Cross-media Event Extraction and Recommendation

Di Lu[1], **Clare R. Voss**[2], **Fangbo Tao**[3], **Xiang Ren**[3], **Rachel Guan**[1], **Rostyslav Korolov**[1], **Tongtao Zhang**[1], **Dongang Wang**[4], **Hongzhi Li**[4], **Taylor Cassidy**[2], **Heng Ji**[1], **Shih-fu Chang**[4], **Jiawei Han**[3], **William Wallace**[1], **James Hendler**[1], **Mei Si**[1], **Lance Kaplan**[2]

[1] Rensselaer Polytechnic Institute
{lud2,jih}@rpi.edu
[2] Army Research Laboratory, Adelphi, Maryland
{clare.r.voss,taylor.cassidy,lance.m.kaplan}.civ@mail.mil
[3] Computer Science Department, University of Illinois at Urbana-Champaign
hanj@illinois.edu
[4] Electrical Engineering Department, Columbia University
sfchang@ee.columbia.edu

Abstract

The sheer volume of unstructured multimedia data (e.g., texts, images, videos) posted on the Web during events of general interest is overwhelming and difficult to distill if seeking information relevant to a particular concern. We have developed a comprehensive system that searches, identifies, organizes and summarizes complex events from multiple data modalities. It also recommends events related to the user's ongoing search based on previously selected attribute values and dimensions of events being viewed. In this paper we briefly present the algorithms of each component and demonstrate the system's capabilities [1].

1 Introduction

Every day, a vast amount of unstructured data in different modalities (e.g., texts, images and videos) is posted online for ready viewing. Complex event extraction and recommendation is critical for many information distillation tasks, including tracking current events, providing alerts, and predicting possible changes, as related to topics of ongoing concern. State-of-the-art Information Extraction (IE) technologies focus on extracting events from a single data modality and ignore cross-media fusion. More importantly, users are presented with extracted events in a passive way (e.g., in a temporally ordered event chronicle (Ge et al., 2015)). Such technologies do not leverage user behavior to identify the event properties of interest to them in selecting new scenarios for presentation.

In this paper we present a novel event extraction and recommendation system that incorporates advances in extracting events across multiple sources with data in diverse modalities and so yields a more comprehensive understanding of collective events, their importance, and their inter-connections. The novel capabilities of our system include:

- **Event Extraction, Summarization and Search**: extract concepts, events and their arguments (participants) and implicit attributes, and semantically meaningful visual patterns from multiple data modalities, and organize them into *Event Cubes* (Tao et al., 2013) based on Online Analytical Processing (OLAP). We developed a search interface to these cubes with a novel back-end event summarization component, for users to specify multiple dimensions in a query and receive single-sentence summary responses.
- **Event Recommendation**: recommend related events based on meta-paths derived from event arguments, and automatically adjust the ranking function by updating the weights of dimensions based on user browsing feedback.

2 Overview

2.1 System Architecture

The overall architecture of our system is illustrated in Figure 1. We first extract textual event information, as well as visual concepts, events and patterns from the raw multimedia documents to construct event cubes. Based on the event cubes, the search interface (Figure 2) returns the document that best

[1] The system demo is available at: `http://nlp.cs.rpi.edu/multimedia/event/navigation_dark.html`

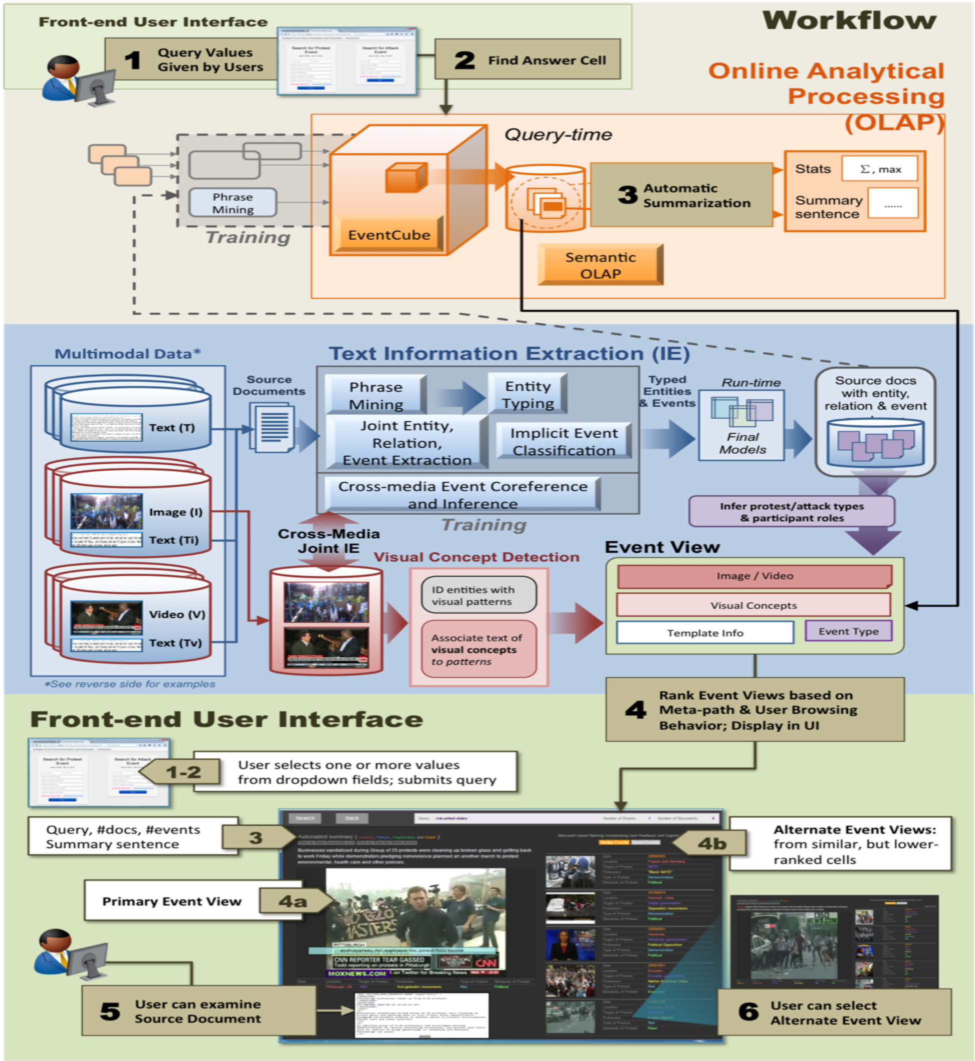

Workflow
Front-end User Interface
1 Query Values Given by Users
2 Find Answer Cell
Online Analytical Processing (OLAP)
Query-time
Phrase Mining
Training
EventCube
3 Automatic Summarization
Stats
Σ, max
Summary sentence
......
Semantic OLAP
Multimodal Data*
Text Information Extraction (IE)
Source Documents
Text (T)
Phrase Mining
Entity Typing
Joint Entity, Relation, Event Extraction
Implicit Event Classification
Cross-media Event Coreference and Inference
Training
Typed Entities & Events
Run-time
Final Models
Source docs with entity, relation & event
Infer protest/attack types & participant roles
Image (I)
Text (Ti)
Video (V)
Text (Tv)
Cross-Media Joint IE
Visual Concept Detection
ID entities with visual patterns
Associate text of visual concepts to patterns
Event View
Image / Video
Visual Concepts
Template Info
Event Type
*See reverse side for examples
4 Rank Event Views based on Meta-path & User Browsing Behavior; Display in UI
Front-end User Interface
1-2 User selects one or more values from dropdown fields; submits query
3 Query, #docs, #events Summary sentence
4b Alternate Event Views: from similar, but lower-ranked cells
4a Primary Event View
5 User can examine Source Document
6 User can select Alternate Event View

Figure 1: System Workflow

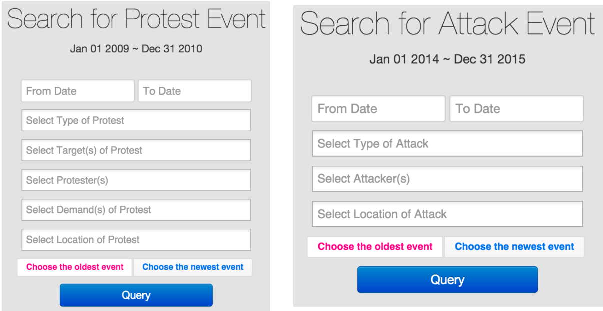
matches the user's query as the first primary event for display. A query may consist of multiple dimensions (Figure 3). The recommendation interface displays multiple dimensions and rich annotations of the primary event, and recommends similar and dissimilar events to the user.

2.2 Data Sets

To demonstrate the capabilities of our system, we use two event types, *Protest* and *Attack*, as our case studies. We collected the following data sets:

- **Protest**: 59 protest incidents that occurred between January 2009 and December 2010, from 458 text documents, 28 images and 31 videos.
- **Attack**: 52 attack incidents that occurred between January 2014 and December 2015, from 812 text documents, 46 images and 6 videos.

3 Event Cube Construction and Search



Search for Protest Event
Jan 01 2009 ~ Dec 31 2010
From Date
To Date
Select Type of Protest
Select Target(s) of Protest
Select Protester(s)
Select Demand(s) of Protest
Select Location of Protest
Choose the oldest event
Choose the newest event
Query

Search for Attack Event
Jan 01 2014 ~ Dec 31 2015
From Date
To Date
Select Type of Attack
Select Attacker(s)
Select Location of Attack
Choose the oldest event
Choose the newest event
Query

(a) Protest Event

(b) Attack Event

Figure 2: Search Interface

3.1 Event Extraction

We apply a state-of-the-art English IE system (Li et al., 2014) to jointly extract entities, relations and events from text documents. This system is based on structured perceptron incorporating multiple levels of linguistic features. However, some important event attributes are not expressed by explicit textual clues. To enrich the profile of each protest event, the system identifies two additional implicit attributes that derive from social movement theories (Della Porta and Diani, 2009; Furusawa, 2006; Dalton, 2013):

- **Types of Protest**, including *demonstration*, *riot*, *strike*, *boycott*, *picket* and *individual protest*.
- **Demands of Protest**, indicating the hidden behavioral intent of protesters, about what they desire to change or preserve, including *pacifist*, *political*, *economic*, *retraction*, *financial*, *religious*, *human rights*, *race*, *justice*, *environmental*, *sports rivalry* and *social*.

We annotated 92 news documents, 59 of which contained protests, to learn a set of heuristic rules for automatically extracting these implicit attributes from the IE outputs of these documents.

3.2 Event Cube Construction and Search

Event extraction helps in converting unstructured texts into structured arguments and attributes (dimensions). However, a user may still need to "drill down," searching back through many documents and changing query selections before finding an item of interest. We use *EventCube* (Tao et al., 2013) to effectively organize and efficiently search relevant events, and measure event similarity based on multiple dimensions. *EventCube* is a general online analytical processing (OLAP) framework for importing any collection of multi-dimensional text documents and constructing text-rich data cube. *EventCube* differs from traditional search engines in that it returns *Top-Cells*, where relevant documents are aggregated by dimension combinations.

We regard each event as a data point associated with multiple dimension values. After a user inputs a multi-dimensional query in the search interface (Figure 2), we build inverted index upon the dimensions in Event Cube to return related events, which provides much more flexible matching compared to keyword search.

4 Multi-media Event Illustration and Summarization

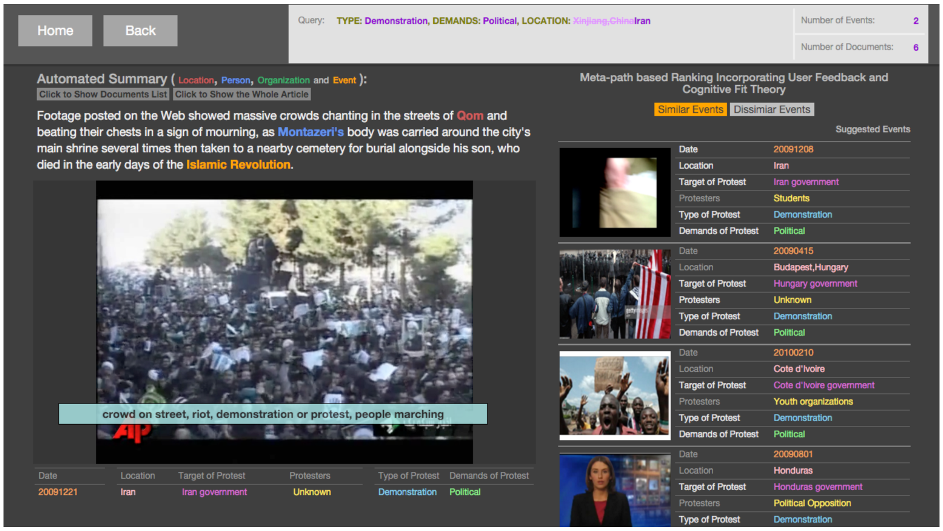
After the search interface retrieves the most relevant event ('primary event'), the user will be directed to the recommendation interface (Figure 2b) and can start exploring various events. The user's initial query is displayed at the top (gray bar) and is updated to capture new user selections. The number of events that match the user's initial query and the number of documents associated with the primary event are displayed in the gray bar, at the far right side. In addition to text IE results, we apply the following multi-media extraction and summarization techniques to illustrate and enrich event profiles.

4.1 Summarization

From each set of relevant documents, we apply a state-of-the-art phrase mining method (Liu et al., 2016) to mine top-k representative phrases. Then we construct an affinity matrix of sentences and apply *spectral clustering* to find several clustering centers (i.e., representative sentences including the most important phrases) as the summary. The user is also provided two options to show the original documents and the document containing the summary.

4.2 Visual Information Extraction

For each event, we retrieve the most representative video/image online using the key-phrases such as date, location and entities as queries. Videos and images are often more impressive and efficient at conveying information. We first apply a pre-trained convolutional neural network (CNN) architecture (Kuznetsova et al., 2012) to extract visual concepts from each video key frame based on the EventNet concept library (Ye et al., 2015). For example, the extracted visual concepts "*crowd on street, riot, demonstration or protest, people marching*" appear when the user's mouse is over the video of the primary event (Figure 2b). Then we adopt the approach described in (Li et al., 2015) which applies CNN and association rule mining technique to generate visual patterns and extract semantically meaningful relations between visual and textual information to name the patterns.



Query: TYPE: Demonstration, DEMANDS: Political, LOCATION: ~~Xinjiang,China~~Iran

Number of Events:	2
Number of Documents:	6

Automated Summary (Location, Person, Organization and Event):

Footage posted on the Web showed massive crowds chanting in the streets of Qom and beating their chests in a sign of mourning, as Montazeri's body was carried around the city's main shrine several times then taken to a nearby cemetery for burial alongside his son, who died in the early days of the Islamic Revolution.

crowd on street, riot, demonstration or protest, people marching

Date	Location	Target of Protest	Protesters	Type of Protest	Demands of Protest
20091221	Iran	Iran government	Unknown	Demonstration	Political

Meta-path based Ranking Incorporating User Feedback and Cognitive Fit Theory

Suggested Events

Date	20091208
Location	Iran
Target of Protest	Iran government
Protesters	Students
Type of Protest	Demonstration
Demands of Protest	Political
Date	20090415
Location	Budapest,Hungary
Target of Protest	Hungary government
Protesters	Unknown
Type of Protest	Demonstration
Demands of Protest	Political
Date	20100210
Location	Cote d'Ivoire
Target of Protest	Cote d'Ivoire government
Protesters	Youth organizations
Type of Protest	Demonstration
Demands of Protest	Political
Date	20090801
Location	Honduras
Target of Protest	Honduras government
Protesters	Political Opposition
Type of Protest	Demonstration

Figure 3: Recommendation Interface

5 Event Recommendation

We rank and recommend events based on **meta paths** (Sun et al., 2011), by representing the whole data set as a heterogeneous network, that is composed of multi-typed and interconnected objects (e.g., events, location, protesters, target of protesters). A meta path is a sequence of relations defined between different object types. For protest events, we define six meta paths: "event-date-event", "event-location-event", "event-target of protest-event", "event-protesters-event", "event-type of protest-event" and "event-demand of protest-event"; and four meta paths for attack events: "event-date-event", "event-location-event", "event-attackers-event" and "event-type of attack-event".

The similarity between two events is the weighted sum of the six meta path similarities. The weights are assigned dynamically by the user's activity:

- When the user clicks on a certain image/video: assign 1.0 to all meta-path similarities.
- When the user clicks on a certain dimension X: 1.0 for the similarity based on Event-X-Event, and 0.2 to other meta-paths.

To illustrate the meta paths, the dimension names of recommended events are highlighted if they share the same dimensions with the primary event. Moreover, the system is switchable between recommending the most similar and most dissimilar events with a toggle button that the user can click.

6 Conclusions and Future Work

In this paper we present a cross-media event extraction and recommendation system which effectively aggregates and summarizes related complex events, and makes recommendations based on user interests. The current system interface incorporates a medium-level human agency (the capacity of an entity to act) by allowing a human user to provide relevance feedback while driving the browsing interests by multi-dimensional recommendations. We plan to follow the cognitive fit theory (Vessey, 1991) and conduct a series of human utility evaluations to formally quantify the impact of each new component and each data modality on enhancing the speed and quality of aggregating and summarizing event-related knowledge, detecting conflicts and errors, and generating alerts.

Acknowledgments

This work was supported by the U.S. ARL NS-CTA No. W911NF-09-2-0053, DARPA Multimedia Seedling grant, DARPA DEFT No. FA8750-13-2-0041 and NSF CAREER Award IIS-1523198. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

References

Russell J Dalton. 2013. *Citizen politics: Public opinion and political parties in advanced industrial democracies*. Cq Press.

Donatella Della Porta and Mario Diani. 2009. *Social movements: An introduction*. John Wiley & Sons.

Katsuto Furusawa. 2006. Participation and protest in the european union and the 'outsider'states. *Contemporary politics*, 12(2):207–223.

Tao Ge, Wenzhe Pei, Heng Ji, Sujian Li, Baobao Chang, and Zhifang Sui. 2015. Bring you to the past: Automatic generation of topically relevant event chronicles. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics (ACL2015)*.

Polina Kuznetsova, Vicente Ordonez, Alexander C Berg, Tamara L Berg, and Yejin Choi. 2012. Collective generation of natural image descriptions. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*.

Qi Li, Heng Ji, Yu Hong, and Sujian Li. 2014. Constructing information networks using one single model. In *Proceedings of the 2014 Conference of the Empirical Methods on Natural Language Processing (EMNLP2014)*.

Hongzhi Li, Joseph G. Ellis, and Shih-Fu Chang. 2015. Event specific multimodal pattern mining with image-caption pairs. *arXiv preprint arXiv:1601.00022*.

Jialu Liu, Xiang Ren, Jingbo Shang, Taylor Cassidy, Clare Voss, and Jiawei Han. 2016. Representing documents via latent keyphrase inference. In *Proceedings of the 25th international conference on World wide web*.

Yizhou Sun, Jiawei Han, Xifeng Yan, Philip S. Yu, and Tianyi Wu. 2011. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. In *Proceedings of 2011 Int. Conf. on Very Large Data Bases*.

Fangbo Tao, Kin Hou Lei, Jiawei Han, ChengXiang Zhai, Xiao Cheng, Marina Danilevsky, Nihit Desai, Bolin Ding, Jing Ge Ge, Heng Ji, et al. 2013. Eventcube: multi-dimensional search and mining of structured and text data. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*.

Iris Vessey. 1991. Cognitive fit: A theory-based analysis of the graphs versus tables literature. *Decision Sciences*, 22(2):219–240.

Guangnan Ye, Yitong Li, Hongliang Xu, Dong Liu, and Shih-Fu Chang. 2015. Eventnet: A large scale structured concept library for complex event detection in video. In *Proceedings of the 23rd ACM International Conference on Multimedia*.