A system for fine-grained aspect-based sentiment analysis of Chinese

Janna Lipenkova
Anacode
`janna.lipenkova@anacode.de`

Abstract

This paper presents a pipeline for aspect-based sentiment analysis of Chinese texts in the automotive domain. The input to the pipeline is a string of Chinese characters; the output is a set of relationships between evaluations and their targets. The main goal is to demonstrate how knowledge about sentence structure can increase the precision, insight value and granularity of the output. We formulate the task of sentiment analysis in two steps, namely unit identification and relation extraction. In unit identification, we identify fairly well-delimited linguistic units which describe features, emotions and evaluations. In relation extraction, we discover the relations between evaluations and their "target" features.

1 Introduction

Whereas most work on sentiment analysis, and especially on less covered languages such as Chinese, is based on probabilistic models and the use of general sentiment lexica, we believe that a holistic approach should also take into account general linguistic knowledge. On the one hand, this allows to leverage the results of several decades of research in theoretical linguistics. On the other hand, the hard-coding of general principles of language structure allows us to create a linguistically adequate training space for further application of probabilistic models.

In the following, we present the "bottom-up" component of our sentiment system which builds opinion representations by a progression along three levels - the lexical, the phrasal and the sentence level. The system has been conceptualized manually and bootstrapped on a corpus of about 1 mio. automotive reviews with an average length of 135 Chinese characters.[1] We use a prebuilt lex-

icon with ca. 2000 entries which contains opinion words, their modifiers, car features as well as a large number of functional categories relevant for the syntactic analysis of phrases and sentences. The performance of the system is evaluated on a testset of 800 annotated sentences. In practice, the presented model is complemented by a probabilistic model which performs topic and polarity classification on the sentence and the document levels; this component will not be described below due to space limitations.

The basic assumption on which the model builds is that language follows rules. Many of these rules have been extensively studied in the linguistic literature and have been taken to a level of abstraction which allows for a straightforward encoding. Incorporating these rules spares us the construction of probabilistic models for the discovery of already established general knowledge about linguistic structure. For example, it has long been observed that Chinese phrase structure is largely head-final (Huang 1982, Li 1990, *i. a.*): nominal modifiers precede their head nouns, whereas degree and negation adverbs normally precede the adjectives or verbs they modify. Due to the relative rigidity of word order in Chinese on the phrasal level, a small set of corresponding phrase-level rules achieves a high coverage on our dataset. Rules do not perform as well on sentence level; nevertheless, some general observations are possible: for example, AP targets precede their APs. These high-level observations form the basis of a sequence classifier which determines whether a sequence of words between two syntactic phrases establishes or disrupts one of the target relations between these phrases.

The paper is structured as follows: after a very brief review of research on aspect-based sentiment analysis (henceforth ABSA), we formulate

---

[1]The reviews were crawled from popular automo-tive sites: `http://www.autohome.com.cn`, `http://auto.16888.com`, `http://auto.qq.com`.

our task and, specifically, present the output format of the system (Section 3). In the second step, we briefly describe the categories used in our lexical resources (Section 4). In the third step, we describe the three levels of processing (Section 5). Finally, we present the evaluation of our system (Section 6).

2 Previous work

ABSA has been exploited as a refined alternative to sentiment analysis on the sentence and the document level: whereas the latter targets the general sentiment or polarity of a piece of text, ABSA outputs a mapping from specific aspects of the discussed topic to their evaluations. Different ABSA approaches have been exploited; thus, Popescu and Etzioni (2005) and Kim and Hovy (2006) present unsupervised algorithms for extracting aspects and determining sentiment in review text. Ding et al. (2008) and Liu (2012) describe approaches based on rules of semantic composition and distance metrics for the identification of relations between aspects and their opinions. Due to the relatively fine granularity of the task, parsing-based approaches have been proposed to capture the aspect/sentiment relations based on sentence structure (Jiang et al. 2011, Boiy and Moens 2009, *i. a.*). Further, the SemEval-2014 task on ABSA (Pontiki et al., 2014) has been addressed with a number of promising approaches and also significantly contributed to a unified understanding of ABSA.

Still, most research is focussed on the English language; for Chinese, most approaches to sentiment analysis are targeted on lexicon construction (e. g. Liu et al. 2013) or sentence/document-level sentiment classification.[2] Only few contributions aim at a finer-grained analysis at the aspect level (Ku et al. (2009), Su et al. (2008)).

3 Task

The goal of aspect-based sentiment analysis is to derive the opinions of a speaker about an entity and its features (Liu, 2012, p. 58). In our framework, opinions can be subclassified into evaluations and emotions. Evaluations express how the author evaluates a specific feature (e. g. *good*, *expensive*), whereas emotions express how the au-

thor feels about a specific feature (e. g. *to please*, *angry*).

We formulate the task in two stages - the identification of syntactic units and the extraction of relations between the syntactic units. Thus, given an opinion statement on a specific product, we "translate" the statement into a set of ($feature, < evaluation|emotion >$) pairs in two processing steps:

1. Build three sets of syntactic units F (features), EV (evaluations) and EM (emotions). For convenience, we will use $E = EM \cup EV$ in cases where the evaluation/emotion distinction is not relevant.

2. For each $e \in E$, find whether it has an opinion target $f \in F$.

A word is in place about the semantic organization of evaluations and emotions in our system. It has long been observed that many evaluation words come with implicit features; for example, the evaluation *beautiful* implicitly contains the feature VisualAppearance. In order to preserve this meaning, we adopt a scalar representation of evaluations (cf. Kennedy and McNally (2005) for a linguistic analysis of scalar expressions): evaluations are represented as pairs of a feature and a numerical value which "maps" the evaluation to some point on the feature scale [-3, 3]. Thus, *beautiful* gets the representation (VisualAppearance, 2), whereas *ugly* gets the representation (VisualAppearance, -2). Similarly, emotions are also represented as pairs of the emotion concept and a numerical value representing the intensity of the emotion (e. g. *angry*: (Anger, 2)).

The final mapping goes from sequences of features to numerical evaluations. In a feature sequence $[f_1, f_2 \ldots f_n]$, features are ordered by the subfeature relation, such that f_i (with $i > 0$) is a subfeature of f_{i-1}. Consider the following feature expression:

(1) 方向盘   的 指针
    steering.wheel DE indicator
    the indicator of the steering wheel

Our representation is [SteeringWheel, Indicator], whereby Indicator is interpreted as a subfeature of SteeringWheel.

Further, implicit features that are contained in associated evaluations are also "moved" into the feature sequence:

---

(2) 方向盘　　　的 指针　很　精准。
steering.wheel DE indicator very precise
The indicator of the steering wheel is very precise.

This sentence contains the evaluation 'precise'. According to the above description, it is decomposed into a feature (Precision) and a positive evaluation. The feature is moved into the feature sequence. The resulting mapping is as follows:

(3) [SteeringWheel, Indicator, Precision] → +2

Thus, instead of limiting ourselves to entities and restricted sets of their immediate features, we adapt a "higher-order" view and allow a hierarchical feature sequence of arbitrary depth. This structure seamlessly integrates implicit features and flexibly captures any granularity that is intended by the author of the text. At the same time, the value of the evaluation is reduced to a single numerical value, which allows for a straightforward aggregation of the final results.

4　Lexical basis

Out lexical resources contain functional and semantic categories. Members of "functional" categories (e. g. conjunctions, phrase-final markers) are only relevant for the syntactic analysis. Semantic categories are relevant for the interpretation of opinions. The top-level semantic categories are:

- Features, e. g. 外观 ('look'), 座椅 ('seat'), 颜色 ('color')
- Evaluations:
  - with implicit features, e. g. 好看 ('beautiful' → VisualAppearance), 便宜 ('cheap' → Price)
  - without implicit features, e. g. 不错 ('not bad'), 一般 ('ordinary'), 还可以 ('OK')
- Emotions, e. g. 赞美 ('admire'), 烦人 ('annoying')
- Degree adverbs and negation words, e. g. 非常 ('very'), 稍微 ('a little bit'), 不 ('not')

Each of these categories is in turn subclassified into more fine-grained classes which capture information about the linguistic use of the subclass members.

5　Processing steps

Figure illustrates the in- and output, the three processing steps as well as the resources involved in these steps.

5.1　Preprocessing

We use the third-party tool jieba[3] for word segmentation and POS tagging; both steps are customized in order to achieve a better performance on domain- and task-specific data. Specifically, the dictionary provided by the tool is intersected with a user-specified dictionary. This user-specified dictionary contains all words from our lexical resources. The user-added words are annotated with customized POS tags, such as 'F' for feature, 'EV' for evaluation etc. The following two examples depict the same sentence as output by jieba without and with customization:

(4)　a.　original jieba output without customization:
后排/vn 空间/n 已经/d 做/v 得/ud
rear.row space　already make DE
很/d 不错/a 了/ul 。/x
very not.bad PFV
The rear space is already quite not bad.

　　b.　after customization:
后排空间/F 已经/d 做/v 得/ud
rear.space　already make DE
很/D 不错/EV 了/ul 。/x
very not.bad　PFV
The rear space is already quite not bad.

Thus, we see that the two words 后排 ('rear row') and 空间 ('space') are merged into one word in the customized output since this combination occurs frequently in automotive texts and has a quasi-lexicalized meaning; the resulting word gets our custom POS tag 'F' (feature). Further, the POS tag of 不错 is changed from the original jieba tag 'a' to the custom tag 'EV' (evaluation).

5.2　Unit identification

In the next step, we identify phrasal units corresponding to features, evaluations, emotions. We use a phrase rule grammar which is based on regular expressions involving the POS tags of the
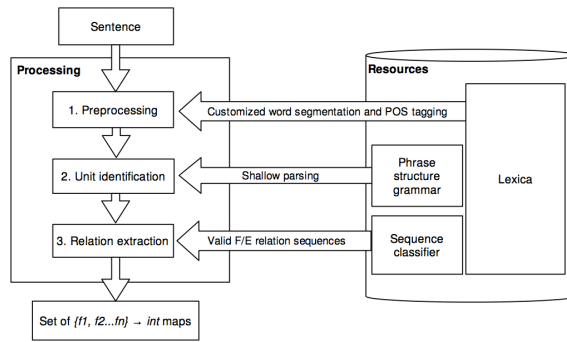
---

[3]`https://github.com/fxsjy/jieba`

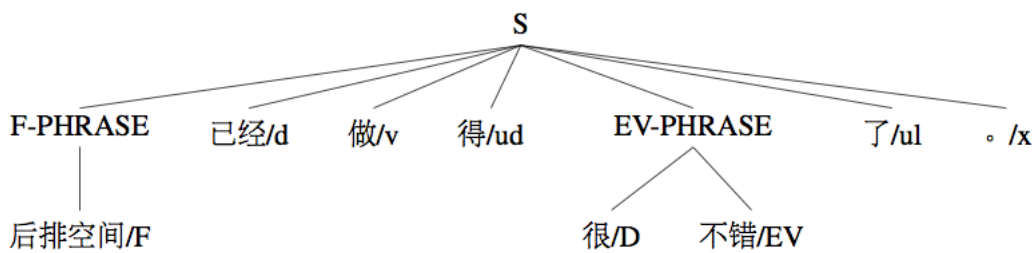Figure 1: Overall architecture of the system

Figure 2: Phrasal analysis of the sentence 后排空间 已经 做 得 很 不错 了。

words. Figure 2 shows the parsed version of example (4b).

In the following, we present some of the most common phrase structures for features and evaluations/emotions that are used in our system.

Feature phrases Besides simple NPs consisting only of one feature word, the most frequent types of feature phrases are phrases with nominal modifiers, coordinated NPs and NPs with pronominal modifiers:

(5) NP modifier:

座椅 的 材料
seat DE material

the material of the seats

(6) 它 的 设计
it DE design
its design

(7) 前排 （跟/和） 后排
front.row (and) rear.row
the front and the rear row

Evaluation and emotion chunks The class of evaluations consists of adjectives, whereas the class of emotions consists both of adjectives and verbs. However, evaluations and emotions get a unified treatment at the unit level, since Chinese stative verbs behave similarly to adjectives: they can also be modified by degree adverbs, used in comparative constructions etc.

Besides simple lexical units, the following are the most frequent phrase types for the E class:

(8) a. Verb or adjective preceded by negation or degree adverb:
很 难受
very difficult.to.bear
very difficult to bear

b. Adjective followed by degree adverb:
小 了 点
small PFV a.bit
a bit small

Evaluations can be coordinated in various ways; for example, coordination can be expressed by simple juxtaposition, with a comma or in the 又 E1 又 E2 construction:

(9) a. juxtaposition / punctuation:
精准 （，）灵活
precise (,) flexible
precise and flexible

58

b. 又 E1 又 E2:
又 　 精准 　 又 　 灵活
CONJ precise CONJ flexible
both precise and flexible

Besides, evaluations are often expressed by so-called "possessed features": the evaluation value is derived from the "amount" to which a feature is possessed by the discussed entity:

(10) 没 有 活力
NEG have vigor
not vigorous

5.3 Relation extraction

After identifying the syntactic units of interest, we proceed with identifying sentence-level relations between these units. In the literature, there are two major approaches to the identification of relations between evaluations and their targets. On the one hand, some authors recur to parsing and identify evaluation targets based on dependency relations (Wu et al. 2009, Jiang et al. 2011, *i. a.*). On the other hand, distance metrics can be used (Ding et al., 2008; Liu, 2012). Since we want to avoid the overhead of full-fledged syntactic parsing, but also want to improve the accuracy of simple distance metrics, we develop a sequence classifier which determines whether a given sequence of words between a feature and an evaluation/emotion phrase indicates a target relation.

The two semantic relations of interest are the *causer* and the *theme* relation. Additionally, the system analyzes a third non-semantic relation – the *topic* – which provides relevant discourse-structural information on the overall aspect discussed in a sentence.

The *causer* relation The causer relation is a fairly well-delimited relation which describes the causer of some state of event. In our model, it is applied to emotions caused by specific features. In the most basic cases, the causer is expressed as subject of one of the causative verbs (让, 令 etc.):

(11) 动力 让 　 我 非常 失望。
power CAUS me very desperate
The power really makes me desperate.

The *theme* relation The *theme* relation is expressed differently for evaluations and emotions. In the case of evaluations, it can be realized as the single argument of an AP or the nominal head of an adjectival modifier:

(12) a. Single argument of an AP:
设计 特别 　 时尚。
design particularly fashionable
The design is particularly fashionable.

b. Nominal head of an adjectival modifier:
特别 　 时尚 　 的 设计
particularly fashionable DE design
a particularly fashionable design

With respect to emotions, the *theme* relation is only relevant for verbs; the feature targets of adjectives are covered by the causer relation. Thus, themes can be expressed as (possibly topicalized) objects of emotion verbs:

(13) a. Object in canonical postverbal position:
我 很 　 喜欢 它 的 设计。
me very like it DE design
I like its design a lot.

b. Topicalized object:
设计 　 很 　 喜欢 , ...
design very like , ...
The design, I like it a lot, ...

5.4 Relation extraction

In the above examples, relations hold between adjoined constituents and can thus be easily recognized. However, in many cases, several words occur between the evaluation/emotion and its target:

(14) 后排空间 已经 做 　 得 很 不错
rear.row space already make DE very
了。
not.bad PFV
The rear space is already quite not bad.

From our corpus, we bootstrap the most frequent sequences that occur between themes and emotions/evaluations, emotions and themes as well as causers and emotions. We then apply a simple classifier for the classification of unseen sequences.

6 Evaluation

The system is evaluated on a testset of 800 sentences annotated for feature, evaluation and emotion phrases and for relations between them. The annotation was carried out according to previously developed annotation guidelines; we worked with

	Precision	Recall
F-phrases	87.43%	85.37%
EV-phrases	89.21 %	84.29%
EM-phrases	88.56%	85.32%

Table 1: Results of unit identification

	Precision	Recall
F-EV relations - theme	89.2%	87.33%
F-EM relations - theme	84.01%	83.10%
F-EM relations - causer	86.49%	87.90%

Table 2: Results of relation extraction

two independent annotators - a native Chinese student without specialized linguistic knowledge and a non-native linguist with very good mastery of the Chinese language. They proceeded in three steps: at the phrase level, the total F-score of inter-annotator agreement was 91.3%. The diverging items were discussed with a third team member to create a unified phrase-level annotation. The reviewed corpus was then annotated for relations between opinion and their targets; in this step, inter-annotator agreement reached 93.4%.

Table 1 shows the results achieved in unit identification; table 2 shows the results achieved for relation extraction on the test set with finalized annotation of F/EV/EM phrases.

7 Outlook

We have shown that the use of a prebuilt lexicon together with the application of general language rules allows to achieve a considerable accuracy in ABSA for Chinese. Currently, the presented system is being extended with a number of more complex sentence-level relations, specifically comparative structures and modal operators. Further,

References

Boiy, Erik and Moens, Marie-Francine. 2009. A machine learning approach to sentiment analysis in multilingual Web texts. *Inf. Retr.* 12(5), 526–558.

Ding, Xiaowen, Liu, Bing and Yu, Philip S. 2008. A Holistic Lexicon-based Approach to Opinion Mining. In *Proceedings of WSDM'08*, WSDM '08, pages 231–240.

Huang, James C.-T. 1982. *Logical relations in Chinese and the theory of grammar*. Ph. D.thesis, MIT, Massachusetts.

Jiang, Long, Yu, Mo, Zhou, Ming, Liu, Xiaohua and Zhao, Tiejun. 2011. Target-dependent Twitter Sentiment Classification. In *Proceedings of ACL'11 - Volume 1*, pages 151–160.

Kennedy, Christopher and McNally, Louise. 2005. Scale structure, degree modification, and the semantics of gradable predicates. *Language* 81, 345 – 381.

Kim, Soo-Min and Hovy, Eduard. 2006. Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text. In *Proceedings of the Workshop on Sentiment and Subjectivity in Text*, SST '06, pages 1–8.

Ku, Lunwei, Huang, Tinghao and Chen, Hsinhsi. 2009. Using Morphological and Syntactic Structures for Chinese Opinion Analysis. In *Proceedings of EMNLP'09*, pages 1260–1269.

Li, Audrey Yen-Hui. 1990. *Order and Constituency in Mandarin Chinese*. Studies in Natural Language and Linguistic Theory, Dordrecht: Kluwer Academic Publishers.

Liu, Bing. 2012. Sentiment Analysis and Opinion Mining.

Liu, Lizhen, Lei, Mengyun and Wang, Hanshi. 2013. Combining Domain-Specific Sentiment Lexicon with Hownet for Chinese Sentiment Analysis. *Journal of Computers* 8(4).

Pontiki, Maria, Galanis, Dimitris, Pavlopoulos, John, Papageorgiou, Harris, Androutsopoulos, Ion and Manandhar, Suresh. 2014. SemEval-2014 Task 4: Aspect Based Sentiment Analysis. In *Proceedings of the SemEval'14*, pages 27–35, Dublin, Ireland: ACL and Dublin City University.

Popescu, Ana Maria and Etzioni, Oren. 2005. Extracting Product Features and Opinions from Reviews. In *Proceedings of HLT & EMNLP'05*, pages 339–346, Stroudsburg, USA.

Su, Qi, Xu, Xinying, Guo, Honglei, Guo, Zhili, Wu, Xian, Zhang, Xiaoxun, Swen, Bin and Su, Zhong. 2008. Hidden Sentiment Association in Chinese Web Opinion Mining. In *Proceedings of WWW'08*.

Wu, Yuanbin, Zhang, Qi, Huang, Xuanjing and Wu, Lide. 2009. Phrase Dependency Parsing for Opinion Mining. In *Proceedings of EMNLP'09*, pages 1533–1541, Stroudsburg, USA.