: The classification accuracy by using general time series features (GTF), fluctuation features (FF), user profile features (UF), and social relation features (SF) separately.

Feature Importance Analysis We study the importance of the features derived from different types of knowledge. Figure 3 shows the classification accuracy achieved by the classifiers trained on each group of features separately. We summarize the results by the following observations:

- *General time series related features are surprisingly not very useful.* Figure 3 shows that the classification accuracy is less than 66% if we only use the general time series features. It implies that it is hard to predict the burst time by only using the time series of the cascade popularity.
- *Fluctuation features are most important.* With the fluctuation features, the classification accuracy on the 1_{st} future time window is near 90%, which is the highest in the four groups of features.
- *The user profile features and social relation features are helpful.* Surprisingly, user profile and social relation features both perform better than general time series features. Only using the user features, the accuracy is around 80% for the 1_{st} future time window prediction, and the figure is 75% for the social relation features only.

Quantitive Comparison with Baselines To quantitatively evaluate how accurate CPB can predict the burst time of cascades, we compare it with the following four baselines.

- **Random.** We randomly select a future time window as the time window in which the burst occurs.
- **Multi-Classification (Multi_C).** We consider the problem of predicting the time window in which the burst occurs as a multi-classification problem.
- **SPIKEM.** SPIKEM (Matsubara et al. 2012) is designed to capture the diffusion patterns of cascades. To make it comparable, we first use SPIKEM to forecast the future volume of a cascade based on its early data, and then identify the burst time window based on the predicted future volume in each future time window.
- **CPB Only Using Time Series Features of the Cascades Popularity (CPB_CP).** To study whether the rich knowledge can improve the prediction performance, we also use

CPB with only the time series features of the cascades popularity as a baseline.

Table 2: Quantitive comparison against four baselines.

μ	Random	Multi_C	CPB_CP	SPIKEM	CPB
2	0.250	0.060	0.056	0.108	0.042
3	0.296	0.132	0.125	0.224	0.065
4	0.312	0.135	0.158	0.227	0.102
5	0.320	0.172	0.186	0.246	0.142
6	0.325	0.178	0.182	0.222	0.144
7	0.330	0.176	0.184	0.254	0.146
8	0.336	0.188	0.192	0.267	0.152

We use the Mean Absolute Error (MAE) computed by $MAE = E(\frac{|I^c - \hat{I}^c|}{l})$ as the evaluation metric. Here I^c denotes the true future time window in which the burst occurs, \hat{I}^c is the predicted time window, and l is the number of future time windows. For ease of comparison, we first fix the number of future time windows l . Then we select the testing samples whose bursts occur in one of the l future time windows. Table 2 gives the results with l from 2 to 8. The figures in bold show the best results. One can see CPB performs significantly better than the four baselines in terms of MAE in all the cases. The MAE increases with the increase of l , which implies bursts in a farther away time window is harder to predict. One can also see although Multi-Classification approach is significantly better than random method, it is less effective than CPB. The performance of SPIKEM is not desirable: even inferior to Multi-Classification method. This is mainly because SPIKEM only utilizes the time series data of the cascade popularity, but cannot capture and handle various knowledge we study. Compared to CPB_CP, the MAE value achieved by CPB decreases by an average of about 30%. It implies that the rich knowledge in information diffusion does help our task.

Conclusion

In this paper, we studied the problem of burst time prediction in cascades and proposed a novel classification based framework CPB by exploring rich knowledge associated to information diffusion. Our solution allows us to predict the cascades with diverse magnitudes and time spans in a unified manner, since we conduct the prediction in the time window granularity by a novel time window based transformation. Extensive evaluations on a real social network dataset demonstrate the effectiveness of CPB.

Acknowledgments

This work is supported in part by NSFC (Nos. 61170189, 61370126, 61202239), the Fund of the State Key Laboratory of Software Development Environment (No. SKLSDE-2013ZX-19), the Innovation Foundation of BUAA for PhD Graduates (No. YWF-14-YJSY-021), Microsoft Research Asia Fund (No. FY14-RES-OPP-105), NSF through grants CNS-1115234, and OISE-1129076, and the Pinnacle Lab at Singapore Management University.

References

- Barabási, A.-L. 2005. The origin of bursts and heavy tails in human dynamics. *Nature* 435:207–211.
- Barabási, A.-L. 2011. *BURSTS: The Hidden Pattern Behind Everything We Do*. Plume, reprint edition.
- Chen, Y.; Amiri, H.; Li, Z.; and Chua, T.-S. 2013. Emerging topic detection for organizations from microblogs. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*.
- Cheng, J.; Adamic, L. A.; Dow, P. A.; Kleinberg, J.; and Leskovec, J. 2014. Can cascades be predicted? In *Proceedings of the 23rd International World Wide Web Conference*.
- Crane, R., and Sornette, D. 2008. Robust dynamic classes revealed by measuring the response function of a social system. *Proceedings of the National Academy of Sciences of the United States of America* 105(41):15649–15663.
- Goel, S.; Anderson, A.; Hofman, J.; and Watts, D. 2013. The structure virality of online diffusion. In *Preprint*.
- Gruhl, D.; Guha, R.; Novak, R. K. J.; and Tomkins, A. 2005. The predictive power of online chatter. In *Proceedings of the 11th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*.
- Hong, L.; Dan, O.; and Davison, B. D. 2011. Predicting popular messages in twitter. In *Proceedings of the 20th International World Wide Web Conference*.
- Hu, X.; Sun, N.; Zhang, C.; and Chua, T.-S. 2009. Exploiting internal and external semantics for the clustering of short texts using world knowledge. In *Proceedings of the 18th ACM conference on Information and knowledge management*.
- Hu, X.; Tang, L.; Tang, J.; and Liu, H. 2013. Exploiting social relations for sentiment analysis in microblogging. In *Proceedings of the sixth ACM International Conference on Web Search and Data Mining*.
- Kleinberg, J. 2002. Bursty and hierarchical structure in streams. In *Proceedings of the 8th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*.
- Kong, S.; Mei, Q.; Feng, L.; Zhao, Z.; and Ye, F. 2014. On the real-time prediction problem of bursting hashtags in twitter. *CoRR abs/1401.2018*.
- Kupavskii, A.; Umnov, A.; Gusev, G.; and Serdyukov, P. 2013. Predicting the audience size of a tweet. In *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media*.
- Li, L.; Liang, C.-J. M.; Liu, J.; Nath, S.; Terzis, A.; and Faloutsos, C. 2011. Thermocast: A cyber-physical forecasting model for data centers. In *Proceedings of the 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*.
- Ma, Z.; Sun, A.; and Cong, G. 2013. On predicting the popularity of newly emerging hashtags in twitter. *Journal of the American Society for Information Science and Technology* 64(7):1399–1410.
- Matsubara, Y.; Sakurai, Y.; Prakash, B. A.; Li, L.; and Faloutsos, C. 2012. Rise and fall patterns of information diffusion: Model and implications. In *Proceedings of the 18th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*.
- Mill, T. C. 1990. *Time Series Techniques for Economists*. Cambridge University Press.
- Myers, S., and Leskovec, J. 2014. The bursty dynamics of the twitter information network. In *Proceedings of the 23th International World Wide Web Conference*.
- Oh, J.; Susarla, A.; and Tan, Y. 2008. Examining the diffusion of user-generated content in online social networks. *Social Science Research Network*.
- Papadimitriou, P.; Dasdan, A.; and Garcia-Molina, H. 2008. Web graph similarity for anomaly detection. In *Proceedings of the 17th International World Wide Web Conference*.
- Parikh, N., and Sundaresan, N. 2008. Scalable and near real-time burst detection from ecommerce queries. In *Proceedings of the 14th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*.
- Petrovic, S.; Osborne, M.; and Lavrenko, V. 2011. Rt to win! predicting message propagation in twitter. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*.
- Pinsen, D. 2012. Predicting the bursting of a market bubble. In <http://finance.yahoo.com/news/predictingbursting-market-bubble-171432469.html>.
- Sakaki, T.; Okazaki, M.; and Matsuo, Y. 2010. Earthquake shakes twitter users: real-time event detection by social sensors. In *Proceedings of the 19th International World Wide Web Conference*.
- Szabo, G., and Huberman, B. A. 2008. Predicting the popularity of online content. *ArXiv e-prints*.
- Vazquez, A.; Oliveira, J. G.; Dezsó, Z.; Goh, K.-I.; Kondor, I.; and Barabási, A.-L. 2006. Modeling bursts and heavy tails in human dynamics. *Physical Review E*(73).
- Wang, S.; Hu, X.; Yu, P. S.; and Li, Z. 2014a. Mmrates: Inferring multi-aspect diffusion networks with multi-pattern cascades. In *Proceedings of the 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*.
- Wang, S.; Xie, S.; Zhang, X.; Li, Z.; Yu, P. S.; and Shu, X. 2014b. Future influence ranking of scientific literature. In *Proceedings of the 2014 SIAM International Conference on Data Mining*.
- Yang, J., and Leskovec, J. 2011. Patterns of temporal variation in online media. In *Proceedings of the fourth ACM International Conference on Web Search and Data Mining*.
- Zhang, J.; Liu, B.; Tang, J.; Chen, T.; and Li, J. 2013. Social influence locality for modeling retweeting behaviors. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence*.
- Zhu, Y., and Shasha, D. 2003. Efficient elastic burst detection in data streams. In *Proceedings of the 9th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*.