Front.Comput.Scireview article manuscript No.
(will be inserted by the editor)

Event Analysis in Social Multimedia: A Survey

Xueliang Liu · Meng Wang · Benoit Huet

Received: date / Accepted: date

Abstract Recent years have witnessed the rapid growth of social multimedia data available over the Internet. The age of huge amount of media collection provides users facilities to share and access data, while it also demands the revolution of data management techniques, since the exponential growth of social multimedia requires more scalable, effective and robust technologies to manage and index them. The event is one of the most important cues to recall people's past memory. The reminder value of an event makes it extremely helpful in organizing data. The study of event based analysis on social multimedia data has drawn intensive attention in research community. In this article, we provide a comprehensive survey on event based analysis over social multimedia data, including event enrichment, detection, and categorization. We introduce each paradigm and summarize related research efforts. In addition, we also suggest the emerging trends in this research area.

Xueliang Liu
School of Computer and Information, Hefei University of Technology Hefei, 230009, China
E-mail: liuxueliang@hfut.edu.cn

Meng Wang
School of Computer and Information, Hefei University of Technology Hefei, 230009, China
E-mail: eric.mengwang@gmail.com

Benoit Huet
Department of Multimedia, EURECOM sophia-antipolis, 06560, France
E-mail: huet@eurecom.fr

1 Introduction

In recent years, we have witnessed the rapid growth of social media collections available over the Internet. A huge amount of media data, such as the photos from a conference that we recently attended, a video clip from a football game or an important game, or eyewitness description that made it to the top list in a newspaper, could be found online. The age of social multimedia provides users facilities to share and access data, while it also demands the revolution of data management techniques[1–6]. All these media content are created by different devices and are shared in different media types (e.g. image, video, text) in different web services. What is common among these media is that they all are captured and convey information about real life events. Efficiently and effectively understanding the heterogeneous event semantic behind these multimedia content and using them to better organize, manage the content is a significant challenge in research community [7].

The event, the public happening that takes place in a given location and time [8,9], is one of the most important cues to recall people's past memory. The reminder value of the event makes it extremely helpful in organizing data. With the development of Web 2.0, many event based information sharing sites, such as last.fm, eventful, upcoming, are appearing online, and a wide variety of events are scheduled and described by these social online services.

In recent years, event based media analysis has attracted much attention, and the study of event based social media analysis could not only continue the methods and algorithms in traditional media analysis, but also leverage the event domain knowledge and ontology to formulate the raised problems deeply. The research community has proposed many different solutions and approaches for modeling, detecting, and processing events [10–12].

In this survey paper, we focus on the problem of mining relationship between events and social media data, while the previous related work has been focused on event enrichment, detection and categorization over social multimedia. Event enrichment [9,13] is the problem of linking the multimedia data to given events, if the media data are taken during events and capture the event content, so that the social media data could be leveraged to illustrate the event by rich modality and from different viewpoints. Event detection [14–16] tries to detect events from a social multimedia stream, and to identify the event attributes such as taken time, location, content. In this problem, there is no a priori knowledge of events. Event categorization [17,18] is the process of categorizing social multimedia based on events. Comparing with the traditional multimedia organization, such as content based clustering, event categorization provides a novel way to organize multimedia data. In this survey, we present a comprehensive detailed study of the research topics above and review the recent advances on social event analysis.

2 Event Enrichment in Social Multimedia

The event is a natural way of referring to any observable occurrence grouping people, place, time, and activities as described in [8]. Event enrichment is the process of enhancing events with pertinent information associated with the event source. Enrichment typically includes the online media from different sources/websites which are generated during events, as well as any known steps to help users grasp the gist. For example, when an event starts, many photos, videos, and microblogs are taken and shared on the social media sites by different participants. Linking these media data from different users to illustrate events in the real world is an efficient way to help user browse multimedia data.

To help users grasp events effectively, some event browsing and searching platforms have been built, which have benefited greatly from social media event content, e.g. eventful.com, upcoming.org, and last.fm. These services sometimes have an explicit connection with media sharing platforms. However, in these web services, less attention is paid to improve the users' experience when searching and browsing content. Actually, automatically associating social media content with known events is a challenging problem owing to the heterogeneous and noisy nature of the data. In recent years, several works have been proposed to investigate how to search event related media data. In this section, we review the work to infer the semantics behind the events and explore social media to illustrate events.

The earlier relevant research focused on the study of news, since the data about news are abundant and easy to collect [19,20]. In [21], to improve the users' attention when reading news articles, a system was proposed to help people reading news by illustrating news story. The application provides mechanisms to select the best illustration for each scene and to select the set of illustrations automatically to improve the story sequence. In [22], an unsupervised approach was presented to describe stories with automatically collected pictures. In this work, semantic keywords are extracted from the story, and used to search an annotated image database. Then a novel image ranking scheme automatically chooses the most important images. In [23], to organize digital photograph collections according to events, the authors presented similarity based methods to cluster digital photos by time and image content. The approach is general and unsupervised, and makes minimal assumptions regarding the structure or statistics of the photo collection. The problem is solved in two steps; first photo similarity is quantified at multiple temporal scales to identify the event clusters, and second the final clusters are determined according to some goodness criteria. In [24], to structure the collections and to generate automatically important summaries, the authors exploited the timing information and develop two photo browsers for collections with thousands of time-stamped digital images. In [25], to browse heterogeneous news sources effectively, a system was developed to address the challenge of extracting "who","what","when" and "where" from a multimodal perspective, which leverages audio-visual information in broadcast news and these embedded in articles, as well as textual cues in both closed captions and raw document content in articles and social media. In [26], a novel method for finding all event related documents across diverse media sources by link analysis. To analyze the connection between documents, the authors use document hyperlinks in the textural content to generate an accurate citation network. On analyzing the ICWSM 2011 Spinn3rdataset [27], it could be observed that hyperlinks across different event related documents account for the majority of the total links, and that such cross event links reveal surprising yet reasonable connections among events that co-evolve over time.

With the popularity of Web 2.0, some social media repositories are used to share users' experience and interests on the Web. These sites also host substantial an amount of user contributed content (e.g., photographs, videos, and textual discription) for a wide variety of real world events in different types and scale. How to illustrate the events with rich social media data has gathered recent attention. In [28, 29], a system was proposed to group photos by events/subevents by visual content and time constraints analysis. In the proposed system, the authors analyzed visual features with bag-of-word method and leveraged time information of events to group a collection of pictures by events and its individual images into subevents. The system allows the users to analyze a collection of images from disk or from internet sharing services such as Picasa [30] and Flickr [31]. In [32], a method was proposed to organize digital photograph collections according to events by clustering method. The authors first clustered photos by their similarity at multiple temporal scales to identify event clusters, and second, the final clusters were determined according to clustering goodness criteria. In [33], the authors proposed an event based photo summarization system which creates a representative subset summary by extracting photos from a larger set. Three properties of the summarization, relevance, diversity and coverage, are calculated using multimodal content and context data, and then used in generating an effective summary.

In [34], the authors proposed a system to present the media content from live music events, assuming a series of concerts are recorded by the same artist. By synchronizing the music clips with audio fingerprint and other metadata, the system gives a novel interface to organize the user contributed content. In [9], Liu et al. developed a framework to associate social media documents with events in two steps: first, time, location, and textual features are extracted via a query of the media candidates on a given event, and then to improve the performance, visual filter is built to remove noisy data. In [35], Twitter messages corresponding to large scale media events were analyzed to improve event reasoning, visualization, and analytic. Based on their research, a visual analytic tool, Vox Civitas, is designed to help journalists and media professionals extract news from large scale aggregations of social media content around broadcast events. The prototype presented in [36] attempted to depict public events using both spatial-temporal context and photo content. In their system, event information are collected as event database, and then photo content model are built for different types of events. With the assistance of such as system, the users could find information about events or activities when they travel in foreign place. Paper [37] provided a way of visualizing events in a static manner. Having extracted semantic information from the main media, the topic, location, and time information related to the event of interest is available for each segment of the audio-visual content. This information is used to mine the latest and hottest related events from social news web services. Relevant events are identified by harvesting the social news website Digg.com. Then, for each event, both relevant tweets on Twitter and compelling images from Google image search are retrieved. The resulting documents are assembled and shown within a static vivid interface featuring both event descriptions, tag cloud and photo collage. In [38], the authors device an approach to enrich microblog data representation by employing machine translation to increase the number of feature from different languages. A web service with similar illustration functionality can be found in EventBurn [39]. It creates a summary of a given hot event from popular services like Twitter, Facebook, and Flickr, but fails to extract events automatically from social media streams.

In the following, we will discuss the work in [9] and [37] in details, both of which essentially present methods on solving the problem of exploring media data to illustrate events.

2.1 Event Enrichment by Multimodal Search

Organizing media data according to events in the real world is the natural way for human to recall his experience [40]. Exploiting event context to solve the management and retrieval problems raising from the social media draws lots of interest in the multimedia community. In this section, we introduce work in [9] to infer the semantics behind the events and explore social media to illustrate events, which focus on building the link between events and social media.

As we know, in recent years some social event repositories are built which help the users share their daily experience effectively. These services, e.g. eventful.com, upcoming.org, last.fm, have provided explicit connection among the event and media sharing platforms. The relationships between scheduled events and media hosted on Flickr can be looked up using special machine tags such as `lastfm:event=XXX` or `upcoming:event=XXX`. The work of [13] has explored the overlap in metadata between four popular web sites, namely Flickr as a hosting web site for photos and Last.fm, Eventful and Upcoming as a documentation of past events. However, since the machine tag is optional when people share media online, much media data are not labeled even if they are actually relevant to an event. Owing to the heterogeneous and noisy nature of the social media data, automatically associating them with known events is a challenging problem.

The problem has been addressed in [9], which focuses on finding as much as possible media resource that has not been tagged with machine tag but that should

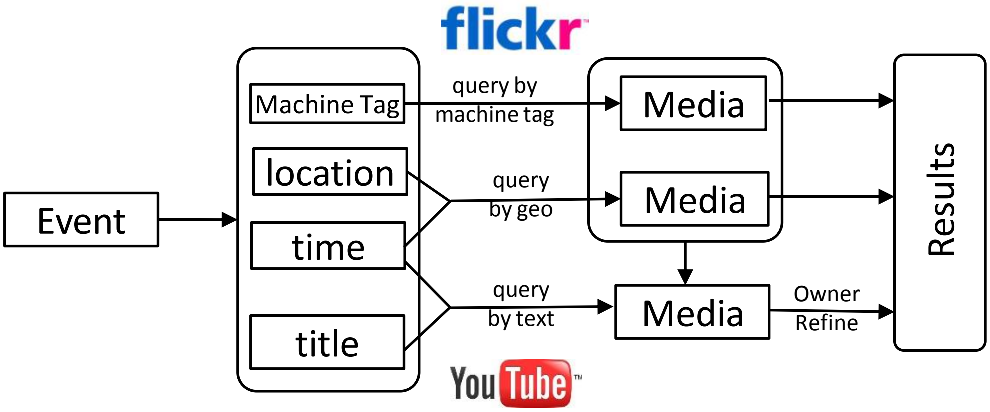

Fig. 1: Overview of enriching event by multimodal search

still be associated with an event. In this work, several approaches are investigated to find those photos/videos to which can be then propagated the rich semantic description of the event improving the recall accuracy of multimedia query for events. The whole framework to enrich event with media data is described in Figure 1. Starting from an event description, three natural attributes can easily be mapped to the media metadata available in Flickr and YouTube [41] and be used as search query in these two sharing platforms: the *what* attribute that represents the event gist, the *where* attribute that gives the geo-coordinates attached to the media item, the *when* attribute that is matched with either the taken date or the uploaded date of the media item. Since querying online with just one of these attributes brings too many noisy results, the combination of two dimensions is helpful to query the media from sharing sites. However, the combination of "title" and "geo tag" is discarded to since there are recurrent annual events with the same title and held in the same location. In the proposal, the two combinations "title" + "time" and "geotag" + "time" are used for performing search query and finding media that could be relevant to a given event. Specifically, on one hand, for all events, the latitude and longitude information is extracted and then perform geographical based query using the Flickr and YouTube API applying a time filter of five days following each event taken date. Since it is possible to have more than one event in the same place, some irrelevant media is retrieved from location based query. On the other hand, title is the most describable and readable information for events. Similarly to geo-tagged query, full text search query is performed on Flickr based on the event title that is extracted from the event description. The retrieved photos are also filtered using a time interval of five days following the event taken time. Due to the well known polysemy problem of textual based query, the title based query also brings lots of irrelevant media data.

To prune the irrelevant media data, a visual content analysis method is developed, which aims at removing the noise images if the visual difference is remarkable enough. Since the media data labeled with machine tag is highly relevant to events and could be obtained easily, they are the best choice as the training samples for filtering noise. The main idea of the algorithm is to measure the visual similarity of media documents. First, a training dataset is built which composes of the media labeled by the event machine tag. And for each image from the location or textual based query, if its minimal similarity to the training set is below a threshold, it will be added to candidates list for illustrating the event. Mathematically, let E as the training photos set, and F as the photos set from location or textual based query. The objective is to select the photos from F which are similar to the photos in E, to additionally enrich the set E illustrating an event. The visual similarity $Dist_v$ between two images is computed as follows:

$$Dist_v(F_j, E_i) = \sum_k |F_j(k) - E_i(k)| \tag{1}$$

The visual features used in the approach are 225D color moments in Lab space, 64D Gabor texture, and 73D Edge histogram [42]. And $F_j(k)$ and $E_i(k)$ are normalized concatenated low level feature vector of the images. F_j is added to the set of media illustrating the event when

$$\exists E_i \in E : L_1(F_j, E_i) < THD_i \tag{2}$$

where THD_i is the threshold which is also learned from training set E. As shown in Equation 3, a strict strategy is used to decide the threshold, which is chosen as the minimal value of similarity of images pairs in training set.

$$THD_i = \min_{\{j\}\backslash i} \sum_k |E_j(k) - E_i(k)| \tag{3}$$

Most of the irrelevant media are filtered by such a method, but some relevant ones are also discarded. In order to recall these photos, the "owner" property in Flickr is exploited and a refinement method is proposed, based on the assumption that people cannot attend more than one event simultaneously. Therefore, all the photos that have been taken by the same owner during the event duration should be assigned to the event. Using this heuristic, it is possible to retrieve photos which do not have any textual/geographical description. And "owner refinement" is the only effective approach to match the event and media data when no enough metadata (such as textual, graphical metadata) is available.

2.2 Event Enrichment by Textual Query Expansion

In this section, we introduce the work of discovering and illustrating events on any given query, and detail the solution to extract and illustrate social events automatically by leveraging the massive social media content [37]. The solution is summarized in Figure 2. Since an event can be defined as something happening at certain time in a given location, the authors start by parsing the query to identify the topic, location and time information with natural language processing algorithm. Rather than detecting events from Twitter data directly, events are queried by crawling and scraping from social news web service. This saves time, computation and storage compared to alternative event detection method. To provide vivid illustration for each event, the relevant tweets are retrieved from Twitter, while the collected data is shown by tag cloud. And photos are retrieved by Google image search engine, and summarized with photo collage/montage technique.

As we know, the three basic properties of an event are location, time, and topic, as stated in [9]. To identify the event semantic of a given query input, the information behind the query can be parsed in the three dimensions. Here, it is assumed that the query input is a noun phrase headed or tailed with complements, such as "The news of the past three days in New York". The structured data is extracted from the noun phrase, where there is a predictable organization of entities and relationships. The issue of extracting structured data from text has been well studied in Natural Language Processing (*NLP*). This process, composed of 3 steps is performed using *NLTK*, a well known *NLP* package [43].

Then, to determine the semantic meaning of each noun phrase, different techniques are employed to extract the location, time, and topic information from the parsed noun phrases. The location is obtained by a query of the DBpedia [44] knowledge database, which provides structured information extracted from Wikipedia. For determining the time information, a script is developed to parse and convert the human readable string, such as "tomorrow", "last week", "Monday" to a time structure. Since it is hard to model the topic in sophisticated way, the nouns are assigned as the topic keywords of the event to search for, if neither time nor location can be determined from it.

In the solution, rather than detecting events from social streams, the events are queried from the social news web service Digg.com. Specifically, the time, location and topic keywords are used as the query parameters to retrieve popular events, ranked by the popularity which is a feature in the web service. In the querying result, many useful metadata can be found, such as "title","submit_date" to describe the property of an event.

Finally, to illustrate events using text and images, the original news content, related tweets from Twitter and images from Google image search, are retrieved and shown in proper way. Specifically, the original textual content could be accessed easily with the "link" metadata in Digg.com. To extract the main body of the web page, the method proposed in [45] is used, to recognize main content in a web page. In order to collect other textual description and comments from different users, the event title is used as the query parameter and the relevant tweets from Twitter are retrieved. To provide a nice and meaningful visualization, a tag cloud is generated to organize the textual data. Besides the tag cloud, a photo collage is generated to illustrate events. Specifically, photos from the Google image search engine is queried with the event title and those photos are filtered out if the cosine distance between their textual metadata and the event title is below a given threshold. With the selected photos, the method proposed in [46] is used to create the photo collage.

In the end, to assist the users viewing the system well, the events are ranked according to the importance, which is measured by the entropy of tag cloud [47] of the events. In details, Let $t \in T$ be a tag from a tag cloud T, the entropy of T can be calculated as

$$entropy(T) = \sum_{t \in T} p(t) log(p(t)) \tag{4}$$

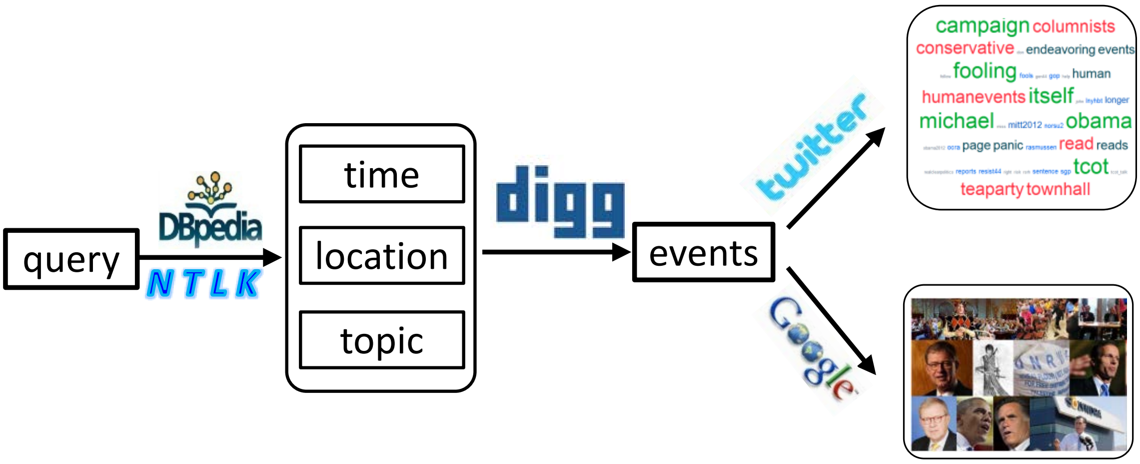

Fig. 2: Overview of event enrichment by textual query expansion

where $p(t) = \frac{weight(t)}{\sum_{x \in T} weight(x)}$. The tag cloud entropy quantifies the disparity of weights between tags. After ranking, the events with more information will be shown first and events with less information shown later.

3 Event Detection in Social Multimedia

Event detection tries to detect events from a social multimedia stream, and to identify event attributes such as taken time, location, content. With the advent of social media, event detection from social media data has gained popularity in the past decade. Generally, typical event discovery methods include clustering, statistical, probabilistic model based methods and some other approaches.

Cluster based method is one of the most important approaches to detect events and receives considerable attention. The event detection by clustering focuses on clustering social media data by visual or spatial-temporal feature and then identifies the cluster as event/non-event in the following process. For example, [48] presented a method to mine events and objects from community photo collections by clustering approaches. In their system, the photos are clustered according to several different modalities (visual and textual features), and the clusters are then classified as objects or events by their durance, since events are usually characterized by a short duration. A very similar framework is proposed to classify the events and landmarks in [49]. In [50, 14], the authors exploited the rich "context" associated with social media content and applied clustering algorithms to identify social events. In [51], the authors presented a scalable graph based multimodal clustering approach to detect events in large collections of multimedia. The proposed approach utilizes example relevant clusters to learn a model of the "the same event" relationship between two items in the multimodal domain and subsequently to organize the items in a graph. In [52], to detect social events, the authors proposed a multimodal clustering algorithm, which uses a known clustering in the currently examined domain to supervise the multimodal fusion and clustering procedure.

Statistical method for event detection involves finding the reproduced pattern among the time series data. Among the various methods that have been developed, static threshold method is the simplest and straightforward one. In this type of approach, events are found when the conditions which are required are met. For example, In [53], focusing on the problem of extracting place and event semantics from tags, a burst scan approach was proposed to extract semantics of tags, unstructured text labels assigned to resources on the web, based on each tag usage pattern. In [15], the authors tried to identify social events from photos stream by burst detection approach, based on the observation that many photos are captured and shared online. They analyzed geotagging information retrieved from online source to determine the bounding box for a set of venues, retrieved a set of photos in the determined bounding box, studied how to design the statistical value for detecting events and designed a sophistic rule to decide the threshold. In [54], to alleviate influence of noise data, the authors combine a translation-based method with a frequency-based method for keyword extraction to mine the user interests. The static threshold method could meet the timeline requirement and provides immediate output results, but it is hard to find out a proper value as the threshold point and fails to provide robust results.

Wavelet based spatial analysis is robust statistical method used in detecting events. In [55], a wavelet based approach was proposed to detect events from social media. At first, the temporal and spatial distributions of tag usage are analyzed by a wavelet transform to suppress noise, so that the tags could be identified if they are related with events. In [56], the authors proposed a method to detect events in Twitter by wavelet based analysis method. The method builds signals for individual words by applying wavelet analysis on the frequency based raw signals of the words, and then filters away the trivial words by studying their corresponding signal auto-correlations. The remaining words are then clustered to form events with a modularity based graph partitioning technique.

Probabilistic based approaches consist of these methods by which the event or other related probabilities are modeled with Bayes inference framework. In particular, the topic models such as latent semantic analysis (LSA) [57], and its probability representation (pLSA) [58] or latent Dirichlet Allocation(LDA) [59] have shown encouraging results in many problems. Specific to event detection, [60] developed a system to combine the popular LDA model with temporal segmentation and spatial clustering for identifying events automatically from a large document collection. In [61], the authors addressed the problem of discovering events from social media data by topic model method. Based on this assumption that the events are the conjoint distribution over the latent topics in a given place, topics are learned from amounts social data using a LDA model. Then, event distribution estimation over the topics is solved using least mean square optimization. In [62], the authors proposed a mixture Gaussian model for burst word extraction in Twitter and employ a novel time dependent hierarchical Dirichlet process (HDP) model [63] for new topic detection. The proposed model not only detects event from previous tweets stream, but also grasps new events. [64] suggested an alternative methodology for event detection using space time scan statistics (STSS). This technique first clusters the tweets by space and time feature, regardless of tweet content, and then evaluates whether clusters relate to space time events by LDA model. Besides these methods, there are several other approaches which are employed to solve event detection problem. For example, in [65], the authors presented a system that is able to classify a stream of social media data into a growing and evolving set of events. In the system, multiple features such as time, geotag, text and visual content are extracted to train an event based classification model. In the proposed system, the emerging new events also could be detected if the photo set is not recognized by the existing event models. In [66], based on the assumption that online social interaction reflects important signals among the participants on the "social affinity" of two photos, Wang et. al. employed an interaction graph to exploit the online social interaction among social items, and used the relationship feature to detect social events.

In addition, there are some other works that study how to exploit the external knowledge on event detection problem. The idea is popularly used in the social event detection task of MediaEval benchmark [11]. For example, in [67], in order to find all soccer clubs and associated stadiums for the given cities in the query, Dbpedia and GeoName [68] were used to extend the query topic and identify an location in the soccer events detection task. In [69], the authors exploited the knowledge from the websites such as last.fm, eventful.com to extract event groundtruth and covert the event detection problem as a matching problem. In [70], the authors investigated the relationship between the spatio-temporal co-occurrences and social ties in a location-based social network, and proposed a method for predicting cooccurrence based on the social ties and habits similarities.

In the following, we will discuss two works that address on mining social events from a media stream. In the first work [15], as much media data are uploaded when an event occurs, a burst detection approach is proposed to target events. In the second work [61], the event semantic is regarded as relevant with the latent topics in human life, and topic model is employed to make decision rule on identifying events.

3.1 Events Detection by Burst Analysis

At first, we review a method to discover events from social media stream by burst detection approach, which detects events by analyzing the uploading at a given place. The heuristic is quite straightforward: It is already known that during events many photos are taken by different people, so if the following two conditions: 1) many photos are taken at a location in a short period 2) many people take photos, meet, an event occurs. As depicted in Figure 3, there are mainly three steps in the whole framework. At first, media data is collected on a given location, and then the time series analysis technique is used to find out the event starting time. Finally, the found events are shown in a proper format.

The approach consists in carefully selecting the date with high number of uploadings and considers those candidate date as the event taken date. More formally, let us consider the time series $\{d_i, i \in [1, T]\}$ that represents the temporal evolution of the photo upload characteristic at a given venue v. The event e starting at time t is detected when the photo upload characteristic is greater than a given threshold.

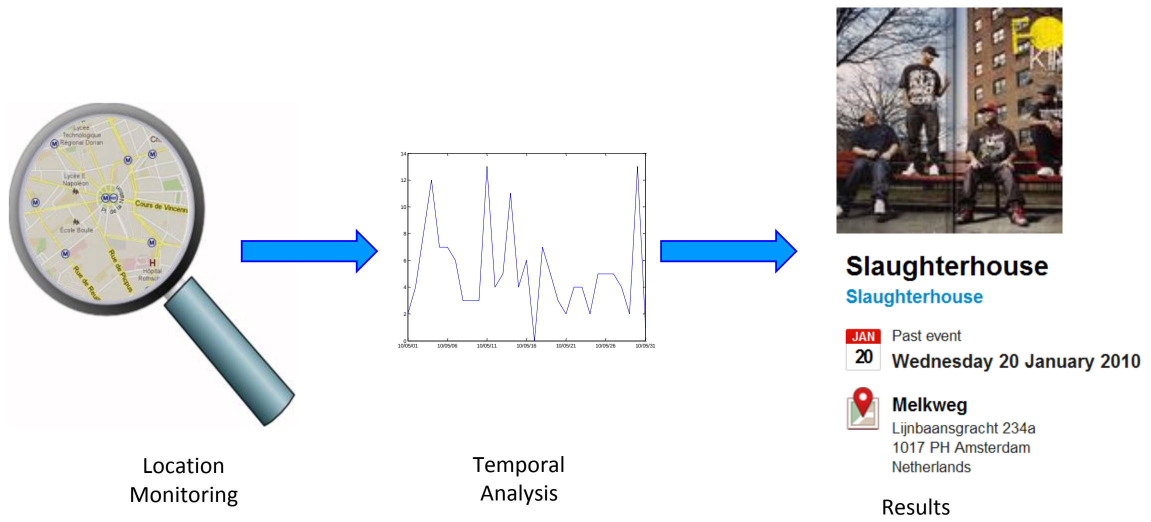

Fig. 3: Overview of event detection by burst detection method

$$e_t = \arg_i(d_i > THD) \tag{5}$$

The authors used different photo uploading characteristics in event detection task. The first one is the number of photo uploaded $d = N_p$. The second one attempts to incorporate the social dimension of the event and accounts for the number of different photo uploaders $d = N_u$. The third characteristic combines the previous ones by pondering the number of photo uploaded with the number of different people that uploaded these photos $d = N_p * N_u$.

As a burst detection method, the threshold plays the key role to the detection results. In this work, the authors thoroughly studied two different rules to choose the threshold and found out that the "median value" of the statistical characteristic is the best method to select a static threshold.

Once an event is identified, the tags of the corresponding photos are extracted to deduce the topic. Specifically, all words from tags and titles of the photos taken on that day are parsed and sorted by their occurrence. The top 15 keywords are kept to infer the topic of individual events. Detected events are manually matched with ground truth events based on date and title.

3.2 Event Detection by Latent Topic Analysis

In this section, we study the method of detecting events from social media by latent topic analysis [61]. It is well known that there are many different concepts in real world. For a given place (a city for example), the set of topics associated with a period of time is stationary. And the event semantic is formulated as a special distribution over these topics, and the event could be discovered by the distribution on these topics. As shown in Figure 4, first the topics are learned from large quantities of data captured at a given location. Then, the least mean square algorithm is employed to estimate the event distribution on a group of validated data samples. Finally the event is detected from a media dataset, if the data fit the distribution over latent topics well.

The method to infer topics among documents used in this work is the LDA model [59] which is a generative probabilistic graphical model to discover topics in documents. In the approach, a collection of geotagged Flickr photos for a given city are retrieved, and the stem words from the title and tags from each photo are used to train the LDA model. When the model is obtained, it is used to infer distribution over these topics on the validated data and to estimate the distribution of the event.

From the LDA model, the distribution of a document d over latent topics can be inferred by equation 6.

$$P(\theta_d|\alpha,\beta,d) = \int\int p(w,z,\theta|\alpha,\beta)dwdz \tag{6}$$

While w is the word vector of a document, z is the topic distribution of a document, and θ is the distribution of the topic.

Following, the distribution of the event over the inferred topics is estimated using validation data, which are the positive media documents of an event. Suppose

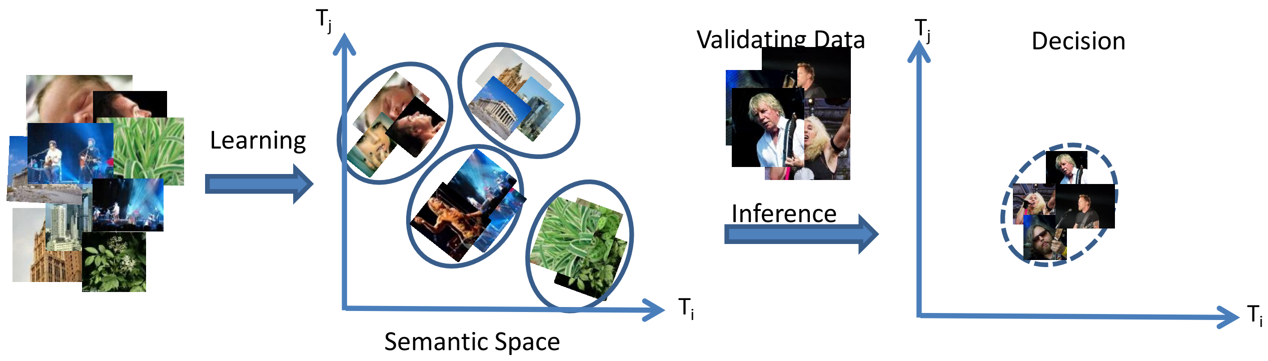

Fig. 4: Overview of event detection by latent topic analysis

D is the inference of the validation data over the latent topics, the event distribution $\mathbf{e}$ can be estimated by least mean square optimization. The objective is to minimize the following equation 7.

$$\mathbf{e} = \operatorname*{argmin}_{e \in R^N} \sum_i Dist(D_i, e) \tag{7}$$

Where the function $Dist$ measures the distance between a validating instance D_i and the event estimation $\mathbf{e}$ over the latent topics. The following standard symmetric KL divergence is used as the distance measure.

$$Dist(D_i, e) = D_{\mathrm{KL}}(D_i \| e) + D_{\mathrm{KL}}(e \| D_i) \tag{8}$$

When the event distribution is estimated from equation 7, it can be used to verify if a new document d is event related or not, according to the rule defined in equation 9.

$$d \;\; is \begin{cases} event, & \text{if } Dist(d, \mathbf{e}) \leq THD \\ noevent, \; otherwise \end{cases} \tag{9}$$

Where the value T is the threshold of the decision function, to decide if a document is relevant or not in the detection process. In practice, the value of T can be inferred from the validation dataset D as follows:

$$THD = k \max_i \{Dist(D_i, \mathbf{e})\} \tag{10}$$

4 Event Categorization in Social Multimedia

The event can be served as a powerful instrument to organize media, thanks to its intrinsically multi-faceted nature. Furthermore, it is the most natural way for human being to store and recall their memories. Categorizing media data by events has drawn much attention in the multimedia research community. Existing approaches on event categorization are based on single modality (e.g., text, visual content) information or multimodal information [35,71,17]. For example, In [17], to categorize social media data by events, the authors employed the naive Bayes method to exploit the social information produced by users in the form of tags, titles and photo descriptions, for classifying pictures into different event categories.

However, the single modality based methods cannot model events well, since the multimodal properties of social events are ignored. In recently years, more works focus on solving the problem by multimodal analysis [17,72]. For instance, the authors of [72] focused on how to assign media data to events. They presented a system to classify a stream of social media data into a growing and evolving set of events, while the similarity of events and media data are modeled by multimodal features. To scale up to the data sizes and data rates in social media applications, a candidate retrieval or blocking step is employed to reduce the number of events that are considered as potential candidates to which the incoming data point could belong to. In [73], the problem of images annotated by events was studied and authors proposed a hierarchical solution to annotate photos by the event semantic automatically using simple temporal cues. In [74], the authors exploited the users provided textual and visual contents associated with social events, and proposed a novel event classification algorithm, which is able to model multi-modality and supervised category label information jointly. While these works model the similarity of social media on the basis of text/time, or the similarity between events and social media with respect to time, location or text features, the importance and effectiveness of these features has not been studied. In [18], the authors addressed the problem of categorizing media data by events, and investigated how to select the representative features and to incorporate the missing attributes in the system. In this paper, different learning approaches are employed to train the event models on different features, and it is found that through experiments on a large set of events, the best discriminant features are tags, spatial and temporal feature. Following, we will detail the work in [17] and in [18] that categorize media by events with textual and multi-modality analysis methods respectively.

4.1 Event Based Media Categorization by Textual Analysis

To organize media effectively and match the users' information, the authors of [17] proposed a framework to categorize media by events based on textual analysis.

In their work, a large event/image set is first created in two steps. To collect events, the YAGO ontology [75] is used to find out the Wikipedia entities having a type "wordnet_event". With this method a list of 138,794 Wikipedia events is retrieved. To collect images, starting from the Wikipedia event page, first the corresponding Flickr groups are crawled by Flickr API. In Flickr, the group is created by the users to represent their interesting in a specific event, thus it could be considered as trustful ground truth. Thanks to the Flickr API, about 30 thousand event related groups are retrieved. Following, for each retrieved group, all images contributed by the group members are collected. The tags from these pictures are extracted and formulated by tf-idf method to represent the pictures. To sum up, the 138,794 events, 2,639,254 pictures, along with 22 Million tags have been gathered.

Following, it is found that there is duplication in the raw event/image set, and the clustering method is used to group the images from duplicated group together in two steps. The first clustering method is performed based on the Wikipedia classes. In this case, all pictures belonging to the Flickr groups having the same Wikipedia class are merged into one cluster. The second clustering is made use of the WordNet classes. Similar to clustering based on Wikipedia classes, all pictures belonging to the Flickr groups having the same WordNet class are merged into one cluster. In addition, since enough data is needed for training the event classifiers, a cleaning step is performed to remove the group clusters not having sufficient photo instances.

Finally, for each event category, a classifier is trained by the Naive Bayes multinomial method implemented in Weka [76]. When training the classifier, the positive examples are represented by the pictures gathered from the corresponding event group, while the negative ones are randomly selected from the pictures corresponding to the rest of the event classes. The number of positive and negative examples is almost equally balanced. For evaluating the performance, the classification accuracy (Acc), precision (P) and recall (R) measures are inspected, when performing 10-fold cross validation on the data set. The promising results of the final evaluation show that proposed event based classification is feasible and confirm the quality of the user provided tags.

4.2 Event Based Media Categorization by Multimodal Fusion

In [18], the event based media classification by multimodal fusion method is investigated. Specifically, the authors studied two fundamental questions that are yet not addressed by previous work: the feature and model selection, as well as handling the issue of potential missing value. These problems are tackled within an event based photo categorization framework. For each event in the dataset, a classification model is trained using the photos originating from that event. To evaluate the effectiveness of different learning approaches and features for modeling events, some supervised learning methods, like KNN, SVM [77], Decision Tree [78] and Random Forest [79], are employed to learn the models based on temporal (date and time), location (geo-coordinates), tag (annotations) and visual features. These features are extracted as follows.

- **Temporal feature** Time is one of the most important components of an event. The temporal feature used in this work is the photo taken time, which is represented by the number of past seconds from Unix epoch. The taken time is compulsory in photo metadata.
- **Location feature** Nowadays, geographical metadata is a common component in social media [80]. The GPS metadata, that is the latitude and longitude coordinate, is extracted as the location feature. GPS information is not always available in photos metadata. To cope with missing value, the method proposed in [81] is employed: the feature vector is filled with zero if the value is missing while a binary flag is added to indicate availability or not to indicate whether the feature value is missing or present.
- **Tag Feature** The Boolean weighting scheme is employed to measure the term's frequency of tags [82]. In detail, for each event a word vocabulary with the 200 most frequent tags is created and the tags in

each photo are projected on the vocabulary, creating a vector. Each dimension in the vector corresponds to a separate term. If a term occurs in the document, its value in the vector is 1, or 0 otherwise. Tag metadata is also not compulsory and the same strategy as on location feature is used to handle missing value. Hence, the tag feature is a 201D vector.

– **Visual Feature** Visual features are also representative for the photo content. In this work, multiple low level visual features that have been popularly used in visual content analysis [83] such as: 64D color histogram, 73D edge histogram and 64D Gabor features are used. The three visual features are concatenated into 201D and normalized during preprocessing.

The authors evaluated the event based media categorizing work on a large scale image dataset. To collect the dataset, the event dataset EventMedia, created by Troncy et al [13] using the linking data techniques is exploited. In EventMedia, the events originate from three large public event repositories (last.fm, eventful and upcoming) and media data connected by the event machine tag are crawled from social media sharing platform such as Flickr or twitter. There are about 1M events in this corpus, illustrated with 1.7M photos. The data is saved in RDF format and can be queried through a SPARQL entry point [84]. Since sufficient exemplars are required for training and testing, the events with at least 40 photos labeled with location metadata are collected. In EventMedia, there are 674 events which meet this condition and are used as the event collection, along with the associated 92K photos.

For each event, a classification model is trained on the positive and negative samples. Building the model on an event basis allows adding new events without affecting previously learned models and reduces the impact of the increasing of events in the dataset. The positive examples of an event are the pictures originating from the event, while the negative ones are randomly selected from the pictures corresponding to the remaining events in the dataset. From the event dataset, the authors randomly select 100 events to train the 1-vs-all classifiers with different learning approaches and features. The area under ROC and PR curve [82] is used as the evaluation criteria.

From the result, some conclusion is conducted. First, the tag is the most representative feature when modeling event, followed closely by the spatial-temporal feature. In addition, the combination of spatial-temporal and tag feature obtains the best performance overall. Second, on modeling the photos features which are very parse and with missing value, the Decision Tree and KNN methods obtain better performance compared with SVM, while the previous approaches are designed to deal with problems with irregular decision boundary.

5 Datasets and Benchmarks for Event Analysis

The developing event analysis methods call for evaluation on large dataset. While some datasets that are popularly used in the research are listed as follows.

5.1 Upcoming Event dataset

The Upcoming dataset [14] contains 9,515 events taken between January 1, 2006, and December 31, 2008. From each event, the images crawled from Flickr tagged with an upcoming event machine tag are also provided. In the dataset, there are around 270k photos, which means on average 28.42 photos per event. The dataset is exploited in social event identification task[66, 14].

5.2 MediaEval Event dataset

The MediaEval social event detection task is an annual forum in social event analysis community. A dataset and several event detection tasks are released every year [85, 11]. For example, in 2014 there are two tasks in MediaEval social event detection: to cluster media by events and to retrieval events by its attributes. The organizers also release two datasets: The first dataset contains 362,578 images which are grouped into 17,834 clusters that represent social events. The second dataset contains 110,541 images but no labeled information is provided. Both dataset are comprised of images collected from Flickr using the Flickr API. In addition, for both datasets the actual image files and their metadata such as username of the uploader, date taken, date uploaded, title, description, tags and geo-location, are made available. All images are covered by a Creative Commons license.

5.3 TrecVid Event dataset

The Multimedia Event Detection (MED) evaluation [86] track is a part of the annual TRECVID Evaluation [87]. The task starts from 2010 and the goal of MED is to assemble core detection technologies into a system that can search multimedia recordings for user defined events based on pre-computed metadata. A collection of Internet multimedia (i.e., clips containing both audio and video streams) is provided to the MED participants. The data consists of publicly available, user generated content posted to the various Internet video hosting sites, such YouTube, DailyMotion [88]. The clips are provided in MPEG-4 formatted files, while the video is encoded to the H.264 standard and audio is encoded using MPEG-4'S Advanced Audio Coding (AAC) standard. The latest dataset contains about 244,000 video clips in 9,911 hours.

6 Open Issues and Challenges

Though serveral exciting achievement in social event analysis is accomplished, there is still some challenging on this problem. Below we list a few important aspects which can be a part of future research.

1. **Personal Events Analysis**
   In media shared web sites, many media data are taken during personal events with less and closed people, such as birthday, wedding. Analyzing personal photo/video albums for understanding the conveyed events is an emerging trend. It is very helpful to discover personal events and find their components in order to offer support for organizing and sharing the event related information. The exploit of the personal events could progress the event based research greatly. How to study the personal events is still an open challenge.
2. **Event Analysis in Big Data**
   Big data becomes a new trend in research community. With the rapid increasing of social media data, there is growing demand for solutions to analyze event relevant information. How to apply the traditional event analysis method to big data is new challenge both in research and industry.
3. **Event Recommendation**
   Event recommendation could help users organize, participate, comment and share offline events such as cocktail parties, seminars and concerts. It plays an important role in recommending the most relative events to users who are likely to participate. The main challenge on event recommendation is how to model the heterogeneous social relation. The event recommendation service is an essential topic in online event system.

7 Conclusion

We have presented a comprehensive survey on the emerging research topic of the event based social media analysis. In this survey, we highlight the current progress on this topic, discuss a number of representative works on event illustration, event detection and event based categorization on social media, and show some future directions alongside. We hold that the research topic will attract more attention in the future, while the event is utilized as the center role in social media retrieval and organization system.

References

1. Mor Naaman. Social multimedia: highlighting opportunities for search and mining of multimedia data in social media applications. *Multimedia Tools and Applications*, 56(1):9–34, 2012.
2. Jiebo Luo, Dhiraj Joshi, Jie Yu, and Andrew Gallagher. Geotagging in multimedia and computer vision - a survey. *Multimedia Tools and Applications*, 51(1):187–211, 2011.
3. Alessandro Vinciarelli, Maja Pantic, and Herve Bourlard. Social signal processing: Survey of an emerging domain. *Image and Vision Computing*, 27(12):1743 – 1759, 2009.
4. Tao Mei, Yong Rui, Shipeng Li, and Qi Tian. Multimedia search reranking: A literature survey. *ACM Comput. Surv.*, 46(3):38:1–38:38, January 2014.
5. Meng Wang, Bingbing Ni, Xian-Sheng Hua, and Tat-Seng Chua. Assistive tagging: A survey of multimedia tagging with human-computer joint exploration. *ACM Comput. Surv.*, 44(4):25:1–25:24, September 2012.
6. Qiang Yang. Three challenges in data mining. *Frontiers of Computer Science in China*, 4(3):324–333, 2010.
7. Haixin Ma, Weining Qian, Fan Xia, Xiaofeng He, Jun Xu, and Aoying Zhou. Towards modeling popularity of microblogs. *Frontiers of Computer Science*, 7(2):171–184, 2013.
8. Utz Westermann and Ramesh Jain. Toward a Common Event Model for Multimedia Applications. *IEEE MultiMedia*, 14(1):19–29, 2007.
9. Xueliang Liu, Raphaël Troncy, and Benoit Huet. Finding Media Illustrating Events. In *ACM International Conference on Multimedia Retrieval*, Trento, Italy, 2011.
10. Ansgar Scherp, Ramesh Jain, Mohan Kankanhalli, and Vasileios Mezaris. Modeling, detecting, and processing events in multimedia. In *Proceedings of the International Conference on Multimedia*, MM '10, pages 1739–1740, 2010.
11. Georgios Petkos, Symeon Papadopoulos, Vasileios Mezaris, Raphael Troncy, Philipp Cimiano, Timo Reuter, and Yiannis Kompatsiaris. Social event detection at mediaeval: a three-year retrospect of tasks and results. In *Proc. ACM ICMR 2014 Workshop on Social Events in Web Multimedia*, 2014.
12. Timo Reuter, Symeon Papadopoulos, Giorgos Petkos, Vasileios Mezaris, Yiannis Kompatsiaris, Philipp Cimiano, Christopher de Vries, and Shlomo Geva. Social event detection at mediaeval 2013: Challenges, datasets, and evaluation. In *Proceedings of the MediaEval 2013 Multimedia Benchmark Workshop*, 2013.
13. Troncy, Raphaël and Malocha, Bartosz and Fialho, André T. S. Linking Events with Media. In *6th International Conference on Semantic Systems*, Graz, Austria, 2010.
14. Hila Becker, Mor Naaman, and Luis Gravano. Learning similarity metrics for event identification in social media. In *Web Search and Data Mining*, 2010.
15. Xueliang Liu, Raphaël Troncy, and Benoit Huet. Using social media to identify events. In *Proceedings of the ACM SIGMM Workshop on Social Media*, WSM '11, pages 3–8, 2011.

16. Chiraz Trabelsi and SadokB. Yahia. A probabilistic approach for events identification from social media rss feeds. In *Database Systems for Advanced Applications*, volume 7827, pages 139–152. 2013.
17. Claudiu S Firan, Mihai Georgescu, Wolfgang Nejdl, and Raluca Paiu. Bringing order to your photos: Event-Driven Classification of Flickr Images Based on Social Knowledge. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, page 189, October 2010.
18. Xueliang Liu and Benoit Huet. Heterogeneous features and model selection for event-based media classification. In *International Conference on Multimedia Retrieval, Dallas, TX, USA, April 16-19, 2013*, pages 151–158.
19. Daniel Billsus and Michael J. Pazzani. A hybrid user model for news story classification. In *Proceedings of the seventh international conference on User modeling*, pages 99–108, June 1999.
20. Hiroyuki Toda and Ryoji Kataoka. A clustering method for news articles retrieval system. In *the international conference on World Wide Web*, page 988, May 2005.
21. Diogo Delgado, João Magalhães, and Nuno Correia. Assisted News Reading with Automated Illustrations. In *ACM conference on Multimedia*, pages 1647–1650, 2010.
22. Dhiraj Joshi, James Z Wang, and Jia Li. The Story Picturing Engine—a system for automatic text illustration. *ACM Transactions on Multimedia Computing Communications and Applications*, 2(1):68–89, 2006.
23. Matthew Cooper, Jonathan Foote, Andreas Girgensohn, and Lynn Wilcox. Temporal event clustering for digital photo collections. In *ACM International Conference on Multimedia*, 2003.
24. Adrian Graham, Hector Garcia-Molina, Andreas Paepcke, and Terry Winograd. Time as essence for photo browsing through personal digital libraries. In *Proceedings of the ACM/IEEE-CS Joint Conference on Digital Libraries*, JCDL '02, pages 326–335, 2002.
25. Brendan Jou, Hongzhi Li, Joseph G. Ellis, Daniel Morozoff-Abegauz, and Shih-Fu Chang. Structured exploration of who, what, when, and where in heterogeneous multimedia news sources. In *Proceedings of the ACM International Conference on Multimedia*, pages 357–360, 2013.
26. Minkyoung Kim, Lexing Xie, and Peter Christen. Event diffusion patterns in social media. In *International AAAI Conference on Weblogs and Social Media*, May, 2012.
27. K. Burton, N. Kasch, and I. Soboroff. The icwsm 2011 spinn3r dataset. In *the Fifth Annual Conference on Weblogs and Social Media*, 2011.
28. Riccardo Mattivi, Jasper Uijlings, Francesco De Natale, and Nicu Sebe. Categorization of a collection of pictures into structured events. In *Proceedings of the 2nd ACM International Conference on Multimedia Retrieval*, page 1, June 2012.
29. Riccardo Mattivi, Jasper Uijlings, Francesco G.B. De Natale, and Nicu Sebe. Exploitation of time constraints for (sub-)event recognition. In *Proceedings of the 2011 Joint ACM Workshop on Modeling and Representing Events*, pages 7–12. ACM, 2011.
30. Google Picasa. http://picasa.google.com.
31. Flickr. http://www.flickr.com.
32. Matthew Cooper, Jonathan Foote, Andreas Girgensohn, and Lynn Wilcox. Temporal event clustering for digital photo collections. *ACM Trans. Multimedia Comput. Commun. Appl.*, 1(3):269–288, August 2005.
33. Pinaki Sinha and R. Jain. Extractive summarization of personal photos from life events. In *IEEE International Conference on Multimedia and Expo*, pages 1–6, July 2011.
34. Lyndon Kennedy and Mor Naaman. Less talk, more rock: automated organization of community-contributed collections of concert videos. In *18th ACM International Conference on World Wide Web*, pages 311–320, Madrid, Spain, 2009.
35. Nicholas Diakopoulos, Mor Naaman, and Funda Kivran-Swaine. Diamonds in the rough: Social media visual analytics for journalistic inquiry. In *2010 IEEE Symposium on Visual Analytics Science and Technology*, pages 115–122, October 2010.
36. Mingyan Gao, Xian-Sheng Hua, and Ramesh Jain. WonderWhat: Real-time Event Determination from Photos. In *20th World Wide Web Conference*, Hyderabad, India, 2011.
37. Xueliang Liu and Benoît Huet. Event representation and visualization from social media. In *Advances in Multimedia Information Processing–PCM 2013*, pages 740–749. Springer International Publishing, 2013.
38. Jiliang Tang, Xufei Wang, Huiji Gao, Xia Hu, and Huan Liu. Enriching short text representation in microblog for clustering. *Frontiers of Computer Science*, 6(1):88–101, 2012.
39. EventBurn. http://www.eventburn.com/.
40. Xiaolong Zheng, Yongguang Zhong, Daniel Zeng, and Fei-Yue Wang. Social influence and spread dynamics in social networks. *Frontiers of Computer Science*, 6(5):611–620, 2012.
41. Youtube. http://www.youtube.com.
42. Xueliang Liu and Benoit Huet. Automatic concept detector refinement for large-scale video semantic annotation. In *IEEE Fourth International Conference on Semantic Computing*, pages 97–100, 2010.
43. Steven Bird, Ewan Klein, and Edward Loper. *Natural Language Processing with Python*. O'Reilly Media, Inc., 2009.
44. Soren Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, and Zachary Ives. Dbpedia: A nucleus for a web of open data. In *In 6th International Semantic Web Conference, Busan, Korea*, pages 11–15, 2007.
45. Sun Chengjie and Guan Yi. A Statistical Approach for Content Extraction from Web Page. *Journal of Chinese Information Processing*, 18(5):17–22, 2004.
46. Tao Mei, Bo Yang, Shi-Qiang Yang, and Xian-Sheng Hua. Video Collage: Presenting a Video Sequence Using a Single Image. *The Visual Computer*, 25(1):39–51, August 2009.
47. Kamel Aouiche, Daniel Lemire, and Robert Godin. Web 2.0 OLAP: From Data Cubes to Tag Clouds. *Lecture Notes in Business Information Processing*, 2009.
48. Till Quack, Bastian Leibe, and Luc Van Gool. World-scale mining of objects and events from community photo collections. In *Proceedings of the 2008 international conference on Content-based image and video retrieval*, page 47, July 2008.
49. S Papadopoulos, C Zigkolis, Y Kompatsiaris, and A Vakali. Cluster-Based Landmark and Event Detection for Tagged Photo Collections. *IEEE Multimedia*, 18(1):52–63, 2011.
50. Hila Becker, Mor Naaman, and Luis Gravano. Event Identification in Social Media. In *12th International Workshop on the Web and Databases*, Providence, USA, 2009.
51. Georgios Petkos, Symeon Papadopoulos, Emmanouil Schinas, and Yiannis Kompatsiaris. Graph-based multimodal clustering for social event detection in large collections of images. In *MultiMedia Modeling*, volume 8325 of *Lecture Notes in Computer Science*, pages 146–158. Springer International Publishing, 2014.
52. Georgios Petkos, Symeon Papadopoulos, and Yiannis Kompatsiaris. Social event detection using multimodal clustering and integrating supervisory signals. In *Proceedings of the 2Nd ACM International Conference on Multimedia Retrieval*, ICMR '12, pages 23:1–23:8, 2012.

53. Tye Rattenbury, Nathaniel Good, and Mor Naaman. Towards automatic extraction of event and place semantics from flickr tags. In *Proceedings of ACM SIGIR conference on Research and development in information retrieval*, page 103, July 2007.
54. Zhiyuan Liu, Xinxiong Chen, and Maosong Sun. Mining the interests of chinese microbloggers via keyword extraction. *Frontiers of Computer Science*, 6(1):76–87, 2012.
55. Ling Chen and Abhishek Roy. Event detection from flickr data through wavelet-based spatial analysis. In *ACM conference on CIKM*, 2009.
56. Jianshu Weng and Bu-Sung Lee. Event detection in twitter. In *International AAAI Conference on Weblogs and Social Media*, 2011.
57. Scott Deerwester, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard Harshman. Indexing by latent semantic analysis. *JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE*, 41(6):391–407, 1990.
58. Thomas Hofmann. Probabilistic latent semantic indexing. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 50–57, August 1999.
59. David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3(4-5):993–1022, May 2003.
60. Chi-Chun Pan and Prasenjit Mitra. Event detection with spatial latent Dirichlet allocation. In *Proceeding of the 11th annual international ACM/IEEE joint conference on Digital libraries*, page 349, June 2011.
61. Xueliang Liu and Benoit Huet. Social event discovery by topic inference. In *Image Analysis for Multimedia Interactive Services (WIAMIS), International Workshop on*, pages 1–4. IEEE, 2012.
62. Xun Wang, Feida Zhu, Jing Jiang, and Sujian Li. Real time event detection in twitter. In Jianyong Wang, Hui Xiong, Yoshiharu Ishikawa, Jianliang Xu, and Junfeng Zhou, editors, *Web-Age Information Management*, Lecture Notes in Computer Science, pages 502–513. Springer Berlin Heidelberg, 2013.
63. Yee Whye Teh, Michael I. Jordan, Matthew J. Beal, and David M. Blei. Hierarchical dirichlet processes. *Journal of the American Statistical Association*, 101, 2004.
64. Tao Cheng and Thomas Wicks. Event detection using twitter: A spatio-temporal approach. *PLoS ONE*, 9(6):e97807, 06 2014.
65. Timo Reuter and Philipp Cimiano. Event-based classification of social media streams. In *Proceedings of the 2nd ACM International Conference on Multimedia Retrieval*, ICMR '12, pages 22:1–22:8, 2012.
66. Yanxiang Wang, Hari Sundaram, and Lexing Xie. Social event detection with interaction graph modeling. In *Proceedings of the 20th ACM International Conference on Multimedia*, MM '12, pages 865–868, 2012.
67. Markus Brenner and Ebroul Izquierdo. Social event detection and retrieval in collaborative photo collections. In *Proceedings of the 2nd ACM International Conference on Multimedia Retrieval*, pages 21:1–21:8, 2012.
68. GeoNames. http://geonames.org/.
69. Xueliang Liu, Benoit Huet, and Raphaël Troncy. Eurecom mediaeval 2011 social event detection task. In *MediaEval*, 2011.
70. Rong-Hua Li, Jianquan Liu, JeffreyXu Yu, Hanxiong Chen, and Hiroyuki Kitagawa. Co-occurrence prediction in a large location-based social network. *Frontiers of Computer Science*, 7(2):185–194, 2013.
71. J. Makkonen, H. Ahonen-Myka, and M. Salmenkivi. Simple semantics in topic detection and tracking. *Information retrieval*, 3(7):347–368, 2004.
72. Timo Reuter and Philipp Cimiano. Event-based classification of social media streams. In *Proceedings of the 2Nd ACM International Conference on Multimedia Retrieval*, pages 22:1–22:8, 2012.
73. Jia-Min Gu, Yi-Leh Wu, Wei-Chih Hung, and Cheng-Yuan Tang. Personal photo organization using event annotation. In *Information, Communications and Signal Processing (ICICS) 2013 9th International Conference on*, pages 1–4, Dec 2013.
74. Shengsheng Qian, Tianzhu Zhang, and Changsheng Xu. Multi-modal supervised latent dirichlet allocation for event classification in social media. In *Proceedings of International Conference on Internet Multimedia Computing and Service*, ICIMCS '14, pages 152:152–152:157, 2014.
75. Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. Yago: A large ontology from wikipedia and wordnet. *Web Semantics: Science, Services and Agents on the World Wide Web*, 6(3):203 – 217, 2008.
76. Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. The weka data mining software: An update. *SIGKDD Explor. Newsl.*, 11(1):10–18, November 2009.
77. Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3):27:1–27, 2011.
78. L. Breiman, J. Friedman, R. Olshen, and C. Stone. *Classification and Regression Trees*. Wadsworth and Brooks, Monterey, CA, 1984.
79. Leo Breiman. Random forests. *Machine Learning*, 45:5–32, 2001.
80. J Hays and A A Efros. IM2GPS: estimating geographic information from a single image. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.
81. Hung-Yi Lo, Kai-Wei Chang, Shang-Tse Chen, and et al. An ensemble of three classifiers for kdd cup 2009: Expanded linear model, heterogeneous boosting, and selective naive bayes. *Journal of Machine Learning Research - Proceedings Track*, 7:57–64, 2009.
82. Christopher D Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval*. 1 edition, July 2008.
83. Tat-Seng Chua, Jinhui Tang, Richang Hong, Haojie Li, Zhiping Luo, and Yan-Tao. Zheng. NUS-WIDE: A Real-World Web Image Database from National University of Singapore. In *Proc. of ACM Conf. on Image and Video Retrieval*, Santorini, Greece, 2009.
84. John Hebeler, Matthew Fisher, Ryan Blace, and Andrew Perez-Lopez. *Semantic Web Programming*. Wiley Publishing, 2009.
85. Georgios Petkos, Symeon Papadopoulos, Vasileios Mezaris, and Yiannis Kompatsiaris. Social event detection at mediaeval 2014: Challenges, datasets, and evaluation. In *MediaEval*, 2013.
86. Jingen Liu, Hui Cheng, Omar Javed, Qian Yu, Ishani Chakraborty, Weiyu Zhang, Ajay Divakaran, Harpreet S Sawhney, James Allan, R Manmatha, et al. Multimedia event detection and recounting. 2013.
87. Paul Over, George Awad, Martial Michel, Jonathan Fiscus, Greg Sanders, Wessel Kraaij, Alan F. Smeaton, and Georges Quvenot. Trecvid 2014 – an overview of the goals, tasks, data, evaluation mechanisms and metrics. In *Proceedings of TRECVID 2014*. NIST, USA, 2014.
88. Dailymotion. http://www.dailymotion.fr.