Computer Science and Information Systems 14(3):823–837 https://doi.org/10.2298/CSIS161229031W

Sentiment information Extraction of comparative sentences based on CRF model

Wei Wang, Guodong Xin*, Bailing Wang*, Junheng Huang, and Yang Liu

School of Computer Science and Technology, Harbin Institute of Technology,
150001, Harbin, China
{wwhit, gdxin,wbl, hithjh}@hit.edu.cn, lyylwhhit@126.com

Abstract. Comparative information mining is an important research topic in the sentiment analysis community. A comparative sentence expresses at least one similarity or difference relation between two objects. For example, the comparative sentence "The space of car A is bigger than that of car B and car C" expresses two comparative relations <car A, car B, space, bigger> and <car A, car C, space, bigger>. This paper introduces conditional random fields model to extract Chinese comparative information and focuses on the task of element extraction from comparative sentences. We use the conditional random fields model to combine diverse lexical, syntactic and semantic features derived from the texts. Experiments show that the proposed method is competitive and domain-independent, with promising results.

Keywords: information extraction, comparative sentence, comparative element.

1. Introduction

Whenever people need to make a decision, they commonly want to know about others' views, attitudes and sentiments. A comparative sentence provides an important insight into how an entity or event is compared to other entities or events, which could effectively help people make decisions. For example, *" X旅馆比Y旅馆更干净，尽管房间的价格相同(Hotel X is cleaner than Hotel Y, although its price is the same as Y.)"*. Such opinion about comparison, directly comes from customers, could provide greater help for those who have potential consumer demands, but also help business executives to automatically track the attitudes and emotions of customers in the on-line forums, determine whether the customers are satisfied with their products and services, and capture the information of competitors. Therefore, the development of effective methods to automatically analyze opinions, especially comparative opinions, has become an urgent need [8,22,15,21,12,25,5].

The processing object of comparative sentiment analysis(SA) is comparative sentences in evaluative texts, the task is to extract and analyze the opinion elements in the comparative sentences, including judging the tendency of each comparative relation and extracting the various elements related to the tendency. These elements include compared entities, compared aspects, comparative words and opinion words. For example, the comparative sentence *"X手机比Y 手机有更好的用户体验.（Phone X has better user experience than phone Y.）"*. We extract 'phone X' as a subject entity(SE), 'phone Y' as object entity (OE), 'user experience' as a compared aspect (CA), and 'better' as an opinion phrase (OP) related to 'phone X', 'than' as a comparative keyword (CK).

The primary task of comparative sentiment analysis is to locate and extract the comparative elements in sentences, and then to determine the emotional tendency of the author for different objects according to the extracted contents. Information extraction in comparative sentences is different from that in regular opinion sentences. It extracts the objects with comparative relation and their shared aspects, rather than extracts a single entity or aspect that is directly evaluated by the author. The comparative relations between entities are usually reflected by the comparative words, so this paper introduces the comparative word candidate features and heuristic position features to improve the system's ability to identify compared entities. In addition, the comparative element extraction has the following problems:

Problem 1: How to fully identify phrase-level elements, for example, a product name may be consisted of a brand name and a model name. If we only extract a part of them, it will cause the lack of information. Therefore, we introduce shallow syntactic features to enhance the ability of system to identify phrase-level elements.

Problem 2: How to distinguish between different types of elements, for example, the POS tags for SEs, OEs and CAs are usually nouns or noun phrases. Therefore, it is difficult to distinguish these three types of elements. But Relative positional relations between them and comparative words have some directive functions.

To sum up, in order to construct a general comparative element extraction system, we introduce some linguistic features and heuristic features, such as shallow syntactic features, comparative word candidate features, and heuristic position features. In the case of no increase of domain knowledge, the performance of comparative element extraction is improved effectively, which shows the effectiveness of the proposed method.

The remainder of the paper is organized as follows, section 2 presents related work. Section 3 describes the method for comparative element extraction from comparative sentences. After that, the experiment results and the future directions are given in section 4.

2. Related work

There are many unsupervised methods [6,14,8,22,21,3,24] for aspect term extraction in review texts. Hu and Liu[6] first study the problem, they extract aspect terms through the association rule mining. And then they employ opinion words to mine infrequent aspect terms. Many of subsequent studies use the relationships between opinion words and aspect words to extract the aspect terms and opinions. In Qiu's[14] work, dependency relations are used as key clues, and the dual propagation approach is proposed to extract aspect words and opinion words by propagating information between opinion words and aspect words. The method is a semi-supervised bootstrapping process because the use of opinion word seeds. The purpose of comparative element extraction is to obtain various components associated with the comparative statement. Jindal and Liu first define the comparative element extraction problem [8] , where they deem a comparative sentence that describes a comparative relation that is consisted of five fundamental elements: comparative keyword, two compared entities, compared aspect and comparative type. They present a new method based on sequence rule mining called as "label sequence rule(LSR)" . The LSR method can extract the elements in a single comparative relation. Yang and Ko[22] propose an alternative approach that marks comparative element candidates based on part of speech (POS) type, then constructs POS sequence patterns for each candidate and treats them as features of machine learning algorithm. These two works are based on context POS information to obtain comparative elements with a certain type of POS. In addition, some researchers use the dictionary including the domain data to mine comparative elements. Xu et al. [21] compile a product dictionary and an attribute dictionary in mobile phone domain by collecting some common product names and attribute names manually in corresponding domain. Feldman et al [3] build a brand dictionary for running shoes and cars respectively and recognize product model by developing a set of regular expressions for the model-names. The approaches based on dictionary are domain-related, which have many limitations.

There are some researchers adopting bootstrapping technique to extract compared entities. Li et al [12] develop a weakly-supervised bootstrapping method for automatic compared entities mining from online comparative questions. In their study, the algorithm starts bootstrapping process with a single extraction pattern. Using it, a set of initial seed comparator pairs are extracted from a question collection. Next, new extraction patterns are generated from comparator pairs and the comparative questions containing comparator pairs. The algorithm iterates until no more new patterns are found from the question collection. In Ding et al study [1], the bootstrapping process starts with a few seed entities. From them, the algorithm iteratively find more entities in a document set. At each iteration, sequence patterns are mined to find more entities based on already found entities. Obviously, their works are weakly-supervised and do not need to label a large number of corpora.

Supervised methods [20,7,23,18,9] often treat aspect and opinion word extraction as a sequence labeling problem, and the Conditional Random Fields (CRF) is one of the most main-stream methods used for sequence labeling tasks. Xing et al [20] select keywords, noun phrases and their position information as the features of CRF model to build up an element extraction model for identifying technical indices in standard technical comparative sentences. Huang et al. [7] first identify compared entities in sentences using CRF model. Then they distinguish compared subjects from compared objects based on the relative position between the entity and keyword. Yin et al.[23] extract aspect terms based on the features of distributed representations of words and dependency paths. They regard multi-hop dependency paths as a sequence of syntactic relations. In learning the embedding features, they not only use the word, but also consider the richer context information, such as neighbor words, and the dependency context information. Wang et al. [18] propose a new model by integrating recursive neural networks and conditional random fields into a unified framework for aspect and opinion term extraction. The recursive neural networks can learns high-level features by utilizing the double propagation of aspect-opinion pairs in the dependency tree. The learned features are input into the CRF model to capture the context of each word for aspect and opinion term extraction.

Representation learning has been successfully applied to natural language processing, such as information extraction, sentiment analysis [19,16,2] and so on. It represents text in different granularities with a low-dimensional dense vector, which includes context semantic information. Wang et al. [19] perform aspect level sentiment classification using an Attention-based LSTM (Long Short-Term Memory) networks. Attention can focus on different parts of a sentence when different aspects are used as input. Tang et al. [16] design a deep memory network with multiple computational layers for aspect level sentiment classification. Each layer of the deep memory network is an attention model with an external memory to calculate the importance of each context word of a given aspect. Dong et al.[2] employ adaptive recursive neural network(AdaRNN) to perform target-dependent Twitter sentiment classification. AdaRNN uses more than one composition functions and adaptively choose them based on the context and linguistic tags.

In the context of comparative element extraction, there are some scholars converting element extraction task into semantic role labeling task. Wang et al [17] define three types of comparative patterns (eg. <entity> <keyword> <entity> <sentiment word>) to describe the relation among comparative elements. They employ the generalization comparative patterns to label comparative elements. Li [11] constructs semantic role parsing trees by utilizing semantic role labeling package and Stanford parser. They calculate the similarity between two sub-trees to label comparative elements. However, the above works can just obtain elements in a single comparative relation.

The work to determine entities preferred by reviewer has also been explored. Ganapathibhotla and Liu [4] primarily deal with context-sensitive sentiments by exploiting external information available on the Web. In this study, we use the Conditional Random Fields (CRF) learning algorithm to identify comparative elements. Lafferty et al [10] first introduce CRF for segmenting and labeling sequence data.

This paper uses the supervised method to extract comparative elements. Compared with the existing studies, our method makes full use of the various lexical, syntactic and heuristic information of the comparative sentence. And multiple key elements of comparative sentences are extracted at the same time.

3. Methods

3.1. Comparative sentence key concepts

Comparative information plays an important role in dealing with some practical problems, such as decision-making, opinion summarization, etc. Here, we give some basic definitions of comparative information mining(CIM) at sentence level.

Definition (comparative information mining): CIM is a problem of finding the comparison information between entities in text documents, which can be decomposed into the following main subtasks: I. Identify comparative sentences. II. Extract comparative elements and relations.

Definition (comparative sentence): A comparative sentence is a sentence that expresses one or more comparative relations between objects, which means that there may be more than one comparative relation in a sentence.

A comparative sentence can be explicit, e.g., *" X电视比Y电视画面清晰.(TV X has a clearer picture than TV Y.)"* or implicit, e.g., *"X手机有摄像功能，而Y 手机没有. (Phone X has a camera function, but phone Y does not have.)"*.

Definition (comparative relation): A comparative relation describes a relation of similarity or difference between two objects on an aspect.

A comparative relation can be formally expressed as a 5-tuple: *(SE, OE, CA, OP, CK)*, which refers to subject entity, object entity, compared aspect, opinion phrase, and comparative keyword. Some elements in a comparative relation can be omitted. For example, in a superlative sentence, object entity is usually being omitted.

Definition (comparative keyword): A comparative keyword is an indicator of comparative relation, for example '比*(than)* ', '相似*(similar)* ', '不同*(different)* ', '最*(most)*

' etc. There are not the specialized morphemes in Chinese, such as the -er/-est suffix, as the comparative characteristics.

Definition (compared entity): A compared entity is an object that is being compared with another object in a sentence, which can be a subject entity or an object entity. A compared entity can be almost anything, e.g., a people, a place, a product, an event, etc.

Definition (compared aspect): A compared aspect is an aspect on which two objects are being compared. An aspect can be explicit or implicit in a sentence, for example, *"钻石的价格高于珍珠.(The price of diamonds is higher than that of pearls.)", and "钻石比珍珠更昂贵.(Diamonds are more expensive than pearls.)"*.

There are two main comparative types: gradable and non-gradable. Gradable comparison describes an order relationship of entities with regard to an aspect. For example, sentences comprising phrases such as '比⋯性能更好(better performance than)', '低于(lower than)', '相比⋯有所提高(improved⋯compared with)' are typically classified to gradable comparison. We further divide gradable comparison into two sub-types, greater or less than comparison, and superlative comparison. The latter generally contains phrases such as 'the most expensive', 'the best quality' etc, for example, the sentence " 在所有手机品牌中, iphone是最受欢迎的.(In all mobile phone brands, iphone is the most popular.) " is a gradable superlative comparison where we extract 'iphone' as a subject entity, 'popular' as an opinion phrase, and 'most' as a comparative keyword.

Non-gradable comparison describes similarity or difference between entities, and does not express the order of entities. We further divide it into three sub-types, similarity comparison, difference comparison and implicit comparison. Non-gradable similarity comparison expresses the similarity of entities by using phrases such as ' 和⋯一样(the same⋯as, as ⋯as)', '和⋯相似(similar to, similarity between)' in a sentence. For example, in a camera review, the sentence "The photo quality of camera X is as good as camera Y." indicates that the similarity in picture quality between camera X and camera Y. Non-gradable difference comparison states the difference of entities on a certain aspect, and does not grade them. Phrases such as 'different from', 'distinguish from', 'difference between' can be the indicator of such type sentence. For example, in the sentence "The screen size of monitor X is different from that of monitor Y", the user expresses the difference between monitor X and monitor Y in screen size, without ordering them based on the size of screen. Non-gradable implicit comparison implicitly states the difference of entities on one or more aspects, for example, the sentence "Entity X has aspect A1, but entity Y does not have.".

3.2. Extraction of comparative elements

Comparative elements In this section, we describe how comparative elements are extracted from comparative sentences. The basic strategy is an integrated lexical, syntactic and semantic features and condition random fields learning approach to extract comparative elements.

There are four types of comparative elements to be extracted in our study: subject entity(*SE*), object entity(*OE*), compared aspect(*CA*), and opinion phrase(*OP*).

Example 1. *"手机X的摄像头比手机Y的更好更实用. (Phone X has a better and more practical camera than phone Y.)"*

Example 2. *"在所有汽车中，Z性能最优越. (The performance of Z is the most superior in all cars.)"*

In Example 1 sentence, *'手机X(phone X)'* is a SE, *'手机Y(phone Y)'* is an *OE*, *'摄像头(camera)'* is a *CA*, *'更好更实用(better and more practical)'* is a *OP*. In Example 2 sentence, 'Z' is a *SE*, *'性能(performance)'* is a *CA*, *'优越(superior)'* is a *OP*.

There are two important problems need to be solved in the task of comparative element extraction: i) whether comparative elements are composed of only a single word; ii) how to distinguish SE, OE and CA that have similar POS tags.

Composition of Elements: comparative elements can be composed of one or more words. For instance, "better and more practical" is composed of multiple words in example 1. If we only extract one word "better" substituted for "better and more practical", some important information will be lost. We thus define that comparative elements can be composed of one or more words.

Distinction of Elements: Subject entity, object entity and aspect are mainly noun or noun phrase. So, we could not effectively distinguish them by only using the POS tags. Fortunately, we find that various elements commonly play different grammatical roles in comparative sentences. For instance, Subject entity is mainly as the subject of sentence, object entity acts as the object, and opinion phrase is as the predicate in the syntax function. Furthermore, we also find that subject entity is usually in the left of a keyword and object entity is in the right of a keyword. These linguistic clues are useful for distinction of SE, OE and CA elements.

Feature representation We introduce various linguistic-related features to extract comparative elements. Several preprocessing steps are executed towards comparative sentences, including word segmentation, POS tagging, phrase syntactic parsing. In this study, we use some basic linguistic features and more advanced ones as follows:

1) Words: A Word is the smallest linguistic unit that expresses natural language semantics. In western phonetic language, there is a clear delimiter between words. In Chinese, there is no obvious delimiter between words. Therefore, we first perform word segmentation for each sentence. Then each word in a sentence is used as a baseline feature of CRF model for Chinese information extraction work.

2) POS tags: part-of-speech tag is also a class of important features. Due to SE, OE and CA are mainly noun. Sentiment words are commonly adjective or verb. Comparative keywords are mainly preposition or adverb. Hence, POS tags are helpful for identifying different types of elements.

3) Chunks: Chunk division, also known as shallow parsing, is used to recognize independent components in a sentence whose structure is relatively simple, such as non-recursive noun phrases, verb phrases etc. The chunk labels are derived from syntactic parsing tree of a sentence. In a comparative sentence, SE and CA can be composed of noun or noun phrase. Keyword and OE usually form a preposition phrase. OP can be adjective or adjective phrase. So, chunk feature can contribute to identify comparative elements in phrase level.

4) Keywords: The keyword candidates in a comparative sentence are labeled by using a set of paired keywords e.g. *"与···不同(different...from)"*. A lexicon of 660 paired keywords is created by counting their co-occurrence frequency, and then pruning manually. The keyword candidates are useful for discriminating SEs from OEs in a comparative sentence.

5) Positions: Most of SEs are in the left of keywords and OEs are in the right of keywords in comparative sentences. By using the heuristic position information between entity and keyword can further distinguish SE from OE.

6) Contexts: The context of a word in a sentence can also affect the type of element. In this study, we use context within the radius of 3 of each target word in a sentence as feature. The context feature is set by feature template of CRF model.

The above linguistic features are automatically extracted by using Stanford segmenter, and Stanford parser.

Conditional random field model Conditional random fields (CRF) [10], which is an undirected probabilistic graphical model, has the following advantages for labeling and segmenting sequence data: i) CRF can effectively exploit the rich, global features of the inputs, and do not need to represent dependencies of the inputs. ii) Context information are taken into account by CRF, e.g., the linear chain CRF predicts sequences of labels for sequences of input samples in natural language processing. iii) Long-range dependencies between the inputs can be represented. The extraction of comparative elements involves multiple entities, rich features from the inputs, and long-range dependencies. Thus CRF is the very appropriate algorithm for modeling it.

In this paper, we adopt CRF++0.53 toolkit to execute training and labeling for model. The features extracted from the feature set are added to the model by setting feature template. Therefore, the feature selection problem is transformed into a feature template selection problem. This paper designs 6 feature templates based on the linguistic related features described above as shown in Table 1.

In Table 1, *w, t* denotes word and POS tag feature respectively. *c, l* represents comparative word candidate and heuristic position feature. *s* denotes shallow parsing feature. In order to verify the effective of syntactic and heuristic features, we build 6 feature templates in the experiments. Followed by the lexical level (baseline) feature template(T1), comparative word candidates are added to T1 template(T2), comparative word candidate and heuristic position and word features(T3), comparative word candidate and heuristic position features are added to T1 template (T4), shallow parsing features are added to T1 template(T5), All features (T6).

4. Experimental evaluation

We conduct various experiments to evaluate the performance of the proposed methods for comparative element extraction task.

4.1. Experiment data

The experiment data is derived from task 2 of the fourth Chinese Opinion Analysis Evaluation (COAE2012) [13]. It consists of consumer reviews of automotive and electronic products. The sentence distribution of the data is shown in Table 2. The ratio of training set, development set and test set is 4: 4: 1.

The COAE2012 task 2 is divided into two sub-tasks. Task 2.1: Identify which sentences are comparative sentences in a given set of sentences. Task 2.2: Extract comparative elements from the identified comparative sentences, including compared entity and compared aspect, and determine the opinion direction of compared entities.

Table 1. Feature Template

Template	Feature	Feature Template
T1	*w,t*	w_n ,t_n $n \in \{-3, ..., 3\}$
		$w_{n-1}w_n$,$t_{n-1}t_n$ $n \in \{0, 1, 2\}$
		w_nw_{n+1} ,t_nt_{n+1} $n \in \{-2, -1, 0\}$
		$t_{n-1}t_n$ t_{n+1} $n \in \{-1, 0, 1\}$
		w_nt_n $n = 0$
T2	*w,t,c*	w_n ,t_n ,c_n $n \in \{-3, ..., 3\}$
		$w_{n-1}w_n$,$t_{n-1}t_n$,$c_{n-1}c_n$ $n \in \{0, 1, 2\}$
		w_nw_{n+1} ,t_nt_{n+1}, c_nc_{n+1} $n \in \{-2, -1, 0\}$
		$t_{n-1}t_n$ t_{n+1}, $c_{n-1}c_n$ c_{n+1} $n \in \{-1, 0, 1\}$
		w_nt_n ,t_nc_n $n = 0$
T3	*w,c,l*	w_n ,c_n ,l_n $n \in \{-3, ..., 3\}$
		$w_{n-1}w_n$,$c_{n-1}c_n$,$l_{n-1}l_n$ $n \in \{0, 1, 2\}$
		w_nw_{n+1} ,c_nc_{n+1}, l_nl_{n+1} $n \in \{-2, -1, 0\}$
		$c_{n-1}c_n$ c_{n+1}, $l_{n-1}l_n$ l_{n+1} $n \in \{-1, 0, 1\}$
		w_nc_n ,c_nl_n $n = 0$
T4	*w,t,c,l*	w_n ,t_n, c_n ,l_n $n \in \{-3, ..., 3\}$
		$w_{n-1}w_n$,$t_{n-1}t_n$,$c_{n-1}c_n$,$l_{n-1}l_n$ $n \in \{0, 1, 2\}$
		w_nw_{n+1} ,t_nt_{n+1} ,c_nc_{n+1}, l_nl_{n+1} $n \in \{-2, -1, 0\}$
		$t_{n-1}t_n$ t_{n+1},$c_{n-1}c_n$ c_{n+1}, $l_{n-1}l_n$ l_{n+1} $n \in \{-1, 0, 1\}$
		w_nt_n ,t_nc_n ,t_nl_n ,c_nl_n $n = 0$
T5	*w,t,s*	w_n ,t_n ,s_n $n \in \{-3, ..., 3\}$
		$w_{n-1}w_n$,$t_{n-1}t_n$,$s_{n-1}s_n$ $n \in \{0, 1, 2\}$
		w_nw_{n+1} ,t_nt_{n+1}, s_ns_{n+1} $n \in \{-2, -1, 0\}$
		$t_{n-1}t_n$ t_{n+1}, $s_{n-1}s_n$ s_{n+1} $n \in \{-1, 0, 1\}$
		w_nt_n ,t_ns_n $n = 0$
T6	*w,t,c,l,s*	w_n ,t_n, c_n ,l_n ,s_n $n \in \{-3, ..., 3\}$
		$w_{n-1}w_n$,$t_{n-1}t_n$,$c_{n-1}c_n$,$l_{n-1}l_n$,$s_{n-1}l_n$ $n \in \{0, 1, 2\}$
		w_nw_{n+1} ,t_nt_{n+1} ,c_nc_{n+1}, l_nl_{n+1}, s_nl_{n+1} $n \in \{-2, -1, 0\}$
		$t_{n-1}t_n$ t_{n+1},$c_{n-1}c_n$ c_{n+1}, $l_{n-1}l_n$ l_{n+1} ,$s_{n-1}s_n$ s_{n+1} $n \in \{-1, 0, 1\}$
		w_nt_n ,t_nc_n ,t_nl_n ,t_ns_n ,c_nl_n,c_ns_n ,l_ns_n *n*=0

Table 2. Sentence distribution of the data

Type	Sentence distribution
Comparatives	1624(16.92%)
Non-comparatives	7976(83.08%)
Total	9600(100%)

Task 2.2 marks three parts: ProductName, FeatureName and Polarity. Our study is similar to task 2.2. In this task, the coverage is used to assess for the consistency. Set *x, y* are the results of different people annotation, coverage is defined as follows:

$$Coverage(x, y) = len(x \cap y)/len(x) * 100\% \tag{1}$$

Where *len(x)* represents the length of *x*, $x \cap y$ is the intersection of *x* and *y*. We set coverage is 0.2 in the experiment.

4.2. Evaluation methods

Task 2.2 is an information extraction task. It is difficult to determine the boundary of ProductName and FeatureName for task 2.2. Therefore, evaluation adopts two indicators: accurate evaluation and coverage evaluation.

Accurate evaluation: the extracted entity exactly matches with the answer. For example, when the answer is 'screen resolution', it is incorrect result to submit either 'screen' or 'resolution'.

Coverage evaluation: the extracted entity has overlap portion with the answer. In the above example, it is correct result if we submit 'screen' or 'resolution'.

4.3. Experimental results

Experimental results of comparative element extraction We use the comparative sentences in the automotive and electronic fields in COAE2012 task 2 to extract the comparative elements, a total of 1600 comparative sentences. Most of these sentences are typical comparative sentences, and a few implicit comparisons. The distribution of comparative elements is shown in Table 3. Stanford parser is used to perform phrase syntactic parsing. The experimental results are an average of 5 fold cross validation. We use two evaluation methods, accurate evaluation and coverage evaluation to measure the performance of system. The experiment results are shown in Table 4 where SUB represents subject entities, OBJ represents object entities, ATTR represents aspect names, OPIN represents evaluation words or phrases.

Table 3. The distribution of comparative elements in two fields

Field	Comparative Sentences	Subject Entity	Object Entity	Compared Aspect	Keyword	Opinion Phrase
Car	800	650	810	836	1421	831
Electronic	800	505	860	687	943	802

Table 4 shows the average result of element extraction in two fields. When introducing all features (T6 template), the results of element extraction are superior to other feature combinations (T1-T5 template). When using T1 template, the performance of system is poor, particularly recall.

Because T1 template contains only lexical level features, such as words and POS tags, these features can provide limited information for classification task, and the information

Table 4. The average results of 5-fold cross validation(%)

Element	Template	Accurate Evaluation			Coverage Evaluation		
		Precision	Recall	F1-score	Precision	Recall	F1-score
SUB	T1	67.43	39.03	48.78	74.91	41.53	53.44
	T2	68.47	41.57	50.99	73.29	47.35	57.53
	T3	73.12	32.00	43.83	76.41	36.44	49.35
	T4	70.25	41.81	51.51	75.66	48.01	61.29
	T5	66.08	37.94	48.21	72.25	42.19	53.12
	T6	71.61	41.36	51.54	80.44	50.31	61.90
OBJ	T1	81.60	66.93	73.36	83.00	69.11	75.42
	T2	81.57	69.83	74.99	84.72	72.02	77.86
	T3	78.05	70.63	73.82	78.77	72.71	75.62
	T4	80.75	73.77	76.90	87.88	76.89	82.02
	T5	81.78	66.13	73.02	83.86	68.25	75.25
	T6	82.22	73.03	77.18	91.69	77.21	83.24
ATTR	T1	72.80	48.38	58.13	78.17	50.04	61.02
	T2	74.43	52.83	61.80	79.96	55.57	65.57
	T3	76.51	39.69	52.27	80.66	42.84	55.96
	T4	73.88	52.11	61.11	78.88	55.31	65.03
	T5	71.36	49.71	58.12	75.06	51.73	61.25
	T6	73.70	51.74	60.80	81.95	55.91	66.47
OPIN	T1	87.12	61.67	72.17	89.15	62.05	73.17
	T2	87.38	64.55	74.19	88.98	66.48	76.10
	T3	88.69	64.26	74.44	88.77	66.10	75.78
	T4	87.45	68.47	76.74	90.51	68.98	78.29
	T5	86.34	62.95	72.70	88.45	64.97	74.91
	T6	87.08	68.85	76.86	89.30	71.15	79.20

obtained contains some noise. T2 template expands features from lexical level to heuristic information. It adds keyword candidate feature that makes the evaluation indicators to be significantly raised. The performance of the T3 template is polarized. On the one hand, the worst performance is gotten for SE and CA identification. Because the T3 template does not contain part of speech tag feature, it is the primary indication of the subject entities and attributes. On the other hand, keyword candidate and heuristic position features are added to T3 template, which improve the performance of OE and OP identification.

T4 template adds keyword candidate and heuristic information on the basis of T1 template, and provides the position information of other elements relative to candidate keywords in the sentence. The recall and F1-score are greatly improved, which show that keyword candidate and heuristic position features are very effective in the comparative element extraction problem.

However, T4 template has limited recognition ability for phrase level elements. Thus, T5 template expands features from lexical level to phrase level. It adds shallow parsing feature which improves the F1-score of CA and OP, but decrease the F1-score of other elements. T6 template, which includes all features, greatly increases the recall and F1-score of system. This proves that various features, such as lexical, syntactic, and heuristic features, are effectively for the comparative element extraction.

Table 4 compares the performance of the system for accurate evaluation and coverage evaluation. The best performance is obtained when we use coverage evaluation. This means that the system can correctly locate the comparative elements, but the ability to accurately identify the boundaries of the elements is limited. As we can see in Table 3, the precision is relatively high, while the recall is low in each of results. One possible reason is that multiple feature decisions improve the precision of system, while reduce the recall. The other reason is that domain knowledge is not introduced into the system. Experimental results show that the annotated results of OE and OP are better than those of SE and CA. Since the positions of OE and OP in the sentence are relatively fixed, they are commonly in the right of keyword or are degree adverbs. While a number of SEs to be omitted, and the positions of CA to be unfixed increase the difficulty of identifying them.

As the conditional random fields is a supervised learning method, there are domain adaptability problems. In order to verify the effectiveness of our method, for the car field, we use the electronic field corpus as training set. Similarly, for the electronic field, we use the car field corpus as training set. The average of these experiments is taken as the final experimental result.

Table 5. The results of domain cross annotation(%)

Element	Template	Accurate Evaluation			Coverage Evaluation		
		Precision	Recall	F1-score	Precision	Recall	F1-score
SUB	T6 Template	58.15	18.51	27.91	64.53	24.74	35.77
OBJ	T6 Template	79.00	58.42	66.65	85.60	61.70	71.71
ATTR	T6 Template	63.53	37.05	45.14	67.75	40.28	50.52
OPIN	T6 Template	80.43	60.27	68.78	82.67	62.59	71.24

By comparing table 5 with table 4, we find that, in Table 5, the model established by domain cross training has a substantial decrease in the performance of element extraction compared to table 4. Among them, the subject entities and compared aspects have the biggest decrease, and the comparative keywords and opinion words have a smaller decrease. The reason is that the subject entities, object entities and compared aspects are domain related. For example, a subject entity or object entity is usually a brand name or product name in a domain, and an aspect is a component or characteristic of a product. Thus, these three elements are domain related. On the other hand, the position of the subject and attribute varies greatly in the sentence, and aspect is not easily distinguished from subject entity. Therefore, domain cross annotation has the greatest impact on the subject entity and attribute. Since the position of object entity is relatively fixed, the recognition performance is better than that of the subject entity and attribute. Because of the small domain correlation of opinion words, and its recognition performance is relatively good.

Compare with COAE2012 evaluation results The average result of element extraction using T6 template in our experiments is compared with the max value of evaluation results in COAE2012. In contrast experiment, the average of 5-fold cross validation is adopted. The result is shown in Table 6, where PROD represents product name(SE and OE), ATTR represents aspect name.

Table 6. The result contrast on COAE2012 data (%)

Element	Method	Accurate Evaluation			Coverage Evaluation		
		Precision	Recall	F1-score	Precision	Recall	F1-score
ATTR	T6 Template	73.70	51.74	60.80	81.95	55.91	66.47
	Max value	66.05	62.52	60.78	77.94	67.51	65.69
PROD	T6 Template	76.92	57.20	64.36	86.07	63.76	72.57
	Max value	67.77	66.05	64.30	82.67	73.58	71.58
PROD+	T6 Template	75.84	55.38	63.17	84.69	61.14	70.54
ATTR	Max value	60.81	53.89	52.55	67.45	58.56	57.00

Table 6 shows that F1-scores of extracting entity, aspect, entity and aspect are higher than the max values of COAE2012, which indicate the proposed method in this paper is effective. Table 6 shows the precision is higher, and the recall is lower in each result. On the one hand, the positions of SEs and CAs in comparative sentences are too flexible to capture, and SEs are often omitted in comparative sentences, which can affect the mean recall of system. On the other hand, we do not introduce any domain knowledge, such as domain knowledge base, domain dictionary in the process of element extraction. If the domain dictionary is introduced in the model training phase, it will improve the recall of the extraction results. However, the cost of the artificial domain dictionary is relatively large and can not be applied to other fields. Therefore, in order to improve the recall of the system, it is necessary to find more effective features of the universal domain to solve the problem. In addition, the indices of all coverage matching of the system are higher than those of the accurate matching. It shows that the system can be more accurate to locate various elements, but the boundary identification is not accurate enough. The reason is mainly from the accumulation of errors in the Natural Language Processing tools at the bottom, including word segmentation, part of speech tagging and syntactic analysis tools. Therefore, the improvement of low-level language processing technology is of great significance for improving the accuracy of information extraction.

Performance analysis of sentences with multiple comparative relations A comparative sentence can contain one or more comparative relations. In the car corpus, 25.4% of sentences contains more than one comparative relation. Therefore, It is necessary to analyze the element extraction performance of these sentences. The experimental results are shown in Table 7.

Table 7. The performance of element extraction in multi-relation sentences(%)

Element	Accurate Evaluation			Coverage Evaluation		
	Precision	Recall	F1-score	Precision	Recall	F1-score
SUB	65.97	43.75	52.61	70.93	44.01	54.32
OBJ	77.62	85.28	81.27	79.42	86.35	82.74
ATTR	76.73	58.54	66.41	78.38	58.90	67.26
OPIN	96.18	64.09	76.92	96.50	65.12	77.76

Table 7 shows that recall and F1-score of accurate evaluation are significantly improved in sentences with multiple comparisons. F1-score of coverage evaluation decrease significantly.

5. Conclusions and Future Work

This paper studies the problem of comparative element extraction in the comparative sentences. Conditional random fields model is employed to extract comparative elements, which fuses various lexical, syntactic and heuristic features. A comparative element extraction model is constructed by using the supervised method. The performance indices of the element extraction are improved. The experiment results show that the shallow syntactic features can effectively identify the phrase-level comparative elements. The comparative keyword candidate features can not only compensate for the lack of comparative words in the training samples, but also make a preliminary locating of other elements. Heuristic position features are helpful to distinguish between elements such as subject entities and object entities. All the features introduced in the model are domain-independent, so the method can be applied directly to other areas. In the future, we plan to find more effective features that represent a sentence to further improve the recall of our system. We also plan to summarize extracted information into an opinion summarization.

References

1. Ding, X., Liu, B., Zhang, L.: Entity discovery and assignment for opinion mining applications. In: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 1125–1134. ACM, Paris, France (2009)

2. Dong, L., Wei, F., Tan, C., Tang, D.: Adaptive recursive neural network for target-dependent twitter sentiment classification. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. pp. 49–54. Association for Computational Linguistics, Baltimore, Maryland, USA (2014)
3. Feldman, R., Fresko, M., Goldenberg, J., Netzer, O., Ungar, L.: Extracting product comparisons from discussion boards. In: Proceedings of the Seventh IEEE International Conference on Data Mining. pp. 469–474. IEEE, Omaha, Nebraska, USA. (2007)
4. Ganapathibhotla, M., Liu, B.: Mining opinions in comparative sentences. In: Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1. pp. 241–248. Association for Computational Linguistics, Manchester, UK (2008)
5. He, H., Li, Z., Yao, C., Zhang, W.: Sentiment classification technology based on markov logic networks. New Review of Hypermedia and Multimedia 22(3), 243–256 (2016)
6. Hu, M., Liu, B.: Mining opinion features in customer reviews. In: Proceedings of the Nineteenth National Conference on Artificial Intelligence. pp. 755–760. Association for the Advancement of Artificial Intelligence, San Jose, California (2004)
7. Huang, G.H., Yao, T.F., Liu, Q.: Mining chinese comparative sentences and relations based on crf algorithm. Application Research of Computers 27(6), 2061–2064 (2010)
8. Jindal, N., Liu, B.: Mining comparative sentences and relations. In: Proceedings of the Twenty-First National Conference on Artificial Intelligence. pp. 1331–1336. Association for the Advancement of Artificial Intelligence, Boston, Massachusetts (2006)
9. Kiritchenko, S., Zhu, X., Cherry, C., Mohammad, S.: Nrc-canada-2014: Detecting aspects and sentiment in customer reviews. In: Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). pp. 437–442. COLING, Dublin, Ireland (2014)
10. Lafferty, J., McCallum, A., Pereira, F.C.N.: Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: Proceedings of the 18th International Conference on Machine Learning. pp. 282–289. ACM, Williamstown, MA, USA (2001)
11. Li, J.J.: Research on the identification of comparative sentences and relations and its application. Master's thesis, Chongqing University, China (2010)
12. Li, S., Lin, C.Y., Song, Y.I., Li, Z.: Comparable entity mining from comparative questions. IEEE Transactions on Knowledge and Data Engineering 25(7), 1498–1509 (2013)
13. Liu, K., Wang, S., Liao, X., Xu, H.: Overview of chinese opinion analysis evaluation. In: Proceedings of the 4st Chinese Opinion Analysis Evaluation. pp. 1–32. Chinese Informaiton Processing Society of China, Nanchang, Jiangxi, China (2012)
14. Qiu, G., Liu, B., Bu, J., Chen, C.: Opinion word expansion and target extraction through double propagation. Computational Linguistics 37(1), 9–27 (2011)
15. Shi, L., Li, S., Jiang, P., Liu, H.: Improving comparative sentence extraction of chinese product reviews by sentiment analysis. Journal of Engineering Science & Technology Review 9(6), 149–156 (2016)
16. Tang, D., Qin, B., Liu, T.: Aspect level sentiment classification with deep memory network. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. pp. 214–224. Association for Computational Linguistics, Austin, Texas, USA (2016)
17. Wang, S., Li, H., Song, X.: Automatic semantic role labeling for chinese comparative sentences based on hybrid patterns. In: Proceedings of the 2010 International Conference on Artificial Intelligence and Computational Intelligence. pp. 378–382. IEEE, Sanya, China (2010)
18. Wang, W., Pan, S.J., Dahlmeier, D., Xiao, X.: Recursive neural conditional random fields for aspect-based sentiment analysis. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. pp. 616–626. Association for Computational Linguistics, Austin, Texas, USA (2016)
19. Wang, Y., Huang, M., Zhu, X., Zhao, L.: Attention-based lstm for aspect-level sentiment classification. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. pp. 606–615. Association for Computational Linguistics, Austin, Texas, USA (2016)

20. Xing, L., Liu, L.: Chinese standard comparative sentence recognition and extraction research. In: Proceedings of the International Conference on Information Engineering and Applications. pp. 415–422. Springer, London, Chongqing, China (2013)
21. Xu, K., Liao, S.S., Li, J., Song, Y.: Mining comparative opinions from customer reviews for competitive intelligence. Decision Support Systems 50(4), 743–754 (2011)
22. Yang, S., KoJindal, Y.: Extracting comparative entities and predicates from texts using comparative type classification. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics. pp. 1636–1644. Association for Computational Linguistics, Portland, Oregon (2011)
23. Yin, Y., Wei, F., Dong, L., Xu, K., Zhang, M.: Unsupervised word and dependency path embeddings for aspect term extraction. In: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence. pp. 2979–2985. Association for the Advancement of Artificial Intelligence, New York, USA (2016)
24. Zhang, L., Liu, B., Lim, S.H., O' Brien-Strain, E.: Extracting and ranking product features in opinion documents. In: Proceedings of the 23rd International Conference on Computational Linguistics. pp. 1462–1470. Association for Computational Linguistics, Uppsala, Sweden (2010)
25. Zhang, W., He, H., Cao, B.: Identifying and evaluating the internet opinion leader community based on k-clique clustering. Neural Computing and Applications 25(3-4), 595–602 (2014)

Wei Wang lecturer at Harbin Institute of Technology, her research interests cover natural language processing, sentiment analysis and information safety.

Guodong Xin lecturer at Harbin Institute of Technology, his research interests cover social networks and network security.

Bailing Wang professor at Harbin Institute of Technology, his research interests cover network security, computer network and social networks.

Junheng Huang associate professor at Harbin Institute of Technology, his research interests cover social networks, natural language processing and network security.

Yang Liu associate professor at Harbin Institute of Technology, his research interests cover network security, computer network and social networks.

Received: December 29, 2016; Accepted: August 25, 2017.