Neurocomputing 181 (2016) 64–75

Contents lists available at ScienceDirect

Neurocomputing

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

Event graph based contradiction recognition from big data collection

Maofu Liu [a,*], Limin Wang [a], Liqiang Nie [b], Jianhua Dai [c], Donghong Ji [d]

[a] Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, College of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065, China
[b] School of Computing, National University of Singapore, Singapore 117417, Singapore
[c] School of Computer Science and Technology, Tianjin University, Tianjin 300072, China
[d] School of Computer, Wuhan University, Wuhan 430072, China

ARTICLE INFO

Article history:
Received 15 March 2015
Received in revised form
16 June 2015
Accepted 18 June 2015
Available online 5 December 2015

Keywords:
Contradiction recognition
Event graph
Big data
High-level event semantic feature
Atomic event
Event semantic role

ABSTRACT

Conventional models relying on similarity utilizing low-level surface statistical, syntactic and lexical semantic features are suboptimal in contradiction recognition, especially for the large-scale data, such as the sensor and traffic data. To tackle this problem, this work treats the text and hypothesis pair as event graph and proposes a novel model based on event graph to recognize contradiction in big data collection. The proposed model is capable of seamlessly sewing the high-level event semantic features corresponding to the conflicting linguistic phenomena to identify contradictory construction. Experimental results show that our event graph based contradiction recognition model achieves significant improvement as compared to state-of-the-art competitors.

1. Introduction

The past decade has witnessed that more and more attentions have been paid to recognizing entailment in big data context, such as the large-scale sensor and traffic data. The textual entailment is mainly due to the variability of semantic expression, which is the fundamental and essential characteristic of natural language. Theoretically, the constraint of textual entailment is less restrictive than logical entailment and a common definition of textual entailment in formal semantics is formulated as a text T entails another text H if the text H is almost likely true in each conceivable circumstance in which the text T is true.

In addition to textual entailment, the textual contradiction is another important type of semantic relation for the textual inference and has been recently received a surge of interest in the computational linguistics community. The textual contradiction is the negation of textual entailment in some way. The contradiction recognition was formally introduced in the fourth recognizing textual entailment (RTE) challenge in 2008 [1–4] since the semantic relations of both entailment and contradiction are the key data of semantics.

Finding the semantic relation of contradicting texts is also foundational for text understanding in big textual data and many attentions have been paid to the textual contradiction recognition in large-scale English text pairs. In fact, there also exists the conflicting phenomena of semantic expressions in the other kinds of natural language, especially Chinese. The recognizing inference in text (RITE) challenge attempts to recognize the textual entailment in Chinese texts and its multi-class (MC) subtask has included the contradiction recognition [5–7]. To the best of our knowledge, despite its value and significance, contradiction recognition from big Chinese data is still in its infancy. In this paper, we seek to understand the ways contradictions occur across Chinese text pairs.

In the big data collected from intelligent transportation system, social media and the other sources, there also exists the contradictory textual evidences to assist in making the decision. With the exponential growth of Chinese online content, especially traditional textual content in Chinese websites and user generated contents (UGCs) in Chinese microblogging and social media services, textual entailment and contradiction are indeed the pervasive phenomena in Chinese texts and worthy of being deeply and thoroughly investigated. In the

* Corresponding author.
E-mail addresses: liumaofu@wust.edu.cn (M. Liu), smile_wlm@163.com (L. Wang), nieliqiang@gmail.com (L. Nie), david.joshua@qq.com (J. Dai), dhji@whu.edu.cn (D. Ji).

http://dx.doi.org/10.1016/j.neucom.2015.06.099
0925-2312/

following text pair gathered from Weibo,[1] a famous microblogging service in China, T0 is part of the initial post about cab-hailing applications by one Weibo user and H0 is the comment to T0 by the other Weibo user.

In Chinese:

T0: 打车软件可以为智能交通系统提供有价值的第一手交通数据。

H0: 打车软件无数据价值。

In English[2]:

T0: The cab-hailing apps can provide valuable first-hand traffic data for the intelligent transportation system.

H0: The cab-hailing apps have no data value.

In the textual entailment recognition, the low-level surface statistical, syntactic and lexical semantic features have been used to estimate the similarity between the specified text pair, but they do not work well in the textual contradiction recognition. The text T0 will be recognized to entail the hypothesis H0 according to their very high similarity, but the right semantic relation between T0 and H0 is contradiction. In order to describe this issue clearly, let us consider the following text pair from RITE-1 dataset in Simplified Chinese as another example.

In Chinese:

T1: 电影《无极》在 12 月 11 日举行首映典礼。

H1: 电影《无极》在 12 月 12 日晚将举行首映典礼。

In English:

T1: The movie premiere of THE PROMISE will be held in December 11.

H1: The movie premiere of THE PROMISE will be held in the evening of December 12.

In above text pair, text T1 and hypothesis H1, will be recognized to hold the paraphrase or entailment because of their very high similarity. However, the intrinsic semantic relation between T1 and H1 is contradiction due to the high-level conflicting temporal information of event 首映典礼 (premiere), 12 月 11 日 (December 11) in text T1 and 12 月 12 日晚 (the evening of December 12) in hypothesis H1. Moreover, in official evaluation of RITE-1, precision and recall of contradiction in RITE systems mainly based on similarity related features are much lower than those of entailment and paraphrase. So, besides textual entailment recognition, the high-level semantic features corresponding to the conflicting linguistic phenomena are also necessary for textual contradiction recognition.

In order to solve the issues mentioned above, in this paper, we introduce the high-level event semantic features to enhance the contraction recognition and propose a model for automatically detecting these types of conflicting constructions.

The contributions of this paper are summarized in threefold:

(1) We specialize in the conflicting expressions in Chinese language and attempt to recognize textual contradiction in big Chinese data.
(2) We propose a novel event graph based model to recognize textual contradiction.
(3) After analyzing event semantic conflicting phenomena in Chinese text pairs, we incorporate event graph based semantic features into our model to enhance the recognition of textual contradiction.

The rest of this paper is organized as follows. Section 2 reviews the related work. Section 3 overviews our event graph based textual contradiction recognition model. Section 4 then presents experiments and discussions. Finally, we conclude our work in Section 5.

2. Related work

Dagan et al. [8,9] defined the concept of textual entailment "*T* entails *H*" as a directional relationship that if a human reading the entailing text *T* would typically infer that the entailed hypothesis *H* is most likely or probably true based on common human understanding of language as well as common background knowledge. The strict logical definition of contradiction is that the texts *A* and *B* are contradictory if there is no possible world in which texts *A* and *B* are both true [10–13]. The looser definition, closely matching the human intuitions of incompatibility in description of the same event, is that the contradiction occurs when two texts or textual fragments are extremely unlikely to be true at the same time [5,6,10]. Textual contradiction is usually embodied in opposite polarities on the whole and the use of negation, antonyms or opposite language structure in detail [11,12].

Condoravdi et al. first recognized that textual entailment, as well as textual contradiction, is very important to robust text understanding, and defined them in formal semantics with the strict logical condition, but they did not report any empirical results [13]. Harabagiu et al. provided the first empirical results for contradiction detection [11]. De Marneffe et al. proposed an appropriate definition of contradiction for natural language processing (NLP) tasks and constructed a typology of contradictions [10]. Almost all the existing studies on contradiction recognition are about text and hypothesis pairs in English [14–17]. In this paper, we are interested in the given Chinese text and hypothesis pairs instead of the English ones.

The machine learning intends to automatically identify complex patterns and make intelligent decisions based on the observed data and the machine learning methods have been applied to classification problems [18,19]. In fact, some machine learning methods also could be used to classify the English text pairs according to the semantic relation in them, since the textual entailment or contradiction recognition is one kind of classification problem [11,20,21]. In this paper, we commit to recognizing textual contradiction in large-scale Chinese text pairs collection by the typical classification model of support vector machine.

[1] http://www.weibo.com.

[2] The English versions for all the examples are translated from the Chinese ones.