Learning to Adapt Credible Knowledge in Cross-lingual Sentiment Analysis

Qiang Chen*,†, Wenjie Li†,*, Yu Lei†, Xule Liu*, Yanxiang He*,‡
*School of Computer Science, Wuhan University, China
†Department of Computing, The Hong Kong Polytechnic University, Hong Kong
*Hong Kong Polytechnic University Shenzhen Research Institute, China
‡The State Key Lab of Software Engineering, Wuhan University, China
*{qchen, xuleliu, yxhe}@whu.edu.cn
†{csqchen, cswjli, csylei}@comp.polyu.edu.hk

Abstract

Cross-lingual sentiment analysis is a task of identifying sentiment polarities of texts in a low-resource language by using sentiment knowledge in a resource-abundant language. While most existing approaches are driven by transfer learning, their performance does not reach to a promising level due to the transferred errors. In this paper, we propose to integrate into knowledge transfer a knowledge validation model, which aims to prevent the negative influence from the wrong knowledge by distinguishing highly credible knowledge. Experiment results demonstrate the necessity and effectiveness of the model.

1 Introduction

With the wide range of business value, sentiment analysis has drawn increasing attention in the past years. The extensive research and development efforts produce a variety of reliable sentiment resources for English, one of the most popular language in the world. These available rich resources become the treasure of knowledge to help conduct or enhance sentiment analysis in the other languages, which is a task known as cross-lingual sentiment analysis (CLSA). In the literature of CSLA, the language with abundant reliable resources is called the source language (e.g., English), while the low-resource language is referred to as the target language (e.g., Chinese). However, in this paper, the situation is a low resource language scenario, where the source language is English, and the target language is Chinese.

The main idea of existing CLSA researches is to first build up the connection between the source and target languages to overcome the language barrier, and then develop an appropriate knowledge transfer approach to leverage the annotated data from the source language to train a sentiment classification model in the target language, either supervised or semi-supervised. In particular, these approaches exploit and convert the knowledge learned from the source language to automatically generate and expand the pseudo-training data for the target language.

The machine translation (MT) service is one of the most common ways used to build the language connection (Wan, 2008; Banea et al., 2008; Wan, 2009; Wei and Pal, 2010; Gui et al., 2014). Although it is claimed in Duh et al. (2011) that the MT service is ripe for CLSA, the imperfect MT quality hinders existing MT-based CLSA approaches from the further advance. In our preliminary study, we find that even the Google translator[1] (i.e., one of the most widely used online MT service (Shankland 2013)) may unavoidably changes the sentiment polarity of the translated text, as illustrated below, with a percentage of around 10%.

*[**Original English Text**]: I am at home on bed rest and desperate for something good to read.*
*[Sentiment Label: **Negative**]*
*[**Translated Chinese Text**]:* 我在家卧床休息和绝望的东西很好看。 {*Meaning: I am in bed to rest at home and feel that desperate things are also good to read.*}*[Sentiment Label: **Positive**]*

The noisy data generated by MT errors for sure will weaken the contribution of the transferred knowledge and even worse may create conflicting knowledge. While it is a critical step in CLSA to localize the sentiment knowledge learned from the source language in the target language, to the best of our knowledge, hardly any previous research has focused on knowledge validation to filter out the noisy knowledge having sentiment changes caused by wrong translations during knowledge transfer.

[1] http://translate.google.com

Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, pages 419–429,
Beijing, China, July 26-31, 2015. ©2015 Association for Computational Linguistics

To reduce the noisy sentiment knowledge introduced into the target language, we are motivated to validate the knowledge transferred from the source language by checking its linguistic distributions and sentiment polarity consistency with the known knowledge in the target language. Different from previous co-training based approaches where two language views recommend knowledge to each other in the same manner, we consider the source language as the "supervisor" and the target language as the "learner". The "supervisor" boosts itself with its own accumulated labeled data (called knowledge) and meanwhile recommends its confident knowledge to the "learner". The "learner" tries to select trustworthy knowledge based on the recommendation to update and expand its training data. Adding a process to efficiently filter out noisy knowledge and retain the self-adaptive and interested new knowledge makes the subsequent boosting process more credible. This is why our approach can outperform state-of-the-art CLSA approaches.

The rest of this paper is organized as follows. Section 2 summarizes the related work. Section 3 explains the proposed model. Section 4 presents experimental results. Finally, Section 5 concludes the paper and suggests future work.

2 Related Work

2.1 Sentiment Analysis

Sentiment has been analyzed in different language granularity, e.g., entity, aspect, sentence and document. This paper focuses on sentiment analysis of online product reviews in the document level.

Existing approaches are generally categorized into lexicon-based and machine learning based approaches (Liu, 2012). Lexicon-based approaches highly depend on sentiment lexicons. Turney (2002) derives the overall phrase and document sentiment scores by averaging the sentiment scores provided in a lexicon over the words included. Similar idea is adopted in (Hiroshi et al., 2004; Kennedy and Inkpen, 2006). Machine learning based approaches, on the other hand, apply classification models. The task-specific features are designed to train sentiment polarity classifiers. Pang et al. (2002) compare the performance of NB, SVM and ME on movie reviews. SVM is found more effective. Gamon (2004) shows that SVM with deep linguistic features can further improve the performance. A variety of other machine learning approaches are also proposed to sentiment classification (Mullen and Collier, 2004; Read, 2005; Hassan and Radev, 2010; Socher et al., 2013).

Cross-domain sentiment classification (CDSC) shares certain common characteristics with cross-lingual sentiment classification (CLSC) (Tan et al., 2007; Li et al., 2009; Pan and Yang, 2010; He et al., 2011a; Glorot et al., 2011). Notice that the gap between source domain and target domain is the main difference between CDSC and CLSC. CLSC copes with two different datasets in two different languages. This difference makes CLSC a new challenge, drawing specific attention to researcher recently.

2.2 Cross-lingual Sentiment Analysis

There are two alternative solutions to cross-lingual sentiment analysis. One is ensemble learning that combines multiple classifiers. The other is transfer learning that develops strategies to adapt the knowledge from one language to the other. Wan (2008) is among the pioneers to develop the ensemble learning solutions, where multiple classifiers learned from different training datasets including those in original languages and translated languages are combined by voting. Most researches, on the other hand, explore transfer learning and focus on knowledge adaptation. For example, Wan (2009) applies a supervised co-training framework to iteratively adapt knowledge learned from the two languages by transferring translated texts to each other. Other similar work includes (Wei and Pal, 2010) and (He, 2011b). All these approaches rely on MT to build language connection.

Meanwhile, the unlabeled parallel data is also employed to fill the gap between two languages. To solve the feature coverage problem with the EM algorithm, Meng et al. (2012) leverage the unlabeled parallel data to learn unseen sentiment words. Similarly, Popat et al. (2013) use the unlabeled parallel data to cluster features in order to reduce the data sparsity problem. Meng et al. (2012) and Popat et al. (2013) also use the unlabeled parallel data to reduce the negative influence of the noisy and incorrect sentiment labels introduced by machine translation and knowledge transfer. However, the parallel data is also a scarce resource.

Some existing transfer learning based CLSA methods have attempted to address the noisy knowledge problem caused by wrong labels by checking label consistency. For example, to filter out the unconfident labels in Chinese, the supervised learning method proposed by (Xu et al., 2011) runs boosting in Chinese by checking consistency between the labels manually annotated in English and predicted by Chinese classifiers on translated Chinese. The work in (Gui et al., 2014) follows the same line although it considers knowledge transferring between two languages. On the contrary, the main focus of our work is to filter out the noisy knowledge having sentiment changes by wrong translations. Actually, both label consistency checking and linguistic distribution checking are important. Any one alone cannot work well. In fact, both of them are considered as the knowledge validation in our work, though the later is our focus.

3 Credible Boosting Model

In this paper, we propose a knowledge validation approach to improve the effectiveness of knowledge transfer without directly using extra parallel data. Our target is to filter out the noisy sentiment labels introduced by MT and the incorrect sentiment labels generated by imperfect classifier in the source language. Here, the knowledge is referred to as a collection of distributed document presentations with sentiment labels that have been verified to be robust in sentiment classification (Le and Mikolov, 2014). A novel credible boosting model, namely CredBoost is proposed to apply transfer-supervised learning with an added self-validation mechanism to guarantee the knowledge transferred highly credible and self-adaptive.

3.1 Problem Description

In a standard cross-lingual sentiment analysis setting, the training data includes labeled English reviews $L_{EN} = \{(x_i^{l_{en}}, y_i)\}_{i=1}^{M}$ and unlabeled Chinese reviews $U_{CN} = \{x_j^{u_{cn}}\}_{j=1}^{N}$, where x_i^k $(k = l_{en} \mathit{or}\ u_{cn})$ represents review i and $y_i \in \{-1, 1\}$ is the sentiment label of review x_i^l. The test data is Chinese reviews $T_{CN} = \{x_s^{t_{cn}}\}_{s=1}^{S}$.

We now introduce the unlabeled data into credBoost's setting. L_{EN} is divided into two disjoint parts L_{EN}^T and L_{EN}^B, where L_{EN}^T for basic training and L_{EN}^B for self-boosting. We translate L_{EN} into Chinese to obtain extra labeled Chinese pseudo-reviews $L_{TrCN} = \{(x_i^{l_{cnTr}}, y_i)\}_{i=1}^{M}$ and U_{CN} into English to obtain extra unlabeled English pseudo-reviews $U_{TrEN} = \{x_j^{l_{enTr}}\}_{j=1}^{N}$. Thereby, we obtain a pair of pseudo-parallel data (U_{CN}, U_{TrEN}).

The task is to use L_{EN} and U_{CN} to train a Chinese classifier to predict sentiment polarity for the test data T_{CN}. It is a standard transfer learning problem. We consider two language views, i.e., source language view D_S and target language view $D_\mathcal{T}$. D_S boosts itself with the labeled English data and recommend translated knowledge to $D_\mathcal{T}$, while D_t selects self-adaptive ones to boost itself.

3.2 Framework of CredBoost

The CredBoost model involves two synchronously boosting views for two languages respectively. During training, one view acts as a "supervisor" that recommends and passes the knowledge to the other view. The same knowledge is also added into its own view for boosting by automatically updating the weights of the labeled data. The other view acts as a "learner" that receives the recommended knowledge and selects the best-suited new knowledge to learn.

As mentioned before, the knowledge transferred through MT is not reliable. The source language view may also make wrong predictions and thus transfer the wrong knowledge to the target language even the translations are correct. Whether or not the "learner" can benefit from its "supervisor" and how much it benefits highly depends on the credibility and adaptiveness of the recommended knowledge accepted by the "learner". Knowledge validation is necessary to ensure the quality of learning. The objective of knowledge validation is to identify the new and acquired knowledge from recommendations. Both language views are iteratively trained until learning converges or reaches the iteration upper bound.

In the **source language view**, at iteration (t), the CredBoost model first uses $L_{EN}^{T(t)}$ to train a basic classifier $\mathcal{C}_{EN}^{(t)}$ and then uses $\mathcal{C}_{EN}^{(t)}$ to predict $L_{EN}^{B(t)}$ and $U_{TrEN}^{(t)}$. Top m and top n instances are sampled from $L_{EN}^{B(t)}$ and $U_{TrEN}^{(t)}$ respectively, by Formula (1) :

$$\begin{aligned} O_{EN}^{(t)} &= \{(x_{i'}^{LB}, \hat{y}_{i'}^{LB})\}_{i'=1}^{m_{en}} \\ TR_{EN}^{(t)} &= \{(x_i^{UTr}, \hat{y}_i^{UTr})\}_{i=1}^{n_{en}} \end{aligned} \tag{1}$$

where $O_{EN}^{(t)}$ denotes the candidates to be added into the training data, and $TR^{(t)}_{EN}$ the knowledge to be recommended to the target language view. We use the source knowledge validation function $V_{\mathcal{S}}(O^{(t)}_{EN})$ to identify the acquired knowledge $K^{(t)}_{'Ac}$ learned in the previous learning process and the new knowledge $K^{(t)}_{'Nw}$ fresh to the current knowledge system from $O^{(t)}_{EN}$. The importance of each training instance is updated according to the performance of prediction by Formula (2) :

$$\begin{aligned}
\omega'^{Ac}_{i'} &= \begin{cases} e^{\epsilon(t)} \cdot \sqrt{\nu^{(t)}_{i'} \cdot c^{(t)}_{i'}} & \text{if } \hat{y}'^{Ac}_{i'} \neq y'^{Ac}_{i'} \\ \sqrt{\nu^{(t)}_{i'} \cdot c^{(t)}_{i'}} & \text{otherwise;} \end{cases} \\
\omega'^{Nw}_{j'} &= \begin{cases} e^{\epsilon(t)} \cdot \log(1 + \sqrt{e} \cdot c^{(t)}_{j'}) & \text{if } \hat{y}'^{Ac}_{j'} \neq y'^{Ac}_{j'} \\ \log(1 + \sqrt{e} \cdot c^{(t)}_{j'}) & \text{otherwise.} \end{cases}
\end{aligned} \tag{2}$$

where $c^{(t)}_{j'}$ is the confidence of an instance given by $\mathcal{C}^{(t)}_{EN}$, thus $\log(1 + \sqrt{e} \cdot c^{(t)}_{j'}) > 1$ is to enhance the weight of new knowledge because of the higher significance contributing to the later learning. $\nu^{(t)}_{i'} (< 1)$ is the adaptiveness score given by the source knowledge validation function $V_{\mathcal{S}}(O^{(t)}_{EN})$. $\epsilon_{(t)} (> 1)$ is the error rate of $\mathcal{C}^{(t)}_{EN}$, thus $e^{\epsilon(t)} > 1$ is to reward the wrongly predicted data in the next iteration. $\hat{y}'^{Ac}_{i'}$ is the label given by $\mathcal{C}^{(t)}_{EN}$ and $y'^{Ac}_{i'}$ is the manually annotated label. For the incorrectly predicted instance, the weight is boosted inversely to the performance of the current classifier. The instance identified as the new knowledge which contributes more to performance improvement is given a reward parameter to enhance its significant in the next training iteration. Data sets update by Formula (3). The training starts with iteration (1), the training data is initially set as $L^{T(1)}_{EN} = L^{T}_{EN}$.

$$\begin{aligned}
L^{T(t+1)}_{EN} &= L^{T(t)}_{EN} \cup K^{(t)}_{'Ac} \cup K^{(t)}_{'Nw} \\
L^{B(t+1)}_{EN} &= L^{B(t)}_{EN} - (K^{(t)}_{'Ac} \cup K^{(t)}_{'Nw})
\end{aligned} \tag{3}$$

In the **target language view**, at iteration (t), the CredBoost model receives the recommended knowledge $TR^{(t)}_{EN}$ and projects it to $O^{(t)}_{CN}$ from the unlabeled Chinese data $U^{(t)}_{CN}$ with the pseudo-parallel data $(U^{(t)}_{CN}, U^{(t)}_{TrEN})$. $O^{CN}_{(t)}$ is validated by the target knowledge validation function $V_{\mathcal{T}}(O^{(t)}_{CN})$ to identify the acquired knowledge $K^{(t)}_{Ac}$ and the new knowledge $K^{(t)}_{Nw}$. $K^{(t)}_{Ac}$ and $K^{(t)}_{Nw}$ are projected to $K^{(t)}_{*Ac}$ and $K^{(t)}_{*Nw}$ from the unlabeled English pseudo-data $U^{(t)}_{TrEN}$. The weight of an instance is updated by Formula (4), and the parameter setting is similar to that in the source language view. The confidence $c^{(t)}_{i}$ is directly transferred from D_s. We reward the validated knowledge to raise their significance in the training data considering they are originally Chinese.

$$\begin{aligned}
\omega^{Ac}_{i} &= \sqrt{c^{(t)}_{i} \cdot \log(1 + \sqrt{e} \cdot v^{(t)}_{i})} \\
\omega^{Nw}_{j} &= e^{\log(1 + \sqrt{e} \cdot c^{(t)}_{j})} = 1 + \sqrt{e} \cdot c^{(t)}_{j}
\end{aligned} \tag{4}$$

We update the data setting by Formula (5). The training data is initially set as $U^{T(1)}_{CN} = U^{T}_{CN}$. The CredBoost model is illustrated in Algorithm 1.

$$\begin{aligned}
L^{(t+1)}_{TrCN} &= L^{(t)}_{TrCN} \cup K^{(t)}_{Ac} \cup K^{(t)}_{Nw} \\
U^{(t+1)}_{CN} &= U^{(t)}_{CN} - (K^{(t)}_{Ac} \cup K^{(t)}_{Nw}) \\
U^{(t+1)}_{TrEN} &= U^{(t)}_{TrEN} - (K^{(t)}_{*Ac} \cup K^{(t)}_{*Nw})
\end{aligned} \tag{5}$$

Algorithm 1 CredBoost Model

```
Input: English labeled data $L^{T}_{EN}$ and $L^{B}_{EN}$, translated English unlabeled data $U_{TrEN}$, translated Chinese data $L_{TrCN}$ and unlabeled Chinese data $U_{CN}$;
Initialize: Weights $W^{(1)}_{EN} = \{1\}^M$ for $L^{T}_{EN}$ and $W^{(1)}_{TrCN} = \{1\}^M$ for $L_{TrCN}$;
For $t = 1, \cdots, T$:
    1. Use $L^{T(t)}_{EN}$ to learn English classifier $C_{EN(t)}$;
    2. Use $\mathcal{C}^{(t)}_{EN}$ to predict $L^{B(t)}_{EN}$ and $U^{(t)}_{TrEN}$ sample top $m$ and top $n$ instances from $L^{B(t)}_{EN}$and $U^{(t)}_{TrEN}$, $O^{(t)}_{EN}$ and $TR^{(t)}_{EN}$;
    3. Validate $O^{(t)}_{EN}$ by knowledge validation function $V_{\mathcal{S}}(O^{(t)}_{EN})$ to identify acquired knowledge $K^{(t)}_{'Ac}$ and new knowledge $K^{(t)}_{'Nw}$, generate the weights for them by Formula (2), then recommend $TR^{(t)}_{EN}$ to $D_{\mathcal{T}}$;
    4. Project $TR^{(t)}_{EN}$ to $O^{(t)}_{CN}$ with pseudo-parallel data $(U^{(t)}_{CN}, U^{(t)}_{TrEN})$, and use knowledge validation function $V_{\mathcal{T}}(O^{(t)}_{CN})$ to identify acquired knowledge $K^{(t)}_{Ac}$ and new knowledge $K^{(t)}_{Nw}$, then generate weights for them by Formula (4);
    5. Update $D_{\mathcal{S}}$ by Formula (2) and $D_{\mathcal{T}}$ by Formula (5);
End For.
Output: Chinese classifier $\mathcal{C}^{(T)}_{CN}$.
```

3.3 Knowledge Validation

Knowledge is familiarity, awareness or understanding of someone or something, such as facts, information or skills, which is acquired through experience or education by perceiving, discovering or learning[2]. It can be implicit or explicit.

In machine learning, natural language knowledge is a continuously improving hypothesis that consists of both semantic and significant domain

[2] Definition from Oxford Dictionary of English, available at: `http://oxforddictionaries.com/view/entry/m_en_us126`.

characters. While language is the expression of semantic, semantic is the carrier of sentiment. Using another word, two texts with more smaller semantic distance have higher probability to share the same sentiment polarity. Choi and Cardie (2008) assert that the sentiment polarity of natural language can be better inferred by compositional semantics. They also suggest that incorporating compositional semantics into learning can improve the performance of sentiment classifiers. Saif et al. (2012) also demonstrate that the addition of extra semantic features can further improve performance.

In order to filter out noisy and incorrect sentiment labels, we propose a knowledge validation approach to reduce these noisy data that hinder the improvement of learning performance. Knowledge validation is a way to identify the acquired knowledge implied in current knowledge system and also the new knowledge fresh to current knowledge system. The knowledge can be represented in the semantic space. (Le and Mikolov, 2014) project documents into a low-dimension semantic space with a deep learning approach, known as document-to-vector (Doc2Vec[3]). Considering that Dov2Vec has been verified to be efficient in many NLP tasks including sentiment analysis, we follow previous research to represent knowledge embedded in product reviews with the vectors generated by Doc2Vec.

Suppose distributed representations (i.e., low-dimensional vectors) of the all reviews including $\{L_{EN}^T, L_{EN}^B, U_{TrEN}\}$ and $\{L_{TrCN}, U_{CN}\}$ are $\{\mathcal{V}(L_{EN}^T), \mathcal{V}(L_{EN}^B), \mathcal{V}(U_{TrEN})\}$ and $\{\mathcal{V}(L_{TrCN}), \mathcal{V}(U_{CN})\}$ respectively. At iteration (t), $\mathcal{V}(L_{EN}^{T(t)})$ is the current knowledge system of the English view and $\mathcal{V}(L_{TrCN}^{(t)})$ is that of the Chinese. The knowledge validation runs separately in the source and target views.

In the **target language view**, at iteration (t), suppose the prediction confidence of the candidate $(x_i^U, \hat{y}_i^U) \in O_{CN}^{(t)}$ is $c_i^{(t)}$. We define the adaptiveness score as the average distance of top ζ_+ semantic distances between the instance x_i^{LB} and the positive cluster of $L_{TrCN}^{(t)}$, denoted as $L_{TrCN}^{(t)+}$, and top $\zeta_-^{(t)} = \zeta_+ \cdot \frac{\mathcal{L}_+^{(t)}}{\mathcal{L}_-^{(t)}}$ semantic distances between x_i^U and the negative cluster, denoted as $L_{TrCN}^{(t)-}$, where $\mathcal{L}_+^{(t)}$ and $\mathcal{L}_-^{(t)}$ are the numbers of the elements in $L_{TrCN}^{(t)+}$ and $L_{TrCN}^{(t)-}$ respectively. The validation parameters are defined by Formula (6), ω_r is the weight of training instance $\mathcal{V}(r)$, $\nu_i^{(t)}$ is the adaptiveness score, and $\mathcal{V}_*^{label} \in \{1, -1\}$ is the validated label which denotes the knowledge belonging to the positive cluster $L_{TrCN}^{(t)+}$ or the negative cluster $L_{TrCN}^{(t)-}$. The validation process is illustrated in Algorithm 2, where the acquired knowledge is $k_{Ac}^{(t)}$, and the new knowledge is $k_{Nw}^{(t)}$.

$$
\begin{aligned}
&\mathcal{D}(\mathcal{V}(x_i^{LB}), \mathcal{V}(r)) = \frac{\mathcal{V}(x_i^{LB})^T \cdot \mathcal{V}(r)}{\parallel \mathcal{V}(x_i^{LB}) \parallel \cdot \parallel \mathcal{V}(r) \parallel} \\
&\Rightarrow \begin{cases} \nu_i^{(t)+} = \frac{1}{\zeta +} \sum\limits_{r \in L_{EN}^{(t)+}} \omega_r \; \mathcal{D}(\mathcal{V}(x_i^{LB}), \mathcal{V}(r)) \\ \nu_i^{(t)-} = \frac{1}{\zeta_-^{(t)}} \sum\limits_{r' \in L_{EN}^{(t)-}} \omega_{r'} \; \mathcal{D}(\mathcal{V}(x_i^{LB}), \mathcal{V}(r')) \end{cases} \\
&\Rightarrow \quad \Delta(\nu_i^{(t)}) = \nu_i^{(t)+} - \nu_i^{(t)-} \\
&\Rightarrow \quad \delta_i^{(t)} = \frac{1}{e^{1+\Delta(\nu_i^{(t)})}} \\
&\Rightarrow \quad \mathcal{V}_*^{label} = \begin{cases} 1 & \text{if } \delta_i^{(t)} > 0.5, \\ -1 & \text{if } \delta_i^{(t)} \leq 0.5. \end{cases} \\
&\Rightarrow \quad \nu_i^{(t)} = \begin{cases} \nu_i^{(t)+} & \text{if } \mathcal{V}_*^{label} = 1, \\ \nu_i^{(t)-} & \text{if } \mathcal{V}_*^{label} = -1. \end{cases}
\end{aligned} \tag{6}
$$

where $\mathcal{D}(\mathcal{V}(x_i^{LB}), \mathcal{V}(r))$ is the Cosine distance between the distributed representations of the two reviews. $\nu_i^{(t)+}$ and $\nu_i^{(t)-}$ are the weighted averages of the semantic distances. $\delta_i^{(t)}$ is the Sigmoid function which computes the probability that the data is distributed in the positive cluster $L_{TrCN}^{(t)+}$.

In the **source language view**, at iteration (t), let's suppose the prediction confidence of candidate $(x_{i'}^{LB}, \hat{y}_{i'}^{LB}) \in O_{EN}^{(t)}$ to be $c_{i'}^{(t)}$. The definitions of validation parameters are similar to those in the target language view. The validation process is illustrated in Algorithm 3. The validation is looser, because the training data and candidates are both in English. This differs from it in the target view.

4 Experiments

4.1 Experimental Setup

We evaluate the proposed CredBoost model on an open cross-lingual sentiment analysis task in NLP&CC 2013[4]. The data set provided is a

[3] Doc2Vec is one of the models implemented in the free python library ***Gensim*** which can be freely downloaded at: `https://pypi.python.org/pypi/gensim`.

[4] NLP&CC is an annual conference of Chinese information technology professional committee organized by Chinese computer Federation (CCF). It mainly focuses on the study and application novelty of natural language processing and Chinese computation. CLSA task is the task 3 of NLP&CC 2013. For more details and open

Algorithm 2 Knowledge Validation $V_{\mathcal{T}}(D_{\mathcal{T}})$

```
Input: Labeled Chinese training data $L^{(t)}_{TrCN}$, weights of labeled data $W^{(t)}_{CN}$ and semantics vectors of all English data for iteration $(t)$: $\{\mathcal{V}(L^{(t)}_{TrCN}), \mathcal{V}(U^{(t)}_{CN})\}$;
Initialize: $K^{(1)}_{'Ac} = \phi$, $K^{(1)}_{'Nw} = \phi$;
For  $x^U_i$ in $O^{(t)}_{CN}$:
    1. Use $L^{(t)}_{TrCN}$ to train a classifier $\mathcal{C}^{(t)}_{CN}$,
       then use $\mathcal{C}^{(t)}_{CN}$ predict $x^U_i$, giving label $y^{CN}_i$;
    2. Get validated label $\mathcal{V}^{label}_*$, positive and
       negative average distances $\nu^{(t)+}_i$, $\nu^{(t)-}_i$
       of $x^U_i$ by fomula (6);
    3. If $\nu^{(t)+}_i < \psi$ and $\nu^{(t)-}_i < \psi$:
            If $\hat{y}^{LB}_i = \mathcal{V}^{label}_*$:
            Then $K^{(t)}_{Nw} \leftarrow K^{(t)}_{Nw} + x^U_i$;
       Else:
            If $\hat{y}^{LB}_i = \mathcal{V}^{label}_* = y^{CN}_i$:
            Then $K^{(t)}_{Ac} \leftarrow K^{(t)}_{Ac} + x^U_i$;
End For.
Output: $K^{(t)}_{Nw}$, $K^{(t)}_{Ac}$.
```

Algorithm 3 Knowledge Validation $V_{\mathcal{S}}(D_{\mathcal{S}})$

```
Input: Weights of labeled data $W^{(1)}_{EN}$ and semantics vectors of all English data for iteration $(t)$:
$\{\mathcal{V}(L^{T(t)}_{EN}), \mathcal{V}(L^{B(t)}_{EN}), \mathcal{V}(U^{(t)}_{TrEN})\}$;
Initialize: $K^{(1)}_{'Ac} = \phi$, $K^{(1)}_{'Nw} = \phi$;
For  $x^{LB}_{i'}$ in $O^{(t)}_{EN}$:
    1. Get validated label $\mathcal{V}^{label}_{'}$, positive and
       negative average distances $\nu^{(t)+}_{i'}$, $\nu^{(t)-}_{i'}$
       of $x^{LB}_{i'}$ by fomula (6);
    2. If $\nu^{(t)+}_{i'} < \psi$ and $\nu^{(t)-}_{i'} < \psi$:
            If $\hat{y}^{LB}_{i'} = \mathcal{V}^{label}_{'}$:
            Then $K^{(t)}_{'Nw} \leftarrow K^{(t)}_{'Nw} + x^{LB}_{i'}$;
       Else:
            If $\hat{y}^{LB}_{i'} = \mathcal{V}^{label}_{'}$:
            Then $K^{(t)}_{'Ac} \leftarrow K^{(t)}_{'Ac} + x^{LB}_{i'}$;
End For.
Output: $K^{(t)}_{'Nw}$, $K^{(t)}_{'Ac}$.
```

collection of bilingual Amazon product reviews in Books, DVD and Music domains. It contains 4,000 labeled English reviews, 4,000 Chinese test reviews, and 17,814, 47,071, 29,677 unlabeled Chinese reviews in three different domains. We randomly select 2,000 unlabeled Chinese reviews in each domain to train classifiers. Besides, the pseudo-data sets described in CredBoost model are translated with Google translator. The data set is summarized in Table 1.

To better illustrate the significance of knowledge validation during knowledge transfer, we compare the proposed method with the following baseline methods:

Lexicon-based (LB): The standard English MPQA sentiment lexicons are translated into Chinese and then utilized together with a small number of Chinese turning words, negations and intensifiers to predict the sentiment polarities of the Chinese test reviews.

Domain		English		Chinese	
		L	U	L	U
Books	Train	4,000	-	-	2,000
	Test	-	-	4,000	-
DVD	Train	4,000	-	-	2,000
	Test	-	-	4,000	-
Music	Train	4,000	-	-	2,000
	Test	-	-	4,000	-

Table 1: Experimental data sets. All data sets are balanced, L represents labeled data and U represents unlabeled data.

Basic SVM (BSVM-CN): The labeled English reviews are translated into Chinese, which are then used as the pseudo-training data to train a Chinese SVM classifier.

Primarily boost transfer learning (BTL-1): The labeled English reviews are used to train the English classifier, which is applied to label the English translations of the unlabeled Chinese reviews. These labeled Chinese reviews obtained via MT together with the Chinese translations of the labeled English reviews are then used as the pseudo-training data to train a Chinese sentiment classifier.

Best result in NLP&CC 2013 (BR2013): This is the best result reported in NLP&CC 2013. Unfortunately, the specification of the method is not available.

Self-boost (SB-CN) in Chinese: The labeled English reviews are translated into Chinese, which are used as the pseudo-training data to train a basic Chinese classifier. This classifier is iteratively refined by choosing the most confidently predicted English reviews to add into the Chinese training data until a predefined iteration number reaches. It can be also considered as a self-adaptive boosting approach.

Iteratively boost transfer learning (BTL-2): This is an enhanced transfer learning method sharing the same learning framework with CredBoost but it ignores knowledge validation. It iteratively transfers the knowledge from English to Chinese. The learning in both languages iteratively boosts themselves separately. The transfer size is 16, comparable to that in CredBoost.

Basic co-training (CoTr): The co-training method proposed in (Wan, 2009) is implemented. It is bidirectional transfer learning. In each

resource, you can available at: `http://tcci.ccf.org.cn/conference/2013/index.html`.

iteration, 10 positive and 10 negative reviews are transferred from one language to the other.

Doc2vec feature CredBoost (dCredB): This method is similar to CredBoost except that document-to-vector is used to generate features when training basic classifiers. The vectors are obtained from both original and translated reviews. The dimension of doc2vec is 300, while the other parameters are set as default.

The baseline methods described above are categorized into three classes: the first four which are preliminary methods, the middle three which are several state-of-the-art models being comparable to our proposed model, and the last one which is a comparison to suggest that the knowledge representation is not the answer to the performance improvement. For all the methods excluding LB and BR2013, we use support vector machines (SVMs) as basic classifiers. We use the Liblinear package (Fan et al., 2008) with the linear kernel[5]. All methods use Unigram+Bigram features to train the basic classifiers, except for dCredB.

4.2 Experimental Result

In this work, there are two main parameters that may significantly influence the performance of our proposed model. They are the new knowledge validation boundary ψ and the validation scale ζ_+ in the training data. We set the values of parameters with the grid search strategy. We first fix initial $\zeta_+ = 14$ to search the best new knowledge validation boundary ψ from an empirical value set $\{0.30, 0.35, 0.40, 0.45, 0.50\}$. We then fix the best $\psi = 0.40$ to check the suitable validation scale ζ_+ from the initial value set $\{6, 8, 9, 10, 11, 12, 14, 16\}$ in which values are comparable with the knowledge transfer scale of CoTr in the training data. Besides, the recommendation size m for English is set to 20 and the recommendation size n for Chinese is set to 40. The final settings are listed in Table 2. The performance is evaluated in terms of accuracy (Ac) defined by Formula (7).

$$Ac(f) = \frac{p^f}{P^f}, \quad Avg\_Ac = \frac{1}{3} \cdot \sum_{f' \in \mathcal{F}} Ac(f') \quad (7)$$

where p^f is the number of correct predictions and P^f is the total number of the test data; $\mathcal{F} \in \{Books, DVD, Music\}$ is the domain set.

Domain	ψ	ζ_+	m	n
Books	0.45	12	20	40
DVD	0.40	12	20	40
Music	0.40	9	20	40

Table 2: Parameter settings of three domains in this paper.

Approaches	Domain			Avg_Ac
	Books	DVD	Music	
LB	0.7770	0.7832	0.7595	0.7709
BSVM-CN	0.7940	0.7995	0.7778	0.7904
BTL-1	0.8010	0.8058	0.7605	0.7891
BR2013	0.7850	0.7773	0.7513	0.7712
SB-CN	0.8400	0.8428	0.8012	0.8280
BTL-2	0.8105	0.8265	0.7980	0.8117
CoTr	0.8025	0.8508	0.7812	0.8115
dCredB	0.6485	0.6753	0.6700	0.6646
CredBoost	**0.8465**	**0.8518**	**0.8093**	**0.8359**

Table 3: Macro performance of all approaches in three domains. All values are accuracies and Avg-Ac represents the average accuracy in three domains.

The performances are reported in Tables 3 and 4. As shown, CredBoost outperforms all the other comparison methods. The first four baselines have poor performances compared to others. This suggests that the CLSA problem cannot be well solved by directly learning from the labeled translated data without any knowledge adaption or knowledge validation. SB-CN, BTL-2 and CoTr employ iterative boosting to adapt knowledge from the source English to the target Chinese without validating the transferred knowledge. They inevitably mis-recommend the massive noisy data into Chinese. CredBoost, in contrast, introduces knowledge validation into transfer learning with iterative boosting. It better adapts knowledge from English to Chinese and thus ensures the credibility of the accepted knowledge. Its best result justifies our assumption.

Specifically, SB-CN leverages both the Chinese training data translated from the labeled English data and the unlabeled Chinese data used for boosting. The boosting in Chinese iteratively selects the trustworthy data with the labels assigned by the Chinese classifier. Our proposed method, however, exploits two different languages simultaneously with an additional boosting step, i.e., it transfers knowledge from English to Chinese during boosting. We then use knowledge validation model to validate the unlabeled Chinese data whose labels are assigned by the English

[5] The parameter setting used in this paper is '-*s 7*'.

Model (Books)	Positive			Negative			Ac
	P	R	F1	P	R	F1	
LB	0.7368	0.8400	0.7850	0.8140	0.7000	0.7527	0.7700
BSVM-CN	0.8249	0.7465	0.7837	0.7685	0.8415	0.8033	0.7940
BTL-1	**0.8537**	0.7265	0.7850	0.7620	0.8755	0.8148	0.8010
BR2013	-	-	-	-	-	-	0.7850
SB-CN	0.8716	0.7975	0.8329	0.8134	**0.8825**	0.8465	0.8400
BTL-2	0.7105	**0.8881**	0.7894	**0.9105**	0.7588	0.8278	0.8105
CoTr	0.8339	0.7555	0.7928	0.7765	0.8495	0.8114	0.8025
dCredB	0.5310	0.6941	0.6017	0.7660	0.6202	0.6854	0.6485
CredBoost	0.8225	0.8640	**0.8427**	0.8705	0.8306	**0.8501**	**0.8465**

Model (DVD)	Positive			Negative			Ac
	P	R	F1	P	R	F1	
LB	0.7648	0.8180	0.7905	0.8044	0.7485	0.7754	0.7832
BSVM-CN	0.7745	0.8450	0.8082	0.8295	0.7540	0.7900	0.7995
BTL-1	0.8282	0.7715	0.7988	0.7861	0.8400	0.8122	0.8058
BR2013	-	-	-	-	-	-	0.7773
SB-CN	**0.8853**	0.7875	0.8335	0.8086	**0.8980**	0.8510	0.8428
BTL-2	0.8525	0.8104	0.8309	0.8005	0.8444	0.8219	0.8265
CoTr	0.8374	0.8705	**0.8536**	**0.8652**	0.8310	0.8478	0.8508
dCredB	0.6070	0.7030	0.6515	0.7435	0.6542	0.6960	0.6753
CredBoost	0.8440	**0.8572**	0.8508	0.8595	0.8465	**0.8530**	**0.8518**

Model (Music)	Positive			Negative			Ac
	P	R	F1	P	R	F1	
LB	0.7387	0.8030	0.7695	0.7842	0.7160	0.7485	0.7595
BSVM-CN	0.8492	0.6755	0.7525	0.7306	0.8800	0.7984	0.7778
BTL-1	0.8437	0.6395	0.7275	0.7097	0.8815	0.7863	0.7605
BR2013	-	-	-	-	-	-	0.7513
SB-CN	**0.8787**	0.6990	0.7786	0.7501	**0.9035**	0.8197	0.8012
BTL-2	0.7285	0.8461	0.7829	0.8675	0.7616	0.8111	0.7980
CoTr	0.8536	0.6790	0.7564	0.7335	0.8835	0.8015	0.7812
dCredB	0.5860	0.7043	0.6397	0.7540	0.6455	0.6955	0.6700
CredBoost	0.7258	**0.8708**	**0.7917**	**0.8928**	0.7653	**0.8241**	**0.8093**

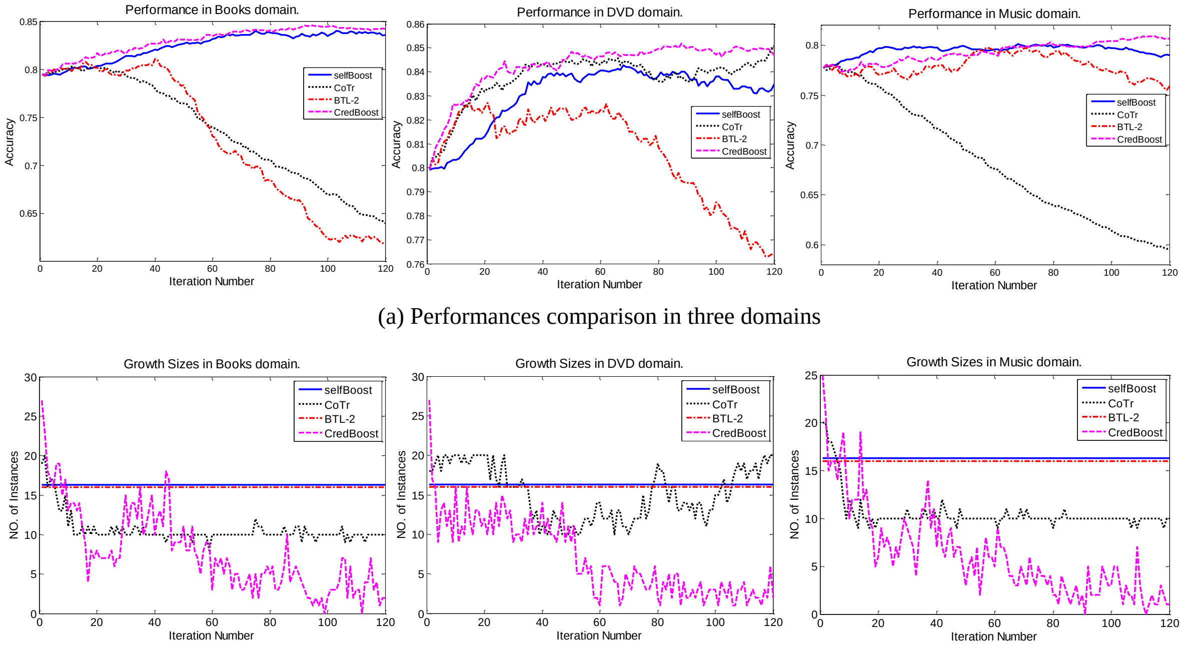
Table 4: Micro performance of all approaches in three domains. P: Precision, R: Recall, F1: micro-F measure, Ac: Accuracy, and - represents unknown. The model in BR2013 is unknown, thus its micro performance is unavailable.

classifier. It is reasonable that a Chinese classifier performs better on Chinese text than an English classifier performs on the translated English text due to the different language distributions and MT errors. However, as shown in Tables 3 and 4, the better performance of our proposed method compared with that of the self-boosting method further suggests the effectiveness of our proposed knowledge validation model.

Figure 1 illustrates the continuous changes of performances vs. the corresponding growth sizes of the training data sets for SB-CN, BTL-2, CoTr, and CredBoost. According to our common sense, noisy data have negative influence on performance improvement. Compared to the other three methods, CredBoost accepts less number of training instances during learning while it achieves more improvement. This verifies the ability of CredBoost that can filter out the noisy data recommended by the English sentiment classifier.

In Figure 1(a), the curves of BTL-2 and CoTr suggest that directly transferring the knowledge recommended from English imports many noisy data into Chinese. It is also obvious that the performance curve of CredBoost implies a stable improvement trend while the other three decrease after certain iterations because of the accumulated negative influence from the noisy data. Figure 1(b) shows CredBoost accepts decreased training instances after certain iterations because the number of "high-quality" instances decrease when learning proceeds. This finding suggests that knowledge validation would rather abandon "less-credible" knowledge with higher probability than easily accept it. Knowledge validation in the proposed model guarantees highly-credible learning when transferring knowledge from English to Chinese. The results also show that CredBoost has great potential to achieve better performance approaching to supervised approaches if more unlabeled Chinese data are available.

Another interesting finding is also observed.



(a) Performances comparison in three domains

(b) Growth sizes comparison in three domains

Figure 1: Performances vs. Growth Sizes for SB-CN, CoTr, BTL-2, and CredBoost in three domains. The similar performance curves of CoTr is also reported in (Gui et al., 2014).

Although document-to-vector represents content semantic well, it cannot determine the sentiment polarity of text well, even when the document-to-vectors that are used to train basic classifiers are learned on the mixture of the translated and original reviews. The superior performance of CredBoost to dCredB suggests that the semantic representation is effective to identify highly-credible acquired knowledge and new knowledge but it alone may not be sufficient enough to model the sentiment information.

We also conduct some other experiments to study the sensitivity of the new knowledge validation boundary ψ and the validation scale ζ_+ in the training data. The experimental results show that the performances with different parameter settings fluctuate around the best result reported in Tables 3 and 4 in a small range. Our model is basically quite stable.

5 Conclusion

In this paper, we propose a semi-supervised learning model, called CredBoost, to address cross-lingual (English vs Chinese) sentiment analysis without direct labeled Chinese data nor direct parallel data. We propose to introduce knowledge validation during transfer learning to reduce the noisy data caused by machine translation errors or inevitable mistakes made by the source language sentiment classifier. The experimental result demonstrates the effectiveness of the proposed model. In the future, we will explore more suitable knowledge representations and knowledge validation in the CredBoost framework.

Acknowledgements

We thank all the anonymous reviewers for their detailed and insightful comments on this paper. The work described in this paper was supported by the Research Grants Council of Hong Kong project (PolyU 5202/12E and PolyU 152094/14E) and the grants from the National Natural Science Foundation of China (61272291, 61472290, 61472291 and 61303115).

References

Carmen Banea and Rada Mihalcea, Janyce Wiebe, Samer Hassan. 2008. Multilingual Subjectivity Analysis Using Machine Translation. In *Proceedings of the 2008 Conference on Empirical Methods in Natual Language Processing*, pages 127-135, Honolulu, October.

Carmen Banea, Yoonjung Choi, Lingjia Deng, Samer Hassan, Michael Mohler, Bishan Yang, Claire

Cardie, Rada Mihalcea, Janyce Wiebe. 2013. CPN-CORE: A Text Semantic Similarity System Infused with Opinion Knowledge. In *Proceedings of the Main Conference and the SHared Task in *SEM 2013*, pages 221-228, Atlanta, Georgia, June 13-14, 2013.

Yejin Choi and Claire Cardie. 2008. Learning with Compositional Semantics as Structural Inference for Subsentential Sentiment Analysis. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 792-801, Honolulu, October 2008.

Kevin Duh and Akinori Fujino and Masaaki Nagata. 2011. Is Machine Translation Ripe for Cross-lingual Sentiment Classification? In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: shortpapers*, pages 429-433, Portland, Oregon, June 19-24, 2011.

Rong-En Fan, Kai-Wei Chang, Cho-Jui Ksieh, Xiang-Rui Wang, Chih-Jen Lin. 2008. LIBLINEAR: A Library for Large Linear Classification. In *Journal of Machine Learning Research*, 9 (2008) 1871-1874.

Micheal Gamon. 2004. Sentiment Classification on Customer Feedback Data: Noisy Data, Large Feature Vectors and the Role of Linguistic Analysis. In *Proceedings of the 20th International Conference on Computational Linguistics*, pages 841-847, CH.

Xavier Glorot, Antoine Bordes, and Yoshua Bengio. 2011. Domain Adaptation for Large-scale Sentiment Classification: A Deep Learning Approach. In *Proceedings of the 28th International Conference on Machine Learning*, pages 513-520, Bellevue, Washington, USA.

Lin Gui, Ruifeng Xu, Qin Lu, Jun Xu, Jian Xu, Bin Liu, Xiaolong Wang. 2014. Cross-lingual Opinion Analysis via Negative Transfer Detection. In *Proceedings of the 52th Annual Meeting of the Association for Computational Linguistics (short paper)*, pages 860-865, Baltimore, Maryland, USA, June 23-25 2014.

Ahmed Hassan and Dragomir Radev. 2010. Identifying Text Polarity Using Random Walks. In *Proceedings of the 48th Annual Meeting on Association for Computational Linguistics*, pages 395-403, Uppsala, Sweden, 11-16 July 2010.

Yulan He, Chenghua Lin, Harith Alani. 2011a. Automatically Extracting Polarity-bearing Topics for Cross Domain Sentiment Classification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Huamn Language Technologies*, pages 123-131, Portland, Oregon, USA.

Yulan He. 2011b. Latent Sentiment Model for Weakly-Supervised Cross-Lingual Sentiment Classification. In *Proceedings of the 33th European Conference on Information Retrieval(ECIR 2011)*, 18-21 Apr 2011, Dublin, Ireland.

KANAYAMA Hiroshi, NASUKAWAA Tetsuya, WATANABE Hideo. 2004. Deeper Sentiment Analysis Using Machine Translation Technology. In *Proceedings of the 20th International Conference on Computational Linguistics*, pages 494-500.

Alistair Kennedy and Diana Inkpen. 2006. Sentiment Classification of Movie and Product Reviews Using Contextual Valence Shifters. *Computational Intelligence*,22(2):110-125.

Quoc Le, Tomas Mikolov. 2014. Distributed Representations of Sentences and Documents. In *Proceedings of the 31th International Conference on Machine Learning*, Beijing, China, 2014. JMLR: W&CP volume 32.

Tao Li, Vikas Sindhwani, Chris Ding, and Yi Zhang. 2009. Knowledge Transformation for Cross-Domain Sentiment Classification. In *Proceedings of the 32th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 716-717, Boston, MA, USA.

Bing Liu. May 2012. *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publisher.

Xinfan Meng, Furu Wei, Xiaohua Liu, Ming Zhou, Ge Xu, Houfeng Wang. 2012. Cross-Lingual Mixture Model for Sentiment Classification. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, pages 572-581, Jeju, Republic of Korea, 8-14 July 2012.

Tony Mullen and Nigel Collier. 2004. Sentiment analysis using support vector machines with diverse inoformation sources. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 412-418, (July 2004) poster paper.

Sinno Jialin Pan and Qiang Yang, Fellow, IEEE. 2010. A Survey on Transfer Learning. In *Journal of IEEE Transactions on Knowledge and Data Engineering*, Vol.22, NO.10, October 2010.

Bo Pang and Lillian Lee, Shivakumar Vaithyanathan. 2002. Thumps Up? Sentiment Classification using Machine Learning Techniques. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing*, pages 79-86, Philadelphia, July 2002.

Kashyap Popat, Balamurali A R, Pushpak Bhattacharyya and Gholamreza Haffari. 2013. The Haves and the Have-Nots: Leverage Unlabeled Corpora for Sentiment Analysis. In *Proceedings of the 51th Annual Meeting of the Association for Computational Linguistics*, pages 412-422, Sofia, Bulgaria, 4-9 August 2013.

Jonathon Read. 2005. Using Emotions to reduce Dependency in Machine Learning Techniques for Sentiment Classification. In *Proceedings of the 43th Annual Meeting on Association for Computational Linguistics Student Research Workshop*, pages 43-48.

Hassan Saif, Yulan He and Harith Alani. 2012. Semantic Sentiment Analysis of Twitter. In *Proceedings of the 11th International Semantics Web Conference ISWC 2012*, Boston, USA.

Stephen Shankland. 2013. Google Translate now serves 200 millon people daily. In *CNET. CBS Interactive Inc*. May 18, 2013.

Richard Socher, Alex Perelygin, Jean Y. Wu, Jason Chuang, Chiristopher D. Manning, Andrew Y. Ng and Christopher Potts. 2013. Recursive Deep Models for Semantics Computationality Over a Sentiment Treebank. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*.

Songbo Tan, Gaowei Wu, Huifeng Tang and Xueqi Cheng. 2007. A Novel Scheme for Domain-transfer Problem in the context of Sentiment Analysis. In *CIKM 2007*, November 6-8, 2007, Lisboa, Portugal.

Peter D. Turney. 2002. Thumps Up or Thumps Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 417-424, Philadelphia, July 2002.

Xiaojun Wan. 2008. Using Bilingual Knowledge and Ensemble Technics for Unsupervised Chinese Sentiment Analysis. In *Proceedings of the 2008 Conference on Empirical Methods in Natual Language Processing*, pages 553-561, Honolulu, October 2008.

Xiaojun Wan. 2009. Co-Training for Cross-Lingual Sentiment Classification. In *Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP*, pages 235-243, Suntec, Singapore, 2-7 August 2009.

Bin Wei and Christopher Pal. 2010. Cross Lingual Adaptation: An Experiment on Sentiment Classifications. In *Proceedings of the 48 Annual Meeting of the Association for Computational Linguistics (short paper)*, pages 258-262, Uppsala, Sweden, 11-16 July 2010.

Ruifeng Xu, Jun Xu and Xiaolong Wang. 2011. Instance Level Transfer Learning for Cross Lingual Opinion Analysis. In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis, ACL-HLT 2011*, pages 182-188, 24 June, 2011, Portland, Oregon, USA.