When food matters: identifying food-related events on Twitter

Eleonora Ciceri, Ilio Catallo, Davide Martinenghi and Piero Fraternali*

Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milan, Italy
first.last@polimi.it

Abstract. Food communities in Twitter are growing every year, and food-related content permeates everyday conversations. Users meet on Twitter to share recipes, give cooking advices or simply inform others about what they are eating. While some of these food-related conversations are not associated with any special occurrence, many conversations take place instead during specific events. The detection of food-related events gives interesting insights: people do not talk only about Halloween and Easter, but they also create their own food-related events, such as the promotion of products (e.g., an online petition to propose the production of bacon-flavored chips) or themed home-made recipes (e.g., a day of recipes dedicated to chocolate). In this paper, we propose an approach that accurately captures food-related content from the tweet live stream, and analyze the detected conversations to identify food-related events. The proposed technique is general as it can be applied to the identification of other thematic events in digital streams.

1 Introduction

Recently, Twitter has received much attention from the research community. It is reported[1] that 500 million tweets are published on a daily basis. Tweets cover a variety of topics, ranging from personal status updates (e.g., *"going to the gym"*) to local and global news (e.g., *"FBI investigating possible corruption at New York prison"*). Tweets may contain hashtags, i.e., words prefixed with the hash symbol #, which allow tweets with similar topics to be identified. Users interested in specific topics can search for relevant tweets by hashtags, which make it particularly easy for users to create conversations about specific events. In the following, we denote by *event* a recognizable happening of limited duration [7]. While some topics are extemporary, news-based, or tied to some specific real-world occurrence, others are always discussed, permeating from everyday conversations and involving large communities. An example is *food*: food bloggers, food celebrities, media channels and common users discuss about themes such as food for holidays, cooking advices for singles, and virtual recipe sharing

* This work is partly funded by the EC's FP7 "Smart H2O" project, and the EU and Regione Lombardia's "Proactive" project

[1] https://about.twitter.com/company

parties. Food conversations, as for other topics with a wide coverage in social media, permeate several events, which originate either within the boundary of the digital community (e.g., `#TacoTuesday`) or in the real world (e.g., `#easter`). Despite the huge adoption of Twitter as a platform for publishing and talking about events, their automatic detection still remains an open problem [4]. Indeed, given the availability of such a diverse assortment of tweets, it is still not completely clear how to automatically recognize a given hashtag (and its related stream of tweets) as being associated with an event.

In this paper, we propose a technique for the automatic detection of *topic-related events*, i.e., events pertaining to a given topic of interest. More precisely, we devise a two-step detection procedure: we first identify hashtags related to a given topic of interest, and then analyze them in order to extract the associated topic-related events. We show that, when applied to food-related events, our method is able to successfully identify relevant events among the top-1000 hashtags, attaining 100% Precision@10, and 80% Precision@172. Moreover, in addition to common food-related celebrations such as `#easter`, the proposed technique also manages to identify more Twitter-specific initiatives, such as `#MeatlessMonday`. Nevertheless, note that our technique is applicable to several other contexts, including disaster management, breaking news and political events.

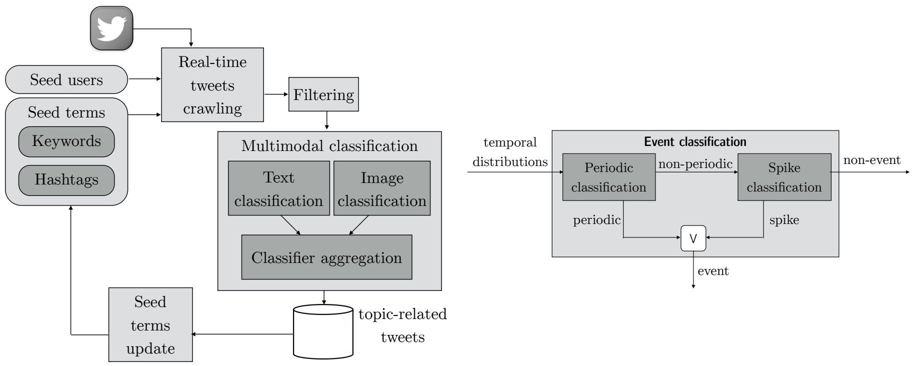
The remainder of this paper is organized as follows. We formally introduce the *topic-related event detection problem* in Section 2. In Section 3, we introduce our process for the retrieval and subsequent identification of topic-related tweets. In Section 4, our approach for the detection of events is presented. In Section 5, we demonstrate the effectiveness of our method on a real-world scenario. In Section 6, we discuss about the related works in the literature, just before our final conclusions and discussion of future works in Section 7.

2 Topic-related event detection: problem statement and proposed approach

Let $\mathcal{T} = \{\theta_1, \dots, \theta_N\}$ denote the tweet set obtained from observing the tweet live stream for a certain amount of time. Each tweet $\theta_j = \langle \omega_j, I_j, \mathcal{H}_j \rangle$ is composed of a textual component ω_j, a (possibly empty) image component I_j, and a set of related hashtags $\mathcal{H}_j$. Moreover, let τ denote a topic of interest. If we indicate with $\mathcal{Y} = \{\mathtt{Y}, \mathtt{N}\}$ the set of relevance classes for the topic τ, we can associate each tweet θ_j with a label $y_j \in \mathcal{Y}$, such that $y_j = \mathtt{Y}$ if tweet θ_j is related to topic τ, and $y_j = \mathtt{N}$ otherwise. By considering the set of the sole relevant tweets $\mathcal{T}^R = \{\theta_j : y_j = \mathtt{Y}\} \subseteq \mathcal{T}$, and defining $\mathcal{H}^R = \bigcup_{j:\theta_j \in \mathcal{T}^R} \mathcal{H}_j$ as the set of hashtags extracted from $\mathcal{T}^R$, we can therefore formulate the *topic-related event detection problem* as that of finding a set of topic-related hashtags $\mathcal{F} \subseteq \mathcal{H}^R$ that are also associated with an event.

In order to solve the event detection problem, we devise the following two-step procedure.

1. **Topic-related tweet retrieval**. Each tweet entering our system is classified as relevant/non-relevant for the topic τ. Specifically, to determine the rele-



(a) Topic-related tweet retrieval pipeline (b) Event detection pipeline

Fig. 1. Conceptual model of the two-step procedure for the detection of topic-related events

vance to topic τ we adopt a multimodal classification approach [23], which combines textual and image classification.

2. **Event classification**. For each hashtag in $\mathcal{H}^R$, we count its daily occurrences to obtain its temporal distribution (which conveys the change of its usage over time). Temporal distributions are used to classify the hashtags as either event-related or event-unrelated.

These two phases are implemented as independent processes, discussed in Section 3 and Section 4, respectively.

3 Topic-related tweet retrieval

The process for the retrieval and classification of topic-related tweets is illustrated in Figure 1(a). The system identifies topic-related tweets from the live stream in three phases: *crawling*, *filtering* and *classification*. Thanks to the presence of a feedback loop, the system automatically follows the topics users are currently discussing, thus adapting the crawling step to the emerging trends in conversations. Let us comment in greater detail on each such step.

3.1 Crawling phase

Let a *seed user* denote a user which was identified by a domain expert as relevant to topic τ. Moreover, let $\mathcal{S}$ be a set of tweets manually labeled as relevant/non-relevant to topic τ. A *seed term* (either keyword or hashtag) is a term that appears frequently in positively labeled tweets and rarely in negatively labeled tweets in $\mathcal{S}$. The crawling module monitors the tweet live stream[2], and retains tweets meeting at least one of the following selection criteria: *i)* authored by a seed user; or *ii)* containing at least one relevant seed term.

[2] Within the limitations of the Twitter's terms of service.

3.2 Filtering phase

The collected tweets proceed in input to the filtering module, which discards a tweet if at least one of the following conditions holds: *i)* the tweet content is not written in English, *ii)* the tweet contains inappropriate words, or *iii)* the tweet contains words belonging to a topic-dependent set of stop words (e.g., "apple" in the case of food).

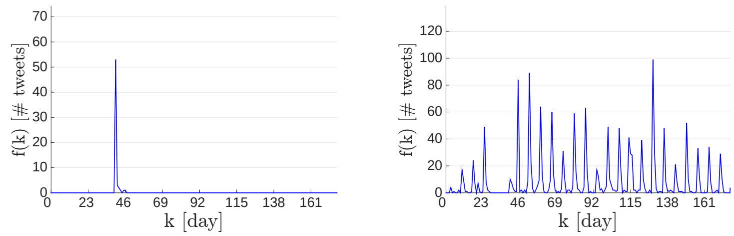
3.3 Classification phase

This step consists of a classification phase, at the end of which each tweet is labeled as relevant/non-relevant to the topic τ. We first disaggregate each tweet $\theta_j \in \mathcal{T}$ in its constituting components ω_j and I_j, and then use a textual and an image classifier to obtain two independent opinions on the relevance of ω_j and I_j to the topic τ. Finally, we merge these opinions to obtain a unique relevance label y_j for the tweet θ_j.

Text classification. We collected a dataset of tweets $\mathcal{T}^\omega$ (such that $\mathcal{T}^\omega \cap \mathcal{T} = \emptyset$) and manually annotated their textual components ω_j with a label $y_j^\omega \in \mathcal{Y}$, which specifies the relevance of ω_j w.r.t. topic τ. Each textual component ω_j is subdivided in terms. User mentions (written as `@username`) and stop words are deleted from the list of extracted terms, since they are not attributable to a specific topic. On the contrary, hashtags (after trimming the `#` symbol off) are kept as discriminative features. Finally, terms are normalized by lowercasing letters and applying Porter stemming [20], and the feature vector x_j^ω is computed according to a TF-IDF approach. To train the classifier and assess its performance, we split $\mathcal{T}^\omega$ in training set $\mathcal{T}^\omega_{\text{train}}$ (60%), cross-validation set $\mathcal{T}^\omega_{\text{CV}}$ (20%) and test set $\mathcal{T}^\omega_{\text{test}}$ (20%). An SVM classifier with RBF kernel is trained on the set $\{(x_j^\omega, y_j^\omega)\}_{j:\theta_j \in \mathcal{T}^\omega_{\text{train}}}$. The combination of the classifier parameters (i.e., the regularization parameter C and the kernel width σ) that guarantees the best performance on the cross validation set $\mathcal{T}^\omega_{\text{CV}}$ is selected, and the classifier performance is computed on the test set $\mathcal{T}^\omega_{\text{test}}$.

Image classification. We collected a dataset of tweets $\mathcal{T}^I$ (such that $\mathcal{T} \cap \mathcal{T}^I = \emptyset$) and manually annotated their image component I_j with a label $y_j^I \in \mathcal{Y}$, which specifies the relevance of I_j w.r.t. topic τ. An equal (and small) amount of positive and negative samples is extracted from $\{I_j\}_{j:\theta_j \in \mathcal{T}^I}$, and their key-points together with the related SIFT descriptors [16] are computed. By applying k-means clustering, we aggregate the extracted descriptors in K clusters, and use the centers of the learned clusters as representative terms: they characterize the visual dictionary $\mathcal{W}$. Each image I_j is then analyzed to extract its feature vector: i) we extract the key-points of I_j and the related descriptors; ii) for each key-point, we select from $\mathcal{W}$ the three most similar terms; iii) we build a histogram of occurrences of the selected terms; iv) we normalize the histogram x_j^I, which represents the feature vector for the image I_j. The set of collected visual samples is subdivided in training set $\mathcal{T}^I_{\text{train}}$ (60%), cross validation set $\mathcal{T}^I_{\text{CV}}$ (20%) and test set $\mathcal{T}^I_{\text{test}}$ (20%). An SVM classifier with RBF kernel is finally trained on the available training set $\{(x_j^I, y_j^I)\}_{j:\theta_j \in \mathcal{T}^I_{\text{train}}}$, and performance is computed on $\mathcal{T}^I_{\text{test}}$.

Classifier aggregation. In case tweet θ_j is made of a single component (i.e., either ω_j or I_j) the aggregation is not necessary. When both text and image



(a) `#NationalCheesecakeDay` (b) `#MeatlessMonday`

Fig. 2. Temporal distributions of a spike event (a) and a periodic event (b)

content exist, we aggregate the classifiers opinions, with the method proposed in [24], which applies Bayesian formalism and belief functions to estimate the aggregated label y_j.

4 Event classification

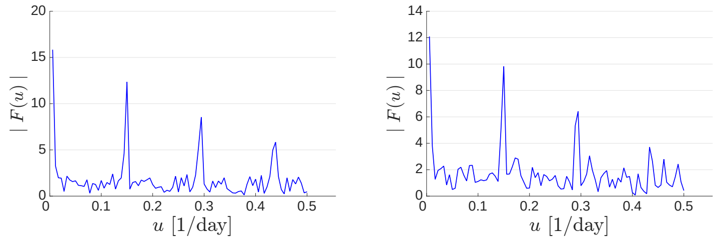
Twitter users track content related to specific topics using hashtags. Some tags are just used to describe content, so that it can be easily classified and retrieved in the future. Other hashtags are meant to track *real-world events* (e.g., earthquakes, holidays, elections) and *social events* (e.g., birthday of a social community).

When an event occurs and users start talking about it, the rate of usage of the related hashtag(s) increases rapidly, and it stays off-the-scale with respect to other common hashtags until either the event ends or the community loses interest in it. To study the rate of usage of hashtags, one can analyze their *temporal distributions*. A temporal distribution is a K-dimensional histogram associated with hashtag H, where the k-th component indicates the number $f(k)$ of tweets produced during day k that contain H. Two examples of temporal distribution are shown in Figure 2.

In this paper, we identify topic-related events by tracking temporal variations in the usage of hashtags. We start from a collection of tweets related to topic τ downloaded as described in Section 3. For each hashtag in the collection, we extract its temporal distribution, and use a supervised approach to decide if the hashtag is related to an event.

4.1 Tracked events

Events discussed on Twitter have different natures. Some events happen once, and generate a large interest (although limited in time). For these events, which we call *Spike Events*, there is a single (and strong) perturbation in the usage of related hashtags. An example of spike event is shown in Figure 2(a). Here, a single activity peak on the hashtag `#NationalCheesecakeDay` was detected, as the Twitter food-related community joined the event by massively publishing cheesecake recipes in a limited amount of time.



(a) `#MeatlessMonday` (b) `#TacoTuesday`

Fig. 3. Fourier transforms of two periodic events

On the other hand, some events are recurring periodically. For these events, which we call *Periodic Events*, there are multiple perturbations in the usage of related hashtags, such that the interest in the event raises periodically, and is null (or low) during the other days. An example of periodic event is shown in Figure 2(b). Here, an activity peak on the hashtag `#MeatlessMonday` can be detected on each Monday, since the event is joined by people that meet virtually every Monday to discuss about meatless recipes.

Figure 1(b) depicts our event classification process. The temporal distributions associated with the hashtags we want to classify as event-related/event-unrelated are fed as an input to a chain of two binary classifiers, the first dedicated to spike event detection, while the second dedicated to periodic event detection. A hashtag (or equivalently its temporal distribution) is labeled as event-related if at least one the classifiers recognizes it.

Feature set. Spike and periodic events have a peculiar temporal distribution which is common for all the events of the same class. However, when it comes to training a classifier for the recognition of event classes, temporal distributions cannot be used as feature vectors: they suffer from *temporal dependence* of subsequent components, and consequently events that clearly belong to the same class but happened in different periods of time would have completely different feature vectors and thus would not help the classifier learn the underlying model. For this reason, we used as feature vector the spectrum of the *Fourier transforms* $|F(u)|$ of the normalized temporal distribution, which describe the frequency components of the signal and are agnostic with respect to the actual time of the events. As an example, Figure 3 shows the Fourier transforms of two periodic events, which happen in different periods but have similar spectrum.

Event classifiers. When it comes to building an annotated dataset to train the classifier, we come up with an *unbalanced training set*, since events are rare if compared to the total number of produced hashtags. Due to the lack of positive samples (i.e., temporal distributions corresponding to events), the classifiers could easily fall into the problem of *overfitting the data*. Thus, we applied the EasyEnsemble algorithm [13], which uses undersampling to rebalance the training set, combined with AdaBoost classifiers [11], since boosting is often robust to overfitting. Finally, to assess the performance of the classifiers on the training and test sets, we applied K-fold cross validation, with $K = 10$.

(a) Topic classifier

Text samples $\mathcal{T}^{\omega}$	Dictionary size	12988
	Positive samples	14234
	Negative samples	14218
	Total samples	**28452**
Image samples $\mathcal{T}^{I}$	Dictionary size	5000
	Positive samples	11759
	Negative samples	11746
	Total samples	**23505**

(b) Spike and periodic event classifier

Samples in $\mathcal{H}_s^R$	Positive samples	2030
	Negative samples	4870
	Total samples	**6900**
Samples in $\mathcal{H}_p^R$	Positive samples	5000
	Negative samples	5890
	Total samples	**10890**

Table 1. Dataset cardinalities

5 Experimental Evaluation

In this section we assess the performance of the proposed topic-related event detection approach. We first show how we can correctly identify topic-related tweets captured from the tweet live stream. Then, we apply event detection to the resulting tweet set, showing that our approach is capable of attaining good performance (measured as Precision@K).

5.1 Topic-related tweet retrieval

In the following, we illustrate the characteristics of the datasets we used to assess the multimodal classifier performance and report classification performance.

Dataset description. We trained the text and image classifiers on, respectively, the textual and image datasets $\mathcal{T}^{\omega}$ and $\mathcal{T}^{I}$, whose cardinalities are reported in Table 1(a). To test our classification approach, we randomly extracted and manually annotated the following sets of samples: *i)* $\tilde{\mathcal{T}}^{\omega}$, composed of 1900 tweets containing only text; *ii)* $\tilde{\mathcal{T}}^{\omega+I}$, composed of 1900 tweets containing both text and images, where $\mathcal{T}^{\omega}$, $\mathcal{T}^{I}$, $\tilde{\mathcal{T}}^{\omega}$, $\tilde{\mathcal{T}}^{\omega+I}$ are all disjoint. Note that some tweets are characterized by ambiguous content, and thus annotating them as relevant or not relevant is difficult for a human annotator too. On our dataset, the inter-annotator agreement is 93.86%.

Classifiers performance. Multimodal classification improves performance with respect to text classification on $\tilde{\mathcal{T}}^{\omega}$ and $\tilde{\mathcal{T}}^{\omega+I}$. Table 2 shows how accuracy, precision, recall and $F1$-measure increase in this scenario.

Text classification performance is insufficient when images are involved, because it is not able to interpret visual content and may misinterpret the text associated with images.

5.2 Event classification

In this section, we assess the performance of the proposed event detection technique on the food-related tweets.

Dataset description. We ran our topic-related tweet retrieval process from June 1, 2014 to June 10, 2015. During that period, the system processed more than 15 million tweets, 9 millions of which were labeled as food-related. The corresponding number of relevant hashtags was 171451. However, only 21451 were associated with a temporal distribution comprising more than 5 tweets and were included in the final set of topic-related hashtags $\mathcal{H}^R$. In order to train the spike classifier, we took a random sample $\mathcal{H}^R_s$ of size 6900 from $\mathcal{H}^R$. Then, we performed a data annotation campaign on the crowdsourcing platform Champagne [6], to label them as event-related/unrelated. Crowd workers were prompted with a sequence of temporal distributions (similar to those in Figure 2), and asked to identify spike events. A different approach was instead required for training the periodic classifier, due to the fact that periodic events are quite rare in $\mathcal{H}^R$. We compensated for this unfavorable situation as follows. We first identified 10 periodic events in $\mathcal{H}^R$. We then used such events to synthetically generate 5000 new positive instances by combining each periodic event with a Gaussian process with mean 0 and variance 0.03, and randomly shifting the temporal distribution within a period of 7 days. Such procedure is similar to what is done in the literature (see, e.g., [18]). Let us denote the resulting dataset as $\mathcal{H}^R_p$. The cardinalities of the two datasets are reported in Table 1(b).

Classifiers performance. The performance of the spike and periodic event classifiers are reported in Table 3. As shown, both classifiers attain high values of $F1$-measure and accuracy, on both the training and test set. In order to further evaluate the effectiveness of our approach, we also tested the proposed event detection technique against a gold standard dataset $\mathcal{H}^R_g$, which we obtained by first ordering hashtags in $\mathcal{H}^R$ by total number of tweets, and then providing a gold label for the first 1000 hashtags. In particular, each hashtag has been assigned a gold label by analyzing different factors, such as the name of the hashtag, its current use on Twitter, the shape of its temporal distribution, and the content of tweets collected by the process. Since the total number of tweets might be intended as a proxy for the success of an event, we believe that testing the proposed technique against the top-1000 hashtags can provide a meaningful insight on its effectiveness in detecting successful events. Since the test was performed against a top-K ranked list, we measured performance by means of a Precision-Recall curve, which depicts the attained precision-recall values as K increases. Figure 4 reports the performance of our technique on $\mathcal{H}^R_g$. As shown, our method correctly identifies the first 14 food-related events. Overall, our method labels 172 events as food-related, which leads to a final precision-recall value of $(0.80, 0.67)$.

<table>
<tr><th></th><th>$\tilde{\mathcal{T}}^{\omega}$</th><th>$\tilde{\mathcal{T}}^{\omega+I}$</th></tr>
<tr><td>Text classification</td><td rowspan="2">Accuracy = 75.22%
Precision = 65.67%
Recall = 80.54%
F1 − measure = 72.35%</td><td>Accuracy = 73.47%
Precision = 63.61%
Recall = 80.30%
F1 − measure = 70.99%</td></tr>
<tr><td>Multimodal classification</td><td>Accuracy = 82.23%
Precision = 79.13%
Recall = 97.22%
F1 − measure = 87.24%</td></tr>
</table>

Table 2. Performance of text classification and multimodal classification

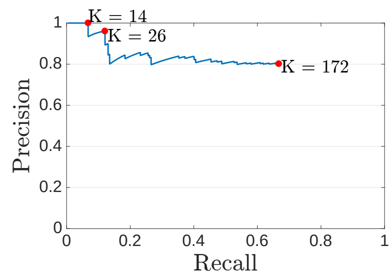


Fig. 4. Recall-Precision curve

Discussion. Table 4 shows the top-10 food-related hashtags retrieved by our event detection pipeline, together with tweet samples showcasing their usage. The list reports: *i)* food-centered social events that are confined in the Twittersphere (`#foodiechats`, `#MeatlessMonday`, `#bandwiches`); *ii)* holidays (`#halloween`, `#easter`) and periodic calendar-based events (`#sunday`, `#tbt` [that is: '*Throwback Thursday*'], `#tgif` [that is: '*Thank God it's Friday* ']) during which users share themed recipes; *iii)* media events (`#espys`), during which people have dinner in front of the TV and share comments about the show and their food; *iv)* food-centered advertising campaign (`#TeamWalmartProduce`). Finally, Figure 5 shows a set of images retrieved by our pipeline and related to the periodic calendar-based events `#TacoTuesday` and `#NationalCheesecakeDay`. This sample shows how our pipeline is able to retrieve high quality multimedia content (thanks to multimodal classification), which could be used, e.g., to summarize the contents shared by Twitter users during the detected events.

6 Related work

A number of recent works in the literature cover the problem of event detection on Twitter. The work in [22] builds a spatiotemporal model to estimate where and when events happened, with specific focus on earthquakes and typhoons. The work in [10] applies a state-of-the-art earthquake detection algorithm to detect earthquake-related tweets in real-time. The demo in [17] proposes a system which identifies in real-time real-world events by detecting bursty keywords. The work in [14] detects unusually crowded regions that can eventually suggest the

	Training set	**Test set**
Spike classifier	`Accuracy` = 92.48% `F1-measure` = 87.99%	`Accuracy` = 92.11% `F1-measure` = 88.14%
Periodic classifier	`Accuracy` = 99.77% `F1-measure` = 99.75%	`Accuracy` = 99.45% `F1-measure` = 99.42%

Table 3. Performance of spike and event classifiers when tested against the training and test set

Hashtag	# tweets	Representative tweet
#foodiechats	28845	@Foodiechats We have Smoked Turkey Sliders, Tandoori Chicken Flatbread Panko Sesame Fish Skewers, and Peach Shortcake! #foodiechats
#MeatlessMonday	26643	Spicy black bean burgers. #MeatlessMonday #food
#TeamWalmartProduce	22421	There's nothing better than a dessert with delicious stone fruit! #ad #TeamWalmartProduce
#sunday	19966	Photo: Sushi treats at the Spice Haat Sunday Brunch #sunday #brunch #sushi
#halloween	16201	Strawberry Ghosts – are these cute! Love the little ghost "tails" on them #halloween #partyfood
#espys	10002	first time i've ever cried while eating pizza. love you, Stuart Scott. #staySTRONG #espys
#tbt	9268	RT @Justelise97: Pancakes + Vanilla Ice Cream #tbt #throwback #foodporn
#easter	8964	RT @FoodEmbassy_: This #Italian #pie has #easter written all over it! Torta Pasquale!! @BBCFood
#bandwiches	7978	Peanut butter and Pearl Jam #bandwiches @midnight
#tgif	6903	Egg whites and PB toast. #postworkout #breakfast #daymaker #tgif #riseandshine #todayisagoodday #smile

Table 4. Top-10 food-related hashtags based on the total number of tweets

occurrence of geo-social events. The work in [1] identifies local events by dividing the timeline of a potential event in time frames, extracting bursty keywords in each time frame and selecting only the keywords that have local spatial distribution. The work in [8] retrieves tweets that contain drug-related keywords and identifies drug-related events as spikes in the number of collected tweets. The work in [25] classifies social events by clustering temporal series having similar shapes. The work in [9] applies a similar approach, with the strong assumption that no event can transgress the boundaries of a day. The work in [7] sequentially retrieves tweets from Twitter and transform them in lists of words, which are then used to cluster keywords according to their density and filter non-local events. The work in [12] manually identifies hashtags related to the *Je Suis Charlie* event and analyze how it relates to the raising counter-events (e.g., *Je Ne Suis Pas Charlie*). The work in [21] performs POS tagging, named entity extraction and extraction of temporal expressions to create classes of events, using unsupervised approaches, attaining a Precision@100 of 90%, a Precision@500 of 66% and a Precision@1000 of 52%. The work in [26] detects composite social events over streams, by using information deriving from similarity between messages in the social stream. The work in [19] analyzes the sentiment of produced tweets to discover real-world events, under the assumption that an event shifts

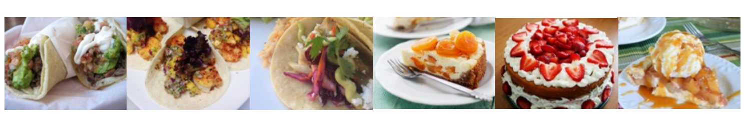


Fig. 5. Images from #TacoTuesday (left) and #NationalCheesecakeDay (right)

the sentiment toward a topic (represented by specific keywords in the content). Events are thus recognized as bursty keywords that shifted the mood of users. This approach achieves 60% recall if the objective is to discover the exact date of an event, and 90% recall if a tolerance of ± 1 day is allowed. The work in [2] uses several topic detection algorithms and an extension of the tf-idf approach over time to recognize emerging bursty topics. For this work, the Recall@N varies between 50% and 90% (depending on the used dataset). Although we rank favorably with comparable works such as [19] and [2], in many cases we cannot directly contrast our approach to what is present in the literature. Indeed, while our technique aims at identifying how hashtags relate to events, a significant percentage of previous works ([22], [10], [14], [1], [7], [26]) focus instead on the problem of spatially localizing such events. A direct comparison is also not possible for those works that try to identify open-domain events, such as [21].

Several works use supervised classification methods to state if content is related to an event. The work in [5] clusters similar messages to perform topic identification, and then classifies content as event-related/event-unrelated, based on temporal features (e.g., deviations from expected message volume), social features (e.g., retweets and mentions), topical features (e.g., focus on a topic) and Twitter-centric features (e.g., hashtag usage). In that work, the F1 measure achieves 83.7% on test set, while Precision@20 is 65%. The work in [3] uses an SVM classifier to select flu-related tweets, to track how flu moves over space and time. The F1-measure achieved by this method is 75.6%. The work in [15] identifies crime and disaster-related events via binary classification, based on Twitter-specific features (e.g., hashtags) and on the presence of event-specific text features (e.g., presence of happening time). Although on a different topic of interest, our approach is competitive with the afore-mentioned classification-based methods available in the literature. Moreover, note that none of the previous works deal with the problem of identifying periodic events, which we showed is an interesting problem in itself and permits unveiling a significant percentage of social events.

7 Conclusions

In this paper, we investigated the problem of topic-related event detection on Twitter, which we cast as a supervised learning problem. We focused on the concrete use case of identifying events that include a food-related component, such as holidays or commercial initiatives. We first induced a multimodal classifier capable of identifying tweets related to the topic of interest, which we used to isolate relevant tweets from the global tweet stream. Events were therefore identified by applying a chain of two classifiers, one for the identification of periodic events and one for the identification of spike events.

The experimental evaluation showed that our approach attains a Precision@10 value of 100%, and a Precision@172 value of 80%, proving therefore competitive with other state-of-the-art approaches available in the literature. Future work will focus on enriching the event classifier feature vector to capture social components, such as user profile characteristics (e.g., authority) and network characteristics (e.g., centrality) and on spatial distribution analysis.

References

1. H. Abdelhaq et al. Eventweet: Online localized event detection from twitter. *VLDB*, 6(12), 2013.
2. L. M. Aiello et al. Sensing trending topics in twitter. *IEEE Transactions on Multimedia*, 15(6), 2013.
3. E. Aramaki et al. Twitter catches the flu: detecting influenza epidemics using twitter. In *EMNLP*, 2011.
4. F. Atefeh and W. Khreich. A survey of techniques for event detection in twitter. *Computational Intelligence*, 2013.
5. H. Becker et al. Beyond trending topics: Real-world event identification on twitter. *ICWSM*, 11, 2011.
6. C. Bernaschina et al. Champagne: a web tool for the execution of crowdsourcing campaigns. In *WWW Companion volume*, 2015.
7. A. Boettcher and D. Lee. Eventradar: A real-time local event detection scheme using twitter stream. In *GreenCom*, 2012.
8. C. Buntain and J. Golbeck. This is your twitter on drugs: Any questions? In *WWW Companion*, pages 777–782, 2015.
9. C. De Boom et al. Semantics-driven event clustering in twitter feeds. In *Making Sense of Microposts (# Microposts2015) (WWW)*, 2015.
10. P. S. Earle et al. Twitter earthquake detection: earthquake monitoring in a social world. *Annals of Geophysics*, 54(6), 2012.
11. Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *JCSS*, 55(1), 1997.
12. F. Giglietto and Y. Lee. To be or not to be charlie: Twitter hashtags as a discourse and counter-discourse in the aftermath of the 2015 charlie hebdo shooting in france. In *Making Sense of Microposts (# Microposts2015) (WWW)*, 2015.
13. H. He et al. Learning from imbalanced data. *TKDE*, 21(9), 2009.
14. R. Lee and K. Sumiya. Measuring geographical regularities of crowd behaviors for twitter-based geo-social event detection. In *ACM SIGSPATIAL, 2010*, 2010.
15. R. Li et al. Tedas: A twitter-based event detection and analysis system. In *ICDE*, 2012.
16. D. G. Lowe. Object recognition from local scale-invariant features. In *Computer vision, 1999*, volume 2, 1999.
17. M. Mathioudakis and N. Koudas. Twittermonitor: trend detection over the twitter stream. In *SIGMOD*, 2010.
18. J. Nonnemaker and H. S. Baird. Using synthetic data safely in classification. In *IS&T/SPIE Electronic Imaging*, pages 72470G–72470G, 2009.
19. G. Paltoglou. Sentiment-based event detection in twitter. *Journal of the Association for Information Science and Technology*, 2015.
20. M. F. Porter. An algorithm for suffix stripping. *Program*, 14(3), 1980.
21. A. Ritter et al. Open domain event extraction from twitter. In *SIGKDD*, 2012.
22. T. Sakaki et al. Earthquake shakes twitter users: real-time event detection by social sensors. In *WWW*, 2010.
23. M. Woźniak et al. A survey of multiple classifier systems as hybrid systems. *Information Fusion*, 16:3–17, 2014.
24. L. Xu et al. Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE TSMC*, 22(3), 1992.
25. J. Yang and J. Leskovec. Patterns of temporal variation in online media. In *WSDM*, 2011.
26. X. Zhou and L. Chen. Event detection over twitter social media streams. *The VLDB Journal*, 23(3), 2014.