CrowdStory: Fine-Grained Event Storyline Generation by Fusion of Multi-Modal Crowdsourced Data

BIN GUO, Northwestern Polytechnical University
YI OUYANG, Northwestern Polytechnical University
CHENG ZHANG, Georgia Institute of Technology
JIAFAN ZHANG, Northwestern Polytechnical University
ZHIWEN YU, Northwestern Polytechnical University
DI WU, Hunan University
YU WANG, University of North Carolina at Charlotte

Event summarization based on crowdsourced microblog data is a promising research area, and several researchers have recently focused on this field. However, these previous works fail to characterize the fine-grained evolution of an event and the rich correlations among posts. The semantic associations among the multi-modal data in posts are also not investigated as a means to enhance the summarization performance. To address these issues, this study presents CrowdStory, which aims to characterize an event as a fine-grained, evolutionary, and correlation-rich storyline. A crowd-powered event model and a generic event storyline generation framework are first proposed, based on which a multi-clue–based approach to fine-grained event summarization is presented. The implicit human intelligence (HI) extracted from visual contents and community interactions is then used to identify inter-clue associations. Finally, a cross-media mining approach to selective visual story presentation is proposed. The experiment results indicate that, compared with the state-of-the-art methods, CrowdStory enables fine-grained event summarization (e.g., dynamic evolution) and correctly identifies up to 60% strong correlations (e.g., causality) of clues. The cross-media approach shows diversity and relevancy in visual data selection.

Concepts: • **Information System Applications: Miscellaneous**

General Terms: Design, Algorithm, Performance

Additional Key Words and Phrases: Event Sensing, Mobile Crowdsourcing, Fine-grained, Storyline, Correlation.

ACM Reference format:
Bin Guo, Yi Ouyang, Cheng Zhang, Jiafan Zhang, Zhiwen Yu, Di Wu, Yu Wang. 2017. CrowdStory: Fine-Grained Event Storyline Generation by Fusion of Multi-Modal Crowdsourcing Data. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 1, 3, Article 55 (2017), 19 pages.
DOI: http://doi.org/10.1145/3130920

1 INTRODUCTION

Social event sensing is important to the society, often presenting the causes, highlights, and evolution of an event. The information obtained is widely used in areas such as public safety, event tracing, and disaster management, among others.

This work is supported by the National Natural Science Foundation of China (grant nos. 61332005, 61373119, and 61402369), and the National Basic Research Program of China (grant no.2015CB352400).
Author's addresses: B. Guo (e-mail: guob@nwpu.edu.cn), No. 127, Youyi-West Rd., Xi'an 710072, China; Y. Ouyang, No. 127, Youyi-West Rd., Xi'an 710072, China; C. Zhang, North Avenue, Atlanta, GA 30332, USA; J. Zhang, No. 127, Youyi-West Rd., Xi'an 710072, China; Z. Yu, No. 127, Youyi-West Rd., Xi'an 710072, China; D. Wu, Hunan University, Changsha 410082, China; Y. Wang, 9201 University City Blvd. Charlotte, NC 28223, USA.

DOI: http://doi.org/10.1145/3130920

With the rapid development of wireless networks and smartphones, more people have gotten used to posting interesting stories about what they see or hear in their lives on social media. For example, microblogs have become one of the most significant media for instant event sensing, public sharing, and rapid propagation of news.

Social media-enabled event sensing consists of two major research tasks: *event detection* and *event summarization*. Event detection aims to detect hot trending events from large-scale crowdsourced data. For example, Sakaki et al. [28] reported earthquakes in real time based on data from Twitter. Each detected event usually corresponds to a tremendous volume of unorganized posts, which can potentially provide further information about the event. However, the large number of event posts is often redundant and noisy and has complex correlations among them, making it difficult and time-consuming to extract the key information on that event. To quickly understand the essence of events, event summarization over large-scale crowdsourced data has been explored in this community. The popular event summarization methods either select representative posts as the summary or rearrange the posts in chronological order. However, these methods cannot characterize every event clue/aspect, the causes and consequences, and the evolution of each clue of the event. These data, however, are important to event sensing and decision-making because they can present the event (e.g., a disaster or a terrorist attack) in a logical, replicable, and comprehensive manner [48]. In the present work, all data relevant to an event are assumed to be accessible by using the existing event detection methods [5, 7, 13, 22, 25, 28, 30, 38, 39] or simply by applying the hashtags in microblogs. Based on this, we present CrowdStory, which provides fine-grained event summarization and event storyline generation. In contrast to previous studies, the proposed technology is developed based on the understanding of multifaceted event clues, the evolution of each clue, and the semantic associations among clues.

In this research, a "clue" is defined as an aspect of an event. In general, it corresponds to a set or a cluster of posts that contains similar keywords and describes the same aspect of an event. For example, for a terrorist attack event, a clue may refer to the scene of the attack, an analysis of the attack, or the rescue efforts, among others. The following three methods are applied to understand the evolution and correlation of the important clues of an event. First, the *multifaceted clues* from a large number of posts about an event are identified. Second, as the event evolves, the event segmentation is studied, and the *evolution of clues* in different sub-events is extracted. Third, the *complex relationships among different clues* of the event are derived, including the time order, causality, and supplementary correlations, among others. The 2015 Tianjin explosion event[1], in which a series of explosions killed 173 people and injured hundreds of others, is used as an example to show the design of CrowdStory. The event happened on August 12, 2015 at a container storage station in the port of Tianjin, China.

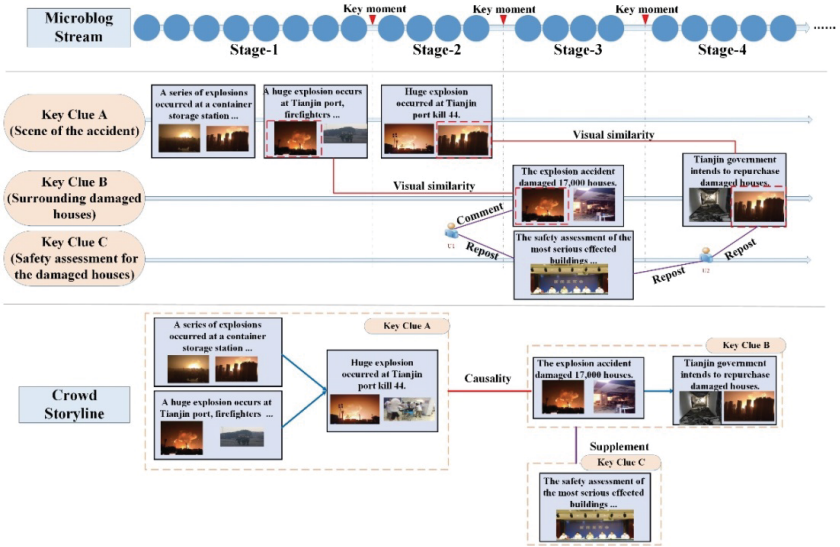

Fig. 1. An illustration of CrowdStory.

[1] https://en.wikipedia.org/wiki/2015_Tianjin_explosions.

Figure 1 partially describes the event by using CrowdStory. The three key clues (A, B, and C) are first identified; the posts in the same clue have similar keywords. Clues A, B, and C, respectively, are relevant to the scene of an outbreaking explosion accident, the houses damaged in the accident, and the safety assessment of the damaged houses. Each clue corresponds to a set of posts. The posts for a clue are segmented by key change moments which indicate the evolution of the event. The inter-clue relationship can also be derived. For example, it is clear that clue A causes clue B, and clue C is supplementary information to clue B. Our proposed technology can provide a storyline that depicts these complicated relationships. Due to the spontaneous nature of crowd contribution in social media, several of the event posts may contain *redundant* information. Therefore, among the large number of event posts, only *selected* multimodal data (texts and images) are used to visually present the story.

To achieve this end, several challenges need to be addressed.

(1) *How to make the summary such that it illustrates the evolution of the multifaceted aspects/clues of the event.* Previous works have only considered the event as a whole and have not carried out fine-grained, clue-level summarization.

(2) *How to identify the complicated correlations among the clues and generate an informative storyline.* The existing studies mostly connect posts in the time order, which cannot characterize the rich correlations among posts.

(3) *How to select representative multimodal data for event visualization.* As previously mentioned, multimodal data selection is needed in event summarization to reduce information redundancy. The challenge is how to identify the important data in posts. Currently, the data selection is mostly based on text diversity, and only the images embedded in the relevant posts are used for visualization. The semantic correlations of the texts and images in different posts are not investigated. To ensure a better visual presentation of events, the *diversity* and *relevancy* of visual contents among posts, which are crucial for information retrieval, should also be considered.

CrowdStory is designed to explore multimodal microblog information and to characterize an event as a fine-grained, evolutionary, and correlation-rich storyline. In particular, the approach has the following contributions:

(1) A fine-grained event storyline generation framework using crowdsourced social media data. We build a crowd-powered event data model and use the multi-modal data as well as the implicit human intelligence (HI) to generate event storylines.

(2) An evolutionary and multi-clue–based approach for fine-grained event summarization. The textual contents are used for clue detection, and the temporal crowd-posting patterns are studied for event segmentation.

(3) Extraction of implicit HI from visual contents and community interactions for inter-clue association identification and storyline generation. The visual and user interaction contexts are used to identify causes and consequences and supplementary correlations among event clues.

(4) A cross-media [42] mining approach to selective visual story presentation. Cross-media mining indicates the exploration of the semantic correlations among different model presentations (e.g., texts, images, and GPS coordinates) of the same sensing object (e.g., an event). In the present work, multimodal data are represented and measured in a shared semantic space and are selected based on both visual diversity and relevancy.

To evaluate CrowdStory, experiments are carried out by using data collected from Sina Weibo[2] and Twitter[3]. The results indicate that our proposed method performs better than the state-of-the-art methods. First, CrowdStory enables fine-grained event summarization (e.g., dynamic evolution) and correctly identifies up to 60% strong correlations (e.g., causality) of clues. Second, the cross-media approach shows both visual diversity and relevancy in event presentation.

2 RELATED WORK

2.1 Crowdsourced Event Summarization

In this section, previous works are discussed based on how they organize and present the summary of an event. The studies are categorized into three types.

The *first* type of works selects representative posts as the summary of an event. Rudra et al. [27] classified tweets to extract situational information on disaster events and then used a content word–based approach to representative post

[2] http://weibo.com
[3] http://www.twitter.com

selection. Corney et al. [6] proposed a method of sport event summarization in which data contributed by fans are selected. Bian et al. [1] used annotation information and images to select posts from crowdsourced data. Schinas et al. [29] proposed a topic modelling technique to capture the relevance of messages to event topics, as well as a graph-based algorithm to select top-ranked images for visual event summary. Kim et al. [15] designed crowdsourcing tasks to generate summaries of events based on commonly used narrative templates. However, these approaches cannot determine the evolution and correlation of posts.

The *second* type of research uses a straight timeline to describe the evolution of an event. For instance, many projects [4, 27, 28, 38] use spike/burst detection to determine important moments and then select posts for each important moment to generate a chronological event summary. Meladianos et al. [24] proposed a graph degeneracy method to detect important event moments. Xu et al. [42] applied an approach that could automatically select a set of representative images to generate a concise visual summary of a real-world event from the Tumblr microblogging platform. Shen et al. [32] used a participant-based event summarization approach that applied a mixture model to detect important sub-events associated with each participant so as to attain better coverage. However, although these works summarize events in a time sequence manner, they cannot characterize other rich correlations (e.g., causality) among different aspects of an event.

To address this issue, the *third* type of studies aims to generate a structural storyline to describe events, usually by applying graph-based methods or probabilistic models. For example, [35, 18, 47] consider the post selection problem as a minimum dominating set problem in a graph, and use the Steiner tree algorithm to extract structural information. Lee et al. [17] proposed a context search method to track the structural context on the fly to build a vein of stories. Dehghani et al. [8] developed a framework to generate a summarized storyline of news events based on the concepts in graph theory. Hua et al. [12] introduced a Bayesian model to generate storylines from massive documents and infer the corresponding hidden relations and topics. Tang et al. [34] presented a nonparametric probabilistic topic model called CHARCOAL, which jointly models news storylines, stories, and topics. In comparison with timeline-based summaries, storylines can better present an event by linking representative posts as a graph. However, these works merely use textual and temporal similarities to establish links among posts; thus, the relations identified are limited.

The present study differs from and potentially outperforms the above works in several aspects. First, the evolution of multifaceted aspects/clues of the event is shown by clue detection and event segmentation. Second, the human intelligence implied in multimodal information, such as images, reposts, and comments, is used to identify multiform semantic associations (e.g., causes and consequences). Third, a cross-media mining approach to selective visual and textual story presentation is proposed. The correlations among images and texts in different posts are measured in a shared semantic space, and images are selected based on both diversity and relevancy.

2.2 Cross-Media Mining

There are some related studies on the use of cross-media mining in multimedia information retrieval. Wu et al. [40] proposed a general cross-media ranking algorithm called Bi-CMSRM for multimedia information retrieval. Yu et al. [43] presented a discriminative coupled dictionary hashing (DCDH) method for fast cross-media retrieval. Zhang et al. [44] developed a supervised multimodal hashing method for similarity search. Song et al. [33] proposed an inter-media hashing (IMH) model for exploring the correlations among multiple media types from different data sources to enable large-scale inter-media retrieval. Wei et al. [37] applied a modality-dependent cross-media retrieval (MDCR) model, in which two couples of projections are learned for different cross-media retrieval tasks instead of one couple of projections. Zhou et al. [45] used sparse coding and matrix factorization for a multimodal search. Wang et al. [36] applied deep learning methods to carry out multimodal retrieval. Shang et al. [31] proposed a deep model to learn the high-level feature representation shared by multiple modalities for cross-media retrieval. Our work differs from these previous studies in that it applies cross-media mining in event storyline visualization.

3 CROWDSTORY: AN OVERVIEW

3.1 Crowd-Powered Event Data Model

The inputs of our fine-grained event characterization model are microblog posts (post) on an event; the output is a generated storyline.

3.1.1 Post data model. Figure 2 shows a post on Sina Weibo depicting scenes of a family reunion after the Paris terrorist attack. A microblog post consists of rich and multimodal data, including the publisher, posting time, textual and visual contents (i.e., images), and inter-post interaction (reposts/comments), among others. We propose a model with 5-tuple post data $post = \langle pid, time, text_c, vis_c, intr \rangle$: pid is the unique identification of the post; $time$ is the posting time; $text_c$ is the textual content of the post; vis_c is a set of images embedded in the microblog; $intr$ refers to pairs of user interactions over the post, including the publisher, comment users, and repost users. The set of posts about an event e is denoted by P, and the number of posts about e is denoted by N.

Fig. 2. Multi-modal microblog posts: an example.

Table 1. Key notations

Notation	Definition
e	An event
P	The set of posts about an event e
N	The number of posts about an event e
S	An event storyline graph
C	a set of clues, and each clue describes one aspect of the event
E_C	A set of undirected edges among clues
SG	A set of stages obtained after event segmentation
cid	A clue in C
G_{cid}	A directed graph that depicts the evolution of a clue
V	A set of nodes to represent the clue
E_V	A set of undirected edges among clues

3.1.2 Event storyline graph. A storyline provides a global view of the development of the story over time. The aim of CrowdStory is to construct an event storyline graph $S = (C, E_c)$ based on the original event post set P. C denotes a set of clues, and each clue describes one aspect of the event. E_c refers to a set of undirected edges among clues, denoting the associations among them. Three types of association among clues are defined.

- ***Causality association***. The word "causality" may have different meanings in different situations. In AI, causality is often linked with probabilistic theory and statistical associations [26]. Here, we refer to the WikiPedia definition[4] of causality as "*the agency or efficacy that connects one process (the cause) with another process or state (the effect)*," which is defined at the qualitative level. In particular, the connected clues in the present work have the relationship of cause and effect.
- ***Supplement association***. This refers to the adjunctive details or background knowledge about a post.
- ***Relevant association***. This means that related posts refer to the same subtopic of the event.

In terms of time, an event can be segmented into a set of stages SG. Each clue $cid \in C$ can be represented as a directed graph $G_{cid} = (V, E_V)$. V is a set of selected nodes representing the clue, which contains both textual and visual contents. Data selection is carried out at the clue stage ($sg \in SG$). E_V is a set of directed edges (indicating the time order). E_V connects $v(v \in V)$ to generate the evolution for the clue. The rich and multimodal information in the post data model is leveraged to generate the storyline, as presented later. Table 1 shows the key notations used in this paper.

3.2 The CrowdStory Framework

Figure 3 presents an overview of the CrowdStory framework, which has four major modules: 1) *data crowdsourcing*, which collects data from various microblogging sites; 2) the *crowdsourced event data model*, which transfers raw data into a unified presentation; 3) *evolutionary clue characterization*, which applies the temporal and textual clustering approach to obtain multifaceted clues about the event and to characterize the intra-evolution of each clue; and 4) *inter-clue association identification and storyline generation*, which explores user interactions and visual contexts to identify the correlations among clues. Two optimization methods are introduced for text selection, and a cross-media mining approach is developed for selection of visual contents based on both data diversity and relevancy.

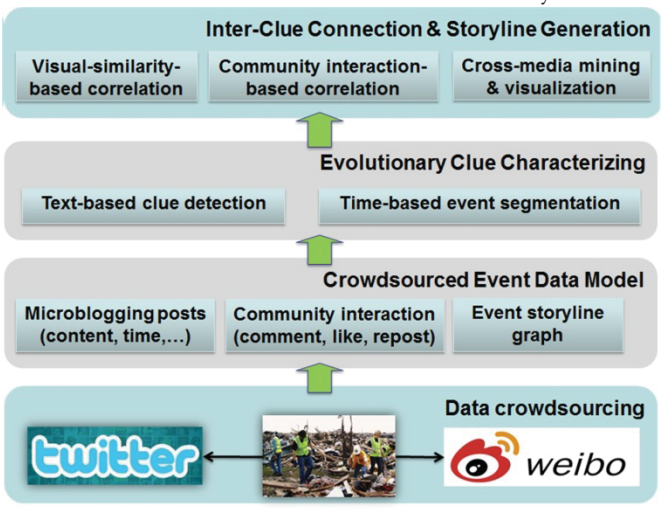

Fig. 3. The CrowdStory Framework.

4 DETAILED DESIGN OF CROWDSTORY

4.1 Storyline Generation with Complementary Data Fusion

The proposed system collects different types of data from microblogs, including texts, images, and user interactions. Each type represents distinct features, which can be potentially complementary to each other. Thus, heterogeneous and

[4] https://en.wikipedia.org/wiki/Causality

complementary microblog data are used for event storyline generation. Figure 4 presents the process of generating the storyline.

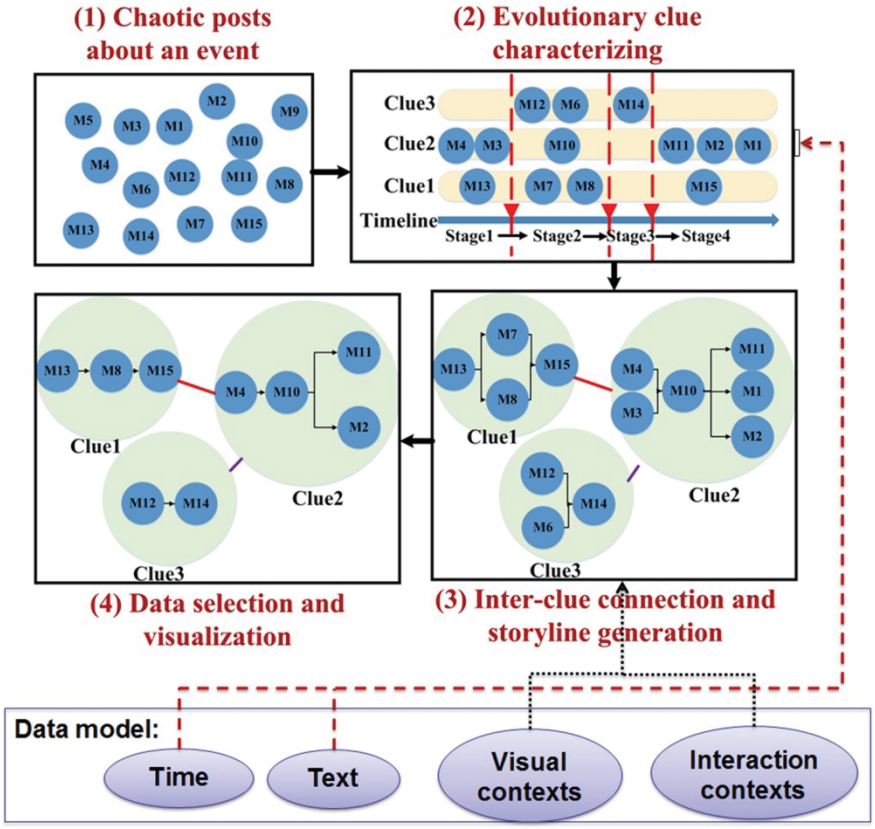

Fig. 4. An illustration of storyline generation with multi-modal, complementary data fusion.

4.1.1 Evolutionary clue characterization. A high-impact social event usually consists of different *clues*. The event evolves and has different important moments. For instance, the meaning of two similar posts published at different times can be different; for instance, in a report on the death toll in an accident, the textual content is similar, but the death toll changes. The existing event summaries based on microblog data are defined at a coarse-grained level and thus fail to depict fine-grained elements, such as various clues and their evolution. To address this challenge, we designed a two-step grouping approach that uses a combination of textual and temporal similarities. Multifaceted clues about the event are first obtained based on textual similarity. Then, temporal clustering is used to segment the clues and characterize their evolution.

Textual-similarity for clue detection. Similar textual contents often refer to the same aspect of an event, which can be used for clue detection. The state-of-the-art methods, namely, TF-IDF [2] and cosine distance, are used to measure the textual similarity.

Event segmentation with temporal distribution features. An essential element in characterizing an event is the distribution of the time stamp of posts, which usually shows the evolution of the event. The post time interval is used to measure the time similarity of two posts. If the interval between them is short, the posts are more likely to refer to the same stage of an event.

After retrieving the textual similarity and temporal distribution features of the posts, a connected graph [3] is used for clue detection (CD) and event segmentation (ES). Suppose N posts about an event are obtained. First, two N^*N matrices *M1 and M2* are built to describe the similarity among different posts. Entry $M_{i,j}$ (for *M1* or *M2*) is the textual similarity

between posts i and j for CD or the absolute time interval between the two posts for ES, respectively. Second, a graph $EvoG = (P, E)$ is constructed based on M, where P is a set of posts on the event, and E is a set of undirected edges. We consider that there is an edge between posts p_i and p_j, if *i*) $M_{i,j}$ is above a threshold δ for CD, or *ii*) $M_{i,j}$ is less than a threshold θ for ES. Third, the depth-first search (DFS) approach is applied to obtain a set of connected subgraphs. Each subgraph represents a clue (for CD) or a stage (for ES) of the event. C denotes the grouped clue and SG denotes the stage set. By examining the connected clues and stages together, the evolution of each clue can be derived, as shown in Fig. 4(2).

Inner-clue connection. The connections among posts within the same clue are called inner-clue connections. To characterize the evolution of each clue over time, connections are established between posts at each stage of a clue, as shown in Fig. 4(3).

4.1.2 Identification of Inter-Clue Association. The extracted clues are not isolated from each other but instead are connected with each other (e.g., clue A causes clue B in Fig. 1). Identifying the connection between clues would help to better represent the evolution of an event. To address this challenge, the implicit human intelligence (HI) hidden in the rich contents of posts is used. This idea is motivated by our observation that the knowledge hidden in the process of data generation, such as the individual and crowd behavior contexts, is often neglected in the understanding of crowdsourced data.

Compared with the textual contents, the visual contents (selected images from the post publisher to show textual contents and individual behavior contexts) and user interaction contexts (crowd behavior contexts) potentially contain rich semantic or high-level information, which can be used to detect connections. For instance, in Fig. 1, all posts in clues A and B refer to the situation at the explosion scene, and clues A and B seem to be irrelevant because their keywords are different. However, both clues contain an image of the same entity (a building on fire), which connects clue A to clue B visually. Overall, two context-based heuristic rules are applied in the association identification, as presented below.

Visual contexts. When two posts are related to the same content, they are more likely to have similar visual contexts (i.e., images). To characterize the relationships between posts, the image similarity is calculated by using scale-invariant feature transform (SIFT) [21], which is a common approach used in image similarity measurement. The advantage of the scale-invariant feature is that the detection performance is not influenced much by the angle of the picture. SIFT is applied to extract feature points from images, and the similarity between two different images is measured by using the match ratio of feature points. $SimV_{i,j}$ represents the visual similarity between the images in posts i and j.

$$SimV_{i,j} = \frac{\sum_{m \in pic_i, n \in pic_j} MR_{m,n}}{|pic_i| \cdot |pic_j|} \tag{1}$$

pic_i and pic_j denote the set of images in posts i and j, respectively; $MR_{m,n}$ is the match ratio of feature points between images m and n; $|pic_i|$ and $|pic_j|$ are the number of images in posts i and j, respectively.

Interaction contexts. The user interaction contexts refer to the interactions (e.g., comments and reposts) between a user and the event posts, which also incorporate rich HI. Compared with machine processing, humans have a high ability to identify the relations among different aspects of an event, such as causality and supplement, as in the case shown in Fig. 1. This is often reflected in the user interactions because people interact over posts on topics that they prefer or have a strong connection with. For instance, as shown in Fig. 1, U1 pays attention to the damaged buildings, comments on a post (about how many houses are damaged in the event) in clue B, and reposts another post (about the safety assessment of the buildings) in clue C. Motivated by the triadic closure theory in social networks [16], we assume that if two posts are closely related, they will have more common users who comment and repost compared with irrelevant posts. The relations between different posts are characterized by measuring the interaction group similarity in Eq. (2).

$$SimI_{i,j} = \frac{\sum_{m_i \in IC_i \cap IC_j} weight_{m_i}}{|IC_i|} + \frac{\sum_{m_j \in IC_i \cap IC_j} weight_{m_j}}{|IC_j|} \tag{2}$$

$SimI_{i,j}$ is the interaction group similarity between posts i and j; IC_i and IC_j are the interaction communities who comment and repost posts i and j, respectively; $|IC_i|$ and $|IC_j|$ refer to the number of users in the interaction community. Note that some users may be more active, commenting on or reposting a wide range of posts. This may result in incorrect

link prediction. To address this issue, the interaction community is regarded as a document, and each user as a word in that document. Thus, TF-IDF is used to calculate the weight of each user, i.e. $weight_u$.

4.1.3 Inter-clue connection. A *clue* consists of a set of *posts*, as shown in Fig. 4(3). Thus, two methods, *post*-level and *clue*-level, are developed to determine the connection between different clues (i.e., inter-clue connection).

1) Post-level integration. The similarity between two clues for each pair of posts is first computed, and then the average value of the post pairs is regarded as the final similarity of the two clues.

2) Clue-level fusion. The images or interactions for the same clue are first aggregated, and then "post" is replaced with "clue" in Eqs. (1) and (2). The visual/interaction similarity at the clue-fusion level can then be obtained.

To avoid excessive or invalid connections, in both methods we connect only the pairs of clues with the highest visual or interaction similarity. The connection time point is the start encountering point of the two clues.

4.2 Multimodal Storyline Visualization

Given the above steps, a large graph is generated from among the posts. However, it is not good to deliver the large number of connected posts to the readers directly. Alternatively, we should select representative data for presentation. The challenge is how to identify important multi-modal data in the post set.

To visualize the evolution of each clue, we should select data from each stage of a clue, and we view the posts in it as a *community*, denoted by $c(sg_i^{cid})$. Here, cid is the clue id, $sg_i \in SG$ is the stage id, and c is a function used to retrieve all posts in sg_i^{cid}. To ensure a multi-modal summary, the semantic correlations among the texts and images in different posts of the clue-stage community are considered. A hybrid method is proposed to address this issue. First, representative texts from among the posts in a stage are selected to optimize the text diversity. Second, a cross-media mining approach is proposed to represent the images and texts in a shared semantic space and to retrieve images within the community based on both visual diversity and relevancy.

4.2.1 Optimized selection of important texts. In general, there are redundant textual contents in each clue, especially in each stage of a clue. Figure 4(3) shows a simple example, in which posts M7 and M8 are within the same stage (i.e., stage 2 in Fig. 4(2)) of Clue 1 and are redundant in texts; thus, only M8 is selected in the final presentation of the event. In the present work, two optimization problems are formulated to remove the redundant data in each clue: one is based on the dominating set (DS) [18], and the other uses the submodular function (SF) [20].

Dominating set. A subset of vertices in a graph is called a dominating set if every vertex in the graph is contained in the set or has a neighbor in it. The minimum-weight dominating set is the dominating set with the minimum weight. In our problem, the original posts weighted by TF-IDF are connected as a graph based on textual and temporal similarities, and the dominating set represents the set of important posts in the graph. This is an NP-hard problem, and a greedy method presented in [18] is used to attain a near-optimal solution to select important posts.

Submodular function. The function can be formulated as Eq. (3).

$$f(S) = \sum_{i \in C \setminus S} \sum_{j \in S} sim(i,j) - \lambda \sum_{i,j \in S: i \neq j} sim(i,j) \tag{3}$$

where C refers to the posts in a stage of a clue (see Fig. 4(2) for examples), S is the set of important posts, $sim(i,j)$ is the cosine similarity between posts i and j, λ is a threshold, and $\lambda \geq 0$. The important node selection problem can be formulated as a submodular function maximization under budget constraint, and we apply the greedy solution (a guaranteed near-optimal solution) presented in [20] to solve it.

4.2.2 Image selection by cross-media mining. Identifying the correlations among different media types is quite challenging because they are within different representation spaces. In the present work, a new method is proposed for the selection of important images for each stage of the clue sg_i^{cid}. First, Latent Dirichlet Allocation (LDA) [2] is used to extract the proportion of subtopics as the raw features of texts. Second, SIFT is applied to extract feature points from images, and the k-means is used to cluster feature points to obtain the "bag of visual words" vocabulary. Based on the learned bag of visual words, a TF-IDF vector can be obtained for each image. Third, the Cross-Media Mining (CMM) algorithm is proposed to attain a multimodal event summary.

As shown in Algorithm 1, the inputs of CMM incorporate the image set I from all posts in sg_i^{cid}, a textual feature vector x generated by LDA over all textual contents in sg_i^{cid}, and a set of visual features $y = \{y_1, y_2, \ldots, y_m\}$, where m is the size of the image set. The output of Algorithm 1 is a selected set of diverse and relevant images O, the size of which is predefined according to the requirements (e.g., there are generally three embedded images for each post in Sina Weibo). To ensure output diversity, the images are first grouped into r clusters by using the k-means, where r is equal to the size of O. Then, the SCM-seq method [44] is used to compute the *hash-code* for both texts and images such that the two media types are expressed in the same feature space. In SCM-seq, training with a set of given text-image pairs is needed to characterize the correlations of textual and visual contents in a shared semantic space. News articles from popular websites are used as the training set to extract the pairs (images and their annotations), which generally show high-quality correlation characterization. The computed hash-code can be used to calculate the hamming distance between each image and the texts in $c(sg_i^{cid})$. Finally, to enhance the image-text relevancy, the semantically closest image in each cluster is selected to form O. Overall, both *image clustering* and *text-image semantic distance measurement* are used to achieve a balance between visual diversity and relevancy.

Algorithm 1 CMM

Input:
Text feature vector x of $c(sg_i^{cid})$;
A set of image features $y = \{y_1, y_2, \ldots y_m\}$ of I;
Candidate image set $I = \{i_1, i_2, \ldots i_m\}$;
Output:
Selected Image set O;

```
C = KMeans(y,r)
Hashcode_t = SCM-Seq(x)
for each c_i ∈ C do
    min = INITIALVALUE
    for each y_m ∈ c_i do
        Hashcode_v = SCM-Seq(y_m)
        dis = HammingDistance(Hashcode_t, Hashcode_v)
        if dis < min then
            min = dis
            inx = m
        end if
    end for
    O = O ⋃ i_inx
end for
return O;
```

5 EVALUATION AND DISCUSSION

5.1 Experiment Design

The performance of CrowdStory is determined by the following methods: 1) *evolutionary clue characterization*, whether the two-step grouping approach with temporal and textual features can result in a fine-grained summary compared with the state-of-the-art methods; 2) *inter-clue association identification*, whether the visual and community interaction contexts can help identify valuable associations among clues; and 3) *multimodal storyline visualization*, whether the present work enables the selection of representative textual and visual contents.

5.1.1 Data sets. Four data sets are used in the tests. The first two are crawled from Sina Weibo, which can be regarded as the Chinese version of Twitter. The first data set is about an explosion accident in Tianjin and covers the period of August 13 to September 12, 2015. The second is about a terrorist attack in Paris on November 13, 2015. Details of these data sets are given in Table 2. They are used to evaluate the performance of fine-grained event summarization and inter-clue association identification. The third data set is a larger-scale open data set for event detection collected from Twitter [23]. In this data set, each event corresponds to a set of clusters, which can be considered as the clues in the present work. Each tweet is assigned a cluster id for labeling to which clue of an event it belongs. Table 3 presents the

details of this data set, which is used to verify the effectiveness of using visual and interaction contexts for association identification. The fourth data set is crawled from the leading news Web sites in China (Sina[5] and Sohu[6]). It includes 1,243 images about the explosion accident in Tianjin, 1,002 images about the terrorist attack in Paris, and the annotations for each image. This data set is used to train SCM-Seq.

Table 2. **Statistic Information of two Crawled Sina Weibo Datasets.**

Events	Original Posts	Images	Reposts	Comments	Timespan
Tianjin	303	1,119	572,446	275,850	2015.08.13-2015.09.12
Paris	512	1,339	456,882	61,612	2015.11.14-2015.12.07

Table 3. **Statistic Information of the Twitter Dataset.**

Event Number	Clusters	Tweets	Images
100	9,466	415,564	1,586,140

Data annotation. Event summarization is subjective. To obtain the ground truth, five human labelers were recruited to manually select representative posts from the first two data sets. The labelers were advised to select representative posts based on the following three questions.

- Is the textual content related to the event?
- Does the textual content reveal something important about the event that the public care about?
- Does the textual content provide supplementary knowledge about the event, such as background or practical tips?

To retrieve the ground truth for the inter-clue association identification, the human labelers were also asked to tag each identified association as one of these four labels: *causality*, *supplement*, *relevant*, and *weak*. The first three are defined in Section 3.1. A new association called *weak* is proposed, which denotes that the two clues relate to the same event, but the link between them is not that critical.

5.1.2 Experiment settings. In the text preprocessing, Jieba[7] is applied to segment Chinese sentences into words and remove stop words. There are three parameters that should be determined in our system: thresholds δ and θ, which are used in clue detection and event segmentation, and λ, which is used in the sub-modular function.

- The identification of similar texts is similar to the clustering problem; thus, internal clustering validation measures are used to determine the suitable value of δ.
- Based on empirical experience, θ is set to 0.1.
- The value of λ is determined by using ROUGE [19] scores, which are widely applied in the evaluation of document summarization performance.

The fourth data set is used to train SCM-Seq. Each image is represented as a 128-dimensional bag of visual SIFT feature vector, and each text description is represented by a 10-dimensional feature vector generated by LDA.

5.1.3 Baselines and metrics. Baselines are defined for the different technical parts of the present work. For the evolutionary event summary, we refer to our method as S_SF and S_DS. S and NS indicate the presence and absence of event segmentation, respectively; DS refers to the dominating set; ST denotes the Steiner tree algorithm [18], and SF is the sub-modular function. The baselines are introduced below.

[5] http://www.sina.com
[6] http://www.sohu.com
[7] https://github.com/fxsjy/jieba

• **NS_DS_ST [18]**. In this approach, DS is used to select representative posts, and ST is applied to add nodes and connections to generate an event storyline. However, this method does not include event segmentation as proposed in the present work. We chose this as the state-of-the-art method. Previous works [35, 18, 47] have reported that it performs better than other event summarization methods.

• **NS_SF [20]**. This approach uses SF for multi-document summarization in which event segmentation is also not considered.

For multimodal data selection, two baselines, namely, Cluster and SCM-Seq, are used in comparison with CMM. The Cluster method selects images that are close to the centroid in each cluster as the final results. SCM-Seq [44] selects images that are close to the textual content in hamming distance without clustering.

The following metrics are also used in the system evaluation.

• **The modified Hubert Γ statistic [14]**. This is used to determine the parameter δ, as defined in Eq. (4):

$$\Gamma = (1/M)\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} P(i,j)\cdot Q(i,j) \tag{4}$$

where $M = N(N-1)/2$, N is the number of posts, P is the textual similarity matrix of the post set, and Q is an $N \times N$ matrix whose (i, j) element is equal to the distance between the representative points of the clusters in which the posts p_i and p_j belong to. In an appropriate clustering result, P and Q will be in close agreement and the value of Γ will be high. Halkidi et al. [11] suggested that the place where the knee occurs is an indication of the number of clusters.

• **Accuracy**. By using the association labels given by the five human labelers as the ground truth, the percentage of strong correlation (causality, supplement, and relevant) can be computed to evaluate the performance of inter-clue association identification.

• **ROUGE**. This set of metrics is widely used for evaluating the performance of document summarization. Previous studies [19] have suggested that ROUGE-1 and ROUGE-L perform well in evaluating very short summaries, whereas ROUGE-1 and ROUGE-2 are well-suited for multi-document summarization. Therefore, ROUGE-1, ROUGE-2, and ROUGE-L are used in the present work.

5.1.4 User study. Another user study was carried out to evaluate the storyline generation. Eighteen students (including 3 females) with ages ranging from 19 to 27 years from the Department of Computer Science of our university were recruited into the study through email lists and posters.

In the first phase, the subjects were asked to evaluate our multimodal data selection method. To validate the performance under rich data environments, only clue-stage sg_i^{cid} with more than 10 images was chosen from the two Sina Weibo data sets; 27 clue stages satisfied this criterion. CMM and two baselines (Cluster and SCM-Seq) were then used to individually select a fixed number of images (three in the test) from each sg_i^{cid} chosen. The textual contents of $c(sg_i^{cid})$ were first presented to the subjects, who were then asked to choose one group of images obtained by the three methods that can best characterize the textual contents. Because there were two clue stages with the same results determined by different methods, a total of 450 (25*18) comparable results were obtained. Finally, the *support ratio* for different methods was computed.

In the second phase, the usage and results of CrowdStory for different events were presented to the subjects, who were then asked to rank the storyline on a 5-point Likert scale (with 5 as the top score) based on the following questions: (1) Does the storyline depict various clues of the event? (2) Do the connections well characterize the associations of clues? (3) Does the storyline present more informative data compared with news articles? (4) Does the storyline show the dynamic evolution of different clues of the event?

5.2 Experiment Results

5.2.1 Parameter selection. Figure 5 shows the value of the modified Hubert Γ statistic for δ. The data indicate that 0.5 is the knee point; thus, δ is set to 0.5 in the experiments. Figure 6 shows the ROUGE-1 and ROUGE-L scores for alternative values of parameter λ on the two crawled data sets. For the Tianjin/Paris data set, 0.9/0.8 is an appropriate value for λ because it leads to the best ROUGE scores.

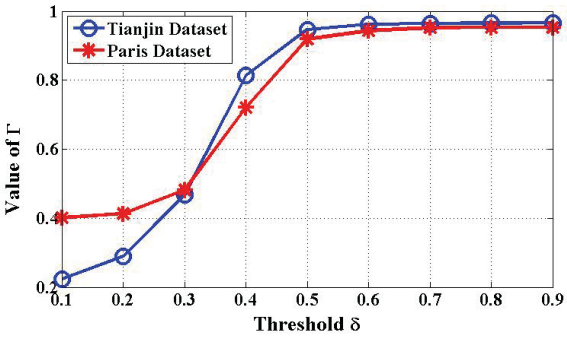

Fig. 5. The value of modified Hubert Γ statistic for each δ.

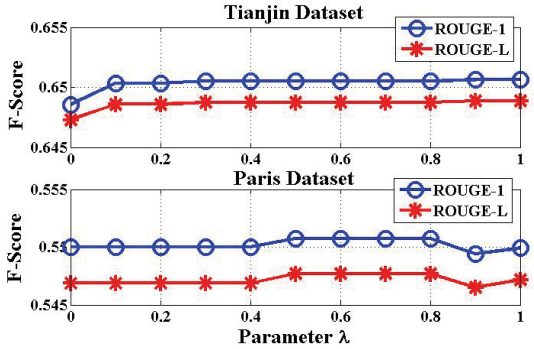

Fig. 6. Performance of parameter λ on the two datasets.

Table 4. The comparison among different summarization methods.

Dataset	Methods	ROUGE-1	ROUGE-2	ROUGE-L
Tianjin	NS_DS_ST	0.22486	0.16811	0.22144
	NS_SF	0.64476	0.58986	0.64336
	S_DS	0.66704	0.61871	0.66583
	S_SF	0.66597	0.62139	0.66476
Paris	NS_DS_ST	0.25612	0.15942	0.24975
	NS_SF	0.50893	0.41435	0.50428
	S_DS	0.57986	0.47076	0.57735
	S_SF	.0.64538	0.54546	0.64263

5.2.2 Evolutionary event summary. Table 4 shows the results of different summarization methods, which indicate that, in general, our methods outperform the baselines. Compared with the methods without event segmentation (i.e., NS_DS_ST and NS_SF), our methods (S_DS and S_SF) improve all three ROUGE scores. Figure 7 shows the number of posts in different summaries. NS_DS_ST clearly selects the minimum number of posts in a summary, which we believe is why NS_DS_ST has the lowest ROUGE scores. Because DS_ST can minimize the textual redundancy among posts, it may also filter different stages of the event that are similar in textual contents. Our event segmentation method can generate a fine-grained summary because it depicts the dynamic evolution of the event.

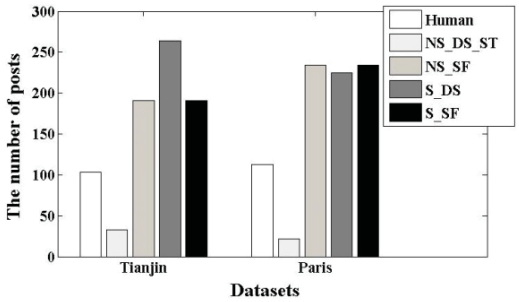

Fig. 7. The number of posts in summary by different methods.

Table 5 shows the results of the usage of NS_DS_ST and S_DS on the summarization of the death toll for the Tianjin explosion event. S_DS clearly generates a finer-grained summary compared with NS_DS_ST; S_DS shows the dynamic evolution of the death toll. This result indicates that a fine-grained summary can be obtained through summarization within each stage. All the above experiment results indicate that our methods outperform the state-of-the-art techniques.

Table 5. An example of the death toll summarized for Tianjin explosion via NS-DS+ST and S-DS.

Methods	Death Toll Summary
NS_DS_ST	2015-08-14 08:29:13| 32 hours after the Tianjin explosion happened, 50 people killed including 17 firefighters.
S_DS	2015-08-15 21:49:17| The death toll of Tianjin port explosion rises to 104.
	2015-08-25 15:31:27| The death toll of Tianjin port explosion accident rises to 135.
	2015-09-11 11:26:43| There are 165 victims found in the explosion accident, and there are still 9 people lost contact.

5.2.3 Association identification. To validate the performance of our association identification approach, statistical analysis was carried out to measure whether the posts with common topics have more visual similarities and interactions. The identification results are further compared with human-labeled data.

Statistical analysis. Posts belonging to the same event often have similar images, whereas the others do not. A total of 1178 images were sampled from 623 tweets in the Twitter data set; these tweets belonged to 375 clusters (41 events). The average image similarity was calculated in three situations: posts of the same cluster (PC), posts on the same event (PSE), and posts on different events (PDE). The semantic association between posts in the three situations clearly ranged from strong to weak. Therefore, if the computed visual/interaction similarity among the posts follows this trend, the assumption that using the two contexts for association identification can be proved.

The average visual similarities for PC, PSE, and PDE are 0.85%, 0.15%, and 0.03%, respectively. The average interaction group similarities for PC, PSE, and PDE are 1.99%, 1.1%, and 0.2%, respectively. All the results indicate that posts of the same cluster have the highest visual or interaction similarities, whereas posts about different events present the lowest similarity. This finding verifies the usability of our inter-clue association identification method. Figure 8 shows the percentage of strong correlations (causality, supplement, and relevance) identified by our method compared with the human-labeled data discussed in Section 5.1.1. The results are all above 40% and reaches 60% in the best case. This indicates that post-level fusion performs better than clue-level fusion because the latter can miss crucial information in the data integration. For the Paris data set, the percentages of correctly identified strong correlations are 12% for causality, 29% for supplement, and 59% for relevance.

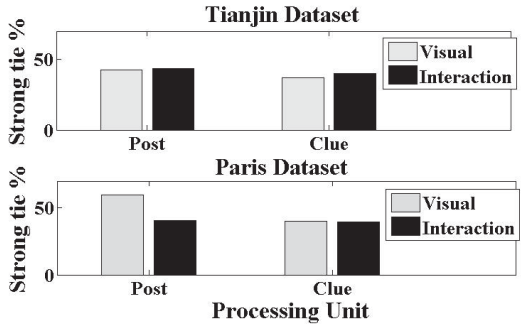

Fig. 8. The accuracy of association identification in the user study.

5.2.4 Multimodal data selection. The user study results indicate that CMM, Cluster, and SCM-Seq receive 39%, 37%, and 24% support (refer to Section 5.1.4 for the introduction), respectively. In the interviews, the most common reasons given by the subjects for their decision are "to the point" and "diversity in coverage." In the following, two cases are presented to explain the experiment results.

Case 1. Figure 9 shows the selected images obtained by using different methods. The text content is about the riot scene in the Republic Square in Paris. In this case, most of the subjects (61%) supported CMM. They explained that the images provided by CMM depicted multiple views of the clashes between the demonstrators and the police. In particular, the second image in CMM matches well the textual content that, "*Even the flowers and candles to honor the dead are destroyed...*"; this shows the power of cross-media mining. The results of SCM-Seq lack diversity. For instance, the latter two images show the same scene taken from different angles.

Case 2. Figure 10 is about the street scene during the Paris terrorist attack, with representative textual contents such as "*a terrorist attack happened in Paris, more than 100 people were killed, and more than 200 people were injured.*" In this case, most of the subjects supported the Cluster approach (67%). Compared with the other two methods, the images in Cluster are more diverse, depicting the street scene from different perspectives, including aftermath hugs, rescue efforts, and flowers and candles of mourning in the street.

Based on these case studies, SCM-Seq can obtain relevant images regarding the query texts but cannot attain result diversity. In contrast, Cluster achieves result diversity but suffers loss of data relevancy. CMM can combine the advantages of the two methods and attain a balance between relevancy and diversity in visual content selection.

Fig. 9. The selected image groups for the three methods in case 1.

Fig. 10. The selected image groups for the three methods in case 2.

5.2.5 Event storyline: an illustration. Figure 11 shows an example of a storyline for the Tianjin explosion event. There is a main clue, which contains the maximum number of posts. The main clue reports the evolution of the death toll throughout the whole explosion. The branches of the main clue are the associated clues. To make the storyline clear and easy to follow, each associated clue only has a representative post linked to the main clue. There are two kinds of associations in this example. The first is *supplement*, e.g., firsthand and background knowledge of chemical storage in China. The second is *relevant*, which pertains to relevant clues referring to the same subtopic, including the progress of rescue efforts, the circumstances of the accident, and the reconstruction after the accident. The details presented in Fig. 11 show the clue (red circle) reporting the evolution of the rescue process for the event. For instance, "*a person named Zhou Ti is rescued at 7:05 on August 14, and he is a firefighter; another survival is rescued several hours later.*"

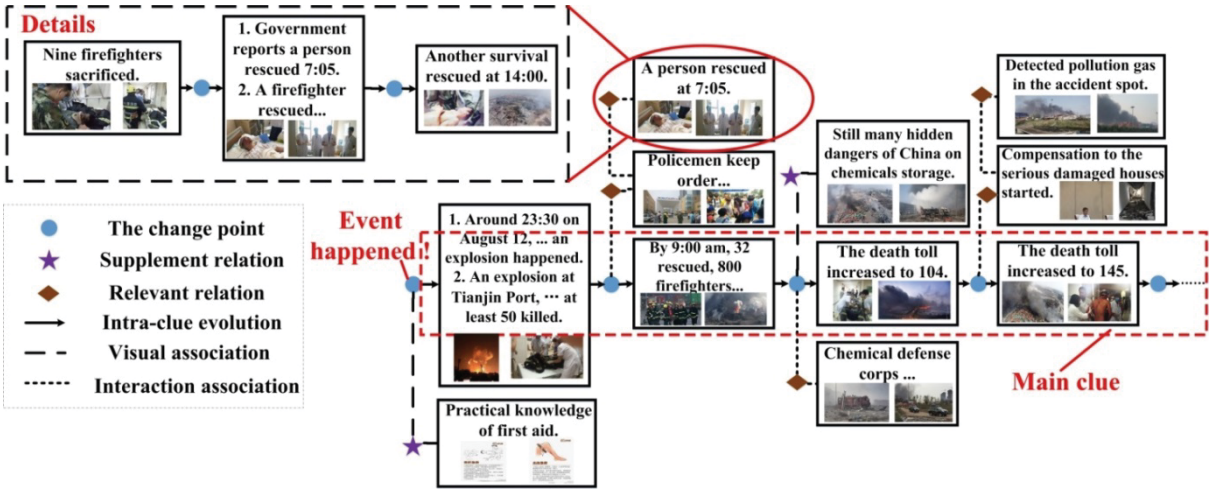

Fig. 11. An illustrative example of a storyline for the Tianjin explosion event.

5.2.6 Subjective user scoring. Figure 12 shows the box plots of the subjective user evaluation of CrowdStory. The median scores for all questions are above 3.5, indicating that the subjects are interested in our work. Cohen's kappa coefficient is used to measure the inter-rater agreement among all the subjects [9]; the obtained values are 0.5 and 0.6 for the Tianjin and Paris data sets, respectively. The values are not quite high as we use the 5-point Likert scale for user scoring. According to [9], the results suggest that the scoring has some consistency among the subjects. The results lead to the following observations: 1) our approach decreases the burden of reading original posts, 2) our method provides rich and comprehensive information, and 3) the associations identified are meaningful. It should be noted that there could be possible bias introduced by merely question answering in the user study. In the future work, we will on one hand improve

the questionnaire design and, on the other hand, have in-depth interviews with more participants from different fields for their feedback to improve our system.

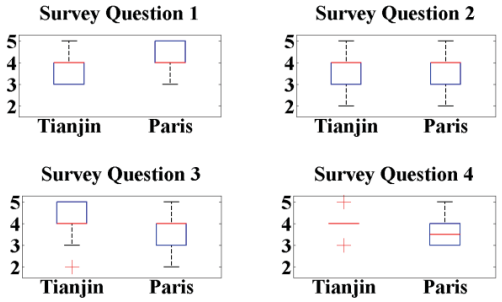

Fig. 12. The subjective evaluation of CrowdStory via a user study.

5.3 Discussions

In this subsection, the research findings and potential future directions to improve our work are discussed.

Learning with crowd intelligence. This study reports our efforts in fine-grained and correlation-rich storyline generation. One of the major contributions of this work is that the implicit human intelligence extracted from post generation and community interaction is leveraged for inter-clue association identification. The crowd intelligence–based learning approach can be further studied and applied in other crowdsourced data understanding systems. For example, in our previous work on FlierMeet [10], crowd intelligence (e.g., crowd-object interaction patterns, social relations, and user preferences) is leveraged for mobile crowdsourced picture quality estimation and semantic tagging. CrowdWiFi [41] leverage crowd intelligence to understand roadside WiFi network and provide opportunistic data services.

Association identification. Human behaviors are complex, and the associations identified by human intelligence are sometimes not that strong. Thus, our approach to association identification should be improved to eliminate weak ties. Traditional news media have the merit of a logical relationship presentation, and semantic relations often exist among the paragraphs of a news article and its embedded links. This can be used as a complementary resource to enhance correlation identification.

Multimodal data selection. In the current system, we use only textual and visual information to select data. However, the sentiments among users can be reflected in reposts and comments, which can lead to crucial information with a shared focus. In future works, we plan to improve the data selection method by using more information from microblogs.

Personalized storyline generation. Graph mining is used to detect clues. However, the number of clues that should be obtained is usually not clear, and the large number of clues also affects the readability of the storyline. For each event, users often show interest in a number of clues. Thus, we intend to improve our method by allowing people to specify several posts as clue seeds toward personalized storyline generation.

Social media knowledge integration. CrowdStory presents a general framework for fine-grained event storyline generation by using crowdsourced social media data. Data from both Sina Weibo (the most popular microblogging platform in China) and Twitter are used in the evaluation. Although it has been reported that the topics in the two social media Web sites rarely overlap [46], we find that influential disaster events (e.g., the Tianjin explosion event) are discussed in both sites. However, the data reported by the two Web sites are often complementary. For example, on Twitter, people abroad may report the experiences of their countries in dealing with similar disaster events. Therefore, combining the data from the two Web sites in the event summary will help provide full knowledge about a particular event. In future research, we plan to study the multisource event knowledge integration problem.

System evaluation. In the current study, four data sets are used to evaluate the effectiveness of our methods. In the future, we intend to increase the experiments and data sets for a more in-depth evaluation and analysis. First, the event data set will be enlarged to show the performance of our methods in various situations. Second, for more objective data labeling, event report data from news Web sites will be used, with the set of news items about an important event grouped and linked based on their strong connections. In the present work, the findings about inter-clue association identification

may be biased because the ground truth is determined by human labelers. Thus, a more objective data labeling method can provide expert-level knowledge and help evaluate the effectiveness of our association identification method.

6 CONCLUSION

This study presents CrowdStory, which investigates the rich crowdsourced data in microblogs for use in event characterization. An evolutionary and multi-clue–based approach that uses textual similarity and temporal contexts is developed to obtain a fine-grained event summary. Visual contents and community interactions are applied in inter-clue association identification and storyline generation. A cross-media mining approach is proposed for visual content selection based on both data diversity and relevancy. Our system is evaluated with the use of crawled event data sets from social media Web sites. The results show that our method has a better summarization performance than the state-of-the-art methods. In addition, user interactions and visual contexts are effective in association identification; in the best case, up to 60% strong correlations (e.g., causality and supplement) can be identified. Finally, the user study results indicate that the cross-media data selection approach is promising. In future works, we intend to enhance our association identification approach to eliminate weak ties. Personalized and human-in-the-loop storyline generation approaches will also be investigated.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (grant nos. 61332005, 61373119, and 61402369), and the National Basic Research Program of China (grant no.2015CB352400).

REFERENCES

[1] J. Bian, Y. Yang, H. Zhang, and T. Chua, "Multimedia Summarization for Social Events in Microblog Stream," IEEE Transactions on multimedia, vol. 17, no. 2, pp. 216–228, 2015.
[2] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," Journal of machine Learning research, vol. 3, no.1, pp. 993-1022, 2003.
[3] F. Buckley and M. Lewinter, "A friendly introduction to graph theory," Prentice Hall, 2003.
[4] D. Chakrabarti and K. Punera, "Event Summarization using Tweets," in Proc. of ICWSM'11, AAAI, 2011, pp. 66–73.
[5] Y. Chen, et al., "Event detection using customer care calls," in Proc. of INFOCOM'13, 2013, pp. 1690-1698.
[6] D. Corney, C. Martin, and A. Göker, "Two sides to every story: Subjective event summarization of sports events using Twitter," in Proc. of the SoMuS ICMR 2014 Workshop, ACM, 2014.
[7] A. Cui, M. Zhang, Y. Liu, et al., "Discover breaking events with popular hashtags in twitter," in Proc. of CIKM'12, ACM, 2012, pp. 1794-1798.
[8] N. Dehghani, M. Asadpour, "Graph-based Method for Summarized Storyline Generation in Twitter," arXiv preprint arXiv:1504.07361, 2015.
[9] J.L. Fleiss, B. Levin, C.P. Myunghee, "The measurement of interrater agreement. Statistical methods for rates and proportions," no. 2, pp. 212-236, 1981.
[10] B. Guo, H. Chen, Z. Yu, X. Xie, S. Huangfu, D. Zhang, "FlierMeet: A Mobile Crowdsensing System for Cross-Space Public Information Reposting, Tagging, and Sharing," IEEE Transactions on Mobile Computing, vol. 14, no. 10, 2015, pp. 2020-2033.
[11] M. Halkidi, Y. Batistakis, and M. Vazirgiannis, "On clustering validation techniques," Journal of intelligent information systems, vol. 17, no. 2-3, pp. 107-145, 2001.
[12] T. Hua, X. Zhang, W. Wang, et al., "Automatical Storyline Generation with Help from Twitter," in Proc. of CIKM'16, ACM, 2016, pp. 2383-2388.
[13] L. Huang and L. Huang, "Optimized Event Storyline Generation based on Mixture-Event-Aspect Model," in Proc. of EMNLP'13, ACL, 2013, pp. 726–735.
[14] L. Hubert and P. Arabie, "Comparing partitions," Journal of classification, vol. 2, no. 1, pp. 193-218, 1985.
[15] J. Kim, A. Monroy-Hernandez, "Storia: Summarizing social media content based on narrative theory using crowdsourcing," in Proc. of CSCW'16. ACM, 2016, pp. 1018-1027.
[16] J. Kleinberg, D. Easley, Networks, Crowds, and Markets, Cambridge University Press, 2010.
[17] P. Lee, et al., "CAST: A Context-Aware Story-Teller for Streaming Social Content," in Proc. of CIKM'14, ACM, 2014, pp. 789-798.
[18] C. Lin, et al., "Generating event storylines from microblogs," in Proc. of CIKM'15, ACM, 2012, pp. 175-184.
[19] C. Y. Lin, "Rouge: A package for automatic evaluation of summaries," in Proc. of ACL'04 workshop, 2004, pp. 1-8.
[20] H. Lin and J. Bilmes, "Multi-document summarization via budgeted maximization of submodular functions," in Proc. of ACL'10, 2010, pp. 912-920.
[21] D. G. Lowe, "Object recognition from local scale-invariant features," in Proc. of ICCV'99, 1999, pp. 1150-1157.
[22] M. Mathioudakis, N. Koudas, "TwitterMonitor: trend detection over the Twitter stream," in Proc. Of SIGMOD'10, ACM, 2010, pp. 1155-1158.
[23] A. J. McMinn, Y. Moshfeghi, and J. M. Jose, "Building a large-scale corpus for evaluating event detection on twitter," in Proc. of CIKM'13, ACM, 2013, pp. 409-418
[24] P. Meladianos, et al., "Degeneracy-based real-time sub-event detection in twitter stream," in Proc. of ICWSM'15, 2015, pp. 248-257.
[25] J. Nichols, J. Mahmud, and C. Drews, "Summarizing sporting events using twitter," in Proc. of IUI'12, ACM, 2012, pp. 189-198.
[26] J. Pearl, "Causality," Cambridge university press, 2009.
[27] K. Rudra, et al., "Extracting Situational Information from Microblogs during Disaster Events: a Classification-Summarization Approach," in Proc. of CIKM'15, ACM, 2015, pp. 583-592.
[28] T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes Twitter users: real-time event detection by social sensors", in Proc. of WWW'10, ACM, 2010, pp. 851-860.
[29] M. Schinas, S. Papadopoulos, Y. Kompatsiaris, et al., "Visual event summarization on social media using topic modelling and graph-based ranking algorithms," in Proc. of ICMR'15, ACM, 2015, pp. 203-210.
[30] M. Schinas, S. Papadopoulos, Y. Kompatsiaris, et al., "StreamGrid: Summarization of Large Scale Events using Topic Modelling and Temporal Analysis," SoMuS@ ICMR, 2014.
[31] X. Shang, H. Zhang, T S. Chua, "Deep learning generic features for cross-media retrieval," International Conference on Multimedia Modeling. Springer

International Publishing, 2016, pp. 264-275.
[32] C. Shen, F. Liu, F. Weng, et al., "A Participant-based Approach for Event Summarization Using Twitter Streams," HLT-NAACL, 2013, pp. 1152-1162.
[33] J. Song J, Y. Yang, Y. Yang, et al., "Inter-media hashing for large-scale retrieval from heterogeneous data sources," in Proc. of SIGMOD'13. ACM, 2013, pp. 785-796.
[34] S. Tang, F. Wu, S. Li, et al., "Sketch the Storyline with CHARCOAL: A Non-Parametric Approach," IJCAI, 2015, pp. 3841-3848.
[35] D. Wang, T. Li, and M. Ogihara, "Generating Pictorial Storylines Via Minimum-Weight Connected Dominating Set Approximation in Multi-View Graphs", in Proc. of AAAI'12, 2012.
[36] W. Wang, et al., "Effective deep learning-based multi-modal retrieval," The VLDB Journal, vol. 25, no. 1, pp. 79-101, 2016.
[37] Y. Wei, Y. Zhao, Z. Zhu, et al., "Modality-dependent cross-media retrieval," ACM Transactions on Intelligent Systems and Technology (TIST), vol. 7, no. 4, pp. 57, 2016.
[38] A. Weiler, M H. Scholl, F. Wanner, et al., "Event identification for local areas using social media streaming data," in Proc. of the ACM SIGMOD Workshop on Databases and Social Networks. ACM, 2013, pp. 1-6.
[39] A. Witayangkurn, T. Horanont, Y. Sekimoto, et al., "Anomalous event detection on large-scale gps data from mobile phones using hidden markov model and cloud platform," in Proc. of UbiComp'13. ACM, 2013, pp. 1219-1228.
[40] F. Wu, X. Lu, Z. Zhang, et al., "Cross-media semantic representation via bi-directional learning to rank," in Proc. of MM'13, ACM, 2013, pp. 877-886.
[41] D. Wu, Q. Liu, Y. Li, et al., "Adaptive Lookup of Open WiFi Using Crowdsensing," IEEE/ACM Transactions on Networking, vol. 24, no. 6, pp. 3634-3647, 2016.
[42] J. Xu, T C. Lu, "Seeing the big picture from microblogs: Harnessing social signals for visual event summarization," in Proc. of IUI'15. ACM, 2015, pp. 62-66.
[43] Z. Yu, F. Wu, Y. Yang, et al., "Discriminative coupled dictionary hashing for fast cross-media retrieval," in Proc. of SIGIR'14, ACM, 2014, pp. 395-404.
[44] D. Zhang and W.J. Li, "Large-Scale Supervised Multimodal Hashing with Semantic Correlation Maximization," in Proc. of AAAI'14, 2014.
[45] Q. Zhang, B. Goncalves, "Topical differences between Chinese language Twitter and Sina Weibo," in Proc. of WWW'16, 2016.
[46] J. Zhou, G. Ding, Y. Guo, "Latent semantic sparse hashing for cross-modal similarity search," in Proc. of SIGIR'14, 2014, pp. 415-424.
[47] W. Zhou, et al., "Generating textual storyline to improve situation awareness in disaster management," in Proc. IRI'14, 2014, pp. 585-592.
[48] L. J. Griffin, "Narrative, event-structure analysis, and causal interpretation in historical sociology," American journal of Sociology, vol. 98, no. 5, pp. 1094-1133, 1993.

Received February 2017; revised May 2017; accepted July 2017.