Toward finer-grained sentiment identification in product reviews through linguistic and ontological analyses

Hye-Jin Min
Computer Science Department
KAIST, Daejeon, KOREA
hjmin@nlp.kiast.ac.kr

Jong C. Park
Computer Science Department
KAIST, Daejeon, KOREA
park@nlp.kaist.ac.kr

Abstract

We propose categories of finer-grained polarity for a more effective aspect-based sentiment summary, and describe linguistic and ontological clues that may affect such fine-grained polarity. We argue that relevance for satisfaction, contrastive weight clues, and certain adverbials work to affect the polarity, as evidenced by the statistical analysis.

1 Introduction

Sentiment analysis have been widely conducted in several domains such as movie reviews, product reviews, news and blog reviews (Pang et al., 2002; Turney, 2002). The unit of the sentiment varies from a document level to a sentence level to a phrase-level, where a more fine-grained approach has been receiving more attention for its accuracy. Sentiment analysis on product reviews identifies or summarizes sentiment from reviews by extracting relevant opinions about certain attributes of products such as their parts, or properties (Hu and Liu, 2004; Popescu and Etzioni, 2005). Aspect-based sentiment analysis summarizes sentiments with diverse attributes, so that customers may have to look more closely into analyzed sentiments (Titov and McDonald, 2008). However, there are additional problems.

First, it is rather hard to choose the right level of detail. If concepts corresponding to attributes are too general, the level of detail may not be so much finer than the ones on a document level. On the other hand, if concepts are too specific, there may be some attributes that are hardly mentioned in the reviews, resulting in the data sparseness problem. Second, there are cases when some crucial information is lost. For ex-

ample, suppose that two product attributes are mentioned in a sentence with a coordinated or subordinated structure. In this case, the information about their relation may not be shown in the summary if they are classified into different upper-level attributes. Consider (1).

(1) a. 옷은 맞지만/맞긴 한데, 색상이 너무 어두워요. osun macciman, sayksangi nemwu etwuweyo. 'It fits me okay, but the color is too dark.' (size: barely positive, color: negative)
b. 생각보다 좀 얇지만, 안에 받쳐 입는 거니까 나름 괜찮은거 같아요. sayngkakpota com yalpciman, aney patchye ipnun kenikka nalum kwaynchanhunke kathayo. 'It's a bit thinner than I thought, but it is good enough for layering.' (thickness: negative but acceptable, overall: positive)

Example (1) shows sample customer reviews about clothes, each first in Korean, followed by a Yale Romanized form, and an English translation. Note that the weight of the polarity in the sentiment about size e.g. in (1a) is overcome by the one about color. However, if the overall sentiment is computed by considering only the number of semantically identical phrases in the reviews, it misses the big picture.

In particular, when opinions regarding attributes are described with respect to expressions whose polarities are dependent on the specific contexts such as the weather or user preference, an overestimated or underestimated weight of the sentiment for each attribute may be assigned. In our example, 얇다/yalpta/'thin' has an ambiguous polarity, i.e., either positive or negative, whose real value depends on the expected utility of the clothes. In this case, the negative polarity is the intended one, as shown in (1b). In order to reflect this possibility, we need to adjust the weight of each polarity accordingly.

In this paper, we propose to look into the kind of linguistic and ontological clues that may in-

fluence the use of polarities, or the relevance for 'satisfaction of purchase' inspired by Kano's theory of quality element classification (Huisko-nen and Pirttila, 1998), the conceptual granularities, and such syntactic and lexical clues as conjunction items and adverbs. They may play significant roles in putting together the identified polarity information, so as to assess correctly what the customers consider most important. We conducted several one-way Analysis of Variance (ANOVA) tests to identify the effects of each clue on deriving categories of polarity and quantification method 2 to see whether these clues can distinguish fine-grained polarities correctly.

Section 2 introduces categories of polarity. Section 3 analyzes ontological and linguistic clues for identifying the proper category. Section 4 describes our method to extract such clues for a statistical analysis. Section 5 discusses the results of the analysis and implications of the results. Section 6 concludes the paper.

2 Categories of polarity

We suggest two more fine-grained categories of polarity, or 'barely positive' (BP) and 'acceptably negative' (AN), in addition to positive (P), negative (N) and neutral (NEU). We distinguish 'barely positive' from normal positive and distinguish 'acceptably negative' from normal negative in order to derive finer-grained sentiments. Wilson and colleagues (2006) identified the strength of news articles in the MPQA corpus, where they separated intensity (low, medium, high) from categories (private states). For the purpose of identifying each attribute's contribution to the satisfaction after purchase, we believe that it is not necessary to have so many degrees of intensity. We argue that the polarity of 'barely positive' may hold attributes that must be satisfied and that 'acceptably negative' may hold those that are somewhat optional.

3 Linguistic and Ontological Analyses

In this section, we discuss linguistic and ontological clues that influence the process of identifying finer-grained polarity. For the purpose of exposition, we build hierarchical and aspect-based review structure as shown in Figure 1. Major aspects include Price, Delivery, Service, and Product. If we go down another level, Product is divided into Quality and Comfortableness. In defining relevant attributes, we consider all the lower-level concepts of major aspects, which

contain the characteristics of the product with a description of the associated sentiment.
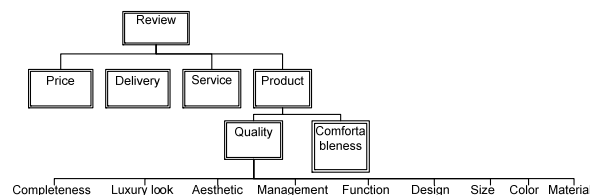

Figure 1. Review structure

Relevance for Satisfaction: We consider relevant attributes that affect the quality and satisfaction of the products as one of the important clues. Quality elements classified by Kano as shown in Table 1 can be base indicators of relevant attributes for satisfaction in real review text. For example, while completeness of the product may become crucial if the product has a defect, it is usually not the case that it would contribute much to the overall satisfaction of the customer.

Quality Elements	Example features
Must-be Quality (MQ)	Durability, Completeness
1-dimension Quality (1DQ)	Design, Color, Material
Attractive Quality (AQ)	Luxurious look

Table 1. Kano's Quality Elements

Conceptual Granularity: The concepts corresponding to attributes have a different level of detail. If the customer wants to comment on some attributes in detail, she could use a fine-grained concept (e.g., the width of the thigh part of the pants) rather than a coarse-grained one (e.g., just the size of the pants). To deal properly with the changing granularity of such concepts, we constructed a domain specific semi-hierarchical network for clothes of the Clothing-Type structure, in addition to the Review structure, by utilizing hierarchical category information in online shopping malls. Figure 2 shows an example for "pants".
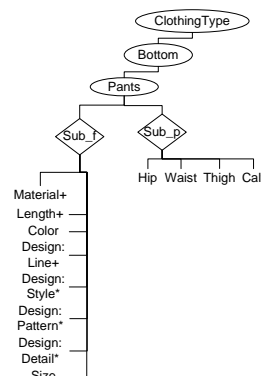

Figure 2. ClothingType structure for pants

Syntactic and Lexical Clues: Descriptions of each attribute in the reviews are often expressed

in a phrase or clause, so that conjunctions, or endings of a word with a conjunctive marker in Korean, play a significant role in connecting one attribute to another. They also convey a subtle meaning of the sentiment about relations between two or more connected attributes. We classified such syntactic clues into 4 groups of likeness (L), contrary (C), cause-effect (CE), and contrary with contrastive markers (CC).

Wilson and colleagues (2006) selected some syntactic clues as features for intensity classification. The selected features are shown to improve the accuracy, but the set of clues may vary to the nature of the given corpus, so that some otherwise useful clues that reflect a particular focused structure may not be selected. We argue that some syntactic clues such as the use of certain conjunctions can be identified manually to make up for the limitation of feature selection.

Adverbs modifying adjectives or verbs such as *too*, and *very* also strengthen the polarity of a given sentiment, so such clues work to differentiate normal positive or negative from 'barely positive' and 'acceptably negative'. Table 2 summarizes linguistic clues in the present analysis.

Clues	Examples	
CONJ/ END	L	-고 -ko 'and'
	C	-지만 -ciman 'but', 그러나 kulena 'however'
	CE	-어서 -ese 'so', 그래서 kulayse 'therefore'
	CC	-긴 –지만 -kin -ciman 'It's …', 'but', 'though'
ADV	Strong	매우 maywu 'very', 너무 nemwu 'too'
	Mild	좀 com 'a little'

Table 2. Syntactic and Lexical Clues

All these three types of clue that appear in the review text may interact with one another. For example, attributes with 'barely positive' tend to be described with a concept on a coarse level, and may belong to Must-be Quality (e,g., size in (1a)). However, if such attributes are negative, customers may explain them with a very fine-grained concept (e.g., the width of thigh is okay, but the calf part is too wide; interaction between relevance for satisfaction and conceptual granularity). They may also use adverbs such as 'too' to emphasize such unexpected polarity information. For emphasis, a contrastive structure can be used to indicate which attribute has a more weight (e.g., 'A but B'; interaction between syntactic clues and relevance for satisfaction). In

addition, an unfocused attribute A may be the attribute with 'acceptably negative' if the polarity of the attribute B is positive. We believe that the interaction between lexical and syntactic clues and relevance for satisfaction are the most important and that this correlation information may be utilized with such fine-grained polarity as 'barely positive' or 'acceptably negative'.

4   Clue Acquisition

We acquired data semi-automatically for each clue from the extracted attributes and their descriptions from 500 product reviews of several types of pants and annotated polarities manually. We obtained raw text reviews from one of the major online shopping malls in Korea[1] and performed a morphology analysis and POS-tagging. After POS-tagging, we collected all the noun phrases as candidates of attributes. We regarded some of them as attributes with the following guidelines and filtered out the rest: 1) NP with frequent adjectives 2) NP with frequent non-functional and intransitive verbs. In the case of subject omission, we converted adjectives or verbs into their corresponding nouns, such as 'thin' into 'thickness'. Hu and Liu (2004) identified attributes of IT products based on frequent noun phrases and Popescu and Etzioni (2005) utilized PMI values between product class (hotels and scanners) and some phrases including product. In our case, we used attributes that belong only to the Product concept in the Review structure, because most attributes we consider are sub-types or sub-attribute of Product. The total number of <attribute, polarity> pairs is 474.

For relevance for satisfaction, we converted extracted attributes into one of the types of Kano's quality elements by the mapping table we built. For conceptual granularity we regarded all the attributes with a depth less than 2 as 'coarse' and those more than 2 as 'fine'. Syntactic and lexical clues are identified from the context information around extracted adjective or verbs by the patterns based on POS information.

5   Statistical Analysis and Discussion

We conducted one-way Analysis of Variance (ANOVA) tests using relevance for satisfaction (ReV), conceptual granularity (Granul), and two linguistic clues, ADV and CONJ/END, in order to assess the effects of each clue on identifying categories of polarity. The ANOVA suggests

---

reliable effects of ReV (F(2,474) = 22.2; p = .000), ADV (F(2, 474) = 41.3; p = .000), and CONJ/END (F(3, 474) = 6.1; p = .000). We also performed post-hoc tests to test significant differences. For ReV, there are significant differences between 'MQ' and '1DQ' (p=.000), and between 'MQ' and 'AQ' (p =.032). AQ is related to 'positive' and MQ to 'acceptably negative' by the result. For ADV, there are significant differences between all pairs (p <.05). For CONJ/END, there are significant differences between 'likeness' and 'contrary' (p = .015), and between 'likeness' and 'contrary with contrastive markers' (p = .025). The 'contrary' and 'contrary with contrastive markers' types of conjunctions are related to 'acceptably negative'.

We also conducted Quantification method 2 to see if these clues can discriminate between BP and P and discriminate between AN and N. The regression equation for distinguishing AN from N is statistically significant at the 5% level (F(7,177) = 12,2; R^2=0.335; Std. error of the estimate = 0.821; error rate for discriminant = 0.21). The coefficients for 'mild' (t^2=30.8), 'contrary' (t^2=17.8) and 'contrary with contrastive markers' (t^2=14.1) are significant.

The results lead us to conclude that we can identify 'acceptably negative' from the clothes reviews by extracting the particular lexical clue, adverbs of 'mild' category and syntactic clue, such as conjunctions of 'contrary', and 'contrary with contrastive markers', or *contrastive weight*. This clue may convey the customer's argumentative intention toward the product, or *argumentative orientation*, for instance, A and B in '*A but B. C*' have different influence on the following discourse C (Elhadad and McKeown, 1990).

Although 'contrary with contrastive markers' plays an important role in identifying 'acceptably negative', it could also be used to identify another type of 'positive' as shown in example (2).

> (2) 좀 두껍다는 생각이 듭니다. 그래도 따뜻하긴 하네요. com twukkeptanun sayngkaki tupnita. kulayto ttattushakin haneyyo. 'It is a bit thick, but it keeps me warm.'

It is a positive feature, but neither fully positive nor barely positive. It seems to be somewhere inbetween. The order of appearance in reviews may also affect the strength of polarity. In addition, particular cue phrases such as ~것만 빼고/kesman ppayko/'except that …' can also convey 'acceptably negative', too.

In the future, we need to assess the importance of each proposed clue relative to others and to the existing ones. We also need to investigate the nature of interactions among linguistic, ontological and relevance for satisfaction clues, which may influence the actual performance for identifying finer-grained polarity.

6 Conclusion and Future Work

We proposed further categories of polarity in order to make aspect-based sentiment summary more effective. Our linguistic and ontological analyses suggest that there are clues, such as 're-levance for satisfaction', 'contrastive weight' and certain adverbials, that work to affect polarity in a more subtle but crucial manner, as evidenced also by the statistical analysis. We plan to find out product attributes that contribute most to modeling the interaction among the proposed clues in effective sentiment summarization.

References

Ana-Maria Popescu and Oren Etzioni 2005. *Extracting Product Features and Opinions from Reviews*. Proc. HLT/EMNLP 2005, 339-346.

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. *Thumbs up? Sentiment classification using machine learning techniques*. Proc. EMNLP.

Ivan Titov and Ryan McDonald 2008. *A Joint Model of Text and Aspect Ratings for Sentiment Summarization*. Proc. ACL-08: HLT, 308-316.

Janne Huiskonen and Timo Pirttila. 1998. *Sharpening logistic customer service strategy planning by applying Kano's quality element classification*. International Journal of Producion Economics, 56-57, 253-260, Elsevier Science B.V.

Michael Elhadad and Kathleen R. McKeown. 1990. *Generating Connectives*. Proc. COLING'97-101.

Minqing Hu and Bing Liu. 2004. *Mining and summarizing customer reviews*. Proc. ACM SIGKDD, 168–177. ACM Press.

Peter D. Turney. 2002. *Thumbs up or thumbs down? Sentiment orientation applied to unsupervised classification of reviews*. Proc. ACL, 417-424.

Theresa Wilson, Janyce Wiebe, and Rebecca Hwa. 2006. *Recognizing Strong and Weak Opinion Clauses*. Computational Linguistics, 22 (2): 73-99.