International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.9, No.8 (2016), pp.73-86
http://dx.doi.org/10.14257/ijsip.2016.9.8.06

Context-Based Value Tracking

Yaoyi Xi[1], Bicheng Li[2*],Yuan Gao[3] and Yongwang Tang[3]

[1]*Luoyang University of Foreign Language*
[2]*College of Computer Science and Technology, Huaqiao University*
[3]*Zhengzhou Information Science and Technology Institute*
WIM_GY@163.com

Abstract

Value tracking aims to capture the changes of attribute values along with the evolution of topic. Existing researches on value tracking only extracted the attribute values chronologically, and took no use of the context information to verify the correctness of the values. This paper proposes a context-based value tracking method. First, extract the candidate attribute values according to the patterns generated by the regular expressions; Second, recognize the temporal expressions in the source sentences of the candidate values based on conditional random fields; Finally, identify the real attribute values according to the temporal features and location features in the context. Experiments on TREC 2013 Knowledge Base Acceleration (KBA) stream corpus and human-build Chinese corpus demonstrate that the proposed method can track the changes of the attribute values effectively.

***Keywords**: Event extraction; Value tracking; Context; Temporal expression; Information extraction; Topic evolution*

1. Introduction

Value Tracking aims to capture the changes of attribute values along with the evolution of topic [1]. Different topics have different attribute types. For example, emergency topics always cause casualties, so they have 'death toll' and 'injury number' attributes. With the development of one topic, the values of some attributes will be always changing. Take the epidemic-related topic as an example, the number of infections is small at the beginning of the epidemic and keeps growing over time. Then the number keeps stable after the government adopted proper measures of prevention and control. In the final stage, the number starts to tail off with the successful manufacturing and using of the vaccine. Tracking the changes of important attribute values in time is helpful for relevant authorities to determine the level of catastrophe and make scientific decisions.

Although event extraction technologies can extract attribute value [3], but it only extracts corresponding value from one single document or sentence. Compared with event extraction, value tracking needs to constantly identify real values in the topic-related documents along with the development of the topic. Existing researches on value tracking [6] only used event extraction technology to extract the attribute values and did not make use of the context information to verify the values' correctness, which resulted higher error rates. In order to solve this problem, this paper proposes a context-based value tracking method, which utilizes the temporal and location features in the context to improve the accuracy of value tracking.

The rest of the paper is organized as follows: In Section 2, we present related work. Section 3, gives a detailed analysis of the importance of temporal and location features in the source context of the attribute value. Section 4, presents a formal description of our

---

[*]Corresponding Author

ISSN: 2005-4254 IJSIP

value tracking approach, while Section 5 describes the experiments and reports the results. Finally, Section 6 summarizes our conclusions.

2. Related Work

Value tracking was first formally proposed by NIST (National Institute of Standards and Technology) in 2012 [1], and was selected as an evaluation track in TREC (Text Retrieval Conference) 2013. In TREC 2013, the organizers gave the definition of value tracking, related data set and evaluation metric. Liu *et. al.*, [6] used the name of an attribute as trigger and selected the sentences that containing the trigger as the source sentences. After that, they estimated the attribute value according to the numerical value in the surrounding words of the trigger. The length of the window is 5. Zhang *et. al.*, [7] translated value tracking into sequence labeling problem and used the given training data to train the CRF (conditional random fields) model [8]. Then they used the CRF model to extract attribute values from the sentences that contain key words of the attribute.

Besides TREC, Wu *et. al.*, [9] conducted similar research when they studied event extraction, but they also only chronologically extracted the attribute values without verifying the correctness of the values.

3. Analysis of the Importance of Context

The temporal features and location features in the attribute's context play an important role in value tracking. Their importance is analyzed as follows.

(1) Temporal Feature

Existing researches neglected the temporal features, which brought two problems: ① The value extracted from one document may be not the real value of the attribute at the publication time of the document; ②There may be a conflict between the values at the same time or at different time.

Figure 1, gives a relevant example. It is a news report published at December 29, 2013, which is about 'Explosions in the Russian city of Volgograd, December, 2013'. In Figure 1, the sentence in bold contains the number of death, but it is not the true value of death toll at December 29, 2013. The first sentence of this report contains the true value, but it is contradict with the value in the bolded sentence. Similar situation also exists in different documents coming at the same time or at different time.

Explosion at Russian train station kills three

2013-12-29 17:54

MOSCOW - An explosion at a train station in the southern Russian city of Volgograd killed three people and wounded six others on Sunday, the Interfax news agency cited a law enforcement source as saying.

The cause of the blast was not immediately clear. An attack by a female suicide bomber in the same Russian city on October 21 killed seven people, in the deadliest such blast outside the volatile North Caucasus region in nearly three years.

Figure 1. Example of Temporal Feature Importance

(2) Location Feature

The importance of location features reflects in two aspects:

① The position of the attribute's source sentence in the document influences the judgment of its authenticity. Generally speaking, the main content of one document and the real attribute value always appear in the first few sentences, so their degree of authenticity is higher than the sentences on the later place.

② The geographical location information contained in the source sentence of the attribute also has influence.

Figure 2, gives a relevant example, which is about 'Japan Earthquake, March 11, 2011'. In Figure 2, the first sentence contains the true number of death. The last sentence also contains the number of death, but it reports a different topic which happened at Ecuador and Colombia, not Japan. From Figure 2, we can see that considering the relative position of the sentence in the document and the topic's occurring geographical location can help to verify the correctness of the attribute value.

Daybreak reveals huge devastation in tsunami-hit Japan

2011-03-12 07:01

TOKYO - Japan confronted devastation along its northeastern coast on Saturday, with fires raging and parts of some cities under water after a massive earthquake and tsunami that likely killed at least 1,000 people.

……

On January 31, 1906, an 8.8-magnitude quake struck **the coasts of Ecuador and Colombia**, killing nearly **1,000 people**.

Figure 2. Example of Location Feature Importance

4. Methodology

Based on the problems resulted from traditional methods, this paper proposes the context-based value tracking method, as shown in Figure 3. First, extract the candidate values according to the patterns generated by the regular expressions; Second, recognize the temporal expressions in the source sentences of the candidate values based on CRF by incorporating basic features, dictionary features and dependency features; Finally, identify the true attribute values according to the temporal features and location features in the context.

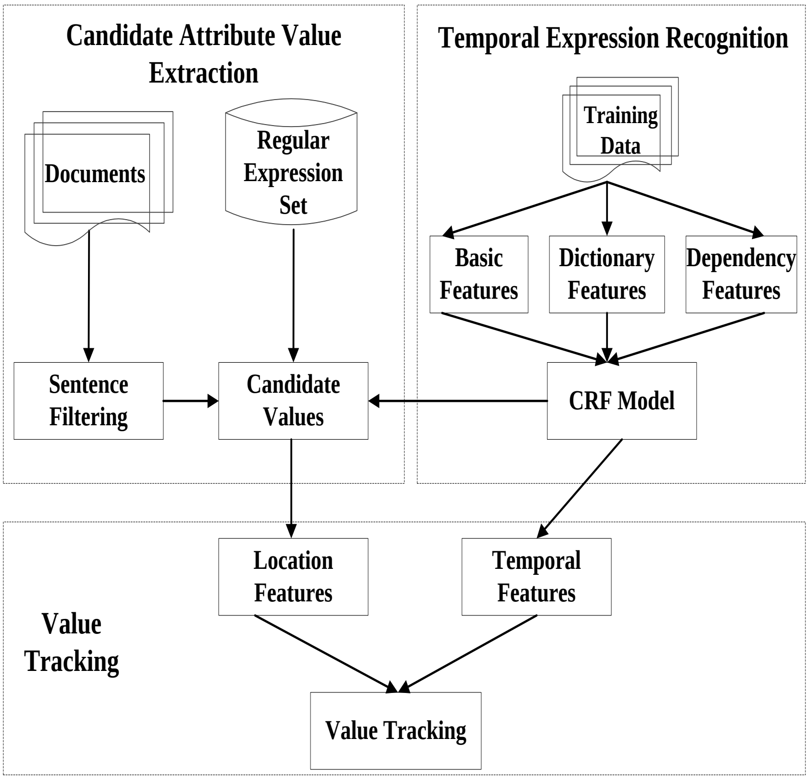


Figure 3. Value Tracking

4.1. Candidate Attribute Value Extraction

In general, the number of the attribute type of interest is small. For example, the main concern of emergency topic is its number of deaths and injuries. The attribute value always emerges in fixed pattern, such as ‘XX people died’, and ‘XX’ is the number of death (usually Arabic numerals). Considering this characteristic, we extract the attribute value through hand-crafted patterns.

(1) Candidate Sentences Selection

The attribute type of interest generally appears in important sentences. We filter out non-important sentences according to the following two conditions: ①The number of words in the sentence is less than 10. ②There are no representing topic terms in the sentence. Representing topic terms can distinguish one topic from another and generally appear in each topic-related document.

(2) Pattern Generation

Regular expression is proposed by Stephen Kleene in 1956. It provides a mechanism that search specific strings from a character set [10] and is widely adopted by researchers to generate patterns [11]. Regular expression is composed of ordinary character and special character. Take the ‘number of deaths’ and ‘number of injuries’ as example, some regular expressions are listed in Figure 4.

```
".*(kill|killed|killing|kills) *(\\d+,\\d+|\\d+).*"
".*(death toll|death tolls|deaths|death).* (\\d+,\\d+|\\d+).*"
".*(injure|injured|injuring|injures|wound|wounded|wounding|woundes)
*(\\d+,\\d+|\\d+).*"
```

Figure 4. Example of Regular Expression

(3) Candidate Attribute Value Extraction

Assume $V = \{V_1, V_2, ..., V_T\}$ is the set of candidate attribute values, where V_t is the subset of V, t is the happening time of the candidate attribute value. At the beginning of the topic, set V to be null. With the development of the topic, extract candidate attribute values according to the hand-crafted regular patterns and add them to V_t at each time t. The element in V_t is a seven- tuple $(v, s, d, t_1, t_2, l_1, l_2)$, where v is the candidate attribute value, s is the source sentence, d is the source document, t_1 is the happening time, t_2 is the publication time of the source document, l_1 is the geographical location of the candidate attribute value， l_2 is the relative position of the source sentence in the source document and it is a number in [0, 1]. Extraction of t_1 and t_2 will be introduced in detail in Section 4.2.

4.2. Temporal Expression Recognition

Temporal expression recognition is helpful to verify the correctness of the candidate attribute values. Compared with the attribute of interest, temporal expression is more flexible and is inconvenient to be recognized by hand-crafted pattern or rules, so we adopt the statistical machine learning based method to recognize them. First, extract the basic features, dictionary features and dependency features. Second, train the CRF model by these features. Finally, recognize the temporal expression in the source sentence of the candidate attribute values by the trained model. The training process of CRF model can be found in [12]. This section focuses on feature extraction and temporal expression normalization.

(1) Feature Extraction

The features used by existing researches on temporal expression recognition can be divided into two categories, namely basic features and dictionary features. We refine the dictionary features and add a third category of feature, *i.e.*, dependency features. Dependency parsing is completed by LTP[1].

①Basic Features

Basic features include word feature (word form and part of speech), n-grams, numerals or not, and whether contained temporal expression separator.

②Dictionary Features

Existing researches on temporal expression recognition simply classified all the time-related words into a single dictionary. However, different time-related words have different function in temporal expression. We divide the time-related words into two dictionaries, namely Time-Unit Dictionary and Time Dictionary. Time-Unit Dictionary contains unit of time, such as year and day *et. al.*, and it always needs to be combined with Arabic numerals to express time, such as 3 years and so on. Time Dictionary contains the remaining time-related words, and the words in it can represent time separately, such as yesterday and afternoon *et. al.*,

③Dependency Features

For convenience, we denote the word in Time Dictionary as time word, and the word in Time-Unit Dictionary as unit word. We use the following 4 dependency features.

For each word, if there are time words on its left, we choose the time word which is nearest to current word and select the dependency path between the two words as the first

1 http://www.ltp-cloud.com/

dependency feature. In a similar way, we select the dependency path on the right as the second dependency feature.

For each word, if there are unit words on its left, we choose the unit word which is nearest to current word and select the dependency path between the two words as the third dependency feature. In a similar way, we select the dependency path on the right as the fourth dependency feature.

(2) Temporal Expression Normalization

After recognition, we can get three kinds of temporal expressions, namely explicitly expressions, implicitly expressions and relatively expressions [13]. Explicit temporal expressions are fully specified. For example, March 11, 1982 is the explicit expression. Implicit expressions can be assigned once the implicit temporal semantics of such expressions are known. For example, the international Labor Day is assigned May 1. The normalization of relative expressions for which a reference time is needed is the most challenging task. For such expressions, we make use of hand-crafted rules and assign the values in an underspecified format depending on the assumed reference time. Figure 5 gives some examples of the normalization of relative expressions.

Today: Publication time of the source document
Now: Publication time of the source document
Yesterday: One day before the publication time of the source document

Figure 5. Example of Temporal Expression Normalization

If there is no temporal expression in the sentence, use publication time of the source document as the happening time of the attribute value.

4.3. Value Tracking

Generally speaking, correct values usually appear in many documents. For example, if a candidate attribute value is reported by multiple news reports, it is more likely to be the real value. According to this characteristic and the importance of the context, we adopt context-based method to track value, with the steps as follows.

Suppose at time t, the candidate attribute value set V_t has been obtained after the extraction of attribute value. For each element in V_t:

Step1: Filter out the candidates which occurred earlier than the topic.

Step2: For non-location attribute, filter out the candidates which occurred in a place different from the geographical location of the topic.

Step3: For each candidate attribute value, sort the documents that contain the value according to the documents' publication time. Then, calculate the document frequency of the value and select the newest document as the value's source document.

Step4: Select the candidate value which has the highest document frequency as the tracked attribute value. If two candidates have the same document frequency, go to Step5. If not, go to Step7.

Step5: Compare the publication time of the two candidates' source documents. We think that the value coming from the later document is more exact. If the two candidates' source documents have the same publication time, go to Step6.

Step6: For each candidate attribute value, it may have multiple source sentences. We calculate the sum of these sentences' relative position, namely $\sum l_2$. A smaller sum indicates the candidate attribute value often appears in a more advanced position in the document and its probability of being correct is larger.

Step7: If there is no value at time t, we think the value at time t is the same as the one in time $t-1$.

Step8: Continue tracking the attribute value in time $t+1$.

5. Experiments

The paper uses the value tracking evaluation data set in TREC 2013 and human-annotated Chinese corpus to verify the validity of the proposed method.

5.1. Datasets

(1) TREC 2013 data set

The data set includes 1040520595 documents, which came from news, blogs, Twitter and so on between Nov. 2011 and Feb. 2013. The detailed information is listed in Table 1. TREC 2013 focused on four attribute types[2], which are Deaths (the number of deaths), Injuries (the number of injuries), Financial Impact (the financial impact of the topic in US dollars) and Locations (the location where the topic happened).

Table 1. Detailed Information of TREC 2013 Data Set

Topic	Topic Type	Start and End Date
2012 Buenos Aires Rail Disaster	accident	2012-02-22-11, 2012-03-03-11
2012 Pakistan garment factory fires	accident	2012-09-11-13, 2012-09-21-13
2012 Aurora shooting	shooting	2012-07-20-06, 2012-07-30-06
Wisconsin Sikh temple shooting	shooting	2012-08-05-15, 2012-08-15-15
Hurricane Isaac (2012)	storm	2012-08-28-16, 2012-09-07-16
Hurricane Sandy	storm	2012-10-24-15, 2012-11-03-15
June 2012 North American derecho	storm	2012-06-29-15, 2012-07-09-15
Typhoon Bopha	storm	2012-11-30-14, 2012-12-10-14
2012 Guatemala earthquake	earthquake	2012-11-07-16, 2012-11-17-16
2012 Tel Aviv bus bombing	bombing	2012-11-21-11, 2012-12-01-11

(2) Chinese corpus

There is no existing standard Chinese test set for value tracking methods. We randomly choose 8 bursty news topics from 4 news websites. Detailed statistics are listed in Table 2. We choose these sites because all of them provide special topic of news edited by professional editors.

According to the characteristic of the corpus, we mainly focus on two types of attribute, namely ‘the number of deaths’ and ‘the number of injuries’. After crawling all linked news stories for each topic, we hired two human annotators to label the number of deaths

2 The initial track also includes the attribute ‘Displaced’ (The number of people who have had to leave their homes as a result of the topic), but it did not give the evaluation result.

and injuries for each topic independently. The annotators were asked to read the news stories of each topic several times to form a general picture on its development. In the next step, each annotator was asked to identify the number of deaths and injuries for each topic independently. The two annotators then met together, reviewed the value collection annotated individually for each topic, and revised them to arrive at a “consensus” attribute value collection for the topic.

Table 2. Detailed Information of Chinese Corpus

	Sina News	Tencent News	Phoenix News	Netease News	Sum
1、Explosions in the Russian city of Volgograd, December, 2013	31	32	74	0	137
2、Crash of Asiana Airlines Flight 214 in San Francisco, July, 2013	120	108	322	78	628
3、Kunming terror attack, March, 2014	0	270	0	0	270
4、Boston Marathon bombing, April, 2013	229	209	224	65	727
5、Crash of Libyan Afriqiyah Airways plane in Tripoli, May, 2010	73	75	0	0	148
6、Oslo explosions and shootings in Norway, July, 2011	256	210	0	345	811
7、Mumbai bomb attacks in India, July, 2011	50	48	0	36	134
8、UN building bombing in Nigeria's capital, August, 2011	46	28	0	20	94

(3) Preprocessing

For TREC 2013 data set, we remove common stop-words and use CoreNLP[3] to label part of speech. Retain the verbs, nouns, and adjectives. After that, convert all the words to lowercase for easy match of regular expressions.

For Chinese corpus, we remove common stop-words and tokens which are neither verbs, nor nouns, nor adjectives from the news articles with the help of NLPIR[4] and identify the geographical location according to the ‘ns’ label.

5.2. Evaluation Metrics

TREC 2013 evaluated a value tracking system according to the expected error with respect to the true attribute value f_{u^*}.

$$EE(u,u^*,\tau) = \frac{1}{t_e - t_s}\int_{t_s}^{t_e} Err(u,u^*,\tau)d\tau \tag{1}$$

$$Err(u,u^*,\tau) = \left| f_{u^*}(\tau) - f_u(\tau) \right| \tag{2}$$

Where t_s is the start time of the topic, t_e is the end time of the topic, $f_{u^*}(\tau)$ denotes the true value of the attribute at time τ, $f_u(\tau)$ denotes the tracked value of the attribute at

3 http://nlp.stanford.edu/research.shtml
4 http://ictclas.nlpir.org/

time τ . The smaller the expected error, the better the performance of the value tracking system is.

5.3. Results

(1) Results on TREC 2013 Data Set

We participated in the value tracking evaluation in TREC 2013, but only used the proposed attribute value extraction method. The results are listed in Table 3. Our method is denoted as 'wim_GY_2013_VT1' [14].

Similar with the other participating systems, the 'wim_GY_2013_VT1' method also did not use context features, but the performance of the 'wim_GY_2013_VT1' method is higher than other systems. From Table 3, we can see that the average expected error of the 'wim_GY_2013_VT1' method is lowest. PRIS-PRISTS1, PRIS-PRISTS2 and PRIS-PRISTS3 are proposed by Zhang *et. al.,* [7]. They used CRF model to extract attribute values, but their performance is not high. This is proof that when the extraction objective satisfies some fixed patterns, the pattern based method can achieve better performance than machine learning methods. ICTNET-ValueTask is proposed by Liu *et. al.,* [6] and it belongs to pattern based methods. However, Liu *et. al.,* directly used the names of the attribute to choose sentences and extracted attribute values from these sentences by finding the Arabic numerals around the attribute names. Their method is fairly rough and the performance is low.

Table 3. Performance Comparison on TREC 2013 Data Set

Systems	Location	Deaths	Injuries	Finnancial impact
PRIS_PRISTS1	18101.8	37880.4	64886.5	9.5251
PRIS_PRISTS2	11864.1	88666.5	106099	25.1175
PRIS_PRISTS3	11796.3	88666.5	106375	42.5367
ICTNET_ValueTask	20038.0	188.495	390.985	13.3539
wim GY 2013_VT1	14483.6	2726.06	410.092	13.3539

(2) Results on Chinese Corpus

Because the method proposed by Zhang *et. al.,* needs training data, we use the former 4 topics as test set and the latter 4 topics as training set.

We compare our proposed context-based method (hereafter Con-VT for short) with the methods proposed by Liu *et. al.,* (hereafter Rule-VT for short) and Zhang *et. al.,* (hereafter CRF-VT for short) respectively. Con-VT takes use of the temporal and location features. To verify the effectiveness of the two features, we adopted the following three ways to revise Con-VT. 1) Complete value tracking just by extracting attribute values (hereafter Con-VT1 for short); 2) On the basis of Con-VT1, complete value tracking by considering temporal features (hereafter Con-VT2 for short); 3) On the basis of Con-VT1, complete value tracking by considering location features (hereafter Con-VT3 for short).

Figure 6, shows the comparison of the overall performance of different methods. The result details on each topic are listed in Table 4. From Figure 6, and Table 4, we can see that Con-VT obviously outperforms the other methods. The major factor that affects Rule-VT and CRF-VT is that they do not recognize temporal expressions, which caused them to be easily affected by noise data. For example, news reports usually retrospect similar events in history, which is noise to value tracking. CRF-VT's expected error is highest and Rule-VT is better than CRF-VT, this is because Rule-VT considers the representative topic terms which reduces the likelihood of being affected by noise data.

In the absence of temporal and location features, Con-VT1 is comparable to Rule-VT. Compared with Con-VT1, the performance of Con-VT2 and Con-VT3 are improved, which verifies the importance of temporal and location features. Con-VT3 is worse than Con-VT2, which demonstrates that compared with location features, temporal features are more important to value tracking.

All of the methods have a better performance on topic 1 and 3 than them on topic 2 and 4. This is because compared with topic 2 and 4, topic 1 and 3 last for a short time and have less sentences that contain attribute values, which is less likely to be affected by noise data.

Although Con-VT performs best, it still has some drawbacks. First, it depends on hand-crafted patterns. The coverage of the patterns will influence the attribute value extraction directly and further influence the performance of value tracking. Second, the performance of temporal expression recognition should be further improved.

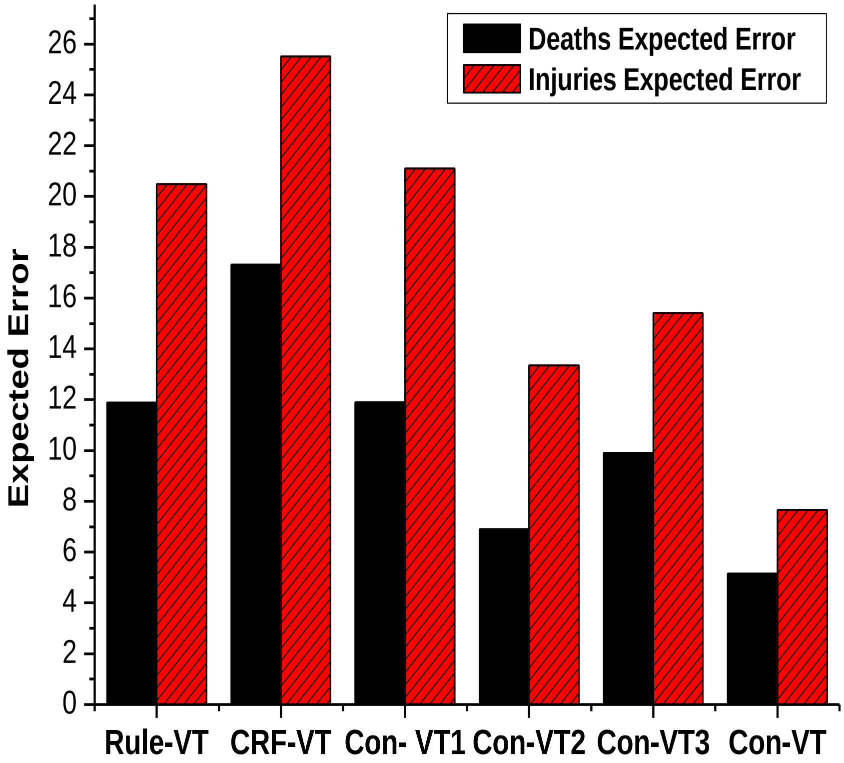


Figure 6. Overall Performance Comparison

Table 4. Performance Comparison on Each Topic

	1、Explosions in the Russian city of Volgograd		2、Crash of Asiana Airlines Flight 214 in San Francisco	
Method	Deaths Expected Error	Injuries Expected Error	Deaths Expected Error	Injuries Expected Error
Rule-VT	4.29	7.57	20.91	26.26
CRF-VT	5.43	8.57	30.23	42.53
Con-VT1	3.43	6.14	22.02	23.85
Con-VT2	1.85	4.29	13.04	16.13
Con-VT3	2.43	5.43	20.15	15.02
Con-VT	**1.57**	**2.71**	**10.36**	**7.57**

	3、Kunming Railway Station terror attack		4、Boston Marathon bombing	
Method	Deaths Expected Error	Injuries Expected Error	Deaths Expected Error	Injuries Expected Error
Rule-VT	4.14	7.5	18.18	40..63
CRF-VT	4.67	8.33	28.9	42.63
Con-VT1	4.17	8.57	17.98	45.88
Con-VT2	1.08	2.83	11.65	30.15
Con-VT3	2.42	4.17	14.63	37.05
Con-VT	**0.17**	**0.75**	**8.55**	**19.63**

5.4. Display the Changes of Attribute Value

Figure 7, displays the tracking result for the attribute ‘Number of Deaths’ of the topic ‘Kunming Railway Station terror attack’. The topic lasted for 12 days. In Figure 7, the X-axis is date and the Y-axis is the number of deaths up to that day. ‘Real Value’ denotes the true number of deaths and ‘Tracked value’ denotes the tracked number of deaths obtained by Con-VT. Both of the true values and tracked values are connected by solid lines. From Figure 7, we can see that except March 1, the tracked values are all the same with the real values, which makes the solid line of ‘Real Value’ covered by the solid line of ‘Tracked Value’. On March 1, the tracked value is wrong. The main reason for this is that the happening time of the topic is about 9:00 p. m. of March 1. As it is in the evening, the related reports are few and relatively short, which weakens the role of the context features. Besides, we can see from Figure 7, that some attributes’ values do not change frequently and after some time they become persistent throughout the remaining course of the topic.

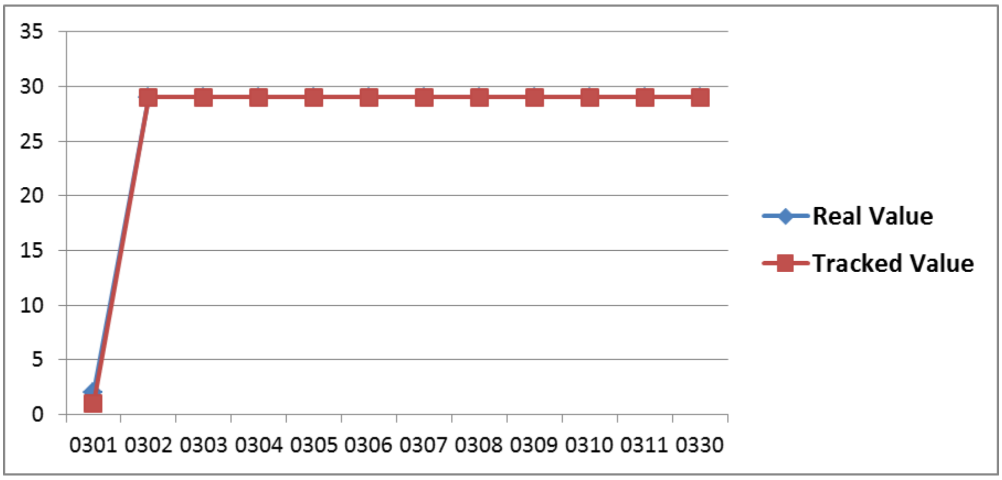


Figure 7. Display of the Tracked Result for the Number of Deaths of Topic 3

6. Conclusion

Value tracking is a new research direction. Existing related researches neglected the context information and only extracted the attribute values chronologically, which resulting a high error rate. This paper not only extracts the attribute values chronologically, but also uses the temporal and location features in the context to refine the extraction result, which improves the performance of value tracking.

In this paper, the extraction of attribute values depends on hand-crafted patterns. The future research may use semi-supervised method to learn the patterns automatically. In addition, the recognition and normalization of temporal expression should be further studied.

Acknowledgments

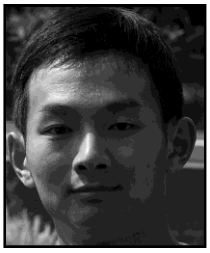
This research is supported by the National Social Science Foundation of China No.14BXW028 and Scientific Research Funds of Huaqiao University.

References

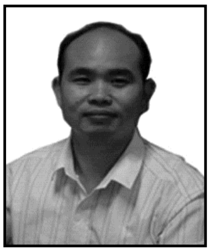
[1] TREC, "TREC Temporal Summarization". http://www.trec-ts.org/, **(2013)**.
[2] J. Aslam, M. E. Abueg, V. Pavlu and T. Sakai, "Trec 2013 temporal summarization", Proceedings of the 22nd Text Retrieval Conference, Gaithersburg: NIST, **(2013)**.
[3] H. Tan, T. Zhao and H. Wang, "Using Multi-level Patterns to Identify Chinese Event Argument Roles", Journal of Computational Information Systems, vol. 3, **(2008)**, pp. 827-834.
[4] W. Liu, F. Liu, D. Wang, N. Ding and X. Wang, "Extraction of Event Elements Based on Event Ontology Reasoning", Intelligent Information and Database Systems. Springer International Publishing, **(2015)**, pp. 577-586.
[5] P. Li, Q. Zhu and G. Zhou, "Joint modeling of argument identification and role determination in Chinese event extraction with discourse-level information", Proceedings of the Twenty-Third international joint conference on Artificial Intelligence, AAAI Press, **(2013)**, pp. 2120-2126.
[6] Q. Liu, Y. Liu, D. Wu and X. Cheng, "ICTNET at Temporal Summarization Track TREC 2013", Proceedings of the 22rd Text Retrieval Conference. Gaithersburg: NIST, **(2013)**.
[7] C. Zhang, W. Xu, F. Meng, H. Li, T. Wu and L. Xu, "The Information Extraction systems of PRIS at Temporal Summarization Track", Proceedings of the 22rd Text Retrieval Conference. Gaithersburg: NIST, **(2013)**.
[8] J. Lafferty, A. McCallum and F. C. N. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data", **(2001)**.
[9] P. Wu, Q. Chen and L. Ma, "Research on Extraction and Integration of Developing Event Based on Analysis of Space-time Information", Journal of Chinese Information Processing, vol. 20, no. 1, **(2006)**, pp. 21-28.

[10] Y. Gong, Q. Liu, X. Shao, C. Pan and H. Jiao, "A novel regular expression matching algorithm based on multi-dimensional finite automata", Proceedings of 15th International Conference on High Performance Switching and Routing. Vancouver: IEEE, **(2014)**, pp. 90-97.
[11] J. Hoffart, F. M. Suchanek, K. Berberich and G. Weikum, "YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia", Proceedings of the Twenty-Third international joint conference on Artificial Intelligence, AAAI Press, **(2013)**, pp. 3161-3165.
[12] J. Lafferty, A. McCallum and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data", Proceedings of 18th International Conference on Machine Learning. Morgan Kaufmann, **(2001)**, pp. 282-289.
[13] F. Schilder and C. Habel. "From temporal expressions to temporal information: Semantic tagging of news messages", Proceedings of the workshop on Temporal and spatial information processing-Volume 13. Association for Computational Linguistics, **(2001)**.
[14] Y. Xi, B. Li, J. Zhou and Y. Tang, "ZZISTI at TREC 2013 Temporal Summarization Track", Proceedings of the 22rd Text Retrieval Conference. Gaithersburg: NIST, **(2013)**.

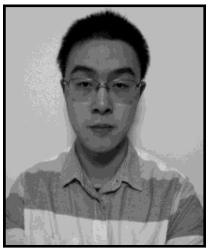
Authors



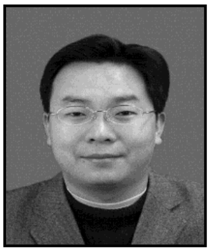
Yaoyi Xi, Ph.D. He has been working on theoretical research and experimental application of text processing technologies. The main research in his Ph.D. program is around topic evolution analysis, topic detection and tracking, and multi-document summarization.



Bicheng Li, Professor, teaches at Huaqiao University as a professor. His teaching subjects include pattern recognition and artificial intelligence, and wavelet transform. His research fields include text analysis and understanding, speech/image/video processing and recognition, and information fusion.



Yuan Gao, Master student, studies at Zhengzhou Information Science and Technology Institute. He received his bachelor degree from Zhengzhou Information Science and Technology Institute in 2012. His research interest covers natural language processing.



Yongwang Tang, Master, received his M.S. degree in control theory and control engineering from Kunming University of Science and Technology in 2008. His research fields include semantic Web technologies, information fusion, and database management.

