Information Sciences 547 (2021) 271–288

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Two-stage three-way enhanced technique for ensemble learning in inclusive policy text classification

Decui Liang *, Bochun Yi

School of Management and Economics, University of Electronic Science and Technology of China, Chengdu 610054, China

ARTICLE INFO

Article history:
Received 11 April 2020
Received in revised form 18 July 2020
Accepted 13 August 2020
Available online 24 August 2020

Keywords:
Three-way decisions
Decision support
Ensemble learning
Convolutional neural networks
Inclusive policy text classification

ABSTRACT

With the development of the social economy, small and medium-sized enterprises (SMEs) play a vital role in promoting economic development. Multiple local governments in China are developing policy recommended platforms in order to help SMEs better understand the inclusive policy. However, these online platforms manually extract the key information from the inclusive policy texts, which takes a lot of time and causes low efficiency. The policy text is composed of some paragraphs and each paragraph corresponds to a topic. When we classify the paragraphs into different topics, there exists a decision risk of text misclassification. Therefore, we design two-stage based three-way enhanced technique to automatically classify these text paragraphs into the predefined categories. At the first stage, by using ensemble learning algorithms, we construct an ensemble convolution neural network (CNN) model in order to ensure the generalization ability and stability of text classification results. Meanwhile, we develop a new weight determination method to integrate the prediction results of all base classifiers according to the accuracy and classification confidence. With the help of three-way decisions (3WD), we assign the samples with poor resolution to the boundary area for secondary classification, which can reduce the decision risk. At the second stage, in order to classify the boundary region samples and improve the overall classification results, we further utilize traditional machine learning method as the secondary classifier. Finally, we develop some comparison experiments to verify our proposed method. The experimental results show that the two-stage three-way enhanced classification framework is valid and obtains a better performance. Our proposed method can effectively support the designment of policy recommended platforms and serve SMEs.

1. Introduction

With the rapid development of economy, small and medium-sized enterprises (SMEs) have been gradually at the center of most industrial activities in the developed and developing countries [30]. According to the result of Ref. [31], it is estimated that SMEs are the predominant form of enterprises, which approximately comprises 99% of all firms. In the course of economic operation, although the government has issued many inclusive policies, few enterprises know the information, which may lead to them miss the development opportunity. In the report to the 19th national congress of the communist party of China (CPC), it was clearly proposed to strengthen innovation support for SMEs and encourage the development

* Corresponding author.
E-mail addresses: decuiliang@126.com (D. Liang), 724309779@qq.com (B. Yi).

https://doi.org/10.1016/j.ins.2020.08.051
0020-0255/

of private enterprises. Subsequently, in order to help SMEs find inclusive policies, many cities have launched policy service recommendation platforms, e.g., Langfang city,[1] Chongqing city[2] and Chengdu city.[3] However, these recommendation platform manually extract the key information of the policy texts from the government website. Take Chengdu city as an example, Chengdu SMEs association is developing policy service recommendation platform based an applet of WeChat, which is shown in Fig. 1b. In Fig. 1a, the staff manually split the policy text into paragraphs and then put paragraphs into the different topics shown in the red boxes of Fig. 1b. Obviously, this manual approach takes a lot of time and causes low efficiency. Therefore, it is necessary to automatically classify these policy text for the government.

The method of assigning paragraphs to different topics can be regarded as text classification. Text classification is able to automatically divide many texts into the predefined categories, which is an important task in Natural Language Process (NLP) and machine learning [19,24]. Many applications of text mining can be modelled as a text classification problem [28], including sentiment analysis [45,46], news classification [3] and so on. For a text, it consists of many words, which is called as unstructured data. However, machine learning algorithms require structured format input [34]. Therefore, one of the tasks of text classification is to represent natural language text with an appropriate format.

Text representation is the process of transforming the unstructured texts into structured data in the format of numerical vectors. In general, there are two main directions in text representation: traditional feature-based methods and deep learning methods [46]. The traditional feature-based methods treat each word as a potential feature, like N-gram, term frequency-inverse document frequency (TF-IDF), etc, which ignore the semantic relation among words and may cause high dimensions. Deep learning methods have achieved remarkable results in many fileds, e.g., image recognition and NLP [2]. Because of its superiority, word embeddings have become an increasingly popular vector representation technique [13]. Mikolov et al. [26,27] proposed two efficient word embedding model, namely, Continuous Skip-gram model (Skip-gram) and Continuous Bag-of-Words model (CBOW). CBOW may be more suitable when the data sets are not too big [4]. Word embeddings are distributed representation methods, which have lower dimension and consider the semantic information. Doc2Vec as the newest representation schemes is an extension of the Word2Vec representation [20]. The purpose of Doc2Vec is to determine an appropriate distributed representation for a single document. Based on these deep learning methods, the paper firstly investigate the numerical vector transformation of the unstructured texts in the policy text classification task.

In recent years, convolutional neural network (CNN) model of deep learning can provide accurate and reliable results for NLP fields [48]. For the policy text classification task, CNN can automatically adjust parameters by means of convolution layer and pooling layer in forward and back propagation. However, there are two challenges: (1) How to ensure the generalization ability and stability of classification results? (2) How to reduce the risk of misclassification? For the first challenge, ensemble learning is exactly a method that can train a set of individual classifiers and has great generalization ability. Nowadays, ensemble learning method is gaining more and more attention in the machine learning and data mining communities [10,16,28,36]. Two of the most common ensemble learning algorithms are bagging and boosting [25]. Therefore, in this paper, we propose an ensemble CNN model for policy text classification based on bagging and boosting, respectively. Generally, for text classification, there are only two corresponding relationships between the element and the category: belonging and not belonging. Obviously, the decision risk of text misclassification is not considered. Hence, with respect to the second challenge, three-way decisions (3WD) method proposed by Yao [41] can provide a new tool for the text classification. The core idea of the 3WD method is to divide an universe into three disjoint regions, including an acceptance decision region, a deferment decision region and a rejection decision region. If the decision maker has enough information, he (or she) can quickly make a decision, i.e., acceptance and rejection. Otherwise, the decision maker can choose to postpone the decision. The methodology of 3WD is widely applied in both theoretic fields and application fields, e.g., decision-theoretic rough set models [40], recommendation system [44] and multiple attribute decision making [15].

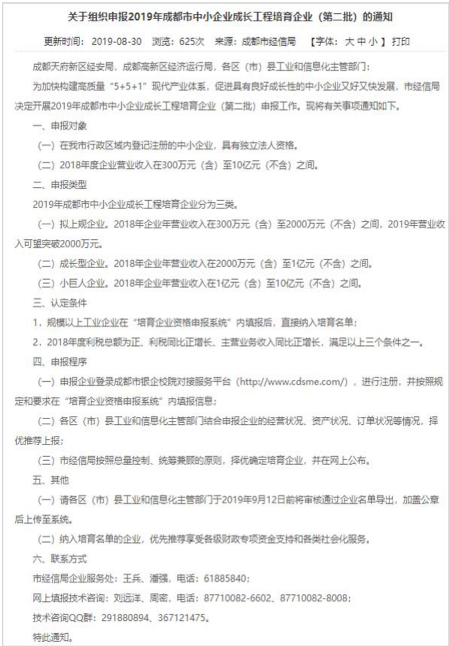
In addition to deep learning methods, some traditional machine learning methods can also get a good performance, such as Support Vector Machines (SVM) [6], Decision Tree (DT) [12], Random Forest (RF) [33]. When we train models, deep learning and traditional machine learning are obviously quite different from text representation viewpoints. In order to improve the overall performance, inspired by the results of Ref. [45], we construct a two-stage policy text classification algorithms with 3WD by combining deep learning and traditional machine learning. Specifically speaking, at the first stage, we train an ensemble CNN model according to the classification confidence and accuracy of each base classifier. Meanwhile, we put forward a new determination method of the weights to integrate the prediction results of all base classifiers. When we compute the classification results of the first stage, we divide the samples with low confidence into boundary region by using 3WD. At the second stage, we further classify the boundary region samples with traditional machine learning algorithms.

In summary, with the help of 3WD, the main innovations of this paper can be summarized into three aspects: (1) We propose a new two-stage policy text classification model framework based on 3WD. (2) For the ensemble learning, we design a new method to determine the weight of each base classifier. (3) After performing many comparative experiments, we have verified that 3WD-AdaCNN-SVM model is the best combination strategy in the inclusive policy text classification.

[1] http://www.sme-lf.com/home/index.

[2] http://smedsc.org.cn/.

[3] http://www.cdsme.com/.



关于组织申报2019年成都市中小企业成长工程培育企业（第二批）的通知

更新时间：2019-08-30 浏览：625次 来源：成都市经信局 【字体：大 中 小】 打印

成都天府新区经安局，成都高新区经济运行局，各区（市）县工业和信息化主管部门：

为加快构建高质量"5+5+1"现代产业体系，促进具有良好成长性的中小企业又好又快发展，市经信局决定开展2019年成都市中小企业成长工程培育企业（第二批）申报工作。现将有关事项通知如下。

一、申报对象

（一）在我市行政区域内登记注册的中小企业，具有独立法人资格。

（二）2018年度企业营业收入在300万元（含）至10亿元（不含）之间。

二、申报类型

2019年成都市中小企业成长工程培育企业分为三类。

（一）拟上规企业。2018年企业年营业收入在300万元（含）至2000万元（不含）之间，2019年营业收入可望突破2000万元。

（二）成长型企业。2018年企业年营业收入在2000万元（含）至1亿元（不含）之间。

（三）小巨人企业。2018年企业年营业收入在1亿元（含）至10亿元（不含）之间。

三、认定条件

1．规模以上工业企业在"培育企业资格申报系统"内填报后，直接纳入培育名单；

2．2018年度利税总额为正、利税同比正增长、主营业务收入同比正增长，满足以上三个条件之一。

四、申报程序

（一）申报企业登录成都市银企校院对接服务平台（http://www.cdsme.com/），进行注册，并按照规定和要求在"培育企业资格申报系统"内填报信息；

（二）各区（市）县工业和信息化主管部门结合申报企业的经营状况、资产状况、订单状况等情况，择优推荐上报；

（三）市经信局按照总量控制、统筹兼顾的原则，择优确定培育企业，并在网上公布。

五、其他

（一）请各区（市）县工业和信息化主管部门于2019年9月12日前将审核通过企业名单导出，加盖公章后上传至系统。

（二）纳入培育名单的企业，优先推荐享受各级财政专项资金支持和各类社会化服务。

六、联系方式

市经信局企业服务处：王兵、潘强，电话：61885840；

网上填报技术咨询：刘远洋、周密，电话：87710082-6602、87710082-8008；

技术咨询QQ群：291880894、367121475。

特此通知。

(a) The original policy text

(b) The segmentation of the policy text

Fig. 1. Policy strategy system of Chengdu SMEs association.

The rest of paper is organized as follows: Section 2 introduces related works. In Section 3, we present two-stage three-way enhanced technique for ensemble learning models. In Section 4, we design some experiments and provide the corresponding discussion in details. Finally, we draw some conclusions and some possible future works in Section 5.

2. Related works

2.1. Three-way decisions (3WD)

3WD is a new decision technology and has attracted extensive attention in various fields [45]. Considering the decision risk and uncertainty, 3WD was initially proposed by Yao [41] based on rough sets and Bayesian decision procedure [39,40]. In recent years, Yao [43] further proposed a generalized trisecting-acting-outcome (TAO) model for 3WD, see Fig. 2.

In Fig. 2, TAO model is consist of three components, i.e., trisecting, acting and outcome evaluation. Specifically, the function of trisecting is to divide a whole into three related and relatively independent parts. The function of acting is to apply a set of strategies to address the three parts. The function of outcome evaluation is to measure the effectiveness of the results. In general, the three related and relatively independent parts are called as positive region, negative region and boundary region of rough set, which are corresponding to acceptance, rejection and deferment strategies, respectively. In order to determine which region the given object belongs to, let $(L, \preceq)$ be a totally ordered set for two elements α, β with a condition of $\beta < \alpha$. Therefore, for an evaluation function $\upsilon : U \rightarrow L$, the three regions can be defined as follows [42]:

$$\begin{cases} POS_{(\alpha,\beta)}(\upsilon) = & \{x \in U | \upsilon(x) \succeq \alpha\} \\ NEG_{(\alpha,\beta)}(\upsilon) = & \{x \in U | \upsilon(x) \preceq \beta\} \\ BND_{(\alpha,\beta)}(\upsilon) = & \{x \in U | \beta \prec \upsilon(x) \prec \alpha\} \end{cases}, \tag{1}$$

where $0 \leqslant \beta < \alpha \leqslant 1$. In this case, $POS_{(\alpha,\beta)}(\upsilon), NEG_{(\alpha,\beta)}(\upsilon), BND_{(\alpha,\beta)}(\upsilon)$ are corresponding to positive region, negative region and boundary region, respectively. The main idea of the 3WD methodology is to divide a universe into three disjoint regions of Eq. (1). Then, different decision strategies can be adopted for the different regions [38,47]. The traditional decision-making method is mainly two-way decision, which definitely determines the category of each object. Unlike two-way decisions, 3WD takes into account the uncertainty caused by the inadequate information and adds a deferment decision. It may reduce high decision risk for compulsively classifying the uncertain objects [21].

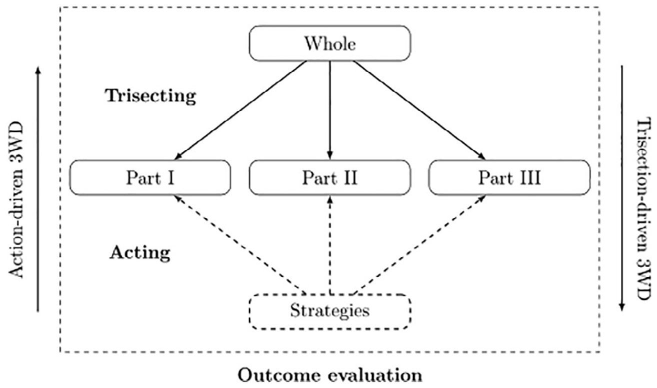


Fig. 2. TAO model of three-way decisions.

2.2. Text classification

In NLP and machine learning, text classification is one of the important tasks [19,24]. In general, there are two main methods in text classification: traditional machine learning methods and deep learning methods. For the traditional machine learning methods, Srivastava et al. [35] used support vector machine (SVM) and random forest (RF) to build a health care text classification system. Galitsky [9] used nearest-neighbor machine learning method and proposed syntactic parse tree to classify the reviews. Kim et al. [18] incorporated semantic concept features into term feature statistics and proposed semantic Naive Bayes learning method. Liu et al. [23] compared four feature selection algorithms and five machine learning algorithms to address multi-class sentiment classification task. These methods focus on feature selection and the performance depends on the effect of feature engineering, which may take a lot of time to train models.

For the deep learning methods, they can learn rich representations and extract complex and abstract features from their input [7]. Elnagar et al. [8] utilized Word2Vec embedding and proposed some varietal Gated Recurrent Unit (GRU) models for Arabic text categorization. Abdi et al. [1] combined Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) to classify a user's opinion expressed in reviews for multi-feature fusion. For the government, it should establish various complaint channels to report quality problems. Based on this thought, Zhong et al. [48]. proposed a Convolutional Neural Network (CNN) approach to classify the short texts into predefined categories. These deep learning methods can automatically capture the texts features and reserve semantic information.

In a nutshell, no matter traditional machine learning methods or deep learning methods, they only consider black-or-white situation. That is to say, the text classification algorithms are classified as belonging to or not belonging to a certain category. This basic idea is the traditional two-way decisions, which may cause high misclassification risks. With the aid of 3WD, we can assign indistinguishable objects to the boundary area and wait for much more information to execute secondary classification.

2.3. Ensemble learning

Ensemble learning can be used to combine the predictions of multiple classifiers in order to achieve a better performance in comparison with a single base classifier. Remarkable improvement of generalization ability can be achieved by using ensemble learning methods [29]. For the ensemble learning, there are mainly two types. The one is Bagging and the other one is AdaBoost. Bagging algorithm uses bootstrap sampling for getting data subset in order to train some base classifiers [32]. The core principle of AdaBoost is to put more emphasis on the weak base classifiers which have high bias error at each boosting iteration [39].

Ensemble learning methods have been successfully applied in a numerous applications. For example, Sun and Gao [36] proposed Adaboost-BP neural network model in the exploration of energy saving potential. Gao et al. [10] combined Adaboost and CNN for head detection in crowded surveillance environment. Onan et al. [28] applied Bagging and AdaBoost methods in numerous machine learning algorithms for keyword extraction and text classification, which is just a one-stage classification. Kang et al. [16] used boosting method of text-based hidden Markov models for text classification. Roshan and Asadi [32] proposed evolutionary multi-objective optimization for imbalanced data sets classification by using Bagging method. Zhang et al. [46] proposed a cost-sensitive ensemble learning in sentiment classification. Compared to the single classifier, ensemble learning can combine the predictions of multiple classifiers to achieve a better performance, which ensure the generalization ability and stability of classification results. Thus, we develop an ensemble CNN model and use it as the first stage classifier in this paper.

3. Three-way enhanced methods for ensemble learning of policy text classification

In this section, based on 3WD, we design two-stage classification model for the policy text and propose three-way enhanced methods for ensemble learning. The most popular ensemble learning methods are Bagging and AdaBoost. Therefore, we present Bagging-based model and AdaBoost-based model, respectively. The architecture of the proposed model is shown in Fig. 3. In Fig. 3, the proposed model contains two stages. In the Stage 1, we choose two ensemble learning methods, i.e., Bagging and AdaBoost, respectively, and CNN model is regarded as base classifier. Inspired by the results of Ref. [45], in this paper, each CNN is trained by its own accuracy and confidence, which will be covered in the next subsection. Because 3WD can provide an effective method to address the uncertainty and indistinguishability, we perform the secondary classification for the samples with low discrimination in the first stage classification results. Hence, in the Stage 2, in order to be different from the deep learning method, we can choose traditional machine learning method, such as SVM and so on, as the secondary classifier. Meanwhile, we can deeply investigate and compare the performance of different combination methods.

3.1. The process of CNN for policy text classification

Since Kim [17] used a simple CNN for the classification tasks in 2014, it has aroused wide concern [19,45,46]. In this paper, we utilize CNN as our base classifier at the first stage of classification. The basic structure of CNN is shown in Fig. 4, which has five layers in total.

In Fig. 4, for the input layer, it is a word embedding matrix, which requires the same dimensions. Each line is a pre-trained Word2Vec for a certain word. If the length of the text is greater than the number of lines, then the text should be truncated. Otherwise, we should add zero rows. For the convolution layer, CNN relies on some filters to capture the features automatically. For the pooling layer, we employ max-pooling method to obtain the text representation. Then, in order to avoid overfitting, we add dropout layer. Finally, because this is a classification task, we use Softmax method to estimate the predicted probability in the output layer.

The CNN designed for the text classification task accepts n-word input text (padded if necessary). The concatenated word vector is represented as $\mathbf{X} \in \mathbb{R}^{u \times v}$, where u is the number of words in the input text and v is the size of the embedding dimension for each word. Therefore, the CNN model will execute convolution operations on $\mathbf{X}$ by means of the convolution filter j whose weight has the dimension $\mathbf{W}^j \in \mathbb{R}^{h \times k}$. Note that h is the window size and k is the number of words of the filter. With respect to the policy text classification, we know that k equals to v. By using the filter $\mathbf{W}^j$, we can obtain a feature c_i generated from a window of words $X_{(i:i+h-1)} \in \mathbb{R}^{h \times k}$ as follows:

$$c_i^j = f(W^j \cdot X_{(i:i+h-1)} + b^j), \tag{2}$$

where b^j is the bias of $\mathbf{W}^j$ and f is a non-linear function such as the Rectified Linear Unit (ReLU). By applying the $j-th$ convolution filter on $\mathbf{X}$, we can obtain a feature map as follows:

$$\mathbf{c}^j = [c_1^j, c_2^j, \ldots, c_i^j, \ldots, c_{u-h+1}^j]. \tag{3}$$

Then, we apply a max-pooling operation over the feature map and obtain a maximum value, i.e., $\hat{c}^j = \max(\mathbf{c}^j)$. In order to capture different features, we adopt multiple convolution filters and each filter complies with the same rule. Then, we can acquire a feature vector as follows:

$$\mathbf{x} = [\hat{c}^1, \hat{c}^2, \ldots, \hat{c}^i, \ldots, \hat{c}^{n_j}], \tag{4}$$

where n_j is the number of filters. To avoid overfitting, we add dropout layer as follows:

$$\mathbf{z} = g(\mathbf{x}). \tag{5}$$

Finally, in order to predict the category of the input text $\mathbf{X}$, the feature vector $\mathbf{z}$ is fed into a fully-connected softmax function. Thus, we can acquire the predicted probability vector $p \in \mathbf{R}^5$ as follows:

$$p = softmax(\mathbf{W}^T\mathbf{z} + b), \tag{6}$$

where $\mathbf{W} \in \mathbb{R}^{n_j \times 5}$ and $b \in \mathbf{R}^5$ are weight matrix and bias for the output layer. Note that the number "5" means the number of classification categories in this paper. For the output layer, because of the multi-classification task, let p_{1st} and p_{2nd} denote the maximum and second highest probability values of the predicted categories. According to the results of Ref. [45], the confidence of CNN model is computed as follows:

$$CF = p_{1st} - p_{2nd}. \tag{7}$$

In Eq. (7), the higher CF, the more confident the CNN can classify the input object. In this paper, we denote ϵ as the confidence threshold. With the aid of 3WD, if $CF \geqslant \epsilon$, then the CNN model can obtain high discrimination degree for the predicted category.

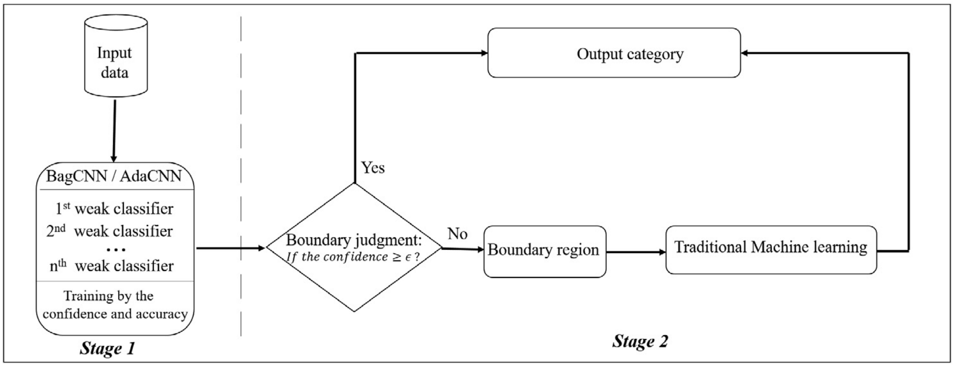


Fig. 3. The architecture of the proposed model.

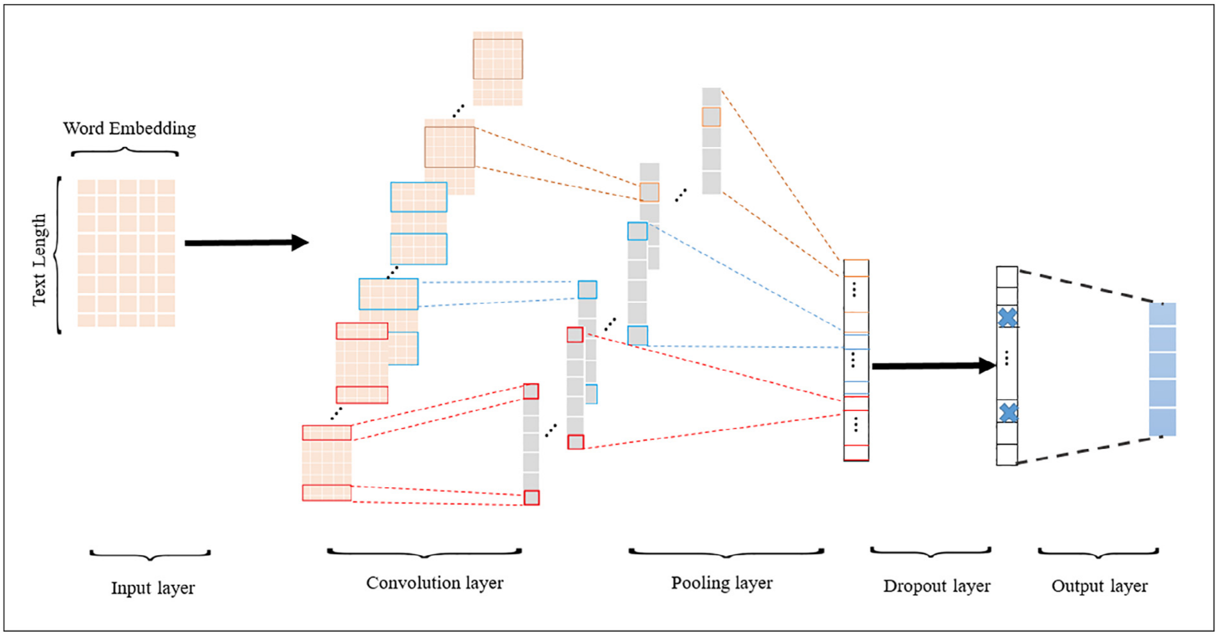


Fig. 4. The architecture of CNN model.

3.2. New ensemble learning for policy text classification at the first stage

For the classification problem, there exist a little contingency in the final prediction. Considering the contingency of classification results, it is necessary to apply many classifiers to acquire the comprehensive performance. Ensemble learning can be used to integrate many base classifiers and to improve generalization performance. Therefore, at the first stage, we utilize CNN as base classifier and deeply investigate two ensemble learning methods of Bagging and AdaBoost, respectively. Meanwhile, we design the corresponding algorithms.

3.2.1. 3WD-BagCNN model

In this subsection, we introduce the ensemble method of Bagging [32], which is a parallel ensemble learning method. Then, we propose 3WD-BagCNN model based on 3WD and CNN. Suppose that the number of samples is m, the number of labels is n and the number of sampling rounds is T. The sample i is represented as (x_i, y_i), which x_i and y_i respectively stands for the text vector and label. According to the Bagging method, we can obtain a total sample set T and each sample set contains m samples. Next, we can concurrently train the base classifier $h_t(1 \leqslant t \leqslant T)$.

For each base classifier h_t, according to Eq. (7), let m_0 be the number of low discrimination samples, which satisfies the condition $CF < \epsilon$. Hence, the classification capacity of the base classifier h_t can be calculated as follows:

$$b_t = 1 - \frac{m_0}{m}. \tag{8}$$

In addition, we count the number of misclassified samples and denote it as m_1. The classification accuracy of the base classifier h_t can be obtained as follows:

$$a_t = 1 - \frac{m_1}{m}. \tag{9}$$

Based on a_t and b_t, we can get the weight of each base classifier h_t as follows:

$$w_t^{Bag} = \frac{a_t \cdot b_t}{\sum_{t=1}^{T} a_t \cdot b_t}. \tag{10}$$

For a new input data x_k, the probability of the base classifier h_t classifying x_k into the category j is $(p_k^j)^t$. By integrating all base classifiers, we can get the prediction probability value of x_k at the first stage as follows:

$$\hat{y}_k^j = \sum_{t=1}^{T} w_t^{Bag} \cdot (p_k^j)^t. \tag{11}$$

Based on the overall prediction $\hat{y}_k^j$, the original method of the text classification just classifies the object x_k according to the maximum value of $\hat{y}_k^j$. For example, if $\hat{y}_k^1$ is the maximum value, then x_k will be classified into the first category. However, this may cause high misclassification risks, especially when the values among $\hat{y}_k^j$ are indistinguishable. Therefore, we should measure the confidence of the ensemble model as follows:

$$CF_k = \hat{y}_{k1st} - \hat{y}_{k2nd}, \tag{12}$$

where $\hat{y}_{k1st}$ and $\hat{y}_{k2st}$ denote the maximum and second highest probability values of the predicted categories. If $CF_k \geqslant \epsilon$, then we can say that the performance of the first stage classification is excellent. Otherwise, the data x_k will be put into the boundary region of 3WD and should be judged once again. In general, the specific steps are shown in Algorithm 1 below.

Algorithm 1 3WD-BagCNN method

Input: The number of sampling round T.
Output: The results of classification at the first stage.
1: **for** $t = 1; t \leqslant T; t++$ **do**
2: Obtain the sample set D_t.
3: Train CNN model h_t on the D_t.
4: Calculate the classification accuracy a_t and capacity b_t by using Eqs. (8) and (9).
end for
5: Calculate the weight of each base classifier by using Eq. (10).
6: For the new data x_k, sum up the results of all base classifiers and obtain $\hat{y}_k^j$ by using Eq. (11).
7: Calculate the confidence CF_k by using Eq. (12).
8: **if** $CF_k \geqslant \epsilon$ **then**
9: End of the classification.
10: **else**
11: Assign x_k to the boundary region and wait for the secondary classification.
end if
12: **return** The results of classification at the first stage.

3.2.2. 3WD-AdaCNN model

In this subsection, we will introduce AdaBoost [39] and propose 3WD-AdaCNN model based on 3WD and CNN. AdaBoost is a serial ensemble learning method. At the end of each round, the weights of the training samples are changed adaptively. In AdaBoost method, for the first sampling round, the initial weights of training samples can be denoted as $w = \{w_i = \frac{1}{m} | i = 1, 2, \ldots, m\}$. The error rate of the base classifier h_t is computed as follows:

$$\sigma_t = \frac{1}{m}\left[\sum_{i=1}^{m} w_i \cdot I(h_t(x_i) \neq y_i)\right], \tag{13}$$

where $I(\cdot)$ stands for the indicator function. Also, we can get the classification capacity of each base classifier h_t by Eq. (8). The weight of each base classifier is determined as follows:

$$w_t^{Ada} = \sqrt{\frac{1}{2} ln\left(\frac{1-\sigma_t}{\sigma_t}\right) \cdot b_t}. \tag{14}$$

The weight of h_t can be used to update the weights of training samples. Let w_i^t be the weight of the data (x_i, y_i) at the round t. Therefore, the weight updating mechanism of AdaBoost method is given below:

$$w_i^{t+1} = \frac{w_i^t}{Z_t} \times \begin{cases} e^{-w_t^{Ada}}, & h_t(x_i) = y_i \\ e^{w_t^{Ada}}, & h_t(x_i) \neq y_i \end{cases}, \tag{15}$$

where Z_t is a normalization factor. For the new data x_k, the probability of the base classifier h_t classifying x_k into the category j is $(p_k^j)^t$. By adding the results of all base classifiers, we can obtain the overall prediction at the first stage as follows:

$$\hat{y}_k^j = \sum_{t=1}^{T} w_t^{Ada} \cdot (p_k^j)^t. \tag{16}$$

Based on the overall prediction result $\hat{y}_k^j$ and Eq. (12), we can utilize the similar way of 3WD to handle x_k. The specific steps are shown in Algorithm 2 below.

Algorithm 2 3WD-AdaCNN method

Input: The number of sampling round T.
Output: The results of classification at the first stage.
1: Initialize the weights of training samples $w = \{w_i = \frac{1}{m} | i = 1, 2, \ldots m\}$.
2: **for** $t = 1; t \leqslant T; t++$ **do**
3: Obtain the sample set D_t according to w.
4: Train CNN model h_t on the D_t.
5: Classify the original data sets by using h_t.
6: Calculate the error rate σ_t by using Eq. (13).
7: **if** $\sigma_t > 0.5$ **then** $w = \{w_i = \frac{1}{m} | i = 1, 2, \ldots m\}$
8: Return step 3.
end if
9: Calculate the weight of each base classifier by using Eq. (14).
10: Update the weights of training samples by using Eq. (15).
end for
11: For the new data x_k, sum up the results of all base classifiers and obtain $\hat{y}_k^j$ by using Eq. (16).
12: Calculate the confidence CF_k by using Eq. (12).
13: **if** $CF_k \geqslant \epsilon$ **then**
14: End of the classification.
15: **else**
16: Assign x_k to the boundary region and wait for the secondary classification.
end if
17: **return** The results of classification at the first stage.

3.3. The boundary region policy text classification at the second stage

For the text classification task, there are usually two directions, i.e., traditional machine learning methods and deep learning methods. Due to the strong learning ability, deep learning methods perform better, especially for the text classification task. However, no matter traditional machine learning methods or deep learning methods, they only consider black-or-white situation and they directly classify objects into a certain category or not. This basic idea may cause high misclassification risks. In order to avoid this problem, with the help of 3WD, we can delay the decisions for the indistinguishable objects and put them into boundary region. After that, we are able to focus on the objects in boundary region and conduct the secondary classification, which may reduce the misclassification risks.

For the two-stage classification model of Fig. 3, we design 3WD-BagCNN and 3WD-AdaCNN models at the first stage, respectively, see Algorithms 1 and 2. With the help of 3WD, at the second stage, we further employ different traditional machine learning methods to improve the performance of the boundary region policy text classification. The specific steps are shown in Algorithm 3 below.

Algorithm 3 The boundary region policy text classification at the second stage with traditional machine learning methods

Input: The type of ensemble learning method.
Output: The final results of classification at the second stage.
Initialize the indexes of the two ensemble learning methods as $N = \{1, 2\}$.
for $i = 1; i \leqslant \#N; i++$ **do**
  if i = 1 **then**
    Execute the Bagging ensemble method.
    Acquire the boundary region data set D_B.
  else
    Execute the AdaBoost ensemble method.
    Acquire the boundary region data set D_B.
  end if
end for
Classify data set D_B by using different traditional machine learning methods.
return The final results of classification at the second stage.

Based on 3WD, we can accurately classify the input objects by using two stages. According to the results of Algorithms 1 and 2, we clearly know how Bagging and AdaBoost methods work at the first stage. At the second stage, in order to capture the text features by using different text representation methods, we adopt the several traditional machine learning methods in Algorithm 3. In the next section, we will conduct some experimental analysis to verify the advantages of our proposed methods.

4. Experimental analysis and discussions

In this section, we conduct some experiments to evaluate our proposed methods of Section 3. We can introduce the data sets and baseline methods in advance. Then, we compare our proposed methods with other methods based on these data sets. Finally, the corresponding sensitivity analysis and discussions are further investigated.

4.1. Data sets and baseline methods

In the experimental analysis, we focus on two policy text data sets. The first one is from Chengdu city shown in Fig. 5a and the second one is from Xiamen city shown in Fig. 5b. These two data sets all consist of five labels. The two data sets are made up of paragraphs which have been manually cut off from the complete policy text, e.g., the result of Fig. 1.

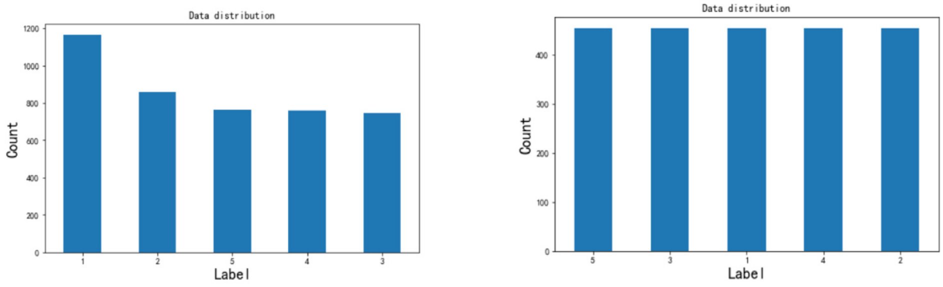
For clarity, the two methods of Section 3, i.e., 3WD-BagCNN method and 3WD-AdaCNN method can be compared with the following baseline methods. They are:

- Support Vector Machine (SVM) [6].
- Logistic Regression (LR) [5].
- K-Nearest Neighbour (KNN) [22].
- Random Forest (RF) [33].
- Decision Tree (DT) [5].
- Convolutional Neural Network (CNN) [17], which is corresponding to the base classifier of our models.
- Recurrent Neural Network (RNN) [1], which is usually for the sequential data tasks and may cause vanishing gradient.
- Long Short-Term Memory (LSTM) [1], which is a variant of RNN. The problem of gradient disappearance can be solved to some extent.
- Bi-directional Long Short-Term Memory (BiLSTM) [24], which is composed of forward LSTM and backward LSTM.
- Gated Recurrent Unit (GRU) [11], which maintains the effect of LSTM while making the structure simpler.

Note that the first five methods are traditional machine learning methods. In order to take advantage of corpus, we use Doc2Vec to represent the input texts for the first five methods. The last five methods belong to deep learning methods with Word2Vec format of the input texts. Except for the ten independent methods, in order to validate the performance of our ensemble models, we also compare another five two-stage classifiers by combining the traditional machine learning models and Bagging or AdaBoost methods. All methods have been list in Table 1.

4.2. Experimental settings

For the ensemble learning, we chose the number of sampling $T = 5$. Because there is no specific corpus about policy texts of SMEs, we crawl more than 30,000 articles with more than 100,000,000 words as our training corpus to train Word2Vec



(a) The policy texts from Chengdu city (b) The policy texts from Xiamen city

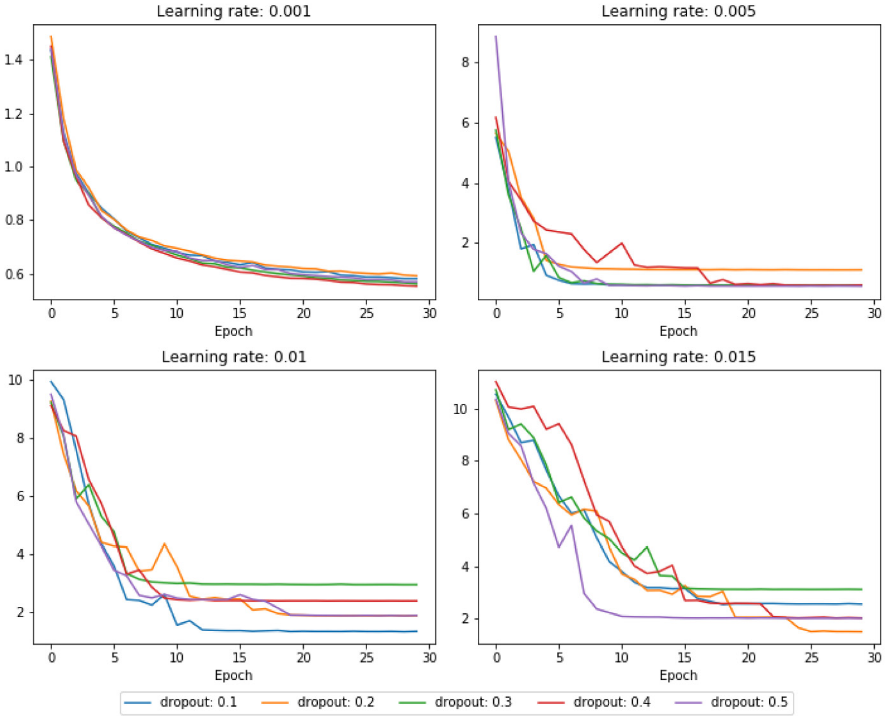
Fig. 5. Policy texts from Chengdu city and Xiamen city.

Table 1
The performance results of our proposed models and baseline methods.

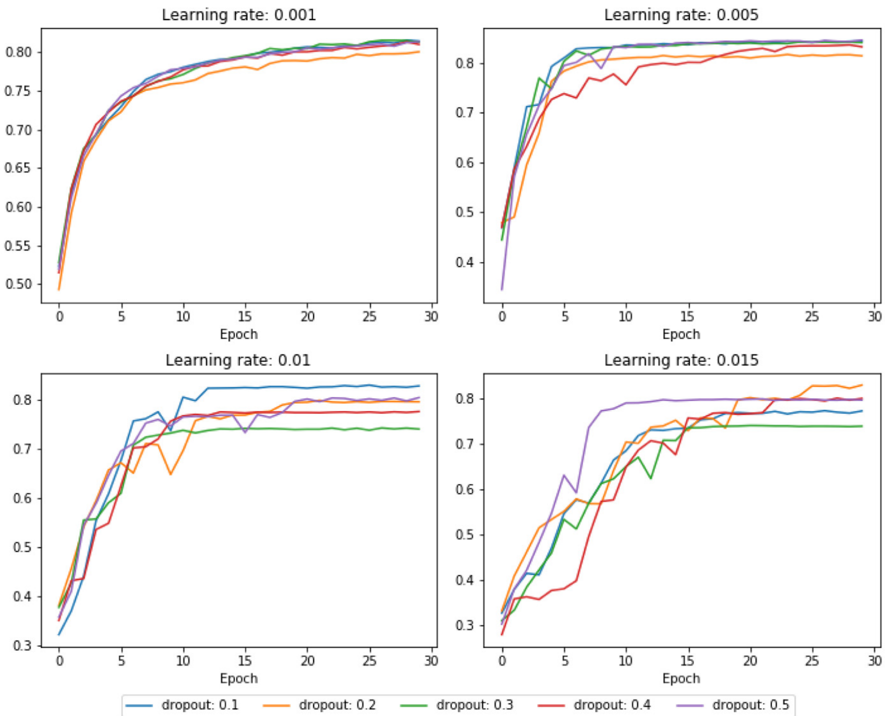
Method	Chengdu		Xiamen	
	Accuracy	Macro-F1	Accuracy	Macro-F1
3WD – AdaCNN – SVM	**0.9418**	**0.9433**	**0.8678**	**0.8670**
3WD – AdaCNN – LR	0.9325	0.9339	0.8348	0.8340
3WD – AdaCNN – KNN	0.9196	0.9226	0.7291	0.7338
3WD – AdaCNN – RF	0.9185	0.9225	0.8018	0.7973
3WD – AdaCNN – DT	0.8941	0.8982	0.7687	0.7666
3WD – BagCNN – SVM	0.9325	0.9341	**0.8326**	**0.8285**
3WD – BagCNN – LR	**0.9336**	**0.9345**	0.8304	0.8282
3WD – BagCNN – KNN	0.9092	0.9118	0.7621	0.7606
3WD – BagCNN – RF	0.9290	0.9304	0.8150	0.8113
3WD – BagCNN – DT	0.9057	0.9078	0.7709	0.7656
SVM	0.9080	0.9125	0.7841	0.7841
LR	0.9080	0.9114	0.7930	0.7952
KNN	0.8370	0.8478	0.6035	0.6195
RF	0.7870	0.7981	0.6894	0.6877
DT	0.7008	0.7125	0.5749	0.5719
CNN	0.8871	0.8909	0.7357	0.7269
RNN	0.8124	0.8143	0.5639	0.5695
LSTM	0.8881	0.8922	0.7291	0.7308
BiLSTM	0.8893	0.8905	0.7489	0.7473
GRU	0.8531	0.8574	0.5947	0.5976
3WD – AdaCNN – CNN	**0.8976**	**0.9010**	**0.7225**	**0.7201**
3WD – BagCNN – CNN	**0.9069**	**0.9088**	**0.7269**	**0.7122**

and Doc2Vec. The Word2Vec and Doc2Vec have dimensionality of 300 and are trained using the CBOW and DBOW, respectively. The activation function is ReLU and the training is done through stochastic gradient descent (SGD). In order to avoid overfitting, we add dropout layer and carry out the early stopping strategy, which can automatically stop training if the training accuracy no longer boosts three times in succession. When training the base classifier, we firstly set the confidence threshold ϵ to 0.3 and then discuss the changing influence of ϵ. In this paper, we employ ensemble learning methods, which can avoid error and contingency at the sampling stage. Therefore, we can use holdout validation. For each algorithm, we repeatedly run five times with randomly partitioned training (80 percent) and testing (20 percent) data. In addition, we randomly sample 20% of training data as validation data. All experiments are implemented in Python3.7 and performed on a computer with Intel (R) Core(TM) i5-8300H CPU and NVIDIA GeForce GTX1050Ti with memory equipped by 16 GB RAM.

The parameters of some baseline methods are as follows: For CNN model, the filter windows are 2, 3, 4 with 200 feature maps each. For RNN, LSTM, BiLSTM and GRU models, we set the number of hidden layer neurons to 100. The batch size of these all deep learning models is equal to 64. For the traditional machine learning models, we use sklearn toolkit in Python. In this case, we set regularization intensity to $1e5$ in LR model and penalty coefficient to 11 in SVM model. Besides, we leave the remaining parameters at their default values.

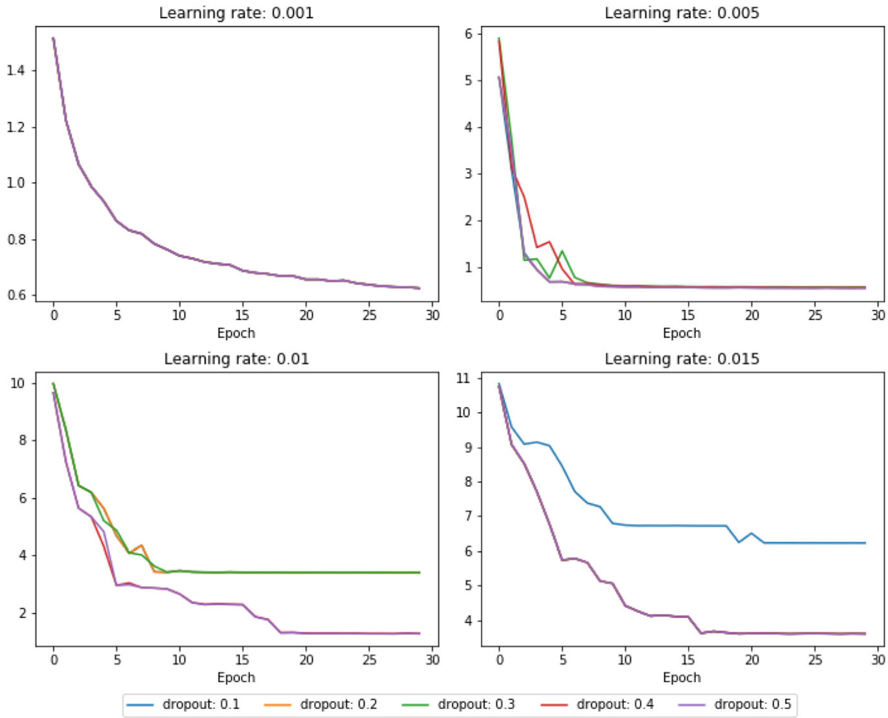


(a) The loss of the validation set

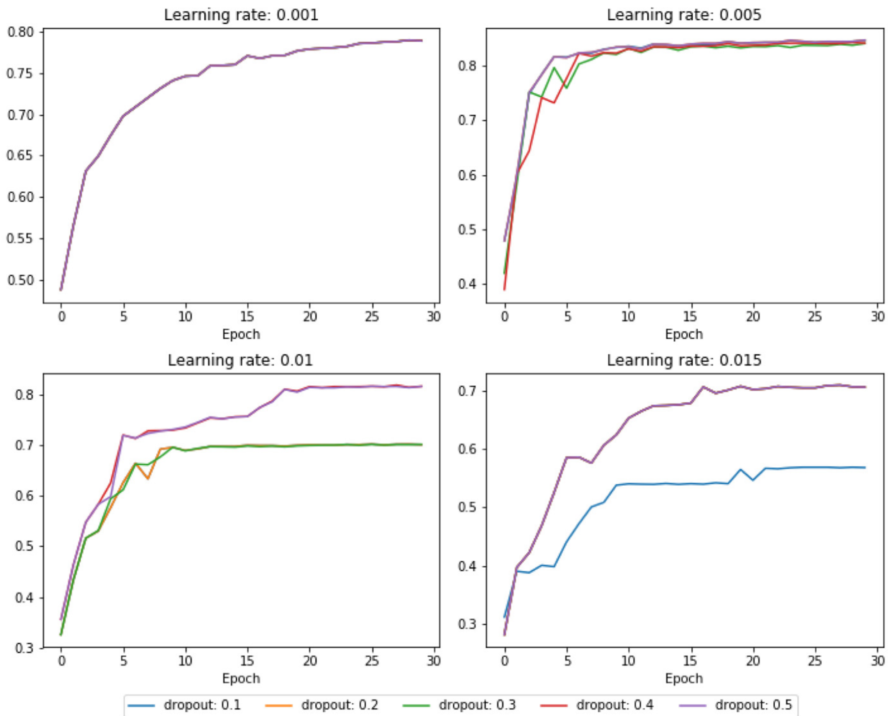


(b) The accuracy of the validation set

Fig. 6. The loss and accuracy of the validation set in 3WD-BagCNN method.



(a) The loss of the validation set



(b) The accuracy of the validation set

Fig. 7. The loss and accuracy of the validation set in 3WD-AdaCNN method.

4.3. Learning rate and dropout rate

In order to determine the learning rate and dropout rate for CNN in the first stage of classification, we use the Chengdu city policy texts to train the 3WD-BagCNN and 3WD-AdaCNN models via SGD based on the validation set. The loss and accuracy of validation set under 3WD-BagCNN and 3WD-AdaCNN methods are shown in Figs. 6 and 7, respectively.

Table 2
The performance results of 3WD-AdaCNN and 3WD-BagCNN methods at the first stage

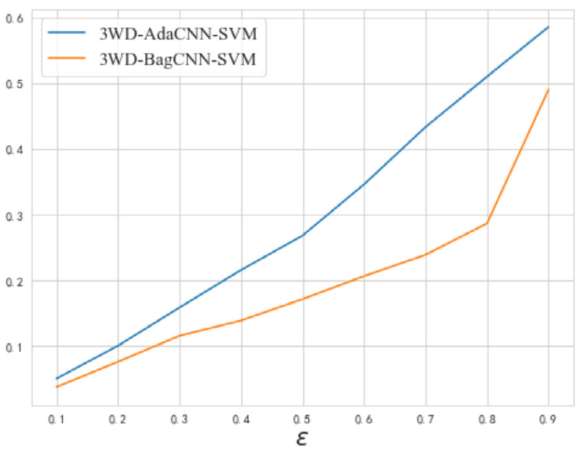
Method	Chengdu			Xiamen		
	Accuracy	Macro-F1	Rate of test set	Accuracy	Macro-F1	Rate of test set
3WD-AdaCNN	**0.9502**	0.9501	84%	**0.8710**	**0.8455**	55%
3WD-BagCNN	0.9500	**0.9515**	88%	0.8468	0.8035	73%

Table 3
The comparisons in Chengdu policy texts at the second stage

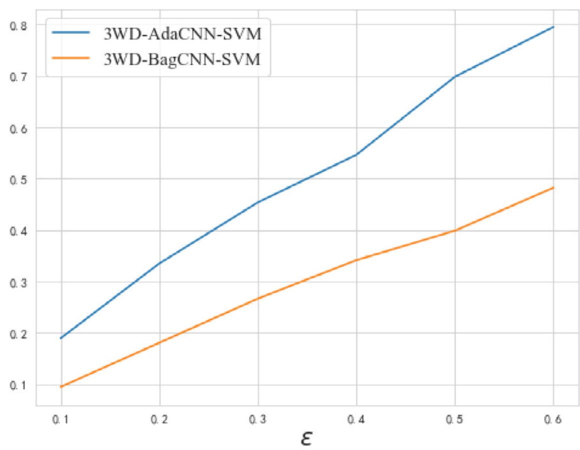
Method	3WD-AdaCNN			3WD-BagCNN		
	Accuracy	Macro-F1	Rate of test set	Accuracy	Macro-F1	Rate of test set
SVM	**0.8971**	**0.8655**	16%	0.7980	0.7792	12%
LR	0.8382	0.7878	16%	**0.8081**	**0.7917**	12%
KNN	0.7574	0.7548	16%	0.5960	0.6171	12%
RF	0.7500	0.6994	16%	0.7677	0.7490	12%
DT	0.5956	0.5801	16%	0.5657	0.5545	12%
CNN	**0.6148**	**0.5290**	16%	**0.5521**	**0.5305**	12%

Table 4
The comparisons in Xiamen policy texts at the second stage

Method	3WD-AdaCNN			3WD-BagCNN		
	Accuracy	Macro-F1	Rate of test set	Accuracy	Macro-F1	Rate of test set
SVM	**0.8641**	**0.8609**	45%	**0.7934**	**0.8178**	27%
LR	0.7913	0.7912	45%	0.7851	0.7910	27%
KNN	0.5583	0.5528	45%	0.5289	0.5445	27%
RF	0.7184	0.6897	45%	0.7273	0.7070	27%
DT	0.6456	0.6275	45%	0.5620	0.5340	27%
CNN	**0.5385**	**0.4873**	45%	**0.4017**	**0.3722**	27%



(a) Chengdu policy texts



(b) Xiamen policy texts

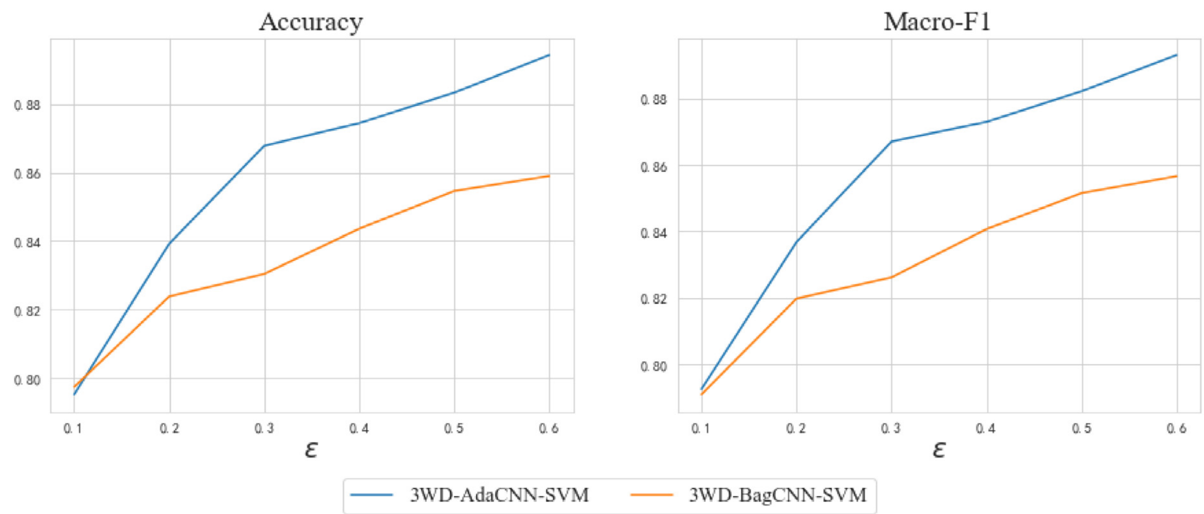
Fig. 8. The rate of test sets at the second stage with the change of ϵ.

From the results of Figs. 6 and 7, we find that no matter what kind of the ensemble method, for the learning rate, there is no doubt that the smaller the learning rate is, the faster the loss and the accuracy of the validation set fall in the beginning. For the dropout rate, the bigger the dropout rate is, the less smooth and the more unstable the curve is. Hence, we select 0.005 as the learning rate and 0.3 as the dropout rate, which can obtain higher accuracy and more smooth curve according to the performance of the validation set.

4.4. Experiment results and discussion

Since we are faced with a multi-class classification, we can mainly use two indicators to measure performance, i.e, *accuracy* and $macro - F1$ [14], which can reflect how well the predictions are balanced among different classes. Thus, based on policy texts of Chengdu city and Xiamen city, the performance results of our proposed models and baseline methods are summarized in Table 1.

(a) Chengdu policy texts



(b) Xiamen policy texts

Fig. 9. The changes of the accuracy and Macro-F1 with the change of ϵ.

In Table 1, we can find that our proposed two-stage models perform better than any other single classifier, no matter what traditional machine learning model uses in the second stage. For example, 3WD-AdaCNN-SVM and 3WD-BagCNN-SVM surpass SVM model both in Chengdu and Xiamen city data. Moreover, among all traditional machine learning methods, 3WD-AdaCNN-SVM and 3WD-BagCNN-SVM achieve better performance. The two-stage strategy combines the advantages of deep learning and traditional machine learning methods, which can improve accuracy overall. In addition, to illustrate the necessary of using traditional machine learning method at the second stage, we conduct a comparative experiment. By considering the accuracy and operating efficiency, we select CNN model as the secondary classifier. Almost all strategies that combine 3WD-AdaCNN, 3WD-BagCNN and traditional machine learning methods are better than 3WD-AdaCNN-CNN and 3WD-BagCNN-CNN. It implies that if we continue to use CNN to classify the boundary region objects, then there is no difference from the first stage in essence. Hence, it is feasible to employ traditional machine learning methods to improve the classification performance of the boundary region data. The results also show the advantage of our strategy.

In this paper, our proposed methods use two-stage classification strategy based on 3WD. Therefore, it is necessary to investigate the difference of the classification results between the two stages. For the first stage, the comparison results

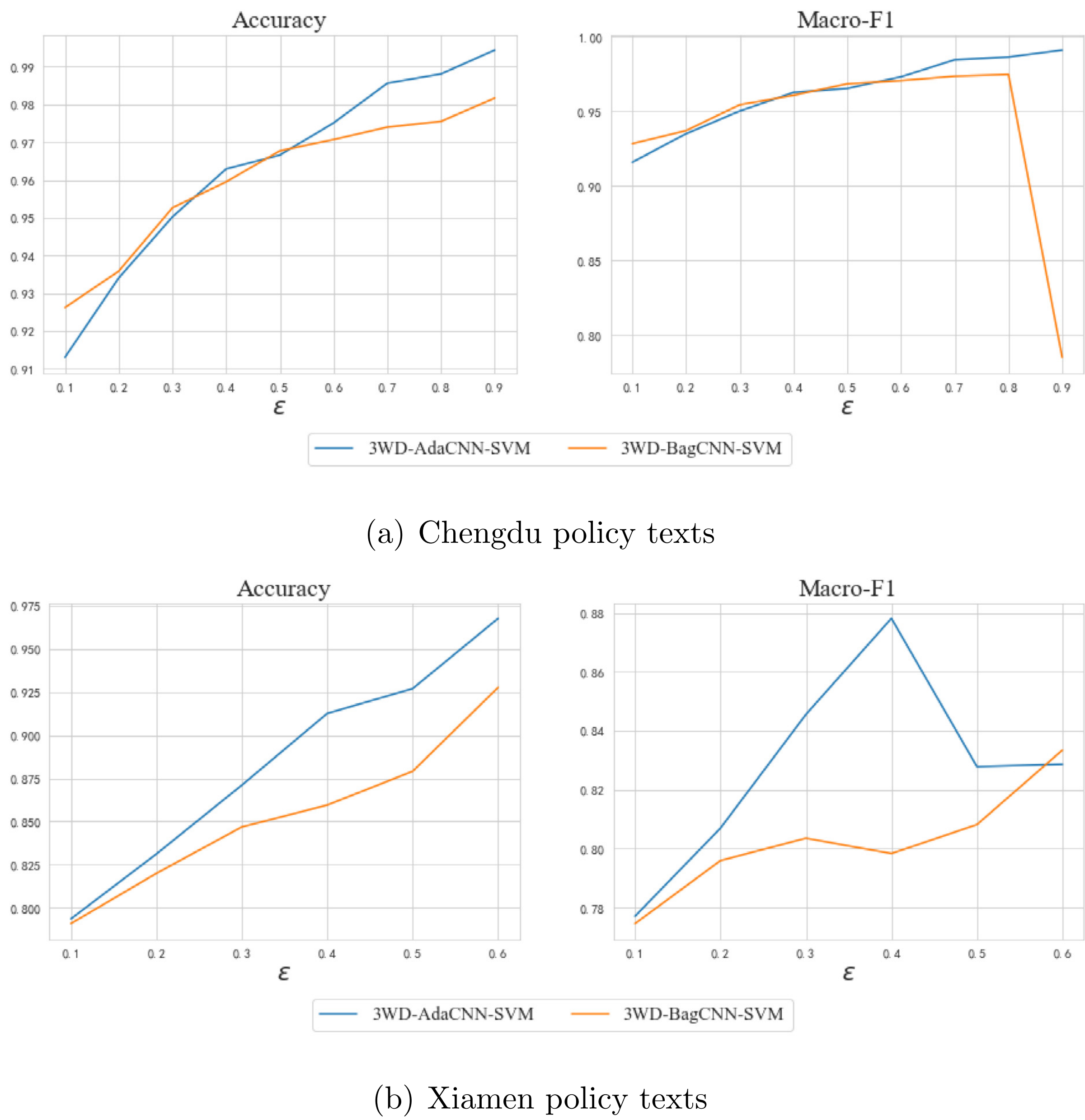
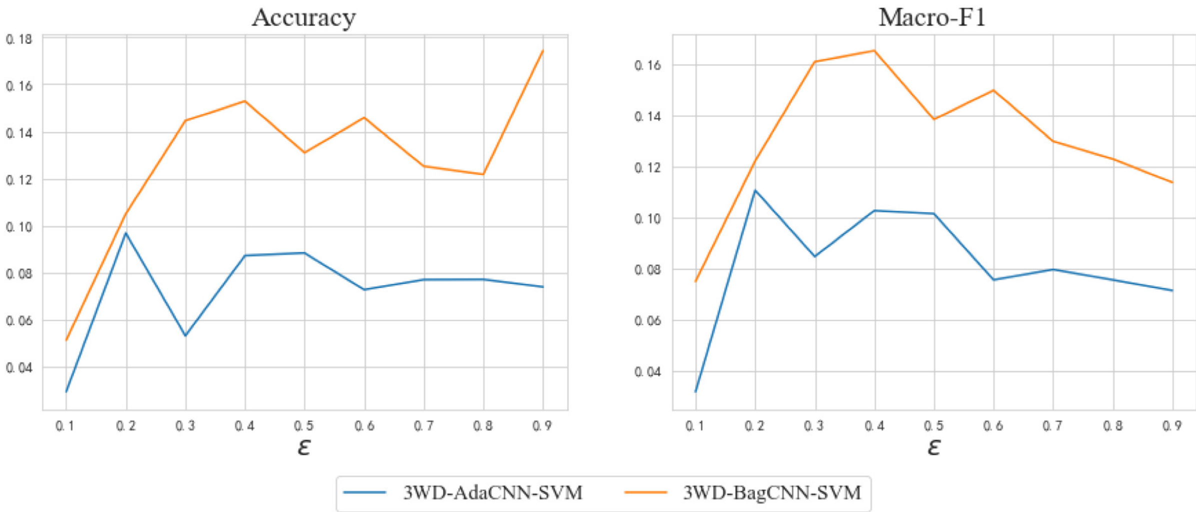


Fig. 10. The changes of the accuracy and Macro-F1 when changing ϵ at the first stage.

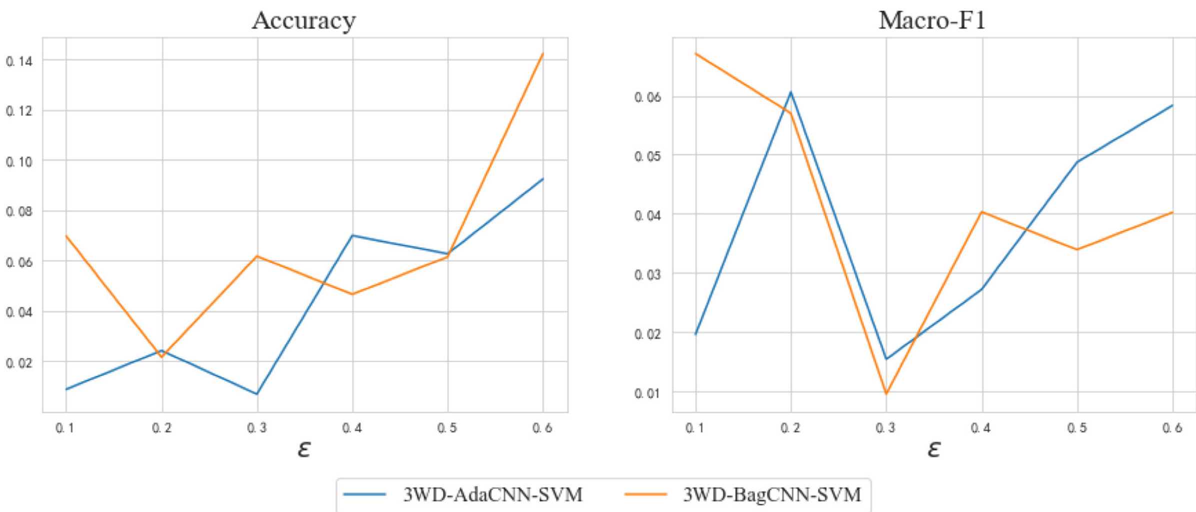
are shown in Table 2. We can see that 3WD-AdaCNN is almost better than 3WD-BagCNN in the two test sets. Besides, if the number of data set is more, then there exists less difference between 3WD-AdaCNN and 3WD-BagCNN methods.

For the second stage, the comparison results are shown in Tables 3 and 4. In this case, CNN method achieves worst performance both in Chengdu and Xiamen policy texts no matter what the ensemble method is used at first stage. Because we have executed the way of feature extraction of CNN in the first stage and filtered the indistinguishable data, if we continue to use CNN model to classify the data, then we can only obtain unsatisfactory results. That is the reason that we choose traditional machine learning method as our secondary classifier of three-way enhanced ensemble learning methods.

In order to explore the influence of the confidence threshold ϵ, we further conduct some experiments. Specifically, we focus on two models, i.e., 3WD-AdaCNN-SVM and 3WD-BagCNN-SVM, because they have better performance in Table 1. When we change the confidence threshold ϵ, the most direct influence is the proportional distribution of the number of test set between two classification stages. If ϵ is smaller, then less samples will be judged at the second stage, see Fig. 8. According to Eq. (12), if ϵ is smaller, then more samples will be classified at the first stage. It means that the classifier of the first



(a) The difference of two-stages in Chengdu policy texts



(b) The difference of two-stages in Xiamen policy texts

Fig. 12. The difference of two-stages when changing ϵ.

stage has greater tolerance. On the contrary, if ϵ is bigger, then more samples will not be classified directly at the first stage and will be conducted the secondary judgment, which means the classifier of the first stage has stricter judgment standards. Besides, for the Bagging and AdaBoost methods, there exist some other differences. Relatively speaking, the confidence threshold ϵ has a greater influence on AdaBoost method both in Chengdu and Xiamen policy texts, which can be reflected in the amplitude of variation of test set rate. The reason is that AdaBoost method is more likely to focus on the misclassified samples.

Then, we further discuss the influence about the classification accuracy and Macro-F1 when we change the confidence threshold ϵ, see Fig. 9.

In Fig. 9, with the incremental increase of the confidence threshold ϵ, the classification accuracy and Macro-F1 can be basically improved. Because the number of Xiamen policy texts is less than that of Chengdu policy texts, the advantage of CNN model can not be taken adequately. Therefore, the overall performance in Xiamen policy texts is not good as that in Chengdu policy texts both in Bagging and AdaBoost methods. Because our proposed methods consist of two stages, we can further discuss the effect on the experiment results when changing the confidence threshold ϵ both in the first stage and the second stage, which are shown in Figs. 10 and 11.

In Fig. 10, we can see that the blue line is almost always above the yellow line at the first stage. Hence, the performance of AdaBoost is better than that of Bagging. At the second stage, in Fig. 11, the two lines intersect each other in the Xiamen policy texts. If ϵ is bigger, then AdaBoost can outperform Bagging sooner or later. However, for the Chengdu policy texts, AdaBoost always outperforms Bagging. In addition, we also compare the difference of the two classification stages, which is shown in Fig. 12.

In Fig. 12, for the Chengdu policy texts, the performance gap between the two stages is less than 0.2 both in 3WD-AdaCNN and 3WD-BagCNN methods and the different capabilities are relatively obvious. For the Xiamen policy texts, even though the capabilities of 3WD-AdaCNN and 3WD-BagCNN methods are fairly close, the performance gap between the two stages is larger when the value of ϵ increases.

5. Conclusions

Considering the decision risk of text misclassification, we develop a two-stage classification framework of inclusive policy text based on 3WD, which can automatically classify these texts. Since Bagging and AdaBoost are two popular ensemble learning methods, we introduce them into our framework and discuss their generalization abilities and stabilities. At the first stage, we adopt CNN model as base classifier. In order to integrate all base classifiers, we propose a new method to determine the weights of base classifiers by considering the accuracy and classification confidence. If the classification confidence is bigger, then the predicted category is more likely to be distinguished precisely. Based on the advantage of deferment decision of 3WD, if the object can not be distinguished accurately at the first stage, then it will be performed secondary classification. With the help of 3WD, we combine deep learning and traditional machine learning methods. Through a series of experiments, we find that the AdaBoost outperforms Bagging in our two-stage classification framework and 3WD-AdaCNN-SVM model is the best combination strategy. Besides, if the number of samples is larger, the performance gap of the classifier of AdaBoost at each stage can be more obvious. In this paper, we just consider the two-stage classification scenario. In the future research work, we will discuss the multistage case in the policy text classification.

CRediT authorship contribution statement

Decui Liang: Methodology, Writing - review & editing. **Bochun Yi:** Writing - original draft, Data curation, Validation, Formal analysis.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is partially supported by the National Natural Science Foundation of China (Nos. 71401026, 71432003, 61773352), the Planning Fund for the Humanities and Social Sciences of Ministry of Education of China (No. 19YJA630042) and the Double First-class Construction Research Support Project of UESTC (No. SYLYJ2019210).

References

[1] A. Abdi, S.M. Shamsuddin, S. Hasan, J. Piran, Deep learning-based sentiment classification of evaluative text based on Multi-feature fusion, Information Processing & Management 56 (4) (2019) 1245–1259.
[2] A. Aldweesh, A. Derhab, A.Z. Emam, Deep learning approaches for anomaly-based intrusion detection systems: A survey, taxonomy, and open issues, Knowledge-Based Systems (2019) 1–19.

[3] D.B. Bracewell, J.J. Yan, F.J. Ren, S. Kuroiwa, Category classification and topic discovery of japanese and english news articles, Electronic Notes in Theoretical Computer Science 225 (2009) 51–65.
[4] R.S.M. Carrasco, M.A. Sicilia, Unsupervised intrusion detection through skip-gram models of network behavior, Computers & Security 78 (2018) 187–197.
[5] M.Y. Chen, Predicting corporate financial distress based on integration of decision tree classification and logistic regression, Expert Systems with Applications 38 (9) (2011) 11261–11272.
[6] P. Chen, L. Yuan, Y. He, S. Luo, An improved SVM classifier based on double chains quantum genetic algorithm and its application in analogue circuit diagnosis, Neurocomputing 211 (2016) 202–211.
[7] G. Ciaparrone, F.L. Snchez, S. Tabik, et al, Deep learning in video multi-object tracking: A survey, Neurocomputing (2019) 1–28.
[8] A. Elnagar, R.A. Debsi, O. Einea, Arabic text classification using deep learning models, Information Processing & Management 57 (1) (2020) 1–17.
[9] B. Galitsky, Machine learning of syntactic parse trees for search and classification of text, Engineering Applications of Artificial Intelligence 26 (3) (2013) 1072–1091.
[10] C. Gao, P. Li, Y. Zhang, et al, People counting based on head detection combining Adaboost and CNN in crowded surveillance environment, Neurocomputing 208 (2016) 108–116.
[11] L. Gao, X. Wang, J. Song, Y. Liu, Fused GRU with semantic-temporal attention for video captioning, Neurocomputing (2019) 1–7.
[12] X.Q. Guan, J.Y. Liang, Y.H. Qian, J.F. Pang, A multi-view OVA model based on decision tree for multi-classification tasks, Knowledge-Based Systems 138 (2017) 208–219.
[13] S. Henry, C. Cuffy, B.T. McInnes, Vector representations of multi-word terms for semantic relatedness, Journal of Biomedical Informatics 77 (2018) 111–119.
[14] D. Hyun, C. Park, M.C. Yang, et al, Target-aware convolutional neural network for target-level sentiment analysis, Information Sciences 491 (2019) 166–178.
[15] F. Jia, P.D. Liu, A novel three-way decision model under multiple-criteria environment, Information Sciences 471 (2019) 29–51.
[16] M. Kang, J. Ahn, K. Lee, Opinion mining using ensemble text hidden Markov models for text classification, Expert Systems with Applications 94 (2018) 218–227.
[17] Y. Kim, Convolutional Neural Networks for Sentence Classification, Eprint Arxiv, (2014) 1–6..
[18] H.J. Kim, J. Kim, J. Kim, P. Lim, Towards perfect text classification with Wikipedia-based semantic Naïve Bayes learning, Neurocomputing 315 (2018) 128–134.
[19] J. Kim, S. Jang, E. Park, S. Choi, Text classification using capsules, Neurocomputing (2019) 1–8.
[20] D. Kim, D. Seo, S. Cho, P. Kang, Multi-co-training for document classification using various document representations: TF?IDF, LDA, and Doc2Vec, Information Sciences 477 (2019) 15–29.
[21] J.H. Li, C.C. Huang, J.J. Qi, et al, Three-way cognitive concept learning via multi-granularity, Information Sciences 378 (2016) 244–263.
[22] Z. Liu, Y. Liu, J. Dezert, Q. Pan, Classification of incomplete data based on belief functions and K-nearest neighbors, Knowledge-Based Systems 89 (2015) 113–125.
[23] Y. Liu, J.W. Bi, Z.P. Fan, Multi-class sentiment classification: The experimental comparisons of feature selection and machine learning algorithms, Expert Systems with Applications 80 (2017) 323–339.
[24] G. Liu, J.B. Guo, Bidirectional LSTM with attention mechanism and convolutional layer for text classification, Neurocomputing 337 (2019) 325–338.
[25] A.S. Khwaja, A. Anpalagan, N. Naeem, B. Venkatesh, Joint bagged-boosted artificial neural networks: Using ensemble machine learning to improve short-term electricity load forecasting, Electric Power Systems Research 179 (2020) 1–7.
[26] T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space, ICLR Workshop, (2013)..
[27] T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, Advances in Neural Information Processing Systems (2013) 3111–3119.
[28] A. Onan, S. Korukoglu, H. Bulut, Ensemble of keyword extraction methods and classifiers in text classification, Expert Systems with Applications 57 (15) (2016) 232–247.
[29] A. Onan, S. Korukoglu, H. Bulut, A hybrid ensemble pruning approach based on consensus clustering and multi-objective evolutionary algorithm for sentiment classification, Information Processing & Management 53 (4) (2017) 814–833.
[30] A. Prashar, Towards sustainable development in industrial small and medium-sized enterprises: An energy sustainability approach, Journal of Cleaner Production 235 (20) (2019) 977–996.
[31] A. Prashar, M. Vijaya Sunder, A bibliometric and content analysis of sustainable development in small and medium-sized enterprises, Journal of Cleaner Production (2019) 1–19.
[32] S.H. Roshan, S. Asadi, Improvement of Bagging performance for classification of imbalanced datasets using evolutionary multi-objective optimization, Engineering Applications of Artificial Intelligence 87 (2020) 1–19.
[33] T. Salles, M. Goncalves, V. Rodrigues, L. Rocha, Improving random forests by neighborhood projection for effective text classification, Information Systems 77 (2018) 1–21.
[34] R.A. Sinoara, J.C. Collados, R.G. Rossi, et al, Knowledge-enhanced document embeddings for text classification 163 (2019) 955–971.
[35] S.K. Srivastava, S.K. Singh, J.S. Suri, Effect of incremental feature enrichment on healthcare text classification system: A machine learning paradigm, Computer Methods and Programs in Biomedicine 172 (2019) 35–51.
[36] W. Sun, Q. Gao, Exploration of energy saving potential in China power industry based on Adaboost back propagation neural network, Journal of Cleaner Production 217 (2019) 257–266.
[37] P. Wang, Q. Liu, X. Yang, et al, Ensemble re-clustering: refinement of hard clustering by three-way strategy, International Conference on Intelligent Science & Big Data Engineering (2017) 1–8.
[38] P. Wang, H. Shi, X. Yang, et al, Three-way k-means: integrating k-means and three-way decision, International Journal of Machine Learning & Cybernetics (2019) 1–11.
[39] C. Xiao, N. Chen, C. Hu, et al, Short and mid-term sea surface temperature prediction using time-series satellite data and LSTM-AdaBoost combination approach, Remote Sensing of Environment 233 (2019) 1–18.
[40] Y.Y. Yao, Decision-theoretic rough set models, Rough Sets and Knowledge Technology (2007) 1–12.
[41] Y.Y. Yao, Three-way decisions with probabilistic rough sets, Information Sciences 180 (3) (2010) 341–353.
[42] Y.Y. Yao, An outline of a theory of three-way decisions, in: Rough Sets and Current Trends in Computing, Springer, Berlin Heidelberg, 2012, pp. 1–18.
[43] Y.Y. Yao, Three-way decision and granular computing, International Journal of Approximate Reasoning 103 (2018) 107–123.
[44] H.R. Zhang, F. Min, B. Shi, Regression-based three-way recommendation, Information Sciences 378 (2017) 444–461.
[45] Y.B. Zhang, Z.F. Zhang, D.Q. Miao, J.Q. Wang, Three-way enhanced convolutional neural networks for sentence-level sentiment classification, Information Sciences 477 (2019) 55–64.
[46] Y.B. Zhang, D.Q. Miao, J.Q. Wang, Z.F. Zhang, A cost-sensitive three-way combination technique for ensemble learning in sentiment classification, International Journal of Approximate Reasoning 105 (2019) 85–97.
[47] Q.H. Zhang, G.H. Pang, G.Y. Wang, A novel sequential three-way decisions model based on penalty function, Knowledge-Based Systems (2019) 1–40.
[48] B.T. Zhong, X.J. Xing, P. Love, et al, Convolutional neural network: Deep learning-based classification of building quality problems, Advanced Engineering Informatics 40 (2019) 46–57.