Semantic-Frame Representation for Event Detection on Twitter

Yanxia Qin*, Yue Zhang†, Min Zhang‡, Dequan Zheng*
*School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China
†Information Systems Technology and Design, Singapore University of Technology and Design, Singapore
‡School of Computer Science and Technology, Soochow University, Suzhou, China
{qolina, dqzheng2007, frcchang}@gmail.com, minzhang@suda.edu.cn

[1]

Abstract—**Unsupervised methods for detecting news events from tweet streams cluster feature representations via their burstiness, and filter out more news worthy clusters as outputs. Words, segments and tweets have been used as event feature representations, with segments being state-of-the-art due to their balance of expressive power and non-sparsity. However, segments do not convey structural event information, making output clusters difficult to understand. We investigate the use of semantic frame elements instead of segments as event features, observing not only better readability, but improvements in both precision and recall thanks to the effect of noise-filtering in frame extraction.**

***Keywords*-tweet; event detection; semantic frame; representation; unsupervised;**

I. Introduction

Event detection from large scale real-time tweets can assist public opinion monitoring, advertising and brand image maintenance. Both supervised and unsupervised methods have been used for tweet event detection. While supervised methods [1], [2] are typically designed to monitor certain event types, such as earthquakes and concerts, unsupervised methods can be used for open domain events. Ritter et al. [3] proposed a semi-supervised event detection method focusing on extracting open domain events in specified structures including named entities, event phrases, time and event category. In this paper, we investigate unsupervised event detection from tweets focusing on detecting open domain events without predefined structures.

Existing methods typically involve two steps. First, features are extracted to represent tweets, and clustered into groups. Second, feature clusters are filtered so that those that are more likely to represent events are used as outputs. Different feature representations have been investigated, which include words [4], [5], segments [6], [7] and tweets themselves [8], [9]. Note that segments are meaningful n-grams yielded by segmenting tweets into non-overlapping phrases based on phrase stickiness. Compared with words and segments, tweets are relatively more sparse and time consuming to process, and hence most unsupervised methods in the literature are based on words and segments.

Segment-based methods are efficient, extracting key concepts from tweets for detecting events. On the other hand, event reports by segment clusters can be difficult for human readers to comprehend, since they do not convey structural information. It is thus typically necessary for the reader to refer to search engines to understand the real news story [6]. A key source of information that is missing from segment-based system outputs is *who did what to whom*, a semantic frame structure that conveys the most crucial elements of events. For example, an output event represented by independent segments “florida, sugar bowl, louisville” gives no information that which team won the game. To address this, we extract semantic frames from tweets by leveraging open information extraction [10], using them as features instead of tweets or segments.

Unit	Event
tweet	gameday! come on in later and watch the gators take down louisville! #sugarbowl
	rt @eancaafootball: retweet if you were impressed by louisville’s huge upset over florida in the sugar bowl! http://t.co/vjqrccvu
segment	*florida, sugar bowl, sec, bowl, win, louisville*
frame	**(louisville, gets biggest win in, program history), (-, goes, florida), (uf, was favored by, 14), (-, go, gators), (-, go, cards)**

Table I
AN EVENT EXAMPLE.

Compared to tweets, frames are much more light-weight serving as features, and can offer the same level of efficiency as segment-based event detection systems. Compared with segments, they can give essential information of events as tweets do. Table I shows an example event from our data, where the real event is a Sugar Bowl football game between Louisville Cardinals and Florida Gators, where Florida had the majority’s expectation to win before the match, but lost in the end. Our frame based method gave a succinct summary of the event. In contrast, it can be very difficult to tell such details from events clusters by the word-/segment-based systems.

In addition to readability, a second advantage of using frames as feature representations is that the extraction of frames requires a basic degree of grammaticality on input tweets, which can help filter out mundane and uninformative tweets, which tweets can cause significant precision losses in word and segment based systems. We assume that tweets are sufficiently redundant, so that at least one frame can be extracted for most events.

We verify the potentials of frames-based tweet event

[1]This is a draft version. Full version at https://ieeexplore.ieee.org/document/8300594

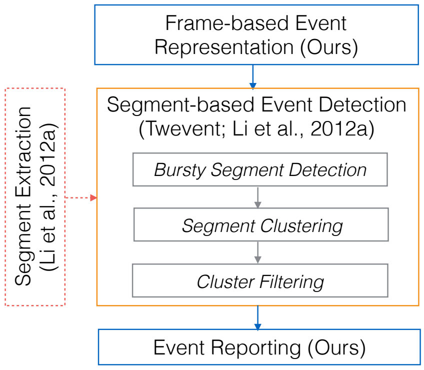

Figure 1. Comparison of our model and Twevent. (Twevent: left red box + middle orange box. Ours: top and bottom blue box + middle box.)

1	Tweet	justin bieber smokes weed?! omg shocking
	Segment	*justin bieber; smokes; weed; shocking*
	Frame	**(justin bieber, smokes, weed)**
2	Tweet	rt @footballtalk_: lewis holtby says moving to tottenham is a chance to 'fulfil his dreams' of premier league football.
	Segment	*rt; @footballtalk; lewis; says; moving; chance; dreams; premier; league; football*
	Frame	**(lewis holtby, says moving to, tottenham); (tottenham, is a chance to fulfil, his dreams);**
3	Tweet	@karennbabess lol no im nice (:
	Segment	*@karennbabess nice*
	Frame	-

Table II
Segments/**FRAMES** EXTRACTED FROM TWEETS.

detection by empirically comparing frame features with segment features of Li et al. [6]. The framework of our method is shown in Figure 1, in comparison with Twevent. The left box indicates segment-based feature extraction in Twevent. The middle box represents event detection method used in Twevent, which takes a stream of segment-based features as inputs and yields a set of segment-based event clusters. In our model, we feed semantic frame elements as features to event detection, and link frame elements in output event clusters back to their original frames for event reporting. Results show that existing open information techniques allow our frame-based method to outperform the segment-based method of Li et al. [6] in both precision and recall.

The contributions of this paper are two-fold. First, we proposed a novel feature extraction method for tweets, which utilizes informative structured frames by open information extraction as features instead of traditional words, segments and tweets. Second, we build a frame-based event detection model and evaluated its effectiveness and high readability of generated events. The code and data will be released at https://github.com/qolina/ialp17-fred.

II. Frame-based Event Representation

Shown in Figure 1, before the main process of segment-based event detection, Twevent [6] partitions each tweet into a set of non-overlapping segments through an optimization process over the stickiness of potential segments by considering 1) length, 2) probability of being anchor texts in Wikipedia[2] and 3) cohesiveness defined with the use of Microsoft Web N-Gram service[3]. The process can be time-consuming due to the use of external resources and inaccurate as they only rely on statistics.

In contrast, we extract segments by finding semantic frame elements. A **frame** is defined as a triple (arg_s, $verb$, arg_o), where $verb$ represents an action and arg_s and arg_o represent the verb's subject and object, respectively. A **frame element** denotes anyone of arg_s, $verb$, arg_o in a frame. Rather than more recent open information extraction systems like Ollie [11] and ClausIE [12], we use ReVerb [10] to extract frames from tweets, for the following reasons. First, ReVerb uses an unsupervised method, which is suitable for Twitter data due to fast-changing Twitter topics and extremely large scale of Twitter data. Second, ReVerb does not require syntactic parsing, which is still a very difficult task for Twitter data. While shallow syntactic analysis like POS and chunking on Twitter have been utilized to extract meaningful frames through syntactic constraints in ReVerb.

[2]http://www.wikipedia.org/
[3]http://web-ngram.research.microsoft.com/info/

Examples of extracted segments and frames from tweets are shown in Table II. In the first tweet, both meaningful segments and frames are successfully extracted from high quality short tweets. As discussed earlier, frame-based representation is superior than segment-based for two reasons. First, structured frames have higher readability. For example, the frame extracted from the second tweet, (lewis holtby, says moving to, tottenham) is easier to understand than independent segments 'lewis', 'says', 'moving' etc. Second, frame extraction can filter noisy uninformative tweets as it requires grammatical quality of tweets. For example, the third tweet in Table II contains no frames as it is a low quality mundane tweet.

III. Event Detection

Frame-element-based event detection is conducted by the segment-based event detection framework [6], which includes bursty segment extraction, segment clustering and cluster filtering. In particular, we first calculate the burstiness of each element to find the bursty elements, and then cluster bursty elements through a k-Nearest Neighbor graph. Here burstiness refers to the relative frequency of segments in a certain time window, as compared with their average frequencies over all time windows. It serves as the basis of most unsupervised news event detection methods. As a final step, event clusters are ranked and filtered by a heuristic score, which uses information from Wikipedia to estimate how likely a segment cluster is for describing a news event. For more details, refer to Twevent [6]. Here we use the system in exactly the same setting.

IV. Event Reporting

We now obtain a set of clusters in time window d. Each cluster c is represented by five most representative elements $E_c = \{e_i | i \in [1, 5]\}$. Since the readability

Unit	Daily Average	Total
raw tweet	3, 604 K	54, 065 K
preprocessed tweet	2, 073 K	31, 097 K
user	1, 359 K	16, 000 K
word	79 K	382 K
segment	288 K	1, 604 K
frame	1, 797 K	14, 948 K
frame element	1, 439 K	14, 957 K

Table III
DATA STATISTICS

of frame elements can be as low as segments, an event reporting step is used, which aims to map each element back to its original frame. In particular, for each element $e \in E_c$, we extract all frames which contain e, denoted as Γ_e. The original frame of e, γ_e^*, is estimated with the frame with largest frequency in current time window d.

$$\gamma_e^* = \arg\max_{\gamma_e}\{f_{\gamma_e,d}|\gamma_e \in \Gamma_e\} \tag{1}$$

Then an event cluster c can be represented as a set of frames $\{\gamma_{e_1}^*, \gamma_{e_2}^*, ..., \gamma_{e_5}^*\}$, where each frame is a meaningful semantic triple. In contrast, the segment-based clusters yields a set of independent n-grams, which do not explicitly state the event structure, and cannot be intuitively linked to a set of salient event-containing tweets. As a result, Li et al. [6] do not take an event reporting step.

V. EXPERIMENTS

A. Data

Our Twitter data are crawled using Twitter public streaming API, and consist of tweets published from Jan. 1st to Jan. 15th, 2013. Summary of the data set is shown in Table III. Comparison between the average number of word, segment and frame per day demonstrates sparsity of frames, and hence the necessities of using frame elements instead of frames as clustering features. A preprocessing step is applied, which includes 1) non-English-tweet removal [13], 2) tweet normalization [14], 3) Twitter specific part-of-speech analysis [15] and 4) noun phrase identification[4] [16].

B. Evaluation

Precision and the number of detected events are used as evaluation metrics. Recall is replaced by the total number of detected events, because it is difficult to identify all events that happened over a period. An output event, represented by a given date and a group of features (e.g. segments for Twevent, frames for our method), is evaluated manually by two annotators (with Cohen's kappa 0.65) labelling whether it is a news event or not. Search engines are allowed to assist the manual decision process. Events that happened both at and before the given date are annotated as true news events, as some events can stay hot in tweets for several days. Given a segment-based event output "florida, sugar bowl, louisville", annotator should confirm that whether there is a sugar bowl game between florida and louisville on given date. If there are more than two team involved in multiple matches in the output, it will be labelled as a false news.

[4]http://www.dcs.shef.ac.uk/$\sim$mark/phd/software/

System	#Events	Precision	Readability
Twevent	107	75.70%	0.30
FrED	114	**85.09**%	0.59

Table IV
COMPARISON OF SEGMENT-/FRAME-BASED EVENT DETECTION MODELS.

Segment	Element	Segment	Element
harry talor	harry and taylor	*sorry*	-
rose bowl	the rose bowl	*pretty good*	-
retire	is retiring at	*555555*	-
please	please tell/check	*xxx*	-

Table V
EXAMPLE N-GRAMS THAT SERVED AS BURSTY SEGMENTS AND FRAME ELEMENTS.

In addition, we utilize an easy strategy to evaluate system performance on event readability. A readability score (0 means "difficult to understand by given output segments/frames" and needs assistance from search engine, 1 means "easy to understand") is assigned to each event. Averaged readability of events is used to estimate readability of different event detection methods.

C. Results and Analysis

The results of out method (FrED) and Twevent [6] are shown in Table IV. Comparing FrED with Twevent, improvements on precision (75.70→85.09) and readability (0.30 → 0.59) verify the effectiveness of our frame-based event detection method versus segment-based method. One of the main reasons for the improvement is that frame detection conducts feature selection by filtering out irrelevant non-frame words. In contrast, the segment-based method relies on bursty segment detection to filter out infrequent phrases, without a refined feature selection step.

Table V shows some examples of bursty segments and frame elements, which provide evidence for better feature selection of FrED. From the table, we can find that meaningful n-grams (left column) can be detected as bursty segments and frame elements by Twevent and FrED respectively. Meaningless n-grams detected as bursty segments by Twevent can be filtered out by FrED with the syntactic constraints. In particular, meaningless frequent segments which are discarded through frame extraction include 1) adjectives such as 'pretty good', 2) interjections such as 'please' and 'sorry', 3) emoticons such as '555555' (crying); 4) noisy words like 'xx'.

To further investigate the reason why FrED outperforms Twevent, we compare the distribution of human-labelled false news events of Twevent and FrED, shown in Figure 2. False news events are classified into three categories: 1) mundane topics such as "follow spree" topics launched by celebrities, in which a celebrity randomly selects a few participating fans to follow. 2) meaningless clusters,

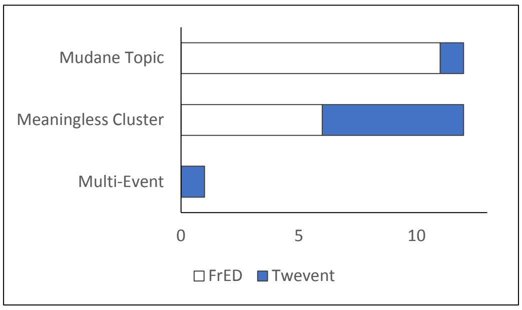

Figure 2. Error analysis of different event detection methods.

System	Event
Twevent	ray lewis; ray; retire; retiring; inspired
FrED$_{ele}$	ray lewis; retiring; is retiring at; is retiring
FrED	(**ray lewis**, told, the team),
	(ray lewis, **retiring**, -),
	(he, **is retiring at**, the end),
	(ray lewis, **is retiring**, -),
Twevent	- (not detected)
FrED$_{ele}$	weed; smoked; bieber; cutting; smokes
	(justin bieber, smokes, **weed**);
	(justin bieber, **smoked**, weed);
FrED	(-, cut for, **bieber**);
	(-, **cutting**, yourself);
	(justin bieber, **smokes**, weed)

Table VI
EXAMPLE OUTPUT EVENTS.

which can catch heterogenous topics, such as a horoscope topic "scorpio; energetic mars; activates more for" and 3) mixed events, which consist of multiple events. For example, an event "derby; stoke; 3-1; 3-0" consists of two sports match of "Manchester City 3-0 Stoke City" and "Derby County 3-1 Middlesbrough". We can see that FrED performs similar with Twevent in distinguishing meaningless clusters. Besides, FrED generates less mix events than Twevent. Most importantly, FrED greatly outperform Twevent in filtering out mundane topics with the help of feature selection by frame extraction.

D. Example Outputs

Table VI shows some event examples, in which FrED gives more readable output. FrED$_{ele}$ is FrED without the event reporting step (Section IV), which presents frame elements in the resulting clusters. Note that there may not be corresponding arg_s or arg_o for a $verb$ in all frames. Events detected by Twevent and the frame element clusters of FrED$_{ele}$ are mostly described by noun phrases without verbs, which cannot show important structural information. In contrast, FrED describe events with frames, which contain verb phrases and are more readable.

VI. CONCLUSION

We proposed a semantic-frame-based representation for feature-pivot Twitter event detection framework. Frames, defined as triplets ($subject$, $verb$, $object$), are structured information units and hence convey more event information than traditional feature like words/segments. Improved precision and readability over a segment-based baseline method show the effectiveness of frame-based event detection method.

REFERENCES

[1] T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes twitter users: real-time event detection by social sensors," in *Proceedings of WWW*, Raleigh, North Carolina, USA, 2010, pp. 851–860.

[2] E. Benson, A. Haghighi, and R. Barzilay, "Event discovery in social media feeds," in *Proceedings of ACL-HLT*, Portland, Oregon, 2011, pp. 389–398.

[3] A. Ritter, Mausam, O. Etzioni, and S. Clark, "Open domain event extraction from twitter," in *Proceedings of the 18th ACM SIGKDD*, 2012, pp. 1104–1112.

[4] A. J. McMinn and J. M. Jose, "Real-time entity-based event detection for twitter," in *Proceedings of CLEF*, Toulouse, France, September 2015, pp. 65–77.

[5] A. Cui, M. Zhang, Y. Liu, S. Ma, and K. Zhang, "Discover breaking events with popular hashtags in twitter," in *Proceedings of the 21st CIKM*, Maui, Hawaii, USA, 2012, pp. 1794–1798.

[6] C. Li, A. Sun, and A. Datta, "Twevent: segment-based event detection from tweets," in *Proceedings of CIKM*, Maui, Hawaii, USA, 2012, pp. 155–164.

[7] Y. Qin, Y. Zhang, M. Zhang, and D. Zheng, "Feature-rich segment-based news event detection on twitter," in *Proceedings of the Sixth IJCNLP*, Nagoya, Japan, October 2013, pp. 302–310.

[8] H. Becker, M. Naaman, and L. Gravano, "Beyond trending topics: Real-world event identification on twitter," in *Proceedings of ICWSM*, 2011.

[9] G. Ifrim, B. Shi, and I. Brigadir, "Event detection in twitter using aggressive filtering and hierarchical tweet clustering," in *Proceedings of he SNOW 2014 Data Challenge*, Seoul, Korea, April 2014, pp. 33–40.

[10] A. Fader, S. Soderland, and O. Etzioni, "Identifying relations for open information extraction," in *Proceedings of EMNLP*, 2011, pp. 1535–1545.

[11] Mausam, M. Schmitz, R. Bart, S. Soderland, and O. Etzioni, "Open language learning for information extraction," in *Proceedings of the EMNLP-CoNLL '12*, 2012, pp. 523–534.

[12] L. Del Corro and R. Gemulla, "Clausie: Clause-based open information extraction," in *Proceedings of the 22nd WWW*, 2013, pp. 355–366.

[13] M. Lui and T. Baldwin, "langid.py: An off-the-shelf language identification tool," in *Proceedings of the ACL 2012 System Demonstrations*, 2012, pp. 25–30.

[14] B. Han, P. Cook, and T. Baldwin, "Automatically constructing a normalisation dictionary for microblogs," in *Proceedings of the EMNLP-CoNLL*, 2012, pp. 421–432.

[15] L. Derczynski, A. Ritter, S. Clark, and K. Bontcheva, "Twitter part-of-speech tagging for all: Overcoming sparse and noisy data," in *Proceedings of RANLP*, 2013.

[16] L. Ramshaw and M. Marcus, "Text Chunking Using Transformation-Based Learning," in *Proceedings of the Third Workshop on Very Large Corpora*, Somerset, New Jersey, 1995, pp. 82–94.