Noname manuscript No.
(will be inserted by the editor)

Mining Urban Events from the Tweet Stream through a Probabilistic Mixture Model

Joan Capdevila · Jesús Cerquides · Jordi Torres

the date of receipt and acceptance should be inserted later

Abstract The geographical identification of content in Social Networks have enabled to bridge the gap between online social platforms and the physical world. Although vast amounts of data in such networks are due to breaking news or global occurrences, local events witnessed by users *in situ* are also present in these streams and of great importance for many city entities. Nowadays, unsupervised machine learning techniques, such as Tweet-SCAN, are able to retrospectively detect these local events from tweets. However, these approaches have limited abilities to reason about unseen observations in a principled way due to the lack of a proper probabilistic foundation. Probabilistic models have also been proposed for the task, but their event identification capabilities are far from those of Tweet-SCAN. In this paper, we identify two key factors which, when combined, boost the accuracy of such models. As a first key factor, we notice that the large amount of meaningless social data requires explicitly modeling non-event observations.Therefore, we propose to incorporate a background model that captures spatio-temporal fluctuations of non-event tweets. As a second key factor, we observe that the shortness of tweets hampers the application of traditional topic models. Thus, we integrate event detection and topic modeling, assigning topic proportions to events instead of assigning them to individual tweets. As a result, we propose WARBLE, a new probabilistic model and learning scheme for retrospective event detection that incorporates these two key factors. We evaluate WARBLE in a data set of tweets located in Barcelona during its festivities. The empirical results show that the model outperforms other state-of-the-art techniques in detecting vari-

Joan Capdevila
Universitat Politècnica de Catalunya (UPC), Barcelona Supercomputing Center (BSC)
E-mail: jc@ac.upc.edu

Jesús Cerquides
Artificial Intelligence Research Institute (IIIA), Spanish National Research Council (CSIC)
E-mail: cerquide@iiia.csic.es

Jordi Torres
Universitat Politècnica de Catalunya (UPC), Barcelona Supercomputing Center (BSC)
E-mail: torres@ac.upc.edu

The final publication is available at Springer via http://dx.doi.org/10.1007/s10618-017-0541-y

ous types of events while relying on a principled probabilistic framework that enables to reason under uncertainty.

Keywords Event detection · Social Networks · Probabilistic Models · Variational Inference

1 INTRODUCTION

Social Networks, such as Twitter or Instagram, have turned citizens into social sensors capable of reporting and spreading interesting events straightaway from their mobile device (e.g. Mumbai terrorist attacks (Stelter and Cohen, 2008), Osama Bin Laden raid (Newman, 2011)). Moreover, through the geographical identification of content, i.e. geo-location, these networks have enabled to associate physical locations to some of these events (Zheng, 2012). Local events refer to happenings witnessed by users at some specific time and place (Lee, 2012), and differ from more general event types in the fact that the latter are not generally spatially bounded (Weng and Lee, 2011; Becker et al., 2011). Identifying automatically such events, their temporal and spatial extent, the social structure, etc. has become an interesting research problem with a broad range of applications (Panagiotou et al., 2016).

Some local events, such as music concerts, protests, conferences, etc., are badly covered by traditional media but of a great importance for many social and economic actors. For example, the city council might want to know about events that have happened in its urban area during the past week, month or year in order to plan future events, prepare communication strategies and arrange logistics. Spreading a team of pollsters over the city might be too costly and still incapable of identifying certain types of events (e.g. unscheduled events) or data dimensions (e.g. social relationships). On the contrary, leveraging social network analysis to automatize the detection and summarization of these local events seems a much more plausible approach.

Twitter has become the *de facto* Social Network to perform this event detection task, mainly because the shortness of tweet messages fosters the quick consumption and spreading of information (Atefeh and Khreich, 2015). In this article, we focus on the retrospective detection of local events from the stream of geo-located tweets. Others have already proved that this subset of tweets is sufficient to precisely uncover various types of local events ranging from earthquakes (Sakaki et al., 2010) to social events (Lee and Sumiya, 2010) or traffic jam (Krumm and Horvitz, 2015). Nonetheless, Twitter poses a set of features that makes the task of event detection particular and calls for novel approaches that go beyond standard topic models (Blei, 2012) used in Topic Detection and Tracking (TDT) (Allan et al., 1998) for news articles. Next, we highlight three of these well-known challenges:

rarity. Event-related publications are masked by tones of non-event data such as *memes*, user conversations or *retweet* activities, making it very hard to uncover interesting patterns (Becker et al., 2011).

text-shortness. The length limit in the textual component of tweets hampers the application of standard text models which rely on the co-occurrence of words such as traditional topic models (Hong and Davison, 2010).

variability. The tweeting activity is not flat along a day (it peaks during late night and falls in early morning, i.e. see Fig. 4a), nor over a urban area (it concentrates in the city center and spreads in suburbs, i.e. see Fig. 4b) (Li et al., 2013).

Capdevila et al. (2017) proposed a technique called Tweet-SCAN capable of dealing with rarity and text-shortness, but unable to capture the temporal and spatial variability of tweets. The technique extends DBSCAN (Density-based spatial clustering of applications with noise) (Ester et al., 1996) to cluster tweets as per their spatial, temporal and textual features. DBSCAN, and by extension Tweet-SCAN, intrinsically address rarity since both distinguish between noise points, associated with non-event tweets, and cluster points, related to event tweets. Tweet-SCAN tackles text-shortness by aggregating tweets with the same hashtag or key term and training traditional topic models like HDP (Teh et al., 2006) with them. However, Tweet-SCAN does not address variability, given that DBSCAN-like algorithms use a single constant threshold to distinguish between noise and cluster points.

McInerney and Blei (2014) presented a probabilistic model that also clusters tweets as per their spatial, temporal and textual features. They leverage text-shortness by learning topics from an external news dataset and transferring them to the model. However, this approach does not explicitly address rarity, nor variability, compromising the overall precision since many discovered clusters will not correspond to any existing event, but to groups of similar non-event tweets.

Against this background, we propose WARBLE, a probabilistic model and learning scheme that explicitly addresses all three challenges. To address rarity, our model groups non-event tweets together in a separate background component. The spatio-temporal features of this background component are preset through empirical backgrounds learned from geo-located tweets prior to the period of interest. These spatio-temporal empirical priors also enable to capture varying tweet densities in space and time. In this manner, we can overcome previous models' shortcomings to detect events in areas/periods of low tweeting activity (e.g. suburbs, off-peak hours) likewise in those of high activity (e.g. downtown, peak hours). Furthermore, by learning topics and events simultaneously the proposed method is able to exclusively use the tweet stream, thus dropping the dependence on an external data set.

Contributions from this work [1] are the following:

1. We present a new probabilistic model, the so-called WARBLE model, which explicitly addresses rarity, text-shortness and variability.
2. We propose a variational inference algorithm to approximate the posterior distribution given the data.
3. We show that the WARBLE model outperforms state-of-the-art event detection techniques in a data set made of geo-located tweets in the city of Barcelona during its local festivities.

The rest of the paper is structured as follows. In section 2, we present related work on event detection in social networks with special focus to local events. In

[1] This is an extended version of an unpublished paper that was presented at the ICML Anomaly Detection Workshop 2016 (Capdevila et al., 2016a). The present work also incorporates event summaries, evaluation in terms of BCubed metrics, further details on the model and learning algorithm as well as the release of the WARBLE code and "La Mercè" datasets.

section 3, we introduce the WARBLE model in full detail. The learning scheme for the background model and the variational inference algorithm are described in section 4. In section 5, we show the detection performance of the proposed model in terms of set matching and BCubed metrics and compare against other state-of-the-art techniques. We conclude this work in section 6 by presenting some remarks and future work.

2 RELATED WORK

Event detection in Twitter has been deeply influenced by the Topic Detection and Tracking (TDT) project (Allan et al., 1998). According to this project, an event is "something that happens at specific time and place with consequences" (Panagiotou et al., 2016). Therefore, many event detection approaches have been based on measuring these consequences to uncover the true occurrence. However, consequences can be extremely diverse, ranging from an increase on the number of publications (Becker et al., 2011) to the use of certain language structures (Ritter et al., 2012), presence of posts about specific subjects (Akbari et al., 2016) or about personal and time-specific topics (Li and Cardie, 2014). In this work, we follow the most common approach to event detection in social networks (Becker et al., 2011; Lee, 2012) which assumes that the consequences of an event are translated into an increase of publications in the network.

Initial techniques for event detection have focused on extending existing TDT approaches for text collections to social networks. For example, authors in (Petrović et al., 2010) have proposed a document-pivot model which represents tweets through the traditional term vector model and scales up nearest neighbor search through locality-sensitive hashing. Others in (Long et al., 2011) have identified Twitter-specific features that determine topical words and detect events by clustering co-occurrent topical words over a graph. The frequency domain has also been explored by Weng and Lee (2011), who proposed to construct wavelet signals from words and performed clustering based on the cross correlation between signals. Lately, authors in (Becker et al., 2011) addressed rarity in the tweet stream by post-processing resulting clusters and deciding whether or not they were event-related through a supervised classifier.

None of the above approaches considered geo-location, hampering the association of discovered clusters to local events. One of the first works to take into account geo-located tweets was an earthquake detection and monitoring system based on Kalman filtering (Sakaki et al., 2010). Nonetheless, this system filters earthquake-related tweets beforehand, limiting its capacity to discover events about other subjects. A different approach to circumvent this issue consists in simply comparing the expected tweeting behavior in a spatio-temporal subregion against the actual behavior. For instance, Lee and Sumiya (2010) defined Regions of Interest (RoI) through a clustering-based space partition method and constantly monitored these subregions to detect abnormal behaviors through outlying indicators. Krumm and Horvitz (2015) employed instead a uniform tessellation to partition the space and a detection scheme that compares the predicted number of tweets against the actual number. A shortcoming with both techniques is that finer partitions tend to perform badly for large events that affects several subregions, while coarser partitions perform poorly for small events (Wong and Neill, 2009). Space-Time Scan Statistic (STSS) methods (Kulldorff et al., 2005) were proposed to overcome these problems and they have been applied to detect spatio-temporal events in Twitter (Cheng and Wicks, 2014). However, all these techniques do not explicitly consider text, limiting the capabilities to identify different types of event inside the monitored subregion.

DBSCAN-like techniques, which have been categorized as bottom-up detection approaches in Wong and Neill (2009), have also been considered for this task due to the noise resilience capabilities (Ester et al., 1996). EventRadar (Boettcher and Lee, 2012) proposed to incorporate the original DBSCAN into a processing pipeline with different stages to detect local events from tweets. Authors in (Gomide et al., 2011; Tamura and Ichimura, 2013) proposed to use instead the spatio-temporal extension called ST-DBSCAN (Birant and Kut, 2007) to detect predefined events (precipitation and dengue) from text filtered tweets. Lately, others (Singh, 2015) extended DBSCAN to incorporate text through cosine similarity over term vectors to discover various types of unspecified events. Capdevila et al. (2017) presented Tweet-SCAN which relies on Jensen-Shannon distance over topic distributions. Topics are learned by pooling tweets per *hashtag* and training a HDP topic model (Teh et al., 2006) from these aggregated documents. However, one of the major limitations of DBSCAN-like approaches to event detection from tweets is that they fail to detect events that are not dense enough. In other words, DBSCAN-like techniques cannot capture varying tweet densities along time and space.

Probabilistic models were already considered for the TDT project. For instance, Li et al. (2005) proposed a generative model that incorporates content and time information in a unified framework with latent events for retrospective event detection. Similarly, Pan and Mitra (2011) adapted the Spatial Latent Dirichlet Allocation (SLDA) (Wang and Grimson, 2008) typically used in image segmentation for spatio-temporal event detection on text. The influence of these methods in Twitter can be found in the work by McInerney and Blei (2014). The model was proposed for uncovering newsworthy events from tweets by using of an external news data set, from which topics were transferred from this external dataset. However, these models assigned a latent event to each tweet without distinguishing between event and non-event tweets. This assumption might compromise the overall precision when performing local event detection in Twitter because events are very rare.

This work extends the probabilistic model presented in (McInerney and Blei, 2014) to effectively deal with rarity by considering non-event tweets as first class citizen. These non-event tweets are explicitly modeled through an empirical background which captures the varying tweeting activities along time and space. Moreover, the fact that we learn distinct spatial precision matrices and temporal precision scalars for each event, enables to overcome a major issue of DBSCAN-like algorithms, that is the inability to capture events with different density levels. By simultaneously learning topics and events, we are also able to mitigate the lack of word co-occurrence problem that arise in traditional topic models. Furthermore, in contrast to models that disregard text, we are capable of distinguishing between events that overlap in space-time but are from different topics.

3 PROBABILISTIC MODEL

In this section we explain how the WARBLE model explicitly addresses rarity, variability and text-shortness. In the remaining, $\mathbb{T}_n$ is a random variable which represents the time, geolocation and message for the n-th tweet, and $\mathbb{T} = \{\mathbb{T}_1, \ldots, \mathbb{T}_N\}$ is the whole collection of observed tweets.

3.1 Modeling rarity through Heterogeneous Mixture Models

The model proposed by McInerney and Blei (2014) is a mixture model in which every mixture component shares the same distributional form. Fig. 1a shows the probabilistic graphical model (PGM) (Koller and Friedman, 2009) for McInerney and Blei proposal. They assume the existence of K latent events. The model assigns to each event k a proportion π_k of the tweets. Furthermore, there is a set of parameters β_k which characterizes the probability distribution function (pdf) of the tweets of that event. Furthermore, for each tweet n, they assume the existence of a latent event, encoded in the discrete hidden variable e_n, from which the data for the n-th tweet is generated. Given e_n, the distribution of $\mathbb{T}_n$ is

$$\mathbb{T}_n \sim f\left(\beta_{e_n}\right) \tag{1}$$

where f is the pdf, common for all mixture components. That is, the only difference between two events k and k' is that their parameters β_k and $\beta_{k'}$ are different, but the functional form of f remains the same among components.

The joint probability distribution for McInerney and Blei's model can be expressed as follows,

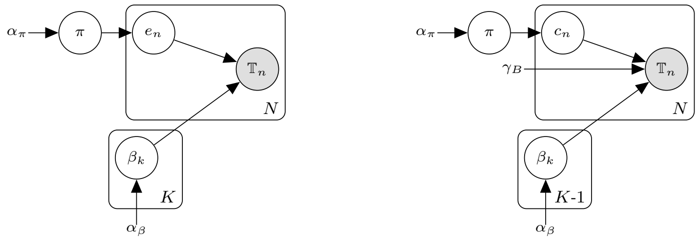
$$p(\mathbb{T}, e, \beta, \pi) = p(\pi|\alpha_\pi) \prod_{n=1}^{N} p(\mathbb{T}_n\,|\beta_{e_n}) p(e_n|\pi) \prod_{k=1}^{K} p(\beta_k|\alpha_\beta) \tag{2}$$

where $p(\pi|\alpha_\pi)$ follows a Dirichlet distribution, $p(e_n|\pi)$ is a Categorical distribution with parameters π and the functional form of $p(\mathbb{T}_n\,|\beta_{e_n})$ is common for all K components. Moreover, the model considers a prior over the event parameters $p(\beta_k|\alpha_\beta)$.

As argued in the introduction, a vast majority of tweets is not event related. We would like to address rarity of event data by introducing a new mixture component, to which we will refer as *background*, which contains those tweets which are not part of any event. In probabilistic terms, it seems clear that the distribution of tweets inside the background component should be widely different from that inside events. McInerney and Blei's model assumes (Eq. 1) that all components follow the same base distribution f, and thus it is unable to deal with the introduction of a background component whose distribution is widely different from that of events.

Accordingly, we propose to generalize McInerney and Blei's model to handle heterogeneous components. To do that, for each component k, we enable a different base function f_k as shown in Eq (3).

$$\mathbb{T}_n \sim f_{e_n}\left(\beta_{e_n}\right). \tag{3}$$



(a) McInerney and Blei's model (b) Simplified WARBLE model

Fig. 1: Probabilistic Graphical Models (PGMs)

Our model fits into the framework proposed by Banfield and Raftery (1993). To the best of our knowledge no application of that framework to event modeling has been reported.

The WARBLE model depicted in Fig. 1b is the PGM representation for an heterogeneous mixture model of tweets in which the K-th component (the background) follows a different statistical distribution. This component corresponds to the background and is represented through a set of parameters γ_B. Moreover, the latent variables are now symbolized through c_n to denote that a tweet might be generated by event components ($c_n < K$) or by background ($c_n = K$).

The joint probability distribution for Fig. 1b can be written as,

$$p(\mathbb{T}, c, \beta, \pi) = p(\pi|\alpha_\pi) \prod_{n=1}^{N} p_{c_n}(\mathbb{T}_n \,|\beta_{c_n}, \gamma_B) p(c_n|\pi) \prod_{k=1}^{K-1} p(\beta_k|\alpha_\beta) \tag{4}$$

where now the tweet distribution depends on the component assignment, $p_{c_n}(\mathbb{T}_n \mid \beta_{c_n}, \gamma_B)$. Moreover, we observe that the background component does not consider a prior over its parameters. The next section provides additional details on how we model the distribution of the background component.

3.2 Modeling variability through a spatio-temporal background

Geo-located social data such as tweets tends to be unevenly distributed through space and time. For example, it is known that users are more likely to tweet during late evening and from highly populated regions (Li et al., 2013). Because of this, we foresee the need to explicitly take this variability into account in order to identify events at peak hours as well as during valleys. This challenge has been deeply studied in classical sensor networks where the spatial scan statistic has been extended to consider non-homogeneous Poisson process as the baseline process (Kulldorff, 1997). It occurs in spatial clustering of trees in forestry, identifying clusters of a particular kind of star in astronomy or geographical clustering of disease in epidemics.

The WARBLE model proposed in Fig. 1b enables to consider a density varying distribution with parameters γ_B for the background component. Here, we propose to model this background through two independent histogram distributions with parameters T_B and L_B, respectively.

The temporal histogram distribution can be represented through a piecewise-continuous function which takes constant values $(T_{B_1}, T_{B_2}, ...T_{B_{I_T}})$ over the I_T contiguous intervals in the variable domain. For example, Fig. 2 shows the 1D-histogram distribution in the temporal range from t_{min} to t_{max}, in which there are I_T intervals of length b. Moreover, we must note that the piecewise function has to be normalized to sum 1 in order to fulfill the properties of probability distributions.

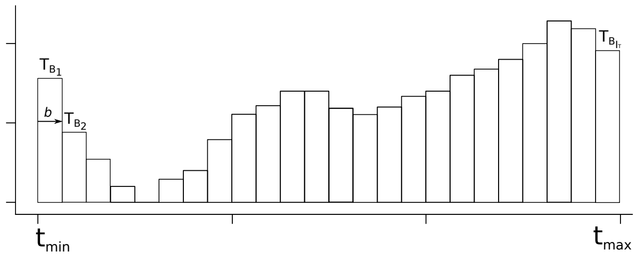


Fig. 2: Temporal histogram distribution $1d\text{-}Hist(.)$

Similarly, the spatial background is modeled through a 2D-histogram distribution over the geographical space, which is represented in a Cartesian coordinate system. The 2d-piecewise-continuous function is expressed through I_L constant values $(L_{B_1}, L_{B_2}, ...L_{B_{I_L}})$ in a grid of squares with size b x b each.

Through these histogram distributions, the WARBLE model can consider different spatio-temporal backgrounds which can be learned from tweets as we will see in section 4.1.

3.3 The complete WARBLE model

We present here the complete WARBLE model to perform event detection from tweets. The probabilistic graphical model in Fig. 3 provides a more detailed version of the model depicted in Fig. 1b.

In the complete WARBLE model tweets $\mathbb{T}_n$ are now represented by their temporal t_n, spatial l_n and textual $w_{n,.}$ features. The parameters β_k for the k-th event comprise the set of variables $\beta_k = \{\tau_k, \lambda_k, \mu_k, \Delta_k, \theta_k\}$. As for the hyperparameters, α_β in Fig. 1b corresponds to the set of hyperparameters m_τ, β_τ, a_λ, b_λ, m_μ, β_μ, ν_Δ, W_Δ, α_θ in Fig. 3. Finally, the hyperparameter of the background component γ_B in Fig. 1b is composed of the hyperparameters for the temporal (T_B) and spatial (L_B) features in Fig. 3. Furthermore, in this detailed model we add two additional variables $\phi = \{\phi_1, \ldots, \phi_T\}$, where ϕ_t encodes parameters of the distribution over words of the t-th topic and θ_K which encodes the distribution over topics for the background component. We also add an additional hyperparameter α_ϕ.

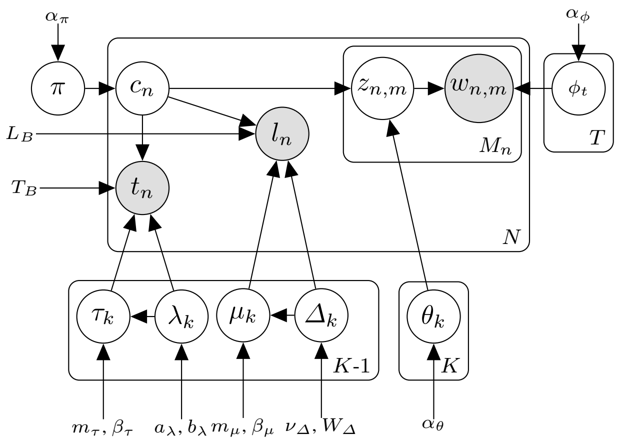


Fig. 3: The WARBLE model in detail

In Eq. (5) we provide the joint probability distribution, which fully describes the WARBLE model in probabilistic terms.

$$p(\mathbb{T}, c, \beta, \pi, \phi, \theta_K) = p(\pi|\alpha_\pi) \prod_{n=1}^{N} p_{c_n}(\mathbb{T}_n\,|\beta_{c_n}, \gamma_B)p(c_n|\pi)$$
$$\prod_{k=1}^{K-1} p(\beta_k|\alpha_\beta)p(\phi|\alpha_\phi)p(\theta_K|\alpha_\beta) \tag{5}$$

In the remaining we specify each of the factors in the right hand side of Eq. (5). As explained above $p(\pi|\alpha_\pi)$ follows a Dirichlet distribution, that is $p(\pi|\alpha_\pi) = Dir(\pi|\alpha_\pi)$. As usual in probabilistic topic models (Blei, 2012), $p(\phi|\alpha_\phi) = \prod_{t=1}^{T} Dir(\phi_t|\alpha_\phi)$ is the product of T Dirichlet distributions with hyperparameter α_ϕ.

As for the tweet probability distribution $p_{c_n}(\mathbb{T}_n\,|\beta_{c_n}, \gamma_B)$, we have that

$$p_{c_n}(\mathbb{T}_n\,|\beta_{c_n}, \gamma_B) = p_{c_n}(t_n|\tau_{c_n}, \lambda_{c_n}, T_B) \cdot p_{c_n}(l_n|\mu_{c_n}, \Delta_{c_n}, L_B)$$
$$p(w_{n,.}|\theta_{c_n}, \phi) \tag{6}$$

Here, the posting time t_n of event-related tweets arises from a Normal distribution $N(.)$ with unknown mean τ_{c_n} and precision λ_{c_n}, and that of non-event tweets is generated by a 1D histogram distribution $Hist(.)$ with parameter T_B, formally

$$p_{c_n}(t_n|\tau_{c_n}, \lambda_{c_n}, T_B) = \begin{cases} Hist(t_n|T_B) & \text{, if } c_n = K \\ N(t_n|\tau_{c_n}, \lambda_{c_n}) & \text{, otherwise.} \end{cases} \tag{7}$$

Similarly, the geographical locations l_n of event-related tweets comes from a multivariate Normal distribution with unknown mean μ_{c_n} and precision Δ_{c_n} and that of non-event tweets is generated by a 2D histogram distribution $Hist(.)$ with parameter L_B:

$$p_{c_n}(l_n|\mu_{c_n}, \Delta_{c_n}, L_B) = \begin{cases} Hist(l_n|L_B) & \text{, if } c_n = K \\ N(l_n|\mu_{c_n}, \Delta_{c_n}) & \text{, otherwise.} \end{cases} \tag{8}$$

Regarding textual features, in WARBLE the m-th word of the n-th tweet is generated as follows. First a topic $z_{n,m}$ is drawn from a Categorical distribution over topics with parameter θ_{c_n}. Then, the word $w_{n,m}$ is sampled from the assigned topic distribution over words with parameter $\phi_{z_{n,m}}$. Formally,

$$p(w_{n,.}|\theta_{c_n}, \phi) = \prod_{m=1}^{M_n} Cat(z_{n,m}|\theta_{c_n})\, Cat(w_{n,m}|\phi_{z_{n,m}}). \tag{9}$$

The prior over event component parameters $p(\beta_k|\alpha_\beta)$ is

$$\begin{aligned} p(\beta_k|\alpha_\beta) =& N(\mu_k|m_\mu, \beta_\mu \Delta_k) W(\Delta_k|\nu_\Delta, W_\Delta) \\ & N(\tau_k|m_\tau, \beta_\tau \lambda_k) G(\lambda_k|a_\lambda, b_\lambda) \\ & Dir(\theta_k|\alpha_\theta) \end{aligned} \tag{10}$$

where the unknown means and precisions are drawn from a Normal-Gamma $N(.)$-$G(.)$ and a Normal-Wishart $N(.)$-$W(.)$. The Dirichlet distribution $Dir(.)$ with hyperparameters α_θ is considered as conjugate prior for the Categorical distribution over topics θ_{c_n}. Similarly, the topic distribution of the background component is also a Dirichlet, $p(\theta_K|\alpha_\beta) = Dir(\theta_K|\alpha_\phi)$, completing the specification of the joint probability distribution.

3.4 Modeling text-shortness through event specific topic proportions

Finally, we note that the detailed WARBLE model presented above integrates clustering and topic modeling, which has lately been found very promising in modeling short and sparse text (Hong and Davison, 2010; Quan et al., 2015).

Following this approach, tweets are clustered into different components c_n as per its temporal t_n, spatial l_n and textual $w_{n,.}$ features, aggregating short text messages into longer pseudo-documents. In our model, these pseudo-documents correspond to the mixture components (events or background).

In contrast to traditional topic modeling where distributions over topics are document-specific (Blei et al., 2003), we here assume that topics $z_{n,m}$ are drawn from component-specific distributions θ_k. This enables to directly obtain topics that are event-related or background-related, providing an interesting approach for automatic event summarization (Long et al., 2011).

4 LEARNING FROM DATA

In this section we describe how to use the WARBLE model to identify a set of events in a region during a period of interest. The procedure assumes the availability of a recorded dataset of tweets from that region and follows two steps. First, we use the tweets previous to the start of the period of interest to derive a background model. Then, we use the tweets recorded during the period of interest to find the most probable assignment of tweets to mixture components.

4.1 Learning the background model

To learn the spatio-temporal background from tweets, we propose to collect tweets previous to the period of interest and within the same region in order to add a sense of typicality to the model.

From the collected tweets, the temporal background is built by first computing the daily histogram with I_T bins. Then, the daily histogram is smoothed by means of a low pass Fourier filter in order to remove high frequency components. The cut-off frequency f_c determines the smoothness of the resulting signal. The normalized and smoothed histogram provides the parameters for the temporal background $T_{B_1}, T_{B_2}, ...T_{B_{I_T}}$.

The spatial background is build following the same procedure. However, geographical location has to be first projected into a Cartesian coordinate system in order to consider locations in a 2-d Euclidean space. The spatial range limits can be determined from the most southwestern and northeastern points. We consider now a two dimensional Gaussian filter with a given variance σ. The resulting 2D-histogram provides the parameter for the spatial background $L_{B_1}, L_{B_2}, ...L_{B_{I_L}}$.

We suggest to set the number of bins for the temporal and spatial histograms as well as the cut-off frequency and variance empirically. Future work will examine how to automatically adjust these parameters.

4.2 Assigning tweets to mixture components

We are interested in finding the most probable assignment of tweets to mixture components, given the data at hand, that is finding c^*

$$c^* = \underset{c}{\operatorname{argmax}}\, p(c|l, t, w; \Gamma) \tag{11}$$

where Γ stands for the model hyperparameters L_B, T_B, α_π, α_θ, α_ϕ, m_τ, β_τ, a_λ, b_λ, m_μ, β_μ, ν_Δ and W_Δ. Exactly assessing c^* is computationally intractable for the WARBLE model. Therefore, we propose to

1. Use mean-field variational Bayesian inference (Fox and Roberts, 2012; Jordan et al., 1999) to approximate $p(X|D; \Gamma)$ (where X stands for the set of random variables containing c, z, π, τ, λ, μ, Δ, θ and ϕ, and D stands for our data, namely l, t, and w) by a distribution $q(X; \eta)$ (where η stands for the variational parameters to be detailed later).
2. Assess c^* from the approximation, that is

$$c^* = \underset{c}{\operatorname{argmax}}\, q(c; \eta) = \underset{c}{\operatorname{argmax}} \int_{X-c} q(X; \eta). \tag{12}$$

In the following we provide detail on each of these two points.

4.2.1 Mean-Field Variational Bayesian inference

Our mean-field variational inference algorithm relies on minimizing the Kullback-Leibler (KL) divergence between $p(X|D;\Gamma)$ and a distribution $q(X;\eta)$ which factorizes as

$$q(X;\eta) = q(\pi)\prod_{t=1}^{T} q(\phi_t)\prod_{n=1}^{N} q(c_n)\prod_{m=1}^{M_n} q(z_{n,m})$$
$$q(\theta_K)\prod_{k=1}^{K-1} q(\tau_k)q(\lambda_k)q(\mu_k)q(\Delta_k)q(\theta_k). \tag{13}$$

The KL divergence is minimized through an iterative coordinate-descent scheme until convergence is reached. Thus, the factors in Eq. (13) are sequentially updated, one factor at a time. The mean-field variational update for the factor corresponding to a random variable x whatsoever is

$$q(x) \propto \exp\left(\int_{X-x} q(X;\eta)\log p(X,D;\Gamma)\right) \tag{14}$$

where $\log p(X,D;\Gamma)$ is the logarithm of the join probability distribution for the WARBLE model defined in Eq. (5) . After all variables have been updated the KL divergence is compared with that of the previous iteration. In case convergence has not been reached yet, another round of updates is started.

We notice that due to the introduction of the background distributions, the model is not conjugate-exponential (Fox and Roberts, 2012; Ghahramani and Beal, 2001). Thus, the updates in Eq. (14) need to be manually derived for each variable. To exemplify the derivations, we include here the development of the most complex update, that of the assignment variable c_n. Since our distribution follows the Bayesian network in Fig. 3, Eq. (14) can be simplified to

$$q(c_n) \propto \exp\left(\int_{Z} q(Z)\log p(c_n,Z,D;\Gamma)\right) \tag{15}$$

where Z is the set of variables in the Markov blanket of c_n, which are π, t_n, τ, λ, l_n, μ, Δ, $z_{n,\cdot}$ and θ.

Given that the right side of Eq. (15) is proportional to the approximate distribution $q(c_n)$, we can disregard terms that do not depend on c_n and express the remaining as a product,

$$q(c_n) \propto f_{\text{prior}}(c_n)\cdot f_{\text{time}}(c_n)\cdot f_{\text{loc}}(c_n)\cdot\prod_{m=1}^{M_n} f_{m\text{-word}}(c_n) \tag{16}$$

where

$$
\begin{aligned}
f_{\text{prior}}(c_n) &= \exp\left(\int_{\pi} q(\pi)\log p(c_n|\pi)\right)\\
f_{\text{time}}(c_n) &= \exp\left(\int_{\tau_{c_n},\lambda_{c_n}} q(\tau_{c_n})q(\lambda_{c_n})\log p(t_n|\tau_{c_n},\lambda_{c_n})\right)\\
f_{\text{loc}}(c_n) &= \exp\left(\int_{\mu_{c_n},\Delta_{c_n}} q(\mu_{c_n})q(\Delta_{c_n})\log p(l_n|\mu_{c_n},\Delta_{c_n})\right)\\
f_{m\text{-word}}(c_n) &= \exp\left(\int_{\theta_{c_n},z_{n,m}} q(\theta_{c_n})q(z_{n,m})\log p(z_{n,m}|\theta_{c_n})\right).
\end{aligned}
\tag{17}
$$

We observe that there are four factors, one for the mixture proportions and one for each tweet feature (posting time, geographical location and text message).

Since c_n is a discrete variable, $q(c_n)$ fits in the functional form of a Categorical distribution with variational parameter c'_n, defined as the normalization of $\tilde{c}'_{nk}$,

$$c'_{nk} = \frac{\tilde{c}'_{nk}}{\sum_{k=1}^{K}\tilde{c}'_{nk}} \tag{18}$$

where $\tilde{c}'_{nk}$ can be obtained from Eq. (16):

$$\tilde{c}'_{nk} = f_{\text{prior}}(k)\cdot f_{\text{time}}(k)\cdot f_{\text{loc}}(k)\cdot \prod_{m=1}^{M_n} f_{m\text{-word}}(k). \tag{19}$$

Note that the background component takes no part in f_{prior} and $f_{m\text{-word}}$, whose expressions can hence be derived following a standard procedure. Thus, we omitted them next.

However, the introduction of a background model entails differences in the spatio-temporal factors f_{loc} and f_{time}, since the background component ($k = K$) follows a different distribution function. Considering the pdf in Eq. (7), the temporal factor can be defined as follows,

$$f_{\text{time}}(k) = \begin{cases} Hist(t_n|T_B) & k = K\\ \exp\left(\int_{\tau_k,\lambda_k} q(\tau_k)q(\lambda_k)\log N(t_n|\tau_k,\lambda_k)\right) & \text{otherwise}\end{cases} \tag{20}$$

and from Eq. (8), the spatial factor is,

$$f_{\text{loc}}(k) = \begin{cases} Hist(l_n|L_B) & k = K\\ \exp\left(\int_{\mu_k,\Delta_k} q(\mu_k)q(\Delta_k)\log N(l_n|\mu_k,\Delta_k)\right) & \text{otherwise}\end{cases} \tag{21}$$

where in each equation the event components are computed from the corresponding Normal distributions and the background component from the Histogram distribution.

Nonetheless, to find a closed-form expression for Eq. (20) we need to derive the approximated distributions for $q(\tau_k)$ and $q(\lambda_k)$. We provide a summary of the functional forms for each variational distribution $q(x)$ in Table 1. Full details on the updates can be found in a technical report (Capdevila et al., 2016b).

Table 1: Functional forms for $q(X)$

$q(x)$	Functional form
$q(\pi)$	$Dir(\pi\|\pi'_k)$
$q(c_n)$	$Cat(c_n\|c'_{nk})$
$q(z_{n,m})$	$Cat(z_{n,m}\|z'_{n,m,t})$
$q(\phi_t)$	$Dir(\phi_t\|\phi'_t)$
$q(\tau_k)$	$N(\tau_k\|m_{\tau_k}, \beta'_{\tau_k}\frac{a'_\lambda}{b'_\lambda})$
$q(\lambda_k)$	$G(\lambda_k\|a'_\lambda, b'_\lambda)$
$q(\mu_k)$	$N(\mu_k\|\mu'_k, \beta'_{\mu_k}\nu' W')$
$q(\Delta_k)$	$W(\Delta_k\|\nu', W')$
$q(\theta_k)$	$Dir(\theta_k\|\theta'_k)$

4.2.2 Using the variational approximation to assign tweets to mixture components

Recall that our objective was to find the most likely assignment of tweets to mixture components using the variational approximation to the posterior shown in Eq. (12). Note that we can take benefit from the fact that $q(X)$ factorizes as shown in Eq. (13) to assess the mixture component for each tweet independently. Thus, the n-th tweet will be assigned to the mixture component which maximizes the Categorical distribution $q(c_n; c'_n)$, that is,

$$c_n^* = \underset{c_n}{\operatorname{argmax}}\, q(c_n; c'_n) = \underset{k}{\operatorname{argmax}}\, c'_{n,k}. \tag{22}$$

5 EXPERIMENTS

In this section, we present the experimental dataset, the evaluation metrics, the WARBLE settings for these experiments, the detection performance of WARBLE and comparative results against other state-of-the-art techniques. The code to reproduce all the experiments can be found in this repository[1].

5.1 Dataset description: “La Mercè 2014”

The availability of datasets for local event detection in Twitter is very limited, hampering the advance of the research field. Because of this, we have crawled and published our own dataset from geo-located in the city of Barcelona during its local festivities on the 24th of September 2014, referred as “La Mercè 2014”. Local events in this set of tweets were tagged by local experts helped with the official calendar of the festivities[2].

[1] https://github.com/jcapde/WARBLE
[2] https://github.com/jcapde/Twitter-DS/tree/master/MERCE/2014

The data set is composed of 2173 tweets out of which 202 belong to 6 distinct real-world events. “La Mercè 2014” events on the 24th of September consisted of a music concert at *Bogatell* beach area, human towers exhibition at *Plaça Sant Jaume*, open day at *MACBA* museum, a food market at *Parc de la Ciutadella*, a wine tasting fair at *Arc de Triomf* and fireworks near *Plaça d’Espanya*. Moreover, experts identified a 7th event which arose in *Bogatell* area during the afternoon as a result of several users reviving the earlier concert.

Tweets were processed beforehand as follows. The posting times were transformed into ordered scalar values by considering *24-09-2014 00:00:00* time-stamp as the reference value. The geographical coordinates, a.k.a latitude and longitude, were transformed into UTM (Universal Transverse Mercator) to work with them as in the euclidean space. Textual messages were cleaned by removing URLs, numbers, emoticons and other special characters. Stopwords in Catalan, Spanish and English were also removed from tweets and all words are converted into lower case.

5.2 Evaluation metrics

The assessment is performed in terms of extrinsic clustering evaluation (Amigó et al., 2009). More specifically, we use common metrics in event detection based on set matching such as purity, inverse purity and F-measure (Yang et al., 1998), but we also propose to consider more robust clustering figures such as the BCubed family (Bagga and Baldwin, 1998).

BCubed metrics, known as BCubed precision, recall and F-measure, defines correctness within a pair of points p and p' as,

$$correctness(p, p') = \begin{cases} 1 & L(p) = L(p') \Longleftrightarrow C(p) = C(p') \\ 0 & otherwise \end{cases} \tag{23}$$

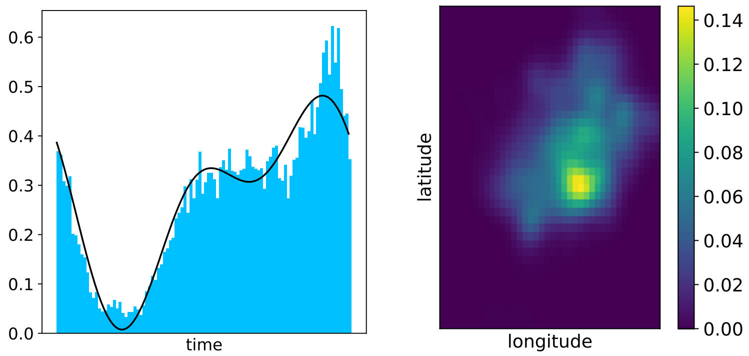
where $L(p)$ corresponds to the label of point p and $C(p)$, to its cluster. Therefore, correctness is one i.f.f. the labels of two points match as well as their clusters. BCubed metrics satisfy desideratum which are not accomplished by set matching metrics (Amigó et al., 2009). For event detection, an interesting properties satisfied by BCubed metrics is the so-called *rag bag*. A metric satisfying rag bag will prefer clusterings in which all “miscellaneous” observations are grouped together into a diverse cluster.

Both family metrics define F-measure to avoid trivial solutions on purity(precision) and inverse purity(recall). Purity or BCubed precision is trivially maximum when each tweet is assigned to a different event and inverse purity or BCubed recall is highest when all tweets are set to the same unique event, respectively (Amigó et al., 2009). Therefore, F-measure, the harmonic mean of both metrics, is proposed to avoid these trivial solutions and become a proper evaluation metric.

5.3 WARBLE settings for “La Mercè 2014”

In this section, we detail the parameters of the WARBLE model as well as the spatio-temporal backgrounds for “La Mercè 2014”. The WARBLE model presented in Section 3 contains several parameters and hyperparameters. Although their optimization is out of the scope of this paper, we have not experimented substantial differences in the results when varying them. The number of components K is set to 8 so that the model is able to capture the 7 events occurring. Following Capdevila et al. (2017), we set the number of topics T to 30.

In addition to “La Mercè 2014” dataset, we also consider tweets previous to the period of interest in order to learn the spatio-temporal backgrounds T_B and L_B as explained in Section 4.1. In particular, we collected tweets from the 20th to the 23th of September 2014 to build the following backgrounds.



(a) Temporal background (b) Spatial background

Fig. 4: Spatio-temporal backgrounds

Fig. 4a shows the daily histogram of tweets in which we observe a valley during the early morning and a peak at night, indicating low and high tweeting activity during these hours, respectively. The 1-d histogram has been computed with $I_T = 100$ bins. Fig. 4a also contains the smoothed histogram distribution (black line) that is used to set the temporal background parameters $T_{B_1}, T_{B_2}, ...T_{B_{I_T}}$.

Fig. 4b is the smoothed histogram for all tweet locations, which give us the parameters for the spatial background $L_{B_1}, L_{B_2}, ...L_{B_{I_L}}$. The 2-d histogram has been computed with $I_L = 1600$ bins. We observe that the most likely areas in the filtered histogram (in bright yellow) correspond to highly dense regions of Barcelona like the city center, while city surroundings are colored in blue indicating lower density of tweets.

We note that the above backgrounds are in accordance with spatio-temporal behaviors founds in other studies (Li et al., 2013).

5.4 Results

First, we assess WARBLE in “La Mercè 2014” dataset through recall figures for each labeled event. Then, we compare its F-measure performance against state-of-the-art techniques such as McInerney & Blei model (McInerney and Blei, 2014) and Tweet-SCAN (Capdevila et al., 2017).

5.4.1 Assessment of WARBLE *in "La Mercè 2014"*

Table 2 summarizes the assessment of WARBLE in "La Mercè 2014" dataset. For each event, set matching recall provides the fraction of relevant tweets that are correctly identified and BCubed recall, shown in parentheses, provides the average correctness. Despite their intrinsic differences, both recall figures show very similar results. We observe that larger events (# tweets), such as concert and fireworks, are correctly identified (high recall) while smaller ones, like museums open day or human towers exhibition, are harder to detect.

However, we notice that the food market and wine tasting exposition could not be discovered at all. We argue that this is because both were all-day events and had fewer tweets in comparison to the rest. Future work could explore to treat all-day events differently, for instance introducing priors for these events with greater temporal variance.

Finally, the resulting mean coordinates (lat, long) and times from the probabilistic model are also coherent with "La Mercè" schedule.

Table 2: Recall figures and spatio-temporal features per event

Event	# tweets	Recall (BCubed)	Location (lat;long)	Time (hh:mm:ss)
Concert	27/28	0.96 (0.93)	41.3931 ± 0.0014; 2.2058 ± 0.0018	02:32:40 ± 0:11:32
Human towers	11/20	0.55 (0.36)	41.3834 ± 0.0013; 2.1775 ± 0.0016	12:46:56 ± 0:08:40
Concert revival	26/30	0.86 (0.76)	41.3926 ± 0.0012; 2.2056 ± 0.0017	13:44:19 ± 0:10:17
Museums open day	18/25	0.72 (0.56)	41.3836 ± 0.0012; 2.1716 ± 0.0044	18:18:33 ± 0:08:27
Fireworks	62/65	0.95 (0.91)	41.3734 ± 0.0015; 2.1496 ± 0.0022	22:11:10 ± 0:06:18

The probabilistic model, apart from spatio-temporal information, also provides information about which topics are linked to each event, enabling automatic event summarization. Topic distributions plotted in Fig. 5, show that each event is mainly about one topic, except for the last one which corresponds to background ($k = K$). Therefore, there are two events whose main topic is number 17, one event for topic 24, another for topic 5 and one last event which is mainly about topic 14.

The content of each topic can be taken out of the corresponding word distributions. Table 3 shows the most probable words for each topic, enabling to understand topics and events. For example, Topic 17 refers to music since words *concert*, *txarango* (local band) and *manel* (local band) are very likely. We have already seen that this topic was linked to two resulting events in Fig. 5 which we can associated with the music concert at *Bogatell* beach area and the revival on the afternoon. We also note that top words in each topic usually refer to the event location, which can be explained from the fact that most tweet messages explicitly mention the place.

5.4.2 Evaluation against state-of-the-art

In what follows, we compare WARBLE from section 3 against other event detection techniques. In particular, we will compare the performance of:

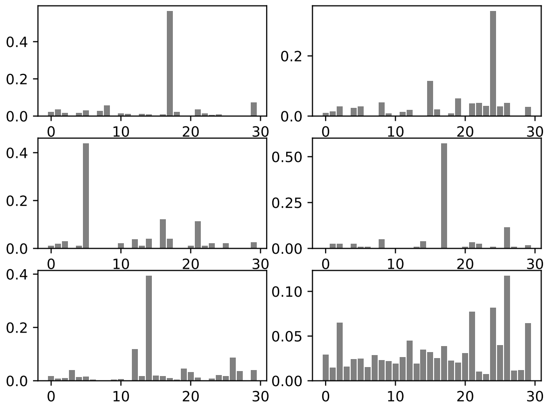


Fig. 5: Topic distributions per event θ_k

Table 3: Most probable words per topic from ϕ_t. English translations in *italics*.

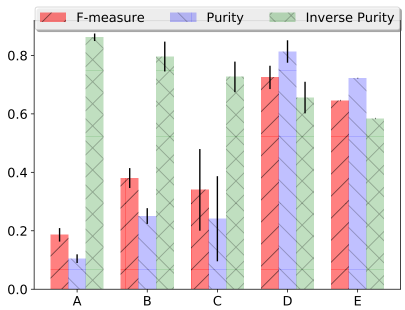
Topic 5	Topic 14	Topic 17	Topic 24	*Topic 26*
museu *museum*	piromusical *fireworks*	platja *beach*	plaça *square*	im *I'm*
macba *MACBA*	plaça *square*	bogatell *Bogatell*	dia *day*	q *that*
contemporani *contemporary*	despanya *from Spain*	txarango *Txarango*	jaume *Jaume*	gran *big*
fan *do*	font *fountain*	concert *concert*	catalunya *Catalonia*	mercé *Mercé*
veient *looking*	poder *power*	manel *Manel*	day *day*	hoy *today*

(A) McInerney & Blei model (McInerney and Blei, 2014), which does not consider background and does not perform simultaneous topic-event learning.
(B) The WARBLE model without simultaneous topic-event learning.
(C) The WARBLE model without modeling background.
(D) The complete WARBLE model.
(E) Tweet-SCAN with $\epsilon_1 = 250m$, $\epsilon_2 = 3600s$, $\epsilon_3 = 0.9$, $\mu = 0.5$, $MinPts = 7$.

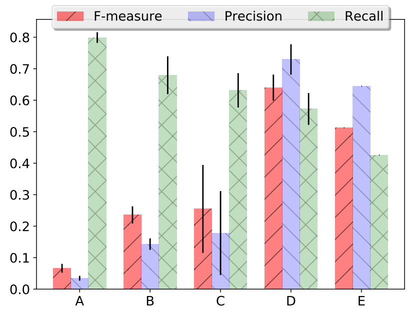
For those models that do not perform simultaneous topic-event learning, the Latent Dirichlet Allocation model (Blei et al., 2003) is separately trained with tweets aggregated by key terms as proposed in (Hong and Davison, 2010).

Fig. 6a shows the results for each event detection model introduced earlier in terms of set matching metrics. Results show that WARBLE outperforms the existing state-of-the-art models (A & E) in terms of F-measure and purity. Moreover, by analyzing the results of models B and C we see a clear synergy between background modeling and simultaneous topic-event learning. Neither of them separately achieves a large increase of the F-measure, but when combined they do. Fig. 6b shows that the same conclusions can be drawn from the analysis of BCubed metrics.

Fig. 7 provides visual insight on the quality of the events detected by each of the alternatives, by drawing tweets in a 3-dimensional space corresponding to the spatial (lat, long) and temporal (time) features. Each tweet is colored with the maximum likelihood event assignment (c_n^*) for that tweet. Moreover, to improve visualization, the most populated cluster, which usually is the background, is plotted with tiny dots



(a) Mercè 2014 - Set matching metrics



(b) Mercè 2014 - BCubed metrics

Fig. 6: Detection performance. (A) McInerney & Blei model (B) WARBLE w/o simultaneous topic-event learning (C) WARBLE w/o background model (D) WARBLE model (E) Tweet-SCAN

for all models, except model A, which fails to capture a clear background cluster. The figure shows that the similarity between hand-labeled data and the WARBLE model can only be compared to that of Tweet-SCAN.

6 CONCLUSIONS

In this paper, we identified three main challenges in event detection from Twitter data, namely rarity, text-shortness and variability. In order to address them, we proposed WARBLE, a new probabilistic model and variational learning algorithm that uncovers real-world events from tweets in an unsupervised manner. The WARBLE model explicitly tackles rarity and variability through a background component, which captures varying tweet densities in time and space. To mitigate text-shortness, our proposal simultaneously learn topics and events making it easier to find word co-occurrences among tweets within the same event. Furthermore, this probabilistic approach to event detection paves the way to reason about unseen observations or partially observed data in a probabilistically well principled way.

The experimental results show that WARBLE outperforms other state-of-the-art techniques in detecting local events from “La Mercè 2014” dataset. Moreover, the evaluation highlights the need to simultaneously consider the spatio-temporal background and joint topic-event learning. The event detection model also provides automatic summarization about the event, enabling to describe different aspects of the event (“When?”, “Where?”, “What?”).

Despite Gaussian distributions are computationally convenient for spatio-temporal features, future work should consider the use of more complex statistical models for these dimension to study the impact of these assumptions in the trade-off between detection accuracy and computational complexity. Furthermore, understanding the influence of hyperparameters in the detection capabilities of the proposed model as well as tunning up them through Bayesian non-parametric, seems a promising avenue for future research in this area.

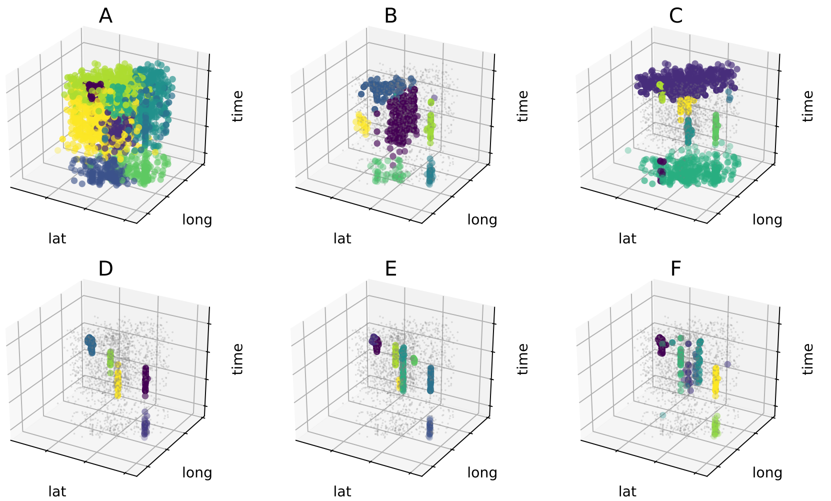


Fig. 7: (A) McInerney & Blei model (B) WARBLE w/o simultaneous topic-event learning (C) WARBLE w/o background model (D) WARBLE model (E) Tweet-SCAN (F) Labeled events

Acknowledgements This work is partially supported by Obra Social "la Caixa", by the Spanish Ministry of Science and Innovation under contract (TIN2015-65316), by the Severo Ochoa Program (SEV2015-0493), by SGR programs of the Catalan Government (2014-SGR-1051, 2014-SGR-118), Collectiveware (TIN2015-66863-C2-1-R) and BSC/UPC NVIDIA GPU Center of Excellence. We would also like to thank the reviewers for their constructive feedback.

References

Akbari M, Hu X, Liqiang N, Chua TS (2016) From tweets to wellness: wellness event detection from Twitter streams. In: Proceedings of the 30th AAAI Conference on Artificial Intelligence

Allan J, Carbonell JG, Doddington G, Yamron J, Yang Y (1998) Topic detection and tracking pilot study final report. In: Proceedings of the DARPA Broadcast News Transcription and Understanding Workshop

Amigó E, Gonzalo J, Artiles J, Verdejo F (2009) A comparison of extrinsic clustering evaluation metrics based on formal constraints. Information Retrieval 12(4):461–486

Atefeh F, Khreich W (2015) A survey of techniques for event detection in Twitter. Computational Intelligence 31(1):132–164

Bagga A, Baldwin B (1998) Algorithms for scoring coreference chains. In: Proceedings of the first International Conference on Language Resources and Evaluation Workshop on Linguistics Coreference, pp 563–566

Banfield JD, Raftery AE (1993) Model-based Gaussian and non-Gaussian clustering. Biometrics 49(3):803–821

Becker H, Naaman M, Gravano L (2011) Beyond trending topics: real-world event identification on Twitter. In: Proceedings of the 5th International AAAI Conference on Weblogs and Social Media (ICWSM)

Birant D, Kut A (2007) ST-DBSCAN: An algorithm for clustering spatial-temporal data. Data and Knowledge Engineering 60(1):208 – 221

Blei DM (2012) Probabilistic topic models. Communications of the ACM 55(4):77–84

Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. Journal of machine Learning research 3(Jan):993–1022

Boettcher A, Lee D (2012) Eventradar: A real-time local event detection scheme using Twitter stream. In: Proceedings of the IEEE International Conference on Green Computing and Communications (GreenCom), IEEE, pp 358–367

Capdevila J, Cerquides J, Torres J (2016a) Recognizing warblers: a probabilistic model for event detection in Twitter. In: the Anomaly Detection Workshop in the International Conference on Machine Learning (ICML)

Capdevila J, Cerquides J, Torres J (2016b) Variational forms and updates for the WARBLE model. Tech. rep., `http://people.ac.upc.edu/jc/papers/warble_technical.pdf`

Capdevila J, Cerquides J, Nin J, Torres J (2017) Tweet-SCAN: An event discovery technique for geo-located tweets. Pattern Recognition Letters 93:58 – 68

Cheng T, Wicks T (2014) Event detection using Twitter: a spatio-temporal approach. PloS one 9(6):1–10

Ester M, Kriegel HP, Sander J, Xu X (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD), vol 96, pp 226–231

Fox CW, Roberts SJ (2012) A tutorial on variational Bayesian inference. Artificial Intelligence Review 38(2):85–95

Ghahramani Z, Beal MJ (2001) Propagation algorithms for variational Bayesian learning. Proceeding of the Advances in Neural Information Processing Systems (NIPS)

Gomide J, Veloso A, Meira W Jr., Almeida V, Benevenuto F, Ferraz F, Teixeira M (2011) Dengue surveillance based on a computational model of spatio-temporal locality of Twitter. In: Proceedings of the 3rd International Web Science Conference (WebSci)

Hong L, Davison BD (2010) Empirical study of topic modeling in Twitter. In: Proceedings of the First Workshop on Social Media Analytics

Jordan MI, Ghahramani Z, Jaakkola TS, Saul LK (1999) An introduction to variational methods for graphical models. Machine learning 37(2):183–233

Koller D, Friedman N (2009) Probabilistic graphical models: principles and techniques. MIT press

Krumm J, Horvitz E (2015) Eyewitness: Identifying local events via space-time signals in Twitter feeds. In: Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM

Kulldorff M (1997) A spatial Scan statistic. Communications in Statistics-Theory and methods 26(6):1481–1496

Kulldorff M, Heffernan R, Hartman J, Assunção R, Mostashari F (2005) A space-time permutation Scan statistic for disease outbreak detection. PLoS Med 2(3)
Lee CH (2012) Mining spatio-temporal information on microblogging streams using a density-based online clustering method. Expert Systems with Applications 39(10):9623 – 9641
Lee R, Sumiya K (2010) Measuring geographical regularities of crowd behaviors for Twitter-based geo-social event detection. In: Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks (LBSN), pp 1–10
Li J, Cardie C (2014) Timeline generation: tracking individuals on Twitter. In: Proceedings of the 23rd international conference on World Wide Web (WWW)
Li L, Goodchild MF, Xu B (2013) Spatial, temporal, and socioeconomic patterns in the use of twitter and flickr. Cartography and Geographic Information Science 40(2):61–77
Li Z, Wang B, Li M, Ma WY (2005) A probabilistic model for retrospective news event detection. In: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, ACM, pp 106–113
Long R, Wang H, Chen Y, Jin O, Yu Y (2011) Towards effective event detection, tracking and summarization on microblog data. In: Web-Age Information Management, Springer, pp 652–663
McInerney J, Blei DM (2014) Discovering newsworthy tweets with a geographical topic model. In: The News Publishing workshop in tthe 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)
Newman N (2011) Mainstream media and the distribution of news in the age of social discovery. Reuters Institute for the Study of Journalism, University of Oxford
Pan CC, Mitra P (2011) Event detection with spatial latent dirichlet allocation. In: Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries, pp 349–358
Panagiotou N, Katakis I, Gunopulos D (2016) Detecting events in online social networks: definitions, trends and challenges, Springer International Publishing, Cham, pp 42–84
Petrović S, Osborne M, Lavrenko V (2010) Streaming first story detection with application to Twitter. In: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pp 181–189
Quan X, Kit C, Ge Y, Pan SJ (2015) Short and sparse text topic modeling via self-aggregation. In: Proceedings of the 24th International Conference on Artificial Intelligence (IJCAI)
Ritter A, Etzioni O, Clark S, et al. (2012) Open domain event extraction from Twitter. In: Proceedings of the 18th international conference on Knowledge discovery and data mining (KDD)
Sakaki T, Okazaki M, Matsuo Y (2010) Earthquake shakes Twitter users: real-time event detection by social sensors. In: Proceedings of the 19th international conference on World Wide Web (WWW)
Singh S (2015) Spatial temporal analysis of social media data. Master's thesis, Technische Universität München

Stelter B, Cohen N (2008) Citizen journalists provided glimpses of Mumbai attacks. URL http://www.nytimes.com/2008/11/30/world/asia/30twitter.html

Tamura K, Ichimura T (2013) Density-based spatiotemporal clustering algorithm for extracting bursty areas from georeferenced documents. In: Proceedings of IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp 2079–2084

Teh YW, Jordan MI, Beal MJ, Blei DM (2006) Hierarchical dirichlet processes. Journal of the American Statistical Association 101(476):1566–1581

Wang X, Grimson E (2008) Spatial latent dirichlet allocation. In: Advances in neural information processing systems (NIPS)

Weng J, Lee BS (2011) Event detection in Twitter. In: Proceedings of the 5th International AAAI Conference on Weblogs and Social Media (ICWSM)

Wong WK, Neill DB (2009) Tutorial on event detection. In: the International Conference on Knowledge Discovery and Data Mining (KDD)

Yang Y, Pierce T, Carbonell J (1998) A study of retrospective and on-line event detection. In: Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval

Zheng Y (2012) Tutorial on location-based social networks. In: the 21st international conference on World Wide Web (WWW)