Multi-Aspect Opinion Polling from Textual Reviews

Jingbo Zhu Huizhen Wang
Natural Language Processing Lab
Northeastern University, China
{zhujingbo,wanghuizhen}@mail.neu.edu.cn

Benjamin K. Tsou
Language Information Sciences Research Centre, City University of Hong Kong
rlbtsou@cityu.edu.hk

Muhua Zhu
Natural Language Processing Lab
Northeastern University, China
zhumuhua@gmail.com

ABSTRACT

This paper presents an unsupervised approach to aspect-based opinion polling from raw textual reviews without explicit ratings. The key contribution of this paper is three-fold. First, a multi-aspect bootstrapping algorithm is proposed to learn from unlabeled data aspect-related terms of each aspect to be used for aspect identification. Second, an unsupervised segmentation model is proposed to address the challenge of identifying multiple single-aspect units in a multi-aspect sentence. Finally, an aspect-based opinion polling algorithm is presented. Experiments on real Chinese restaurant reviews show that our opinion polling method can achieve 75.5% precision performance.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database Application – *data mining*; I.2.7 [**Artificial Intelligence**]: Natural Language Processing – *text analysis*.

General Terms

Algorithms, Experimentation

Keywords

Opinion polling, review mining, sentiment analysis

1. INTRODUCTION

With the growing availability of opinion-rich resources such as product reviews and personal blogs, an opinion poll is a popular way to present quantitative indication of user positive and negative opinions on social issues or products. In practice, for example, to investigate public opinions on restaurants, we care more about some particular aspects such as *food* or *service* instead of overall ratings. Nowadays people increasingly express their opinions in the form of textual reviews without explicit numeral ratings, and often present differing opinions on several aspects simultaneously in the same review. To handle such case, aspect-based sentiment analysis techniques are required [1][2].

In this paper we explore the problem of aspect-based opinion polling automatically from raw textual reviews without explicit ratings. An aspect-based opinion polling system takes as input a set of textual reviews and some predefined aspects, and identifies the polarity of each aspect from each review to produce an opinion poll. Here we first provide a Chinese restaurant review example to describe the aspect-based opinion polling task.

Chinese Restaurant Review	
Sentence-1	环境不错，菜品一般，很贵。 *The environment is nice. The quality of food is so so. The food is very expensive.*
Sentence-2	服务我很欣赏。 *I like their service very much.*

Figure 1: A restaurant review example of two sentences

From the above review, a generated aspect-based opinion poll of five [aspect, polarity] pairs involving *environment*, *discount policy*, *food*, *charge* and *service* aspects, is shown in Table 1.

Table 1. A generated opinion poll

Opinion Poll Form			
Aspect	Polarity		
	Positive(+)	Neutral	Negative(-)
Environment	√		
Discount Policy[1]		√	
Food			√
Charge			√
Service	√		

2. ASPECT-RELATED TERM LEARNING

Like features used in previous studies on sentiment analysis [3][4], in this study we consider two types of *aspect-related terms* (ARTs) for aspect identification, including word-type features such as *nouns*, *verbs*, *adjectives, adverbs*, and multi-word terms. To extract multi-word terms from unlabeled data, we utilize the *C-value* method [5] which takes as input a review set and produces a list of multi-word terms ranked in the descending order of *Cvalue score*. The Cvalue score of a multi-word term t can be calculated [5] by:

- If t is not contained by any other terms

$$Cvalue(t) = \log(|t|) \times frq(t),$$

- Otherwise

[1] The user did not provide any information on the aspect discount policy. In this case, we consider the polarity of the aspect *discount policy* as neutral (i.e. unknown).

CIKM'09, November 2-6, 2009, Hong Kong, China.

$$Cvalue(t) = \log(|t|)(frq(t) - \frac{1}{n(L)}\sum_{l\in L} frq(l))$$

where $|t|$ denotes the number of words contained by t, $frq(t)$ indicates the frequency of occurrence of t in the corpus, L is the set of multi-word terms containing t, and $n(L)$ denotes the number of terms in S.

A *bootstrapping* method starts learning with a small number of seed ARTs for each aspect under the help of unlabeled data. Bootstrapping can be viewed as iterative clustering where in each learning cycle the most valuable candidate is chosen to augment the current seed set, and the learning procedure continues until the predefined stopping criterion is satisfied. We utilize the *RlogF* metric [6] to evaluate each candidate ART t by

$$\text{RlogF}(t) = \log frq(t,T) \times R(t,T), \quad (1)$$

where T is the current seed set, $frq(t,T)$ is the frequency of co-occurrence of t and T within a limited context (i.e., k words to left or right of t), $frq(t)$ is the frequency of co-occurrence of t in the corpus, and $R(t,T)=frq(t,T)/frq(t)$.

For the purpose of aspect identification based on these learned ART sets, we need to assign each learned ART with an important score that indicates the degree of its ability of reflecting the corresponding aspect. The importance score $SAB_i(t)$ of an ART t for the i^{th} aspect can be measured by means of a rank function as

$$SAB_i(t) = 1 - \frac{r_i(t)}{|S_i|}. \quad (2)$$

where $S_i=\{t_{i1}, t_{i2}, \ldots, t_{ik}\}$ is the ART set of the i^{th} aspect produced by SAB. Notice that t_{ij} is learned in the j^{th} iteration, $|S_i|$ indicates the number of ARTs in S_i, and $r_i(t)$ represents the rank of t in S_i, indicating in which iteration it was learned. A higher $SAB_i(t)$ value indicates that t is a more important ART for the i^{th} aspect.

Typically, we apply the bootstrapping method to learn an ART set for each aspect independently. We refer to this technique as *single-aspect bootstrapping* (SAB). However, we find that many ARTs occur in two or more SAB output lists, which means these ARTs belong to two or more aspects, named *multi-aspect ARTs*. The ambiguity degree $AD(t)$ of a multi-aspect ART t can be measured by means of an entropy-like function of ranks of t in m SAB output lists as

$$AD(t) = \frac{-\sum_{i=1}^{m} \frac{r_i(t)}{\sum_{1\le j\le m} r_j(t)} \log \frac{r_i(t)}{\sum_{1\le j\le m} r_j(t)}}{\log m} \quad (3)$$

where $S=\{S_1, S_2, \ldots, S_m\}$ represents ART sets for m aspects generated by SAB. The denominator is used for normalization. A higher $AD(t)$ value indicates that the ART t is more ambiguous.

By considering two above factors *SAB importance score* and *ART ambiguity degree* together, we present a new method for ART learning, namely *multi-aspect bootstrapping* (MAB). The importance score $MAB_i(t)$ of an ART t for the i^{th} aspect can be calculated by

$$MAB_i(t) = SAB_i(t) \times (1 - AD(t)) \quad (4)$$

Algorithm 1: Multi-Aspect Bootstrapping Learning
Input: initial aspect seed sets $S=\{S_1, S_2, \ldots, S_m\}$ for m aspects, and pool of unlabeled data U
Stage 1: Candidate ART Extraction
Extract nouns, verbs, adjectives, adverbs and top-n multi-word terms recognized by C-value method from U to form a candidate ART set Ω for bootstrapping.
Stage 2: **Single-Aspect Bootstrapping Learning**
Start learning with the seed set S_i for the i^{th} aspect;
Repeat
1. Use Equation (1) to calculate *RlogF* score of each candidate in Ω;
2. Select the candidate with the highest *RlogF* score to augment S_i, and remove it from Ω;

Until the predefined stopping criterion[2] is met.
Stage 3: **Re-scoring**
For each ART set S_i produced by SAB
1. Use Equation (4) to calculate *MAB* score of every ART in S_i;
2. Sort ARTs in the descending order of *MAB* score to produce final S_i.

Output: Final generated ART sets S^* for m aspects.

Figure 2. The multi-aspect bootstrapping algorithm

3. SENTENCE SEGMENTATION MODEL

Many sentences in real reviews often involve two or more aspects. Let's revisit the example shown in Fig. 1. The first sentence contains three single-aspect segments: an environment-segment (环境不错/ *the environment is nice*), a food-segment (菜品一般/ *the quality of food is so so*), and a charge-segment (很贵/ *the food is very expensive*). We refer to such sentences as *multi-aspect sentences*. To handle such case, we propose a *multi-aspect sentence segmentation* (MAS) model that takes as input a multi-aspect sentence and produces multiple single-aspect segments. A single-aspect segment might be a sub-sentence[3], or a combination of some consecutive sub-sentences. Let $C=c_1c_2\ldots c_n$ be a sentence C consisting of n sub-sentences, and let $U=u_1u_2\ldots u_k$ be a segmentation of C consisting of k single-aspect segments. The MAS model can be formalized by introducing a new criterion function $J(C,U)$ which aims to evaluate each candidate segmentation and assign a score to it.

$$U^* \overset{def}{=} \underset{U}{\text{argmax}} J(C,U) \quad (5)$$

It seems an appealing solution by incorporating aspect information of segments into the design of the criterion function $J(C,U)$. In this work, aspect identification of a segment is implemented by using ART sets $S=\{S_1, S_2, \ldots, S_m\}$ for m aspects produced by SAB or MAB method. The j^{th} aspect score of a

[2]The bootstrapping learning can end when a desirable number of ARTs have been learned.

[3] The separation mark between two adjacent sub-sentences is defined as a comma or a semicolon in a sentence.

segment u_i in U can be computed by suming the importance scores of all ARTs of the j^{th} aspect in the segment, that is

$$\phi_j(u_i) = \sum_{t \in u_i} score_j(t) \quad (6)$$

where t is an ART. The score(.) function represents SAB(.) or MAB(.). The most likely aspect j^* of the segment u_i is given by

$$j^* = \arg\max_j \phi_j(u_i) \quad (7)$$

To determine whether two adjacent segments in U are associated with different aspects, we use an indicator function $\delta(u_i,u_j)$ whose value is 1 if segments u_i and u_j are labeled as two different aspects, and 0 otherwise. $\delta(u_0,u_1)$ is assumed to be 1. The criterion function $J(C,U)$ can be designed by

$$J(C,U) = \sum_{1 \le i \le k} \delta(u_{i-1}, u_i) \times \phi_{j^*}(u_i) \quad (8)$$

In the final generated segmentation U^*, we adjust the aspect of each segment based on the following rule: *If the aspect score of the most likely aspect of a segment* u_i *is less than a predefined threshold, the aspect of the segment* u_i *is considered as NULL.*

4. ASPECT-BASED OPINION POLLING

An aspect-based opinion poll is described as a two-dimensional form filled with *3m* (aspect, polarity) pairs and their corresponding numbers, involving *m* aspects and three polarities (i.e. positive, negative and neutral). For each (aspect, polarity) pair, its associated number indicates how many textual reviews express it. The aspect-based opinion polling can be summarized as follows:

Algorithm 2: Aspect-based Opinion Polling
Input: a review set and *m* predefined aspects
Step1: For each review

1. Use the MAS method to implement the aspect segmentation of each sentence. Each multi-aspect sentence is segmented into multiple single-aspect segments;
2. Identify the aspect and the polarity of each segment, and then generate the (aspect, polarity) pair expressed by each segment;
3. Count the number of each (aspect, polarity) pair generated from this review;
4. If there is polarity conflict problem for one aspect, its predominant polarity is determined from the maximum of the number of competing pairs;
5. Output *m* final (aspect, polarity) pairs for this review.

Step2: Count the number of each (aspect, polarity) pair generated from the review set, and build an aspect-based opinion poll *P* by filled with the number of each (aspect, polarity) pair.
Output: an aspect-based opinion poll

Figure 3. Aspect-based opinion polling algorithm

In this work, a readily available Chinese sentiment lexicon released by Hownet[4] is used for polarity analysis. This lexicon contains 3730 positive words and 3116 negative words. Thirteen negation words such as "不/not" are used in handling negation. We adopt the sentiment-lexicon-based method [7] for segment polarity analysis, in which the semantic orientation value of a segment is to sum the polarity values of all sentiment words in the segment. The polarity value of a positive word is empirically set to +1, and set to -1 for a negative word. The resulting semantic orientation value of a segment indicates its corresponding polarity, that is, >0 for positive, <0 for negative, and equals to 0 for neutral.

[4] http://www.keenage.com/html/c_index.html

5. EXPERIMENTAL EVALUATION

We evaluate our aspect-based opinion polling method on a corpus of real Chinese restaurant reviews taken from the website *DianPing.com*, which contains 13,358 reviews (54,747 sentences in total) for 100 restaurants. In the preprocessing step we utilized the NEUCSP[5] tool to implement Chinese word segmentation and POS tagging.

We constructed a test set to evaluate the effectiveness of each polling method, which contains 3,325 reviews (11,606 sentences in total) for 25 randomly chosen restaurants. To form the gold standard, two human judges were asked to label each review with five appropriate (aspect, polarity) pairs involving five aspects (i.e. *environment, favorable policy, food, charge* and *service*) and their corresponding polarities (i.e. *positive, negative* or *neutral*). For the disagreements between two judges, a third human judge acted as an adjudicator. The rest of reviews were used as unlabeled data for bootstrapping-based ART learning.

In C-value method, we considered *bigrams*, *trigrams* and *4-grams* to be candidate multi-word terms. The *l* value of the limited context (used in Equation (1)) was empirically set to 5. The bootstrapping learning process stops when 2000 ARTs for each aspect have been learned. In each bootstrapping algorithms, five seeds were initially provided for each aspect as follows:

Table 2. Initial aspect seed sets for bootstrapping

Aspect	*Seeds*
Environment	环境/*environment*, 豪华/*luxury*, 装修/*decoration*,嘈杂/*noisy*, 吵闹/*noisy*
Discount Policy	打折/*discount*, 免费/ *free*, 赠券/*coupon* 优惠/*on sale*, 赠送/ *gift and rebate*
Food	食物/*food*, 口味/*taste*, 油腻/*oily,* 好吃/*delicious*, 正宗/*authentic*
Charge	价格/*price*, 贵/*expensive*, 便宜/*cheap*, 买单/*pay the bill*, 性价比/*cost performance*
Service	服务员/waiter, 体贴/*considerate*, 服务/service,周到/*good service* 热情/*friendly*

In the evaluation, we test four automatic opinion polling methods: 1)**MAB+MAS** method adopts MAB for ART learning, and MAS model for multi-aspect sentence segmentation; 2)**MAB+Full-Stop** method adopts MAB for ART learning, and simply considers the whole sentence as a single single-aspect segment

[5] NEUCSP is a Chinese word segmentation and POS tagging tool at (http://www.nlplab.com/chinese/source.htm)

(i.e. namely full-stop-based method); 3)**SAB+MAS** method adopts SAB for ART learning, and MAS for multi-aspect sentence segmentation; 4)**SAB+Full-Stop** method adopts SAB for ART learning, and full-stop-based method for multi-aspect sentence segmentation.

In the gold standard, each review was manually labeled with five appropriate (aspect polarity) pairs. Each automatic method also assigned five (aspect polarity) pairs to each review. In such case, actually the *precision* and *recall* metrics are same. To evaluate the effectiveness of each polling method, we adopted *precision* as the performance evaluation metric which measures the fraction of automatically identified (aspect, polarity) pairs that are correct.

Table 3. Results of different automatic methods for aspect-based opinion polling

Methods	Precision (%)
MAB+MAS	75.5
MAB+Full-Stop	70.3
SAB+MAS	71.2
SAB+Full-Stop	66.4

Table 3 shows the effectiveness of each automatic method for aspect-based opinion poll generation. MAB+MAS method achieves the highest average precision of 75.5%, which does not use labeled data. Experimental results show that in this task MAB outperforms SAB, and MAS is better than full-stop-based method.

6. RELATED WORK

Aspect-based sentiment summarization [2][8][9] aims to produce a summary expressing the aggregated sentiment for each aspect and supporting textual evidence. From real restaurant reviews we find that many sentences express two or more aspects. However, the issue of multi-aspect sentence segmentation is seldom mentioned in previous studies. We propose here a MAS model to segment a multi-aspect sentence into multiple single-aspect segments. Another difference is that our work focuses on aspect-based opinion poll generation instead of textual summarization generation. Some researchers [10][11] have applied bootstrapping method for learning subjective words or sentiment patterns, in the same fashion of our single-aspect bootstrapping method. We proposed a multi-aspect bootstrapping method for learning aspect-related terms for each aspect from unlabeled data. More recent work [12][13] has expanded polarity analysis on a multi-point scale under a ranking or ordinal regression framework. That is the most similar work to us. However, these previous studies focused on supervised or semi-supervised learning techniques for rating inference problem, which needs labeled data for training. However, in our work, no explicitly labeled data is used for training.

7. CONCLUSION AND DISCUSSION

This paper presents an unsupervised approach to aspect-based opinion polling from raw textual reviews. From experimental results, we found a *mixed opinion problem* that the number of segments expressing positive on an aspect (e.g. food aspect) is approximately equal to the number of other segments expressing negative on it in the same review. In such case, the predominant polarity of this aspect cannot be simply considered as positive or negative. The alternative way is to assign a degree score to each polarity expressed in such reviews.

Actually our approach can be applied to aspect-based English language opinion poll generation if one of readily available English sentiment lexicons such as *SentiWordNet*[6] is used for polarity analysis. Since most readily available sentiment lexicons are general-purpose knowledge bases, it is worth studying how to automatically transfer a general-purpose sentiment lexicon in real domain applications to achieve better performance.

8. ACKNOWLEDGEMENTS

This work was supported in part by the National Science Foundation of China (60873091).

9. REFERENCES

[1] Pang B. and Lee L. 2008. Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, Vol. 2, Nos. 1-2(2008) 1-135

[2] Hu M. and Liu B. 2004. Mining and summarizing customer reviews. In Proceedings of the 2004 ACM SIGKDD international conference on knowledge discovery and data mining, pp168-177.

[3] Pang, B. Lee L., and Vaithyanathan S. 2002. Thumbs up? Sentiment classification using machine learning techniques. In proc. of EMNLP02

[4] Riloff E., Patwardhan S., and Wiebe J. 2006. Feature subsumption for opinion analysis. In Proceedings of EMNLP06, pp.440-448

[5] Frantzi K., Ananiadou S., Mima H. 2000. Automatic recognition of multi-word terms: the C-value/NC-value method. International Journal on Digital Libraries, p115-130

[6] Riloff E. and Jones R. 1999. Learning dictionaries for information extraction by multi-Level bootstrapping, Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI-99)

[7] Wan X. 2008. Using bilingual knowledge and ensemble techniques for unsupervised Chinese sentiment analysis. In Proceedings of EMNLP08, pp553-561

[8] Gamon M., Aue A., Corston-Oliver S., and Ringger E. 2005. Pulse: mining customer opinions from free text. In Proceedings of the 6th international symposium on intelligent data analysis, pp.121-132

[9] Zhuang L., Jing F. and Zhu X. Y. 2006. Movie review mining and summarization. In Proceedings of the 15th ACM international conference on information and knowledge management, pp43-50

[10] Riloff, E., Wiebe J., and Wilson T. 2003. Learning subjective nouns using extraction pattern Bootstrapping. In Prof. of CoNLL-03

[11] Zagibalov T., and Carroll J. 2008. Unsupervised Classification of Sentiment and Objectivity in Chinese Text. In Proceedings of Proceedings of the Third International Joint Conference on Natural Language Processing, pp304-311

[12] Pang B. and Lee L. 2005. Seeing stars: exploiting class relationships for sentiment categorization with respect to rating scales. In Proceedings of the ACL05, pp115-124

[13] Snyder B. and Barzilay R. 2007. Multiple aspect ranking using the good grief algorithm. In Proceedings of NAACL/HLT-2007, pp300-307

[6] http://sentiwordnet.isti.cnr.it/.