Fine Granular Aspect Analysis using Latent Structural Models

Lei Fang[1] and **Minlie Huang**[2]
State Key Laboratory of Intelligent Technology and Systems,
Tsinghua National Laboratory for Information Science and Technology,
Department of Computer Science and Technology,
Tsinghua University, Beijing 100084, PR China.
[1]`fang-l10@mails.tsinghua.edu.cn`
[2]`aihuang@tsinghua.edu.cn`

Abstract

In this paper, we present a structural learning model for joint sentiment classification and aspect analysis of text at various levels of granularity. Our model aims to identify highly informative sentences that are aspect-specific in online custom reviews. The primary advantages of our model are two-fold: first, it performs document-level and sentence-level sentiment polarity classification jointly; second, it is able to find informative sentences that are closely related to some respects in a review, which may be helpful for aspect-level sentiment analysis such as aspect-oriented summarization. The proposed method was evaluated with 9,000 Chinese restaurant reviews. Preliminary experiments demonstrate that our model obtains promising performance.

1 Introduction

Online reviews have been a major resource from which users may find opinions or comments on the products or services they want to consume. However, users sometimes might be overwhelmed, and not be able to read reviews one by one when facing a considerably large number of reviews, and they may be not be satisfied by only being served with document-level reviews statistics (that is, the number of reviews with 1-star, 2-star, ..., respectively). Aspect-level review analysis may be alternative for addressing this issue as aspect-specific opinions may more clearly, explicitly, and completely describe the quality of a product from different properties.

Our goal is to discover informative sentences that are consistent with the overall rating of a review, and simultaneously, to perform sentiment analysis at aspect level. Notice, that a review with a high rating (say, 4/5 stars) may contain both negative and positive opinions, and the same to a review with a very low rating (say, 1/2 star). From our point of view, each review has a set of sentences that are informative and coherent to its overall rating. To perform fine granular sentiment analysis, the first step is to discover such coherent content.

Many information needs require the systems to perform fine granular sentiment analysis. Aspect-level sentiment analysis may be more useful for users to have a global picture of opinions on the product's properties. Furthermore, different users may have different preferences on different aspects of a product. Taking the reviews on mobile phones as an example, female users may focus more on the appearance while male users may lay more emphasis on the hardware configuration; younger users prefer to the app or game resources while older users may just pay attention to the basic function of calling or messaging.

In recent years, there has been much work focused multilevel sentiment classification using structural learning models. Yi (2007) extends the standard conditional random fields to model the local sentiment flow. Ryan (2007) proposed structured models for fine-to-coarse sentiment analysis. Oscar (2011) proposed to discover fine-grained sentiment with hidden-state CRF(Quattoni et al., 2007). Yessenalina (2010) deployed the framework of latent structural SVMs(Yu and Joachims., 2009) for multilevel sentiment classification. As for aspect level rating, ranking, or summarization, Benjamin(2007) em-

ployed the good grief algorithm for multiple aspect ranking and the extensions of the generative topic models were also widely studied, such as (Titov and McDonald., 2008; Brody and Elhadad., 2010; Wang et al., 2010; Li et al., 2011; Lu et al., 2011; Jo and Oh., 2011; Lin and He, 2009).

In this paper, we build a general structural learning model for joint sentiment classification and aspect analysis using a latent discriminate method. Our model is able to predict the sentiment polarity of document as well as to identify aspect-specific sentences and predict the polarity of such sentences. The proposed method was evaluated with 9,000 Chinese restaurant reviews. Preliminary experiments demonstrate that our model obtains promising performance.

2 Model

2.1 Document Structure

We assume that the polarity of document is closely related to some aspects for the reason that people are writing reviews to praise or criticize certain aspects. Therefore, each informative sentence of the document characterizes one aspect, expressing aspect specific polarity or subjective features. Similar to previous work on aspect analysis (Wang et al., 2010) and multi-level sentiment classification (Yessenalina et al., 2010), we define the aspect as a collection of synonyms. For instance, the word set {"value", "price", "cost", "worth", "quality"} is a synonym set corresponding to the aspect "price". For each document, an aspect is described by one or several sentences expressing aspect specific polarity or subjective information.

Let document be denoted by x, and $y \in \{+1, -1\}$ represents the positive or negative polarity of the document, s is the set of informative sentences, in which each sentence is attached with certain aspect $a_i \in A = \{a_1, ..., a_k\}$. Yessenalina (2010) chooses a sentence set that best explains the sentiment of the whole document while the s here retain this property. Let $\Psi(x, y, s)$ denote the joint feature map that outputs the features describing the quality of predicting sentiment y using the sentence set s.

Let x^j denote the j-th sentence of document x, and a^j is the attached aspect of x^j. In spirit to (Yessenalina et al., 2010), we propose the following formulation of the discriminate function

$$\vec{w}^T \Psi(x, y, s) =$$
$$\frac{1}{N(x)} \sum_{j \in s} \left(y \cdot \vec{w}_{pol_{a^j}}^T \psi_{pol}(x^j) + \vec{w}_{subj_{a^j}}^T \psi_{subj}(x^j) \right)$$

where $N(x)$ is the normalizing factor, $\psi_{pol}(x^j)$ and $\psi_{subj}(x^j)$ represents the polarity and subjectivity features of sentence x^j respectively. $\vec{w}_{pol}$ and $\vec{w}_{subj}$ denote the weight for polarity and subjectivity features. To be specific for each aspect, we have $\vec{w}_{pol_a}$ and $\vec{w}_{subj_a}$ representing the vector of feature weight for aspect a to calculate the polarity and subjectivity score.

$$\vec{w}_{pol}^T = \begin{bmatrix} \vec{w}_{pol_{a_0}}^T \\ \vdots \\ \vec{w}_{pol_{a_k}}^T \end{bmatrix}, \quad \vec{w}_{subj}^T = \begin{bmatrix} \vec{w}_{subj_{a_0}}^T \\ \vdots \\ \vec{w}_{subj_{a_k}}^T \end{bmatrix}$$

To make prediction, we have the document-level sentiment classifier as

$$h(x; \vec{w}) = \underset{y=\pm 1}{\operatorname{argmax}} \max_{s \in S(x)} \vec{w}^T \Psi(x, y, s)$$

where $S(x) = \{s \subseteq 1, \ldots, |x| : |s| \leq f(|x|)\}$, $f(|x|)$ is a function that depends only on the number of sentences in x, as illustrated in (Yessenalina et al., 2010). Therefore, for each sentence x^j, we compute the joint subjectivity and polarity score with respect to aspect a and label y as

$$score(x^j, a, y) = y \cdot \vec{w}_{pol_a}^T \psi_{pol}(x^j) + \vec{w}_{subj_a}^T \psi_{subj}(x^j)$$

we then assign aspect a^j to sentence x^j if

$$a^j = \underset{a \in A}{\operatorname{argmax}} \, score(x^j, a, y)$$

After sorting $score(x^j, a^j, y)$ in decreasing order and taking summation by selecting the top $f(|x|)$ (or fewer, if there are fewer than $f(|x|)$ that have positive joint score) sentences as the total score for each $y \in \{+1, -1\}$, we then predict y with the higher joint score as the sentiment of the whole document. This formulation of $\vec{w}^T \Psi(x, y, s)$ and classifier explains that for each sentence, the assigned aspect has the highest score over other aspects.

2.2 Feature Space

In our model, we use **bag-of-words** features. In order to obtain a model that is jointly trained, and satisfy the condition that the overall polarity of document should influence the sentiment of extracted informative sentences. We denote the weight vector modeling the polarity of entire document as $\vec{w}_{doc}$, as follows:

$$\vec{w}^T \Psi(x, y, s) =$$

$$\frac{y}{N(x)} \left(\sum_{j \in s} (\vec{w}_{pol_{aj}}^T \psi_{pol}(x^j) + \vec{w}_{doc}^T \psi_{pol}(x^j)) \right)$$

$$+ \frac{1}{N(x)} \left(\sum_{j \in s} \vec{w}_{subj_{aj}}^T \psi_{subj}(x^j) \right) + y \cdot \vec{w}_{doc}^T \psi_{pol}(x)$$

2.3 Training

We trained our model using the latent structural SVMs (Yu and Joachims., 2009).
OP1:

$$\min_{\vec{w}, \xi \geq 0} \frac{1}{2} ||w||^2 + \frac{C}{N} \sum_{i=1}^{N} \xi_i$$

$$s.t. \forall i :$$

$$\max_{s \in S(x_i)} \vec{w}^T \Psi(x_i, y_i, s) \geq$$

$$\max_{s' \in S(x_i)} \vec{w}^T \Psi(x_i, -y_i, s') + \triangle(y_i, -y_i, s') - \xi_i$$

We define $\triangle(y_i, -y_i, s') = 1$, that is, we view document level sentiment classification loss as the loss function. It should be noticed that OP1 is nonconvex. To circumvent the optimization difficulty, we employ the framework of structural SVMs (Tsochantaridis et al., 2004) with latent variables proposed by Yu (2009) using the CCCP algorithm (Yuille and Rangarajan., 2003). In terms of the formulation here, since the true informative sentence set is never observed, it is a hidden or latent variable. Thus, we keep s_i fixed to compute the upper bound for the concave part of each constraint, and rewrite the constraints as

$$\xi_i \geq \max_{s' \in S(x_i)} \vec{w}^T \Psi(x_i, -y_i, s') - \vec{w}^T \Psi(x_i, y_i, s_i) + 1$$

After that, we have y_i completed with the latent variable s_i as if it is observed. For each training example, starting with an initialization sentence set in which each sentence is with an aspect label, the training procedure alternates between solving an instance of the structural SVM using the s_i and predicting a new sentence until the learned weight vector $\vec{w}$ converges. In our work, we use the performance on a validation set to choose the halting iteration, as is similar to Yessenalina (2010).

2.4 Model Initialization

To initialize the informative sentence set, following the experiment result of Yessenalina (2010), we set $f(|x|) = 0.3 * |x|$, that is, we only select the top 30% of the total sentences as the set of informative part of the document. The normalizing factor is set as $N(x) = \sqrt{f|x|}$, as Yessenalina (2010) demonstrates that square root normalization can be useful. To analyze the aspect of each sentence, we need to give an initial guess of the aspect and sentiment for each sentence.

Sentence level sentiment initialization : To initialize the sentence level sentiment, we employ a rule based method incorporating positive and negative sentiment terms, with adversative relation considered.

Sentence aspect assignment initialization : Obviously, if a synonym of aspect a occurs in sentence x^l, we assign aspect a to x^l, and add x^l to an aspect specific sentence set P_a. For sentence x^l without any aspect term, we set a as the aspect label if

$$a = \operatorname*{argmax}_{a' \in A} similarity(x^l, P_{a'})$$

We select the sentences whose sentiment is consistent with the overall rating of a review as the initial guess of the latent variable.

3 Experiments

In this section, we evaluate our model in terms of document and sentence level sentiment classification, we also analyze the performance of aspect assignment for each sentence. The model is evaluated on the Chinese restaurant reviews crawled from Dianping[1]. Each of the reviews has an overall rating ranging from one to five stars. To be specific, we consider a review as positive if its rating is greater

---

than or equal to 4 stars, or negative if less than or equal to 2 stars. The corpus has 4500 positive and 4500 negative reviews. Data and an implementation of our model are publicly available[2].

We train 5 different models by splitting these reviews into 9 folds. Two folds are left out as the testing set, and each model takes 5 folds as training set, 2 folds as validation set, and the performance is averaged. Besides, we also manually label 100 reviews, in which each sentence is labeled as positive or negative corresponding to certain aspect or with no aspect description. On average, each review has 9.66 sentences. However, only 21.5% of the total sentences can be assigned to aspect by directly matching with aspect terms, which explains that keywords based aspect sentiment analysis may fail. For restaurant reviews, we pre-defined 11 aspects, and for each aspect, we select about 5 frequently used terms to describe that aspect. Table 1 shows some examples of the aspect synonym set used in this paper:

Aspect	Synonym Set
Taste	味道"taste", 口味"flavor"
Price	价格"price" , 价钱"cost"
Dishes	菜品"dishes", 菜式"cuisine"
Ingredients	食物"food" , 食材"ingredients"
Facility	设施"facility", 座位"seat"
Location	位置"location",
Environment	环境"environment",
	装修"decoration"
Service	服务"service" , 服务员"waiter"
	态度"attitude"

Table 1: Samples of Aspect Synonym.

Document level sentiment classification We compare our method with previous work on sentiment classification using standard SVM(Pang et al., 2002). Our model yields an accuracy of 94.15% while the standard SVM classifier yields an accuracy of 90.35%. Clearly, our model outperforms the baseline on document level sentiment classification.

Sentence level sentiment classification Our method can extract a set of informative sentence that are coherent to the overall rating of a review. The evaluation of sentence-level sentiment classification is based on manual annotation. We

sample 100 reviews, and present the extracted 300 sentences to annotators who have been asked to assign positive/negative/non-related labels. Among the sentences, 251 correctly classified as positive or negative while 49 are misclassified. And, 38 sentences of the 49 sentences have mix opinions or are non-subjective sentences.

Aspect Assignment To evaluate the accuracy of aspect assignment, we compare the predicted aspect labels with the ground truth (manual annotation). As some of sentences have explicit aspect terms and can be easily identified, we only consider those sentences without aspect words. In the extracted 300 sentences, 78 sentences have aspect terms, and for the rest, our model assigns correct aspect labels to 44 sentences while random guess only maps 21 sentences with right labels.

4 Conclusion and Future Work

In this paper, we address the task of multilevel sentiment classification of online custom reviews for fine granular aspect analysis. We present a structural learning model based on struct-SVM with latent variables. The informative sentence set is regarded as latent variable, in which each sentence is attached with certain aspect label. The training procedure alternates between solving an instance of the standard structural SVM optimization and predicting a new sentence set until the halting condition is satisfied. In addition, our model is a enough general model which can be easily extended to other domains. Preliminary experiments demonstrate that our model obtains promising performance.

There are several possibilities to improve our model. For future work, we propose to incorporate prior knowledge of latent variables to the model. One possible way is to reformulate the loss function by taking the predicted aspect of the extracted sentences into consideration. Another is to introduce confidence score to the extracted sentences, such that the learned support vectors that are labeled with higher confidence shall assert more force on the decision plane.

References

Samuel Brody and Noemie Elhadad. 2010. An unsupervised aspect-sentiment model for online reviews. In *Proceedings of Annual Conference of the North American Chapter of the ACL, (NAACL)*.

Yohan Jo and Alice Oh. 2011. Aspect and sentiment unification model for online review analysis. In *Proceedings of Conference on Web Search and Data Mining (WSDM)*.

Peng Li, Yinglin Wang, Wei Gao, and Jing Jiang. 2011. Generating aspect-oriented multi-document summarization with event-aspect model. In *Proceedings of Conference on Empirical Methods in Natural Language Processing, (EMNLP)*.

Chenghua Lin and Yulan He. 2009. Joint sentiment/topic model for sentiment analysis. In *Proceedings of the conference on Information and knowledge management(CIKM)*.

Bin Lu, Myle Ott, Claire Cardie, and Benjamin Tsou. 2011. Multi-aspect sentiment analysis with topic models. In *The ICDM' 2011 Workshop on Sentiment Elicitation from Natural Text for Information Retrieval and Extraction*.

Yi Mao and Guy Lebanon. 2007. Isotonic conditional random fields and local sentiment flow. In *Proceedings of Advances in Neural Information Processing Systems (NIPS)*.

Ryan McDonald, Kerry Hannan, Tyler Neylon, Mike Wells, and Jeff Reynar. 2007. Structured models for fine-to-coarse sentiment analysis. In *Proceedings of Annual Meeting of the Association for Computational Linguistics, (ACL)*.

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? sentiment classification using machine learning techniques. In *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

A. Quattoni, S. Wang, L.-P Morency, M. Collins, and T. Darrell. 2007. Hidden-state conditional random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Benjamin Snyder and Regina Barzilay. 2007. Multiple aspect ranking using the good grief algorithm. In *Proceedings of Annual Conference of the North American Chapter of the ACL, (NAACL)*.

Oscar Täckström and Ryan McDonald. 2011. Discovering fine-grained sentiment with latent variable structured prediction models. In *Proceedings of Annual European Conference on Information Retrieval , (ECIR)*.

Ivan Titov and Ryan McDonald. 2008. A joint model of text and aspect ratings for sentiment summarization. In *Proceedings of Annual Meeting of the Association for Computational Linguistics, (ACL)*.

Ioannis Tsochantaridis, Thomas Hofmann, Thorsten Joachims, and Yasemin Altun. 2004. Support vector machine learning for interdependent and structured output spaces. In *Proceedings of the International Conference on Machine Learning, (ICML)*.

Hongning Wang, Yue Lu, and Chengxiang Zhai. 2010. Latent aspect rating analysis on review text data: A rating regression approach. In *Proceedings of the International Conference on Knowledge Discovery and Data Mining (KDD)*.

Ainur Yessenalina, Yisong Yue, and Claire Cardie. 2010. Multi-level structured models for document-level sentiment classification. In *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Chun-Nam John Yu and Thorsten Joachims. 2009. Learning structural svms with latent variables. In *Proceedings of the International Conference on Machine Learning, (ICML)*.

A. L. Yuille and Anand Rangarajan. 2003. The concave-convex procedure (cccp). *Neural Computation*, 15:915–936.