

# Employing Personal/Impersonal Views in Supervised and Semi-supervised Sentiment Classification

Shoushan Li<sup>†‡</sup> Chu-Ren Huang<sup>†</sup> Guodong Zhou<sup>‡</sup> Sophia Yat Mei Lee<sup>†</sup>

<sup>†</sup>Department of Chinese and Bilingual  
Studies  
The Hong Kong Polytechnic University  
{shoushan.li, churenhuang,  
sophiaym}@gmail.com

<sup>‡</sup>Natural Language Processing Lab  
School of Computer Science and Technology  
Soochow University, China  
gdzhou@suda.edu.cn

## Abstract

In this paper, we adopt two views, personal and impersonal views, and systematically employ them in both supervised and semi-supervised sentiment classification. Here, personal views consist of those sentences which directly express speaker's feeling and preference towards a target object while impersonal views focus on statements towards a target object for evaluation. To obtain them, an unsupervised mining approach is proposed. On this basis, an ensemble method and a co-training algorithm are explored to employ the two views in supervised and semi-supervised sentiment classification respectively. Experimental results across eight domains demonstrate the effectiveness of our proposed approach.

## 1 Introduction

As a special task of text classification, sentiment classification aims to classify a text according to the expressed sentimental polarities of opinions such as 'thumb up' or 'thumb down' on the movies (Pang et al., 2002). This task has recently received considerable interests in the Natural Language Processing (NLP) community due to its wide applications.

In general, the objective of sentiment classification can be represented as a kind of binary relation  $R$ , defined as an ordered triple  $(X, Y, G)$ , where  $X$  is an object set including different kinds of people (e.g. writers, reviewers, or users),  $Y$  is another object set including the target objects (e.g. products, events, or even some people), and  $G$  is a subset of the Cartesian product  $X \times Y$ . The concerned relation in sentiment classification is  $X$ 's evaluation on  $Y$ , such as 'thumb up', 'thumb down', 'favorable',

and 'unfavorable'. Such relation is usually expressed in text by stating the information involving either a person (one element in  $X$ ) or a target object itself (one element in  $Y$ ). The first type of statement called personal view, e.g. 'I am so happy with this book', contains  $X$ 's "subjective" feeling and preference towards a target object, which directly expresses sentimental evaluation. This kind of information is normally domain-independent and serves as highly relevant clues to sentiment classification. The latter type of statement called impersonal view, e.g. 'it is too small', contains  $Y$ 's "objective" (i.e. or at least criteria-based) evaluation of the target object. This kind of information tends to contain much domain-specific classification knowledge. Although such information is sometimes not as explicit as personal views in classifying the sentiment of a text, speaker's sentiment is usually implied by the evaluation result.

It is well-known that sentiment classification is very domain-specific (Blitzer et al., 2007), so it is critical to eliminate its dependence on a large-scale labeled data for its wide applications. Since the unlabeled data is ample and easy to collect, a successful semi-supervised sentiment classification system would significantly minimize the involvement of labor and time. Therefore, given the two different views mentioned above, one promising application is to adopt them in co-training algorithms, which has been proven to be an effective semi-supervised learning strategy of incorporating unlabeled data to further improve the classification performance (Zhu, 2005). In addition, we would show that personal/impersonal views are linguistically marked and mining them in text can be easily performed without special annotation.

In this paper, we systematically employ personal/impersonal views in supervised and semi-supervised sentiment classification. First, an unsupervised bootstrapping method is adopted to automatically separate one document into personal and impersonal views. Then, both views are employed in supervised sentiment classification via an ensemble of individual classifiers generated by each view. Finally, a co-training algorithm is proposed to incorporate unlabeled data for semi-supervised sentiment classification.

The remainder of this paper is organized as follows. Section 2 introduces the related work of sentiment classification. Section 3 presents our unsupervised approach for mining personal and impersonal views. Section 4 and Section 5 propose our supervised and semi-supervised methods on sentiment classification respectively. Experimental results are presented and analyzed in Section 6. Section 7 discusses on the differences between personal/impersonal and subjective/objective. Finally, Section 8 draws our conclusions and outlines the future work.

## 2 Related Work

Recently, a variety of studies have been reported on sentiment classification at different levels: word level (Esuli and Sebastiani, 2005), phrase level (Wilson et al., 2009), sentence level (Kim and Hovy, 2004; Liu et al., 2005), and document level (Turney, 2002; Pang et al., 2002). This paper focuses on the document-level sentiment classification. Generally, document-level sentiment classification methods can be categorized into three types: unsupervised, supervised, and semi-supervised.

Unsupervised methods involve deriving a sentiment classifier without any labeled documents. Most of previous work use a set of labeled sentiment words called seed words to perform unsupervised classification. Turney (2002) determines the sentiment orientation of a document by calculating point-wise mutual information between the words in the document and the seed words of ‘*excellent*’ and ‘*poor*’. Kennedy and Inkpen (2006) use a term-counting method with a set of seed words to determine the sentiment. Zagibalov and Carroll (2008) first propose a seed word selection approach and then apply the same term-counting method for Chinese sentiment classifications. These unsupervised approaches are believed to be domain-independent for sentiment classification.

Supervised methods consider sentiment classification as a standard classification problem in which labeled data in a domain are used to train a domain-specific classifier. Pang et al. (2002) are the first to apply supervised machine learning methods to sentiment classification. Subsequently, many other studies make efforts to improve the performance of machine learning-based classifiers by various means, such as using subjectivity summarization (Pang and Lee, 2004), seeking new superior textual features (Riloff et al., 2006), and employing document subcomponent information (McDonald et al., 2007). As far as the challenge of domain-dependency is concerned, Blitzer et al. (2007) present a domain adaptation approach for sentiment classification.

Semi-supervised methods combine unlabeled data with labeled training data (often small-scaled) to improve the models. Compared to the supervised and unsupervised methods, semi-supervised methods for sentiment classification are relatively new and have much less related studies. Dasgupta and Ng (2009) integrate various methods in semi-supervised sentiment classification including spectral clustering, active learning, transductive learning, and ensemble learning. They achieve a very impressive improvement across five domains. Wan (2009) applies a co-training method to semi-supervised learning with labeled English corpus and unlabeled Chinese corpus for Chinese sentiment classification.

## 3 Unsupervised Mining of Personal and Impersonal Views

As mentioned in Section 1, the objective of sentiment classification is to classify a specific binary relation:  $X$ 's evaluation on  $Y$ , where  $X$  is an object set including different kinds of persons and  $Y$  is another object set including the target objects to be evaluated. First of all, we focus on an analysis on sentences in product reviews regarding the two views: personal and impersonal views.

The personal view consists of personal sentences (i.e.  $X$ 's sentences) exemplified below:

- I. Personal preference:  
E1: *I love this breadmaker!*  
E2: *I disliked it from the beginning.*
- II. Personal emotion description:  
E3: *Very disappointed!*  
E4: *I am happy with the product.*
- III. Personal actions:

E5: *Do not waste your money.*

E6: *I have recommended this machine to all my friends.*

The impersonal view consists of impersonal sentences (i.e.  $Y$ 's sentences) exemplified below:

I. Impersonal feature description:

E7: *They are too thin to start with.*

E8: *This product is extremely quiet.*

II. Impersonal evaluation:

E9: *It's great.*

E10: *The product is a waste of time and money.*

III. Impersonal actions:

E11: *This product not even worth a penny.*

E12: *It broke down again and again.*

We find that the subject of a sentence presents important cues for personal/impersonal views, even though a formal and computable definition of this contrast cannot be found. Here, subject refers to one of the two main constituents in the traditional English grammar (the other constituent being the predicate) (Crystal, 2003)<sup>1</sup>. For example, the subjects in the above examples of E1, E7 and E11 are 'I', 'they', and 'this product' respectively. For automatic mining the two views, personal/impersonal sentences can be defined according to their subjects:

**Personal sentence:** the sentence whose subject is (or represents) a person.

**Impersonal sentence:** the sentence whose subject is not (does not represent) a person.

In this study, we mainly focus on product review classification where the target object in the set  $Y$  is not a person. The definitions need to be adjusted when the evaluation target itself is a person, e.g. the political sentiment classification by Durant and Smith (2007).

Our unsupervised mining approach for mining personal and impersonal sentences consists of two main steps. First, we extract an initial set of personal and impersonal sentences with some heuristic rules: If the first word of one sentence is (or implies) a personal pronoun including 'I', 'we', and 'do', then the sentence is extracted as a personal sentence; If the first word of one sentence is an impersonal pronoun including 'it', 'they', 'this', and 'these', then the sentence is extracted as an impersonal sentence. Second, we apply the classifier which is trained with the initial set of personal and impersonal sentences to classify the remaining sentences. This step aims to classify the sentences without pronouns

(e.g. E3). Figure 1 shows the unsupervised mining algorithm.

---

**Input:**

The training data  $D$

**Output:**

All personal and impersonal sentences, i.e. sentence sets  $S_{personal}$  and  $S_{impersonal}$ .

**Procedure:**

- (1). Segment all documents in  $D$  to sentences  $S$  using punctuations (such as periods and interrogation marks)
- (2). Apply the heuristic rules to classify the sentences  $S$  with proper pronouns into,  $S_{p1}$  and  $S_{i1}$
- (3). Train a binary classifier  $f_{p-i}$  with  $S_{p1}$  and  $S_{i1}$
- (4). Use  $f_{p-i}$  to classify the remaining sentences into  $S_{p2}$  and  $S_{i2}$
- (5).  $S_{personal} = S_{p1} \cup S_{p2}$ ,  $S_{impersonal} = S_{i1} \cup S_{i2}$

---

Figure 1: The algorithm for unsupervised mining personal and impersonal sentences from a training data

#### 4 Employing Personal/Impersonal Views in Supervised Sentiment Classification

After unsupervised mining of personal and impersonal sentences, the training data is divided into two views: the personal view, which contains personal sentences, and the impersonal view, which contains impersonal sentences. Obviously, these two views can be used to train two different classifiers,  $f_1$  and  $f_2$ , for sentiment classification respectively.

Since our mining approach is unsupervised, there inevitably exist some noises. In addition, the sentences of different views may share the same information for sentiment classification. For example, consider the following two sentences: 'It is a waste of money.' and 'Do not waste your money.' Apparently, the first one belongs to the impersonal view while the second one belongs to personal view, according to our heuristic rules. However, these two sentences share the same word, 'waste', which conveys strong negative sentiment information. This suggests that training a single-view classifier  $f_3$  with all sentences should help. Therefore, three base classifiers,  $f_1$ ,  $f_2$ , and  $f_3$ , are eventually derived from the personal view, the impersonal

---

<sup>1</sup> The subject has the grammatical function in a sentence of relating its constituent (a noun phrase) by means of the verb to any other elements present in the sentence, i.e. objects, complements, and adverbials.

view and the single view, respectively. Each base classifier provides not only the class label outputs but also some kinds of confidence measurements, e.g. posterior probabilities of the testing sample belonging to each class.

Formally, each base classifier  $f_l$  ( $l=1,2,3$ ) assigns a test sample (denoted as  $x_l$ ) a posterior probability vector  $\bar{P}(x_l)$ :

$$\bar{P}(x_l) = \langle p(c_1 | x_l), p(c_2 | x_l) \rangle^t$$

where  $p(c_l | x_l)$  denotes the probability that the  $l$ -th base classifier considers the sample belonging to  $c_l$ .

In the ensemble learning literature, various methods have been presented for combining base classifiers. The combining methods are categorized into two groups (Duin, 2002): fixed rules such as voting rule, product rule, and sum rule (Kittler et al., 1998), and trained rules such as weighted sum rule (Fumera and Roli, 2005) and meta-learning approaches (Vilalta and Drissi, 2002). In this study, we choose a fixed rule and a trained rule to combine the three base classifiers  $f_1$ ,  $f_2$ , and  $f_3$ .

The chosen fixed rule is product rule which combine base classifiers by multiplying the posterior possibilities and using the multiplied possibility for decision, i.e.

$$\text{assign } y \rightarrow c_j$$

$$\text{where } j = \arg \max_i \prod_{l=1}^3 p(c_l | x_l)$$

The chosen trained rule is stacking (Vilalta and Drissi, 2002; Džeroski and Ženko, 2004) where a meta-classifier is trained with the output of the base classifiers as the input. Formally, let  $x'$  denote a feature vector of a sample from the development data. The output of the  $l$ -th base classifier  $f_l$  on this sample is the probability distribution over the category set  $\{c_1, c_2\}$ , i.e.

$$\bar{P}(x'_l) = \langle p(c_1 | x'_l), p(c_2 | x'_l) \rangle$$

Then, a meta-classifier is trained using the development data with the meta-level feature vector  $x^{meta} \in R^{2 \times 3}$

$$x^{meta} = \langle \bar{P}(x'_{l=1}), \bar{P}(x'_{l=2}), \bar{P}(x'_{l=3}) \rangle$$

In our experiments, we perform stacking with 4-fold cross validation to generate meta-training data where each fold is used as the development data and the other three folds are used to train the base classifiers in the training phase.

## 5 Employing Personal/Impersonal Views in Semi-Supervised Sentiment Classification

Semi-supervised learning is a strategy which combines unlabeled data with labeled training data to improve the models. Given the two-view classifiers  $f_1$  and  $f_2$  along with the single-view classifier  $f_3$ , we perform a co-training algorithm for semi-supervised sentiment classification. The co-training algorithm is a specific semi-supervised learning approach which starts with a set of labeled data and increases the amount of labeled data using the unlabeled data by bootstrapping (Blum and Mitchell, 1998). Figure 2 shows the co-training algorithm in our semi-supervised sentiment classification.

---

### Input:

The labeled data  $L$  containing personal sentence set  $S_{L-personal}$  and impersonal sentence set  $S_{L-impersonal}$

The unlabeled data  $U$  containing personal sentence set  $S_{U-personal}$  and impersonal sentence set  $S_{U-impersonal}$

### Output:

New labeled data  $L$

### Procedure:

Loop for  $N$  iterations until  $U = \emptyset$

- (1). Learn the first classifier  $f_1$  with  $S_{L-personal}$
  - (2). Use  $f_1$  to label samples from  $U$  with  $S_{U-personal}$
  - (3). Choose  $n_1$  positive and  $n_1$  negative most confidently predicted samples  $A_1$
  - (4). Learn the second classifier  $f_2$  with  $S_{L-impersonal}$
  - (5). Use  $f_2$  to label samples from  $U$  with  $S_{U-impersonal}$
  - (6). Choose  $n_2$  positive and  $n_2$  negative most confidently predicted samples  $A_2$
  - (7). Learn the third classifier  $f_3$  with  $L$
  - (8). Use  $f_3$  to label samples from  $U$
  - (9). Choose  $n_3$  positive and  $n_3$  negative most confidently predicted samples  $A_3$
  - (10). Add samples  $A_1 \cup A_2 \cup A_3$  with the corresponding labels into  $L$
  - (11). Update  $S_{L-personal}$  and  $S_{L-impersonal}$
- 

Figure 2: Our co-training algorithm for semi-supervised sentiment classification

After obtaining the new labeled data, we can either adopt one classifier (i.e.  $f_3$ ) or a combined classifier (i.e.  $f_1 + f_2 + f_3$ ) in further training and testing. In our experimentation, we explore both of them with the former referred to as **co-training and single classifier** and the latter referred to as **co-training and combined classifier**.

## 6 Experimental Studies

We have systematically explored our method on product reviews from eight domains: book, DVD, electronic appliances, kitchen appliances, health, network, pet and software.

### 6.1 Experimental Setting

The product reviews on the first four domains (book, DVD, electronic, and kitchen appliances) come from the multi-domain sentiment classification corpus, collected from <http://www.amazon.com/> by Blitzer et al. (2007)<sup>2</sup>. Besides, we also collect the product views from <http://www.amazon.com/> on other four domains (health, network, pet and software)<sup>3</sup>. Each of the eight domains contains 1000 positive and 1000 negative reviews. Figure 3 gives the distribution of personal and impersonal sentences in the training data (75% labeled data of all data). It shows that there are more impersonal sentences than personal ones in each domain, in particular in the DVD domain, where the number of impersonal sentences is at least twice as many as that of personal sentences. This unusual phenomenon is mainly attributed to the fact that many objective descriptions, e.g. the movie plot introductions, are expressed in the DVD domain which makes the extracted personal and impersonal sentences rather unbalanced.

We apply both support vector machine (SVM) and Maximum Entropy (ME) algorithms with the help of the SVM-light<sup>4</sup> and Mallet<sup>5</sup> tools. All parameters are set to their default values. We find that ME performs slightly better than SVM on the average. Furthermore, ME offers posterior probability information which is required for

combination methods. Thus we apply the ME classification algorithm for further combination and co-training. In particular, we only employ Boolean features, representing the presence or absence of a word in a document. Finally, we perform  $t$ -test to evaluate the significance of the performance difference between two systems with different methods (Yang and Liu, 1999).

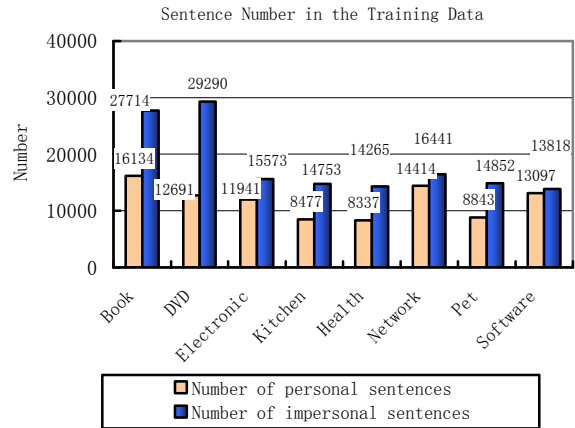


Figure 3: Distribution of personal and impersonal sentences in the training data of each domain

### 6.2 Experimental Results on Supervised Sentiment Classification

4-fold cross validation is performed for supervised sentiment classification. For comparison, we generate two random views by randomly splitting the whole feature space into two parts. Each part is seen as a view and used to train a classifier. The combination (two random view classifiers along with the single-view classifier  $f_3$ ) results are shown in the last column of Table 1. The comparison between random two views and our proposed two views will clarify whether the performance gain comes truly from our proposed two-view mining, or simply from using the classifier combination strategy.

Table 1 shows the performances of different classifiers, where the single-view classifier  $f_3$  which uses all sentences for training and testing, is considered as our baseline. Note that the baseline performances of the first four domains are worse than the ones reported in Blitzer et al. (2007). But their experiment is performed with only one split on the data with 80% as the training data and 20% as the testing data, which means the size of their training data is larger than ours. Also, we find that our performances are similar to the ones (described as fully supervised results) reported in Dasgupta and Ng (2009) where the same data in the four domains are used and 10-fold cross validation is performed.

<sup>2</sup> <http://www.seas.upenn.edu/~mdredze/datasets/sentiment/>

<sup>3</sup> Note that the second version of multi-domain sentiment classification corpus does contain data from many other domains. However, we find that the reviews in the other domains contain many duplicated samples. Therefore, we re-collect the reviews from <http://www.amazon.com/> and filter those duplicated ones. The new collection is here:

[http://lft.cbs.polyu.edu.hk/~lss/ACL2010\\_Data\\_SSli.zip](http://lft.cbs.polyu.edu.hk/~lss/ACL2010_Data_SSli.zip)

<sup>4</sup> <http://svmlight.joachims.org/>

<sup>5</sup> <http://mallet.cs.umass.edu/>

Domain	Personal View Classifier $f_1$	Impersonal View Classifier $f_2$	Single View Classifier (baseline) $f_3$	Combination (Stacking) $f_1 + f_2 + f_3$	Combination (Product rule) $f_1 + f_2 + f_3$	Combination with two random views (Product rule)
Book	0.7004	0.7474	0.7654	0.7919	<b>0.7949</b>	0.7546
DVD	0.6931	0.7663	0.7884	0.8079	<b>0.8165</b>	0.8054
Electronic	0.7414	0.7844	0.8074	0.8304	<b>0.8364</b>	0.8210
Kitchen	0.7430	0.8030	0.8290	0.8555	<b>0.8565</b>	0.8152
Health	0.7000	0.7370	0.7559	0.7780	<b>0.7815</b>	0.7548
Network	0.7655	0.7710	0.8265	0.8360	<b>0.8435</b>	0.8312
Pet	0.6940	0.7145	0.7390	0.7565	<b>0.7665</b>	0.7423
Software	0.7035	0.7205	0.7470	<b>0.7730</b>	0.7715	0.7615
<b>AVERAGE</b>	<b>0.7176</b>	<b>0.7555</b>	<b>0.7823</b>	<b>0.8037</b>	<b>0.8084</b>	<b>0.7858</b>

Table 1: Performance of supervised sentiment classification

From Table 1, we can see that impersonal view classifier  $f_1$  consistently performs better than personal view classifier  $f_2$ . Similar to the sentence distributions, the difference in the classification performances between these two views in the DVD domain is the largest (0.6931 vs. 0.7663).

Both the combination methods (stacking and product rule) significantly outperform the baseline in each domain (p-value<0.01) with a decent average performance improvement of 2.61%. Although the performance difference between the product rule and stacking is not significant, the product rule is obviously a better choice as it involves much easier implementation. Therefore, in the semi-supervised learning process, we only use the product rule to combine the individual classifiers. Finally, it shows that random generation of two views with the combination method of the product rule only slightly outperforms the baseline on the average (0.7858 vs. 0.7823) but performs much worse than our unsupervised mining of personal and impersonal views.

### 6.3 Experimental Results on Semi-supervised Sentiment Classification

We systematically evaluate and compare our two-view learning method with various semi-supervised ones as follows:

**Self-training**, which uses the unlabeled data in a bootstrapping way like co-training yet limits the number of classifiers and the number of views to one. Only the baseline classifier  $f_3$  is used to select most confident unlabeled samples in each iteration.

**Transductive SVM**, which seeks the largest separation between labeled and unlabeled data through regularization (Joachims, 1999). We implement it with the help of the SVM-light tool.

**Co-training with random two-view generation** (briefly called **co-training with random views**), where two views are generated by randomly splitting the whole feature space into two parts.

In semi-supervised sentiment classification, the data are randomly partitioned into labeled training data, unlabeled data, and testing data with the proportion of 10%, 70% and 20% respectively. Figure 4 reports the classification accuracies in all iterations, where **baseline** indicates the supervised classifier  $f_3$  trained on the 10% data; both **co-training and single classifier** and **co-training and combined classifier** refer to co-training using our proposed personal and impersonal views. But the former merely applies the baseline classifier  $f_3$  trained the new labeled data to test on the testing data while the latter applies the combined classifier  $f_1 + f_2 + f_3$ . In each iteration, two top-confident samples in each category are chosen, i.e.  $n_1 = n_2 = n_3 = 2$ . For clarity, results of other methods (e.g. **self-training, transductive SVM**) are not shown in Figure 4 but will be reported in Figure 5 later.

Figure 4 shows that **co-training and combined classifier** always outperforms **co-training and single classifier**. This again justifies the effectiveness of our two-view learning on supervised sentiment classification.

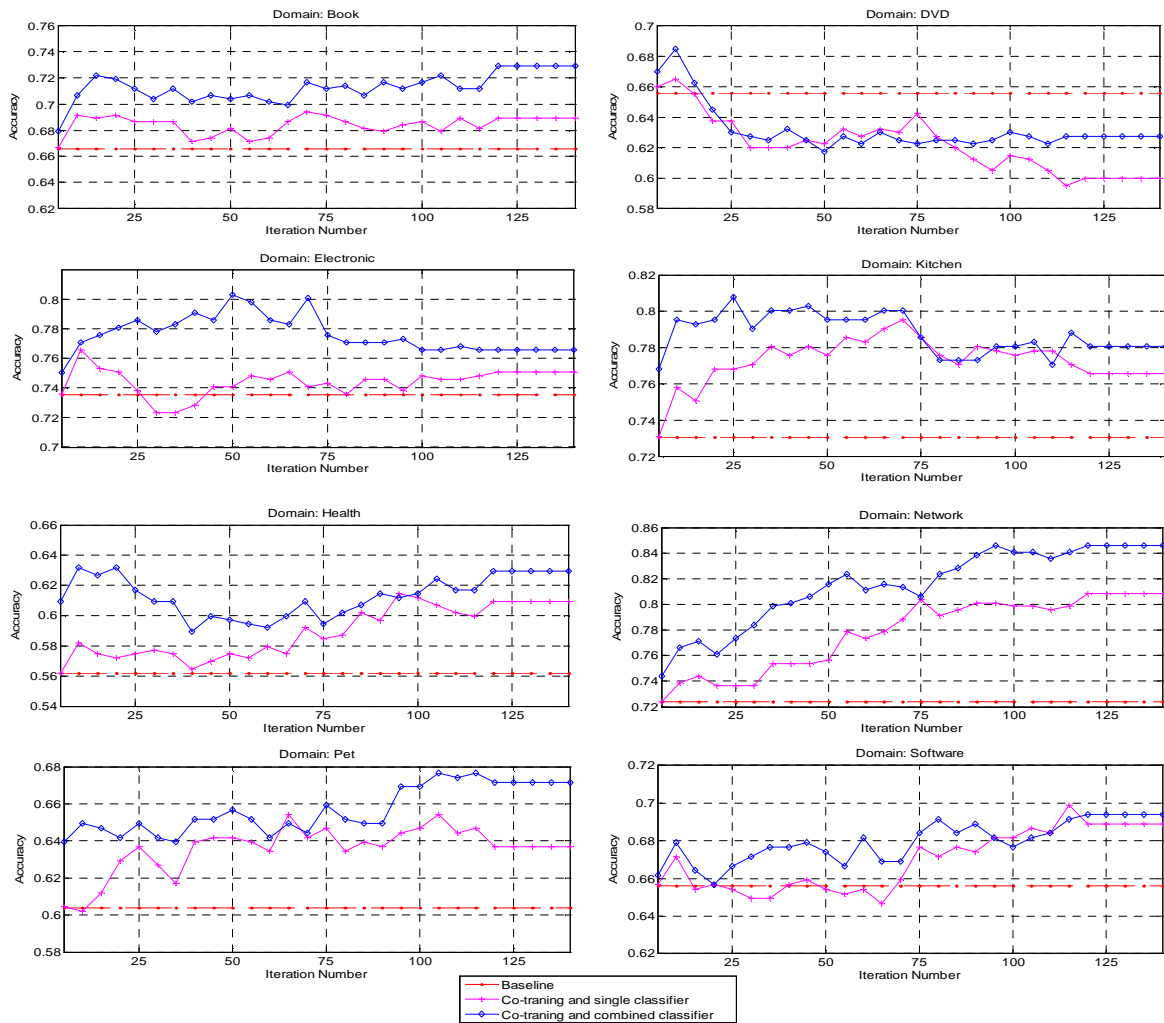


Figure 4: Classification performance vs. iteration numbers (using 10% labeled data as training data)

One open question is whether the unlabeled data improve the performance. Let us set aside the influence of the combination strategy and focus on the effectiveness of semi-supervised learning by comparing the baseline and **co-training and single classifier**. Figure 4 shows different results on different domains. Semi-supervised learning fails on the DVD domain while on the three domains of book, electronic, and software, semi-supervised learning benefits slightly ( $p\text{-value} > 0.05$ ). In contrast, semi-supervised learning benefits much on the other four domains (health, kitchen, network, and pet) from using unlabeled data and the performance improvements are statistically significant ( $p\text{-value} < 0.01$ ). Overall speaking, we think that the unlabeled data are very helpful as they lead to about 4% accuracy improvement on the average except for the DVD domain. Along with the supervised combination strategy, our approach can significantly improve the

performance more than 7% on the average compared to the baseline.

Figure 5 shows the classification results of different methods with different sizes of the labeled data: 5%, 10%, and 15% of all data, where the testing data are kept the same (20% of all data). Specifically, the results of other methods including **self-training**, **transductive SVM**, and **random views** are presented when 10% labeled data are used in training. It shows that **self-training** performs much worse than our approach and fails to improve the performance of five of the eight domains. **Transductive SVM** performs even worse and can only improve the performance of the “software” domain. Although **co-training with random views** outperforms the baseline on four of the eight domains, it performs worse than **co-training and single classifier**. This suggests that the impressive improvements are mainly due to our unsupervised two-view mining rather than the combination strategy.

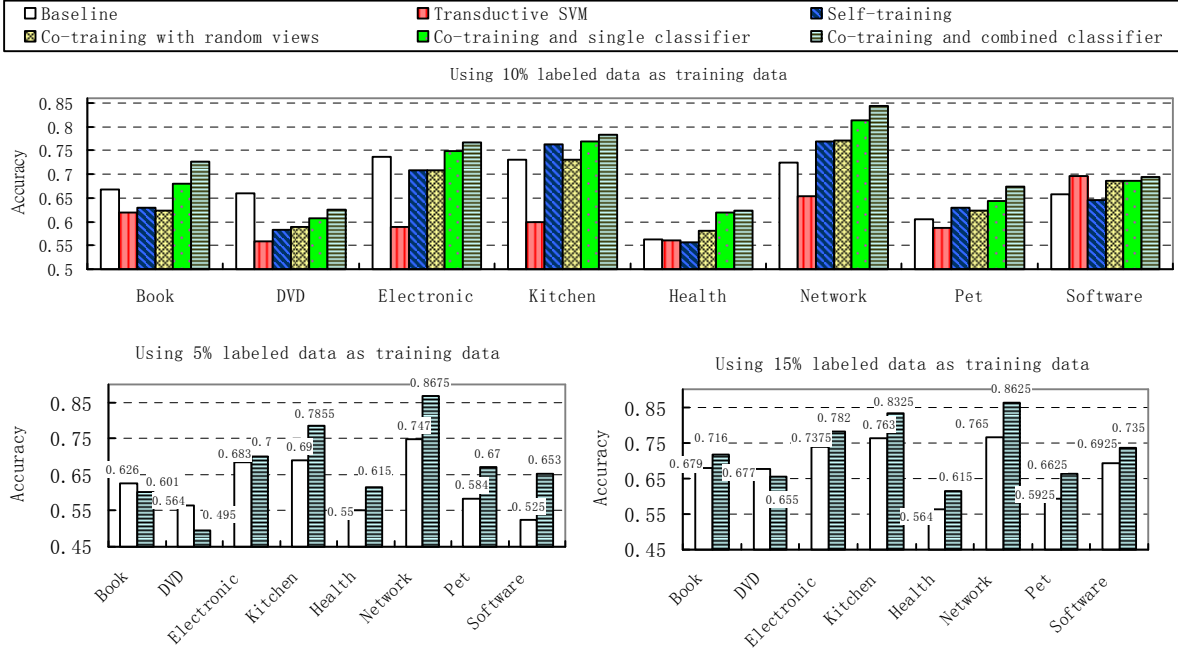


Figure 5: Performance of semi-supervised sentiment classification when 5%, 10%, and 15% labeled data are used

Figure 5 also shows that our approach is rather robust and achieves excellent performances in different training data sizes, although our approach fails on two domains, i.e. book and DVD, when only 5% of the labeled data are used. This failure may be due to that some of the samples in these two domains are too ambiguous and hard to classify. Manual checking shows that quite a lot of samples on these two domains are even too difficult for professionals to give a high-confident label. Another possible reason is that there exist too many objective descriptions in these two domains, thus introducing too much noisy information for semi-supervised learning.

The effectiveness of different sizes of chosen samples in each iteration is also evaluated like  $n_1 = n_2 = n_3 = 6$  and  $n_1 = 3, n_2 = n_3 = 6$  (This assignment is considered because the personal view classifier performs worse than the other two classifiers). Our experimental results are still unsuccessful in the DVD domain and do not show much difference on other domains. We also test the co-training approach without the single-view classifier  $f_3$ . Experimental results show that the inclusion of the single-view classifier  $f_3$  slightly helps the co-training approach. The detailed discussion of the results is omitted due to space limit.

#### 6.4 Why our approach is effective?

One main reason for the effectiveness of our approach on supervised learning is the way how

personal and impersonal views are dealt with. As personal and impersonal views have different ways of expressing opinions, splitting them into two separations can filter some classification noises. For example, in the sentence of “*I have seen amazing dancing, and good dancing. This was TERRIBLE dancing!*”. The first sentence is classified as a personal sentence and the second one is an impersonal sentence. Although the words ‘*amazing*’ and ‘*good*’ convey strong positive sentiment information, the whole text is negative. If we get the bag-of-words from the whole text, the classification result will be wrong. Rather, splitting the text into two parts based on different views allows correct classification as the personal view rarely contains impersonal words such as ‘*amazing*’ and ‘*good*’. The classification result will thus be influenced by the impersonal view.

In addition, a document may contain both personal and impersonal sentences, and each of them, to a certain extent, provides classification evidence. In fact, we randomly select 50 documents in the domain of kitchen appliances and find that 80% of the documents take both of them express explicit opinions. That is to say, the two views provide different, complementary information for classification. This qualifies the success requirement of co-training algorithm to some extent. This might be the reason for the effectiveness of our approach on semi-supervised learning.



## 7 Discussion on Personal/Impersonal vs. Subjective/Objective

As mentioned in Section 1, personal view contains  $X$ 's "subjective" feeling, and impersonal view contains  $Y$ 's "objective" (i.e. or at least criteria-based) evaluation of the target object. However, our technically-defined concepts of personal/impersonal are definitely different from subjective/objective: Personal view can certainly contain many objective expressions, e.g. 'I bought this electric kettle' and impersonal view can contain many subjective expressions, e.g. 'It is disappointing'.

Our technically-defined personal/impersonal views are two different ways to describe opinions. Personal sentences are often used to express opinions in a direct way and their target object should be one of  $X$ . Impersonal ones are often used to express opinions in an indirect way and their target object should be one of  $Y$ . The ideal definition of personal (or impersonal) view given in Section 1 is believed to be a subset of our technical definition of personal (or impersonal) view. Thus impersonal view may contain both  $Y$ 's objective evaluation (more likely to be domain independent) and subjective  $Y$ 's description.

In addition, simply splitting text into subjective/objective views is not particularly helpful. Since a piece of objective text provides rather limited implicit classification information, the classification abilities of the two views are very unbalanced. This makes the co-training process unfeasible. Therefore, we believe that our technically-defined personal/impersonal views are more suitable for two-view learning compared to subjective/objective views.

## 8 Conclusion and Future Work

In this paper, we propose a robust and effective two-view model for sentiment classification based on personal/impersonal views. Here, the personal view consists of subjective sentences whose subject is a person, whereas the impersonal view consists of objective sentences whose subject is not a person. Such views are lexically cued and can be obtained without pre-labeled data and thus we explore an unsupervised learning approach to mine them. Combination methods and a co-training algorithm are proposed to deal with supervised and semi-supervised sentiment classification respectively. Evaluation on product reviews from eight domains shows that our approach

significantly improves the performance across all eight domains on supervised sentiment classification and greatly outperforms the baseline with more than 7% accuracy improvement on the average across seven of eight domains (except the DVD domain) on semi-supervised sentiment classification.

In the future work, we will integrate the subjectivity summarization strategy (Pang and Lee, 2004) to help discard noisy objective sentences. Moreover, we need to consider the cases when both  $X$  and  $Y$  appear in a sentence. For example, the sentence "I think they're poor" should be an impersonal view but wrongly classified as a personal one according to our technical rules. We believe that these will help improve our approach and hopefully are applicable to the DVD domain. Another interesting and practical idea is to integrate active learning (Settles, 2009), another popular but principally different kind of semi-supervised learning approach, with our two-view learning approach to build high-performance systems with the least labeled data.

## Acknowledgments

The research work described in this paper has been partially supported by Start-up Grant for Newly Appointed Professors, No. 1-BBZM in the Hong Kong Polytechnic University and two NSFC grants, No. 60873150 and No. 90920004. We also thank the three anonymous reviewers for their invaluable comments.

## References

- Blitzer J., M. Dredze, and F. Pereira. 2007. Biographies, Bollywood, Boom-boxes and Blenders: Domain Adaptation for Sentiment Classification. In *Proceedings of ACL-07*.
- Blum A. and T. Mitchell. 1998. Combining labeled and unlabeled data with co-training. In *Proceedings of COLT-98*.
- Crystal D. 2003. *The Cambridge Encyclopedia of the English Language*. Cambridge University Press.
- Dasgupta S. and V. Ng. 2009. Mine the Easy and Classify the Hard: Experiments with Automatic Sentiment Classification. In *Proceedings of ACL-IJCNLP-09*.
- Duin R. 2002. The Combining Classifier: To Train Or Not To Train? In *Proceedings of 16th International Conference on Pattern Recognition (ICPR-02)*.
- Durant K. and M. Smith. 2007. Predicting the Political Sentiment of Web Log Posts using

- Supervised Machine Learning Techniques Coupled with Feature Selection. In *Processing of Advances in Web Mining and Web Usage Analysis*.
- Džeroski S. and B. Ženko. 2004. Is Combining Classifiers with Stacking Better than Selecting the Best One? *Machine Learning*, vol.54(3), pp.255-273, 2004.
- Esuli A. and F. Sebastiani. 2005. Determining the Semantic Orientation of Terms through Gloss Classification. In *Proceedings of CIKM-05*.
- Fumera G. and F. Roli. 2005. A Theoretical and Experimental Analysis of Linear Combiners for Multiple Classifier Systems. *IEEE Trans. PAMI*, vol.27, pp.942-956, 2005
- Joachims, T. 1999. Transductive Inference for Text Classification using Support Vector Machines. *ICML1999*.
- Kennedy A. and D. Inkpen. 2006. Sentiment Classification of Movie Reviews using Contextual Valence Shifters. *Computational Intelligence*, vol.22(2), pp.110-125, 2006.
- Kim S. and E. Hovy. 2004. Determining the Sentiment of Opinions. In *Proceedings of COLING-04*.
- Kittler J., M. Hatef, R. Duin, and J. Matas. 1998. On Combining Classifiers. *IEEE Trans. PAMI*, vol.20, pp.226-239, 1998
- Liu B., M. Hu, and J. Cheng. 2005. Opinion Observer: Analyzing and Comparing Opinions on the Web. In *Proceedings of WWW-05*.
- McDonald R., K. Hannan, T. Neylon, M. Wells, and J. Reynar. 2007. Structured Models for Fine-to-coarse Sentiment Analysis. In *Proceedings of ACL-07*.
- Pang B. and L. Lee. 2004. A Sentimental Education: Sentiment Analysis using Subjectivity Summarization based on Minimum Cuts. In *Proceedings of ACL-04*.
- Pang B., L. Lee, and S. Vaithyanathan. 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques. In *Proceedings of EMNLP-02*.
- Riloff E., S. Patwardhan, and J. Wiebe. 2006. Feature Subsumption for Opinion Analysis. In *Proceedings of EMNLP-06*.
- Settles B. 2009. Active Learning Literature Survey. *Technical Report 1648*, Department of Computer Sciences, University of Wisconsin at Madison, Wisconsin.
- Turney P. 2002. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In *Proceedings of ACL-02*.
- Vilalta R. and Y. Drissi. 2002. A Perspective View and Survey of Meta-learning. *Artificial Intelligence Review*, 18(2): 77-95.
- Wan X. 2009. Co-Training for Cross-Lingual Sentiment Classification. In *Proceedings of ACL-IJCNLP-09*.
- Wilson T., J. Wiebe, and P. Hoffmann. 2009. Recognizing Contextual Polarity: An Exploration of Features for Phrase-Level Sentiment Analysis. *Computational Linguistics*, vol.35(3), pp.399-433, 2009.
- Yang Y. and X. Liu. 1999. A Re-Examination of Text Categorization methods. In *Proceedings of SIGIR-99*.
- Zagibalov T. and J. Carroll. 2008. Automatic Seed Word Selection for Unsupervised Sentiment Classification of Chinese Text. In *Proceedings of COLING-08*.
- Zhu X. 2005. Semi-supervised Learning Literature Survey. *Technical Report Computer Sciences 1530*, University of Wisconsin – Madison.