Meaning Matters: Senses of Words are More Informative than Words for Cross-domain Sentiment Analysis

Raksha Sharma, Sudha Bhingardive, Pushpak Bhattacharyya
Dept. of Computer Science and Engineering
IIT Bombay, Mumbai, India
{raksha,sudha,pb}@cse.iitb.ac.in

Abstract

Getting labeled data in each domain is always an expensive and a time consuming task. Hence, cross-domain sentiment analysis has emerged as a demanding research area where a labeled source domain facilitates classifier for an unlabeled target domain. However, cross-domain sentiment analysis is still a challenging task because of the differences across domains. A word which is used with positive polarity in the source domain may bear negative polarity in the target domain or vice versa. In addition, a word which is used very scarcely in the source domain may have high impact in the target domain for sentiment classification. Due to these differences across domains cross-domain sentiment analysis suffers from negative transfer. In this paper, we propose that senses of words in place of words help to overcome the differences across domains, which in turn leads to improvement in cross-domain sentiment classification accuracy. Results show that senses of words provide a better sentiment classifier in the unlabeled target domain in comparison to words for 12 pairs of source and target domains.

1 Introduction

Generally users do not explicitly indicate sentiment orientation (positive or negative) of the reviews posted by them on the Web, it needs to be predicted from the text, which has led to plethora of work in the field of Sentiment Analysis (SA)(Esuli and Sebastiani, 2005; Breck et al., 2007; Li et al., 2009; Prabowo and Thelwall, 2009; Taboada et al., 2011; Cambria et al., 2013; Rosenthal et al., 2014). Most of the proposed techniques for sentiment analysis are based on the availability of labeled train data considering that train data and test data belong to the same domain. Easy access of Internet has made the users to post their experiences with any product, service or application very frequently. Consequently, there is a high increase in number of domains in which sentimental data is available. Getting labeled data manually in each domain is not feasible due to the cost incurred in annotation process.

Cross-domain SA provides a solution to build a classifier in the unlabeled target domain from a labeled source domain. But, due to the differences across domains, cross-domain SA suffers from negative transfer. A word which is used with positive polarity in the source domain may bear negative polarity in the target domain or vice versa. We call such words changing polarity words. In most of the cases, the difference in polarity occurs due to the use of the word with different senses. Example of such changing polarity word is as follows.

1a. His behavior is very cheap. (Negative)

1b. Jet airways provides very cheap flight tickets. (Positive)

In the first case, the word cheap is used with the sense of very poor quality, while in second case the sense is relatively low in price. Use of cheap with different senses is making it to have opposite sentiment polarity. On the other hand, a word which is used very scarcely (low impact) in the source domain might be used very frequently (high impact) in the target domain for sentiment classification. We call such words missing words. In most of the cases, this difference occurs due to the use of different synonymous words having the same sense. Example of such missing word is as follows.
2a. This mobile phone has a decent battery. (Positive)

2b. This place has satisfactory food options. (Positive)

The words decent and satisfactory are synonyms of each other as per Princeton WordNet-3.1. But, the domains of the sentences in 2a and 2b are different, in first case it is mobile domain, while in the second case it is restaurant domain. The words decent and satisfactory can be used interchangeably to express positive opinion in different domains.

In this paper, we propose that use of senses in place of words helps to overcome the differences across domains for cross domain SA. A word like cheap which is negative for human behavior and positive for tickets can be identified with the correct polarity orientation in the domain if we use the respective sense of the cheap in place of the word cheap. On the other hand, if a word is missing in the source domain, but its synonym is present in the target domain then the weight learned by a supervised classifier for the synonym in the source domain can be transferred (reused) to the target domain. In this way, use of senses of words in place of words is overcoming the data sparsity problem. The exemplified words good and decent are holding the same polarity orientation, that is, positive. They can be represented by the same sense as they belong to the same synset in WordNet.

In summary, use of senses (synset-IDs) of words in place of words reduces the amount of negative transfer from labeled source domain to unlabeled target domain to an extent, which in turn results into a more accurate classifier in the target domain. In this paper, we show the effectiveness of senses over words across four domains, viz., DVD (D), Electronics (E), Kitchen (K), and Books (B). Results show that the sense-based classifier in the target domain is more accurate than word-based classifier for 12 pairs of the source and target domains.

2 Related Work

Sentiment analysis within a domain has been widely studied in literature, where the train and test datasets are assumed to be from the same domain (Pang et al., 2002; Turney, 2002; Ng et al., 2006; Kanayama and Nasukawa, 2006; Breck et al., 2007; Pang and Lee, 2008; Li et al., 2009; Taboada et al., 2011). This kind of sentiment analysis where the train and test data are from the same domain is known as in-domain SA. Balamurali et al., (2011) have shown that use of senses in places of words improves the performance of in-domain SA significantly. Though they did not explore senses as features for cross-domain SA. Performance of the sentiment analysis systems drops severely when the test data is from some other domain than the train domain. This drop in performance occurs due to the differences across domains. Hence, getting a high accuracy classifier in the unlabeled target domain from a labeled source domain is a challenging task.

Domain adaptation for cross-domain sentiment classification has been explored by many researchers (Jiang and Zhai, 2007; Ji et al., 2011; Saha et al., 2011; Xia et al., 2013; Bhatt et al., 2015; Zhou et al., 2014; Glorot et al., 2011). Most of the works have focused on learning a shared low dimensional representation of features that can be generalized across different domains. Glorot et al., (2011) proposed a deep learning approach which learns to extract a meaningful representation for each review in an unsupervised fashion. Zhou et al., (2014) also proposed a deep learning approach to learn a feature mapping between cross-domain heterogeneous features as well as a better fea-

---

ture representation for mapped data to reduce the bias issue caused by the cross-domain correspondences. Although, our approach also focuses on shared representation of the source and target, but it observes the senses of words instead of words to reduce the differences created by words across domains. Our approach can handle the use of a word with opposite polarity orientation between source and target domains. In addition, it can deal with words which are missing in the source domain, but significant for target domain. Identification of changing polarity words and missing words by the use of senses across domains makes our approach more robust. Our approach do not determine a low dimensional representation of features (words) mathematically, hence it is less computationally expensive. However, it uses a manually complied resource, that is, WordNet to obtain senses of words.

3 Dataset

In this paper, we have shown impact of word’s sense in cross-domain SA for four domains, viz., DVD (D), Electronics (E), Kitchen (K), and Books (B). Data for all four domains is taken from the amazon archive (Blitzer et al., 2007). Each domain has 1000 positive and 1000 negative reviews. Table 3 shows the total number of reviews per domain and an average number of words per review in each domain.

<table>
<thead>
<tr>
<th>Domain</th>
<th>No. of Reviews</th>
<th>Avg. Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electronics (E)</td>
<td>2000</td>
<td>110</td>
</tr>
<tr>
<td>DVD (D)</td>
<td>2000</td>
<td>197</td>
</tr>
<tr>
<td>Kitchen (K)</td>
<td>2000</td>
<td>93</td>
</tr>
<tr>
<td>Books (B)</td>
<td>2000</td>
<td>173</td>
</tr>
</tbody>
</table>

Table 1: Dataset Statistics

4 Experimental Setup

In all four domains, dataset is divided into two parts, train (80%) and test (20%). Classifier is trained on the train data from source domain and results are reported on the test data from the target domain. We use SVM algorithm (Tong and Koller, 2001) to train a classifier with all the source and target pairs reported in the paper. We have presented comparison between word-based cross-domain SA and sense-based cross-domain SA. In case of word-based SA, a set of unique words (unigrams) in the training corpus make a feature set. In case of sense-based cross-domain SA, a set of unique synset-IDs (senses) form a set of features. In order to annotate the corpus with senses, we have used IMS (It Makes Sense), which is a publicly available supervised English all-words word sense disambiguation (WSD) system. There were a few instances of words which IMS failed to annotate with sense, for these instances we considered the word as feature.

5 Results

Total 12 pairs of source and target domains are possible with 4 domains. We have extensively validated our hypothesis that use of senses in place of words provides a more accurate sentiment classification system for an unlabeled target domain for 12 pairs of source and target domains. Figure 1 shows the classification accuracy obtained with word-based and sense-based systems in the target domain. The classification algorithm (SVM) and the training corpus are the same in both cases, though the difference lies in the representation of data. In case of word-based system, corpus is seen as a sequence of words, while in case of sense-based system, corpus is seen as a sequence of senses. In other words, in sense-based system words are replaced with their respective senses. For all 12 pairs of source and target domains sense-based system performs better than word-based system. Though in case of $D \rightarrow E$ and $E \rightarrow D$, difference in accuracy is low. DVD and electronics are two very different domains unlike electronics and Kitchen, or DVD and books. DVD dataset contains reviews about music albums, while electronics dataset contains reviews about electronic products. This difference in types of reviews make them to share less number of words. Table 2 shows the percent (%) of common words among the 4 domains. The percent of common unique words are common unique words divided by the summation of unique words in the domains individually. However, consistent improvement in accuracy for all 12 pairs validates our hypothesis that use of senses in place of words re-

2The dataset is available at: http://www.cs.jhu.edu/mdredze/datasets/sentiment/index2.html

3We use SVM package libsvm, which is available in java-based WEKA toolkit for machine learning: http://www.cs.waikato.ac.nz/ml/weka/downloading.html

4Available at: http://www.comp.nus.edu.sg/nlp/software.html
Figure 1: Accuracy obtained with word-based and sense-based systems in the target domain.

<table>
<thead>
<tr>
<th>Source Target</th>
<th>Word-based</th>
<th>Sense-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>E - D</td>
<td>15</td>
<td>22</td>
</tr>
<tr>
<td>E - K</td>
<td>17</td>
<td>14</td>
</tr>
<tr>
<td>E - B</td>
<td>14</td>
<td>22</td>
</tr>
<tr>
<td>D - K</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>D - B</td>
<td>17</td>
<td>14</td>
</tr>
<tr>
<td>K - B</td>
<td>17</td>
<td>17</td>
</tr>
</tbody>
</table>

Table 2: Common unique words between the domains in percent (%).

<table>
<thead>
<tr>
<th>Domain</th>
<th>Word-based</th>
<th>Sense-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>79</td>
<td>81.25</td>
</tr>
<tr>
<td>D</td>
<td>76.25</td>
<td>79</td>
</tr>
<tr>
<td>K</td>
<td>85</td>
<td>86</td>
</tr>
<tr>
<td>B</td>
<td>75.75</td>
<td>79</td>
</tr>
</tbody>
</table>

Table 3: In-domain sentiment classification accuracy in %.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Word-based</th>
<th>Sense-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>79</td>
<td>81.25</td>
</tr>
<tr>
<td>D</td>
<td>76.25</td>
<td>79</td>
</tr>
<tr>
<td>K</td>
<td>85</td>
<td>86</td>
</tr>
<tr>
<td>B</td>
<td>75.75</td>
<td>79</td>
</tr>
</tbody>
</table>

Table 3 shows the in-domain sentiment classification accuracy obtained with words and senses as features. Here, the classification algorithm is the same for all four domains, but the train and test data are from the same domain. For example, if classifier is trained in the electronics domain, results are reported on the test data from the electronics domain only. Balamurali et al., (2011) have shown that senses are better features than words for in-domain SA. They have reported results in tourism and health domains. The results reported in Table 3 for electronics, DVD, kitchen and book domains also validate the hypothesis proposed by Balamurali et al., (2011). We can consider the accuracies reported in the Table 3 as the skyline for cross-domain SA. Though the gap in accuracy obtained under cross-domain (cf. Figure 1) settings and in-domain (cf. Table 3) settings is very high, yet use of senses tries to fill the gap to an extent.

6 Conclusion

Cross-domain sentiment analysis is a challenging task due to the differences across domains. Direct application of a classifier on a domain other than the domain of the training data degrades the classification accuracy. In this paper, we propose that use of senses of words in place of words helps to overcome the differences across domain to an extent, which in turn leads to a more accurate classifier in the target domain from a labeled source domain. Senses are able to assign correct weights to changing polarity words and missing words across domains under supervised classification settings. Results have shown that sense-based cross-domain system outperforms the word-based system by 3% on an average. We have shown impact of senses of words in comparison to words for 12 pairs of source and target domains using 4 domains. In future, senses of words can be combined with other features and techniques to reduce the gap between skyline and the reported accuracy for cross-domain SA.

---

5Since the dataset in tourism (600 documents) and health (1000 documents) domains is very small in size, we have not reported results with these domains for cross-domain SA.
References


