

# An Iterative Reinforcement Approach for Fine-Grained Opinion Mining

**Weifu Du**

Haerbin Institute of Technology  
Haerbin, China

duweifu@software.ict.ac.cn

**Songbo Tan**

Institute of Computing Technology  
Beijing, China

tansongbo@software.ict.ac.cn

## Abstract

With the in-depth study of sentiment analysis research, finer-grained opinion mining, which aims to detect opinions on different review features as opposed to the whole review level, has been receiving more and more attention in the sentiment analysis research community recently. Most of existing approaches rely mainly on the template extraction to identify the explicit relatedness between product feature and opinion terms, which is insufficient to detect the implicit review features and mine the hidden sentiment association in reviews, which satisfies (1) the review features are not appear explicit in the review sentences; (2) it can be deduced by the opinion words in its context. From an information theoretic point of view, this paper proposed an iterative reinforcement framework based on the improved information bottleneck algorithm to address such problem. More specifically, the approach clusters product features and opinion words simultaneously and iteratively by fusing both their semantic information and co-occurrence information. The experimental results demonstrate that our approach outperforms the template extraction based approaches.

## 1 Introduction

In the Web2.0 era, the Internet turns from a static information media into a platform for dynamic information exchanging, on which people can express their views and show their selfhood. More and more people are willing to record their feelings (blog), give voice to public affairs (news review), express their likes and dislikes on products (product review), and so on. In the face of the volume of sentimental information available on the Internet continues to increase, there is growing interest in helping people better find, filter, and manage these resources.

Automatic opinion mining (Turney et al., 2003; Ku et al., 2006; Devitt et al., 2007) can play an important role in a wide variety of more flexible and dynamic information management tasks. For example, with the help of sentiment analysis system, in the field of public administration, the administrators can receive the feedbacks on one policy in a timelier manner; in the field of business, manufacturers can perform more targeted updates on products to improve the consumer experience.

The research of opinion mining began in 1997, the early research results mainly focused on the polarity of opinion words (Hatzivassiloglou et al., 1997) and treated the text-level opinion mining as a classification of either positive or negative on the number of positive or negative opinion words in one text (Turney et al., 2003; Pang et al., 2002; Zagibalov et al., 2008;). With the in-depth study of opinion mining, researchers committed their efforts for more accurate results: the research of sentiment summarization (Philip et al., 2004; Hu et al., KDD 2004), domain transfer problem of the sentiment analysis (Kanayama et al., 2006; Tan et al., 2007; Blitzer et al., 2007; Tan et al., 2008; Andreevskaia et al., 2008; Tan et al., 2009) and fine-grained opinion mining (Hatzivassiloglou et al., 2000; Takamura et al., 2007; Bloom et al., 2007; Wang et al., 2008; Titov et al., 2008) are the main branches of the research of opinion mining. In this paper, we focus on the fine-grained (feature-level) opinion mining.

For many applications (e.g. the task of public affairs review analysis and the products review analysis), simply judging the sentiment orientation of a review unit is not sufficient. Researchers (Kushal, 2003; Hu et al., KDD 2004; Hu et al., AAAI 2004; Popescu et al., 2005) began to work on finer-grained opinion mining which predicts the sentiment orientation related to different review features. The task is known as feature-level opinion mining.

In feature-level opinion mining, most of the existing researches associate product features and opinion words by their explicit co-occurrence. Template extraction based method (Popescu et al., 2005) and association rule mining based method (Hu et al., AAAI 2004) are the representative ones.

These approaches did good jobs for identifying the review features that appear explicitly in reviews, however, real reviews from customers are usually complicated. In some cases, the review features are implicit in the review sentences, but can be deduced by the opinion words in its context. The detection of such hidden sentiment association is a big challenge in feature-level opinion mining on Chinese reviews due to the nature of Chinese language (Qi et al., 2008). Obviously, neither the template extraction based method nor the association rule mining based method is effective for such cases. Moreover, in some cases, even if the review features appear explicitly in the review sentences, the co-occurrence information between review features and opinion words is too quantitatively sparse to be utilized. So we consider whether it is a more sensible way to construct or cluster review feature groups and opinion words groups to mine the implicit or hidden sentiment association in the reviews.

The general approach will cluster the two types of objects separately, which neglects the highly interrelationship. To address this problem, in this paper, we propose an iterative reinforcement framework, under which we cluster product features and opinion words simultaneously and iteratively by fusing both their semantic information and sentiment link information. We take improved information bottleneck algorithm (Tishby, 1999) as the kernel of the proposed framework.

The information bottleneck approach was presented by Tishby (1999). The basic idea of the approach is that it treats the clustering problems from the information compressing point of view, and takes this problem as a case of much more fundamental problem: *what are the features of the variable  $X$  that are relevant for the prediction of another, relevance, variable  $Y$ ?* Based on the information theory, the problem can be formulated as: *find a compressed representation of the variable  $X$ , denoted  $C$ , such that the mutual information between  $C$  and  $Y$  is as high as possible, under a constraint on the mutual information between  $X$  and  $C$ .* For our case, take the hotel reviews as example,  $X$

is one type of objects of review features (e.g. facilities, service, surrounding environment, etc) or opinion words (e.g. perfect, circumspect, quiet, etc), and  $Y$  is another one. Given some review features (or opinion words) gained from review corpus, we want to assemble them into categories, conserving the information about opinion words (or review features) as high as possible.

The information bottleneck algorithm has some benefits, mainly including (1) it treats the trade-off of precision versus complexity of clustering model through the rate distortion theory, which is a sub-field of information theory; (2) it defines the “distance” or “similarity” in a well-defined way based on the mutual information. The efficiency of information bottleneck algorithm (Slonim and Tishby, 2000) motivates us to take it as the kernel of our framework. As far as we know, this approach has not been employed in opinion mining yet.

In traditional information bottleneck approach, the distance between two data objects is measured by the Jensen-Shannon divergence (Lin, 1991), which aims to measure the divergence between two probability distributions. We alter this measure to integrate more semantic information, which will be illustrated in detail in the following sections, and the experimental result shows the effectiveness of the alteration.

It would be worthwhile to highlight several aspects of our work here:

- We propose an iterative reinforcement framework, and under this framework, review feature words and opinion words are organized into categories in a simultaneous and iterative manner.
- In the process of clustering, the semantic information and the co-occurrence information are integrated.
- The experimental results on real Chinese web reviews demonstrate that proposed method outperforms the template extraction based algorithm.

## 2 Proposed Algorithm

### 2.1 The Problem

In product reviews, opinion words are used to express opinion, sentiment or attitude of reviewers. Although some review units may express general opinions toward a product, most review units are

regarding to specific features of the product.

A product is always reviewed under a certain feature set  $F$ . Suppose we have got a lexical list  $O$  which includes all the opinion expressions and their sentiment polarities. For the feature-level opinion mining, identifying the sentiment association between  $F$  and  $O$  is essential. The key points in the whole process are as follows:

- get opinion word set  $O$  (with polarity labels)
- get product feature set  $F$
- identify relationships between  $F$  and  $O$

The focus of the paper is on the latter two steps, especially for the case of hidden sentiment association that the review features are implicit in the review sentences, but can be deduced by the opinion words in its context. In contrast to existing explicit adjacency approaches, the proposed approach detects the sentiment association between  $F$  and  $O$  based on review feature categories and opinion word groups gained from the review corpus.

To this end, we first consider two sets of association objects: the set of product feature words  $F = \{f_1, f_2, \dots, f_m\}$  and the set of opinion words  $O = \{o_1, o_2, \dots, o_n\}$ . A weighted bipartite graph from  $F$  and  $O$  can be built, denoted by  $G = \{F, O, R\}$ . Here  $R = [r_{ij}]$  is the  $m \times n$  link weight matrix containing all the pair-wise weights between set  $F$  and  $O$ . The weight can be calculated with different weighting schemes, in this paper, we set  $r_{ij}$  by the co-appearance frequency of  $f_i$  and  $o_j$  in clause level.

We take  $F$  and  $O$  as two random variables, and the question of constructing or clustering the object groups can be defined as finding compressed representation of each variable that reserves the information about another variable as high as possible. Take  $F$  as an example, we want to find its compression, denoted as  $C$ , such that the mutual information between  $C$  and  $O$  is as high as possible, under a constraint on the mutual information between  $F$  and  $C$ .

We propose an iterative reinforcement framework to deal with the tasks. An improved information bottleneck algorithm is employed in this framework, which will be illustrated in detail in the following sections.

## 2.2 Information Bottleneck Algorithm

The information bottleneck method (IB) was presented by Tishby et al. (1999). According to Shannon's information theory (Cover and Thomas,

1991), for the two random variables  $X, Y$ , the mutual information  $I(X;Y)$  between the random variables  $X, Y$  is given by the symmetric functional:

$$I(X;Y) = \sum_{x \in X, y \in Y} p(x)p(y|x) \log \frac{p(y|x)}{p(y)} \quad (1)$$

and the mutual information between them measures the relative entropy between their joint distribution  $p(x, y)$  and the product of respective marginal distributions  $p(x)p(y)$ , which is the only consistent statistical measure of the information that variable  $X$  contains about variable  $Y$  (and vice versa). Roughly speaking, some of the mutual information will be lost in the process of compression, e.g.  $I(C,Y) \leq I(X,Y)$  ( $C$  is a compressed representation of  $X$ ).

This representation is defined through a (possibly stochastic) mapping between each value  $x \in X$  to each representative value  $c \in C$ . Formally, this mapping can be characterized by a conditional distribution  $p(c|x)$ , inducing a soft partitioning of  $X$  values. Specifically, each value of  $X$  is associated with all the codebook elements ( $C$  values), with some normalized probability.

The IB method is based on the following simple idea. Given the empirical joint distribution of two variables, one variable is compressed so that the mutual information about the other variable is preserved as much as possible. The method can be considered as finding a minimal sufficient partition or efficient relevant coding of one variable with respect to the other one. This problem can be solved by introducing a Lagrange multiplier  $\beta$ , and then minimizing the functional:

$$L[p(c|x)] = I(C, X) - \beta I(C, Y) \quad (2)$$

This solution is given in terms of the three distributions that characterize every cluster  $c \in C$ , the prior probability for this cluster,  $p(c)$ , its membership probabilities  $p(c|x)$ , and its distribution over the relevance variable  $p(y|c)$ . In general, the membership probabilities,  $p(c|x)$  is "soft", i.e. every  $x \in X$  can be assigned to every  $c \in C$  in some (normalized) probability. The information bottleneck principle determines the distortion measure between the points  $x$  and  $c$  to be the  $D_{KL}[p(y|x) || p(y|c)] = \sum_y p(y|x) \log \frac{p(y|x)}{p(y|c)}$ , the

Kullback-Leibler divergence (Cover and Thomas, 1991) between the conditional distributions  $p(y|x)$

and  $p(y|c)$ . Specifically, the formal optimal solution is given by the following equations which must be solved together.

$$\begin{cases} p(c|x) = \frac{p(c)}{Z(\beta, x)} \exp(-\beta D_{KL}[p(y|x) \| p(y|c)]) \\ p(y|c) = \frac{1}{p(c)} \sum_x p(c|x) p(x) p(y|x) \\ p(c) = \sum_x p(c|x) p(x) \end{cases} \quad (3)$$

Where  $Z(\beta, x)$  is a normalization factor, and the single positive (Lagrange) parameter  $\beta$  determines the ‘‘softness’’ of the classification. Intuitively, in this procedure the information contained in  $X$  about  $Y$  ‘squeezed’ through a compact ‘bottleneck’ of clusters  $C$ , that is forced to represent the ‘relevant’ part in  $X$  w.r.t to  $Y$ .

An important special case is the ‘‘hard’’ clustering case where  $C$  is a deterministic function of  $X$ . That is,  $p(c|x)$  can only take values of zero or one. This restriction, which corresponds to the limit  $\beta \rightarrow \infty$  in Eqs 3 meaning every  $x \in X$  is assigned to exactly one cluster  $c \in C$  with a probability of one and to all the others with a probability of zero. This yields a natural distance measure between distributions which can be easily implemented in an agglomerative hierarchical clustering procedure (Slonim and Tishby, 1999).

$$\begin{cases} p(c|x) = \begin{cases} 1, & \text{if } x \in c \\ 0, & \text{otherwise} \end{cases} \\ p(y|c) = \frac{1}{p(c)} \sum_{x \in c} p(x, y) \\ p(c) = \sum_{x \in c} p(x) \end{cases} \quad (4)$$

The algorithm starts with a trivial partitioning into  $|X|$  singleton clusters, where each cluster contains exactly one element of  $X$ . At each step we merge two components of the current partition into a single new component in a way that locally minimizes the loss of mutual information about the categories. Every merger,  $(c_i, c_j) \Rightarrow c_*$ , is formally defined by the following equation:

$$\begin{cases} p(c_*|x) = \begin{cases} 1, & x \in c_i \text{ or } x \in c_j \\ 0, & \text{otherwise} \end{cases} \\ p(y|c_*) = \frac{p(c_i)}{p(c_*)} p(y|c_i) + \frac{p(c_j)}{p(c_*)} p(y|c_j) \\ p(c_*) = p(c_i) + p(c_j) \end{cases} \quad (5)$$

The decrease in the mutual information  $I(C, Y)$  due to this merger is defined by

$$\delta I(c_i, c_j) \equiv I(C_{\text{before}}, Y) - I(C_{\text{after}}, Y) \quad (6)$$

When  $I(C_{\text{before}}, Y)$  and  $I(C_{\text{after}}, Y)$  are the information values before and after the merger, respectively. After a little algebra, one can see

$$\delta I(c_i, c_j) \equiv (p(c_i) + p(c_j)) \cdot D_{JS}[p(y|c_i), p(y|c_j)] \quad (7)$$

Where the functional  $D_{JS}$  is the Jensen-Shannon divergence (Lin, 1991), defined as

$$D_{JS}[p_i, p_j] = \pi_i D_{KL}[p_i \| \hat{p}] + \pi_j D_{KL}[p_j \| \hat{p}] \quad (8)$$

where in our case

$$\begin{cases} \{p_i, p_j\} \equiv \{p(y|c_i), p(y|c_j)\} \\ \{\pi_i, \pi_j\} \equiv \left\{ \frac{p(c_i)}{p(c_*)}, \frac{p(c_j)}{p(c_*)} \right\} \\ \hat{p} = \pi_i p(y|c_i) + \pi_j p(y|c_j) \end{cases} \quad (9)$$

By introducing the information optimization criterion the resulting similarity measure directly emerges from the analysis. The algorithm is now very simple. At each step we perform ‘‘the best possible merger’’, i.e. merge the clusters  $\{c_i, c_j\}$  which minimize  $\delta I(c_i, c_j)$ .

### 2.3 Improved Information Bottleneck Algorithm for Semantic Information

In traditional information bottleneck approach, the distance between two data objects is measured by the difference of information values before and after the merger, which is measured by the Jensen-Shannon divergence. This divergence aims to measure the divergence between two probability distributions. For our case, the divergence is based on the co-occurrence information between the two variables  $F$  and  $O$ .

While the co-occurrence in corpus is usually quantitatively sparse; additionally, Statistics based

on word-occurrence loses semantic related information. To avoid such reversed effects, in the proposed framework we combine the co-occurrence information and semantic information as the final distance between the two types of objects.

$$D(X_i, X_j) = \alpha D_{semantic}(X_i, X_j) + (1 - \alpha) \delta I(X_i, X_j)$$

where  $\{X_i \in F \wedge X_j \in F\} \vee \{X_i \in O \wedge X_j \in O\}$

(10)

In equation 10, the distance between two data objects  $X_i$  and  $X_j$  is denoted as a linear combination of semantic distance and information value difference. The parameter  $\alpha$  reflects the contribution of different distances to the final distance.

---

**Input:** Joint probability distribution  $p(f, o)$

**Output:** A partition of  $F$  into  $m$  clusters,  $\forall m \in \{1, \dots, |F|\}$ , and a partition of  $O$  into  $n$  clusters  $\forall n \in \{1, \dots, |O|\}$

1.  $t \leftarrow 0$
  2. Repeat
    - a. Construct  $CF^t \leftarrow F^t$
    - b.  $\forall i, j=1, \dots, |CF^t|, i < j$ , calculate
$$d_{ij}^t \leftarrow \alpha D_{semantic}(cf_i^t, cf_j^t) + (1 - \alpha) \delta I(cf_i^t, cf_j^t)$$
    - c. for  $m \leftarrow |CF^t| - 1$  to 1
      - 1) find the indices  $\{i, j\}$ , for which  $d_{ij}^t$  is minimized
      - 2) merge  $\{cf_i^t, cf_j^t\}$  into  $cf_*^t$
      - 3) update  $CF^t \leftarrow \{CF^t - \{cf_i^t, cf_j^t\}\} \cup \{cf_*^t\}$
      - 4) update  $d_{ij}^t$  costs w.r.t  $cf_*^t$
    - d. Construct  $CO^t \leftarrow O^t$
    - e.  $\forall i, j=1, \dots, |CO^t|, i < j$ , calculate
$$d_{ij}^t \leftarrow \alpha D_{semantic}(co_i^t, co_j^t) + (1 - \alpha) \delta I(co_i^t, co_j^t)$$
    - f. for  $n \leftarrow |CO^t| - 1$  to 1
      - 1) find the indices  $\{i, j\}$ , for which  $d_{ij}^t$  is minimized
      - 2) merge  $\{co_i^t, co_j^t\}$  into  $co_*^t$
      - 3) update  $CO^t \leftarrow \{CO^t - \{co_i^t, co_j^t\}\} \cup \{co_*^t\}$
      - 4) update  $d_{ij}^t$  costs w.r.t  $co_*^t$
    - g.  $t \leftarrow t + 1$
  3. until  $(CF^t = CF^{t-1} \text{ and } CO^t = CO^{t-1})$
- 

**Figure 1: Pseudo-code of semantic information bottleneck in iterative reinforcement framework**

The semantic distance can be got by the usage of lexicon, such as WordNet (Budanitsky and Hirst,

2006). In this paper, we use the Chinese lexicon HowNet<sup>1</sup>.

The basic idea of the iterative reinforcement principle is to propagate the clustered results between different type data objects by updating their inter-relationship spaces. The clustering process can begin from an arbitrary type of data object. The clustering results of one data object type update the interrelationship thus reinforce the data object categorization of another type. The process is iterative until clustering results of both object types converge. Suppose we begin the clustering process from data objects in set  $F$ , and then the steps can be expressed as Figure 1. After the iteration, we can get the strongest  $n$  links between product feature categories and opinion word groups. That constitutes our set of sentiment association.

### 3 Experimental Setup

In this section we describe our experiments and the data used in these experiments.

#### 3.1 Data

Our experiments take hotel reviews (in Chinese) as example. The corpus used in the experiments is composed by 4000 editor reviews on hotel, including 857,692 Chinese characters. They are extracted from [www.ctrip.com](http://www.ctrip.com). Each review contains a user's rating represented by "stars", the number of the star denotes the user's satisfaction. The detailed information is illustrated in Table 1,

**Table 1: The detail information of corpus**

User's rating	Number
1 star	555
2 star	1375
3 star	70
4 star	2000

Then we use ICTCLAS<sup>2</sup>, a Chinese word segmentation software to extract candidate review features and opinion words.

Usually, adjectives are normally used to express opinions in reviews. Therefore, most of the existing researches take adjectives as opinion words. In the research of Hu et al. (2004), they proposed that

<sup>1</sup> <http://www.keenage.com/>

<sup>2</sup> [www.searchforum.org.cn](http://www.searchforum.org.cn)

other components of a sentence are unlikely to be product features except for nouns and noun phrases. Some researchers (Fujii and Ishikawa, 2006) targeted nouns, noun phrases and verb phrases. The adding of verb phrases caused the identification of more possible product features, while brought lots of noises. So in this paper, we follow the points of Hu’s, extracting nouns and noun phrases as candidate product feature words.

Take the whole set of nouns and noun phrases as candidate features will bring some noise. In order to reduce such adverse effects, we use the function of Named Entity Recognition (NER) in ICTCLAS to filter out named entities, including: person, location, organization. Since the NEs have small probability of being product features, we prune the candidate nouns or noun phrases which have the above NE taggers.

**Table 2: The number of candidate review features and opinion words in our corpus**

Extracted Instance	Total	Non-Repeated
Candidate review feature	86,623	15,249
Opinion word	26,721	1,231

By pruning candidate product feature words, we get the set of product feature words  $F$ . And the set of opinion words  $O$  is composed by all the adjectives in reviews. The number of candidate product feature words and opinion words extracted from the corpus are shown as Table 2:

### 3.2 Experimental Procedure

We evaluate our approach from two perspectives:

- 1) Effectiveness of product feature category construction by mutual reinforcement based clustering;
- 2) Precision of sentiment association between product feature categories and opinion word groups;

## 4 Experimental Results and Discussion

### 4.1 Evaluation of Review Feature Category Construction

To calculate agreement between the review feature category construction results and the correct labels, we make use of the Rand index (Rand, 1971). This

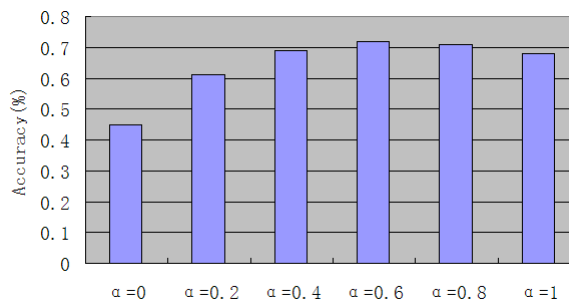
allows for a measure of agreement between two partitions,  $P_1$  and  $P_2$ , of the same data set  $D$ . Each partition is viewed as a collection of  $n*(n-1)/2$  pair wise decisions, where  $n$  is the size of  $D$ . For each pair of points  $d_i$  and  $d_j$  in  $D$ ,  $P_i$  either assigns them to the same cluster or to different clusters. Let  $a$  be the number of decisions where  $d_i$  is in the same cluster as  $d_j$  in  $P_1$  and in  $P_2$ . Let  $b$  be the number of decisions where the two instances are placed in different clusters in both partitions. Total agreement can then be calculated using

$$Rand(P_1, P_2) = \frac{a + b}{n(n-1)/2} \quad (11)$$

In our case, the parts of product feature words in the pre-constructed evaluation set are used to represent the data set  $D$ ;  $a$  and  $b$  represent the partition agreements between the pairs of any two words in the parts and in the clustering results respectively.

In equation 10, the parameter  $\alpha$  reflects the respective contribution of semantic information and co-occurrence information to the final distance. When  $\alpha = 0$  or  $\alpha = 1$ , the co-occurrence information or the semantic information will be utilized alone.

In order to get the optimal combination of the two type of distance, we adjust the parameter  $\alpha$  from 0 to 1 (stepped by 0.2), and the accuracy of feature category construction with different  $\alpha$  are shown in Figure 2:



**Figure 2: The accuracy of review feature category construction with the variation of the parameter  $\alpha$**

From this figure, we can find that the semantic information ( $\alpha=1$ ) contributes much more to the accuracy of review feature category construction than the co-occurrence information ( $\alpha=0$ ), and when  $\alpha=0$ , the approach is equivalent to the traditional information bottleneck approach. We consider this is due partly to the sparseness of the cor-

pus, by enlarging the scale of the corpus or using the search engine (e.g. google etc), we can get more accurate results.

Additionally, by a sensible adjust on the parameter  $\alpha$  (in this experiment, we set  $\alpha$  as 0.6), we can get higher accuracy than the two baselines ( $\alpha = 0$  and  $\alpha = 1$ ), which indicates the necessity and effectiveness of the integration of semantic information and co-occurrence information in the proposed approach.

#### 4.2 Evaluation of Sentiment Association

We use precision to evaluate the performance of sentiment association. An evaluation set is constructed manually first, in which there are not only the categories that every review feature word belong to, but also the relationship between each category and opinion word. Then we define precision as:

$$Precision = \frac{\text{number of correctly associated pairs}}{\text{number of detected pairs}} \quad (12)$$

A comparative result is got by the means of template-extraction based approach on the same test set. By the usage of regular expression, the nouns (phrase) and gerund (phrase) are extracted as the review features, and the nearest adjectives are extracted as the related opinion words. Because the modifiers of adjectives in reviews also contain rich sentiment information and express the view of customs, we extract adjectives and their modifiers simultaneously, and take them as the opinion words.

**Table 3: Performance comparison in sentiment association**

Approach	Pairs	Precision
Template extraction	27,683	65.89%
Proposed approach	141,826	78.90%

Table 3 shows the advantage of our approach over the extraction by explicit adjacency. Using the same product feature categorization, our sentiment association approach get a more accurate pair set than the direct extraction based on explicit adjacency. The precision we obtained by the iterative reinforcement approach is 78.90%, almost 13 points higher than the adjacency approach. This indicates that there are a large number of hidden sentiment associations in the real custom reviews,

which underlines the importance and value of our work.

## 5 Conclusions and Further Work

In this paper, we propose a novel iterative reinforcement framework based on improved information bottleneck approach to deal with the feature-level product opinion-mining problem. We alter traditional information bottleneck method by integration with semantic information, and the experimental result demonstrates the effectiveness of the alteration. The main contribution of our work mainly including:

- We propose an iterative reinforcement information bottleneck framework, and in this framework, review feature words and opinion words are organized into categories in a simultaneous and iterative manner.
- In the process of clustering, the semantic information and the co-occurrence information are integrated.
- The experimental results based on real Chinese web reviews demonstrate that our method outperforms the template extraction based algorithm.

Although our methods for candidate product feature extraction and filtering (see in 3.1) can partly identify real product features, it may lose some data and remain some noises. We'll conduct deeper research in this area in future work. Additionally, we plan to exploit more information, such as background knowledge, to improve the performance.

## 6 Acknowledgments

This work was mainly supported by two funds, i.e., 0704021000 and 60803085, and one another project, i.e., 2004CB318109.

## References

- A. Andreevskaia, S. Bergler. When Specialists and Generalists Work Together: Overcoming Domain Dependence in Sentiment Tagging. ACL 2008.
- A. Budanitsky and G. Hirst. Evaluating wordnetbased measures of lexical semantic relatedness. Computational Linguistics, 32(1):13–47, 2006.

- A. Devitt, K. Ahmad. Sentiment Polarity Identification in Financial News: A Cohesion-based Approach. ACL 2007.
- A. Fujii and T. Ishikawa. A system for summarizing and visualizing arguments in subjective documents: Toward supporting decision making. The Workshop on Sentiment and Subjectivity in Text ACL2006. 2006.
- A. Popescu and O. Etzioni. Extracting product features and opinions from reviews. HLT-EMNLP 2005.
- B. Liu, M. Hu, and J. Cheng. Opinion observer: analyzing and comparing opinions on the web. WWW 2005.
- B. Pang and L. Lee. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. ACL 2005.
- B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up? Sentiment classification using machine learning techniques. EMNLP 2002.
- B. Philip, T. Hastie, C. Manning, and S. Vaithyanathan. Exploring sentiment summarization. In AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications (AAAI tech report SS-04-07). 2004.
- B. Wang, H. Wang. Bootstrapping Both Product Features and Opinion Words from Chinese Customer Reviews with Cross-Inducing. IJCNLP 2008.
- D. Kushal, S. Lawrence, and D. Pennock. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. WWW 2003.
- H. Kanayama, T. Nasukawa. Fully Automatic Lexicon Expansion for Domain-oriented Sentiment Analysis. EMNLP 2006
- H. Takamura, T. Inui. Extracting Semantic Orientations of Phrases from Dictionary. NAACL-HLT 2007.
- I. Titov, R. McDonald. Modeling online reviews with multi-grain topic models. WWW 2008.
- L. Ku, Y. Liang and H. Chen. Opinion Extraction, Summarization and Tracking in News and Blog Corpora. AAAI-CAAW 2006.
- J. Blitzer, M. Dredze, F. Pereira. Biographies, Bollywood, Boom-boxes and Blenders: Domain Adaptation for Sentiment Classification. ACL 2007.
- J. Lin. Divergence Measures Based on the Shannon Entropy. IEEE Transactions on Information theory, 37(1):145–151, 1991.
- K. Bloom and N. Garg and S. Argamon. Extracting Appraisal Expressions. NAACL-HLT 2007.
- M. Hu and B. Liu. Mining and summarizing customer reviews. KDD 2004.
- M. Hu and B. Liu. Mining opinion features in customer reviews. AAAI 2004.
- N. Slonim, N. Tishby. Agglomerative information bottleneck. NIPS 1999.
- N. Slonim and N. Tishby. Document Clustering Using word Clusters via the Information Bottleneck Method. SIGIR 2000.
- N. Slonim and N. Tishby. The power of word clusters for text classification. ECIR 2001.
- N. Tishby, F. Pereira, W. Bialek. The information bottleneck method. 1999, arXiv: physics/0004057v1
- P. Turney. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. ACL 2002.
- P. Turney and M. Littman. Measuring Praise and Criticism: Inference of Semantic Orientation from Association. ACM Transactions on Information Systems, 2003,21(4): 315-346.
- Q. Su, X. Xu, H. Guo, Z. Guo, X. Wu, X. Zhang, B. Swen and Z. Su. Hidden sentiment association in Chinese web opinion mining. WWW 2008.
- S. Tan, G. Wu, H. Tang and X. Cheng. A novel scheme for domain-transfer problem in the context of sentiment analysis. CIKM 2007.
- S. Tan, Y. Wang, G. Wu and X. Cheng. Using unlabeled data to handle domain-transfer problem of semantic detection. SAC 2008.
- S. Tan, X. Cheng, Y. Wang and H. Xu. Adapting Naive Bayes to Domain Adaptation for Sentiment Analysis. ECIR 2009.
- T. Cover and J. Thomas. Elements of Information Theory. John Wiley & Sons, New York, 1991.
- T. Zagibalov, J. Carroll. Automatic Seed Word Selection for Unsupervised Sentiment Classification of Chinese Text. Coling 2008.
- V. Hatzivassiloglou and K. McKeown. Predicting the semantic orientation of adjectives. ACL 1997.
- V. Hatzivassiloglou and J. Wiebe. Effects of adjective orientation and gradability on sentence subjectivity. Coling 2000.
- W. Rand. Objective criteria for the evaluation of clustering methods. Journal of the American Statistical Association, 66, 846-850. 1971