

EKNOT: Event Knowledge from News and Opinions in Twitter

Min Li^{1*}, Jingjing Wang^{1*}, Wenzhu Tong¹, Hongkun Yu¹,
Xiuli Ma², Yucheng Chen¹, Haoyan Cai¹ and Jiawei Han¹

¹Department of Computer Science, University of Illinois at Urbana-Champaign, IL, USA

²School of Electronics Engineering and Computer Science, Peking University, Beijing, China

¹{minli3, jwang112, wtong8, hyu50, ychen233, hcai6, hanj}@illinois.edu, ²xlma@pku.edu.cn

Abstract

We present the EKNOT system that automatically discovers major events from online news articles, connects each event to its discussion in Twitter, and provides a comprehensive summary of the events from both news media and social media's point of view. EKNOT takes a time period as input and outputs a complete picture of the events within the given time range along with the public opinions. For each event, EKNOT provides multi-dimensional summaries: a) a summary from news for an objective description; b) a summary from tweets containing opinions/sentiments; c) an entity graph which illustrates the major players involved and their correlations; d) the time span of the event; and e) an opinion (sentiment) distribution. Also, if a user is interested in a particular event, he/she can zoom into this event to investigate its aspects (sub-events) summarized in the same manner. EKNOT is built on real-time crawled news articles and tweets, allowing users to explore the dynamics of major events with minimal delays.

1 Introduction

Massive information from news media and social media is more easily accessible than ever in this big data era. News agencies deliver real-time status of major events, covering every aspect with fairly formal news articles. In the meantime, people post their comments and raise discussions via social media sites such as Twitter. The different natures of these two sources provide a complementary view of an event: A reasonably objective and accurate presentation of an event, and a view full of opinions and sentiments from the public. With such observation, we develop the system EKNOT which effectively discovers events from news and connects each event to its discussion in Twitter. Given a time period, we intend to answer the following questions: 1) What are the major events; 2) who are the key players in each event; 3) how do people talk about each event and what are their opinions; 4) when is the event and how long does the event last; 5) what are the multiple aspects (sub-events) if the event is rather big and influential? And what are the answers to the above questions for each aspect? By summarizing news and Twitter simultaneously, EKNOT provides informative and comprehensive

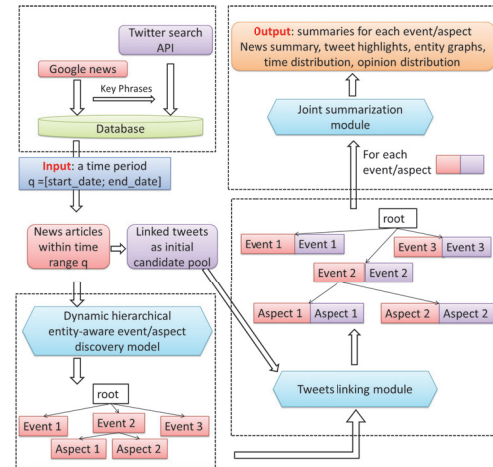


Figure 1: System Architecture. The red background indicates news side and the purple background indicates Twitter side.

summaries on the event level for users to digest the information effectively, which advances existing related research (systems)(Guo et al. 2013; Shi, Ifrim, and Hurley 2014; Tsagkias, de Rijke, and Weerkamp 2011).

2 System Overview

Figure 1 illustrates the system architecture of EKNOT containing four major modules: data collection, event/aspect discovery, tweets linking and joint summarization. The system requires a time period as input and outputs the summaries for the events and aspects within the time range.

In data collection module, EKNOT constantly crawls data from Google news. For each news article, the module extracts the keywords and queries the twitter API to obtain an initial pool of relevant tweets. Entities of the type person, organization or place are also extracted to further characterize the corpus.

Upon user query, the event/aspect discovery phase will identify major events/aspects and represent them by event/aspect descriptors containing word/time/entity distributions. A dynamic hierarchical entity-aware discovery model proposed in our recent work (Wang et al. 2015) is em-

*equal contribution

ployed to learn the optimal representation of the news events and their multiple aspects in a top-down manner. The model integrates text, entities and time with the intuition that an event/aspect must be coherent in all these dimensions. We then retrieve the most relevant tweets from the initial pool by standard information retrieval with BM25¹ using the event's word distribution as the query.

After event discovery in news and tweets, we co-rank the two sources to obtain event news summaries and tweet highlights in the joint summarization module. The co-ranking phase considers a) content and temporal *consistency* with the event/aspect; b) *coherence* between the news summary and tweet summary; c) *coverage* and *diversity* of the news(tweet) summary; and d) whether the tweet summary contains substantial *opinions/sentiments* and represents a general trend of the public. EKNOT combines Co-HITS (Deng, Lyu, and King 2009) and the Maximal Marginal Relevance (MMR) principle (Carbonell and Goldstein 1998) to determine the top ranked news and tweets. The sentiment analysis was achieved by a two-step classification model (Barbosa and Feng 2010).

Together with the news summaries and tweet highlights, EKNOT also provides the following information to facilitate event understanding:

- **Entity Graphs:** EKNOT generates the entity graph based on the computed entity distributions. The relative importance of the entity and correlation between two entities are denoted by the node size and edge width.
- **Time Spans:** EKNOT utilizes the time distribution in the descriptor to approximate the time span.
- **Opinion Graph:** EKNOT shows the positive and negative percentage of the event tweets in a pie chart.

3 Demonstration Plan

We start the demonstration by issuing a time period query. EKNOT first returns the event display page which displays all the major events within this period. Users can choose any event to see the event details and inspect its aspects. “Event Details” page displays the summary of the event. On the “Event Aspects” page, a list of aspects will be displayed as illustrated in Figure 2. This figure displays two sample aspects related to the March 2015 Apple Event: one is about Apple Watch and the other about MacBook. We are able to see the time spans on the top, news summaries on the left, tweet highlights on the right, following by an entity graph and a pie chart showing the sentiment distribution.

Acknowledgment

Research was sponsored in part by the U.S. Army Research Lab. under Cooperative Agreement No. W911NF-09-2-0053 (NSCTA), National Science Foundation IIS-1017362, IIS-1320617, and IIS-1354329, HDTRA1-10-1-0120, and grant 1U54GM114838 awarded by NIGMS through funds provided by the trans-NIH Big Data to Knowledge (BD2K) initiative (www.bd2k.nih.gov), and MIAS, a DHS-IDS Center for Multimodal Information Access and Synthesis at

¹http://en.wikipedia.org/wiki/Okapi_BM25

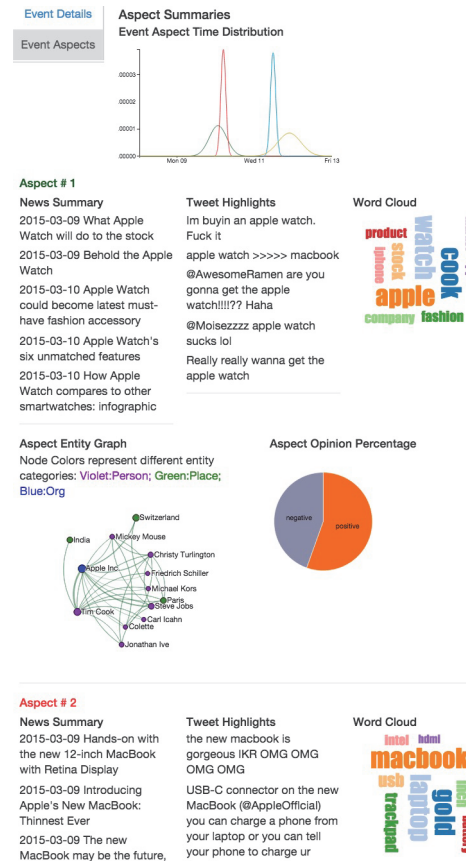


Figure 2: Aspect Details

UIUC. Xiuli Ma is supported by the National Natural Science Foundation of China under Grant No.61103025 and China Scholarship Council.

References

- Barbosa, L., and Feng, J. 2010. Robust sentiment detection on twitter from biased and noisy data. COLING Posters.
- Carbonell, J., and Goldstein, J. 1998. The use of mmr, diversity-based reranking for reordering documents and producing summaries. SIGIR, 335–336.
- Deng, H.; Lyu, M. R.; and King, I. 2009. A generalized co-hits algorithm and its application to bipartite graphs. KDD, 239–248.
- Guo, W.; Li, H.; Ji, H.; and Diab, M. T. 2013. Linking tweets to news: A framework to enrich short text data in social media. ACL, 239–249.
- Shi, B.; Ifrim, G.; and Hurley, N. 2014. Insight4news: Connecting news to relevant social conversations. ECMLPKDD, 473–476.
- Tsagkias, M.; de Rijke, M.; and Weerkamp, W. 2011. Linking online news and social media. WSDM, 565–574.
- Wang, J.; Tong, W.; Yu, H.; Li, M.; Ma, X.; Cai, H.; and Han, J. 2015. Mining multi-aspect reflection of news events in twitter: Discovery, linking and presentation. ICDM.