Neurocomputing 169 (2015) 169–179

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Fusing cross-media for topic detection by dense keyword groups

Weigang Zhang [a], Tianlong Chen [b], Guorong Li [c,*], Junbiao Pang [d], Qingming Huang [b,c], Wen Gao [a,e]

[a] School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China
[b] Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China
[c] School of Computer and Control Engineering, University of Chinese Academy of Sciences, Beijing 100049, China
[d] College of Metropolitan Transportation, Beijing University of Technology, Beijing 100124, China
[e] Institute of Digital Media, Peking University, Beijing 100871, China

ARTICLE INFO

Article history:
Received 9 May 2014
Received in revised form
22 November 2014
Accepted 11 February 2015
Available online 12 June 2015

Keywords:
Topic detection
Cross-media
Dense keyword group
Near-duplicate keyframe
Web video

ABSTRACT

Events are real-world occurrences that lead to the explosive growth of web multimedia content such as images, videos and texts. Efficient organization and navigation of multimedia data in the topic level can boost users' understanding and enhance their experience of the events that have happened. Due to the potential application prospects, multimedia topic detection has been an active area of research with notable progress in the last decade. Traditional methods mainly focus on single media, so the results only reflect the characteristics of one certain media and topic browsing was not comprehensive enough. In this paper, we propose a method of utilizing and fusing rich media information from web videos and news reports to extract weighted keyword groups, which are used for cross-media topic detection. Firstly by utilizing the video-related textual information and the titles of news articles, a maximum local average score is proposed to find coarse weighted dense keyword groups; after that, textual linking and visual linking are applied to refine the keyword groups and update the weights; finally, the documents are re-linked with the refined keyword groups to form an event-related document set. Experiments are conducted on cross-media datasets containing web videos and news reports. The web videos are from Youku, YouTube's equivalent in China, the news reports from sina.com, some of which contain topic-related images. The experimental results demonstrate the effectiveness and efficiency of the proposed approach.

1. Introduction

The rapid development of the Internet and multimedia technologies brings in massive multimedia data such as news, images, and videos. In the last decade much effort has been devoted to making this enormous amount of multimedia information easy to access by users. Previous researches mainly focus on near-duplicate image [6,15,19,22,23] or video [5,13,20] detection so as to achieve high accuracy of multimedia retrieval. However, without effective abstracting technologies, it is hard and time-consuming for users to digest the key information about their concerned hot topics which always lead to a surge of multimedia data in a short time. Integrating multimedia information into topics provides users with an easy way of understanding real-world events that have happened, and therefore enhances user experience. Topic detection is a technology to summarize information from unstructured multimedia data.

The objective of topic detection is to discover new or previously unidentified events concerning something unique that have happen at specific time or places [7]. In our paper, we treat topic detection and event detection equally due to the following two reasons. First, topics and events unfold over time and space with different hot degrees and various duration, thus making it difficult to distinguish between topics and events; second, traditional methods of topic detection is to uncover previously unidentified events, and then to extract topics on analysis of these detected events. Therefore, events are topics to some extent.

Topic detection detects topics in news articles or blog posts [1–4,11,18]. Chen et al. [11] extracted hot terms from the text by combining TF*PDF and the Age Theory. On the basis of these hot terms, key sentences are identified and grouped into clusters, each cluster being deemed as a topic. The method in [4] extracts the documents that are highly related to the bursty features based on timestamp. The extracted documents are then segmented to construct the event hierarchy. Sun et al. [3] integrated event-related queries, news articles, and blog posts through the notion of

* Corresponding author.
E-mail addresses: wgzhang@hit.edu.cn (W. Zhang), chentianlong0702@hotmail.com (T. Chen), liguorong@ucas.ac.cn (G. Li), junbiao_pang@bjut.edu.cn (J. Pang), qmhuang@ict.ac.cn (Q. Huang), wgao@pku.edu.cn (W. Gao).

http://dx.doi.org/10.1016/j.neucom.2015.02.083
0925-2312/

query profile, and grouped the query profiles into event fragments. Fang et al. [27] attempted to exploit appropriate priors to generate topic aspect-oriented summarization for news articles.

Textual topic detection certainly cannot satisfy the need of users for vividly browsing and understanding the hot topics of the ever-growing amount of web multimedia data. There are lots of researches on multimedia topic detection in recent years and the web video topic detection is one of the hottest. Web videos, besides including rich visual content, also contain abundant textual information such as titles, tags, surrounding texts, etc. Liu et al. [14] proposed a bipartite graph model for topic detection on web videos which was achieved by two steps: coarse topic filtering and fine topic re-ranking. Cao et al. [8] proposed an algorithm based on salient trajectory extraction from a topic evolution link graph for topic discovery. Hong et al. [17] summarized the content of videos by analyzing the tags associated with key-shots which are established and ranked on the basis of near-duplicate keyframe detection. Shao et al. [10] presented a star-structured K-partite graph-based co-clustering and ranking framework for web video topic discovery and visualization. Recently, some works [25,26,28,30] have been proposed to fuse multi-modal features for topic detection. He et al. [25] proposed a simple and effective topic detection model called the temporal Discriminative Probabilistic Model (DPM) to suffice for both offline and online topic detection tasks. Ma et al. [26] employed semantic video attributes to complete the complex event detection in videos. Location information, such as GPS data or other geographical data, is usually used for mobile landmark search [12,24]. But in [28], Liu et al. used geographical tags to discover the cultures, scenes, and human behaviors from geo-tagged videos and find the hot topics in the particular geographical regions. Liu et al. [30] proposed a generative probabilistic model named Preference-Topic Model (PTM) to introduce the dimension of user preferences to enhance the insufficient textual information for discovering topics at a satisfactory granularity on sparse data, in which the video content is rich but the textual descriptive information usually sparse. Zhang et al. [29] proposed a new approach to detect burst novel events and predict their future popularity simultaneously in microblogging.

These web video topic detection methods are mostly based on the clustering of tags or keyframes. The results of directly clustering, however, may not be satisfactory due to the deficiency of text information (noisiness, incompleteness and inconsistency) and the low accuracy of visual content analysis.

In fact, the aforementioned methods of text based topic detection or web video based topic detection are just to mine the information of topic from a certain facet for existing works mostly concentrate on single media. However, the occurrences of real-world hot topic always lead to large amounts of cross-media data, and the correlation between different media has yet to be fully mined. Fig. 1 shows an event example which is Bush being attacked by shoes. The right part consists of the event-related news reports and the left part the web videos uploaded by users. We can find that the news reports with a wealth of textual

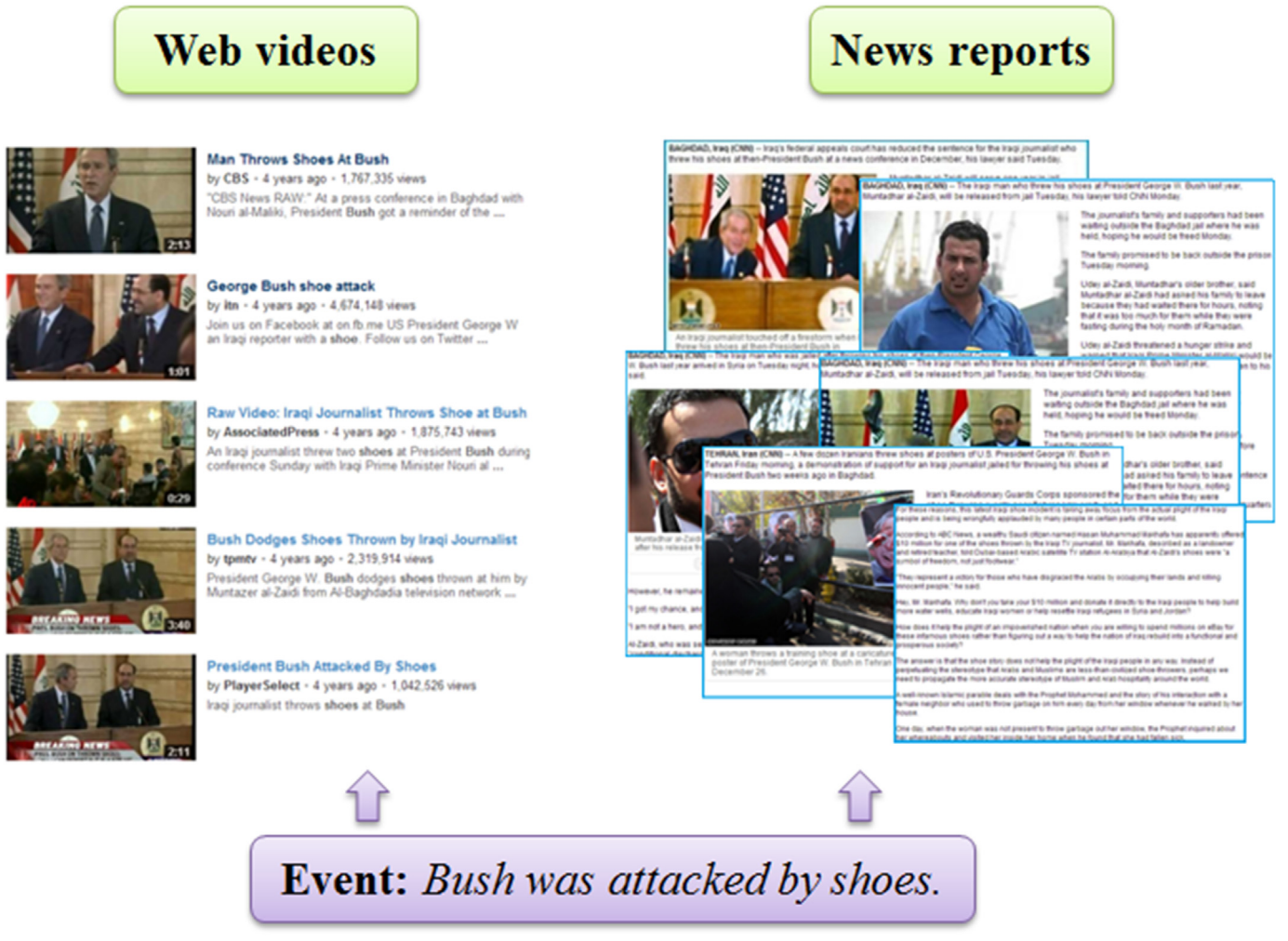

Fig. 1. An example of cross-media data about a certain event.