

Article

Cooperative Hybrid Semi-Supervised Learning for Text Sentiment Classification

Yang Li ¹, Ying Lv ², Suge Wang ^{1,3,*}, Jiye Liang ^{1,3}, Juanzi Li ⁴ and Xiaoli Li ⁵

¹ School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China; YangLi_sxu@163.com (Y.L.); ljy@sxu.edu.cn (J.L.)

² School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China; lvying2016@shu.edu.cn

³ Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, Taiyuan 030006, China

⁴ Computer Science Department, Tsinghua University, Beijing 100084, China; ljjz@keg.cs.tsinghua.edu.cn

⁵ Institute for Infocomm Research, A*Star, Singapore 138632, Singapore; xlli@i2r.a-star.edu.sg

* Correspondence: wsg@sxu.edu.cn; Tel.: +86-351-701-0566

Received: 21 November 2018; Accepted: 21 January 2019; Published: 24 January 2019



Abstract: A large-scale and high-quality training dataset is an important guarantee to learn an ideal classifier for text sentiment classification. However, manually constructing such a training dataset with sentiment labels is a labor-intensive and time-consuming task. Therefore, based on the idea of effectively utilizing unlabeled samples, a synthetical framework that covers the whole process of semi-supervised learning from seed selection, iterative modification of the training text set, to the co-training strategy of the classifier is proposed in this paper for text sentiment classification. To provide an important basis for selecting the seed texts and modifying the training text set, three kinds of measures—the cluster similarity degree of an unlabeled text, the cluster uncertainty degree of a pseudo-label text to a learner, and the reliability degree of a pseudo-label text to a learner—are defined. With these measures, a seed selection method based on Random Swap clustering, a hybrid modification method of the training text set based on active learning and self-learning, and an alternately co-training strategy of the ensemble classifier of the Maximum Entropy and Support Vector Machine are proposed and combined into our framework. The experimental results on three Chinese datasets (COAE2014, COAE2015, and a Hotel review, respectively) and five English datasets (Books, DVD, Electronics, Kitchen, and MR, respectively) in the real world verify the effectiveness of the proposed framework.

Keywords: text sentiment classification; semi-supervised learning; seed selecting; training data updating; alternately co-training

1. Introduction

The convenient interactivity of Web 2.0 technology can enable users to easily interact with each other [1]. The web has become a huge platform for aggregating and sharing information on which people are almost unrestricted to express their opinions, attitudes, feelings, and emotions. The tremendous amount of reviews on the web have been posing a real challenge to the academic community in effectively analyzing and processing text data. Text sentiment classification (TSC) aims to automatically analyze the sentiment orientation of a standpoint, view, attitude, mood, and so on [2–8]. This information can provide potential value to governments, businesses, and users themselves [9–11]. Firstly, it could help governments to understand the public opinions and social phenomenons so as to perceive society trends. Secondly, it could help businesses to capture commercial opportunities,

to managing their online corporate reputation, making precision advertising, and planning business strategies of their e-commerce website. Thirdly, as there are often too many reviews for customers to go through, an automatic text sentiment classifier may be very helpful to a customer to quickly know the review orientations (e.g., positive/negative) about some product to aid customers' decision-making online or offline. The traditional supervised learning algorithms rely on an availability of large-scale training data for sentiment orientation classification [2,12]. Constructing a large-scale and high-quality training data set is a labor-intensive and time-consuming task.

Deep learning, as a kind of supervised learning method, intrinsically learns a useful representation from a large-scale training dataset without human effort [13]. Recently, it was used for text sentiment classification only with a small amount of annotated data [14,15]. The so-called weak supervised learning is proposed in these works. The basic idea of weak supervised learning is to firstly train a deep neural network by using a weakly annotated large-scale external resource, and then the parameters of the deep neural network are adjusted by using the small amount of annotated data.

In this paper, we focus on the problem of text sentiment classification within a small amount of annotated data and without any external resource. Self-learning and active learning are two kinds of important semi-supervised learning techniques by utilizing unlabeled data which are embedded into the learning framework in different ways. Self-learning tries to automatically annotate samples from unlabeled data, and then selects some high-confidence samples to add them into the training dataset in each learning cycle. No manual control is demanded in this process [16–20]. For example, some research proposed self-learning methods for automatically acquiring domain-specific features by using the pseudo-labeled samples with high confidence [20,21]. However, the pseudo-labels with high confidence are not necessarily correct. Label error will be transferred and accumulated in the self-learning process. Unlike no human intervention of self-learning, active learning selects some samples with some expected characteristics from unlabeled data by using a certain strategy to submit them to the domain experts for annotation. The samples with manual labels are added into the training dataset to improve the training data quality. Hence, the active-learning technology is applied into text sentiment classification to improve the quality of the training text set [5,22–24].

In the early works of active-learning for text sentiment classification, some unlabeled texts are randomly selected and annotated as the seed text set, which are used to train the initial classifier of semi-supervised learning [25]. However, randomly selecting unlabeled texts cannot ensure the coherence distribution of them with the whole text data [24,25]. To make sample selection more targeted, different measures were proposed for selecting the so-called informative texts to be annotated according to the different application contexts [5,23,24,26,27].

Co-training is another kind of learning strategy in semi-supervised learning. Under the hypothesis of sufficient and redundant views, the standard co-training algorithm respectively trains two classifiers under two views by using the original labeled dataset. In the co-training stage, each classifier selects some unlabeled samples, which have higher annotated confidence given by the classifier, to add into the training dataset of another classifier. The classifiers are then updated with their respective training dataset [17,28,29]. Co-training can be implemented on multiple component learners of different types to allow performance of their respective advantages, or on multiple component learners of the same type under different views. Some related research shows that integrating heterogeneous learners may achieve better performance than integrating homogeneous learners. Much research showed that the Maximum Entropy model (ME) and Support Vector Machine (SVM) had better performance for text sentiment classification [2,30–32]. For classification problems, ME aims to learn a condition probability distribution of the decision variable about the condition variables under some constraints (also called features) from the training data by following the maximum entropy principle, and SVM aims to learn the optimal classification hyperplane from the training samples.

Focusing on the insufficiency of labeled texts, we propose a synthetic semi-supervised learning framework for text sentiment classification in this paper. In the framework, some measures related to texts are defined, and many techniques, such as Random Swap (RS) clustering, active-learning,

self-learning, and heterogeneous co-training, are synthetically adopted for constructing the initial seed text set, updating the training text set and training the component classifiers, respectively. The final classifier is obtained by integrating the component classifiers.

The remainder of this paper is organized as follows. Section 2 introduces the related works. Section 3 describes the proposed semi-supervised learning framework for text sentiment classification in detail. Section 4 briefly describes the procedures of text sentiment classification. For evaluating the effectiveness of the proposed framework, Section 5 gives the experiment setup, and Section 6 shows the experimental results. Section 7 concludes the paper.

2. Related Works

In this section, we briefly review the related research, in which some techniques of semi-supervised learning, such as self-learning, active-learning, and co-training, are adopted for text sentiment classification.

2.1. Semi-Supervised Learning for Text Sentiment Classification

Seed selection: In a semi-supervised learning framework, the seed set is used to train an initial classifier. However, the quality of seeds may seriously impact the performance of the final classifier. The seeds are selected from all original unlabeled texts by using random sampling under a pre-given threshold, which is used to control the scale of seeds [5,25]. For data with complexity distributions, it is difficult for simple random sampling to guarantee the distribution representativeness of seeds to the data population. Therefore, other methods, such as clustering and the sentiment lexicon-based method, are considered for seed selection [5,22].

Self-learning: The performance of self-learning mainly depends on the selection strategy of the unlabeled samples which take part in the training process. Mallapragada et al. proposed the method of self-learning in which the high confidence samples are gradually added into the training dataset and the previous classifier is linearly combined with the current obtained classifier to promote the classifier performance [18]. Wiegand et al. adopted a rule-based classifier constructed by using a sentiment polarity lexicon and a set of linguistic rules to discriminate the sentiment polarities of unlabeled texts [33]. The texts with annotated sentiment polarities are then used to train a SVM classifier. Qiu et al. proposed the self-supervised, lexicon-based, and corpus-based (SELC) model, which uses a sentiment lexicon to assist the self-learning process [19]. An obvious characteristic of this kind of method is that they need the external resource support of sentiment words, which are mostly domain-specific.

Active learning: Active learning is another kind of semi-supervised learning strategy. It actively chooses some unlabeled samples to be annotated by domain experts, and adds them into the training dataset to improve the learner's performance on new data [5,26,34–36]. Zhou et al. proposed a novel active deep network (ADN) method [35]. ADN is constructed by restricted Boltzmann machines (RBM) based on labeled reviews and an abundance of unlabeled reviews. In the learning framework, they applied active learning to identify reviews that should be labeled as training data, then used the selected labeled reviews and all unlabeled reviews to train ADN architecture. Kranjc et al. proposed an active-learning strategy that selects the nearest samples to the SVM classification hyperplane to annotate the active-learning process for text sentiment classification of stream data [37].

Combination of self-learning and active learning: Some research combine self-learning and active learning for text sentiment classification. Hajmohammadi et al. proposed a semi-supervised learning model in which self-learning and active learning are combined for cross-linguistic text sentiment classification [22]. In their model, unlabeled data are firstly translated from the target language into the source language. These translated data are then augmented into the initial training data in the source language by using active learning and self-learning. A sample density measure for avoiding the selection of outlier samples from unlabeled data is also used into the active-learning algorithm.

2.2. Co-Training for Text Sentiment Classification

Co-training: Li et al. proposed a co-training method with dynamic regulation in semi-supervised learning [38], in which various random subspaces are dynamically generated to deal with the unbalanced class distribution problem, and used it for unbalanced text sentiment classification. Yang et al. presented a novel adaptive multi-view selection (AMVS) co-training method for emotion classification [24]. Two kinds of important distributions, the distribution of feature emotional strength and the distribution of view dimensionality, were proposed to construct multiple discriminative feature views. On the basis of such two distributions, several feature views are iteratively selected from the original feature space in a cascaded way, and the corresponding base classifiers are trained on these views to build a dynamic and robust ensemble. Xia et al. proposed a dual-view co-training algorithm based on dual-view BOW (bags-of-words) representation for semi-supervised sentiment classification [3]. In dual-view BOW, antonymous reviews are automatically constructed by a pair of bags-of-words with opposite views. They pairwisely made use of the original and antonymous views in the training, bootstrapping, and testing process, all based on a joint observation of two views. The experimental results demonstrated the advantages of their approach. Wan et al. proposed a bilingual co-training approach under both English and Chinese views to improve the accuracy of corpus-based polarity classification of Chinese product reviews based on additional unlabeled Chinese data [39]. The co-training algorithm was applied to learn two component classifiers, and such two component classifiers were finally combined into a single sentiment classifier.

3. Proposed Method

In this section, several measures, such as the cluster similarity degree, the cluster uncertainty degree, and the reliability degree, which are related to unlabeled or pseudo-label samples are firstly defined. The whole framework of the proposed method is then described. For the convenience of expression, some symbols and their corresponding meanings are listed in Table 1.

Table 1. Some symbols and their corresponding meanings.

Symbol	Corresponding Meaning
S	total text set
L	labeled text set
U	unlabeled text set
S_d	seed set
x, \dots, y, \dots	text
cl	cluster
$d(x, y)$	distance measure
$sim(x, y)$	similarity measure
$En(x)$	Shannon entropy of x decided by a classifier
var	variance of a random variable
$rel^F(x)$	reliability measure of x decided by a classifier F
$cl_{2\gamma}^{rel}$	high reliability dataset obtained by using self-learning
$cl_{2\beta}^{inf}$	high uncertainty dataset for manual annotation obtained by using active-learning

3.1. Several Related Measures

In our method, several measures for depicting texts from different perspectives, such as the similarity, uncertainty, and reliability, need to be defined.

As we know, the initial seed set is the base of semi-supervised learning. Its representativeness to the distribution of whole unlabeled texts is crucial to the performance of the final classifier obtained based on semi-supervised learning. The local representativeness of an unlabeled text could be reflected by measuring its average similarity with all texts in its local region. The local regions of a text can be obtained by clustering.

Definition 1. Let x be a text, and cl be a cluster of texts. We define the cluster similarity degree $sim(x, cl)$ of x with cl as the average similarity of x , with all texts in the cluster cl .

$$sim(x, cl) = \frac{1}{|cl|} \sum_{y \in cl} sim(x, y) \quad (1)$$

where $sim(x, y) = e^{-d(x,y)}$, and $d(x, y)$ denotes a distance measure of document x and document y .

It should be noted that the similarity $sim(x, y)$ between two texts is defined as a strictly monotone decreasing function of distance which reaches its maximum, 1, at $d(x, y) = 0$, and approaches 0 as the distance becomes larger. This definition is in accord with people's intuition. By Formula (1), one can see that the cluster similarity of a text will degrade into the similarity between two texts, while the cluster only contains a text. Evidently, the cluster similarity of a text is a boundary measure with $0 \leq sim(x, cl) \leq 1$. The larger the cluster similarity of a text is, the more powerful the representativeness to the texts in the cluster is.

Let $x \in U \subseteq S$ be an unlabeled text, $C = \{1, 0\}$ be the two-class label set of texts, and F be a learner/classifier with a probability output from S to C . Evidently, $F(x)$ can be considered as a random variable on C . Suppose $p(x)$ is the probability of $F(x)$ at 1 assigned by F . The probability of $F(x)$ at 0 then equals to $1 - p(x)$. We know that the Shannon entropy of $F(x)$, denoted as $En(x)$, is defined as $En(x) = -p(x) \log p(x) - (1 - p(x)) \log(1 - p(x))$, which depicts the uncertainty degree of $F(x)$, i.e., the uncertainty degree of "x belongs in a prior class".

In accordance with the discussion above, for a text x , a cluster cl , and a learner/classifier F , we took a fusion of Shannon entropy of $F(x)$ and the cluster similarity of x with respect to cl to depict the uncertainty of x from the angle of F .

Definition 2. Let $x \in U$ be an unlabeled text, cl be a cluster of texts, and $En(x)$ be the Shannon entropy of x decided by some classifier F . We defined the cluster uncertainty degree $H(x, cl)$ of x with respect to the cluster cl as:

$$H(x, cl) = sim(x, cl) \times En(x) \quad (2)$$

The property of Shannon entropy means that the larger $En(x)$ is, the harder it is to judge which x belongs to which class for a learner F . If having to assign a class label to x under this occasion, the risk of making mistakes will increase. By Definition 2, $H(x, cl)$ integrates $En(x)$ and $sim(x, cl)$, an annotated sample with larger cluster uncertainty which may simultaneously possess larger Shannon entropy and larger cluster similarity. In other words, a sample with higher cluster uncertainty is a sample of powerful representation and difficult judgement for a learner.

Let F be a classifier (e.g., SVM) which can implicitly generate a probability distribution to an unlabeled text x on the class labels. In general, F then assigns a corresponding class label to x according to the biggest value of the probability distribution. Of course, one hopes to measure the reliability of this method of annotation. For a binary classification problem, the annotated probability distribution of x has the form $P(x) = \{p(x), 1 - p(x)\}$. From the property of the function $p(x)(1 - p(x))$, it follows that the smaller the value of $p(x)(1 - p(x))$ is, the larger the difference of $p(x)$ and $1 - p(x)$ is, i.e., the larger the annotated reliability of x is intuitively. It should be noticed that $p(x)(1 - p(x))$ is precisely the variance of the two-point distribution. Thus, we can define the reliability of the annotated label of x by F as follows.

Definition 3. Let x be an unlabeled text and F be a classifier which assigns a probability distribution $P(x) = \{p(x), 1 - p(x)\}$ to x on the two-class labels. The F reliability degree $rel^F(x)$ of x is defined as:

$$rel^F(x) = 1 - varP(x) = 1 - p(x) + p(x)^2 \quad (3)$$

It should be noticed that $\text{rel}^F(x)$ is a boundary function of $p(x)$ with the boundary $[\frac{3}{4}, 1]$, and it reaches the maximum value 1 at $p(x) = 0$ or $p(x) = 1$; and it reaches the minimum value $\frac{3}{4}$ at $p(x) = \frac{1}{2}$.

3.2. Learning Framework

There are three key parts of the proposed framework, including seed selecting, training data updating, and co-training in the semi-supervised training process. To facilitate understanding, a schematic diagram of seed selecting, training dataset updating, and the co-training process is given in Figure 1.

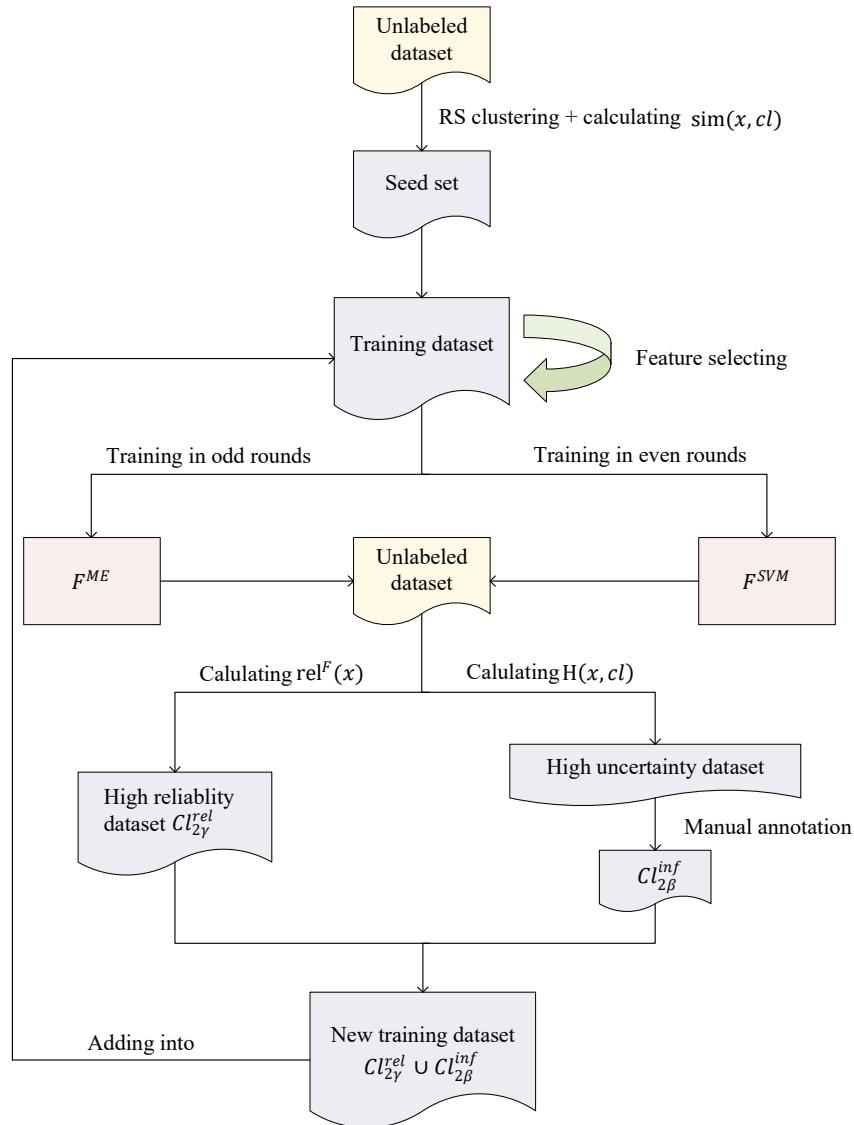


Figure 1. Schematic diagram of the learning framework.

3.2.1. Seed Selection Algorithm

In this subsection, we select a certain number of unlabeled texts from U to construct a seed set. The number of seeds is controlled by a pre-given seed-text ratio α .

To enhance the representativeness of seeds, the original unlabeled data are firstly clustered by using the RS algorithm [40]. On this basis, seeds are selected from each cluster by using the cluster similarity measure of a text defined by Formula (1) under a pre-specified percentage threshold α . The seed selection algorithm is described in Algorithm 1.

Algorithm 1 Acquiring a seed set based on RS and the cluster similarity degree $\text{sim}(x, cl)$

Require: S (each text $x \in S$ is processed by using the software for data preprocessing), the seed-text ratio α .

Ensure: the seed set $Sd; \{\langle x, \text{sim}(x, cl) \rangle\}_{x \in S}$.

Step 1. $Sd = \emptyset$;

Step 2. Vectorization using word2vec for each document $x \in S$;

Step 3. Obtain the cluster number N_{cl} and the clustering result $\{cl_i\}_{i=1}^{N_{cl}}$ of S by using the RS algorithm [40];

Step 4. For each $i = 1, 2, \dots, N_{cl}$

Step 5. For each document $x \in cl_i$

Step 6. Calculate the cluster similarity degree $\text{sim}(x, cl_i)$ according to Formula (1);

Step 7. End for

Step 8. Rank all texts of cl_i in descending order of the cluster similarity degree $\text{sim}(x, cl_i)$;

Step 9. Select text set Sd_i on the top $\alpha * |cl_i|$ of $\text{sim}(x, cl_i)$;

Step 10. Obtain the seed subset Sd_i by manually annotating each text in Sd_i ;

Step 11. $Sd = Sd \cup Sd_i$;

Step 12. End for

Step 13. End

3.2.2. Training Data Update Process

After Algorithm 1, a seed set is generated. By using the seed set, an initial classifier can be trained. In the subsequent learning process, some texts from U with their pseudo- or real labels, which are annotated by a learner or by domain experts, are constantly selected and added into the seed set to update the training dataset. The updating process of the training dataset is alternately performed in two ways—active learning and self-learning.

(1) Updating the training dataset by using active-learning:

In an iterative round, each unlabeled text, x from U , is assigned a probability distribution $P(x)$ on the class label space by a current component classifier F . According to Formula (2), we can calculate the cluster uncertainty degree $H(x, cl)$ of each x , and then respectively select the top β positive and negative texts to add into the training dataset with their real labels given by domain experts.

(2) Updating the training dataset by using self-learning:

In an iterative round, each unlabeled text, x from U , is assigned a probability distribution $P(x)$ on the class label space by a current component classifier, F . According to Formula (3), we can calculate the reliability degree $rel^F(x)$ of each x , and then respectively select the top γ positive and negative texts to add into the training dataset with their pseudo-labels given by F .

3.2.3. Co-Training Strategy

In our framework, co-training is put on two learning algorithms of F^{ME} and F^{SVM} in a rotation pattern. In each iterative round, the classifier generated by F^{ME} or F^{SVM} annotates all the unlabeled texts. Some high-uncertainty pseudo-label texts are selected by active-learning for manual annotation and also added into the training dataset, while some pseudo-label texts with higher reliability degrees are directly selected by self-learning and added into the training dataset. The updated training dataset is used to retrain the learner generated by another learning algorithm. The two co-training units of F^{ME} and F^{SVM} learners are integrated into a final classifier when the training process ends.

In each iterative round, feature selection is performed on the updated training dataset and the feature space is then updated. In each of the odd iterative rounds, the learner F^{ME} is trained on the current training dataset. In each of the even iterative rounds, the learner F^{SVM} is trained on the current training dataset. By using active learning and self-learning, a high-uncertainty dataset, $cl_{2\beta}^{inf}$ is obtained

for manually annotating the high-reliability dataset $cl_{2\gamma}^{rel}$, and the new training dataset $cl_{2\beta}^{inf} \cup cl_{2\gamma}^{rel}$, respectively. The new training dataset is added into the old training dataset. The above-mentioned training process is alternately performed until the termination condition is met. An algorithm of co-training with active learning and self-learning is also given in Algorithm 2.

Algorithm 2 Co-training with self-learning and active learning.

Require: Seed set Sd , unlabeled text set $U \subseteq S$, text set $\{< x, sim(x, cl) >\}_{x \in U}$ in which each text is attached with its cluster label, text number threshold 2β for active learning, text number threshold 2γ for self-learning, the number of iterations $2N$.

Ensure: Classifier F^{ME} and classifier F^{SVM}

Step 1. $L = Sd; i = 1;$

Step 2. Obtain the new feature set T by using the improved fisher's criterion [41] on L , and re-express L on T ;

Step 3. Train F^{ME} on $L; F = F^{ME};$

Step 4. For each $x \in U$

Step 5. Obtain the probability distribution $P(x)$ of x on the label set by F ;

Step 6. Calculate the cluster uncertainty degree $H(x, cl)$ by Formula (2) and calculate the reliability degree $rel^F(x)$ by Formula (3);

Step 7. End for

Step 8. Rank all texts of U in descending order of $H(x, cl)$, and select and manually annotate the positive and negative texts on the top β of U to obtain the high-uncertainty positive and negative text sets cl_{β}^{inf+} and cl_{β}^{inf-} , respectively, the high-uncertainty text set $cl_{2\beta}^{inf} = cl_{\beta}^{inf+} \cup cl_{\beta}^{inf-}$;

Step 9. Rank all texts of U in descending order of $rel^F(x)$, select the positive and negative texts on the top γ of U to obtain the high-reliability positive and negative text sets cl_{γ}^{rel+} and cl_{γ}^{rel-} respectively, the high-reliability text set $cl_{2\gamma}^{rel} = cl_{\gamma}^{rel+} \cup cl_{\gamma}^{rel-}$;

Step 10. $L = L \cup cl_{2\beta}^{inf} \cup cl_{2\gamma}^{rel}; U = U - cl_{2\beta}^{inf} - cl_{2\gamma}^{rel};$

Step 11. If $i = 2 * N + 1$

Step 12. Go to step 22;

Step 13. End if

Step 14. $i = i + 1;$

Step 15. If $i \bmod 2 = 0$

Step 16. Obtain the new feature set T by using the improved fisher's criterion [41] on L , and re-express L on T ;

Step 17. Train F^{SVM} on $L; F = F^{SVM};$

Step 18. Go to step 4;

Step 19. Else

Step 20. Go to step 2;

Step 21. End If

Step 22. End.

3.2.4. Ensemble Classifier

After the co-training process, two component classifiers, F^{ME} and F^{SVM} , can be obtained. For an unlabeled text x and a class label y , suppose that the annotation probabilities of x are $p^{ME}(y|x)$ and $p^{SVM}(y|x)$ assigned by F^{ME} and F^{SVM} , respectively, where the classification function is defined as follows:

$$F(x) = \hat{y} = \arg \max_y \{p^{ME}(y|x) + p^{SVM}(y|x)\} \quad (4)$$

4. Procedures of Text Sentiment Classification

In this section, the procedures of text sentiment classification are briefly described.

Step 1: Data preprocessing. Non-textual information is removed, followed by removal of stop words and stemming. For English datasets, this is using the Natural Language Toolkit (NLTK) (<http://www.nltk.org>). Chinese datasets, on the other hand, are preprocessed using the NLPIR-ICTCLAS platform (<http://ictclas.nlpir.org/index.html>). Finally, documents are represented using word2vec (<https://github.com/Embedding/Chinese-Word-Vectors>) for Chinese datasets, or GloVe (<https://nlp.stanford.edu/projects/glove/>) for English datasets.

Step 2: Dataset partitioning. In order to test the performance of the proposed method, we performed five-fold cross-validation on the datasets in different domains. Each dataset was randomly split into five text subsets. During each part of the five-fold cross-validation process, a single text subset was retained for testing, and the other four text subsets were merged as the unlabeled text set, from which the seed set was selected.

Step 3: Seed set acquisition. As discussed in Section 3.2.1, the seed set was acquired by using Algorithm 1.

Step 4: Feature selection and text representation. As mentioned in Algorithm 2, the features were selected on the current training dataset by using the improved Fisher's criteria [41]. The feature-weight presence measure was used to construct the vector representation of texts.

Step 5: Updating the training dataset and co-training. The training dataset was updated and the component classifiers co-trained in each iterative round by using Algorithm 2.

Step 6: Constructing the ensemble classifier. The final classifier was constructed by using Formula (4).

Step 7: Classifying the test texts. All the test texts were classified by using the final classifier.

5. Experiment Design

In this section, we introduce the experiment datasets, evaluation measures, and training patterns.

5.1. Experimental Data and Parameter-Setting

For verifying the effectiveness of the proposed method, we conducted multi-group experiments on eight text datasets as follows.

Chinese datasets. COAE2014 (<http://tsaop.com:8066/web/resource.html>) and COAE2015 (<http://tsaop.com:8066/web/resource.html>), are two Weibo text datasets for Task 4 and Task 2 of the Chinese opinion analysis evaluation in 2014 and 2015, respectively. COAE2014 contains 6000 Weibo texts on the domains of Mobile telephone, Jadeite, and Insurance. COAE2015 contains 6846 Weibo sentences on more than a dozen domains. Another Chinese dataset, Hotel review (https://download.csdn.net/download/sinat_30045277/9862005) data, is a part of the Hotel sentiment corpus constructed by the Songbo Tan research team of the Computation Technology Institute of Chinese Academy of Sciences. We selected 3000 texts from the original dataset.

English datasets. Four product-review datasets [42] were obtained from Amazon, which belonged to the four domains of Books, DVD, Electronics, and Kitchen, respectively. Another English dataset was the Movie Review (MR) (<http://www.cs.cornell.edu/people/pabo/movie-review-data/>), which was selected if the rating was in stars or a numerical score. We implemented experiments on a polarity dataset, v2.0.

Each experiment dataset was divided into either the training set or testing set. According to the method introduced in Section 3, a few documents from the training set were selected as the initial seeds under the presupposed parameters $\alpha = 0.1$, and the documents of the rest took their sentiment labels out as unlabeled documents. The numbers of documents for the different types of each experiment dataset are shown in Table 2, where WP is the number of positive training texts; WN is the number of negative training texts; SE is the number of initial seed texts; U is the number of unlabeled texts;

PT is the number of positive testing texts; and NT is the number of negative testing texts. According to the literature [22,43], the thresholds of text number are $2\beta = 2$ and $2\gamma = 4$ for active learning and self-learning, respectively; the total number of iterations on the Chinese dataset and the English dataset are N = 200 and N = 120 for Algorithm 2, respectively. The distance function $d(x,y)$ is the Euclidean distance in Formula (1). The feature dimension is 1000 [41].

The ME classifier was implemented with Mallet Toolkits (<http://mallet.cs.umass.edu/>). The SVM classifier adopted the LIBSVM (<https://www.csie.ntu.edu.tw/~cjlin/libsvm/>).

Table 2. Numbers of samples of different types in eight experiment datasets.

	Dataset	WP	WN	SE	U	PT	NT
Chinese dataset	COAE2014	3000	3000	480	4320	600	600
	COAE2015	3423	3423	548	4870	684	684
	Hotel review	1500	1500	240	2160	300	300
Enlish dataset	Books	1000	1000	200	1400	200	200
	DVD	1000	1000	200	1400	200	200
	Electronics	1000	1000	200	1400	200	200
	Kitchen	1000	1000	200	1400	200	200
	MR	1000	1000	200	1400	200	200

5.2. Evaluation Measures

Four classical measures—Precision, Recall, F-value, and Accuracy—generally used in text classification evaluation were adopted in this paper. By PP (PN), RP (RN), and FP (FN), we mean the Precision, Recall, and F-value of a method on the positive (negative) testing texts, respectively. By “Acc” we mean the accuracy of a method on the testing texts. These evaluation measures can be calculated by the following formulas, respectively:

$$RN(Recall) = \frac{TN}{FP + TN} \quad (5)$$

$$PN(Precision) = \frac{TN}{FN + TN} \quad (6)$$

$$FN(F1 - measure) = \frac{2 \times RN \times PN}{RN + PN} \quad (7)$$

$$RP(Recall) = \frac{TP}{TP + FN} \quad (8)$$

$$PP(Precision) = \frac{TP}{TP + FP} \quad (9)$$

$$FP(F1 - measure) = \frac{2 \times RP \times PP}{RP + PP} \quad (10)$$

$$Acc(Accuracy) = \frac{TP + TN}{TP + FP + FN + TN} \quad (11)$$

where TP (true positive) denotes the number of testing texts whose true class is positive, where they are classified into the positive class; FP (false positive) denotes the number of testing texts whose true class is negative, but they are classified into the positive class; FN (false negative) denotes the number of testing texts whose true class is positive, but they are classified into the negative class; TN (true negative) denotes the number of testing texts whose true class is negative, and they are classified into the negative class.

5.3. Training Pattern Design

In this study, we designed four training patterns generated by using single or cooperative training to verify the effectiveness of the proposed method. The abbreviations and meanings of these training patterns are listed below.

(1) CAS-SVM: Only train a SVM classifier by using the seed set acquired from Algorithm 1, and using the unlabeled data with active learning and self-learning in the subsequent iterative process.

(2) CAS-ME: Only train a ME classifier by using the seed set acquired from Algorithm 1, and using the unlabeled data with active learning and self-learning in the subsequent iterative process.

(3) Random-CT: Cooperatively train classifiers of ME and SVM by using the seed set generated only by simple random sampling from the unlabeled data, and using the unlabeled data with active learning, self-learning, and co-training in the subsequent iterative process.

(4) CASCT: Cooperatively train classifiers of ME and SVM by using the seed set obtained by Algorithm 1, and using the unlabeled data with active learning, self-learning, and co-training in the subsequent iterative process.

6. Experimental Results and Analysis

According to the procedures of text sentiment classification and the training patterns described in Sections 4 and 5, respectively, we conducted the following experiments on eight text datasets. It should be noted that all of the experimental results are in the case of five-fold cross-validation.

6.1. On the Seed Selection Method

To inspect the effectiveness of the seed selection method Algorithm 1, we compared it with the random selection method on the three Chinese datasets. The experimental results are shown in Table 3 and Figure 2a–c.

Table 3. Experimental results of seed selection based on Algorithm 1 and random method.

Dataset	Training Pattern	RN	RP	PN	PP	FN	FP	Acc
COAE2014	Random-CT	0.725	0.741	0.737	0.729	0.731	0.735	0.733
	CASCT	0.728	0.79	0.748	0.772	0.738	0.781	0.761
COAE2015	Random-CT	0.666	0.672	0.673	0.671	0.666	0.669	0.670
	CASCT	0.698	0.753	0.739	0.714	0.718	0.733	0.726
Hotel review	Random-CT	0.789	0.836	0.829	0.802	0.807	0.817	0.813
	CASCT	0.847	0.834	0.837	0.846	0.842	0.839	0.841

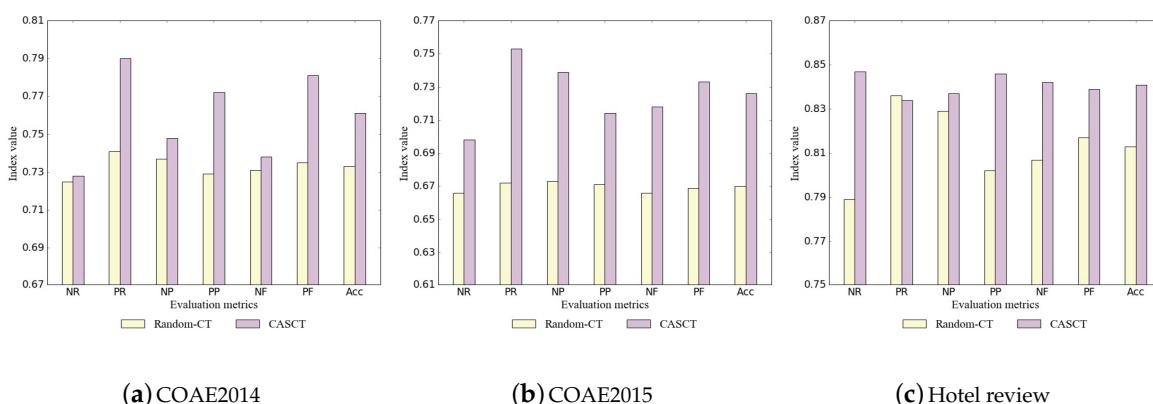


Figure 2. Experimental results of seed selection based on Algorithm 1 and random method.

From Table 3 and Figure 2a–c, we can see that the proposed seed selecting strategy is significantly superior to the random sampling method in every index for the entire Chinese datasets, except the

index RP on the hotel review dataset. We present the schematic diagrams of clustering results after dimension reduction in Figure 3a–c.

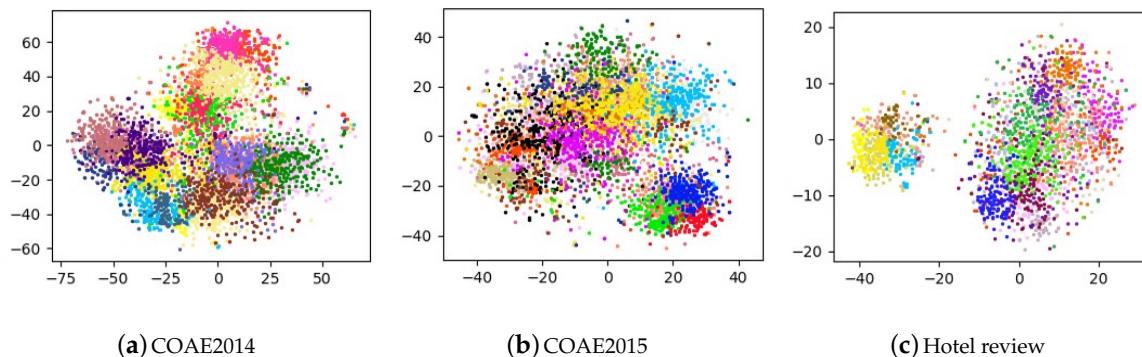


Figure 3. Experimental results of the RS clustering on three Chinese datasets.

6.2. Comparative Experiments on English Datasets

We conducted experiments to evaluate the performance on classification accuracy of our method, CASCT (Section 5.3) by comparing it with the following methods on the five English datasets.

Active learning [35]: An active learning method in which the unlabeled reviews with higher uncertainty are manually annotated, and the SVM classifier is iteratively retrained on the updated training dataset.

Self-training-S [43]: A self-training and active learning approach in which the subspace-based technique is utilized to explore a set of good features, and some informative samples are selected to be annotated by the learner. The top two informative samples are also selected for manual annotation in each iteration.

Dual-view Co-training [3]: A dual-view co-training approach which utilizes dual-view representation by observing two opposite sides of one review.

Active Deep Network (ADN) [35]: Constructed by RBM with unsupervised learning based on labeled reviews and an abundance of unlabeled reviews.

Information AND (IADN) [35]: A variant of ADN. It introduces the information density into ADN for choosing the manual labeled reviews.

LCCT [44]: A Lexicon-based and Corpus-based co-training model for semi-supervised sentiment classification.

The evaluation results of performed experiments are shown in Table 4. It should be noted that the column “Avg. (four domains from Amazon)” in Table 4 is the average value of a method on four dataset domains—Books, DVD, Electronics, and Kitchen—from Amazon.

Table 4. Sentiment classification performance of different methods on English datasets.

Approach	Books	DVD	Electronics	Kitchen	Avg. (Four Domains from Amazon)	MR
Active-learning	0.586	0.58	0.633	0.681	0.62	-
Self-training-S	0.748	0.738	0.765	0.788	0.76	-
Dual-view Co-training	0.721	0.738	0.769	0.780	0.752	-
ADN	0.69	0.716	0.768	0.775	0.737	-
IADN	0.697	0.722	0.779	0.782	0.745	-
LCCT	-	-	-	-	-	0.815
CASCT	0.744	0.771	0.778	0.80	0.773	0.839

From Table 4, we can see that the proposed approach gains a major improvement of nearly 15% over the average accuracy of Books, DVD, Electronics, and Kitchen datasets than the baseline method (Active-learning). Our approach gained the most significant improvement on the DVD dataset, with 3.3% higher than the best accuracy from other methods. We roughly ranked the average accuracy

of the eight semi-supervised methods as follows: CASCT > Self-training-S > Dual-view Co-training > IADN > AND > Active learning on four domain datasets, and CASCT > LCCT on the MR dataset. Among these methods being compared, Self-training-S obtained the highest sentiment classification accuracy, and gained a flat performance with our method CASCT for the Book dataset. The main reason for this is that the top two informative samples were selected for manual annotation in each iteration in its implementation. IADN also had a flat result with our method for the Electronics dataset because the two most uncertain samples were selected and labeled in each iteration. Our approach, considering both the informative and uncertain samples, significantly outperforms all the other methods in regard to DVD, Kitchen, and MR domain datasets.

6.3. Experiments of the Designed Patterns on Chinese Datasets

To inspect the effectiveness of the proposed method, a group of experiments were conducted on three Chinese datasets to compare the performances of the four patterns designed in Section 5.3. The experimental results are shown in Table 5.

Table 5. Experimental results of the different training patterns on Chinese datasets.

Dataset	Training Pattern	RN	RP	PN	PP	FN	FP	Acc
Hotel review	CAS-SVM	0.825	0.839	0.838	0.829	0.831	0.833	0.832
	CAS-ME	0.841	0.845	0.844	0.842	0.843	0.843	0.843
	CASCT	0.847	0.834	0.837	0.846	0.842	0.839	0.841
COAE2014	CAS-SVM	0.701	0.788	0.739	0.755	0.72	0.771	0.748
	CAS-ME	0.725	0.79	0.748	0.77	0.736	0.78	0.76
	CASCT	0.728	0.79	0.748	0.772	0.738	0.781	0.761
COAE2015	CAS-SVM	0.692	0.698	0.683	0.69	0.69	0.692	0.691
	CAS-ME	0.688	0.738	0.725	0.703	0.705	0.72	0.713
	CASCT	0.698	0.753	0.739	0.714	0.718	0.733	0.726

As we know, the performance of a SVM classifier relies more on the so-called support vectors which are close to the classification hyperplane. The modeling method of ME aims to learn the class distribution of a population from the training dataset. The performance of a ME classifier relies on not only the label accuracy of the training data, but also on the training data size.

From Table 5, comparing with CAS-ME and CAS-SVM, CASCT gained the best values under all indexes on the datasets COAE2014 and COAE2015, except the dataset, “Hotel review”. For the “Hotel review” dataset, CASCT obtained the best values under the indexes RN and PP, and CAS-ME achieved the optimal values under the other indexes RP, PN, FN, FP, and Acc. Because the dataset “Hotel review” had the smallest size among the three datasets, the same iteration times of the training process means that we could obtain larger-proportion training data to the entire data for “Hotel review”. Thus, we guess this reason results in a flat performance of CAS-ME with CASCT for the “Hotel review” dataset. However, along with the increases of data size, such as COAE2014 and COAE2015, the advantage of CASCT is then displayed.

To inspect the stability of the pattern CSACT, we implemented it with incremental iterations of the alternative cooperative training strategy based on the hybrid of active learning and self-learning on the COAE2015 dataset. The experimental result is shown in Figure 4.

From Figure 4, we can see that the performance gets better and better with increasing iteration time. Another advantage of CASCT can also be seen from Figure 4, where it was able to obtain a stable result by being iterated 300 times. This means that CASCT is able to achieve a better result without larger-proportion training data in the semi-supervised learning process.

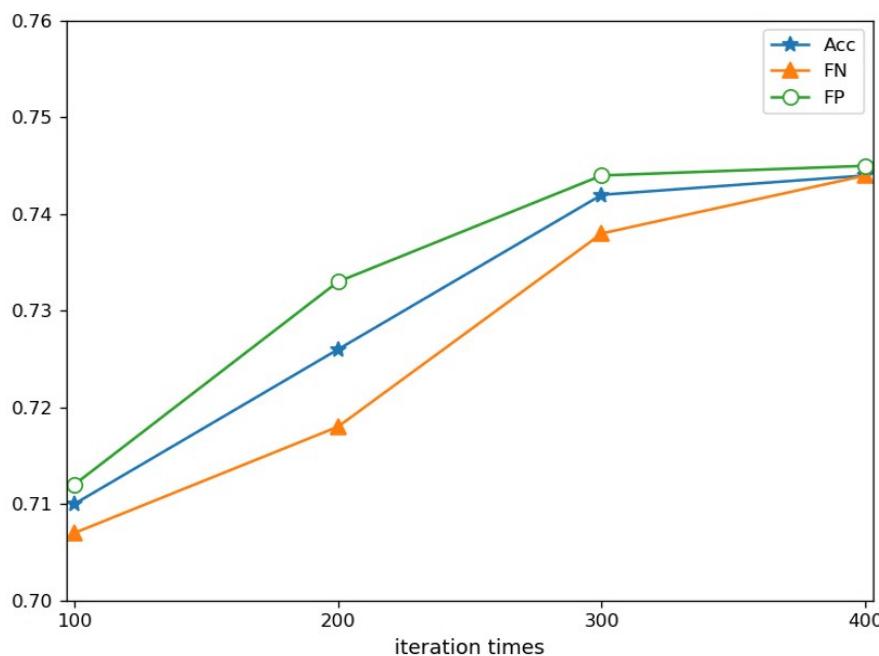


Figure 4. Experimental result of CSACT on COAE2015.

7. Conclusions

The effectiveness of most supervised learning methods often needs the guarantee of a large-scale and high-quality labeled training dataset. However, in reality, such a labeled training dataset is not easy to get. Semi-supervised learning is an effective method for coping with the insufficiency of labeled data in machine learning. In this paper, we proposed a cooperative semi-supervised learning method based on the hybrid mechanism of active learning and self-learning for text sentiment classification. The main characteristics of the proposed method are summarized as follows.

(1) For seed selection, the cluster similarity measure of a sample was firstly defined to depict the representativeness of the sample to a cluster. Unlike the random selection often adopted by many semi-supervised learning methods, we clustered the data by utilizing the RS algorithm, then selected the seeds from each cluster according to the cluster similarity measure and a pre-set ratio threshold. Such a clustering, along with the cluster similarity-based method, can keep the distribution consistency of the seed set with all the data, and it may weaken the model bias phenomenon to a certain extent.

(2) In the semi-supervised learning framework, some unlabeled samples were selected to be added into the training dataset by the current learner after an annotating procedure was executed in some way. The label quality of the added samples are crucial for guaranteeing the performance of the final obtained classifier. To this end, we proposed three measures of a sample: the cluster similarity degree, the cluster uncertainty degree with respect to the learner, and the reliability degree with respect to a learner. These measures were embedded into active learning and self-learning procedures to select expected unlabeled samples to improve the label quality of the training dataset.

(3) In the training process, we designed an alternative cooperation strategy with two kinds of heterogeneous learning algorithms, ME and SVM. Two corresponding component classifiers were also integrated as the final classifier.

A series of experiments were done to verify the effectiveness of the proposed method. The experimental results on eight real-world datasets—COAE2014, COAE2015, Hotel review, Books, DVD, Electronics, Kitchen, and MR—showed that: (a) The proposed seed-selection method based on clustering plus the cluster similarity of a sample is superior to the random selection method; (b) the proposed method outperforms some existing active-learning methods or cooperative strategies;

and (c) the cooperative training strategy combining ME and SVM is superior to the non-cooperative training strategy.

In past research, text sentiment classification has been a rather difficult problem due to its dependence on many different factors, such as languages, domains, text characteristics, and training data. Thus, a single technique might not work well. Some new technologies, such as representation learning and deep learning, are worthy of testing in our future research.

Author Contributions: Y.L. (Yang Li) and S.W. conceived of the proposed approach and analyzed the results; Y.L. (Yang Li) designed the algorithms and wrote the paper; Y.L. (Yang Li) and Y.L. (Ying Lv) performed the experiments; Y.L. (Yang Li), S.W., J.L. (Jiye Liang), X.L. and J.L. (Juanzi Li) discussed the data and revised the manuscript. All authors have read and approved the final manuscript.

Funding: The works described in this paper are supported by the National Natural Science Foundation of China (NSFC Nos. 61632011, 61573231, 61432011, 61672331) and the Key research and development projects of Shanxi Province, China (201803D421024).

Acknowledgments: The authors would like to thank all anonymous reviewers.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Abbasi, A.; Chen, H.; Salem, A. Sentiment analysis in multiple languages: feature selection for opinion classification in Web forums. *ACM Trans. Inf. Syst.* **2008**, *26*, 1–34. [[CrossRef](#)]
2. Pang, B.; Lee, L.; Vaithyanathan, S. Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of the Empirical Methods in Natural Language Processing, Philadelphia, PA, USA, 6–7 July 2002; pp. 79–86.
3. Xia, R.; Wang, C.; Dai, X.Y.; Li, T. Co-training for semi-supervised sentiment classification based on dual-view bags-of-words representation. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Nanchang, China, 29–30 October 2015; pp. 1054–1063.
4. Giatsoglou, M.; Vozalis, M.G.; Diamantaras, K.; Vakali, A.; Sarigiannidis, G.; Chatzisavvas, K.C. Sentiment analysis leveraging emotions and word embeddings. *Expert Syst. Appl.* **2017**, *69*, 214–224. [[CrossRef](#)]
5. Wu, F.; Huang, Y.; Yan, J. Active sentiment domain adaptation. In Proceedings of the Meeting of the Association for Computational Linguistics, Vancouver, BC, Canada, 30 July–4 August 2017; pp. 1701–1711.
6. Fernández-Gavilanes, M.; Álvarez López, T.; Juncal-Martínez, J.; Costa-Montenegro, E. Unsupervised method for sentiment analysis in online texts. *Expert Syst. Appl.* **2016**, *58*, 57–75. [[CrossRef](#)]
7. Bandhakavi, A.; Wiratunga, N.; Padmanabhan, D.; Massie, S. Lexicon based feature extraction for emotion text classification. *Pattern Recognit. Lett.* **2017**, *93*, 133–142. [[CrossRef](#)]
8. Colhon, M.; Vlăduțescu, S.; Negrea, X. How objective a neutral word is? A neutrosophic approach for the objectivity degrees of neutral words. *Symmetry* **2017**, *9*, 280. [[CrossRef](#)]
9. Daniel, M.; Rui, F.N.; Horta, N. Company event popularity for financial markets using twitter and sentiment analysis. *Expert Syst. Appl.* **2016**, *71*, 111–124. [[CrossRef](#)]
10. Schumaker, R.P.; Jarmoszko, A.T.; Labedz, C.S. Predicting wins and spread in the Premier League using a sentiment analysis of twitter. *Dicis. Support Syst.* **2016**, *88*, 76–84. [[CrossRef](#)]
11. Nguyen, T.H.; Shirai, K.; Velcin, J. Sentiment analysis on social media for stock movement prediction. *Expert Syst. Appl.* **2015**, *42*, 9603–9611. [[CrossRef](#)]
12. Kim, Y. Convolutional Neural Networks for sentence classification. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, 25–29 October 2014; pp. 1746–1751.
13. Sun, X.; Li, C.; Ren, F. Sentiment analysis for Chinese microblog based on deep neural networks with convolutional extension features. *Neurocomputing* **2016**, *210*, 227–236. [[CrossRef](#)]
14. Lee, G.; Jeong, J.U.; Seo, S.; Kim, C.Y.; Kim, C.; Kang, P. Sentiment classification with word localization based on weakly supervised learning with a convolutional neural network. *Knowl.-Based Syst.* **2018**, *152*, 70–82. [[CrossRef](#)]
15. Wei, Z.; Guan, Z.; Long, C.; He, X.; Deng, C.; Wang, B.; Quan, W. Weakly-supervised deep embedding for product review sentiment analysis. *IEEE Trans. Knowl. Data Eng.* **2018**, *30*, 185–197.

16. Zhou, Z.H.; Zhan, D.C.; Yang, Q. Semi-supervised learning with very few labeled training examples. In Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 22–23 July 2007; pp. 675–680.
17. Zhou, Z.H. Semi-supervised learning by disagreement. *Knowl. Inf. Syst.* **2010**, *24*, 415–439. [[CrossRef](#)]
18. Mallapragada, P.K.; Jin, R.; Jain, A.K.; Liu, Y. SemiBoost: boosting for semi-supervised learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 2000–2014. [[CrossRef](#)]
19. Qiu, L.; Zhang, W.; Hu, C.; Zhao, K. SELC: A self-supervised model for sentiment classification. In Proceedings of the ACM Conference on Information and Knowledge Management, Hong Kong, China, 2–6 November 2009; pp. 929–936.
20. He, Y.; Zhou, D. Self-training from labeled features for sentiment analysis. *Inf. Process. Manag.* **2011**, *47*, 606–616. [[CrossRef](#)]
21. Ortigosa-Hern.; Ndez, J.; Rodr.; Guez, J.D.; Alzate, L.; Lucania, M.; Inza, I.; Lozano, J.A. Approaching sentiment analysis by using semi-supervised learning of multi-dimensional classifiers. *Neurocomputing* **2012**, *92*, 98–115. [[CrossRef](#)]
22. Hajmohammadi, M.S.; Ibrahim, R.; Selamat, A.; Fujita, H. Combination of active learning and self-training for cross-lingual sentiment classification with density analysis of unlabelled samples. *Inf. Sci.* **2015**, *317*, 67–77. [[CrossRef](#)]
23. Zhu, X.; Goldberg, A.B.; Brachman, R.; Dietterich, T. Introduction to semi-supervised learning. *Semi-Superv. Learn.* **2009**, *3*, 130. [[CrossRef](#)]
24. Yang, Z.; Liu, Z.; Liu, S.; Min, L.; Meng, W. Adaptive multi-view selection for semi-supervised emotion recognition of posts in online student community. *Neurocomputing* **2014**, *144*, 138–150. [[CrossRef](#)]
25. Settles, B. Active learning literature survey. *Univ. Wis.* **2010**, *39*, 127–131.
26. Li, S.; Xue, Y.; Wang, Z.; Zhou, G. Active learning for cross-domain sentiment classification. In Proceedings of the International Joint Conference on Artificial Intelligence, Beijing, China, 3–19 August 2013; pp. 2127–2133.
27. Tan, Z.; Li, B.; Huang, P.; Ge, B.; Xiao, W. Neural relation classification using selective attention and symmetrical directional instances. *Symmetry* **2018**, *10*, 357. [[CrossRef](#)]
28. Jiang, Z.; Zhang, S.; Zeng, J. *A Hybrid Generative/Discriminative Method for Semi-Supervised Classification*; Elsevier Science Publishers B. V.: Amsterdam, The Netherlands, 2013; Volume 37, pp. 137–145.
29. Zhang, Y.; Wen, J.; Wang, X.; Jiang, Z. Semi-supervised learning combining co-training with active learning. *Expert Syst. Appl.* **2014**, *41*, 2372–2378. [[CrossRef](#)]
30. Wang, G.; Sun, J.; Ma, J.; Xu, K.; Gu, J. Sentiment classification: The contribution of ensemble learning. *Decis. Support Syst.* **2014**, *57*, 77–93. [[CrossRef](#)]
31. Xia, R.; Zong, C.; Li, S. Ensemble of feature sets and classification algorithms for sentiment classification. *Inf. Sci.* **2011**, *181*, 1138–1152. [[CrossRef](#)]
32. Catal, C.; Nangir, M. A sentiment classification model based on multiple classifiers. *Appl. Soft Comput.* **2017**, *50*, 135–141. [[CrossRef](#)]
33. Wiegand, M.; Klenner, M.; Klakow, D. Bootstrapping polarity classifiers with rule-based classification. *Lang. Resour. Eval.* **2013**, *47*, 1049–1088. [[CrossRef](#)]
34. Tong, S.; Koller, D. Support vector machine active learning with applications to text classification. *J. Mach. Learn. Res.* **2002**, *2*, 999–1006.
35. Zhou, S.; Chen, Q.; Wang, X. Active deep learning method for semi-supervised sentiment classification. *Neurocomputing* **2013**, *120*, 536–546. [[CrossRef](#)]
36. Fu, Y. A survey on instance selection for active learning. *Knowl. Inf. Syst.* **2013**, *35*, 249–283. [[CrossRef](#)]
37. Kranjc, J.; Smailović, J.; Podpečan, V.; Grčar, M.; Žnidaršič, M.; Lavrač, N. Active learning for sentiment analysis on data streams: Methodology and workflow implementation in the CrowdFlows platform. *Inf. Process. Manag.* **2015**, *51*, 187–203. [[CrossRef](#)]
38. Li, S.; Wang, Z.; Zhou, G.; Lee, S.Y.M. Semi-supervised learning for imbalanced sentiment classification. *J. R. Stat. Soc.* **2008**, *172*, 530.
39. Wan, X. Bilingual co-training for sentiment classification of chinese product reviews. *Comput. Linguist.* **2011**, *37*, 587–616. [[CrossRef](#)]
40. Fränti, P. Efficiency of random swap clustering. *J. Big Data* **2018**, *5*, 13. [[CrossRef](#)]
41. Wang, S.; Li, D.; Song, X.; Wei, Y.; Li, H. A feature selection method based on improved fisher’s discriminant ratio for text sentiment classification. *Expert Syst. Appl.* **2011**, *38*, 8696–8702. [[CrossRef](#)]

42. Blitzer, J.; Dredze, M.; Pereira, F. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, Prague, Czech Republic, 25–27 June 2007.
43. Wei, G.; Li, S.; Xue, Y.; Meng, W.; Zhou, G. Semi-supervised sentiment classification with self-training on feature subspaces. In Proceedings of the Workshop on Chinese Lexical Semantics, Zhengzhou, China, 20–22 May 2014; pp. 231–239.
44. Yang, M.; Tu, W.; Lu, Z.; Yin, W.; Chow, K. LCCT: A semi-supervised model for sentiment classification. In Proceedings of the Human Language Technologies: The 2015 Annual Conference of the North American Chapter of the ACL, Budapest, Hungary, 24–26 November 2015; pp. 546–555.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).