Journal of Computer Science & Technology, Volume 19, Number 1, April 2019

- ORIGINAL ARTICLE -

Semi-Supervised Target-Dependent Sentiment Classification for Micro-Blogs

Clasificación de Sentimientos Semi-Supervisada y Dependiente de Objetivo para Micro-Blogs

Shadi I. Abudalfa[1] and Moataz A. Ahmed[2]

Collage of Computer Science and Engineering,
King Fahd University of Petroleum and Minerals, Dhahran, Kingdom of Saudi Arabia
[1]shadi_abudalfa@hotmail.com
[2]moataz@kfupm.edu.sa

Abstract

The wealth of opinions expressed in micro-blogs, such as tweets, motivated researchers to develop techniques for automatic opinion detection. However, accuracies of such techniques are still limited. Moreover, current techniques focus on detecting sentiment polarity regardless of the topic (target) discussed. Detecting sentiment towards a specific target, referred to as target-dependent sentiment classification, has not received adequate researchers' attention. Literature review has shown that all target-dependent approaches use supervised learning techniques. Such techniques need a large number of labeled data. However, labeling data in social media is cumbersome and error prone. The research presented in this paper addresses this issue by employing semi-supervised learning techniques for target-dependent sentiment classification. Semi-supervised learning techniques make use of labeled as well as unlabeled data. In this paper, we present a new semi-supervised learning technique that uses less number of labeled micro-blogs than that used by supervised learning techniques. Experiment results have shown that the proposed technique provides comparable accuracy.

Keywords: Social Opinions, Sentiment Analysis, Target-Dependent, Polarity Classification, Semi-Supervised Learning.

1. Introduction

Social platforms include currently a huge repository of considerable opinions that are gathered from whole strata of society. These opinions play an important role in dominating and shaping decisions in sovereign fields such as politics, economics, sports, etc. Thus, automatic detection of opinion polarities expressed in micro-blogs (such as tweets) is a highly desirable service and its importance increased significantly with availability of numerous free social platforms [1].

There have been many techniques proposed in the literature for opinion mining especially in short text, known as micro-blogs. Most of such techniques are target-independent, that is they identify the sentiment expressed in a given micro-blog without taking into consideration the topic (target) being discussed in the micro-blog. Such techniques may misclassify some micro-blogs that include more than one target since they always assign same opinion polarity to the micro-blog regardless which target is considered. Recently, more accurate techniques have been proposed to detect a sentiment towards each target included in the micro-blog, referred to as target-dependent sentiment classification techniques [2].

Despite the various techniques proposed for target-dependent sentiment classification, their accuracies are still limited. Moreover, to the best of our knowledge, all of these techniques use supervised learning [3].Supervised learning needs a large set of labeled data. However, labeling micro-blogs is a time-consuming process that includes manual annotation, which tends to result in inaccurate sentiment classifications due to human errors and biased decisions. Using automated tools [4] for annotating micro-blogs may also affect the classification accuracy since the effectiveness of such tools are still limited. Another issue that may arise with supervised learning is the possibility of overfitting as well [5] which decreases accuracy of classifying unseen samples.

In this work, we address this gap by using semi-supervised learning for target-dependent sentiment classification. We experimented with various semi-supervised techniques and analyzed their performance. Subsequently, we propose a new technique that uses partially labeled data. We experimentally validated the new technique.

The rest of this paper is organized as follows.

Section 2 gives a background for target-dependent sentiment classification and semi-supervised learning techniques. Section 3 presents a review of the literature. Section 4 explains the proposed semi-supervised technique. Section 5 describes the experiment environment. Section 6 discusses the experiment results and provides due analysis. Section 7 presents some threats to the validity of our findings. Finally, Section 8 concludes the paper and discusses some suggestions for future work.

2. Background

Sentiment analysis, also known as opinion mining, comes under umbrella of natural language processing [6][7]. It is also one of the active research areas in text mining which has gained much attention nowadays. The main goal of sentiment analysis is identifying polarity of opinions [8]. Sentiment analysis includes numerous subtopics such as polarity classification [9], subjectivity detection [10], review summarization [11], and rumor detection [12][13]. Our research focuses specifically on polarity classification, also referred to as sentiment classification.

Target-independent sentiment classification fails to assign correct opinion to a micro-blog that includes more than one topic (target). Recently, some research works are proposed for dealing with target-dependent sentiment to increase accuracy of detecting opinion polarity expressed in micro-blogs. Systems of target-dependent sentiment classification detect opinion polarity expressed in micro-blog by focusing on the requested target which is included in the micro-blog. Thus, target-dependent strategy may detect different opinions for the same micro-blog by considering the interested target. Whereas, target-independent strategy assigns always same opinion polarity to the micro-blog regardless number of included targets. As a result of, target-independent strategy may lead to degrade classification accuracy.

For example, when we try to analyze a micro-blog “I prefer Samsung mobile more than iPhone” by using target-independent strategy, the detected opinion polarity will be “positive” since the micro-blog includes only positive phrases (“prefer” and “more than”). While, the detected opinion polarity will be more accurate when using target-dependent strategy by identifying “positive” outcome for “Samsung” target and “negative” output if the interested target is “iPhone”.

We can categorize machine learning techniques that are used with sentiment analysis into three classes: Supervised, unsupervised and semi-supervised. Supervised learning techniques use only labeled micro-blogs for training sentiment classifier [14] while unsupervised methods use only unlabeled micro-blogs such as lexicon-based techniques [15]. The third machine learning category [16], which includes semi-supervised learning techniques, uses both labeled and unlabeled micro-blogs for training sentiment analysis tools [17]. Semi-supervised learning techniques possess a desirable characteristic for building sentiment analysis system since they do not require annotating large numbers of micro-blogs for training.

Semi-supervised K-means (SK-means) is based on using labeled samples for selecting initial values of centroids [18] when applying K-means method. Label propagation model is an improved k-Nearest-Neighbor (kNN) method for finding closer unlabeled samples that are similar to labeled ones [19].Quasi-Newton semi-supervised support vector machines (QN-S3VM) is an extended model of support vector machines (SVM) for making SVM mimics a semi-supervised model by using both labeled and unlabeled samples during training phase instead of using only labeled ones [20].

Self-training, also called self-learning or self-labeling, is a technique used to learn from unlabeled data [16]. In this technique, a model is trained using a supervised learning technique by using labeled data points. Then, that model is used to detect sentiment polarities of the unlabeled data points. All unlabeled data points that generate high confidence predictions are added to the labeled dataset. After that, a model is trained again using the supervised learning technique using the larger labeled dataset. This process repeats for many rounds to hit the best accuracy.

3. Literature Review

Numerous approaches have been proposed in the literature for increasing the accuracy of target-independent sentiment classification. Dong et al. [21] integrated target information with recursive neural network to exploit the power of deep learning. Quan and Ren [22] proposed a similarity based approach to provide more fine grained sentiment analysis.

Vo and Zhang [23] generated word2vec features that are suitable for target-dependent sentiment classification. They reported efficiency of using these features by applying SVM with sentiment classification. Using the same dataset, they compared their work with previous works [2][21] [24]. They were able to show that their feature set was superior to those of the other works. More recently, Tang et al. [25] proposed target-dependent long short-term memory (LSTM) models where they reported accuracy improvement by 0.4% over the work of Vo and Zhang [23]; there were no improvement in terms of the macro-F1 score.

Gated neural networks (GNN) has been employed with targeted sentiment classification [26] for increasing classification accuracy. A comparison with a baseline technique [23] is presented by combining three datasets that include the dataset used by the baseline work. Recently, Wang et al. [27] use recursive neural networks (RNN) for classifying micro-blog streams based on target-dependent sentiment classification. To deal with the scarcity of labeled data, they used data clustering for partitioning unlabeled micro-blogs and selecting randomly some micro-blogs from the resulted groups. The selected micro-blogs are labeled manually for training RNN. A very recent work used bi-directional gated recurrent unit for improving accuracy of target-dependent sentiment classification [28] which increased the classification accuracy to 72.3% in comparison with previous related works.

For a comprehensive survey on the subject, the reader is advised to consult the recent work by Abudalfa and Ahmed [3]. The survey concludes that all of the approaches proposed for target-independent sentiment classification use supervised learning techniques. Additionally, the survey reveals that exploiting semi-supervised learning techniques in sentiment analysis has not received enough attention from researchers. More specifically, the survey notes that there is no reported research that uses semi-supervised learning techniques for target-dependent sentiment classification.

In this research, we address the scarcity of labeled data by applying a semi-supervised technique. The technique also resulted in better accuracy due to its ability to avoid overfitting.

4. Solution Approach

4.1 Overview

Our objective in this research is to investigate the suitability of employing semi-supervised learning techniques in target-dependent sentiment classification. We applied various semi-supervised learning techniques as we discuss in Section 6. We selected mainly semi-supervised K-means [29], label propagation models [30], Quasi-Newton semi-supervised support vector machines [31], and Self-Training [16]. We avoided applying semi-supervised techniques that increases time complexity sharply such as semi-supervised deep learning [32]. Results show that some techniques perform better than others. However, accuracies were not that satisfactory. Consequently, we adopted and adapted the self-training technique with linear logistic regression to improve the accuracy. Our proposed technique adds a new level to assess the confidence in the labeling done during the self-training. This extra level, namely SVME, was shown to improve the overall accuracy.

We used two methods for calculating confidence values. The first one, namely SelfTrP, is based on calculating probabilities resulted when predicting the unlabeled data point. While, the other method, namely SelfTrH, uses formula (Eq. 1) provided by S. Ravi [33]:

$$|d - \mu_y| < \sqrt{\tfrac{1}{1-\delta}}\sigma_y \qquad \text{(Eq. 1)}$$

where d is the distance from data point x to decision boundary hyperplane of M (data trained on labeled data), y is the label of x according to hyperplane of M, μ is mean of distances to hyperplane of M, σ is standard deviation of distances to hyperplane of M, and δ is a threshold.

We applied this formula to construct a decision boundary of three dimensions for fitting the classification problem which includes three classes (positive, negative, and neutral). This method adds the unlabeled sample to labeled training set and removes it from the unlabeled set.

4.2 Improved Self-training with Probabilities

Unlike some techniques proposed in the literature which are restricted to only deal with problems of binary classification, our proposed technique (ImproveSelfTrP) can handle multiclass sentiment classifications. For the sake of illustration and without loss of generality, we focus the discussion on classifying micro-blogs into three categories: positive, negative, and neutral. Fig. 1 illustrates a general framework of our proposed technique.

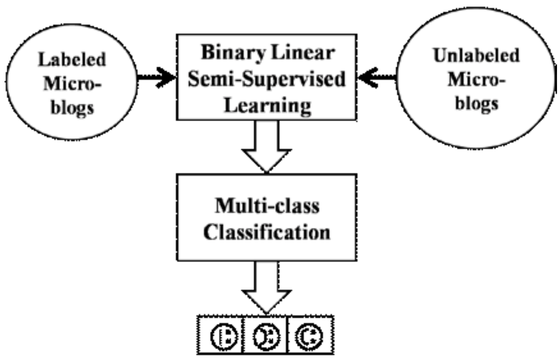

Fig. 1 General framework of proposed solution

Fig.2 depicts the flowchart of the model training mechanism. The input includes X_u, X_l, and L. X_u refers to the set of unlabeled data points. X_l refers to the set of labeled data points. L refers to the set of labels that are corresponding to X_l. Thereby, if there is a data point $v \in X_l$, then L_v is the corresponding label for *v*.

Our proposed technique uses two classifiers, namely, SVMT and SVME. SVMT denotes the classifier that is used for applying self-training semi-supervised learning. While, SVME denotes a classifier that is used for detecting misclassified data points. As shown in Fig. 2, SVMT is trained and used at each iteration to predict the originally labeled data points. Correctly labeled data points are marked as *true* classified. Mislabeled data points are marked as *false* classification. SVME is then trained by using these two classes: *true* and *false* classification.

When predicting the testing data (as shown in Fig. 3), we use both SVMT and SVME for making the final decision. If SVME classified the selected data point as *true* classified. Then, the final decision is selected based on the original prediction result provided by SVMT classier. Otherwise, we selected different sentimental label. For example, if the problem deals with three classes we select the sentimental label that provided confidence in the middle between the highest confidence and lowest confidence that are provided by SVMT classier when predicting the corresponding data point.

SVME is meant to improve classification accuracy by trying to predict misclassified data points before identifying sentiments expressed in micro-blogs. SVME is trained by collecting two data classes. The first class includes all labeled data points that are classified correctly by using SVMT model. While, the other class contains all labeled data points that are misclassified when applying SVMT model.

We can use any classifier when building SVMT and SVME models. However, we used in this work a linear logistic regression. The linear logistic regression requires an input parameter *C* (an inverse of regularization strength) which has an effect on its accuracy. However, any classifiers that would be used for SVMT and SVME would be sensitive to some own parameters.

Moreover, our proposed technique uses input parameter *P* for specifying confidence level of transferring predicted data point from unlabeled data into labeled set during each training round of self-training technique. We should initialize this parameter before running the system and selecting the best value that performs high classification accuracy.

Fig.3 shows how our proposed technique predicts opinion polarity expressed in each micro-blog. We use SVMT for detecting the three prediction probabilities (P_+, P_-, P_o) toward input micro-blog. P_+ value determines confidence probability of identifying the predicted data point as positive sentiment, while P_- and P_o refer to confidence probabilities of identifying negative and neutral opinions respectively.

Performance of the detection model is based on accuracy of SVME classifier when predicting whether a data point *m* is converged to misclassification or classified correctly. If SVME predicts *m* as correctly classified (*CLS* = true), then the sentiment outcome will belong normally to opinion polarity which gives maximum confidence probability ($\max(P_+,P_-,P_o)$).In the example shown in the figure the sentiment which gives the maximum confidence probability is positive. While, if *m* is predicted as misclassified (*CLS* = false) then our technique will select the sentiment outcome that corresponding to the second maximum ($\max2(P_+,P_-,P_o)$) confidence probability. In the example shown in the figure the sentiment which gives the 2nd maximum confidence probability is negative.

Mechanism of working the proposed technique is based on our expectation that the third value of confidence probabilities (the smallest confidence probability) leads usually to wrong sentiment outcome. Thus, the sentiment outcome will be belonged usually to the sentiment corresponding to the maximum confidence probability, if SVME model predicts the micro-blog as correctly classified. Otherwise we should change it to the other sentiment which is corresponding to the second maximum confidence probability.

5. Experiment Setup

We conducted many experiments to test the classification accuracy using semi-supervised learning. The development tools and hardware platform specifications are described in Table 1 and Table 2 respectively. We developed all experiments by using a dataset that is collected by Dong et al. [21] and has been utilized by many other researchers[1].

The dataset consists of 6248 tweets for training and 692 tweets for testing. The distribution of sentiment polarities of micro-blog (in both training and testing data) is 25% are positive tweets, 25% are negative tweets, and the rest 50% are neutral tweets. We use same word2vec embeddings2 that are designed as feature attributes by D. Vo and Y. Zhang [23] for developing all experiments. We also used LIBLINEAR[3] for scaling feature attributes.

[1]https://github.com/duytinvo/ijcai2015
[2]https://code.google.com/p/word2vec
[3]http://www.csie.ntu.edu.tw/~cjlin/liblinear/

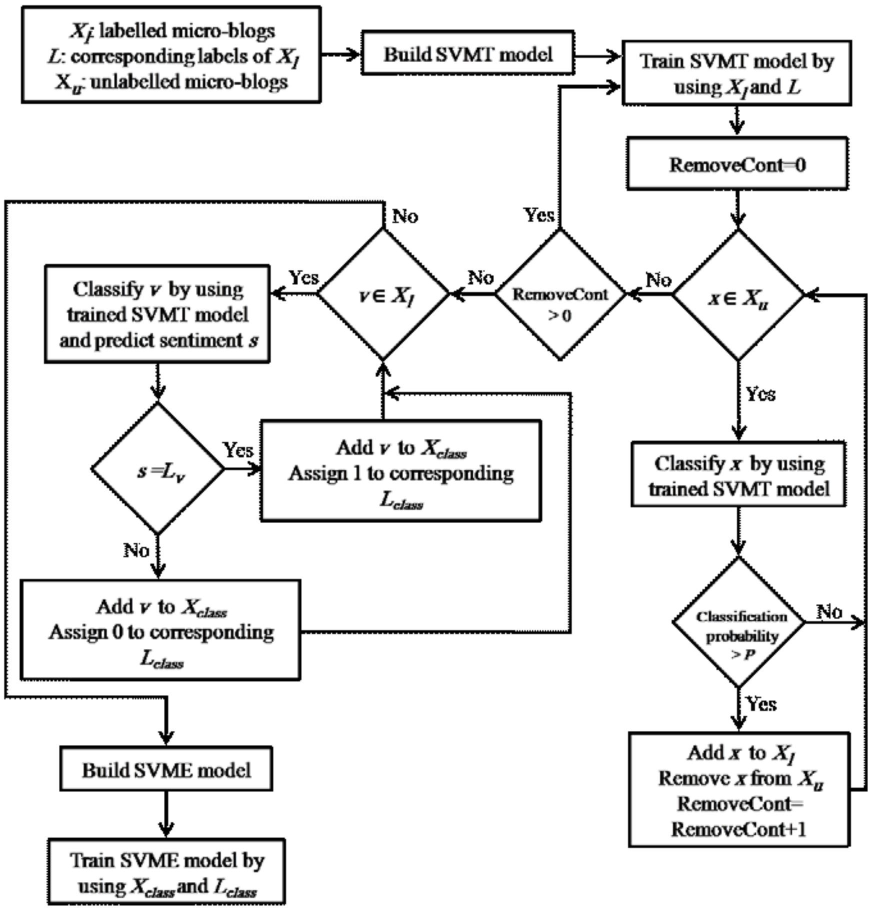

Fig. 2 Flowchart of training models

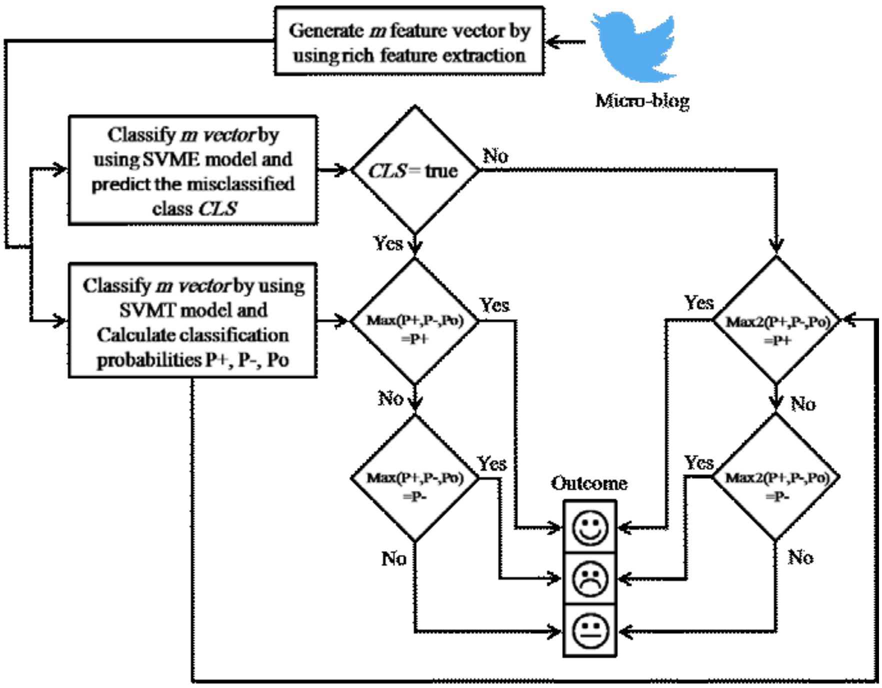

Fig. 3 Flowchart of detecting sentiment of micro-blog by using proposed technique

Table 1 Tools and programs

Tool	Version	Purpose
Python	2.7	Extracting Features, building and learning models for developing experiments, classifying micro-blogs, and computing results.
Anaconda	4.2.0	Open data science platform powered by Python for providing development environment that facilitates developing our experiments.
Spyder	2.3.8	Graphical platform for editing, testing and debugging Python codes.
LibLinear	2.1	Scaling and learning data for building SVM models.
QN-S3VM	2012	Building and learning semi-supervised SVM models.
MS Excel	2016	Analyzing data.
Vim	7.4	Text editor for editing huge training and testing data files.

Table 2: Platform specifications

Component	Specification
CPU	Intel(R) Core (TM) i7-3720 3.40 GHZ
Memory	8.00 GB
OS	Windows 8 (64-bit)

Based on our experiment work, we initialized C parameter to 0.009 when building all SVM models for providing high classification accuracy. We also fixed value of *P* parameter to 0.9 when generating all results that assess efficiency of our proposed technique.

We used the Quasi-Newton Optimization Framework[4] for applying the QN-S3VM experiments. That tool offered through this framework supports only binary classifiers. Thereby, we used multiclass strategies [34] to allow classifications into three classes.

When applying semi-supervised learning technique (SK-means), we used Cosine distance measure that performed better results based on our experiment work in comparison with other distance measures such as Euclidian.

To compare performance of the proposed technique with others, we used two measures: Classification accuracy and Macro-F1 score. Classification accuracy is the ratio of micro-blogs that are classified correctly to all ones [35]. The F1-score (also known as F-score or F-measure) is the average of precision and recall, and its best score is 1 while the worst score is 0. Precision is the ratio of micro-blogs that are correctly classified as positive to all micro-blogs classified as positive. Recall (which also known as true positive rate) is the ratio of micro-blogs that are classified correctly as positive to all positive micro-blogs. F1-score is used with binary classification problem which includes only two classes (positive and negative). Thus, we used macro-F1 score to calculate F1-score for multiclass classification (more than two classes) [36].

[4]http://www.fabiangieseke.de/index.php/code/qns3vm

6. Experiment Results

In this section, we discuss the performance of various semi-supervised learning techniques, along with ours, for target-dependent sentiment classification. It is worth noting that in order to statistically compare our proposed solution with other techniques, we repeated each run twelve times for each technique. We report the highest values achieved. We opted to report the highest values so that we can compare to others who only reported their highest results.

Table 3 describes all models that are evaluated in this work. All supervised learning models are reported in previous related work [25] except SSWE which is proposed by Tang et al. [24] and reported by Vo at al. [23] as comparable model. The rest of Table 3 describes all evaluated and improved semi-supervised learning techniques in comparison with our technique.

Table 4 shows results of our experiments in comparison with previous related works. The last part in the table illustrates accuracy and macro-F1 score of predicating sentiments by using semi-supervised learning models. We run out each experiment 12 times by selecting different parts of labeled data from the training set. Each experiment is executed also with different ratio of labeled data. The ratio is changed incrementally from 0.01 to 0.63 with increasing step equals 0.02. The reported results indicate to the highest accuracy and macro-F1 score that are achieved by training each model with the lowest ratio of labeled micro-blogs.

It is worth noting that semi-supervised learning techniques have been known in the literature to improve classification accuracy in comparison with supervised learning techniques; however, they use same number of labeled samples plus extra unlabeled ones. In this work, we evaluate efficacy of our proposed semi-supervised learning technique by using less number of labeled micro-blogs than that used with supervised learning techniques.

The experiment results show that SK-means technique provides the worst results in comparison with other semi-supervised learning models. This observation discloses that the used dataset is complex and the three classes (positive, negative, and neutral) are not well separated. Thus, we can conclude that applying data clustering for classifying the used dataset does not provide competitive results. The experiment results show also that using Cosine distance when applying SK-means technique to the used dataset achieves the best results in comparison with using other distance measures such as Euclidean distance.

All experiments which are conducted for applying the label propagation technique illustrate that this technique can sharply decrease the ratio of the labeled data that is needed to apply semi-supervised learning. In spite of its limited classification accuracy, this technique outperforms SK-means technique. Additionally, we noticed that using radial basis function (RBF) kernel with label propagation technique increases classification accuracy in comparison with using kNN kernel.

As shown in Table 4, our proposed technique (ImproveSelfTrP) provides comparative accuracy in comparison with other evaluated supervised leaning models while uses partially labeled data. It outperforms all evaluated semi-supervised learning models. It is clear that our proposed solution provides the highest classification accuracy which is achieved also by using a very recent work employed a deep learning model (Bi-GRU). It is interesting also to clarify that the proposed semi-supervised learning technique archived the highest classification accuracy by using only 45% of labeled data. Whereas, learning other supervised learning models with this ratio results in lower classification accuracy. The proposed technique does not provide the best macro-F1 score. However, it outperforms (in terms of both accuracy and macor-F1 score) a recent deep learning model TC-LSTM (macro-F1 score equals 69.5%) which provided an increase in classification accuracy equals 71.5%.

Our explication for not achieving the best macro-F1 score when applying our proposed model tends to that the used dataset has large number of neutral tweets. The proposed technique is designed to predict neutral tweets more accurately. It selects the second maximum of prediction probabilities which works better when classifying initially neutral tweets incorrectly. Thus, our technique will correct the outcome to neutral sentiment which is related usually to the second maximum of prediction probabilities (P_+, P_-, and P_o). While, positive and negative tweets may be misclassified when they lead alternately to lowest or largest prediction. As a result of this, classification accuracy is increased while the macro-F1 score may not match this improvement.

We conducted an experiment to evaluate the impact of the ration of the labeled data on the accuracy. Fig. 4 illustrates results of applying ImproveSelfTrP when using different ratios of labeled data along with the corresponding confidence interval among the 12 runs. The figure shows that the mean of classification accuracies is improved gradually when increasing ratio of labeled data. While, the accuracy did not improve significantly when increasing the ratio more than 29%.The results converge also to high confidence since each confidence interval is so small.

It is worth also mentioning that the semi-supervised learning model S3VMOVR outperforms all supervised learning models except Bi-GRU in terms of both classification accuracy and macro-F1 score. S3VMOVR outperforms the supervised learning model Target-dep+ which uses same feature attributes and provides the second best macro-F1 score (equals to 69.9%). It also outperforms the supervised learning model TC-LSTM which provided the second best classification accuracy (equals to 71.5%) over all supervised learning techniques that are used for target-dependent sentiment classification.

On the other hand, S3VMOVR model is not a robust model since its accuracy is sensitive to initializing a random parameter that is used with Quasi-Newton optimization method. While, our proposed technique does not use any random parameters and it is more robust when using same ratio of labeled and unlabeled micro-blogs. Moreover, when using QN-S3VM models we need to initialize two additional parameters (λ and λ') which make using QN-S3VM model more difficult when finding the optimum values of all these parameters.

7. Threats to Validity

In our experiments, we used a very popular dataset used in the literature. The dataset contained 25% positive, 25% negative, and 50% neutral sentiments. It is not conclusive that applying our techniques to other datasets with different distributions would result in the same classification accuracy. We could not investigate this matter further because there was no other relevant public datasets available.

Moreover, our experiments revealed that results are sensitive to initial values that are used when setting the parameters. Additionally, the ratio of labeled data and the selected labeled samples affect significantly on the classification accuracy.

Table 3 Description of all compared methods

Method	Description	Class
SSWE	Sentiment-specific word embedding model [24].	S
SVM-indep	SVM classifier uses only target-independent features.	S
SVM-dep	SVM classifier uses target-independent features concatenated with target-dependent features [2].	S
RecursiveNN	Standard recursive neural network with target-dependent dependency tree [21].	S
AdaRNN-w/oE	Adaptive recursive neural network (RNN) [21].	S
AdaRNN-w/E	Adaptive recursive neural network (RNN) [21].	S
AdaRNN-comb	Adaptive recursive neural network (RNN) [21].	S
Target-dep	SVM classifier uses rich target-independent and target-dependent features [23].	S
Target-dep+	SVM classifier uses rich target-independent, target-dependent, and sentiment lexicon features [23].	S
LSTM	Long short-term memory model (recurrent neural network) uses Glove vector. It classifies target-dependent sentiment based on target independent strategy [25].	S
TD-LSTM	Target-Dependent LSTM [25].	S
TC-LSTM	Target-Connection LSTM [25].	S
Bi-GRU	Bi-directional gated recurrent unit for target-dependent sentiment classification [28].	S
SK-means	Semi-supervised K-means algorithm with Cosine distance.	SM
LabelProK	Label propagation model by using kNN kernel.	SM
LabelProR	Label propagation model by using RBF kernel.	SM
LabelSpK	Label spreading model by using kNN kernel.	SM
LabelSpR	Label spreading model by RBF kernel.	SM
S3VMOvOVote	QN-S3VM with OVO strategy. The voting strategy is used to select the most dominant perdition.	SM
S3VMOvR	QN-S3VM with OVR strategy.	SM
SelfTrH	Self-training with SVM technique that uses distance from the hyperplane for calculation confidence. The used formula inspired from research work [33].	SM
SelfTrP	Self-training with SVM technique that uses prediction probability for calculating prediction confidence.	SM
ImproveSelfTrP	Our proposed technique	SM

Class: **S**=Supervised learning technique, **SM**= Semi-supervised learning technique.

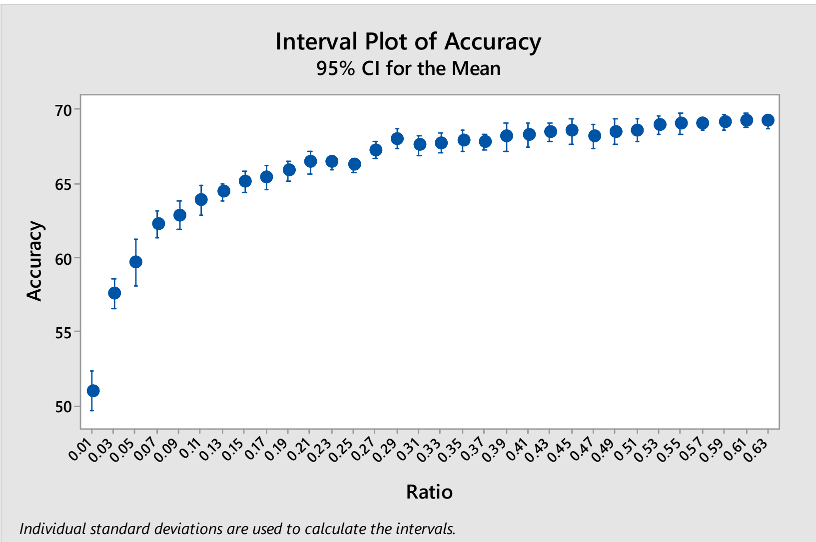

Fig. 4 Effect of changing ratio of labeled data

Table 4 Comparing different techniques for target-dependent sentiment classification

Method, year	Setting	Accuracy	Macro-F1	Ratio
SSWE, 2014		62.4	60.5	100%
SVM-indep, 2011		62.7	60.2	100%
SVM-dep, 2011		63.4	63.3	100%
RecursiveNN, 2014		63.0	62.8	100%
AdaRNN-w/oE, 2014		64.9	64.4	100%
AdaRNN-w/E, 2014		65.8	65.5	100%
AdaRNN-comb, 2014		66.3	65.9	100%
Target-dep, 2015		69.7	68.0	100%
Target-dep+, 2015		71.1	69.9	100%
LSTM, 2016		66.5	64.7	100%
TD-LSTM, 2016		70.8	69.0	100%
TC-LSTM, 2016		71.5	69.5	100%
Bi-GRU, 2018		72.3	70.5	100%
SK-means	Cosine distance measurement	46.8	43.0	37%
LabelProK	kNN kernel, neighbours #=1	56.4	53.6	1%
LabelProR	RBF kernel, gamma =0.07	60.8	55.4	7%
LabelSpK	kNN kernel, neighbours #=7	59.8	53.6	27%
LabelSpR	RBF kernel, gamma= 0.19	61.4	56.6	5%
S3VMOVOvote	linear kernel, lamda=0.045	70.5	68.4	61%
S3VMOVR	linear kernel, lamda=0.025	71.7	70.0	63%
SelfTrH	C=0.009, Threshold=0.81	70.8	67.9	59%
SelfTrP	C=0.009, Prob Threshold=0.9	72.1	69.5	45%
ImproveSelfTrP♦	C=0.009, P=0.9	72.3	69.7	45%

♦ Proposed solution

8. Conclusion and Future Work

In this work, we address the problem of accuracy limitation with target-dependent sentiment classification as well as the need to huge labeled data for training current techniques. We evaluated different semi-supervised learning techniques and conducted many experiments to compare their performance. As an outcome of the comparison, we were able to propose a new semi-supervised learning technique for decreasing number of labeled micro-blogs that are needed for training target-dependent sentiment classification model.

This work can be extended in different directions. It is worth investigating optimization techniques such as genetic algorithms for finding the global optimum values of parameters that are used for building semi-supervised learning models. This work can be extended also by testing the performance of using other semi-supervised learning techniques. It would be also interesting to develop methods for determining the minimal required number of labeled micro-blogs that can be used for achieving the best accuracy. Such micro-blogs should form a representative sample adequate enough to classify the overall input data. Moreover, future work could investigate developing cluster-based technique for partitioning input micro-blogs and selecting specific samples with high confidence for providing high classification accuracy.

Additionally, another research direction may improve both classification accuracy and macro-F1 score by combining more than one semi-supervised learning techniques such as merging our technique (ImproveSelfTrP) with S3VMOVR. In the same manner, extending label propagation technique may improve classification accuracy by using lower ratio of labeled data.

Acknowledgements

The authors wish to acknowledge King Fahd University of Petroleum and Minerals (KFUPM) for providing the facilities to carry out this research. The authors would like also to thank Dr. Yue Zhang and Mr. Duy-Tin Vo for answering questions related to their public dataset and code implementation that have been used in this work. Moreover, the authors thank Mr. Fabian Gieseke for answering some enquiries about his developed tool which has been used also in this research work.

Competing interests

The authors have declared that no competing interests exist.

References

[1] F. Pozzi, E. Fersini, E. Messina, and B. Liu, *Sentiment Analysis in Social Networks*, 1st ed, Elsevier Inc., 2017.

[2] L. Jiang, M. Yu, M. Zhou, X. Liu, and T. Zhao, “Target-dependent Twitter Sentiment Classification,” *Proc. the 49th Annual Meeting of the Association for Computational Linguistics*, pp. 151–160, Portland, Oregon, June 19-24, 2011.

[3] S. Abudalfa, and M. Ahmed, “Survey on Target Dependent Sentiment Analysis of Micro-Blogs in Social Media,” the 9th IEEE GCC Conference & Exhibition, Manama, Bahrain, 7-10 May 2017.

[4] K. Veselovska, J. Hajic, Jr. and J. Sindlerova, "Creating Annotated Resources for Polarity Classification in Czech,” *Proc. KONVENS 2012 (PATHOS 2012 workshop)*, Vienna, September 21, 2012.

[5] I. Tetko, D. Livingstone, A. Luik, "Neural network studies. 1. Comparison of Overfitting and Overtraining," *J. Chem. Inf. Comput. Sci.*, vol. 35, no. 5, pp. 826–833, 1995.

[6] B. Pang, and L. Lee , “Opinion mining and sentiment analysis,” *Foundations and Trends in Information Retrieval*, vol. 2, no. 1-2, pp. 1-135, 2008.

[7] A. Shoufan, and S. Al-Ameri “Natural Language Processing for Dialectical Arabic: A Survey,” *in Second Workshop on Arabic Natural Language Processing*, July 26-31, 2015, pp. 36-48.

[8] M. Tsytsarau, and T. Palpanas, “Survey on mining subjective data on the web,” *Data Mining and Knowledge Discovery*, vol. 24 no. 3, pp. 478- 514, 2012.

[9] B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up? Sentiment classification using machine learning techniques,” *in the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2002, pp. 79-86.

[10] J. Wiebe, T. Wilson, and M. Bell, “Identifying collocations for recognizing opinions,” *in ACL’01 Association for Computational Linguistics Workshop on Collocation: Computational Extraction*, Analysis, and Exploitation, 2001, pp. 24-31.

[11] G. Somprasertsri, and P. Lalitrojwong, “Mining Feature- Opinion in Online Custumer Reviews for Opinion Summarization,” *Journal of Universal Computer Science*, vol. 16, pp. 938-955, 2010.

[12] A. Zubiaga, M. Liakata, R. Procter, K. Bontcheva, and P. Tolmie, “Towards Detecting Rumours in Social Media,” *in AAAI Workshop on AI for Cities*, 2015, pp. 380-390.

[13] K. Wu, S. Yang, and Kenny Q. Zhu, “False Rumors Detection on Sina Weibo by Propagation Structures,” *in the 31st International Conference on Data Engineering*, 2015, pp. 651-662.

[14] R. Groot, “Data Mining for Tweet Sentiment Classification,” Master thesis, Utrecht University, July 23, 2012.

[15] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas, “Sentiment strength detection in short informal text,” *Journal of the American Society for Information Science and Technology*, vol. 61, no. 12, pp. 2544-2558, 2010.

[16] R. Mihalcea, “Co-training and Self-training for word sense disambiguation,” Proc. of the Eighth Conference on Computational Natural Language Learning, 2004.

[17] M. Yang, W. Tu, Z. Lu, W. Yin, and K.-P. Chow, “LCCT: A Semi-supervised Model for Sentiment Classification,” *Human Language Technologies: The 2015 Annual Conference of the North American Chapter of the ACL*, pp. 546–555, Denver, Colorado, May 31 – June 5, 2015.

[18] X. Wang, C. Wang, and J. Shen, “Semi–supervised K-Means Clustering by Optimizing Initial Cluster Centers,” *Web Information Systems and Mining*, vol. 6988 of the series Lecture Notes in Computer Science pp. 178-187, 2011.

[19] D. Zhou, O. Bousquet, T. Lal, J. Weston, and B. Schölkopf, “Learning with local and global consistency,” *Advances in Neural Information Processing Systems 16*, 2004.

[20] F. Gieseke, A. Airola, T. Pahikkala, and O. Kramer, “Sparse Quasi-Newton Optimization for Semi-Supervised Support Vector Machines,” *Technical Report*, 2012.

[21] L. Dong, F. Wei, C. Tan, D. Tang, M. Zhou, and K. Xu, “Adaptive Recursive Neural Network for Target-dependent Twitter Sentiment Classification,” *Proc. of the 52nd Annual Meeting of the Association for Computational Linguistics (Short Papers)*, pages 49–54, Baltimore, Maryland, USA, June 23-25 2014.

[22] C. Quan, and F. Ren, “Target Based Review Classification for Fine-Grained Sentiment Analysis,” *International Journal of Innovative Computing*, Information and Control, vol. 10, no. 1, February 2014.

[23] D. Vo and Y. Zhang, “Target-Dependent Twitter Sentiment Classification with Rich Automatic Features,” *the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI)*, 2015.

[24] D. Tang, F. Wei, N. Yang, M. Zhou, T. Liu, and B. Qin, “Learning sentiment-specific word embedding for Twitter sentiment classification,” *Proc. the 52nd Annual Meeting of the Association for Computational Linguistics*, pp. 1555–1565, 2014.

[25] D. Tang, B. Qin, X. Feng, and T. Liu, “Effective LSTMs for Target-Dependent

Sentiment Classification," *COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pp. 3298–3307, Osaka, Japan, December 11-17 2016.
[26] M. Zhang, Y. Zhang, and D. Vo, "Gated Neural Networks for Targeted Sentiment Analysis," *the 30th AAAI Conference on Artificial Intelligence*, Phoenix, Arizona, USA. Association for the Advancement of Artificial Intelligence, 2016.
[27] Y. Wang, H. Gao, and S. Geng "A Target-Dependent Sentiment Analysis Method for Micro-blog Streams," *18th Asia-Pacific Web Conference, Web Technologies and Applications, APWeb 2016, Part{II}*, pp. 30-42, Suzhou, China, September 23-25, 2016.
[28] M. Jabreel and A. Moreno, "Target-dependent Sentiment Analysis of Tweets using a Bi-directional Gated Recurrent Unit," *Smart Innovation, Systems and Technologies, SIST*, vol. 85, 2018.
[29] S. Basu, A. Banerjee, and R. Mooney, "Semi-supervised Clustering by Seeding", *Proc. of the 19th International Conference on Machine Learning (ICML-2002)*, pp. 19-26, Sydney, Australia, July 2002.
[30] X. Zhu, and Z. Ghahramani, "Learning from labeled and unlabeled data with label propagation," *Technical Report CMU-CALD-02-107*, Carnegie Mellon University, 2002.
[31] F. Gieseke, A. Airolab, T. Pahikkalab, and O. Kramera, "Fast and Simple Gradient-Based Optimization for Semi-Supervised Support Vector Machines," 2013.
[32] Y. Chen, X. Zhu, and S. Gong, "Semi-supervised deep learning with memory," in The European Conference on Computer Vision (ECCV), September 2018.
[33] S. Ravi, "Semi-supervised Learning in Support Vector Machines," *Project Report*, Princeton University, 2014.
[34] M. Aly, "Survey on Multiclass Classification Methods," *Technical Report*, California Institute of Technology, November 2005.
[35] C. Metz, "Basic principles of ROC analysis" (PDF). *Semin Nucl Med.*, vol. 8, no. 4, pp. 283–98, October 1978.
[36] S. Parambath, N. Usunier, and Y. Grandvalet, "Optimizing F-measures by cost-sensitive classification," *Proc. Neural Information Processing Systems (NIPS)*, pp. 2123–2131, 2014.

Citation: S.I. Abudalfa and M.A. Ahmed. *"Semi-Supervised Target-Dependent Sentiment Classification for Micro-Blogs"*, Journal of Computer Science & Technology, vol. 19, no. 1, pp. 55–65, 2019.

DOI: 10.24215/16666038.19.e06

Received: August 21, 2018 **Accepted:** March 14, 2019