The Constitution of a Fine-Grained Opinion Annotated Corpus on Weibo

Liao Jian[1], Li Yang[1], and Wang Suge[1,2]

[1] School of Computer & Information Technology, Shanxi University, Taiyuan, 030006, China
[2] Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, Taiyuan, 030006, China

Abstract. Sentiment analysis on social media represented by Weibo is one of the hotspot research problems in NLP. A comprehensive and systematic fine-grained annotated corpus plays a significance role. In this paper, considering the characteristics of Weibo, we focus on the constitution of a fine-grained, hierarchical opinion annotated corpus and design a set of labelling specification. We manually annotate the opinion sentences with a part of ones containing hidden opinion which can be useful for implicit sentiment analysis. Then a fine-grained aspect extraction, namely opinion triples like <object, attribute, polarity> is finished for aspect-level sentiment research. Moreover, we establish an evaluation method for the task of fine-grained aspect extraction which has been applied in evaluation for years. The corpus was used in the task of COAE2015, and it will be a useful resource for the related research on social media sentiment analysis.

Keywords: weibo corpus, fine-grained opinion annotation, implicit opinion annotation, evaluation

1 Introduction

Weibo is one of the most popular social media which enjoys a rapid development in China. People can freely express their opinion, attitude and emotion with it. According to a report released by China Internet Network Information Center(CNNIC)[1], the quantity of Weibo users totaled 204 million and nearly half of the weibo's updates are sent via mobile phone. Recent trends in Weibo's development have led to a proliferation of studies on social media analysis. A well-designed, fine-grained and large-scale annotated corpus can be benefit to lots of research and applications of social media sentiment analysis.

Up to now, many works focused in building sentiment annotated corpus and have made some progress. Based on Twitter data, Pak et al.[8] established a subjective corpus containing positive and negative sentiment text. Ptaszynski et al.[9] builded a corpus based on a large-scale Japanese blogs, they labeled both sentiment and emotion including sentiment polarity and emotion icon. In Jacob's study[4], a CRF-based model which can effectively reduce human workload was adopted to extract expressions of opinion object to build an object dictionary in interdisciplinary datasets.

Several achievement have been done in establishing Chinese sentiment annotated corpus. Xu et al.[13] labeled nearly 40,000 sentences with sentiment polarity based on Chinese textbooks for primary schools, screenplays and literature periodicals. After analysing characteristics of the annotated corpus with million-words level, some statistical results like the distribution of sentiment and its transfer pattern was given in their study. Based on this research, they designed an framework for product reviews labelling, and annotated the attributions of product and their corresponding opinion expressions in word-level, sentence-level and document-level respectively[14]. For entity-level labelling, Deng et al.[3] have published an entity-level corpus named MPQA3.0, they annotated entities and event targets on the basis of previous work. Dai et al.[2] annotated 5,000 reviews of product with opinion object and sentiment polarity, they also labelled some implicit sentiment, default object and polarity transfer. But on the one hand their corpus suffered from the lack of hierarchy relationship between opinion objects and attributes, on the other hand the product reviews are not much diversity comparing to Weibo data. Recent developments of Weibo have heightened the need for social media oriented annotated corpus. Yao et al.[15] established a emotion annotated corpus contains 14,000 pieces of weibo and 45,431 sentences. They labelled the sentences with 7 kinds of emotion class and 3 kinds of emotion strength.

In this paper, we design a framework for fine-grained, hierarchical, social media oriented annotation and its corresponding specifications. Compared to other resource construction method, our framework can effectively reduce the manual work by using the classifier and automatic aspect extraction method as the initial labelling and assistance. Besides, a random cross-validation process is introduced to minimize the difference of human cognition. Using the framework, we complete a sentence-level opinion corpus's annotation which contains 15,679 pieces of weibo, 20,154 sentences, and a hierarchical aspect-level opinion corpus with 13,787 sentences, 24,093 pairs of triples annotated. We also annotate a part of sentences containing implicit opinions. Moreover, we present a set of evaluation methods for the task of fine-grained opinion aspect extraction. These methods have been used in the fifth to seventh Chinese Opinion Analysis Evaluation[11, 12, 6].

The rest of the paper is structured as follows: We will introduce the specifications in the next section. The main process of annotation will be described in section 3 and a detail analysis of the corpus will be given in section 4. We then present the method and indexes of evaluation in section 5. The sixth part shows the result and analysis of Chinese Opinion Analysis Evaluation 2015 which used the corpus as task dataset. Conclusion and future work are presented in the last section.

2 Specification of annotation

2.1 Granularity of annotation

The authors of Weibo express their opinions or emotions with brief words. These sentences usually contain dense opinions or emotions but suffered from ill-formed expression. Therefore, a large-scale corpus with detailed, fine-grained annotation plays a important role for social media oriented research. In this paper, we choose sentence-level as the basic granularity of annotation, then a hierarchical aspect-level annotation of triples<object, attribute, polarity> will be done on the set of sentences which contain opinions

2.2 Annotation of implicit opinion

The expressions of opinion in Weibo text are consist of explicit and implicit ones. Explicit opinion refer to the phrase or sentence express sentiment polarity with sentiment words which can be effectively identified through a sentiment dictionary, while the implicit ones are made up without sentiment words. Liu[7] defined the implicit opinion as a kind of objective statement contains the general opinion or comparative one which express a subjective feeling of satisfy or unsatisfy. Usually, there are two ways for expressing implicit opinion, namely factual description based and rhetorical description based. For example.

1. 今天是第五天了，还在维修中。
(Today is the fifth day, it's still in maintenance.)
2. 没完没了的会议引发了一场暴风雨。
(The endless meeting sparked a storm.)

The first sentence expresses a negative opinion by using a factual statement while the second one is metaphorical.

According to our previous survey on the dataset containing nearly 20,000 sentences which collected from a famous car reviews website named Autohome⋆ in China, nearly one third of the sentences contain implicit expression. Implicit opinions are easier to detect and to classify than explicit ones for the lack of sentiment words, and most of the current research has focused on explicit opinions[7]. For that reason, we add the labelling of implicit opinion into our corpus for related research in the future.

2.3 Classification of annotated opinion

In this paper, we focus in annotating the sentiment polarity of sentences and triples, and the explicit/implicit class of opinion sentence. We define the classification of annotation as the following table 1.

⋆ http://www.autohome.com.cn

Table 1: The classification of annotated opinion

Opinion Class	Sentiment Polarity	Annotated Label
Explicit	positive	1
	negative	2
	mixed	3
Implicit	positive	4
	negative	5
	mixed	6

3 Construction of sentiment corpus

3.1 Selection of original data and preprocessing

We use a dataset with 10 million weibos which was crawled from the largest and most popular social media named Weibo in China as the original data. This dataset was also used as task data for new sentiment word identification in COAE2014[12]. Comparing to other source, the Weibo dataset is domain independent which covers a wide coverage such as digital product, cars review, food, travel, entertainment and so on. Moreover, a large number of weibos' contents are suffered from arbitrary and ill-formed expression. Therefore, cleaning and preprocessing of the original data need to be finished firstly to reduce the noise. It is consist of two steps.

Filtering. Remove all the weibos which are unsuitable for research. We use a rule-based filter to finish this task. The rules mainly includes, 1)weibos with the length less than 10 Chinese characters; 2)all emoticons or punctuation composed ones; 3)the ones containing mobile phone number or QQ(a famous IM software in China) number, because this kind of weibos are mostly advertisement; 4)completely duplicate ones.

Sentence segmenting. We use the sentence-level as our basic granularity of annotation. For that reason, we segment all the weibos into sentences by using the following rules. 1)split the weibo by the segment characters set S = {".", "。", "!", "?", ":", "..."}; 2)if the period(。) or dot(.) is between two numbers or characters, skip the segment process for it is usually means a number or abbreviation; 3)preserve the first sentence only if it follows with repeating segment characters in S.

3.2 Process of annotation

The main process of annotation in this paper is consist of two stages. Firstly, in the sentence-level opinion annotate stage, we use a dictionary-based opinion sentence identification method proposed in Song's research[10] to select candidate opinion sentences, then send the filtered result to annotators for manual annotation and cross validation. Then we adopt the method proposed by Liao[5] to extract opinion aspect pairs on the selected opinion sentences in stage 1. This pairs can be used as a initial labelling and reference for annotators in aspect-level triplet annotate stage, and a cross validation for consistency is completed after that. The framework of annotate process is shown as the following figure 1.

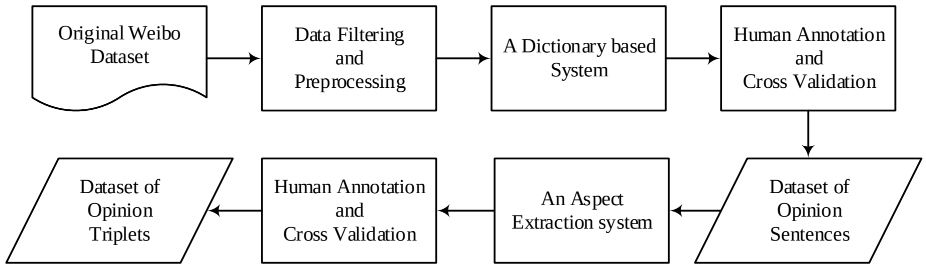

Fig. 1: The process of annotation

Annotation of opinion sentence In this stage, we distinguish the opinion sentences from the dataset and annotate sentiment polarity label. Compared to previous work, we add the annotation of implicit opinion. The basic rules for labelling are defined as, if the sentence which contains opinion has explicit sentiment words, we consider it as a explicit opinion sentence and label the corresponding polarity. While a sentence will be regarded as implicit opinion one if the containing opinion is expressed by factual description or rhetorical description like metaphor and personification.

For example, “坑爹的苹果，坑爹的ipad4。”(Ceezy Apple, ceezy ipad4.) will be marked as a explicit negative sentence for it has an explicit sentiment word “坑爹”(Ceezy). While the sentence “你脑壳遭门夹了才会在太升南路买iphone！”(How stupid are you buying an iPhone on south Taisheng road.) expresses a negative opinion using a metaphor “脑壳遭门夹了”(stupid like the head is crashed by a door).

Annotation of opinion triplet In the fine-grained opinion aspect triplet annotate stage, we try to extract all the opinion triplets<object, attribute, polarity> in the opinion sentences. During the annotation, we define a set of rules in order to achieve a clear, complete, high-consistency result. The rules are defined as follows.

Apposition. Multi objects, attributes or opinion expressions concatenate each other with hyphenation “，” or “、” will be regarded as a whole aspect. While the aspects concatenated by conjunction “和” or “与”(both means “and” in Chinese) are divided into different triplets.

Example 1. “苹果，三星能赚取全球手机市场99% 的利润，国产手机无底线的降价，引来的必然是恶性的市场竞争，更别谈质量、创新了。”(Apple, Samsung earned 99% profits of the global mobile phone market. The price reducing of domestic cellphone without limit will result in cut-throat competition inevitably, let alone quality, innovation.)

Example 2. “我从2002年底用摩托罗拉A系列一直用到2010 年10 月，原因在于A系列各代的易用性和继承性特别好!”(I used the MOTOROLA series A from the end of 2002 to October 2010 for the reason that each generation of series A enjoyed an excellent usability and continuity.)

Comparative. If there is a comparative opinion in sentence, we label the triplets in pairs and the polarity of them depend on the result of comparison.

Example 3. “今天又试了试新帕萨特，感觉比迈腾好，至少动力系统和降噪方面比迈腾强。”(I tried the new PASSAT today, feel better than MAGOTAN, at least the power system and noise reduction are better than its.)

Attributive. If there are more than one attributive word modify the core aspect, we preserve the longest range.

Example 4. “本田newCIVIC同大众Scirocco有着多么相似的出色风格！”(How similar the excellent style are between the Honda newCIVIC and Volkswagen Scirocco.)

Coreference. The demonstrative pronoun will be preserved for the disambiguation of targets.

Example 5. “#新车播报#日前，沃尔沃2013款s60 正式上市销售，这个车是在美国测试中秒杀一众德日豪华车。”(#New Cars Boardcasting#Now, the Volvo s60 2013 is available, this car achieved the best in a evaluation in US and defeat other German and Japanese premium cars overwhelmingly.)

The solutions for example 1-5 are shown as the following table 2. “null” means the aspect is default in current sentence.

Consistency of annotation The annotation of corpus needs collaboratively work. The result of labelling of different annotator is unavoidable different. We design the following process try to reduce the inconsistency.

Mark the unlabelled sentences set as S_u, labelled sentences set as S_l, annotators set as H, the annotator's labelling can be marked as $L(s,h)$, $s \in S$, $h \in H$, $L(\cdot) = \{1,2,3,4,5,6\}$.

Stage of opinion sentence annotation.

(1) Divide the unlabelled sentences set S_u into k groups, for each instance s_{ki} in S_{uk}, an annotator labels the sentiment polarity. After that, we get an initial labelled set $S_l = \{S_{l1}, S_{l2}, \ldots, S_{lk}\}$, in which $S_{lk} = \{L(s_{ki},h) | s_{ki} \in S_{uk}, h \in H\}$.

(2) For each subset S_{lk} in S_l, we send it to another annotator randomly for relabelling. Then a new labelled sentence set $S'_l = \{S'_{l1}, S'_{l2}, \ldots, S'_{lk}\}$ are produced. Select the different set $S_{dk} = \{s_{ki} | L(s_{ki},h) \neq L(s_{ki},h')\}$ of S_{lk} and S'_{lk}. Integrate the subset by $S_l = S_l \cup (S_{lk} \cap S'_{lk})$ and $S_d = S_d \cup S_{dk}$.

(3) Redo step 2 in S_d and get the different set S'_d, the instance in S'_d will be abandoned for it is hardly to get an agreement. Then add the rest subset to labelled set by $S_l = S_l \cup (S_d/S'_d)$.

Stage of fine-grained aspect annotation.

(1) Use the labelled sentence set S_l in previous stage as the initialization of unlabelled set S_u in current stage. Similarly, divide S_u into k groups, and get the

Table 2: The examples of annotation

Rule	Example	Opinion Object	Opinion Attribute	Polarity
Apposition	Example 1	国产手机 (domestic cellphone)	质量、创新 (quality, innovation)	-1
	Example 2	A系列 (series A)	易用性 (usability)	-1
		A 系列 (series A)	继承性 (continuity)	-1
Comparative	Example 3	新帕萨特 (new PASSAT)	null	1
		迈腾 (MAGOTAN)	null	-1
		新帕萨特 (new PASSAT)	动力系统 (power system)	1
		迈腾 (MAGOTAN)	动力系统 (power system)	-1
		新帕萨特 (new PASSAT)	降噪 (noise reduction)	1
		迈腾 (MAGOTAN)	降噪 (noise reduction)	-1
Comparative	Example 4	本田newCIVIC (Honda newCIVIC)	风格 (style)	1
		大众Scirocco (Volkswagen Scirocco)	风格 (style)	1
Coreference	Example 5	这个车 (this car)	null	1

initial labelled set $S_l = \{S_{l1}, S_{l2}, \ldots, S_{lk}\}$, in which $S_{lk} = \{triple(s_{ki}, h) | s_{ki} \in S_{uk}, h \in H\}$, $triple(s_{ki}, h)$ means the opinion aspect triple <object, attribute, polarity>.

(2) Send each subset S_{lk} in S_l to another annotator randomly for proofreading. Then Integrate the new validated set $S'_l = \{S'_{l1}, S'_{l2}, \ldots, S'_{lk}\}$ into S_{lk} by $S_l = S_l \cup S'_{lk}$.

4 Detail analysis of the corpus

We have completed a corpus with 15,679 weibos, 20,154 sentences which covers a wide range such as automobile, electronics, mobile phone, food, entertainment and so on. Among the opinion sentences, we annotate 13,787 sentences with 24,093 pairs of opinion aspect triplets. A detail analysis like sentiment distribution, explicit/implicit proportion, lexical diversity will be shown as follows.

4.1 Sentiment distribution of opinion in sentences and aspects

The statistical result of sentiment distribution of opinion in sentences and aspect triplets is shown as the following table 3.

Table 3: Sentiment distribution of opinion in sentences and aspects

Statistical Item	**Amount/Proportion**	**Negative**	**Positive**	**Mixed**	**Total**
Total Opinion Sentences	Amount	12306	6209	1639	20154
	Proportion	0.611	0.308	0.081	1
Explicit Opinion Sentences	Amount	11610	5139	1236	17985
	Proportion	0.646	0.286	0.068	1
Implicit Opinion Sentences	Amount	696	1070	403	2169
	Proportion	0.321	0.493	0.186	1
Opinion Aspect Triplets	Amount	16136	7602	355	24093
	Proportion	0.670	0.316	0.015	1

We can easily figure it out that the majority opinion in Weibo contents is positive from table 3. However, users prefer to express negative or mixed opinion through implicit expression. This phenomenon is in accordance with the general cognitive pattern that people will speak highly of others directly, while a more euphemistic tone will be adopted when delivering criticism or alternative viewpoints. In consideration of the proportion of opinion aspect triplets, comparing to sentence-level opinion, triplets express more certain sentiment polarity with less mixed opinion ones, for the triplet enjoyed a finer granularity. This can be somehow a support for the idea that aspect-level fine-grained opinion expressions can accurately portray the sentiment information of a document.

4.2 Analysis of lexical diversity in fine-grained opinions

The lexical diversity can effectively describe the coverage and generalization of the corpus in fine-grained aspect annotation. The analysis of lexical diversity is shown as table 4.

Table 4: Lexical diversity in fine-grained opinions

Statistical Item	**Total Opinion**		**Explicit Opinion**		**Implicit Opinion**	
	Object	**Attribute**	**Object**	**Attribute**	**Object**	**Attribute**
Amount	5437	3549	4598	3306	803	468
AWF	4.43	6.79	4.22	5.87	2.70	4.63

AWF means the average word frequency. From table 4 we know that the lexical diversity of implicit is more abundant then explicit for the reason that, 1)it has a less average word frequency($2.70 < 4.22$); 2)each opinion attribute in implicit triplets corresponds 1.71(803/468) objects while the ratio is 1.39(4598/3306) in explicit ones.

4.3 Analysis of consistency

Consistency of opinion sentence annotation. We introduce the κ-coefficient to measure the consistency of opinion sentence annotation. For the labelling of

11 annotators, we calculate the κ value on total, explicit, implicit opinion respectively. The result and analysis is given as follows, see table 5.

Table 5: κ-coefficient of opinion sentence annotation

Measure Index	**Total Opinion**	**Explicit Opinion**	**Implicit Opinion**
κ-coefficient	0.924	0.961	0.687

The κ-coefficient of opinion annotation is satisfied in general but it suffered a low performance in implicit ones. The reason is that there is no clear definition of implicit expression in linguistics, we only set a few rough rules to restrict it to a wide range, so the annotators have to identify the ambiguous concept from their own cognition.

Consistency of fine-grained aspect triplets annotation. To measure the consistency of this stage, We use the evaluation method mentioned in section 5, regarding the initial labelled result as testing set S_t and the second validated result as key set S_k. Calculate the index value of fully match and fuzzy match respectively. The result is shown as following table 6. The "all fully/fuzzy" means

Table 6: Consistency of fine-grained aspect triplets annotation

Match Type	**Mirco**			**Marco**		
	Precision	**Recall**	**F1**	**Precision**	**Recall**	**F1**
All Fully	0.776	0.588	0.669	0.814	0.758	0.785
All Fuzzy	0.815	0.622	0.706	0.866	0.807	0.835
Object Fully	0.825	0.628	0.713	0.870	0.811	0.839
Object Fuzzy	0.858	0.655	0.743	0.911	0.849	0.879
Attribute Fully	0.827	0.630	0.715	0.876	0.816	0.845
Attribute Fuzzy	0.845	0.645	0.731	0.898	0.837	0.867

the opinion object and attribute of instance in testing set must fully/fuzzy match the corresponding item of key instance at the same time and their sentiment polarity is the same. Namely, O_t matches O_k, A_t matches A_k and $P_t = P_k$, $< O_t, A_t, P_t >\in S_t$, $< O_k, A_k, P_k >\in S_k$. Similarly, "object fully/fuzzy" means O_t matches O_k and $P_t = P_k$. "Attribute fully/fuzzy" means A_t matches A_k and $P_t = P_k$.

From table 6 we can see that the consistency of fine-grained aspect triplets annotation is hardly satisfied, the main problem is the different cognition on the hierarchical relation between objects and attributes, the boundary of aspects and so on.

5 Evaluation method for fine-grained opinion extraction

To evaluate the performance of fine-grained opinion extraction system, we design two match patterns, namely fully match and fuzzy match on three evaluate target as opinion object-polarity, opinion attribute-polarity, object-attribute-polarity using the indexes of precision, recall and F1. Denote the extracted aspect in testing set S_t as x, the human labelled aspect in key set S_k as y, the corresponding sentiment polarity as P_t, P_k respectively, some definition is shown as follows.

Def.1, Fully match. If $(x \subseteq y) \wedge (y \subseteq x) \wedge (P_t = P_k)$ is true, then x fully matches y.

Def.2, Fuzzy match. If $(x \subseteq y) \vee (y \subseteq x) \wedge (P_t = P_k)$ is true, then x fuzzy0 matches y.

For example, suppose the key aspect is "屏幕分辨率"(screen resolution) while the testing aspect is "屏幕"(screen) or "分辨率"(resolution), it is not considered as fuzzy match but a fully one.

Def.3, Coverage match. If the coverage value of the overlap between x and y greater than threshold, then x matches y. Coverage is defined as,

$$Coverage(x, y) = \frac{len(x \cap y)}{len(y)} \tag{1}$$

$len(\cdot)$ means the length of the input string, $x \cap y$ is the overlap of x and y. The threshold of coverage is usually set as 0.2, 0.5 or 0.8**.

For example, suppose the key aspect is "屏幕分辨率"(screen resolution) while the testing aspect is "屏幕"(screen) and current threshold is 0.2, then the coverage value of this instance is 0.4 and it is regarded as a valid match.

6 Corpus for Chinese opinion analysis evaluation

Part of the corpus established in this paper was used as the task data for social media oriented opinion analysis, including opinion sentence identification and opinion aspects extraction in the seventh Chinese opinion analysis evaluation(COAE2015). Unlike previous evaluation, there are two type of evaluation on each task this year, namely resource limited and resource unlimited. The former focus on the proposed methods and models by limit the using of unofficial resources, while the focal point of the latter is the performance of each participating system.

6.1 Result of opinion sentence identification and classification

There are 19 teams with 44 results submitted in this task. 13 teams with 18 submissions are resource limited while the number of unlimited is 15 and 26. We list the best and medium result of all the submissions in following table 7.

** The coverage match is equivalent to fully match when the coverage is 1.

Table 7: Result of opinion sentence identification and classification

Resource	**System**	**Mirco**			**Marco**		
		Precision	**Recall**	**F1**	**Precision**	**Recall**	**F1**
Limited	Best	0.821	0.733	0.771	0.830	0.629	0.654
	Medium	0.616	0.539	0.582	0.556	0.438	0.471
Unlimited	Best	0.840	0.785	0.811	0.878	0.667	0.695
	Medium	0.615	0.564	0.606	0.555	0.477	0.495

All listed data above are the best/medium result of **each single index** among the submissions. From table 7 we can know that, 1)on the task of Chinese opinion sentence identification, most teams achieve a acceptable performance but still need to improve on mixed opinion sentences. This kind of data usually contains both positive and negative opinions which is hard to be recognized correctly. 2)Compared with limited resource evaluation, there is only a 4% improvement in average of the unlimited resource. In some sense, it demonstrates that only adding extra resource will hardly work when dealing with social media text.

6.2 Result of opinion aspects extraction

The number of participating teams in this task is much less than the previous one, however the aspect extraction is harder and more meaningful for a fine-grained opinion analysis will accurately portray short text like the social media data. There are 10 teams with 20 results submitted and 5 teams with 6 submissions are resource limited. 15 unlimited results were submitted by 9 teams. Results can be seen in the following table 8.

From table 8 and some analysis on submissions, we can see that,

(1) Compared to previous result on opinion aspect extraction task, the performance of this year decreased 10% in average, using Weibo data which is domain independent instead of some product BBS may be the main reason.

(2) Similar as the task of opinion sentence identification, the performance on mixed opinions needs improving. However, the result is still acceptable for there is only 1.47% of the triplets containing mixed opinion in aspect-level granularity.

(3) Analysis on all the submissions shows that 20.1% of the total submitted answers reached an agreement by more than three teams. This phenomenon demonstrates that there exists great difference among the teams.

(4) Like the previous task, extra resource provide only 5% improvement in average. For social media oriented data, we shall lie our focal point on models and representation.

7 Conclusion and future work

In this paper, we design a framework for fine-grained, hierarchical, social media oriented labelling and its corresponding specifications. We complete a sentence-level opinion corpus's annotation which contains 15,679 pieces of weibo, 20,154

Table 8: Result of opinion sentence identification and classification

Resource	Match Type	System	Mirco			Marco		
			Precision	**Recall**	**F1**	**Precision**	**Recall**	**F1**
Limited	All Fuzzy	Best	0.092	0.244	0.130	0.062	0.197	0.086
		Medium	0.054	0.184	0.073	0.044	0.138	0.056
	All Fully	Best	0.068	0.179	0.097	0.048	0.162	0.066
		Medium	0.039	0.139	0.053	0.032	0.104	0.041
	Object Fuzzy	Best	0.119	0.316	0.168	0.080	0.257	0.112
		Medium	0.074	0.250	0.100	0.061	0.187	0.078
	Object Fully	Best	0.088	0.236	0.125	0.062	0.211	0.085
		Medium	0.058	0.192	0.077	0.046	0.137	0.060
	Attribute Fuzzy	Best	0.125	0.323	0.176	0.084	0.295	0.117
		Medium	0.075	0.293	0.103	0.060	0.201	0.080
	Attribute Fully	Best	0.116	0.298	0.164	0.078	0.270	0.109
		Medium	0.066	0.265	0.092	0.052	0.189	0.073
Unlimited	All fuzzy	Best	0.131	0.288	0.177	0.097	0.222	0.134
		Medium	0.081	0.222	0.108	0.051	0.158	0.073
	All Fully	Best	0.112	0.232	0.151	0.083	0.190	0.115
		Medium	0.065	0.165	0.081	0.041	0.120	0.058
	Object Fuzzy	Best	0.166	0.368	0.223	0.124	0.309	0.169
		Medium	0.125	0.285	0.138	0.078	0.217	0.098
	Object Fully	Best	0.144	0.298	0.194	0.106	0.243	0.147
		Medium	0.089	0.224	0.103	0.061	0.160	0.073
	Attribute Fuzzy	Best	0.159	0.426	0.213	0.114	0.352	0.159
		Medium	0.115	0.292	0.159	0.072	0.213	0.103
	Attribute Fully	Best	0.151	0.399	0.203	0.109	0.325	0.152
		Medium	0.103	0.276	0.143	0.068	0.200	0.098

sentences including explicit or implicit opinions. In addition, a hierarchical aspect-level opinion corpus with 13,787 sentences, 24,093 pairs of triples are established for aspect-level sentiment analysis. A dataset based on our labelled corpus was used as a task dataset in COAE2015 and effectively promote the development of related research on social media sentiment analysis. This dataset is now available on the website(http://115.24.12.5/web/resource.html) for academia. In future work, we will build a larger and more detailed corpus for implicit sentiment analysis.

Acknowledgement

The authors would like to thank all the students' hard work who participate the corpus's labelling including Zhao Celi, Zhang Jin, Xu Chaoyi, Guo Xiaomin, Zhang Jun, Li Min, Qiao Pei, Mu Wanqing, Wang Jia, Wang Jie and Lv Ying. Also thank all anonymous reviewers for their valuable comments and suggestions which have significantly improved the quality and presentation of this paper. This work was supported by the National High-Tech Research and Development Program (863 Program) (2015AA011808); the National Natural Science Foundation of China (61432011, 61573231, 61175067, 61272095, U1435212); the Shanxi Province Returned Overseas Research Project (2013-014); the Shanxi Province Science and Technology Basic Condition Platform Construction (2015091001-0102).

References

1. CNNIC: Statistical report on internet development in China (2015)
2. Dai, M., Zhu, Z., Li, S., Zhou, G.: Corpus construction on opinion information extraction in chinese. Journal of Chinese Information Processing 29(4), 67 (2015)
3. Deng, L., Wiebe, J.: Mpqa 3.0: An entity/event-level sentiment corpus. In: proceedings of the Conference of the North American Chapter of the Association of Computational Linguistics: Human Language Technologies (2015)
4. Jakob, N., Gurevych, I.: Extracting opinion targets in a single-and cross-domain setting with conditional random fields. In: proceedings of Empirical Methods in Natural Language Processing. pp. 1035–1045. Association for Computational Linguistics (2010)
5. Liao, J., Wang, S., Li, D., Zhang, P.: The bag-of-opinions method for car review sentiment polarity classification. Journal of Chinese Information Processing 29(3), 113 (2015)
6. Liao, X., Wang, s., Huang, M.: Overview of chinese opinion analysis evaluation 2015. In: proceedings of Chinese Opinion Analysis Evaluation 2015, Luoyang, China. pp. 5–26 (2015)
7. Liu, B.: Sentiment analysis and subjectivity. Handbook of natural language processing 2, 627–666 (2010)
8. Pak, A., Paroubek, P.: Twitter as a corpus for sentiment analysis and opinion mining. In: proceedings of the International Conference on Language Resources and Evaluation. vol. 10, pp. 1320–1326 (2010)
9. Ptaszynski, M., Rzepka, R., Araki, K., Momouchi, Y.: Automatically annotating a five-billion-word corpus of japanese blogs for sentiment and affect analysis. Computer Speech & Language 28(1), 38–55 (2014)
10. Song, Y.: Design and Implementation of a Rule-based Comparative Opinion Mining System. Master's thesis, Shanxi University, Taiyuan, China (2014)
11. Tan, S., Wang, s., Liao, X., Li, W.: Overview of chinese opinion analysis evaluation 2013. In: proceedings of Chinese Opinion Analysis Evaluation 2013, Taiyuan, China. pp. 5–33 (2013)
12. Tan, S., Wang, s., Xu, W., Yan, X., Liao, X.: Overview of chinese opinion analysis evaluation 2014. In: proceedings of Chinese Opinion Analysis Evaluation 2014, Kunming, China. pp. 4–25 (2014)
13. Xu, L., Lin, H., Zhao, J.: Construction and analysis of emotional corpus. Journal of Chinese Information Processing 22(1), 116–122 (2008)
14. Xu, R., Xia, Y., Wong, K.F., Li, W.: Opinion annotation in on-line chinese product reviews. In: proceedings of the International Conference on Language Resources and Evaluation. vol. 8, pp. 26–30 (2008)
15. Yao, Y., Wang, S., Xu, R., Liu, B., Gui, L., Lu, Q., Wang, X.: The construction of an emotion annotated corpus on microblog text. Journal of Chinese Information Processing 28(5), 83 (2014)