Electronic Preprint for Journal of Information Processing Vol.24 No.6

Regular Paper

Non-linear Time-series Analysis of Social Influence

Thinh Minh Do[1,a)] Yasuko Matsubara[1,b)] Yasushi Sakurai[1,c)]

Received: December 29, 2015, Accepted: February 4, 2016

Abstract: Given a large collection of time-evolving online user activities, such as Google Search queries for multiple keywords of various categories (celebrities, events, diseases, etc...), which consist of d keywords/activities, for l countries/locations of duration n, how can we find patterns and rules? For example, assume that we have the online search volume for "Harry Potter", "Barack Obama" and "Amazon", for 232 countries/territories, from 2004 to 2015, which include external shocks, sudden change of search volume, and more. How do we go about capturing non-linear evolutions of local activities and forecasting future patterns? Our goal is to analyze a large collection of time-evolving sequences, and moreover, to find the answer for the following issues: (a) Are there any important external shocks/events relating to the keywords in the sequences? (b) If there are, can we automatically detect them? (c) Are there any countries/territories which have different reacts to the global trend? In this paper, we present Δ-SPOT, a unifying analytical non-linear model for large scale web search data; as well as an efficient and effective fitting algorithm, which solves the above problems. Δ-SPOT can also forecast long-range future dynamics of the keywords/queries. Extensive experiments on real data show that our method outperforms other effective methods of non-linear mining in terms of accuracy in both fitting and forecasting.

Keywords: time-series data, automatic mining

1. Introduction

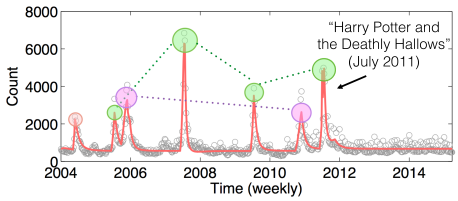
Online news, blogs, SNS and many other web search services have been speedily developing and playing a very important part in information searching. Our goal is to detect patterns, rules and outliers in a huge set of web search data, consisting of tuples of the form: *(query, location, time)*. For example, assume that we have the online search activities for "Harry Potter", in 232 countries/territories from 2004 to 2015.

So, how can we find meaningful information, such as the external shocks/events that happened during the period of 11 years? In the case that such events happened, do they have any relationship between each others (frequent/cyclic events or not?) Also, can we detect global/local-level patterns? Are there countries/locations that react differently from the global trend? Can we forecast the dynamics of future events?

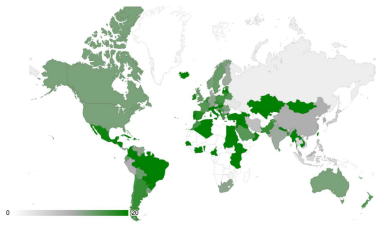
In this paper, we propose Δ-SPOT, a unifying analytical non-linear model which is sense-making, scalable and parameter-free, and provides a good summary of large collections of local online activities. Intuitively, we wish to solve the following problem:

Informal Problem. Given a large collection of online activities, which consists of d keywords in l locations of duration n with missing values and external shocks, we want to

- detect external shocks (important events in reality),
- find global and local patterns, and
- forecast future activities.



(a) Fitting result of Δ-SPOT



(b) World-wide reaction

Fig. 1 Modeling power of Δ-SPOT: (a) It automatically detects external events of keyword "Harry Potter" and (b) the world-wide reaction to the release of the last episode of "Harry Potter" movie series.

Especially, we want to capture all these features automatically and effectively.

Preview of Our Results. Figure 1 (a) shows the search volume for the keyword "Harry Potter" from 2004 to 2015 (11 years) as grey circles, and our fitted model, as solid red line. Our method automatically spots seven big cyclic/non-cyclic events that relate to "Harry Potter". For example, (a) the biennial release (in July) of "Harry Potter" movies and books (shown as green circles), which corresponds to the major publication of works on "Harry Potter", (b) the release of new episodes of "Harry Potter" movie series, held in November (shown as purple circles), and (c) non-cyclic spike in May (shown in red circle).

[1] Graduate School of Science and Technology, Kumamoto University, Kumamoto 860–8555, Japan
a) do@dm.cs.kumamoto-u.ac.jp
b) yasuko@cs.kumamoto-u.ac.jp
c) yasushi@cs.kumamoto-u.ac.jp

Figure 1 (b) shows the world-wide reaction to the release of the last episode of "Harry Potter". It is clearly shown that most of the active (high reaction-level) countries have huge number of fans of "Harry Potter" (For example: the U.S., European countries, English native speaking countries, etc.)

Contributions. In this paper, we propose Δ-SPOT, a unifying analytical non-linear model for large-scale online user activities. Our method has the following desirable properties:

(1) **Sense-making**: Our method can detect external shocks which are related to real-time events, such as the cyclic sporting occasions, or celebrities relating special events.
(2) **Automatic**: Thanks to our modeling framework, our method is fully automatic, requiring no manual tuning, where the goal is to minimize the cost of the resulting modeling.
(3) **Scalable**: Our method scales linearly to the input size.
(4) **Parameter-free**: Δ-SPOT requires no parameters or specialized tuning.

Outline. The rest of the paper is organized in the conventional way. Next, we describe related work, followed by our proposed model and algorithms, experiments, discussion and conclusions.

2. Related Work

We provide a survey of the related literature, which falls broadly into two categories: (a) Pattern discovery in time series and (b) Social activity analysis.

Pattern Discovery in Time Series. In recent years, there has been a huge interest in mining time-stamped data [9], [18], [21]. Traditional approaches typically use linear methods, such as auto-regression (AR), linear dynamical systems (LDS), TBATS [8] and their variants [3], [5], [6], [20]. TriMine [12] is a scalable method for forecasting complex time-stamped events, while, [10] developed AutoPlait, which is a fully-automatic mining algorithm for co-evolving sequences. Rakthanmanon et al. [17] proposed a similarity search algorithm for "trillions of time series" under the DTW distance.

Social Activity Analysis. Analyses of epidemics and social media have attracted a lot of interest. The work described in Ref. [13] studied the rise and fall patterns in the information diffusion process through online social media. Prakash et al. [16] described the setting of two competing products/ideas spreading over a network, and provided a theoretical analysis of the propagation model for arbitrary graph topology. FUNNEL [14] is a non-linear model for spatially co-evolving epidemic tensors, while, EcoWeb [11] is the first attempt to bridge the theoretical modeling of a biological ecosystem and user activities on the Web. For online activity analysis, Gruhl et al. [2] explored online "chatter" (e.g., blogging) activity, and measured the actual sales ranks on Amazon.com, while Ginsberg et al. [1] examined a large number of search engine queries tracking influenza epidemics.

Contrast to the Competitors. Table 1 illustrates the relative advantages of our method. Only our Δ-SPOT matches all requirements, while

- The SI model (and SIR, SIRS, SKIPS [19], etc.) can compress the data into a fixed number of parameters, and capture the dynamics of epidemiological data, however, it cannot describe periodic events, and is incapable of forecasting.
- The traditional AR, ARIMA and related forecasting methods including AWSOM [15], PLiF [7] and TriMine [12] are *fundamentally* unsuitable for our setting, because they are based on linear equations, while we employ *non*-linear equations. Moreover, most of them require parameter tuning.
- FUNNEL [14] is a comfortable non-linear model for time-evolving tensor mining. However, it cannot detect cyclic external shocks and was applied for epidemic sequences.

Table 1 Capabilities of approaches. Only our approach meets all specifications.

	SI/++	AR/++	FUNNEL	Δ-SPOT
Non-linear	√		√	√
Outliers detection			√	√
Online activities				√
Cyclic events/shocks				√
Local analysis			√	√
Parameter-free			√	√
Forecasting		√	√	√

Table 2 Symbols and definitions.

Symbol	Definition
d	Number of keywords/queries
l	Number of locations/countries
n	Duration of sequences
$\mathcal{X}$	3rd-order tensor ($\mathcal{X} \in \mathbb{N}^{d \times l \times n}$)
$\boldsymbol{x}_{ij}$	Local-level sequence of keyword i in location j i.e., $\boldsymbol{x}_{ij} = \{x_{ij}(t)\}_{t=1}^{n}$
$\bar{\boldsymbol{x}}_i$	Global-level sequence of keyword i i.e., $\bar{\boldsymbol{x}}_i = \sum_{j=1}^{l} \boldsymbol{x}_{ij}$
$S_{ij}(t)$	Count of (S)usceptibles i in location j at time t
$I_{ij}(t)$	Count of (I)nfectives i in location j at time t
$V_{ij}(t)$	Count of (V)igilants i in location j at time t
$\mathbf{B_G}$	Base global matrix ($d \times 4$)
$\mathbf{B_L}$	Base local matrix ($d \times l$)
$\mathbf{R_G}$	Growth effect global matrix ($d \times 2$)
$\mathbf{R_L}$	Growth effect local matrix ($d \times l$)
$\mathcal{S}$	External shock tensor i.e., $\mathcal{S} = \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_k\}$
$\mathcal{F}$	Complete set of Δ-SPOT i.e., $\mathcal{F} = \{\mathbf{B_G}, \mathbf{B_L}, \mathbf{R_G}, \mathbf{R_L}, \mathcal{S}\}$

3. Proposed Model

3.1 Intuition behind Our Method

Assume that we receive time-stamped activities of the form *(query, location, time-tick)*. We then have a collection of sequences with d unique queries/keywords, l locations/countries with duration n. We can treat this set of $d \times l$ sequences as a 3rd-order tensor, i.e., $\mathcal{X} \in \mathbb{N}^{d \times l \times n}$, where the element $x_{ij}(t)$ of $\mathcal{X}$ shows the total number of entries of the i-th keyword in the j-th country at time-tick t. For example, ('Olympics', 'US', 'August 3-9, 2008'; 36), means that the search volume for 'Olympics' in 'US' on 'August 3-9 in 2008' is '36'.

We refer to each sequence of the i-th keyword in the j-th location: $\boldsymbol{x}_{ij} = \{x_{ij}(t)\}_{t=1}^{n}$, as a "local/country"-level web search sequence. Similarly, we can turn these local sequences into "global/world"-level web search sequences: $\bar{\boldsymbol{x}}_i = \{\bar{x}_i(t)\}_{t=1}^{n}$, where $\bar{x}_i(t)$ shows the total count of the i-th keyword at time-tick t, i.e., $\bar{x}_i(t) = \sum_{j=1}^{l} x_{ij}(t)$.

Preliminary Observations. Here, let us provide the reader with several important observations.

- **(P1) Basic Trends:** We assume that the popularity size of each activity evolves over time. The popularity size corre-

sponds to the aggregated volume of each user who is interested in each topic in each country. For example, the event that Barack Obama became the 44th president of the US (please see Fig. 5 (a)), is a very important event, not only for the US citizens, but also for people around the world. Many users will spend time searching for Obama's biography or his career. Furthermore, they will share the stories to their friends; and eventually, this leads to an exponential growth in popularity size.

- **(P2) Area Specificity:** For each topic, users all over the world have different types of reaction towards it. Due to social network connection condition, or some specific reasons, there may be a huge spike in some countries at a time-tick, while nothing happens in others. For example, as shown in Fig. 1 (b), users all over the world react differently to the release of the last episode of "Harry Potter" movie series. Mostly, users from countries, in which Harry Potter are popular, are highly active to search for the episode's information.
- **(P3) Population Growth Effect:** We also find that there exists a sudden change of popularity base size in some sequences. For example, the number of searches for "Amazon" sharply rises from 2010 until now. We call this behavior the population growth effect. This phenomenon consists of different behavior compared to the external shock effect. We will discuss it further in Section 3.3 (Fig. 4).
- **(P4) External Shock Events:** One of our main goals is to detect the external shocks that refer to real-time events. The search volume for a keyword sharply rises when some events (special publication, championship, performance, etc...) relating to that keyword happened. For example, Fig. 1 (a) shows seven big events corresponding to the release date of Harry Potter movies and books, from the first spike in 2004 (movie, episode 3), until the last one in 2011 (movie, episode 7 - part 2). Most importantly, we observe that some external shocks have got the cyclic property. These cyclic events happen at the same time-tick of a specific window-size (i.e., one year, two years, etc..), within the same duration. It is important to extract these cyclic pattern from the large set of all external events. In other words, we want to capture the periodical (annual, biennial, quadrennial) events, which provides a highly sensitive view of big events in reality.

Summary. In this paper, we propose a new model, namely, Δ-SPOT, which tries to incorporate all the above important properties that we observed in the real dataset. Consequently, we would like to capture the following properties:

- (**P1**): basic trends
- (**P2**): area specificity and sensitivity
- (**P3**): population growth effect
- (**P4**): cyclic external events

3.2 Δ-SPOT - with a Single Sequence

We begin with the simplest case, assuming that we are given a single sequence. The model we propose has nodes (=users) of three classes:

- Susceptible: nodes in this class can get influenced by their neighboring nodes who have searched for it. In other words, citizens of this class are always ready to search for the keywords.
- Infective: nodes who already searched for the keywords, also capable of influencing other available nodes (share, or tell the story about the keywords), namely, transmitting the interest in the topic to the citizens in the susceptible class.
- Vigilant (i.e., busy/unavailable): nodes in this class do not get condition to search for the information (no network connection, no free time to care about the topic), so they are immune to the influence of the trend.

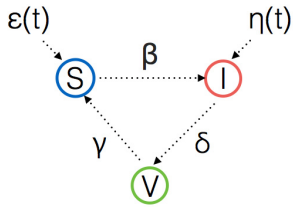


Fig. 2 Δ-SPOT diagrams: classes of population: susceptibles, infectives, and vigilants.

Figure 2 shows a diagram of our model, in which,

- β represents the rate of effective contacts between citizens in infective and susceptible classes;
- δ is the rate at which infected citizens lost interest in the topic and quit searching for it;
- γ is the immunization loss probability for a change in status: being ready to search for the topic.

We also introduce two more parameter, $\epsilon(t)$ and $\eta(t)$, to represent the external shock effect and growth effect, respectively. The idea is that the number of the susceptible class $S(t)$ is the count of users available for infection, and if there is an external shock event at time-tick t, the infection becomes stronger than usual. Therefore, each infective pair would lead to a new infective citizen, and will eventually cause a major spike. With respect to the growth effect, it starts at time t_η and make the number of infectives rise quickly to a new base.

Model 1 (Δ-SPOT-single) Our model can be described as the following equations:

$$\begin{aligned} S(t+1) &= S(t) - \beta S(t)\epsilon(t)I(t)(1+\eta(t)) + \gamma V(t) \\ I(t+1) &= I(t) + \beta S(t)\epsilon(t)I(t)(1+\eta(t)) - \delta I(t) \\ V(t+1) &= V(t) + \delta I(t) - \gamma V(t) \end{aligned} \quad (1)$$

where, the growth effect started at time t_η and $\eta(t)$ is defined as:

$$\eta(t) = \begin{cases} 0 & (t < t_\eta) \\ \eta_0 & (t \geq t_\eta) \end{cases}$$

In addition, we introduce the temporal susceptible rate, $\epsilon(t)$, which is defined as follows:

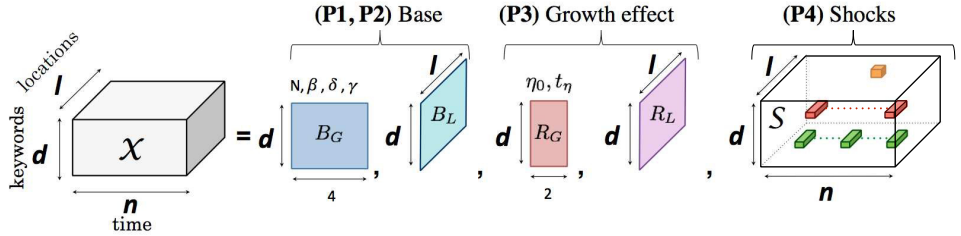
$$\epsilon(t) = 1 + \sum_{i=1}^{k} f(t; s_i)$$

$$f(t; s) = \begin{cases} \epsilon_0 & (t_s + t_p\lceil t/t_p \rceil < t < t_s + t_p\lceil t/t_p \rceil + t_w) \\ 0 & (else) \end{cases}$$

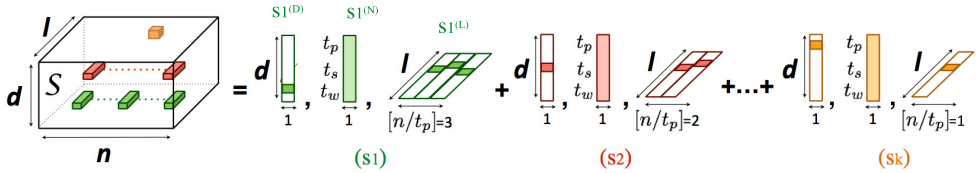
where, k is the number of shocks, and if $k = 0$, then $\epsilon(t) = 1$.

Here, each external shock consists of $s = \{t_p, t_s, t_w, \epsilon_0\}$, i.e.,

- t_p: Periodicity of the event (if $t_p = \infty$, the event is non-cyclic).



(a) Δ-SPOT structure (i.e., $\mathcal{F} = \{\mathbf{B_G}, \mathbf{B_L}, \mathbf{R_G}, \mathbf{R_L}, \mathcal{S}\}$)



(b) External shock tensor (i.e., $\mathcal{S} = \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_k\}$)

Fig. 3 Δ-SPOT structure: (a) important properties extracted from tensor $\mathcal{X}$. Also, (b) external shock tensor $\mathcal{S}$ consists of a set of k components.

- t_s: Starting point of the event.
- t_w: Duration of the event.
- ϵ_0: Strength of the external shock.

3.3 Δ-SPOT - with multi-evolving sequences

So far we have seen how Δ-SPOT captures the dynamics of a single sequence. The next question is: how can we apply Δ-SPOT to multiple time-evolving sequences in $\mathcal{X}$, and capture the individual behavior of d keywords in l locations/countries?

We want to extract the main trends and external patterns of co-evolving sequences $\mathcal{X} \in \mathbb{N}^{d \times l \times n}$, and make a good representation of $\mathcal{X}$. **Figure 3** shows our modeling framework. Given a tensor $\mathcal{X}$, it extracts important patterns with respect to the following aspects, base properties of global and local trends $\mathbf{B_G}$, $\mathbf{B_L}$, population growth effect $\mathbf{R_G}$, $\mathbf{R_L}$, and external shock events $\mathcal{S}$.

Definition 1 (Complete set of Δ-SPOT) Let $\mathcal{F}$ be a complete set of parameters (namely, $\mathcal{F} = \{\mathbf{B_G}, \mathbf{B_L}, \mathbf{R_G}, \mathbf{R_L}, \mathcal{S}\}$) that describe the global/local patterns of the sequences in $\mathcal{X}$.

Next, we will see each property in detail.

(P1) Base trends and global influence. Basically, we assume that the following parameters are the same for all l countries.

Definition 2 (Base-global matrix $\mathbf{B_G}$ ($d \times 4$)) Let $\mathbf{B_G}$ be the set of global parameters of d keywords/queries, where $\{N_i, \beta_i, \delta_i, \gamma_i\}$ is the parameter set of the i-th keyword, and $N_i = S_i(t) + I_i(t) + V_i(t)$.

For example, the potential infection rate of each keyword (e.g., "Harry Potter", "Amazon") should be the same for US and JP.

(P2) Area specificity. Next, we also want to analyze and explain location-specific patterns and trends in $\mathcal{X}$. For example, what is the difference of users reaction for keyword "Ebola" between the U.S. (US) and Nepal (NP)? Our answer is: their behavior is similar, except for the "local sensitivity" of the sequence. Specifically, we share the parameters of the global-level matrices for all l countries. with one exception, N_{ij}, which describes the total population of users for keyword i in the j-th country. Specifically, we set the invariant, $N_{ij} = S_{ij}(t) + I_{ij}(t) + V_{ij}(t)$.

Definition 3 (Base-local matrix $\mathbf{B_L}$ ($d \times l$)) Let $\mathbf{B_L}$ be a parameter set of the potential population of d keywords and l countries, i.e., $\mathbf{B_L} = \{b^{(L)}{}_{ij}\}_{i,j=1}^{d,l}$, where $b^{(L)}{}_{ij}$ is the potential population of susceptibles of the i-th keyword in the j-th country.

This parameter corresponds to the fraction of individuals who are likely to be infected by the trend. For example, US has more users than NP, because they have more capacities for network connection.

(P3) Population growth effect. The growth effect appears due to the launch of new products and services that raise the interest of users, which should have the same starting time all over the world.

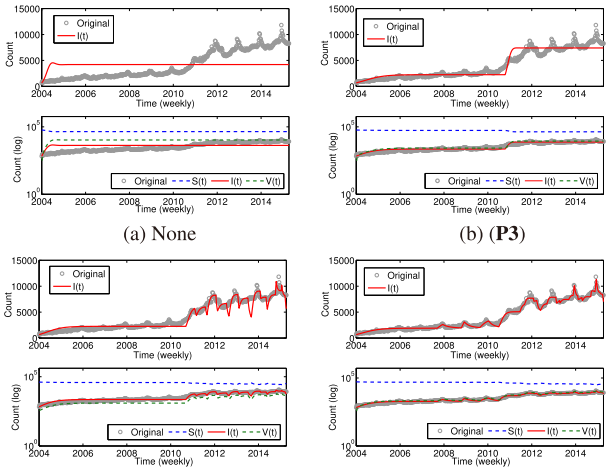
Definition 4 (Growth-global $\mathbf{R_G}$ ($d \times 2$)) Let $\mathbf{R_G}$ be the set of global growth effect parameters of d keywords/queries, where $\{\eta_{0i}, t_{\eta i}\}$ is the parameter set of the i-th keyword.

The growth effect has the same starting time, but different growth rate for l countries.

Definition 5 (Growth-local $\mathbf{R_L}$ ($d \times l$)) Let $\mathbf{R_L}$ be a parameter set of the potential population of d keywords and l countries, i.e., $\mathbf{R_L} = \{r^{(L)}{}_{ij}\}_{i,j=1}^{d,l}$, where $r^{(L)}{}_{ij}$ is the population growth rate of the i-th keyword in the j-th country.

(P4) Cyclic external events. We describe one external shock event in terms of three aspects, *(keyword, country, time)*, for example, "Harry Potter, world-wide, Jul 15-21,2007". To describe each external shock event, we create a new parameter set, namely external shock tensor $\mathcal{S}$, which consists of a set of k external shock events, as described in Fig. 3 (b). i.e., $\mathcal{S} = \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_k\}$ A single external shock event $\mathbf{s}$ can be described as three components: $\mathbf{s} = \{\mathbf{s^{(D)}}, \mathbf{s^{(N)}}, \mathbf{s^{(L)}}\}$.

- The $(d \times 1)$ size component $\mathbf{s^{(D)}}$, which represents the external view for d keywords/queries.
- The (3×1) size component $\mathbf{s^{(N)}}$, which describes the periodicity (t_p), the starting time (t_s), and the duration (t_w) of the external event.
- The $(\lceil n/t_p \rceil \times l)$ size component $\mathbf{s^{(L)}}$, which expresses the strength of the external shocks of one event in l countries,



(a) None (b) (**P3**) (c) (**P4**) (d) (**P3**) + (**P4**)

Fig. 4 Influence of combining growth effect and external shock effect: compared with the case of using (a) none of above, (b) only growth effect, (c) only external shock effect, and (d) the combination of both effects. Clearly, (d) fits the data very well.

where $\lceil n/t_p \rceil$ is the number of shocks belonging to that event.

Consequently, we have:

Definition 6 (External shock tensor $\mathcal{S}$) Let $\mathcal{S}$ be a 3rd-order tensor of k external shock events, i.e., $\mathcal{S} = \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_k\}$, where the matrices show the parameters in terms of three components.

Figure 4 compares the fitting results of the keyword "Amazon", in four different cases to demonstrate the influence of the growth effect (**P3**) and external shocks (**P4**). The result shows the benefit of treating the growth effect differently from external shock effect, as well as combining these two effects to achieve good fitting results (See Fig. 4 (d))[*1].

4. Algorithm

In this section, we describe our fitting algorithm, Δ-SPOT-FIT. Our goal is to extract the important patterns of online user activities from $\mathcal{X}$. More specifically, the problem that we want to solve is as follows:

Problem 1 **Given** a tensor $\mathcal{X}$ of $(keyword, country, time)$ triplets, **Find** a compact description that best summarizes $\mathcal{X}$, that is, $\mathcal{F} = \{\mathbf{B_G}, \mathbf{B_L}, \mathbf{R_G}, \mathbf{R_L}, \mathcal{S}\}$.

We want to find a good representation $\mathcal{F}$ to solve the problem. The essential questions are: (a) How can we estimate the parameter set that best captures the dynamics and patterns in $\mathcal{X}$? (b) How should we decide the number of external shocks k? (c) How can we treat an event to be cyclic or not?

4.1 Model Quality and Data Compression

We propose an intuitive coding scheme, which is based on the minimum description length (MDL) principle. Here, it follows the assumption that the more we can compress data, the more we can detect its hidden patterns.

Model Description Cost. The description complexity of model parameter set consists of the following terms,

- The number of keywords d, locations l, and time-ticks n require $\log^*(d) + \log^*(l) + \log^*(n)$ bits[*2].
- The model parameter set of the global base ($\mathbf{B_G}$), global growth effect ($\mathbf{R_G}$), and local base, growth effect($\mathbf{B_L}, \mathbf{R_L}$), matrices require $d \times 4$, $d \times 2$, $d \times l$ parameters, respectively, i.e., $Cost_M(\mathbf{B_G}) + Cost_M(\mathbf{R_G}) + Cost_M(\mathbf{B_L}) + Cost_M(\mathbf{R_L}) = c_F \cdot d(4 + 2 + l)$, where c_F is the floating point cost[*3].

Similarly, the model description cost of the external shock tensor $\mathcal{S} = \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_k\}$ consists of the following:

- The number of external shocks k requires $\log^*(k)$ bits.

Also, for each shock $\mathbf{s}$, it requires $Cost_M(\mathbf{s}) = Cost_M(\mathbf{s^{(D)}}) + Cost_M(\mathbf{s^{(N)}}) + Cost_M(\mathbf{s^{(L)}})$, more specifically,

- The shock-keyword vector $\mathbf{s^{(D)}}$ requires $\log(d)$ bits.
- The shock-time vector $\mathbf{s^{(N)}} = \{t_p, t_s, t_w\}$ requires $3 \cdot \log(n)$.
- The shock-location matrix $\mathbf{s^{(L)}}$ requires $|\mathbf{s^{(L)}}| \cdot (\log(d) + \log(l) + \log(n) + c_F)$, where, $|\cdot|$ describes the number of non-zero elements.

Consequently, the model cost of the external shock tensor $\mathcal{S} = \{\mathbf{s}_1, \cdots, \mathbf{s}_k\}$ is $Cost_M(\mathcal{S}) = \log^*(k) + \sum_{i=1}^{k} Cost_M(\mathbf{s}_i)$.

Data Coding Cost. Once we have decided the full parameter set $\mathcal{F}$, we can encode the data $\mathcal{X}$ with a given parameters $\mathcal{F}$: $Cost_C(\mathcal{X}|\mathcal{F}) = \sum_{i,j,t=1}^{d,l,n} \log_2 p^{-1}_{Gauss(\mu,\sigma^2)}(x_{ij}(t) - I_{ij}(t))$,

where, $x_{ij}(t)$is the elements in $\mathcal{X}$, and $I_{ij}(t)$ is the estimated count of infections (i.e., Model 1)[*4].

Data Compression Equation. Consequently, the total code length for $\mathcal{X}$ with respect to a given parameter set $\mathcal{F}$ can be described as follows:

$$\begin{aligned} &Cost_T(\mathcal{X}; \mathcal{F}) = \log^*(d) + \log^*(l) + \log^*(n) \\ &+Cost_M(\mathbf{B_G}) + Cost_M(\mathbf{B_L}) + Cost_M(\mathbf{R_G}) \\ &+Cost_M(\mathbf{R_L}) + Cost_M(\mathcal{S}) + Cost_C(\mathcal{X}|\mathcal{F}) \end{aligned} \quad (2)$$

Thus our next goal is to minimize the above function.

4.2 Multi-layer Optimization

Until now, we have seen how we can measure the goodness of the representation of $\mathcal{X}$, if we are given a candidate parameter set $\mathcal{F}$. The next question is, how to find an optimal solution of the full parameter set: $\mathcal{F} = \{\mathbf{B_G}, \mathbf{B_L}, \mathbf{R_G}, \mathbf{R_L}, \mathcal{S}\}$. As described in Section 3.3, Δ-SPOT consists of multiple parameter sets, each of which explains either the local or global pattern of web search sequence in $\mathcal{X}$. For example, the base and growth effect matrices $\mathbf{B_G}$, $\mathbf{R_G}$ explain the global-level behavior of each keyword search volume, while the matrices $\mathbf{B_L}$, $\mathbf{R_L}$ describes the local-level trends. Also, the extra tensor $\mathcal{S}$ consists of both global and local parameters.

In order to estimate these model parameters, we propose a multi-layer optimization algorithm, to search for the optimal solution in terms of both the global-level and local-level parameters. The idea is that we split parameter set $\mathcal{F}$ into two subsets, i.e., $\mathcal{F}_\mathcal{G}$ and $\mathcal{F}_\mathcal{L}$, each of which corresponds to a global/local-level parameter set, and try to fit the parameter sets separately. Our algorithm

[*1] Here, the parameter values are: $\beta = 5.014 \times 10^{-1}$, $\delta = 4.675 \times 10^{-1}$, $\gamma = 5.211 \times 10^{-1}$, $\eta_0 = 1.605 \times 10^{-1}$, $t_\eta = 343$ (the growth effect starts from time-tick 343).

[*2] Here, $\log^*$ is the universal code length for integers.

[*3] We used 4×8 bits in our setting.

[*4] Here, μ and σ^2 are the mean and variance of the distance between the original and estimated values.

Algorithm 1 Δ-SPOT($\mathcal{X}$)

Input: Tensor $\mathcal{X}$ ($d \times l \times n$)
Output: Full parameters, i.e., $\mathcal{F} = \{\mathbf{B_G}, \mathbf{B_L}, \mathbf{R_G}, \mathbf{R_L}, \mathcal{S}\}$
/* Parameter fitting for global-level sequences */
$\{\mathcal{F}_\mathcal{G}\}$ =GlobalFit ($\mathcal{X}$);
/* Parameter fitting for local-level sequences */
$\{\mathcal{F}_\mathcal{L}\}$ =LocalFit ($\mathcal{X}, \mathcal{F}_\mathcal{G}$);
return $\mathcal{F} = \{\mathcal{F}_\mathcal{G}, \mathcal{F}_\mathcal{L}\}$;

Algorithm 2 GlobalFit($\mathcal{X}$)

Input: Tensor $\mathcal{X}$
Output: Set of global-level parameters $\mathcal{F}_\mathcal{G}$ i.e., $\mathcal{F}_\mathcal{G} = \{\mathbf{B_G}, \mathbf{R_G}, \mathcal{S}\}$
for $i = 1 : d$ **do**
  Create $\bar{\boldsymbol{x}}_i$ from $\mathcal{X}$; /* Global sequence $\bar{\boldsymbol{x}}_i$ of i-th keyword */
  /* Initialize external shocks for keyword i */
  $\mathbf{s}_i = \emptyset$;
  while improving the cost **do**
    $\mathbf{b^{(G)}_i} = \arg\min_{\mathbf{b^{(G)}_i}'} \mathbf{Cost_C}(\bar{\mathbf{x}}_\mathbf{i} | \mathbf{b^{(G)}_i}', \mathbf{r^{(G)}_i}, \mathbf{s_i})$; /* Base */
    $\mathbf{r^{(G)}_i} = \arg\min_{\mathbf{r^{(G)}_i}'} \mathbf{Cost_C}(\bar{\mathbf{x}}_\mathbf{i} | \mathbf{b^{(G)}_i}, \mathbf{r^{(G)}_i}', \mathbf{s_i})$; /* Growth */
    $\mathbf{s}_i = \emptyset$; /* Initialize values */
    /* Find external shocks for keyword i */
    while improving the cost **do**
      $\mathbf{s} = \arg\min_{\mathbf{s}'} Cost_C(\bar{\boldsymbol{x}}_i | \mathbf{b^{(G)}_i}, \mathbf{r^{(G)}_i}, \{\mathbf{s_i} \cup \mathbf{s}'\})$;
      $\mathbf{s}_i = \mathbf{s}_i \cup \mathbf{s}$;
    end while
  end while
  /* Update parameter set of i-th keyword */
  $\mathbf{B_G} = \mathbf{B_G} \cup \mathbf{b^{(G)}_i}$; $\mathbf{R_G} = \mathbf{R_G} \cup \mathbf{r^{(G)}_i}$;
  $\mathcal{S} = \mathcal{S} \cup \mathbf{s}_i$;
end for
return $\mathcal{F}_\mathcal{G} = \{\mathbf{B_G}, \mathbf{R_G}, \mathcal{S}\}$;

consists of the following two phases:

- GlobalFit: find good global-level parameters for $\{\bar{\boldsymbol{x}}_i\}_{i=1}^{d}$, i.e., $\mathcal{F}_\mathcal{G} = \{\mathbf{B_G}, \mathbf{R_G}, \mathcal{S}\}$
- LocalFit: find good local-level parameters: for $\{\boldsymbol{x}_{ij}\}_{i,j=1}^{d,l}$, i.e., $\mathcal{F}_\mathcal{L} = \{\mathbf{B_L}, \mathbf{R_L}, \mathcal{S}\}$

Here, the global sequence of the i-th keyword: $\bar{\boldsymbol{x}}_i$ can be described as the sum of the d local sequences, i.e., $\bar{\boldsymbol{x}}_i(t) = \sum_{j=1}^{l} \boldsymbol{x}_{ij}(t)$. Algorithm 1 shows an overview of Δ-SPOT to find the full set of Δ-SPOT parameters given a tensor $\mathcal{X}$.

4.2.1 Global-level Parameter Fitting

Given a tensor $\mathcal{X}$, our sub-goal is to find the optimal global-level parameter set: $\mathcal{F}_\mathcal{G}$, to minimize the cost function (i.e., Eq. (2)). So how can we fit the parameters (i.e., the base and growth parameters) as well as simultaneously estimate the appropriate number of external shocks? To find a good basic parameter set for $\mathcal{X}$, we have to filter out the external shocks. Also, a good external-shock filter requires a well estimated model. We escape this circular dependency by applying an iterative method that employs external shocks detection and filtering, and model fitting in an alternating way until the cost function reaches a minimum value.

Algorithm. Algorithm 2 is a detailed algorithm of the global-level fitting. Given a tensor $\mathcal{X}$, it creates a set of d global sequences: $\{\bar{\boldsymbol{x}}_i\}_{i=1}^{d}$. It tries to fit the global-level parameter set, as well as find the appropriate number of external-shocks. We use the *Levenberg-Marquardt (LM)* [4] algorithm to minimize the cost function. Note that the extra tensor $\mathcal{S}$ consists of k entries $\{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_k\}$, This algorithm can find only the global-level entry, which consists of (*keyword*, *time*). The local-level entries can be computed by local-level parameter fitting, as shown next in Algorithm 3.

Algorithm 3 LocalFit($\mathcal{X}, \mathbf{B_G}, \mathbf{R_G}, \mathcal{S}$)

Input: (a) Tensor $\mathcal{X}$, (b) global-level parameter set $\mathcal{F}_\mathcal{G}$
Output: Set of local-level parameters, i.e., $\mathcal{F}_\mathcal{L}$
while improving the cost **do**
  /* For each local sequence $\boldsymbol{x}_{ij}$ of i-th keyword in j-th country */
  for $i = 1 : n$ **do**
    for $j = 1 : l$ **do**
      $b^{(L)}{}_{ij} = \arg\min_{b^{(L)}{}'_{ij}} Cost_C(\boldsymbol{x}_{ij} | \mathbf{B_G}, \mathbf{R_G}, b^{(L)}{}'_{ij}, \mathcal{S})$;
      $r^{(L)}{}_{ij} = \arg\min_{r^{(L)}{}'_{ij}} Cost_C(\boldsymbol{x}_{ij} | \mathbf{B_G}, \mathbf{R_G}, r^{(L)}{}'_{ij}, \mathcal{S})$;
    end for
  end for
  for each external shock $\mathbf{s}$ in $\mathcal{S}$ **do**
    Update $\mathbf{s}$ to minimize the cost /* Local participation rate */
  end for
end while
return $\mathcal{F}_\mathcal{L} = \{\mathbf{B_L}, \mathbf{B_G}, \mathcal{S}\}$;

Also, the cost function (2) includes the cost of local-level parameters such as $\mathbf{B_L}$, $\mathbf{R_L}$ but these terms are independent of the global model fitting. Hence, we can simply consider them to be constant.

4.2.2 Local-level Parameter Fitting

Given a set of $d \times l$ local-level sequences, $\{\boldsymbol{x}_{ij}\}_{i,j=1}^{d,l} \in \mathcal{X}$, and a set of global-level parameters, $\mathcal{F}_\mathcal{G}$, our next goal is to fit the individual parameters of each disease in each state, that is, $\mathcal{F}_\mathcal{L} = \{\mathbf{B_L}, \mathcal{S}\}$. We propose an iterative optimization algorithm (see Algorithm 3). Our algorithm searches for the optimal solution with respect to the base local matrix $\mathbf{B_L}$ and the local-level external shocks $\mathcal{S}$, so that the total coding cost is minimized.

Lemma 1 The computation time of Δ-SPOT is $O(dln)$.

Proof 1 To create the global-level sequences from $\mathcal{X}$, the algorithm requires $O(dln)$ time. For global-level parameter fitting, it needs $O(\#iter \cdot k \cdot dn)$ time, where $\#iter$ is the number of iterations, k shows the number of external shocks. Similarly, for the local-level parameter fitting, it needs $O(\#iter \cdot k \cdot dln)$ time to fit the parameters. Note that $\#iter$, k are small constant values that are negligible. Thus, the complexity is $O(dln)$.

5. Experiments

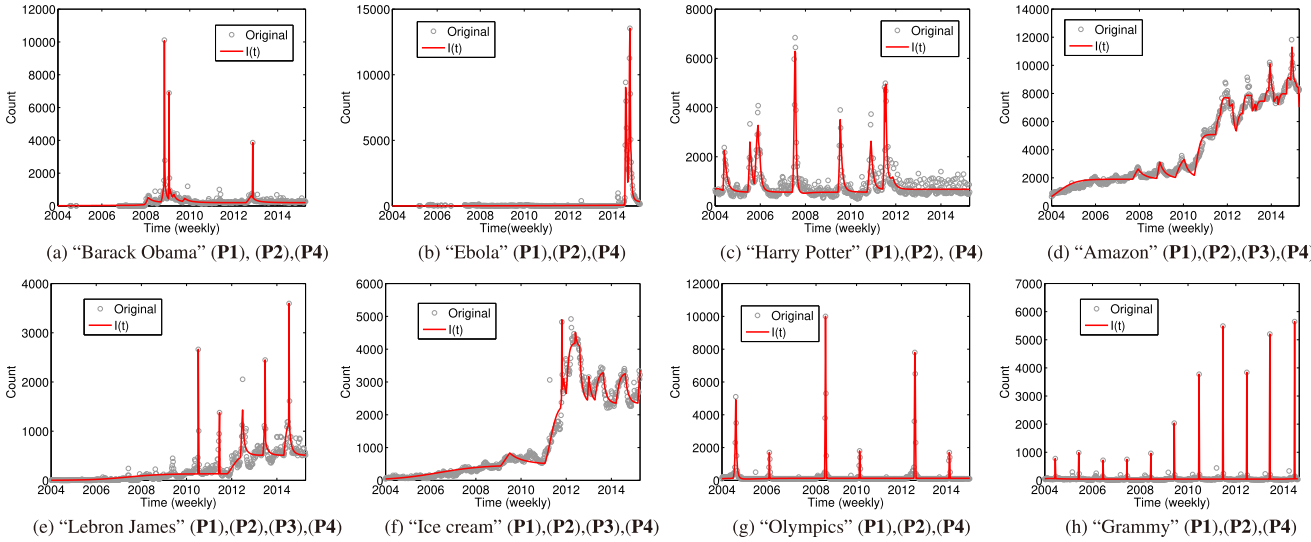
In this section we demonstrate the effectiveness of Δ-SPOT with real dataset. The experiments were designed to answer the following questions:

Q1 *Sense-making*: Can our method help us understand the given input online activities?
Q2 *Accuracy*: How well does our method match the data?
Q3 *Scalability*: How does our method scale in terms of computational time?

Dataset Description. We performed experiments on the following three real datasets.

(1) *GoogleTrends*: This dataset consists of the volume of searches for queries (i.e., keywords) in various topics (i.e., events, celebrities, movies, etc..) on Google[*5] from January

[*5] http://www.google.com/insights/search/



(a) "Barack Obama" (**P1**), (**P2**),(**P4**) (b) "Ebola" (**P1**),(**P2**),(**P4**) (c) "Harry Potter" (**P1**),(**P2**), (**P4**) (d) "Amazon" (**P1**),(**P2**),(**P3**),(**P4**)

(e) "Lebron James" (**P1**),(**P2**),(**P3**),(**P4**) (f) "Ice cream" (**P1**),(**P2**),(**P3**),(**P4**) (g) "Olympics" (**P1**),(**P2**),(**P4**) (h) "Grammy" (**P1**),(**P2**),(**P4**)

Fig. 5 Global fitting results for 8 queries in GoogleTrends dataset of different topics (celebrities, sporting events, awards, movies, etc.)

2004 to January 2015, collected in 232 countries. Each query represents the search volumes that are related to keywords over time (in weekly basis).

(2) *Twitter*: We used more than 7 million Twitter[*6] posts covering an 8-month period from June 2011 to January 2012. We selected the 10,000 most frequently used hashtags.

(3) *MemeTracker*: This dataset covers three months of blog activity from August 1 to October 31 2008[*7], It contains short quoted textual phrases ("memes"), each of which consists of the number of mentions over time. We choose 1,000 phrases in blogs with the highest volume in a 7-day window around their peak volume.

5.1 Sense-making

In this experiment, we demonstrate how effective Δ-SPOT can be in terms of data fitting, external events detection and other important properties. We demonstrate the global fitting results of three datasets:

(1) **Figure 5** shows the results of model fitting on 8 trending keywords/queries in various categories.

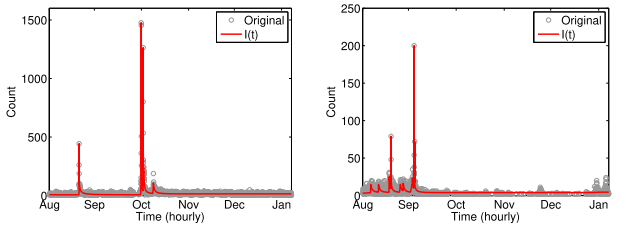
(2) **Figure 6** shows the results of two popular hashtags "#apple" and "#backtoschool".

(3) **Figure 7** shows the results of two phrases ("meme")[*8] from the *MemeTracker* dataset.

In all above figures, we show the original sequences (i.e., black dots) and estimated sequences: $I(t)$ (i.e., red line) in linear-linear scales. Also, we made several important observations, which correspond to the properties mentioned above.

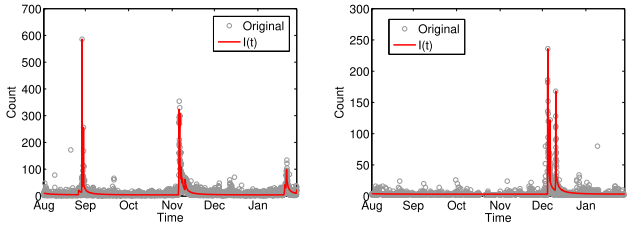
- (**P1**) Base trends and global influence: As shown in Figs. 5, 6 and 7, our proposed model successfully captures long-range non-linear dynamics of user activities, as well as fit the data sequences in high accuracy.

[*6] http://twitter.com/
[*7] http://memetracker.org/
[*8] Meme#3: "yes we can yes we can"
Meme#16: "joe satriani is a great musician but he did not write or have any influence on the song viva la vida we respectfully ask him to accept our assurances of this and wish him well with all future endeavours"



(a) "#apple" (**P1**),(**P2**),(**P4**) (b) "#backtoschool" (**P1**),(**P2**),(**P4**)

Fig. 6 Global fitting results for 2 hashtags in Twitter dataset.



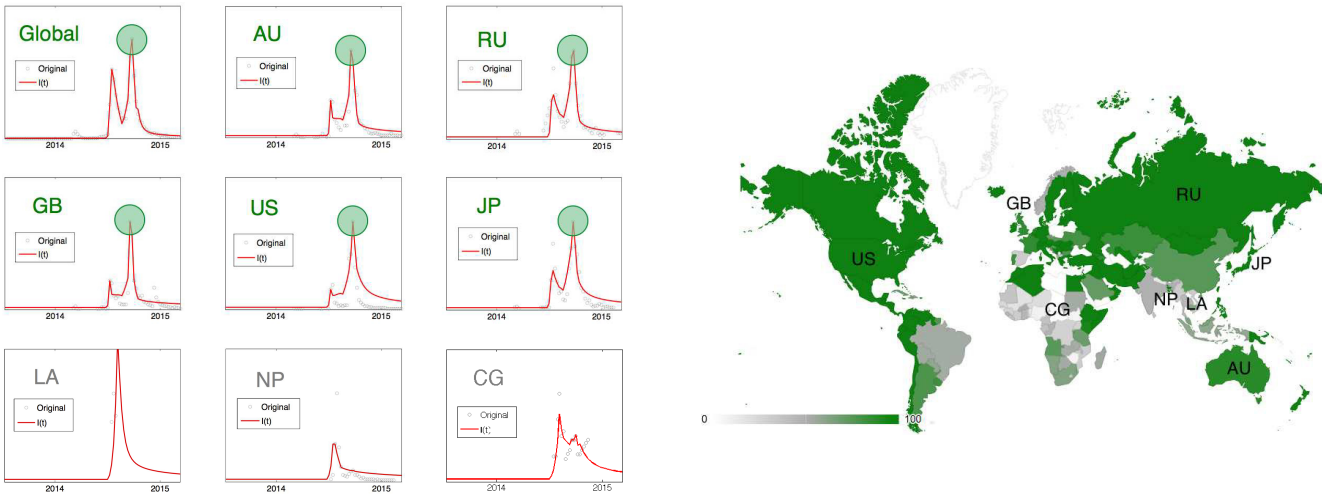
(a) Meme#3 (**P1**),(**P2**),(**P4**) (b) Meme#16 (**P1**),(**P2**),(**P4**)

Fig. 7 Global fitting results for 2 memes in MemeTracker dataset.

- (**P2**) Area specificity: Δ-SPOT can find the local dynamics of each query. For example, **Fig. 8** (a) shows the local fitting results for keyword "Ebola" of *GoogleTrends* dataset; in which, we detected some countries (AU, RU, GB, US, JP) that behave similar to the global trend (i.e., the world reaction to the burst of Ebola Virus in 2014, shown in green circles). Besides, we can also detect several outliers from the countries which have less capacities of network connection (LA, NP, CG).
- (**P3**) Population growth effect: In Fig. 5 (d, e, f), our model can detect the population growth effect, also describe its behavior separately from the external shock effect.
- (**P4**) Cyclic external events: Δ-SPOT can capture important external events relating to the keywords, including the cyclic events.

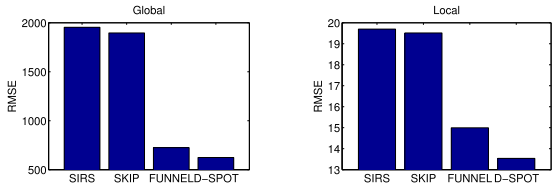
5.2 Accuracy

Next, we discuss the quality of Δ-SPOT in terms of fitting accuracy. We used the fitting result for keyword "Amazon", and



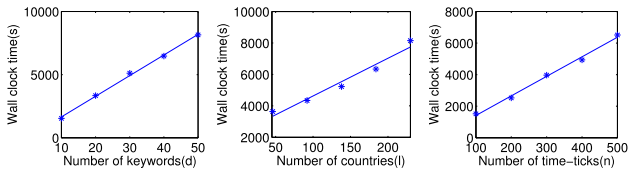
(a) Original/fitted sequences for "Ebola" (b) World-wide reaction

Fig. 8 Local fitting power of Δ-SPOT for the keyword "Ebola" which refers to the Ebola Virus bursting in 2014 (shown in green circles). (a) It can capture the local similar behaviors in Australia (AU), Russia (RU), the U.K. (GB), the U.S. (US) and Japan (JP). It can also capture local outliers in Laos (LA), Nepal (NP) and DR Congo (CG), in comparison to the global trend. And we have a clearer observation in (b) the world map of user reaction to the disease burst in 2014.



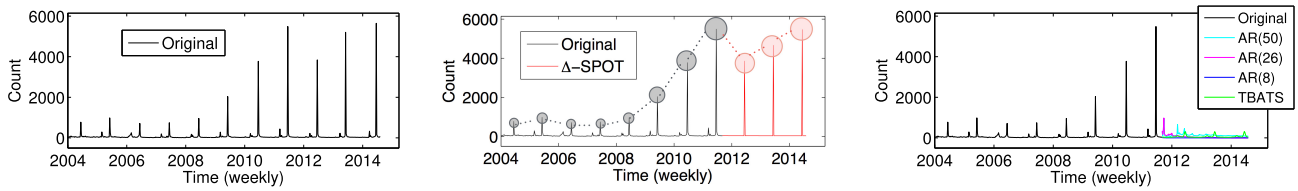
(a) Global fitting (b) Local fitting

Fig. 9 Fitting accuracy (RMSE) for Δ-SPOT (lower is better).



(a) Keywords (d) (b) Countries (l) (c) Duration (n)

Fig. 10 Δ-SPOT scales linearly: wall clock time vs. dataset size ($d \times l \times n$).



(a) Original sequence "Grammy" (b) Forecasted with Δ-SPOT (c) Forecasted with other methods (i.e., AR, TBATS)

Fig. 11 Forecasting result: we train the model parameters using first 400 time-ticks of the sequences and do forecasting the remaining part.

compared Δ-SPOT with the standard SIRS model, SKIPS [19], and FUNNEL [14]. **Figure 9** (a) shows the root-mean-square error (RMSE) between the original and estimated counts of the global sequences $\{\bar{\boldsymbol{x}}_i(t)\}_{i,t}^{d,n}$. Similarly, Fig. 9 (b) shows the results of the local counts $\{\boldsymbol{x}_{ij}(t)\}_{i,j,t}^{d,l,n}$, (i.e., each keyword in each country, at each time-tick). A lower value indicates a better fitting accuracy. Note that the SIRS model cannot capture seasonal dynamics, SKIPS has the ability to capture periodic patterns, while FUNNEL can capture external shocks. Moreover, they are not intended to detect growth effect. As shown in the figures, the SIRS model and SKIPS failed to capture the complicated patterns of data sequences, FUNNEL cannot detect cyclic external events, while our method achieved those properties with high fitting accuracy.

5.3 Scalability

We also evaluated the scalability of Δ-SPOT, and verified the complexity of our method, which we discussed in Lemma 1, in Section 4. **Figure 10** shows the computational cost of Δ-SPOT in terms of the dataset size. We varied the dataset size with respect to (a) keywords d, (b) countries l, and (c) duration n. As shown in Fig. 10, Δ-SPOT is linear with respect to data size.

6. Δ-SPOT at work

As we described in the previous section, Δ-SPOT is capable of analyzing online activities of various categories. Here, we discuss the most important and challenging task of Δ-SPOT, namely, forecasting the future dynamics of co-evolving activities. **Figure 11** shows results of our forecasting in relation to keyword "Grammy". The goal here is to predict the search volume of this keyword in the future. We trained the model parameters by using the 400 time-ticks of the sequence (solid black lines in the figure), and then forecasted the following years (solid red lines). Δ-SPOT can predict the time-tick, the duration and the relative strength of incoming external events, which refer to the annual Grammy Awards, held every February.

We also compared Δ-SPOT with the auto regressive (AR) model, and TBATS model. We applied several regression coefficients: $r = 8, 26, 50$ for AR. In Fig. 11 (a), (b), (c), we show the original sequences, and the forecast results of Δ-SPOT and AR with TBATS, respectively. Our method achieves high forecasting accuracy: we can predict the next three spikes relating to the next three Grammys. Whereas, AR and TBATS failed to forecast future patterns.

7. Conclusion

In this paper, we presented Δ-SPOT, an intuitive model for mining large scale time-evolving online activities. Δ-SPOT demonstrates all the following desirable properties:

(1) It is **effective**: it can detect important hidden events that match the reality.
(2) It is **automatic**: it requires no training set and no domain expertise, thanks to our coding scheme.
(3) It is **scalable**: Δ-SPOT is linear to the data size (i.e., $O(dln)$).
(4) It is **practical**: Δ-SPOT can undertake long-range forecasting and outperforms existing methods.

Acknowledgments This work was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research Number JP15H02705, JP16K12430, JP26730060, JP26280112, and the MIC/SCOPE #162110003.

References

[1] Ginsberg, J., Mohebbi, M., Patel, R., Brammer, L., Smolinski, M. and Brilliant, L.: Detecting influenza epidemics using search engine query data, *Nature*, Vol.457, pp.1012–1014 (2009).
[2] Gruhl, D., Guha, R., Kumar, R., Novak, J. and Tomkins, A.: The predictive power of online chatter, *KDD*, pp.78–87 (2005).
[3] Jain, A., Chang, E.Y. and Wang, Y.-F.: Adaptive stream resource management using kalman filters, *SIGMOD*, pp.11–22 (2004).
[4] Levenberg, K.: A method for the solution of certain non-linear problems in least squares, *Quarterly Journal of Applied Mathmatics*, Vol.II, No.2, pp.164–168 (1944).
[5] Li, L., Liang, C.-J.M., Liu, J., Nath, S., Terzis, A. and Faloutsos, C.: Thermocast: A cyber-physical forecasting model for data centers, *KDD* (2011).
[6] Li, L., McCann, J., Pollard, N. and Faloutsos, C.: Dynammo: Mining and summarization of coevolving sequences with missing values, *KDD* (2009).
[7] Li, L., Prakash, B.A. and Faloutsos, C.: Parsimonious linear fingerprinting for time series, *PVLDB*, Vol.3, No.1, pp.385–396 (2010).
[8] Livera, A.M.D., Hyndman, R.J. and Snyder, R.D.: Forecasting time series with complex seasonal patterns using exponential smoothing, *Journal of the American Statistical Association*, Vol.106, No.496, pp.1513–1527 (2011).
[9] Matsubara, Y., Li, L., Papalexakis, E.E., Lo, D., Sakurai, Y. and Faloutsos, C.: F-trail: Finding patterns in taxi trajectories, *PAKDD*, pp.86–98 (2013).
[10] Matsubara, Y., Sakurai, Y. and Faloutsos, C.: Autoplait: Automatic mining of co-evolving time sequences, *SIGMOD* (2014).
[11] Matsubara, Y., Sakurai, Y. and Faloutsos, C.: The web as a jungle: Non-linear dynamical systems for co-evolving online activities, *WWW*, pp.721–731 (2015).
[12] Matsubara, Y., Sakurai, Y., Faloutsos, C., Iwata, T. and Yoshikawa, M.: Fast mining and forecasting of complex time-stamped events, *KDD*, pp.271–279 (2012).
[13] Matsubara, Y., Sakurai, Y., Prakash, B.A., Li, L. and Faloutsos, C.: Rise and fall patterns of information diffusion: model and implications, *KDD*, pp.6–14 (2012).
[14] Matsubara, Y., Sakurai, Y., van Panhuis, W.G. and Faloutsos, C.: FUNNEL: Automatic mining of spatially coevolving epidemics, *KDD*, pp.105–114 (2014).
[15] Papadimitriou, S., Brockwell, A. and Faloutsos, C.: Adaptive, hands-off stream mining, *VLDB*, pp.560–571 (2003).
[16] Prakash, B.A., Beutel, A., Rosenfeld, R. and Faloutsos, C.: Winner takes all: Competing viruses or ideas on fair-play networks, *WWW*, pp.1037–1046 (2012).
[17] Rakthanmanon, T., Campana, B.J.L., Mueen, A., Batista, G.E.A.P.A., Westover, M.B., Zhu, Q., Zakaria, J. and Keogh, E.J.: Searching and mining trillions of time series subsequences under dynamic time warping, *KDD*, pp.262–270 (2012).
[18] Sakurai, Y., Matsubara, Y. and Faloutsos, C.: Mining and forecasting of big time-series data, *SIGMOD*, pp.919–922 (2015).
[19] Stone, L., Olinky, R. and Huppert, A.: Seasonal dynamics of recurrent epidemics, *Nature*, Vol.446, pp.533–536 (Mar. 2007).
[20] Tao, Y., Faloutsos, C., Papadias, D. and Liu, B.: Prediction and indexing of moving objects with unknown motion patterns, *SIGMOD*, pp.611–622 (2004).
[21] Zoumpatianos, K., Idreos, S. and Palpanas, T.: Indexing for interactive exploration of big data series, *SIGMOD*, pp.1555–1566 (2014).

Thinh Minh Do is a Master course student at Graduate School of Science and Technology, Kumamoto University, Japan. He obtained his B.E. degree from Kumamoto University in 2015. His research interests include web mining and stream processing. He achieved Rakuten Award at WebDB Forum 2015 and Student Travel Award at ACM SIGMOD/PODS 2016.

Yasuko Matsubara is an Assistant Professor in the Department of Computer Science and Electrical Engineering at Kumamoto University, Japan. She obtained her B.S. and M.S. degrees from Ochanomizu University in 2007 and 2009 respectively, and her Ph.D. from Kyoto University in 2012. She was a Visiting Researcher at Carnegie Mellon University during 2011–2012 and 2013–2014. Her research interests include time-series data mining and non-linear dynamic systems.

Yasushi Sakurai is a Professor at Kumamoto University. He obtained his B.E. degree from Doshisha University in 1991, and his M.E. and Ph.D. degrees from Nara Institute of Science and Technology in 1996 and 1999, respectively. In 1998, he joined NTT Laboratories, and became a Senior Research Scientist in 2005. He was a Visiting Researcher at Carnegie Mellon University during 2004–2005. He received two KDD best paper awards in 2008 and 2010. His research interests include time-series analysis, web mining, and sensor data processing.

(Editor in Charge: *Yusuke Miyao*)