International Journal of Approximate Reasoning 105 (2019) 85–97

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

A cost-sensitive three-way combination technique for ensemble learning in sentiment classification ☆

Yuebing Zhang [a,b], Duoqian Miao [a,b], Jiaqi Wang [a,b], Zhifei Zhang [a,c,*]

[a] *Department of Computer Science and Technology, Tongji University, Shanghai 201804, China*
[b] *Key Laboratory of Embedded System and Service Computing, Ministry of Education, Shanghai 201804, China*
[c] *State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China*

ARTICLE INFO

Article history:
Received 14 March 2018
Received in revised form 12 October 2018
Accepted 31 October 2018
Available online 6 November 2018

Keywords:
Three-way decisions
Ensemble learning
Sentiment classification

ABSTRACT

Deep neural networks (DNN) have achieved remarkable results in sentiment classification. Some ensemble methods of DNN models and traditional feature-based models are proposed recently. However, to the best of our knowledge, most of the works use traditional ensemble combination techniques, e.g. voting and stacking, which are designed for weak base classifiers. So far many base classifiers, e.g. DNN, have been able to achieve good results in sentiment classification tasks, so there should be a new ensemble combination technique designed for strong base classifiers. To address this issue, we proposed a cost-sensitive combination technique using sequential three-way decisions (3WD), which is named S3WC. In S3WC, base classifiers are arranged in a linear arrangement, and a gate mechanism is constructed in each step to divide the objects into three groups, i.e., positive region, negative region and boundary region, which respectively correspond to acceptance, rejection and deferment in sequential 3WD. Each object is grouped by minimizing its total cost consisting of misclassification cost and time cost. The objects in boundary region require more information to decrease the misclassification cost, so they are reclassified by the subsequent base classifiers in order to obtain more information, while the time cost increases. In the experiment, we apply S3WC to DNN models and traditional feature-based models on five benchmark datasets, and compare its performance with traditional ensemble combination techniques. The experimental results show that S3WC outperforms any of its base classifiers in terms of classification accuracy, and the total cost of S3WC is lower than that of the existing ensemble combination techniques (e.g. majority-voting, weighted-voting, meta-learning).

1. Introduction

Text sentiment analysis (a. k. a. opinion mining) that analyzes people's opinions and emotions from text is an active research field in natural language processing (NLP) [1]. The existing studies of sentiment classification, which is an important part of sentiment analysis, can be mainly grouped into two categories: lexicon-based and corpus-based approaches [2]. Lexicon-based approaches typically use a dictionary of sentiment words and phrases, and incorporate intensification and negation to compute a sentiment score for each text [3]. Corpus-based approaches treat sentiment classification as a special

☆ This paper is part of the Virtual special issue on Uncertainty in Granular Computing, Edited by Duoqian Miao and Yiyu Yao.

* Corresponding author at: Department of Computer Science and Technology, Tongji University, Shanghai 201804, China.
E-mail addresses: yuebing_zhang@hotmail.com (Y. Zhang), zhifeizhang@tongji.edu.cn (Z. Zhang).

https://doi.org/10.1016/j.ijar.2018.10.019
0888-613X/© 2018 Elsevier Inc. All rights reserved.

case of text categorization problem [4], which utilize machine learning methods to extract reasonable features from texts and feed into a classifier to predict the sentiment [5].

The existing studies of corpus-based approaches are dominated by two main directions: traditional feature-based methods and deep learning methods. Traditional feature-based methods extract manually designed features from the text, e.g., N-gram (unigrams, bigrams, trigrams), Part-of-Speech (POS), term frequency-inverse document frequency (TF-IDF), and then use the features and a classification model to complete the sentiment classification task. The common classification models include naive bayes (NB), maximum entropy (ME), support vector machine (SVM) [4], SVM with NB features (NB-SVM) [6], etc. Deep learning models have achieved remarkable results in computer vision [7] and speech recognition [8] in recent years. Within NLP, many works with deep learning models have focused on learning word embeddings (a. k. a. word vectors) with neural language models [9] and performing composition over these word embeddings for classification [10]. Word embeddings have lower dimension than bag of words (BoW), and semantic information of words is encoded into such a dense representation. Moreover, phrase vectors and document vectors are presented [11] to find a good representation for each phrase or document, respectively. Convolutional neural networks (CNN) are popular deep learning models which utilize layers with convolving filters that are applied to local features [12]. CNN models have been shown to be effective for NLP and have achieved excellent results in semantic parsing [13], search query retrieval [14], sentence modeling [15], and other traditional NLP tasks [10]. Kim [16] trained a simple CNN with one layer of convolution on top of word vectors obtained from an unsupervised neural language model for sentence-level classification, and achieved excellent results on multiple benchmarks.

Some ensemble methods are proposed for sentiment classification recently [17,18], but they all use traditional combination techniques, e.g. voting and stacking, which are designed for weak base classifiers. In this paper we use sequential 3WD to construct a cost-sensitive combination technique for strong base classifiers. The methodology of three-way decisions (3WD) [19] is widely applied in many theoretic fields, such as management sciences [20], social judgement theory, fuzzy sets theory [21], hypothesis testing in statistics, attribute reduction [22,23] and knowledge granulation [24,25]. 3WD are also widely used in numerous application fields, including medical decision making [26], peering review process, government decision [27], Email spam filtering [28], face recognition [29], clustering analysis [30] and classification [31,32]. To give a formal description of 3WD, Yao [33] presented a general overview on existing 3WD researches, and extended the rough sets-based 3WD decisions to a much wider frontier, which outlines a unified theory of 3WD. Furthermore, Yao [34] extended 3WD to sequential 3WD, in which the cost of obtaining required information is considered.

The remainder of this paper is organized as follows. In Section 2, we introduce some related works on sentiment classification. In Section 3, we present the structure and process of S3WC. In Section 4 we report the experimental results and analysis. Finally, we make a conclusion in Section 5.

2. Related work

2.1. Sentiment classification

Sentiment classification is a fundamental and important study area in sentiment analysis. It hammers at detecting the sentiment polarity of a sentence [35] (or a document [36]) based on its textual content. Taking a panoramic view of this area, there are two main directions for sentiment classification: lexicon-based approaches and corpus-based approaches. Lexicon-based approaches typically use sentiment dictionary, intensification and negation to compute a sentiment score for each text. Sentiment words and phrases are marked with sentiment polar and sentiment strength in sentiment dictionary. There are two kinds of sentiment dictionaries according to universality. One is a universal sentiment dictionary which is applicable to almost all fields; the other is a domain sentiment dictionary which is applicable to specific fields. Turney [37] proposed a simple but representative lexicon-based method to classify reviews into recommended or not recommended. The classification of a review is predicted by the average semantic orientation of its phrases, and the semantic orientation of a phrase is calculated by the mutual information between the given phrase and the word “excellent” minus the mutual information between the given phrase and the word “poor”. Ding et al. [38] applied negation words (e.g., not, never, cannot) and contrary words (e.g., but) to improve the performance of lexicon-based method. Thelwall et al. [39] developed *SentiStrength* with sentiment lexicon and linguistic rules for detecting sentiment strength of tweets. In lexicon-based sentiment classification, the problem of contextual polarity is a major cause for classification error. Cho et al. [40] presented a data-driven method of adapting sentiment lexicon to diverse domains. In essence, the method compared the positive/negative review's dictionary word occurrence ratio with the positive/negative review ratio itself to determine which sentiment words to be removed and which sentiment words' polarity to be switched.

As for corpus-based approaches, Pang et al. [4] pioneered to treat sentiment classification as a special case of text categorization and applied three machine learning methods (NB, ME and SVM), and SVM with bag-of-word features achieved the best performance in their experiment. Many studies are inspired by this idea and focus on designing effective features to obtain a better performance on sentiment classification. Wang et al. [6] presented NB-SVM for sentiment classification on movie and product reviews, which is a SVM variant using NB log-count ratios as feature values. Paltoglou et al. [41] and Kim et al. [42] studied the feature weights by investigating variants weighting functions from information retrieval. Katz et al. [43] presented *ConSent*, a novel context-based approach which was effective both in noiseless and noisy text. Salvador et al. [44] used meta-learning to combine and enrich several baseline methods (bag of words, n-grams, lexical resource-based classifier) aiming at cross-domain polarity classification. Mohammad et al. [45] combined active learning and self-training to handle cross-lingual sentiment classification.

2.2. Deep learning approaches for sentiment classification

In recent years, deep learning models have been widely used in NLP. In the beginning, neural network models were used to learn distributed vector representations of word, paragraph [9], and document [11]. Tang et al. [46] proposed sentiment-specific word embeddings for sentiment classification. Meanwhile, some word embeddings for specific tasks like Twitter sentiment classification were proposed [47]. Neural networks were further applied to sentiment classification. Kim [16] trained a simple CNN for sentence classification, and Kalchbrenner [15] subsequently applied dynamic k-max pooling over time to generalize the original max pooling in traditional CNN. Recurrent neural networks (RNN) [48] and some extensions, such as bidirectional recurrent neural networks (BRNN) [49] and gates recurrent neural networks (GRNN) [50], were applied to sentiment classification. In addition to the commonly used neural networks in computer vision, Zhao et al. [51] proposed a self-adaptive hierarchical sentence model which is dedicated to text classification.

2.3. Ensemble methods for sentiment classification

Ensemble methods combine a set of base classifiers in order to achieve a better performance in comparison with a single base classifier. Rokach [52] provide an overview of ensemble methods in classification tasks and think the classification using ensemble methods is based on two dimensions: how predictions are combined (rule-based and meta-learning), and how learning process is done (concurrent and sequential).

Regarding the first dimension, rule-based approaches treat the predictions from base classifiers by a rule (e.g. majority voting, weighted combination, etc.), meta-learning techniques use the predictions from base classifiers as features for a meta-learning model. As explained in Xia et al. [53], weighted combinations of feature sets is effective in sentiment classification tasks, since the weights of ensemble represent the relevance of different feature sets (e.g. n-grams, POS, etc.) instead of assigning relevance to each features individually. Fersini et al. [54] proposed several variants of voting rules and validated their effectiveness in a variety of datasets. In a different work, Fersini et al. [55] compared the majority voting rule with other approaches, and reported that average rule can ensure a better performance than other considered rules. A meta-classifier ensemble model was evaluated by Xia et al. [56] which obtained performance improvements. Aue and Gamon [57] proposed an adaptive meta-learning model which offers a relatively low adaptation effort to new domain. Besides, both rule based and meta learning ensemble models can be enriched with extra knowledge as illustrated in Xia and Zong [56]. Araque et al. [17] used both feature ensemble and classifier ensemble for sentiment analysis, in which feature ensemble includes word embedding from deep learning models and surface features (e.g. n-grams, POS, etc.), classifier ensemble includes rule-based technique and meta-learning technique.

As for the second dimension, concurrent models divide the original dataset into several subsets and train multiple classifiers on the subsets respectively in a parallel fashion. The most popular concurrent technique is bagging. On the contrary, sequential models do not divide the dataset but there is an iteration process in training steps, which taking advantage from previous iterations of training steps to improve the performance of global classifier. Boosting is a very common sequential technique. Sehgal et al. [58] used bagging and other classification algorithms in sentiment analysis tasks and reported that the sentiment evolution and the stock value trend are closely related. Fersini et al. [54] also used bagging techniques in sentiment analysis tasks and showed several experimental results including associated model complexity. Some researchers [59, 60] have reported that bagging techniques are robust to noisy data, while boosting techniques are quite sensitive. Wang et al. [18] studied the suitability of bagging and boosting techniques. Moreover, a different ensemble technique, random subspace, which divides training dataset in feature space instead of instance space, was studied and compared with bagging and boosting in this work.

2.4. Three-way decisions

Most of the existing ensemble methods for sentiment classification use traditional combination techniques, e.g. voting and stacking (a. k. a. meta-learning), which are proposed for weak base classifiers. With the development of machine learning methods for sentiment classification, the base classifiers used in ensemble method can achieve high classification accuracy (generally higher than 70%). It is inappropriate to use traditional combination techniques to combine these strong base classifiers. In this paper, we use sequential three-way decisions to construct a cost-sensitive ensemble combination technique for strong base classifiers. Three-way decisions, which give the inspiration for this paper, are a new interpretation of rules in rough set theory [19]. The applications of three-way decisions are widely concerned in many fields. Yu et al. proposed a three-way clustering method for tree-based incremental overlapping clustering [30]. Zhou et al. discussed the applications of three-way decisions in Email spam filtering [28]. Liu et al. applied three-way decisions into investment decision [61] and government decision [27]. Yang and Yao investigated the applications of three-way decisions in multi-agent decision [62]. Li et al. introduced three-way decisions to face recognition and proposed a cost-sensitive sequential three-way decisions method for face recognition [29]. Li et al. proposed a DNN-based sequential granular feature extraction method and a cost-sensitive sequential three-way decisions strategy [63]. Zhang and Min combined three-way decisions

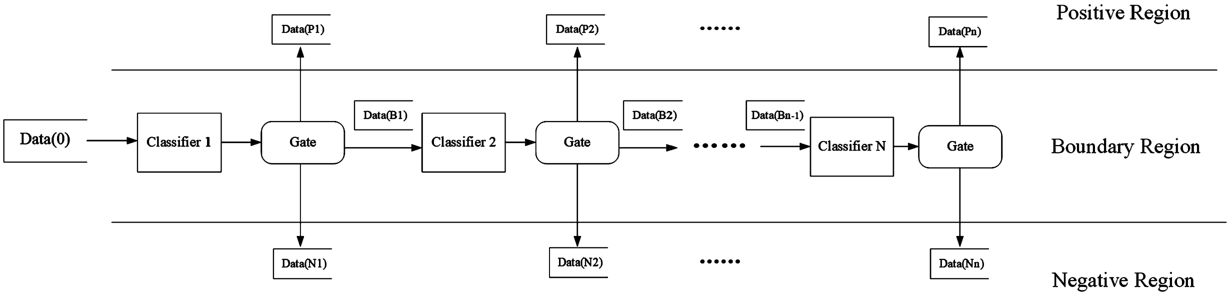

Fig. 1. The structure of 3w-combination technique.

with random forest, and applied three-way decisions into recommender systems [64]. Several 3WD-based ensemble methods were proposed for clustering [65] and classification [66], but they just combined 3WD and ensemble learning to address clustering or classification tasks, instead of using 3WD to design a new ensemble strategy. Li et al. [66] proposed a 3WD-based two-phase classification model, which seems similar with our work. Actually, the proposed work in this paper is a 3WD-based ensemble combination technique, while Li et al. applied a simple ensemble learning approach to implement the second phase of the classification process. Furthermore, as for the model architecture, our method is N-phase while it is two-phase; as for cost matrix, our method is data-based while it is empirical; as for metrics, our method considers the total cost including misclassification cost and time cost, while it only considers misclassification cost (classification accuracy). The above works enrich the theoretic foundation of three-way decisions, and indicate that three-way decisions are applicable for many practical decision problems.

3. The proposed method

3.1. Structure

In this section, we present S3WC for the ensemble methods in sentiment classification tasks. S3WC is specific to strong base classifiers and its architecture is illustrated in Fig. 1. In sequential 3WD theory, decision results in each step are divided into three parts, namely, acceptance, rejection and deferment. This can be simply mapped to binary sentiment classification, which corresponds to positive, negative and boundary. The objects in boundary region require more information to be classified into positive and negative categories, which guarantees a low misclassification cost. In this work we construct a component named “gate” to divide the objects into three regions by minimizing the overall cost consisting of misclassification cost and time cost in each step. Afterwards, the next base classifier is used to reclassify the objects in boundary region. After all the objects have obtained their determinate decisions (acceptance or rejection), the classification process is terminated.

3.2. Gate mechanism

Gate is the core of S3WC, and it determines which objects are divided into positive and negative regions and which objects flow into the next decision step. Gate can also be treated as a map mechanism between a object and its most suitable base classifier. In terms of classification accuracy, a gate is effective if it satisfies:

$$Acc_i(Data(i-1)) < Acc_i(Data(P_i) + Data(N_i)) \tag{1}$$

where $Acc_i(Data(i-1))$ represents the prediction accuracy of $classifier_i$ on $Data(i-1)$.

We propose two strategies for the gate mechanism: mincost-gate and proportion-gate.

3.2.1. Mincost-gate

In mincost-gate, the objects are divide into three regions by minimizing the total cost consisting of misclassification cost and time cost. The misclassification cost in the i-th gate is calculated by the loss matrix [33] shown in Table 1. Let e_i denote the classification error rate of the i-th base classifier on the dev set, n denote the number of base classifier. The losses in Table 1 are set as follows:

$$\begin{aligned}
&\lambda^i_{PP} = \lambda^i_{NN} = e_i \\
&\lambda^i_{PN} = \lambda^i_{NP} = 1 \\
&\lambda^i_{BP} = \lambda^i_{BN} = \begin{cases} e_i + \frac{1}{\sum_{k=i+1}^{n} \frac{1}{e_k}}, & i = 1, 2, ..., n-1 \\ \frac{1+e_i}{2}, & i = n \end{cases}
\end{aligned} \tag{2}$$

Table 1
λ^i_{PP}, λ^i_{BP} and λ^i_{NP} denote the losses incurred for deciding an object to respectively positive region, boundary region and negative region in the i-th gate, when the gold label of this object is positive. λ^i_{PN}, λ^i_{BN} and λ^i_{NN} denote the losses incurred for taking the same actions in the i-th gate when the gold label of this object is negative.

	P	N
P	λ^i_{PP}	λ^i_{PN}
B	λ^i_{BP}	λ^i_{BN}
N	λ^i_{NP}	λ^i_{NN}

Firstly, we prove the loss values defined in equation (2) satisfy the assumptions in three-way decisions [33] which are:

$$\lambda^i_{PP} < \lambda^i_{BP} < \lambda^i_{NP}, \lambda^i_{NN} < \lambda^i_{BN} < \lambda^i_{PN}, \tag{3}$$

$$(\lambda^i_{PN} - \lambda^i_{BN})(\lambda^i_{NP} - \lambda^i_{BP}) > (\lambda^i_{BN} - \lambda^i_{NN})(\lambda^i_{BP} - \lambda^i_{PP}), \tag{4}$$

For strong base classifiers, we make the assumption $0 < e_i < 0.3$, for $i = 1, 2, ..., n$, then (3) is obviously satisfied since $e_i + \frac{1}{\sum_{k=i+1}^{n} \frac{1}{e_k}} \leq e_i + e_n < 0.6$. For (4):

$$\begin{aligned} (4) &\Leftrightarrow 1 - (e_i + \frac{1}{\sum_{k=i+1}^{n} \frac{1}{e_k}}) > e_i + \frac{1}{\sum_{k=i+1}^{n} \frac{1}{e_k}} - e_i \\ &\Leftrightarrow \frac{1}{\sum_{k=i+1}^{n} \frac{1}{e_k}} < \frac{1 - e_i}{2} \\ &\Leftarrow e_n < \frac{1 - e_i}{2}, \end{aligned} \tag{5}$$

$e_n < \frac{1-e_i}{2}$ is obviously satisfied since $0 < e_i < 0.3$. Furthermore, the definition of loss values in equation (2) makes $\lambda^i_{BP} - \lambda^i_{PP}$, $\lambda^i_{BN} - \lambda^i_{NN}$ negatively correlated with $n - i$, positively correlated with e_k, for $k = i + 1, ..., n$, which means the greater the number of base classifiers after the current, and the lower their classification error rate, then the samples are more likely to be divided into boundary region in current step.

Let Pr_P, Pr_N respectively denote the prediction probabilities of the positive and negative categories given by the current base classifier, and $Pr_P + Pr_N = 1$. The misclassification cost in i-th step mc_cost_i is calculated by:

$$\begin{aligned} mc\_cost_i(D_P) &= \lambda^i_{PP} Pr_P + \lambda^i_{PN} Pr_N, \\ mc\_cost_i(D_N) &= \lambda^i_{NP} Pr_P + \lambda^i_{NN} Pr_N, \\ mc\_cost_i(D_B) &= \lambda^i_{BP} Pr_P + \lambda^i_{BN} Pr_N \end{aligned} \tag{6}$$

where D_P, D_N, D_B denote the decisions to assign the object to positive region, negative region and boundary region, respectively.

The total cost in i-th step $cost_i$ is calculated by:

$$\begin{aligned} cost_i(D_P) &= mc\_cost_i(D_P) + \sum_{j=1}^{i} t\_cost_j, \\ cost_i(D_N) &= mc\_cost_i(D_N) + \sum_{j=1}^{i} t\_cost_j, \\ cost_i(D_B) &= mc\_cost_i(D_B) + \sum_{j=1}^{i+1} t\_cost_j \end{aligned} \tag{7}$$

where t_cost_j denotes the time cost incurred for j-th base classifier to classify the object. The time cost of deciding the object to boundary region in i-th step includes not only the time cost of the first i base classifiers, but also the next base classifier. We select the decision with minimal total cost as the optimal decision for the i-th decision step.

Table 2
Statistics of the five data sets used in this paper. N counts the number of instances, $dist(+,-)$ lists the class distribution, l represents the average sentence length, dev is the size of dev set and $test$ is the size of test set. CV means 3-fold cross validation, in which one is training set, one is dev set and one is test set.

Data	N	$dist(+,-)$	l	dev	$test$
IMDB	50000	(0.5, 0.5)	255	0.1	0.5
MR	10662	(0.5, 0.5)	18	CV	CV
CR	3780	(0.64, 0.36)	17	CV	CV
SUBJ	10000	(0.5, 0.5)	21	CV	CV
MPQA	10462	(0.32, 0.68)	3	CV	CV

3.2.2. Proportion-gate

In proportion-gate, fixed proportion of objects are selected into next base classifier. The objects are chosen according to their belief scores from low to high. For example, the proportion-gate with proportion of 0.1 selects 10% objects with lowest belief score to flow into next base classifier. The parameter proportions are fitting on the dev set. The belief score bs is calculated by the prediction probabilities given by the base classifiers:

$$bs = |Pr_P - Pr_N| \tag{8}$$

4. Experiments

In this section, experiments are constructed to evaluate the performance of the proposed method.

4.1. Datasets and baseline methods

We test our model on five benchmark datasets. Summary statistics of the datasets are in Table 2. We describe each dataset in detail below:

- IMDB[1]: This is a movie reviews dataset for binary sentiment classification. The dataset provides a set of 25,000 highly polar movie reviews for training, and 25,000 for testing [67].
- MR[2]: Movie reviews dataset in which each instance is a sentence [68]. The objective is to classify each review by its overall sentiment polarity, either positive or negative.
- CR[3]: Annotated customer reviews of 14 products obtained from Amazon [69]. The task is to classify each customer review into positive and negative categories.
- SUBJ: Subjectivity dataset where the task is to classify a sentence as being subjective or objective [70].
- MPQA[4]: Phrase level opinion polarity detection subtask of the MPQA dataset [71].

We use six baseline methods as base classifiers for S3WC, including three neural network models (CNN, LSTM, MLP) and three traditional sentiment classification models (SVM, MNB, NB-SVM). We compare the performance of S3WC with the existing ensemble combination techniques (majority-voting, weighted-voting and meta-learning).

4.2. Experimental settings

In CNN we use: one convolution layer and one max-pooling layer, filter windows of 3 with 200 feature maps, dropout rate of 0.2, mini-batch size of 64, activation function of Relu. In LSTM we use: dropout rate of 0.2, mini-batch size of 64, activation function of Relu. In MLP we use: two hidden layers with 64 hidden units, dropout rate of 0.5, mini-batch size of 64, activation function of Tanh. Training is done through stochastic gradient descent over shuffled mini-batches with the adadelta update rule. In SVM and NB-SVM we use linear kernel. Since feature engineering is not the focus of this work, we do not use any pre-trained word embeddings for model training. In neural network models, word embeddings are randomly initialized. In other models, we use unigram features. In majority-voting, the prediction labels are used for voting (hard voting). In weighted-voting, the weights are set according to the classification accuracies of the base classifiers on the dev set, and the prediction probabilities are used for voting (soft voting). In meta-learning, SVM with linear kernel is applied as the meta-learning model. In the mincost-gate of S3WC, t_cost_j defined in Eq. (7) is calculated by $t\_cost_j = 200 \times rt_j$, where rt_j denotes the real time (seconds) for j-th base classifier to classify an object. Each experiment was performed 10 times.

[1] http://ai.stanford.edu/~amaas/data/sentiment/.
[2] http://cs.cornell.edu/people/pabo/movie-review-data/.
[3] http://cs.uic.edu/~liub/FBS/sentiment-analysis.html.
[4] http://www.cs.pitt.edu/mpqa.

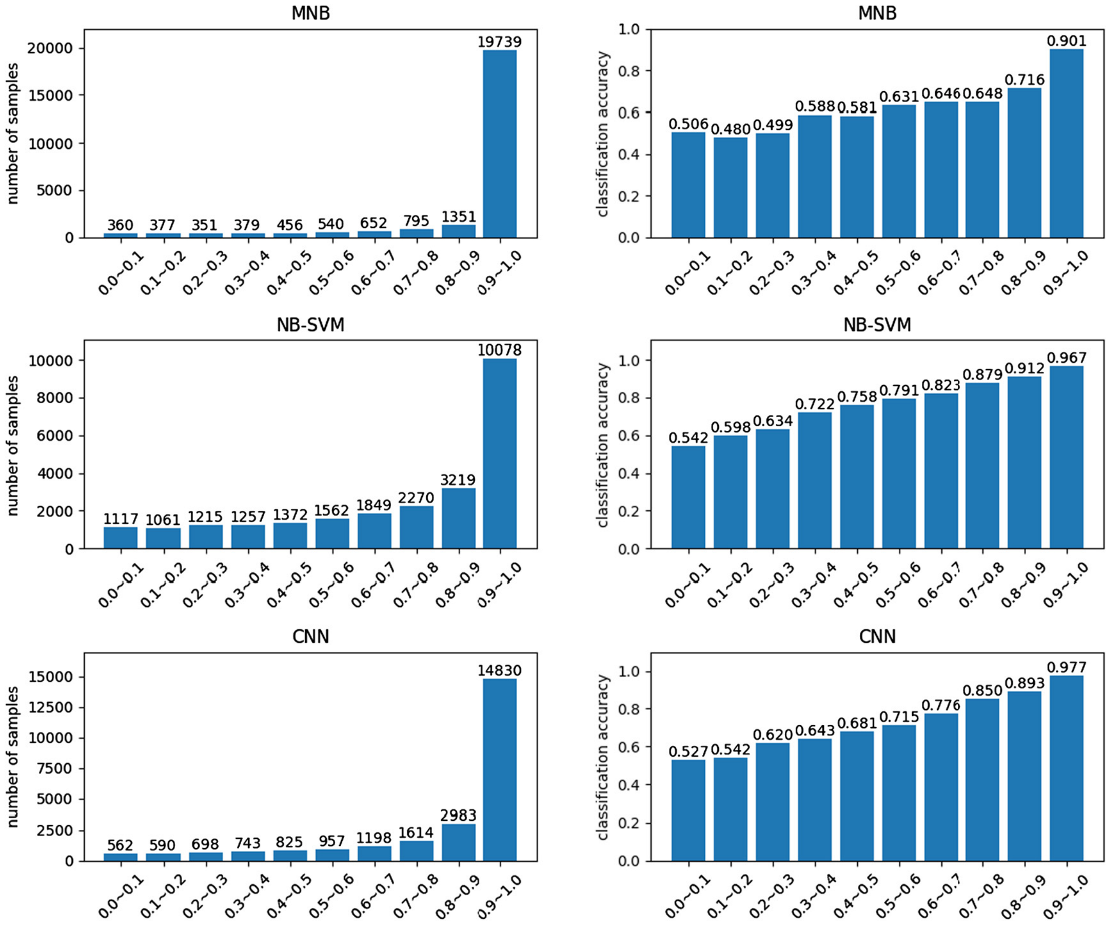

Fig. 2. The distribution of sample size and classification accuracy in different belief score intervals of MNB, NB-SVM, CNN on IMDB dataset.

As we all know, the performance of neural network models is sensitive to the parameters. We do not perform any dataset-specific parameter tuning other than early stopping, so that some base classifiers do not achieve the same results as other works. In this work, we focus on the performance comparison between S3WC and its base classifiers, and the performance comparison between S3WC and other ensemble combination techniques, whether a base classifier achieves its best performance on a dataset is not the concern of this work.

4.3. Accuracy analysis

We study how the belief score defined in Eq. (8) affects the classification accuracy. We divide the samples into 10 groups by the belief scores, and study the sample size and classification accuracy of different belief intervals. The results are shown in Fig. 2. We can see that the belief scores of most samples are in $[0.9, 1.0]$. Moreover, the higher the belief score, the higher the classification accuracy. The results in Fig. 2 illustrate the possibility of improving the classification accuracy by S3WC, because in S3WC the samples are assigned to a base classifier with a relatively high belief score.

The performance of each methods in terms of classification accuracy is shown in Table 3. To statistically measure the significance of performance difference, pairwise t-tests at 5% significance level are conducted between the methods. The results in Table 3 show that, in terms of the average accuracy, S3WC outperforms any of its base classifiers in all cases; and S3WC is statistically superior to any of its base classifiers in most cases according to the results of t-tests. As the number of base classifiers increases, the performance of S3WC is getting better, regardless of the performance of the new base classifier itself. For example, MLP is inferior to LSTM and CNN on MR dataset (MLP: 71.45%, LSTM: 74.64%, CNN: 75.03%), but we obtained a better result when pushing MLP into $S3WC_{CL}$ ($S3WC_{CL}$: 75.54%, $S3WC_{CLM_L}$: 75.97% for proportion-gate). Although the improvement of accuracy is not obvious, considering the new base classifier is much weaker than the existing ones, it's satisfactory to see that the accuracy did not decline. In most cases, S3WC with proportion-gate outperforms S3WC with mincost-gate in terms of classification accuracy.

The classification accuracies of S3WC and other ensemble combination techniques are shown in Fig. 3. The experimental results show that both majority-voting and weighted-voting have a common problem, i.e., heavy dependence on the

Table 3

Performance (mean±std) of each methods in terms of classification accuracy. ● indicates that S3WC is statistically superior to any of its base classifiers (pairwise t-test at 5% significance level). $S3WC_{LM_L M_N}$ means S3WC of LSTM, MLP and MNB, $S3WC_{CN_S S}$ means S3WC of CNN, NB-SVM and SVM, $S3WC_6$ means the S3WC of all six baseline methods. P-Gate, M-Gate denote respectively proportion-gate and mincost-gate. The order of base classifiers in S3WC is consistent with the subscript (e.g. the order of base classifiers in $S3WC_{CLM_L}$ is CNN-LSTM-MLP).

Model		IMDB	MR	CR	SUBJ	MPQA
SVM		85.00 ± 0.51	72.39 ± 0.27	75.84 ± 0.44	88.20 ± 0.22	82.75 ± 0.42
NB-SVM		85.57 ± 0.39	76.59 ± 0.43	79.08 ± 0.61	90.96 ± 0.20	82.15 ± 0.37
MNB		84.20 ± 0.31	76.12 ± 0.22	78.73 ± 0.64	91.06 ± 0.19	80.99 ± 0.44
MLP		84.95 ± 0.79	71.45 ± 0.68	77.11 ± 0.89	89.68 ± 0.52	82.46 ± 0.70
LSTM		87.23 ± 0.77	74.64 ± 0.78	77.23 ± 0.78	89.62 ± 0.55	82.56 ± 0.77
CNN		88.90 ± 0.71	75.03 ± 0.66	79.77 ± 0.88	88.84 ± 0.55	82.88 ± 0.81
$S3WC_{N_S S}$	M-Gate	86.23 ± 0.40●	77.14 ± 0.35●	79.23 ± 0.67	91.00 ± 0.30●	83.10 ± 0.51
	P-Gate	86.45 ± 0.47●	77.00 ± 0.41	78.96 ± 0.51	90.92 ± 0.32	83.19 ± 0.40
$S3WC_{M_L M_N}$	M-Gate	86.40 ± 0.72●	76.66 ± 0.56	79.52 ± 0.72●	91.10 ± 0.47	82.82 ± 0.69
	P-Gate	86.47 ± 0.70●	76.40 ± 0.43	80.00 ± 0.59●	91.10 ± 0.56	82.77 ± 0.60
$S3WC_{LM_L}$	M-Gate	87.79 ± 0.69●	75.03 ± 0.77	78.02 ± 0.47●	90.22 ± 0.61	83.00 ± 0.79
	P-Gate	87.90 ± 0.71●	75.26 ± 0.69●	79.31 ± 0.64●	90.40 ± 0.35●	83.13 ± 0.80●
$S3WC_{LN_S}$	M-Gate	88.00 ± 0.59●	76.63 ± 0.71	80.03 ± 0.57●	91.03 ± 0.35	83.18 ± 0.77●
	P-Gate	88.60 ± 0.53●	76.81 ± 0.57	80.23 ± 0.88●	91.18 ± 0.33	83.04 ± 0.60
$S3WC_{CN_S}$	M-Gate	89.45 ± 0.65	77.00 ± 0.67●	81.55 ± 0.45●	91.00 ± 0.46	83.20 ± 0.89
	P-Gate	89.77 ± 0.58	77.02 ± 0.47●	82.08 ± 0.67●	91.10 ± 0.20	83.15 ± 0.80
$S3WC_{CL}$	M-Gate	89.44 ± 0.80	75.67 ± 0.72●	80.92 ± 0.55●	89.99 ± 0.51	83.07 ± 0.58
	P-Gate	89.81 ± 0.82●	75.54 ± 0.64	81.27 ± 0.76●	90.04 ± 0.60	83.38 ± 0.61●
$S3WC_{CLM_L}$	M-Gate	89.69 ± 0.29	75.99 ± 0.81●	81.32 ± 0.87●	90.06 ± 0.57	83.22 ± 0.88
	P-Gate	89.86 ± 0.42●	75.97 ± 0.55●	81.73 ± 0.50●	90.28 ± 0.45	83.59 ± 0.90●
$S3WC_{CN_S LM_N}$	M-Gate	89.78 ± 0.59●	76.94 ± 0.62●	82.21 ± 0.41●	91.00 ± 0.57●	83.29 ± 0.77●
	P-Gate	89.91 ± 0.71●	77.13 ± 0.74●	83.01 ± 0.70●	91.46 ± 0.49●	83.76 ± 0.80●
$S3WC_6$	M-Gate	89.96 ± 0.72●	77.04 ± 0.55●	82.86 ± 0.51●	91.58 ± 0.32●	83.96 ± 0.89●
	P-Gate	**90.15±0.72●**	**77.58±0.63●**	**83.33±0.43●**	**92.05±0.66●**	**84.86±0.73●**

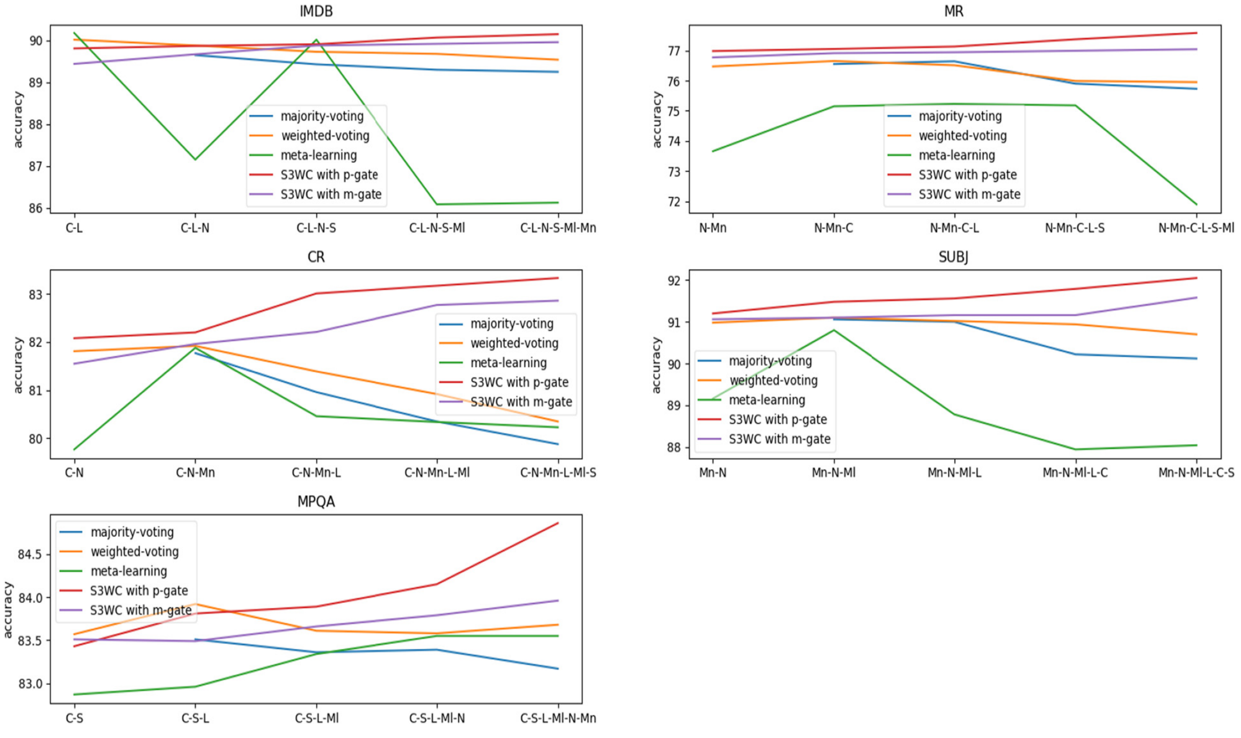

Fig. 3. The performance of ensemble combination techniques for different base classifiers. C-L-N-S represents the ensemble of CNN, LSTM, NB-SVM and SVM, other marks on the abscissa are named in a similar way. The base classifiers are ranked from highest to lowest according to their classification accuracy on the dev set. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

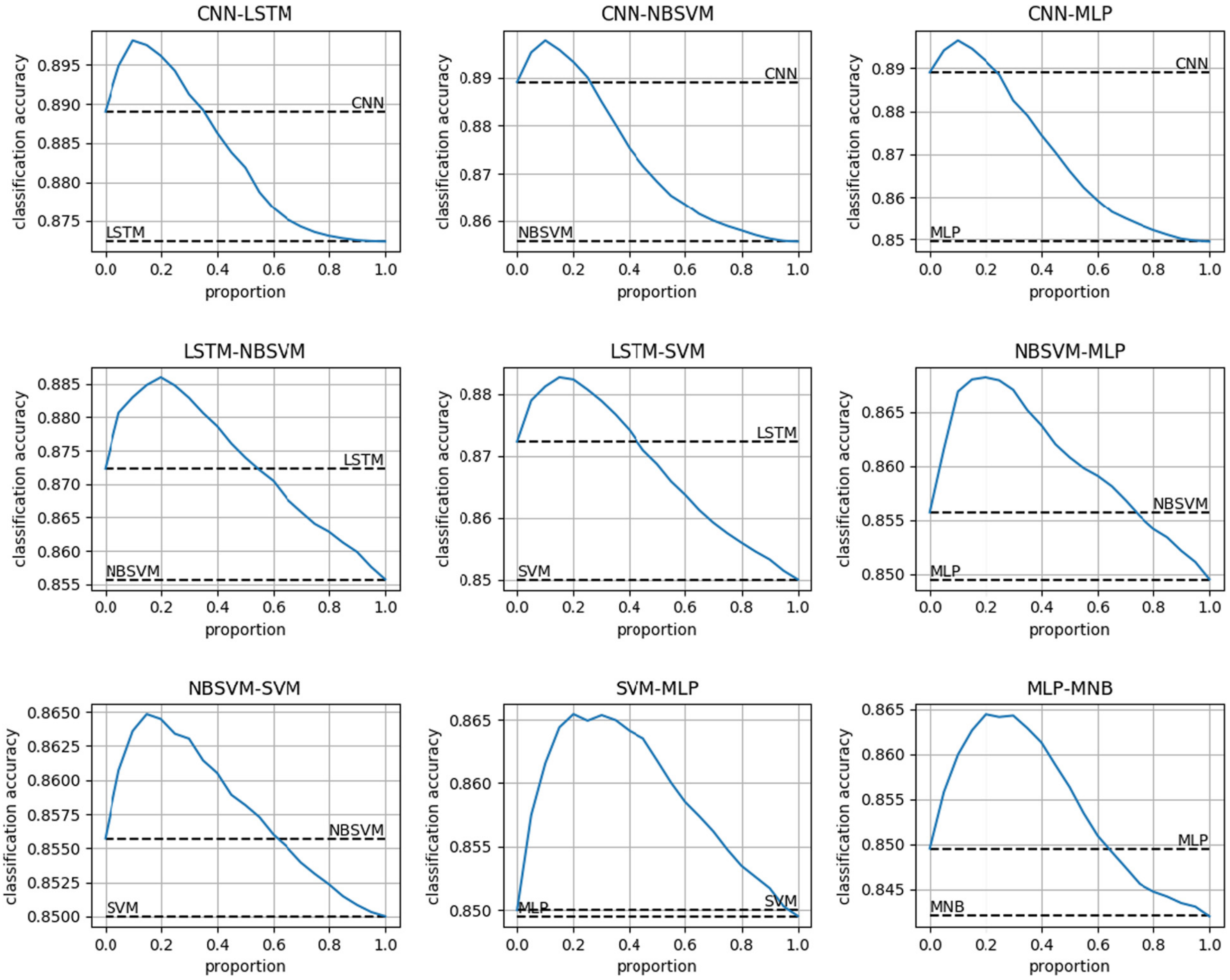

Fig. 4. Classification accuracy of 3w-combination with proportion-gate under different proportion parameters.

performance of the base classifiers. Majority-voting and weighted-voting have good performance if and only if all the base classifiers have good performance. Their performance will deteriorate if a new base classifier, which performs worse than the average of the existing ones, is added. For example, the accuracy of weighted-voting dropped from 90.02% to 89.50% when MLP is added into CNN-LSTM. Meta-learning has the same problem, furthermore, the robustness of meta-learning is even worse than voting. The performance of meta-learning is very unstable when handling the base classifiers with different performance. In contrast, S3WC has good performance and better robustness in terms of classification accuracy. When a new base classifier is added into S3WC, the overall performance will be better regardless of whether performance of the new base classifier is better than the average of the existing ones.

We discuss and explain the robustness of S3WC here. In the existing ensemble combination techniques including majority-voting, weighted-voting and meta-learning, the predictions given by each base classifiers will be more or less adopted. However, some poor performing base classifiers provide “noise predictions”, which reduce the overall prediction accuracy. Even the best base classifier will give some “noise predictions”. In S3WC, gate mechanism is applied to filter the “noise predictions” according to the prediction belief score, and predictions with low belief scores will be ignored. Therefore, S3WC is more robust than the existing ensemble combination techniques when a poor performing base classifiers is added, and the accuracy of S3WC increases steadily as the number of base classifiers increases.

We study the effect of the proportion parameter on the accuracy of S3WC with proportion-gate in the case of 2-model combination. As shown in Fig. 4, the accuracy of S3WC with proportion-gate varies with the proportion parameter, but different combinations of models produce similar accuracy curves. Proportion-gate S3WC with a suitable proportion parameter can achieve a better accuracy than its base classifiers. The optimal proportion varies in $[0.1, 0.3]$ as shown in Fig. 4, and in fact the results of Algorithm 1 are within this interval.

Algorithm 1 Proportion parameter fitting algorithm.

```
Input: Base classifiers $C_1, C_2, ..., C_n$; Proportion-gate dividing algorithm $PG$; Dev data $D_d = \{(\mathbf{x}_1^d, y_1), (\mathbf{x}_2^d, y_2), ..., (\mathbf{x}_k^d, y_k)\}$;
Output: Set of proportion parameter $pp = \{\overline{p}_1, \overline{p}_2, ..., \overline{p}_{n-1}\}$
  Calculate the classification accuracy of base classifiers on dev data: $acc_1, acc_2, ..., acc_n$.
  Sort $C_1, C_2, ..., C_n$ by $acc$, get $C_{(1)}, C_{(2)}, ..., C_{(n)}$, where $acc_{(1)} \geq acc_{(2)} \geq ... \geq acc_{(n)}$.
  for $t = 1$ to $n - 1$ do
    $\overline{p}_t = \text{argmax}_{p_t \in (0,1)}\, accuracy(PG(\{C_{(1)}, C_{(2)}, ..., C_{(t+1)}\}, \{\overline{p}_1, ..., \overline{p}_{t-1}, p_t\}), D_d)$
  end for
  $pp = \{\overline{p}_1, \overline{p}_2, ..., \overline{p}_{n-1}\}$
```

Table 4
Runtime analysis of S3WC and baselines.

Dataset	Method	Extra training time(s)	Predicting time(s)	Total time(s)
IMDB	majority-voting	–	112.1	112.1
	weighted-voting	22.3	112.1	134.4
	meta-learning	111.3	111.5	222.8
	S3WC with m-gate	22.3	15.8	38.1
	S3WC with p-gate	79.2	20.9	100.1
MR	majority-voting	–	44.8	44.8
	weighted-voting	9.1	44.5	53.6
	meta-learning	45.1	44.9	90.0
	S3WC with m-gate	10.2	13.7	23.9
	S3WC with p-gate	39.5	6.9	46.4
CR	majority-voting	–	5.1	5.1
	weighted-voting	1.3	5.1	6.4
	meta-learning	5.0	5.0	10.0
	S3WC with m-gate	1.5	1.6	3.1
	S3WC with p-gate	3.2	1.6	4.8
SUBJ	majority-voting	–	24.8	24.8
	weighted-voting	5.1	25.0	30.1
	meta-learning	24.5	24.4	48.9
	S3WC with m-gate	5.9	6.9	12.8
	S3WC with p-gate	17.8	4.5	22.3
MPQA	majority-voting	–	1.6	1.6
	weighted-voting	0.6	1.8	2.4
	meta-learning	1.6	1.5	3.1
	S3WC with m-gate	0.7	0.3	1.0
	S3WC with p-gate	1.7	0.3	2.0

4.4. Runtime analysis

The time cost of the ensemble methods in sentiment classification tasks consists of three parts: training time of base classifiers, extra training time and predicting time. Extra training time means the training time other than base classifiers training, e.g., the time for learning the parameters used in combination techniques. The base classifiers' training time of S3WC is equal to that of the baseline combination techniques. We study the runtime of S3WC and baseline combination techniques in Table 4. S3WC with mincost-gate is superior to other methods in terms of running time. In S3WC, the objects do not require the predictions of all base classifiers in most cases, and therefore the prediction time of S3WC is less than that of the baseline combination techniques. S3WC with proportion-gate uses more extra training time than S3WC with mincost-gate, which means learning the proportion parameters is more time-consuming than calculating the total cost. S3WC uses more extra training time than voting, and less predicting time than both voting and meta-learning. In general, S3WC is slightly superior in terms of the total time cost.

4.5. Total cost analysis

We study the performance of S3WC in terms of the total cost in Fig. 5. For the sake of convenience in display, we divide the total cost by the number of objects to get the average cost per object on each datasets. As shown in Fig. 5, S3WC is superior to the baseline combination techniques in terms of the total cost on all datasets used in this work. It means that S3WC minimizes the total cost while ensuring good classification accuracy. S3WC with proportion-gate is superior to S3WC with mincost-gate in terms of the classification accuracy as shown in Table 3, but it is inferior in terms of the total cost.

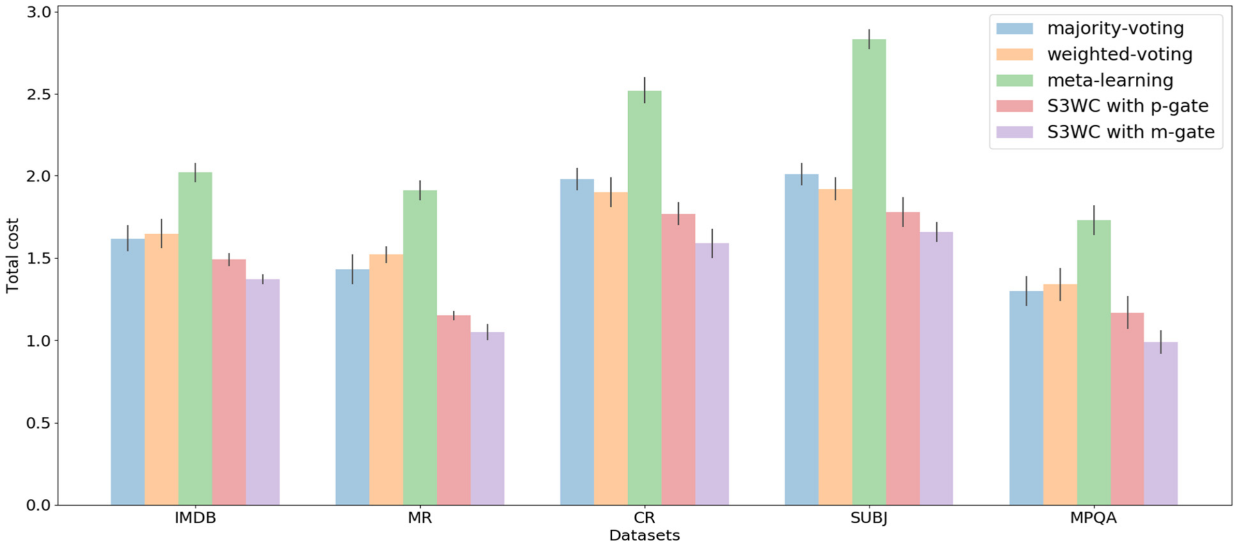

Fig. 5. The total cost of S3WC and the baseline combination techniques.

5. Conclusion

Inspired by the methodology of sequential three-way decisions, we proposed a cost-sensitive ensemble combination technique named S3WC which is designed for strong base classifiers in sentiment classification tasks. S3WC has a sequential structure, and in each step a gate is used to divide the objects into three groups: positive, negative and boundary regions. Instead of adopting the predictions of all base classifiers, S3WC selects only one base classifier to predict for each object. We propose two strategies for the gate mechanism: mincost-gate and proportion-gate. Mincost-gate divides the objects by minimizing the total cost consisting of misclassification cost and time cost. Proportion-gate divides the objects according to their prediction belief scores and fixed proportion of objects are grouped into the boundary region. The experimental results show that S3WC achieves a higher accuracy than any of its base classifiers, and S3WC performs better than the existing ensemble combination techniques (e.g. majority-voting, weighted-voting, meta-learning) in terms of accuracy, robustness, time complexity and the total cost in most cases. Proportion-gate outperforms mincost-gate in terms of accuracy in most cases, but mincost-gate is superior to proportion-gate in terms of time complexity and the total cost. In the future work, we will apply S3WC technique to more tasks, and experiment S3WC on more kinds of base classifiers, including homogeneous and heterogenous.

Acknowledgements

The work described in this paper has been supported by the Ministry of Science and Technology (Grant No. 213), the Natural Science Foundation of China (Grant No. 61673301), the Major Project of Ministry of Public Security (Grant No. 20170004).

References

[1] C.D. Manning, H. Schütze, et al., Foundations of Statistical Natural Language Processing, vol. 999, MIT Press, 1999.
[2] D.Y. Tang, B. Qin, F.R. Wei, L. Dong, T. Liu, M. Zhou, A joint segmentation and classification framework for sentence level sentiment classification, IEEE/ACM Trans. Audio Speech Lang. Process. 23 (11) (2015) 1750–1761.
[3] S.M. Kim, E. Hovy, Automatic detection of opinion bearing words and sentences, in: Companion Volume to the Proceedings of the International Joint Conference on Natural Language Processing (IJCNLP), vol. 8, 2005.
[4] B. Pang, L. Lee, S. Vaithyanathan, Thumbs up?: sentiment classification using machine learning techniques, in: Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing (EMNLP '02), 2002, pp. 79–86.
[5] J. Zhao, K. Liu, G. Wang, Adding redundant features for crfs-based sentence sentiment classification, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP '08), 2008, pp. 117–126.
[6] S. Wang, C.D. Manning, Baselines and bigrams: simple, good sentiment and topic classification, in: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, 2012, pp. 90–94.
[7] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: International Conference on Neural Information Processing Systems, 2012, pp. 1097–1105.
[8] A. Graves, A. Mohamed, G. Hinton, Speech recognition with deep recurrent neural networks, in: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 2013, pp. 6645–6649.
[9] T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, in: Advances in Neural Information Processing Systems, 2013, pp. 3111–3119.
[10] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, P. Kuksa, Natural language processing (almost) from scratch, J. Mach. Learn. Res. 12 (Aug) (2011) 2493–2537.

[11] Q. Le, T. Mikolov, Distributed representations of sentences and documents, in: Proceedings of the 31st International Conference on Machine Learning (ICML-14), 2014, pp. 1188–1196.

[12] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324.

[13] S.W.T. Yih, X. He, C. Meek, Semantic parsing for single-relation question answering, in: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, 2014, pp. 643–648.

[14] Y.L. Shen, X.D. He, J.F. Gao, L. Deng, G. Mesnil, Learning semantic representations using convolutional neural networks for web search, in: Proceedings of the 23rd International Conference on World Wide Web, 2014, pp. 373–374.

[15] N. Kalchbrenner, E. Grefenstette, P. Blunsom, D. Kartsaklis, A convolutional neural network for modelling sentences, in: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, 2014, pp. 212–217.

[16] Y. Kim, Convolutional neural networks for sentence classification, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014, pp. 1746–1751.

[17] O. Araque, I. Corcuera-Platas, J.F. Sánchez-Rada, C.A. Iglesias, Enhancing deep learning sentiment analysis with ensemble techniques in social applications, Expert Syst. Appl. 77(C) (2017) 236–246.

[18] G. Wang, J.S. Sun, J. Ma, K.Q. Xu, J. Gu, Sentiment classification: the contribution of ensemble learning, Decis. Support Syst. 57 (1) (2014) 77–93.

[19] Y.Y. Yao, Three-way decisions with probabilistic rough sets, Inf. Sci. 180 (3) (2010) 341–353.

[20] R. Goudey, Do statistical inferences allowing three alternative decisions give better feedback for environmentally precautionary decision-making?, J. Environ. Manag. 85 (2) (2007) 338–344.

[21] Y.Y. Yao, S. Wang, X.F. Deng, Constructing shadowed sets and three-way approximations of fuzzy sets, Inf. Sci. 412–413 (Supplement C) (2017) 132–153.

[22] X.Y. Zhang, D.Q. Miao, Three-way attribute reducts, Int. J. Approx. Reason. 88 (2017) 401–434.

[23] J. Qian, C.Y. Dang, X.D. Yue, N. Zhang, Attribute reduction for sequential three-way decisions under dynamic granulation, Int. J. Approx. Reason. 85 (2017) 196–216.

[24] H. Fujita, T.R. Li, Y.Y. Yao, Advances in three-way decisions and granular computing, Knowl.-Based Syst. 91 (2016) 1–3.

[25] Y.Y. Yao, Three-way decision and granular computing, Int. J. Approx. Reason. 103 (2018) 107–123.

[26] J.T. Yao, N. Azam, Web-based medical decision support systems for three-way medical decision making with game-theoretic rough sets, IEEE Trans. Fuzzy Syst. 23 (1) (2015) 3–15.

[27] D. Liu, T.R. Li, D.C. Liang, Three-way government decision analysis with decision-theoretic rough sets, Int. J. Uncertain. Fuzziness Knowl.-Based Syst. 20 (supp01) (2012) 119–132.

[28] B. Zhou, Y. Yao, J. Luo, Cost-sensitive three-way email spam filtering, J. Intell. Inf. Syst. 42 (1) (2014) 19–45.

[29] H.X. Li, L.B. Zhang, B. Huang, X.Z. Zhou, Sequential three-way decision and granulation for cost-sensitive face recognition, Knowl.-Based Syst. 91 (2016) 241–251.

[30] H. Yu, C. Zhang, G.Y. Wang, A tree-based incremental overlapping clustering method using the three-way decision theory, Knowl.-Based Syst. 91 (2016) 189–203.

[31] Y. Zhang, J.T. Yao, Gini objective functions for three-way classifications, Int. J. Approx. Reason. 81 (2017) 103–114.

[32] X.D. Yue, Y.F. Chen, D.Q. Miao, J. Qian, Tri-partition neighborhood covering reduction for robust classification, Int. J. Approx. Reason. 83 (2017) 371–384.

[33] Y.Y. Yao, An outline of a theory of three-way decisions, in: Rough Sets and Current Trends in Computing, Springer, 2012, pp. 1–17.

[34] Y. Yao, X. Deng, Sequential three-way decisions with probabilistic rough sets, Inf. Sci. 180 (3) (2010) 341–353.

[35] L. Wang, F.J. Ren, D.Q. Miao, Multi-label emotion recognition of weblog sentence based on bayesian networks, IEEJ Trans. Electr. Electron. Eng. 11 (2) (2016) 178–184.

[36] Y.H. Rao, Q. Li, X.D. Mao, L. Wenyin, Sentiment topic models for social emotion mining, Inf. Sci. 266 (2014) 90–100.

[37] P.D. Turney, Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews, in: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, 2002, pp. 417–424.

[38] X.W. Ding, B. Liu, P.S. Yu, A holistic lexicon-based approach to opinion mining, in: Proceedings of the 2008 International Conference on Web Search and Data Mining, ACM, 2008, pp. 231–240.

[39] M. Thelwall, K. Buckley, G. Paltoglou, Sentiment strength detection for the social web, J. Assoc. Inform. Sci. Technol. 63 (1) (2012) 163–173.

[40] H. Cho, S. Kim, J. Lee, J.-S. Lee, Data-driven integration of multiple sentiment dictionaries for lexicon-based sentiment classification of product reviews, Knowl.-Based Syst. 71 (2014) 61–71.

[41] G. Paltoglou, M. Thelwall, A study of information retrieval weighting schemes for sentiment analysis, in: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, 2010, pp. 1386–1395.

[42] H.D. Kim, C.X. Zhai, Generating comparative summaries of contradictory opinions in text, in: Proceedings of the 18th ACM Conference on Information and Knowledge Management, 2009, pp. 385–394.

[43] G. Katz, N. Ofek, B. Shapira Consent, Context-based sentiment analysis, Knowl.-Based Syst. 84 (2015) 162–178.

[44] M. Franco-Salvador, F.L. Cruz, J.A. Troyano, P. Rosso, Cross-domain polarity classification using a knowledge-enhanced meta-classifier, Knowl.-Based Syst. 86 (2015) 46–56.

[45] M.S. Hajmohammadi, R. Ibrahim, A. Selamat, H. Fujita, Combination of active learning and self-training for cross-lingual sentiment classification with density analysis of unlabelled samples, Inf. Sci. 317 (Supplement C) (2015) 67–77.

[46] D.Y. Tang, F.R. Wei, B. Qin, N. Yang, T. Liu, M. Zhou, Sentiment embeddings with applications to sentiment analysis, IEEE Trans. Knowl. Data Eng. 28 (2) (2016) 496–509.

[47] Y.F. Ren, R. Wang, D.H. Ji, A topic-enhanced word embedding for Twitter sentiment classification, Inf. Sci. 369 (2016) 188–198.

[48] O. Irsoy, C. Cardie, Opinion mining with deep recurrent neural networks, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014, pp. 720–728.

[49] S.W. Lai, L.H. Xu, K. Liu, J. Zhao, Recurrent convolutional neural networks for text classification, in: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015, pp. 2267–2273.

[50] D.Y. Tang, B. Qin, T. Liu, Document modeling with gated recurrent neural network for sentiment classification, in: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2015, pp. 1422–1432.

[51] H. Zhao, Z.D. Lu, P. Poupart, Self-adaptive hierarchical sentence model, in: Proceedings of the 24th International Conference on Artificial Intelligence, 2015, pp. 4069–4076.

[52] L. Rokach, Ensemble Methods for Classifiers, Springer US, Boston, MA, 2005, pp. 957–980.

[53] R. Xia, C.Q. Zong, S.S. Li, Ensemble of feature sets and classification algorithms for sentiment classification, Inf. Sci. 181 (6) (2011) 1138–1152.

[54] E. Fersini, E. Messina, F. Pozzi, Sentiment analysis: Bayesian ensemble learning, Decis. Support Syst. 68 (Supplement C) (2014) 26–38.

[55] E. Fersini, E. Messina, F. Pozzi, Expressive signals in social media languages to improve polarity detection, Inf. Process. Manag. 52 (1) (2016) 20–35.

[56] R. Xia, C.Q. Zong, A pos-based ensemble model for cross-domain sentiment classification, in: International Joint Conference on Natural Language Processing, 2011, pp. 614–622.

[57] A. Aue, M. Gamon, Customizing sentiment classifiers to new domains: a case study, in: International Conference on Recent Advances in Natural Language Processing, 2005, pp. 33–39.

[58] V. Sehgal, C. Song, Sops: Stock prediction using web sentiment, in: IEEE International Conference on Data Mining Workshops, 2007, in: ICDM Workshops, 2008, pp. 21–26.
[59] P. Melville, N. Shah, L. Mihalkova, R.J. Mooney, Experiments on ensembles with missing and noisy data, Lect. Notes Comput. Sci. 3077 (2004) 293–302.
[60] J. Prusa, T.M. Khoshgoftaar, D.J. Dittman, Using ensemble learners to improve classifier performance on tweet sentiment data, in: IEEE International Conference on Information Reuse and Integration, 2015, pp. 252–257.
[61] D. Liu, Y.Y. Yao, T.R. Li, Three-way investment decisions with decision-theoretic rough sets, Int. J. Comput. Intell. Syst. 4 (1) (2011) 66–74.
[62] X.P. Yang, J.T. Yao, Modelling multi-agent three-way decisions with decision-theoretic rough sets, Fundam. Inform. 115 (2–3) (2012) 157–171.
[63] H.X. Li, L.B. Zhang, X.Z. Zhou, B. Huang, Cost-sensitive sequential three-way decision modeling using a deep neural network, Int. J. Approx. Reason. 85 (2017) 68–78.
[64] H.R. Zhang, F. Min, Three-way recommender systems based on random forests, Knowl.-Based Syst. 91 (2016) 275–286.
[65] H. Yu, Q.F. Zhou, A cluster ensemble framework based on three-way decisions, in: International Conference on Rough Sets and Knowledge Technology, Springer, 2013, pp. 302–312.
[66] W.W. Li, Z.Q. Huang, X.Y. Jia, Two-phase classification based on three-way decisions, in: International Conference on Rough Sets and Knowledge Technology, Springer, 2013, pp. 338–345.
[67] A.L. Maas, R.E. Daly, P.T. Pham, D. Huang, A.Y. Ng, C. Potts, Learning word vectors for sentiment analysis, in: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 2011, pp. 142–150.
[68] B. Pang, L. Lee, Seeing stars: exploiting class relationships for sentiment categorization with respect to rating scales, in: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, 2005, pp. 115–124.
[69] M.Q. Hu, B. Liu, Mining and summarizing customer reviews, in: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2004, pp. 168–177.
[70] B. Pang, L. Lee, A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts, in: Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, 2004, p. 271.
[71] J. Wiebe, T. Wilson, C. Cardie, Annotating expressions of opinions and emotions in language, Lang. Resour. Eval. 39 (2) (2005) 165–210.