Multimedia Tool and Application manuscript No.
(will be inserted by the editor)

Xueliang Liu* · **Benoit Huet**

Event-based Cross Media Question Answering

Received: date / Accepted: date

Abstract User generated content, available in massive amounts on the Internet, is receiving increased attention due to its many potential applications. One of such applications is the representation of events using multimedia data. In this paper, an event-based cross media question answering system, which retrieves and summarizes events on a given topic is proposed. In orther words, we present a framework for leveraging social media data to extract and illustrate social events automatically on any given query. The system is built in three steps. First, the input query is parsed semantically to identify the topic, location, and time information related to the News of interest. Then, we use the parsed information to mine the latest and hottest related News from social news web services. Third, to identify a unique event, we model the News content by latent Dirichlet Allocation and cluster the News using the DBSCAN algorithm. In the end, for each event, we retrieve both textual and visual content of News that refer the same event. The resulting documents are shown within a vivid interface featuring both event description, tag cloud and photo collage.

Keywords events, social media, illustration, cross media, question answering

* the work was performed while the first author was working towards his PhD at EURECOM.

X Liu
Hefei University of Technology, China
E-mail: liuxueliang@hfut.edu.cn

B Huet
EURECOM, Sophia-Antipolis, France
Tel.:+334-9300-8179
Fax: +334-9300-8200
E-mail: benoit.huet@eurecom.fr

1 Introduction

The rapid development of Web 2.0 technologies has led to the surge of research activities using the rapidly growing social media data. How to leverage the explosion of this vast amount of data to benefit web users at large is, however, still an open and challenging problem. An event is one of the most important cues for people to recall activities, whether private such as birthdays and weddings, or public such as concerts and world news. The reminder value of an event makes it extremely helpful for organizing data[24, 29]. As a result, event-based media analysis has recently drawn much attention within the multimedia research community.

In this paper, we address the problem of retrieving and summarizing events from social media data, and propose a novel event-based cross media question answering framework to extract and illustrate social events automatically from a given text query. To solve the problem, we use natural language processing algorithm to parse the input query semantically and extract the most popular events from social news web service. Then, the obtained events are depicted through a multi-modal faceted representation composed of a textual description, a tag cloud and a photo collage, providing the viewer with a rich informative panel about events.

Fig. 1: The snapshot of a query example results

The novelty of our proposed framework is twofold. First, as opposed with other methods that mine events from social media data directly [26], we cluster events from social news website such as Google News. This has considerable advantages: lower additional time cost, computation and storage are needed and the results can be achieved in real time. Furthermore, the powerful Google search engine can be used to find the most popularly event

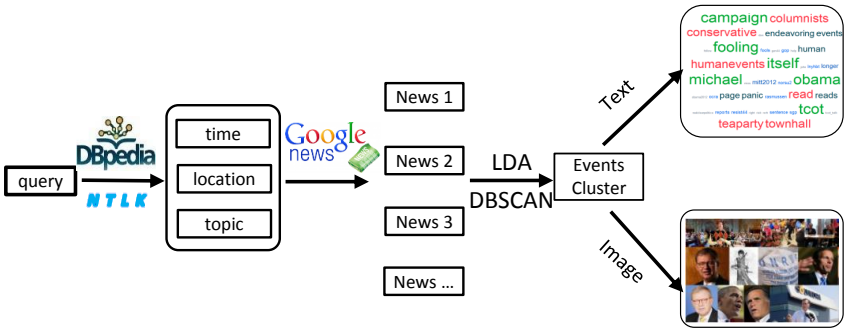

Fig. 2: Overview of the proposed framework

candidates. Secondly, we collect relevant media documents from different media sources and use tag cloud and photo collage to provide vivid interface illustrating as visually as possible each event. This enables content from different view to be shown at the same time, broadening the information available for viewers. The result is generated automatically on any given query. As an example, a subset of the events for a given query "New York in the past three days" can be found in Figure 1.

This paper is organized as follows: we review the related work in Section 2, and describe our proposed method in Section 3. Experimental results are presented in Section 4. Finally, we summarize the paper and propose future work in Section 5.

2 Related Work

The rapid development of Web 2.0 technologies has led to the surge of research activities using the rapidly growing social media data. How to leverage the explosion of this vast amount of data to benefit web users has drawn intensive attention [25,9,13,28]. Especially, Twitter has become one of the most important tools for people to share their interest, their personal issues, their views and their experiences as well as comment on other's or chat. Some research has been done to find events directly from Twitter post [26, 22]. In [26], the authors studied how to employ a wavelet-based techniques to detect events from Twitter stream. A similar method can be found in [5] to detect events from Flickr time series data. In [22], the authors investigate how to filter the tweets to detect seismic activity as it happens. In [3], a system is proposed to detect emerging topics from social streams and illustrate the topics with the corresponding information in multiple modalities. In [14], the authors proposed a method to detect events from Flickr dataset by modeling multi-modality features. Quack *et al.* [21] presented methods to mine events and object from community photo collections by clustering approaches. In [5], a wavelet based approach is proposed to detect events from social media data. Pan and Mitra [20] developed a system to combine the

popular LDA model with temporal segmentation and spatial clustering for automatically identifying events from a large document collection. In [8], the authors focused on building a Naive Bayes event models which classify photos as either relevant or irrelevant to given events. Illustrating events with media addresses the problem of how to leverage vivid multi-modal content to share experience. In [16], the authors studied users' uploading behaviors on Flickr and matched concert events with photos based on different modalities; such as text/tags, time, and geo-location. It results in an enriched photo set which better illustrates events. A similar framework involving more modalities is proposed in [4] for enriching event descriptions. In [7], to improve the users' attention when reading news articles, a system was proposed to help people reading news by illustrating the news story. In [11], an unsupervised approach was presented to describe stories with automatically collected pictures. In this framework, semantic keywords are extracted from the story, and used to search an annotated image database. Then a novel image ranking scheme automatically choose the most important images. In [30], a Text-to-Picture system was developed that synthesizes a picture from natural language text without limitation. The system firstly identified "picturable" text units by natural language processing, then searched for the most likely image parts conditioned on the text, and finally optimized picture layout conditioned on both the text and image parts. Our framework can be considered as a cross media Question Answering (QA) system focusing on events. Multimedia query answering [10,19] involves image, video and audio QA, which aims to return precise images, video clips, or audio fragments as answers to users questions. In [12], a system is proposed to leverage YouTube video collections as a source of reference to fulfill the task of presenting precise reference videos based on user's question. The summarization [27] function of our proposal is quite similar as the web service EventBurn[1], which could create a summary of a given hot event from data collected on popular services like Twitter, Facebook, and Flickr. However, as opposed to our system, it fails to extract events automatically from social media streams.

3 Our Proposal

Our framework, as shown in Figure 2, extracts and illustrates public events by leveraging the social media data directly. Since events can be defined as something happening at a given location and time, we start by parsing the query input to identify its topic, location and time information using the natural language processing algorithm. Rather than detecting events from the Twitter stream directly, events can be obtained by searching, crawling and scraping social news web services, which is an efficient way that saves time, computation and storage compared to alternative detection processes. In the retrieved results, it can be observed that there are much news that originates from different sources and illustrates the events from different viewpoints. To provide a more comprehensive review of the results, we employ Latent Dirichlet Allocation algorithm to mine the latent topic from these news, and

[1] http://www.eventburn.com/

then the news are clustered by DBSCAN approach. We developed a system to demonstrate the proposed framework. Each event is depicted with a text cloud and photo collage generated from the text and images found in the news source.

3.1 Semantic Query Parsing

As defined in [1], an event refers to a specific thing that happens at a specific period and place. The three basic properties of events are location, time, and topic, as stated in [16]. To identify the meaning of a given query input, we would like to extract the information in those three dimensions. Here, we assume that the query input is a noun phrase headed or tailed with complements, such as for example "New York in the past three days". We extract the structured data from this noun phrase, where there is a predictable organization of entities and relationships. The question 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*[2], a well-known *NLP* package. First, the input text is segmented into words using a tokenizer. Then, each word is tagged with part-of-speech tag (*POS*), which provides the lexical categories for words. With the *POS* tag, we use the *RegexpParser chunker* in *NLTK* to divide the query string in syntactically correlated groups of words, and generate the chunker tree to identify each sub-noun phrases for the input string. The process is depicted in Figure 3.

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 through a query of the DBpedia[3] database, which provides structured information extracted from Wikipedia. If geographical metadata, such as "geo:point", "geo:area", are found for a noun phrase, the corresponding location is kept as reference for the events to collect. For determining the temporal information, we develop a script which parses and converts the human readable string, such as "tomorrow","last week", "Monday" to a time structure. We use the DBpedia API and our script to parse the sub noun phrases in order to obtain the location and time information addressed by the query. Since it is hard to model the topic in sophisticated way, we assign the nouns as the topic keywords of the event to search for, if neither time or location can be determined from it.

3.2 News Extraction

Recently, there has been some research focusing on detecting events from social media data, such as detecting events from Twitter stream [26], or Flickr [5,15]. Indeed, useful information can be mined from community-contributed data. However, in these methods, huge amount of data have

[2] http://nltk.org/
[3] http://dbpedia.org

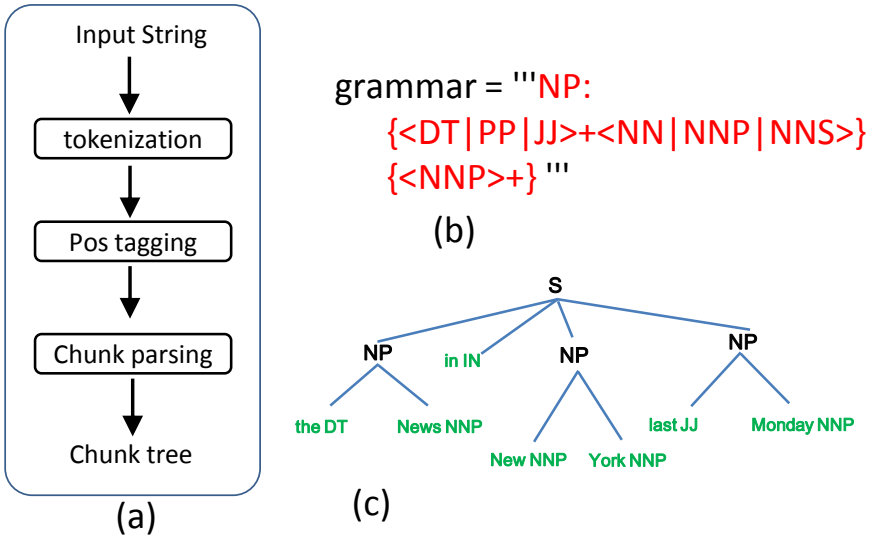

Fig. 3: Semantic parsing using NLP. (a) the flowchart; (b) the grammar used for chunker parsing; (c) an example of Chunk Tree for input query "the News in New York last Monday";

to be downloaded from the web service. Therefore, it takes lots of resources to store the data and lot of time to process it.

Currently, the world's happenings are collected as News by communication services and broadcasted publicly to massive audiences through various channels. Some web services, such as Google News[4] or Digg[5] have been developed to organize the news data into structured data. In this paper, rather than detecting events from social streams, we aim at querying the events from the social news web service Google News, hence benefiting for their structuring process. Google News is a computer-generated news site that aggregates headlines from more than 50,000 news sources worldwide, groups similar stories together. The content is selected and ranked by factors like how often and on what sites a story appears on the Web. It also provides an API to help developer query online news based on different characteristics. We take the time, location and topic keywords as the query parameters to retrieve popular events, ranked according to their popularity. In the returned results, many useful metadata can be found for each event, such as "title","submit_date" to describe the property of events. We also extract the "link" metadata that refers the original content of the news, so that the original content could be retrieved and be used in the text description in the final result.

[4] http://news.google.com
[5] http://digg.com/

Fig. 4: A cluster example in Google New, where the news from different sources are grouped by topic

Although in this paper, the news is extracted from Google News, the whole framework is flexible and can be extended to handle more events directories, such as Digg or any such public API easily.

3.3 News Clustering

Google News has clustered the story articles together if there are from the different sources. As an example, if one searches for "Obama" in order to read the latest news about President Obama, he/she will obtain a story cluster for a topic along with encapsulated events, one of which is as shown in figure 4. When browsing on Google News, one can click the link to see "all 1663 articles" about this topic, in order to read perspectives from different news sources.

However, Google News does not produce perfect clustering results, it clusters near-duplicate news together if they originate from different sources. In a query resulting list, there are still lots of items that describe the same event in different viewpoints. To provide a more comprehensive view of the results, we cluster the items in the results using the DBSCAN algorithm where each of the news is represented as the distribution of topics mined by Latent Dirichlet Allocation (LDA).

At first, we employ the LDA model to represent the News. LDA is a popular method for modeling term frequency occurrences for documents in a given corpus. The basic idea of LDA is to describe a document as mixture of

different topics. It allows sets of observations to be explained by unobserved groups which explain why some parts of the data are similar. In representing each document as a topic distribution (actually a vector), LDA reduces the feature dimensionality from number of distinct words appeared (in a corpus) to the number of topics. In our problem, we use the bag of textual words model to represent news document, some preprocessing is performed before learning the models. For instance, we extract the main body of the web page by the method proposed in [6]. We also filter the stopwords and words in the query for each documents, and employ the bag of textual words feature to present each of News content, which is used as the input parameters of LDA.

After the learning process, each of the News is taken as the distribution p over the latent topics. Then, we employ the DBSCAN algorithm to cluster the document entities. DBSCAN is a density-based clustering algorithm since it finds a number of clusters starting from the estimated density distribution of corresponding inputs. DBSCAN provides a ideal solution for this problem since we do not know the number of events in the news dataset. As opposed to k-means, DBSCAN does not require one to specify the number of clusters in the data as prior. To calculate the distance of news representation when clustering, the Kullback Leibler(KL) divergence is employed. However, KL divergence is not a symmetric measure. Therefore, we employ the standard symmetric version that follows as the distance measure:

$$\begin{aligned} Dist(p,q) &= D_{\mathrm{KL}}(p\|q) + D_{\mathrm{KL}}(q\|p) \\ &= \sum_i p(i) \log \frac{p(i)}{q(i)} + \sum_i q(i) \log \frac{q(i)}{p(i)} \end{aligned} \tag{1}$$

where p_i and p_j are a feature pair in the dataset.

When the cluster processing is finished, to effectively assist the users viewing the system, we would like to show the events according to their importance. In details, we measure the importance of each event by the entropy of its textual words. In details, let $w \in e$ be a word from an article in the event cluster, the entropy of event E can be calculated as

$$entropy(E) = \sum_{w \in e} p(w) log(p(w)) \tag{2}$$

where $p(w) = \frac{Cont(w)}{\sum_{w \in E} Cont(w)}$, is the probability of word w in all of the articles of an event cluster.

The event entropy quantifies the disparity of weights between words. The lower the entropy, the more interesting the corresponding an event is. After ranking, the events with more information will be shown first and events with less information shown later. Obviously the more articles an event is reported in, the more important it is.

3.4 Media Illustration

Years of multimedia research have shown that it is easier and more accurate for the computer to identify specific pattern compared with abstract concept.

To find media illustrating events, a query is specifically tailored for each event. To illustrate events using text and images, we retrieve the original visual and textual content of News that refer the same events, and show them in a appealing and informative visual format.

The event is fully described in the original web page of the latest news, to show the current state of an event. For the latest news, we extract the "link" metadata in Google News and crawl the original web page. To extract the main body of the web page, the method proposed in [6] is employed. It recognizes the main content based on character number under the assumption that characters accumulate more in the main content than in other part of web page.

For each event, we would like to collect other textual description. The textual data is retrieved the original web pages of all the news. To provide a nice and meaningful visualization, we use a tag cloud to organize the textual data. Tag cloud is commonly used on geographic maps to represent the relative size of cities in terms of relative typeface size, which is formulated as a histogram which can represent the frequency of over hundreds of items. In tag cloud, the importance of each tag is shown with different format. This format is useful for quickly perceiving the most prominent terms. For each event, we segment each word in the articles into tags, count the frequency of each tag and generate the tag cloud with tags in different font size and color.

Besides the tag cloud, we also provide another vivid visualization called photo collage [18], to illustrate events. Photo collage is a visual clustering technique that can depict the event with different points of view. To generate the photo collage, we retrieve the key photos in each of the article in an event cluster. With the selected photos, the methods proposed in [18] is used to create the photo collage for each events. The solution is to formulate the generation of collage as a unified energy minimization problem in which each of steps like extracting salient regions of interest (ROIs) from these images, and seamlessly arranging ROIs on a given canvas with the temporal structure is represented by an energy term.

Table 1: Events found for query "Obama in US last week", performed on Oct. 29th, 2012

Event Title
Obama Reportedly Unaware of World Leader Phone Tapping
Obama to meet with House GOP on immigration
Michelle Obama Recycles Dress She First Wore in 2009, Makes it
Obama, Netanyahu discuss Iran, Middle East talks by phone
Obama Says He Picked Comey Among Many to Lead FBI
No, Barack Obama's Twitter Account Wasn't Hacked
Syrian hackers claim to hack Obama's Twitter account
Obama vs. Putin – is the Cold War about to return to the Middle East?
The Unbearable Lightness of Obama

4 Experimental Results

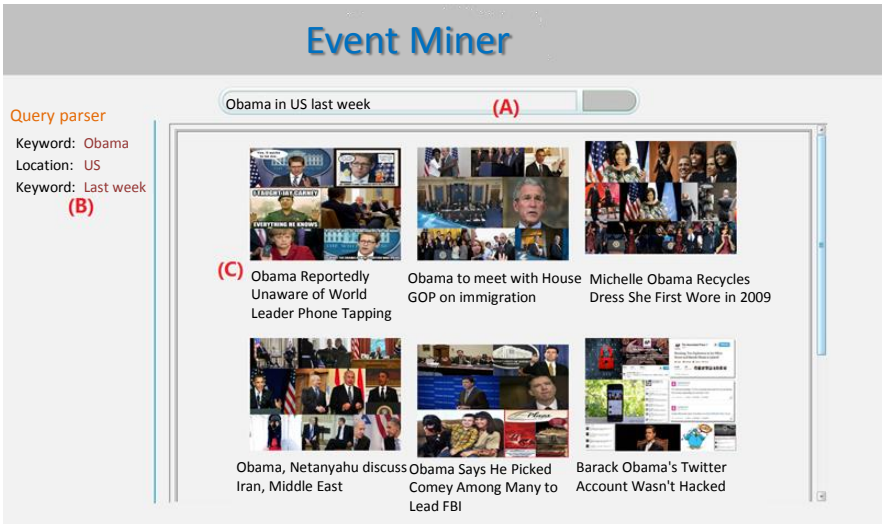

Fig. 5: The web service to show the query results, (A): query words. (B): parsed semantics (C): Events Thumbnails

Our system can be used to extract and illustrate events from a query input. For a given query, we first parse it to find out the topics, location, and time information. It should be noticed that for some queries, one or more semantic dimension is missing, and the default parameter will be set. For example, in the query "the Christmas in London", no temporal information is provided. The default value are "in the last 7 days" for time, to ensure the timeliness of events, and "worldwide" for location, and "" for topic, to reduce the limitation of event querying. The motivation for using "in the last 7 days" as default value for time, is that recent news are more likely to be the topic of interest when not specified. We then use the parsed topic, time, and location to query news from Google News. Then the clustering methods proposed in Section 3.3 is employed to mine the events. After ranking, the top 9 events are kept as hot candidates according to their ranking promoted by the users. Based on our proposal, a web service called Event Miner is built to help user browse media data from events, as shown in Figure 5. In summary, there are three parts in this web service. When a user types the query string, "Obama in US last week" for example, in part (A), our service provides the parsed location, time, and topic keyword (part (B)), and find corresponding events as listed in Table 1. All of the events are depicted by thumbnails in the combined form of photo and event title (part (C)). All of the events are ranked by their importance as described in Section 3.4.

When navigating to a specific event, the event illustration as shown in Figure 6 is provided, giving the user information related on that event solely.

Although currently limited to a simple layout, which can easily be improved by a graphic designer, it is composed of five individual parts to help people understand the event well.

The event title is shown in part (A), which is the highest level description for the events.

To provide the users with detailed information about the event, we also parse the original news web page, and mine the main textual content part, as shown in part (B).

We parse the title and time metadata from the obtained events, use the parameters to query from Twitter the comments from different users, then use the tag clouds to show the results. The tag cloud is presented in part (C). We can clearly see that the larger size content such as "Obama", "Comey", "FBI", matches the event topic "Obama Says He Picked Comey Among Many to Lead FBI" very well.

Besides text visualization, we also retrieve photos with Google image search, and filter the ones that can not be matched based on their textual metadata. All of the matched photos are collaged in the same layout. The photo collage is shown in part (D). From the figure, it could be found that most of photos are relevant to the event; "Obama Says He Picked Comey Among Many to Lead FBI".

To conclude, while tag cloud and photo collage in part (C) and (D) provide an attractive visualization with abstract content, the textual content in part (B) gives concrete description that would assist the users to understand the event in detail.

Fig. 6: An event sample, entitled with "Obama Says He Picked Comey Among Many to Lead FBI". The event is depicted by four parts; (A): title, (B): text description, (C): tag cloud, (D): photo collage.

We collected a dataset to quantify the performance of the proposed News clustering methods. On the given query "Obama in US last week", we collect the the top 370 News documents in Google News web service (The query was performed on 19th Oct, 2013). To obtain the ground truth, the dataset is manually labeled by 5 students who do not know the purpose of this work. It is hard for the annotators to find the general agreement when labeling the data, since many news could be explained in different ways. And the labels from majority is used as the final ground truth. In the end, 52 events are found in the dataset by these students, such as "the new health policy","veterans military", "activities in Middle East", "surveillance program", "general administration duties" and so on. In addition, the dataset is very imbalanced, for example, there are 71 News describing the most observable event "the new health policy" with very different viewpoints, and 15 news describing the situation of "veterans military" problems. However, 41 out of the 52 events, only one News item could be found. The imbalance of dataset gives challenges to the cluster methods.

We use the main content of News as documents and conduct our news cluster methods on this dataset. The 256d Tf-idf feature is employed to represent the document and the number of topics in LDA is set as 50. It should also be noticed that we just consider the distribution of documents on latent topics and the number of topics in our method is NOT a key parameter in the system. Our News cluster methods is compared with the baseline feature tf-idf, and baseline cluster methods KMean++[2], as well as the spectral cluster[23]. The experiments are conducted on a PC platform . The results are evaluated by normalized mutual information (NMI) [17], and reported in Table 2.

Table 2: Event Cluster Results

	TF-IDF	LDA
Kmean++	0.344	0.243
spectral cluster	0.360	0.375
DBSCAN	0.387	0.516

From table 2, we can see that the overall performance is not as good as expected and should be improved further. With TF-IDF representation, the three cluster methods provide comparable performance, with the NMI value 0.344, 0.360 and 0.387 respectively. However, the cluster methods exhibit very difference performance when conducted on LDA feature. The NMI of Kmean++ degrades from 0.344 to 0.243, and spectral cluster does not change too much (with NMI from 0.360 to 0.375). Interestingly, the proposed DBSCAN cluster method obtains the best performance (with its NMI increasing from 0.387 with TF-IDF to 0.516 with LDA).

In addition, the running time of the proposed approach on the collected dataset as well as methods we compare it with are reported in Table3. With the same feature space (either Tf-idf or LDA), the DBSCAN outperforms Kmean++ and spectral cluster methods. However, it also should be notice that it takes about more 50 times by LDA features compared with Tf-idf

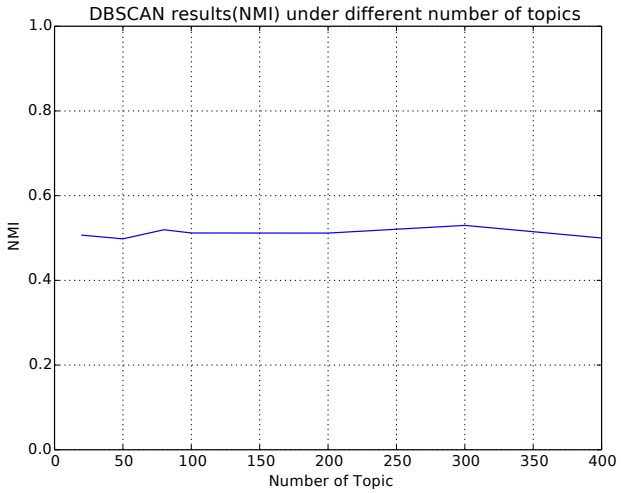

Fig. 7: Clustering results performance with different number of latent topics. We train the LDA model with the topic number from a pool [20, 50, 80, 100, 200, 300, 400], and the final NMI is very stable, which suggests that our approach is robust with respec to the number of topics.

feature. Actually most of the time is spent on training the LDA model. And the following clustering runs faster since the feature dimension reduces from 256d to 50d. To conclude, the proposal can detect events in nearly real time and achieve competitive performance.

Table 3: Event Clustering Time Consumption

	TF-IDF	LDA
Kmean++	0.151	5.051
DBSCAN	0.056	4.956
spectral cluster	0.192	5.092

In the experiments, we also study the impact of the number of latent topic when training the LDA model. We train the model with different number of topics (ranging from 20 to 400) and then cluster the results with DBSCAN. The results are evaluated using NMI, and range from 0.498 to 0.529, as shown in Figure 7. Those results show that the proposed approach is robust to the different number of latent topics.

5 Conclusion

We proposed an original cross media question answering framework leveraging on social media data (News, Media Sharing Platform and Microblog) to

extract and to depict public events. The process is done automatically, without any human assistance, on a given textual query. First, natural language processing is employed to parse and make sense of the input query. Then, a social news web service is employed as the querying sources. To extract relevant events, we employed Latent Dirichlet Allocation to represent the content of News and DBSCAN approach to cluster News into distinctive Events. We present the results with an attractive visual format combining tag clouds and photo collage. We demonstrated how our novel approach addresses the event retrieval and illustrating problem and showed its effectiveness on a real world example.

Acknowledgments

This work was supported by the 973 Program of China (No. 2013CB329604), and the European Commission under contracts FP7-287911 LinkedTV and FP7-318101 MediaMixer .

References

1. J. Allan, J. Carbonell, G. Doddington, J. Yamron, and Y. Yang. Topic detection and tracking pilot study: Final report. In Proceedings of the DARPA Broadcast News Transcription and Understanding Workshop, pages 194–218, Lansdowne, VA, USA, Feb. 1998. 007.
2. D. Arthur and S. Vassilvitskii. k-means++: the advantages of careful seeding. In SODA '07: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, pages 1027–1035, Philadelphia, PA, USA, 2007. Society for Industrial and Applied Mathematics.
3. B.-K. Bao, W. Min, J. Sang, and C. Xu. Multimedia news digger on emerging topics from social streams. In Proceedings of the 20th ACM international conference on Multimedia, MM '12, pages 1357–1358, 2012.
4. H. Becker, D. Iter, M. Naaman, and L. Gravano. Identifying content for planned events across social media sites. In ACM conference on WSDM, 2012.
5. L. Chen and A. Roy. Event detection from flickr data through wavelet-based spatial analysis. In ACM conference on CIKM, 2009.
6. S. Chengjie and G. Yi. A statistical approach for content extraction from web page. Journal of Chinese Information Processing, 18(5):17–22, 2004.
7. D. Delgado, J. a. Magalhães, and N. Correia. Assisted News Reading with Automated Illustrations. In ACM conference on Multimedia, pages 1647–1650, 2010.
8. C. S. Firan, M. Georgescu, W. Nejdl, and R. Paiu. Bringing order to your photos: Event-Driven Classication of Flickr Images Based on Social Knowledge. In Proceedings of the 19th ACM international conference on Information and knowledge management, page 189, New York, USA, Oct. 2010.
9. Y. Gao, M. Wang, Z.-J. Zha, J. Shen, X. Li, and X. Wu. Visual-textual joint relevance learning for tag-based social image search. IEEE Transactions on Image Processing, 22(1):363–376, Jan 2013.
10. R. Hong, M. Wang, G. Li, L. Nie, Z.-J. Zha, and T.-S. Chua. Multimedia question answering. MultiMedia, IEEE, 19(4):72–78, 2012.
11. D. Joshi, J. Z. Wang, and J. Li. The Story Picturing Engine—a system for automatic text illustration. ACM Transactions on Multimedia Computing Communications and Applications, 2(1):68–89, 2006.
12. G. Li, Z. Ming, H. Li, and T.-S. Chua. Video reference: question answering on youtube. In Proceedings of the 17th ACM international conference on Multimedia, pages 773–776, 2009.

13. H. Li, J. Tang, Y. Wang, and B. Liu. Looking into the world on google maps with view direction estimated photos. Neurocomput., 95:72–77, Oct. 2012.
14. X. Liu, B. Huet, and R. Troncy. Eurecom@ mediaeval 2011 social event detection task. In MediaEval, 2011.
15. X. Liu, B. Huet, and R. Troncy. Eurecom@ MediaEval 2011 social event detection task. In Proceedings of the MediaEval 2011 Workshop, 2011.
16. X. Liu, R. Troncy, and B. Huet. Finding Media Illustrating Events. In ACM International Conference on ICMR, Trento, Italy, 2011.
17. C. D. Manning, P. Raghavan, and H. Schütze. Introduction to Information Retrieval. 1 edition, July 2008.
18. T. Mei, B. Yang, S.-Q. Yang, and X.-S. Hua. Video Collage: Presenting a Video Sequence Using a Single Image. The Visual Computer, pages 39–51, 2009.
19. L. Nie, M. Wang, Y. Gao, Z.-J. Zha, and T.-S. Chua. Beyond text qa: Multimedia answer generation by harvesting web information. IEEE Transactions on Multimedia, 15(2):426–441, Feb 2013.
20. C.-C. Pan and P. Mitra. Event detection with spatial latent Dirichlet allocation. In Proceeding of the 11th annual international ACM/IEEE joint conference on Digital libraries, page 349, New York, USA, June 2011.
21. T. Quack, B. Leibe, and L. 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, New York, USA, July 2008.
22. T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In International conference on World Wide Web, Raleigh, North Carolina, USA, 2010.
23. J. Shi and J. Malik. Normalized cuts and image segmentation. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 22(8):888–905, 2000.
24. M. Wang, R. Hong, G. Li, Z.-J. Zha, S. Yan, and T.-S. Chua. Event driven web video summarization by tag localization and key-shot identification. IEEE Transactions on Multimedia, 14(4):975–985, Aug 2012.
25. M. Wang, G. Li, Z. Lu, Y. Gao, and T.-S. Chua. When amazon meets google: Product visualization by exploring multiple web sources. ACM Trans. Internet Technol., 12(4):12:1–12:17, July 2013.
26. J. Weng and F. Lee. Event detection in twitter. In AAAI conference on Weblogs and Social Media, Barcelona, Spain, 2011.
27. I. Yahiaoui, B. Mérialdo, and B. Huet. Comparison of multi-episode video summarization algorithms. EURASIP Journal on applied signal processing Special issue on multimedia signal processing - Volume 2003 N1, January 2003, 01 2003.
28. Z.-J. Zha, L. Yang, T. Mei, M. Wang, Z. Wang, T.-S. Chua, and X.-S. Hua. Visual query suggestion: Towards capturing user intent in internet image search. ACM Trans. Multimedia Comput. Commun. Appl., 6(3):13:1–13:19, Aug. 2010.
29. Z.-J. Zha, H. Zhang, M. Wang, H. Luan, and T.-S. Chua. Detecting group activities with multi-camera context. IEEE Transactions on Circuits and Systems for Video Technology, 23(5):856–869, May 2013.
30. X. Zhu, A. B. Goldberg, M. Eldawy, C. R. Dyer, and B. Strock. A Text-to-Picture Synthesis System for Augmenting Communication. In Proceedings of the 22nd national conference on Artificial intelligence, number 2, pages 1590–1595, 2007.