Research Article

Improved Extension Neural Network and Its Applications

Yu Zhou, 1, 2 Lian Tian, 1 and Linfei Liu 1

1 School of Electric Power, North China University of Water Resources and Electric Power, Zhengzhou 450045, China
2 Key Laboratory of Innovation Method and Decision Management System of Guangdong Province, Guangzhou 510641, China

Correspondence should be addressed to Yu Zhou; zhouyu_beijing@126.com

Received 10 September 2013; Revised 24 November 2013; Accepted 18 December 2013; Published 16 January 2014

Abstract

Extension neural network (ENN) is a new neural network that is a combination of extension theory and artificial neural network (ANN). The learning algorithm of ENN is based on supervised learning algorithm. One of important issues in the field of classification and recognition of ENN is how to achieve the best possible classifier with a small number of labeled training data. Training data selection is an effective approach to solve this issue. In this work, in order to improve the supervised learning performance and expand the engineering application range of ENN, we use a novel data selection method based on shadowed sets to refine the training data set of ENN. Firstly, we use clustering algorithm to label the data and induce shadowed sets. Then, in the framework of shadowed sets, the samples located around each cluster centers (core data) and the borders between clusters (boundary data) are selected as training data. Lastly, we use selected data to train ENN. Compared with traditional ENN, the proposed improved ENN (IENN) has a better performance. Moreover, IENN is independent of the supervised learning algorithms and initial labeled data. Experimental results verify the effectiveness and applicability of our proposed work.

1. Instruction

There are some recognition and classification problems whose features are defined over an interval of values in our world. For example, boys can be defined as a class of men from age 1 to 12, and the permitted operation voltages of a specified motor may be between 100 and 120 V. For these problems, it is difficult to implement an appropriate classification method using the current artificial neural networks (ANNs). Therefore, a new topology of neural network, extension neural network (ENN), is proposed by Wang and Hung [1] to solve these problems. This new neural network is a combination of extension theory [2] and ANN. The ENN uses a modified extension distance (ED) to measure the similarity between the objects and the class centers. It can quickly and stably learn to categorize input patterns and permit the adaptive processes to access significant new information. Moreover, the ENN has shorter learning time and a simpler structure than traditional ANNs. There have been some successful applications based on ENN in the field of pattern recognition, fault diagnosis, classification, cluster analysis, and so on [1, 3–8].

Just like any other ANNs based on supervised learning, the training data set is the most important and critical factor to ensure the supervised learning performance of ENN. In other words, supervised learning works well only when we have perfect training data set. The training data set should be typical and informative, which can represent the environment of applications. On the other hand, for supervised learning, the label of training data is of paramount importance. However, in many real-world applications, the number of labeled data available for training purpose is limited. This is because labeled data are often difficult, expensive, or time consuming to obtain as they require the efforts of experienced human annotators [9]. So the crucial issue in the field of classification and recognition applications of ENN is how to achieve the best possible performance ENN classifier with low quality training data set. Training data selection is one effective method which addresses this issue by selecting the most informative data and labeling the valuable data for training. Training data selection in the supervised learning, often referred to as the data reduction, aims at deciding which training samples from training data set should be retained for further use during the learning process. Data selection can
result in increased capabilities and generalization properties of learning model and shorter time of learning process, reducing space complexity of classification problem and solving imbalanced training data problem [10–14].

In this work, with the purpose of improving the performance of ENN and widening the application range of ENN, we use a novel data selection method based on shadowed sets [15, 16] to refine the training data of ENN. In shadowed sets-based data selection method, core data and boundary data are proposed. Firstly, the given unlabeled training data are partitioned into clusters and the optimal fuzzy matrix of training data is acquired by using clustering algorithm. Then, the corresponding shadowed sets are induced. On the basis of training data and shadowed sets, core data and boundary data could be formed automatically. Finally, the training data set of ENN could be selected from core data and boundary data. Here, the use of clustering algorithm serves two purposes. One is partitioning the given unlabeled training data into clusters and labeling the training samples. The other is producing shadowed sets induced by corresponding fuzzy sets. With our proposed data selection method, when the ENN is trained with selected training data, the performance of the ENN improves significantly as opposed to the ENN that does not receive selected input data for training. In addition, this method is independent of the supervised learning algorithms. In particular, ENN can work well even in the case when there is no labeled data available, which can expand engineering applied scope of ENN.

The rest of this paper is organized as follows. In Section 2, related works are presented. Section 3 introduces the relevant concepts and theoretical basis. Section 4 presents mechanisms and algorithm of our proposed IENN in detail. Section 5 reports on some experiments and discusses some observations. Finally, concluding comments and future research work are offered in Section 6.

2. Related Works

Two factors have direct influence on the supervised learning of ANNs, (1) the network complexity and (2) the training data set [17]. Complexity refers to the number of independent parameters associated with the ANN model and many pruning algorithms are used to tackle this problem, but there is no effective solution for the second problem. In order to enhance the performance of ANNs, many researchers pay more attention to refine and preprocess the original training data to improve the quality of training data. How to improve the quality of training data set is an open problem. So far, to improve the performance of ANNs by using training sample pretreatment technique has achieved some positive results and has become a hot topic. In the following we summarize the related research works from three aspects.

(1) About Data Selection Method. In the supervised learning, it is obvious that some data points in a given set of training data will have a greater effect on the performance of ANNs. We should select these influential data to train ANNs. Selecting training data at random from the sample space is the most popular method followed. Despite its simplicity, this method does not guarantee that the training would be optimal. Sometimes, unwanted effects such as overfitting or information loss by deleting examples are produced from the training sets using random data selection method. To overcome the problems of random data selection method, several new techniques have been introduced for improving the performance of ANNs. In [18], the authors derive a method for selecting exemplars for training multilayer feedforward network and explore the benefits of selective data handling. In [19], the authors propose a methodology to update Learning Vector Quantization (LVQ) prototypes by using a selected subset of the available training data. In [20], the authors introduce a training data selection method based on fuzzy clustering to enhance the supervised learning performance of BP network. In [21], the authors present a pairing method that selects the training data by finding the nearest data of the different classes, and this method can reduce the size of training data and guarantee general ability of the multilayer neural network. In [17], the authors present a data selection that is specific to a difference boosting neural network. A data selection method based on data exchange and data replication is proposed for probabilistic neural network in [22].

(2) About Data Types of Training Data. There also exist some research works about the selection of data types, especially for problems of classification and recognition of ANN based on supervised learning. Different researchers stress on different data types. In [23], the authors verify the importance of boundary data by experiments. They select boundary data as training data set and demonstrate its advantages. In [21], the authors use boundary data and some random data as training data set. Literature [24] puts more emphasis on the representative samples which are close to class centers. In [20] the authors present three methods to select data point that are center-based selection, border-based selection, and hybrid selection, and the authors use experiments to prove that the boundary data and central data used together as training data is better than just using boundary data or central data.

(3) About Dealing with Noise Data. In some situations, training data contains noise data and outliers. In [25], based on the search thought and the Hamming distance, the authors put forward a new approach to eliminate morbid samples that can efficiently eliminate noise data and improve the generalization ability of forward neural network. The network may not converge if there are morbid samples in training data. For solving this problem, an eliminating algorithm of voting that base mode similarity calculating is proposed in [26]. Mislabeled training data is an important source of noise. In [27], the authors offer a new approach which can identify the mislabeled data and handle such data by using a preliminary filtering procedure.

Since performance and reliability of the output of ENN depend on the quality of training data, the sufficient amount of informative samples is essential to learn a good ENN classifier. For ENN, in many real-world engineering applications, we cannot get perfect training data to serve the supervised learning. Most training data include some redundant examples, unlabeled examples, noise data, or less useful
examples. So if we have no effective strategy to select the useful data to achieve good performance of ENN, we have to use the original training data. In this case, the training of ENN may become difficult and inevitably cause performance degradation. Although a variety of data selection methods have been so far proposed, no single approach can be fit for improving the performance of ENN. To solve this problem, in this paper, based on the characteristics of ENN and our past study [28], we use shadowed-sets-based data selection method to refine the training data. The data selection method based on shadowed sets is very suitable for improving the performance of ENN. With this method, the performance of ENN is improved. For improved ENN (IENN), there is no any requirement on the size of initial labeled samples. In particular, the best case to use IENN is when there are no (or very small number) prelabeled samples available so that traditional ENN cannot be used. Hence, IENN relaxes the restricted conditions of ENN in engineering application. Compared to traditional ENN, our proposed method reduces the computation consumption, saves training time, and guarantees generalization ability.

3. Theoretical Background

3.1. Outline of ENN. ENN is a new topology of neural network and the structure of ENN is depicted in Figure 1. There are two connection weights between input neuron nodes and output neuron nodes; one connection weight represents the up bound (maximum value) for the classical domain of the feature and the other connection weight represents the low bound (minimum value) for the required domain of the feature. The connection weights between jth input node and kth output node are \( w_{kj}^{\text{L}} \) and \( w_{kj}^{\text{U}} \), respectively. ENN use \( ED \) (as shown in (1)) to measure similarities between tested data and class centers. \( ED \) can describe the distance between a point \( x \) and an interval \( \langle w^L, w^U \rangle \) as shown in Figure 2.

The learning of the ENN can be seen as supervised learning. The major steps of learning algorithm can be described as follows:

\[
ED = \frac{|x - (w^{\text{U}} + w^{\text{L}})/2| - (w^{\text{U}} - w^{\text{L}})/2|}{(w^{\text{U}} - w^{\text{L}})/2} + 1. \tag{1}
\]

Let training data set be \( X \equiv \{X_1, X_2, \ldots, X_N\} \), where \( N_p \) is the size of training data set. The ith instance in training data set is \( X^i = \{x_{i1}^{\text{L}}, x_{i2}^{\text{L}}, \ldots, x_{in}^{\text{L}}\} \), where \( n \) is the total number of the feature of instances, and the label of the ith instance is \( p \). To evaluate the ENN performance, the error rate function \( E_T \) is defined below:

\[
E_T = \frac{N_m}{N_p}, \tag{2}
\]

where \( N_m \) is the total error number.

First, set the initial connection weights according to the range of classical domains and calculate the initial class centers of each class. Then input training data set. Suppose the ith \( (i = 1, 2, \ldots, N_p) \) training instance is read. Use the \( ED \) to calculate the distance between the training instance \( X^p_i \) and the kth class as follows:

\[
ED_k = \sum_{j=1}^{n} \frac{\left[|x^p_{ij} - z_{kj}| - (w_{kj}^{\text{U}} - w_{kj}^{\text{L}})/2\right]}{\left|(w_{kj}^{\text{U}} - w_{kj}^{\text{L}})/2\right|} + 1, \tag{3}
\]

where \( z_{kj} = (w_{kj}^{\text{U}} + w_{kj}^{\text{L}})/2 \).

Find the \( k^* \), such that \( ED_{k^*} = \text{Min}\{ED_k\} \). If \( k^* = p \), then input the next training instance and calculate the \( ED \); otherwise, update the centers and weights of the \( p \)th and the \( k^* \)th classes as (4)-(5). Repeat the above steps until training process has converged or the total error rate \( E_T \) has arrived at a preset value.

(a) Update the centers of the \( p \)th and the \( k^* \)th classes:

\[
\begin{align*}
z_{\text{new}}^{\text{old}} &= z_{\text{old}}^{\text{old}} + \eta \left(x^p_{ij} - z_{\text{old}}^{\text{old}}\right), \\
z_{k^*j}^{\text{new}} &= z_{k^*j}^{\text{old}} - \eta \left(x^p_{ij} - z_{k^*j}^{\text{old}}\right),
\end{align*}
\tag{4}
\]
Mathematical Problems in Engineering

3.2. Summary of Shadowed Sets Theory. In this section, we briefly recall the concepts of shadowed sets [15, 16]. Conventional uncertainty models like fuzzy sets tend to capture vagueness exclusively through membership values (MVs). This poses a dilemma of excessive precision in describing imprecise phenomenon. The notion of shadowed sets tries to solve this problem of selecting the optimum level of resolution in precision. The motivation behind the debate on excessive precision of fuzzy sets is the conceptual shortcoming associated with precise numeric values of membership used to describe vague concepts. While there is hardly any difficulty in assigning membership grades close to 1 or 0, a lot of uncertainty prevails during the assignment of membership grade of 0.5. Based on this central idea, Pedrycz developed the concept of shadowed sets to improve the observability and interpretability of vague phenomenon.

Consider a fuzzy set as depicted in Figure 3. We attempt to modulate the membership values on the lines of three-valued logic by elevating and reducing some MVs and balancing the uncertainty thus introduced. The elevation, reduction, and balance of uncertainty are quite radical. We try to disambiguate the concept represented by the original fuzzy set by promoting a few of the MVs to one and reducing a few other MVs to zero. Such enhancement of contrast reduces the number of computations as compared to the fuzzy framework. In order to maintain the overall level of vagueness, some other region is defined as the zone of uncertainty. Provision is made so that this particular area of the universe of discourse has intermediate membership values on a unit interval of [0, 1] but left undefined. Rather than a single value, the entire unit interval can be marked as a nonnumeric model of membership grade. Note that to induce a shadowed set, a fuzzy set must accept a specific threshold level. Effectively, this transforms the domain of discourse into clearly marked zones of vagueness. This mapping is called a shadowed set and is defined as

$$A : X \rightarrow [0, [0, 1], 1]$$,

where $X$ is a given universe of discourse.

The codomain of $A$ consists of three components, that is, 0, 1, and the unit interval [0, 1]. They can be treated as degrees of membership of elements to $A$. These three quantification levels come with an apparent interpretation. All elements for which $A(x)$ assumes 1 (denoted by core($A$)) are called a core of the shadowed set—they embrace all elements that are fully compatible with the concept conveyed by $A$. The elements of $X$ for which $A(x)$ attains zero (denoted by exclusion($A$)) are excluded from $A$. The elements of $X$ for which we have assigned the unit interval (denoted by shadow($A$)) are completely uncertain—we are not at position to allocate any numeric membership grade. Therefore, we allow the usage of the unit interval, which reflects uncertainty meaning that any numeric value could be permitted here. In essence, such element could be excluded (we pick up the lowest possible value from the unit interval), could exhibit partial membership (any number within the range from 0 to 1), or could be fully allocated to $A$. Given this extreme level of uncertainty (nothing is known and all values are allowed), we call these elements shadows and hence the name of the shadowed set. An illustration of a shadowed set is included in Figure 3.

One of the most appealing questions concerns the computation of the threshold $\alpha$. An optimization based on balance of vagueness is proposed in [15, 16]. Referring to Figure 4, reduction of some MVs to 0 and elevation to 1 should be compensated by marked indecision in the other zones or increased uncertainty in MVs in the form of a unit interval [0, 1] over particular ranges of $A$. A particular threshold $\alpha$ is selected for the quantification process and is expressed in terms of the relationship

$$V(\alpha) = \left| \int_{-\infty}^{a_1} A(x) \, dx + \int_{a_1}^{+\infty} (1 - A(x)) \, dx - \int_{a_1}^{a_1} A(x) \, dx \right|,$$  

where $a_1$ is a given threshold for the quantification process.

Figure 3: The fuzzy set inducing a shadowed set via a threshold.
that is,
\[ \Omega_1 + \Omega_2 = \Omega_3, \]  
(8)
where \( A \) is the fuzzy set under consideration. In other words, the threshold \( \alpha \) located in \([0, 1/2]\), if selected properly, should zero \((7), V(\alpha) = 0\). The optimization task comes in the form
\[ \alpha_{\text{opt}} = \arg \min \alpha \ V(\alpha). \]  
(9)

The three terms on the right hand side of (7) correspond to regions \( \Omega_1, \Omega_2, \) and \( \Omega_3 \) in Figure 4. The parameters \( a_1 \) and \( a_2 \) denote the boundaries in the integral, delineating the regions in the figure where the membership values are below the threshold \( \alpha \) and above the threshold \( 1 - \alpha \).

In the discrete domain, (7) gets transformed to
\[ V(\alpha_i) = \sum_{X_k \in \text{shadow}_i \cap \omega_{\text{core}}} u_{i_k} + \sum_{X_k \in \text{shadow}_i \cap \omega_{\text{core}}} (u_{i_{\text{max}}} - \alpha_i) \]
\[ -\text{card} \{ X_k \ | \ \alpha_i < u_{i_k} < u_{i_{\text{max}}} - \alpha_i \}, \]
(10)
such that
\[ \alpha_i = \alpha_{\text{opt}} = \arg \min \alpha \ V(\alpha), \]  
(11)
where \( u_{i_k}, u_{i_{\text{min}}} \) and \( u_{i_{\text{max}}} \) denote, respectively, the discrete, the lowest, and the highest MVs to the \( i \)th class.

4. Our Proposed Improved Neural Network (IENN) Mechanisms

4.1. Mechanisms of Data Selection Method Based on Shadowed Sets. Based on past studies and the characteristics of ENN, we know that two types of data points are very important as training data in the field of classification problems: the data points close to class centers and the data points close to borders between classes. Hence, we select these two types of data as training data to train ENN. In the framework of shadowed sets, we establish two important data types—core data and boundary data.

**Definition 1** (core data). Assume that the number of classes is equal to \( c \). Core data is defined as
\[ \text{core data} = \{ x \ | \ \exists x \in \text{core}(A_i) \}, \quad i = 1, 2, \ldots, c. \]  
(12)

Core data are located in the central part of each class. The data points in core data are usually representative and valuable for supervised learning of ENN.

**Definition 2** (boundary data). Assume that the number of classes is equal to \( c \). Boundary data is defined as

boundary data
\[ = \{ x \ | \ \exists x \in \text{shadow}(A_i) \cap x \in \text{shadow}(A_j) \}
\]
\[ \cup \cdots \cup \{ x \ | \ \exists x \in \text{shadow}(A_i) \cap x \in \text{shadow}(A_j) \}
\]
\[ \cap x \in \text{shadow}(A_k) \cup \cdots \cup \{ x \ | \ \exists x \in \text{shadow}(A_i) \cap x \in \text{shadow}(A_k) \cap \cdots \cup \cdots \}. \]
(13)

Boundary data consist of data points around the borders between two or more classes. Boundary data are quite likely to lie near the decision boundaries of classes; they can be regarded as “confusing samples,” which require more attention in classification problems. Sketch map of core data and boundary data is shown in Figure 5.

An illustration of core data and boundary data is included in Figure 6 (the number of classes is 4). An artificial data set is shown in Figure 6(a). Figure 6(b) is the clustering result of the artificial data set after running Fuzzy C-Means clustering algorithm (FCM). The core data and boundary data are clearly shown in Figure 6(c). From Figure 6 we can find that core data and boundary data form an obvious hierarchy of data set. Core data can provide informative and representative samples for the updating of class centers of ENN in the process of supervised learning. Boundary data can guarantee the generalization ability of ENN. Moreover, in the process of training, boundary data are helpful for the updating of up weights and low weights.

Figure 7 shows the data selection model for improving the supervised learning performance of ENN. The critical step of data selection is the establishment of core data and boundary data. The prerequisite of the establishment of core data and boundary data is to induce shadowed sets from raw data set.

The heuristic of our proposed data selection method is similar to some existing works. In [29], the authors choose the valuable training samples which are closest to the current
classification boundary. The intuitive ideas of closest-to-boundary criterion and our boundary data selection are similar. The difference is that our boundary data selection tries to find those samples using clustering information instead of classification. Actually, using clustering in data selection is not new and several works have been done [20, 30]. Different to them, the use of clustering algorithm plays two important roles in shadowed-set-based data selection method. One is labeling the training samples. Hence, our proposed data selection method is independent of the supervised learning algorithms and the initial labeled data for training purpose. The other is inducing the shadowed sets. Some other existing works [24, 31] put more emphasis on the representative samples, which are basic idea of our core data selection. For some classifiers such as support vector machine (SVM) and radial basis function (RBF) neural network, just using core data to train classifier model could get better performance. However, for ENN, these data selection methods have no significant effects for improving performance. This is because the boundary data can express the maximum and minimum values of the interval, which are very meaningful for the training of ENN.

Since our method uses clustering algorithm to label the data, there is no requirement on the size of initial labeled samples. Hence, the best case to use our method is when there are no labeled samples available so that traditional supervised-based selection methods cannot be used. It is worth noting that the capture of selected data with our proposed method is fully automated by using optimal threshold of shadowed sets. We can select training data only from core data and boundary data batch by batch, which can increase the efficiency of data selection. Moreover, our method reduces the computational complexity by selecting a batch of unlabeled sample instead of one sample. Note that, for existing works [31, 32], even a batch of samples are selected at each iteration, still more computational effort is needed compared with ours. The reason is that the existing works are based on classifiers which requires training process, while our method is based on clustering which does not require training process [20]. At the same time, our proposed data selection method is specially fit for improving the performance of ENN. The main reason is that the endpoint values and the central values of classical domain interval of the features are needed to update the training of ENN, and core data and boundary data can just provide useful information to lead the learning of ENN. Another reason is that due to application of shadowed set, the uncertainty data and less informative data can be effectively eliminated. And the implement of core data and boundary data makes the selected sample data highly reliable.

4.2. The Algorithm of IENN. The detailed algorithm of IENN is described below.

**Step 1.** Run clustering algorithm and obtain optimal fuzzy partition matrix $U = [u_{ik}]$ of available training data set, $i = 1, 2, \ldots, c$, $k = 1, 2, \ldots, N$. Here $c$ stands for the number of classes; $N$ is the size of data set.

**Step 2.** From fuzzy partition matrix $U = [u_{ik}]$, we can get $c$ fuzzy sets.
Step 3. In terms of (10)-(11), compute optimal thresholds $\alpha_i$, $i = 1, 2, \ldots, c$.

Step 4. Obtain $c$ shadowed sets induced by $c$ fuzzy sets and $c$ thresholds.

Step 5. Convert $U = [u_{ik}]$ into corresponding shadowed matrix $U' = [u'_{ik}], u'_{ik} \rightarrow \{0, 1, [0, 1]\}$.

Step 6. According to $U'$ and Definitions I and 2, establish the core data and boundary data.

Step 7. Select data. The basic principle of selection is choosing the data closer to centers and boundaries. For this purpose, two control parameters $\beta_1$ (between 0 and 1) and $\beta_2$ (between 1 and 2) are introduced. These two parameters can control the size of selection data. Using control parameters, shadowed partition matrix can be updated. It can be expressed in (14).

$$\text{If } u_{ik} > 1 - \beta_1 \ast \alpha_i, \text{ then } u'_{ik} = 1;$$
$$\text{if } u_{ik} < \beta_2 \ast \alpha_i, \text{ then } u'_{ik} = 0;$$
$$\text{if } \beta_2 \ast \alpha_i \leq u_{ik} \leq 1 - \beta_2 \ast \alpha_i, \text{ then } u'_{ik} = [0, 1].$$ (14)

Step 8. Use selected data as training data set to train ENN.

Step 9. Set up the structure of the ENN.

Step 10. From training data set, build matter-models of each class and set the connection weights between input nodes and output nodes:

$$R_k = \begin{bmatrix} N_k & c_1, V_{k1} \\ c_2, V_{k2} \\ \vdots \\ c_n, V_{kn} \end{bmatrix}, \quad k = 1, 2, \ldots, n. \quad (15)$$

In the extension theory, $c_j$ is the $j$th feature of $N_k$ and $V^L_{kj} = \langle w^L_{kj}, w^U_{kj} \rangle$ are the classical domains of the $k$th class ($N_k$) about the $j$-th feature $c_j$. The interval of classical domains can be directly obtained from previous requirement or determined from training data as follows:

$$w^L_{kj} = \text{Min}_{x_{ij} \in N_k} \{ x_{ij} \},$$
$$w^U_{kj} = \text{Max}_{x_{ij} \in N_k} \{ x_{ij} \}. \quad (16)$$

Step 11. Calculate the initial center of each class:

$$Z_k = \{ z_{k1}, z_{k2}, \ldots, z_{kn} \},$$
$$z_{kj} = \frac{(w^L_{kj} + w^U_{kj})}{2}, \quad k = 1, 2, \ldots, n, \ j = 1, 2, \ldots, n. \quad (17)$$

Step 12. Read the $i$th training instance and its label $p$:

$$X^p_i = \{ x_{i1}^p, x_{i2}^p, \ldots, x_{in}^p \}, \quad p \in n. \quad (18)$$

Step 13. Using (3) calculate $ED$ between the training instance $X^p_i$ and the $k$th class.
Figure 8: The two-dimensional synthetic dataset and its core data points and boundary data points.
**Step 14.** Find $k^*$, such that $ED_{k^*} = \text{Min}\{ED_k\}$. If $k^* = p$ then go to Step 16; otherwise, go to Step 15.

**Step 15.** Update the weights of the $p$th and the $k^*$th classes according to (4)-(5).

**Step 16.** Repeat Steps 12 to 15; if all instances have been classified, then a learning epoch is finished.

**Step 17.** Stop, if the learning process has converged or the total error rate $E_T$ has arrived at a preset value; otherwise, return to Step 12.

**Step 18.** Use the IENN to recognize and classify the object pattern.

### 5. Experiment Study

#### 5.1. Illustrative Experiments Using Synthetic Data Set.

As shown in Figure 8(a), the synthetic data set is two-dimensional and includes 60 data points. The number of classes of this data set is 3. All data points have no label and we use clustering algorithm (FCM) to label these data points for the supervised learning of ENN. The clustering results are described in Figure 8(b). Another important role of clustering algorithm is the production of shadowed sets. Here 3 experiments are conducted and the testing data set of all experiments is the full synthetic data set (60 instances).

In experiment 1, for ENN, we use all data set to train the ENN. In experiment 2, for IENN, we use selected data to train ENN. In experiment 3, the same number of training data as experiment 2 is selected at random to train ENN. Compute optimal thresholds, $\alpha_1 = 0.1258$, $\alpha_2 = 0.1051$, and $\alpha_3 = 0.1471$, respectively. Using control parameters control the size of selected data. The control parameter $\beta_2$ of all experiments in this paper is equal to 1. When $\beta_1 = 1$, we can get 35 core data points and 14 boundary data points, when $\beta_1 = 0.5$, 21 core data points and 14 boundary data points can be captured, and when $\beta_1 = 0.3333$, we can get 14 core data points and 14 boundary data points, as shown in Figures 8(c), 8(d), and 8(e), respectively.

We chose $\beta_1 = 0.3333$ to conduct experiments. The learning curves of ENN and IENN are shown in Figure 9, and Table I exhibits the results of comparative experiments. From Figure 9 and Table I, we can get the following conclusions. (1) Comparing with traditional random data selection method, our proposed data selection is more specifically suited to refine the training data of ENN. (2) Comparing with traditional ENN, the improved ENN (IENN) not only can keep the classification accuracy and short learning time, but also can guarantee generalization ability. By the abovementioned research experiments, we can prove that the shadowed-sets-based selection method to refine the training data of ENN is effective and feasible, and the comprehensive performance of traditional ENN is enhanced.

#### 5.2. Simulation Experiments Using Iris Data Set.

In this section, we will be concerned with the Iris data set. There
are 150 instances in the Iris data set. The Iris data set can be divided into three categories with distinguishing variables being the length and width of sepal and petal. To verify the performance of the IENN, three test cases are given in the following.

**Case 1.** Consider that if the system randomly chooses 75 instances from the Iris data as the training data set, and let the rest of the Iris data be the testing data. For ENN, training data set is used directly to train ENN. For IENN, use selected data to train ENN. According to (10)-(11) we can compute optimal thresholds $\alpha_1 = 0.2612$, $\alpha_2 = 0.0415$, and $\alpha_3 = 0.2006$, as shown in Figure 10. Table 2 shows the comparison results of IENN with traditional ENN and other typical ANNs. From Table 2, it is clearly found that ENN performs better than other typical ANNs, and the performance of ENN is further improved with the using of data selection method based on shadowed sets. We can see that the IENN has better generalization ability; specifically, less training data can get higher accurate rate. It should be noted that ENN and IENN have the same learning speed (9 epochs), but the IENN has less training data, so the total learning time of IENN is more economical.

**Case 2.** Consider that if the training data set contains 150 training instances (i.e., the full Iris data) and the testing data set is equal to the training data set. Table 3 compares the performance of ENN and IENN. It can be seen from Table 3 that the generalization ability of IENN is improved. In another word, the proposed method can keep the generalization ability of network using less training data. Figure 11 shows the learning curves of the ENN and IENN.

**Case 3.** To objectively compare the performance of ENN and IENN, 10-fold cross validation experiment is provided in Case 3. The Iris data set is divided into 10 subsets. Each
Table 2: Comparison of the classification performance of various neural networks.

<table>
<thead>
<tr>
<th>ANN model</th>
<th>Structure</th>
<th>No. of training data</th>
<th>No. of connections</th>
<th>Learning speed (epochs)</th>
<th>Training error</th>
<th>Testing error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceptron</td>
<td>4-3</td>
<td>75</td>
<td>12</td>
<td>200</td>
<td>0.173</td>
<td>0.213</td>
</tr>
<tr>
<td>MLP</td>
<td>4-4-3-3</td>
<td>75</td>
<td>37</td>
<td>50</td>
<td>0.027</td>
<td>0.040</td>
</tr>
<tr>
<td>PNN</td>
<td>4-75-3</td>
<td>75</td>
<td>525</td>
<td>1</td>
<td>0.0</td>
<td>0.053</td>
</tr>
<tr>
<td>CPN</td>
<td>4-20-3</td>
<td>75</td>
<td>140</td>
<td>60</td>
<td>0.107</td>
<td>0.160</td>
</tr>
<tr>
<td>ENN</td>
<td>4-3</td>
<td>75</td>
<td>24</td>
<td>9</td>
<td>0.0</td>
<td>0.040</td>
</tr>
<tr>
<td>IENN</td>
<td>4-3</td>
<td>67</td>
<td>24</td>
<td>9</td>
<td>0.0</td>
<td>0.027</td>
</tr>
</tbody>
</table>


Table 3: Comparison of the classification performance of ENN and IENN.

<table>
<thead>
<tr>
<th>Model</th>
<th>No. of training data</th>
<th>Learning speed (epochs)</th>
<th>Training error</th>
<th>Testing error</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENN</td>
<td>150</td>
<td>4</td>
<td>0.0267</td>
<td>97.33%</td>
</tr>
<tr>
<td>IENN</td>
<td>132</td>
<td>2</td>
<td>0.0</td>
<td>97.33%</td>
</tr>
</tbody>
</table>

Figure 11: The learning curves of Iris data of proposed IENN.

Table 4: Experiment results of ENN and IENN of 10-fold cross validation using Iris data set.

<table>
<thead>
<tr>
<th>Classifiers</th>
<th>ENN</th>
<th>IENN</th>
</tr>
</thead>
<tbody>
<tr>
<td>The average number of training data</td>
<td>135</td>
<td>120.2</td>
</tr>
<tr>
<td>Average learning speed (epochs)</td>
<td>13.1</td>
<td>7.7</td>
</tr>
<tr>
<td>Average training error</td>
<td>0.0222</td>
<td>0.0185</td>
</tr>
<tr>
<td>Average accuracy of test training</td>
<td>96.67%</td>
<td>97.33%</td>
</tr>
</tbody>
</table>

5.3. A Practical Engineering Application. A real-world engineering application is taken as an example to illustrate the applicability of IENN. The background of the engineering is the classification and recognition of mine gas emission of the working face. Based on the degree of influence of various factors upon the gas emission of the working face in coal mines and agreement with past study [33], the following five main factors are chosen as the features for the classification and recognition of gas emission. These features are the burial depth of coal seam (m), coal seam thickness (m), coal seam gas content (m$^3$·d$^{-1}$), average daily schedule of working face (m⋅d$^{-1}$), and average daily production (t⋅d$^{-1}$). The measured data set comes from literature [33] (as listed in Table 5). In this application, the measured data are unlabeled, so the traditional ENN cannot be implemented directly. However, IENN is most appropriate for this type of application. Firstly, we run unascertained clustering algorithm and get the optimal fuzzy matrix. Table 6 shows the membership degrees of every sample. Then, from Table 6 we can get 3 fuzzy sets. Compute optimal thresholds, $\alpha_1 = 0.142$, $\alpha_2 = 0.145$, and $\alpha_3 = 0.044$, respectively. Based on the fuzzy sets and optimal thresholds, 3 shadowed sets can be induced. The shadowed matrix $U' = [u'_{ik}], u'_{ik} \rightarrow \{0, 1, [0, 1]\}$ is listed in Table 7.

From Table 7, we can select 7 core data points and 2 boundary data points. Use selected data to train ENN. The result is that the training time of IENN is quite economical, that is, only 1 epoch. At the same time, the training error of IENN is 0. Using all data (15 instances) test the trained IENN, and the test accuracy is 100%. This application’s results show that IENN is feasible and practical. From this application, we can find that our proposed work in this paper could expand the engineering application range of ENN.

6. Conclusions and Future Work

To achieve the best possible supervised learning performance of ENN, in this paper, we use a novel training data selection.
Table 5: The measured data of gas emission.

<table>
<thead>
<tr>
<th>No. of sample</th>
<th>The burial depth of coal seam (m)</th>
<th>Coal seam thickness (m)</th>
<th>Coal seam gas content (m$^3$ ⋅ d$^{-1}$)</th>
<th>Average daily schedule of working face (m ⋅ d$^{-1}$)</th>
<th>Average daily production (t ⋅ d$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>420</td>
<td>1.8</td>
<td>2.14</td>
<td>4.13</td>
<td>1751</td>
</tr>
<tr>
<td>2</td>
<td>408</td>
<td>2</td>
<td>1.92</td>
<td>4.42</td>
<td>1825</td>
</tr>
<tr>
<td>3</td>
<td>411</td>
<td>2.0</td>
<td>2.15</td>
<td>4.16</td>
<td>1527</td>
</tr>
<tr>
<td>4</td>
<td>456</td>
<td>2.2</td>
<td>2.40</td>
<td>4.51</td>
<td>2104</td>
</tr>
<tr>
<td>5</td>
<td>432</td>
<td>2.3</td>
<td>2.58</td>
<td>4.67</td>
<td>2078</td>
</tr>
<tr>
<td>6</td>
<td>531</td>
<td>2.9</td>
<td>3.35</td>
<td>3.68</td>
<td>2288</td>
</tr>
<tr>
<td>7</td>
<td>550</td>
<td>2.9</td>
<td>3.61</td>
<td>4.02</td>
<td>2325</td>
</tr>
<tr>
<td>8</td>
<td>527</td>
<td>2.5</td>
<td>2.80</td>
<td>3.28</td>
<td>1979</td>
</tr>
<tr>
<td>9</td>
<td>563</td>
<td>3.0</td>
<td>3.68</td>
<td>3.53</td>
<td>2410</td>
</tr>
<tr>
<td>10</td>
<td>516</td>
<td>2.8</td>
<td>3.22</td>
<td>3.45</td>
<td>2242</td>
</tr>
<tr>
<td>11</td>
<td>604</td>
<td>6.2</td>
<td>4.03</td>
<td>2.64</td>
<td>3354</td>
</tr>
<tr>
<td>12</td>
<td>634</td>
<td>6.5</td>
<td>4.80</td>
<td>2.92</td>
<td>3620</td>
</tr>
<tr>
<td>13</td>
<td>590</td>
<td>5.9</td>
<td>4.21</td>
<td>2.85</td>
<td>3139</td>
</tr>
<tr>
<td>14</td>
<td>607</td>
<td>6.1</td>
<td>4.34</td>
<td>2.77</td>
<td>3087</td>
</tr>
<tr>
<td>15</td>
<td>640</td>
<td>6.3</td>
<td>4.67</td>
<td>2.75</td>
<td>3412</td>
</tr>
</tbody>
</table>

Table 6: The membership degrees of the samples.

<table>
<thead>
<tr>
<th>No. of sample</th>
<th>Class 1</th>
<th>Class 2</th>
<th>Class 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.879</td>
<td>0.104</td>
<td>0.017</td>
</tr>
<tr>
<td>2</td>
<td>0.853</td>
<td>0.123</td>
<td>0.024</td>
</tr>
<tr>
<td>3</td>
<td>0.845</td>
<td>0.132</td>
<td>0.023</td>
</tr>
<tr>
<td>4</td>
<td>0.91</td>
<td>0.079</td>
<td>0.011</td>
</tr>
<tr>
<td>5</td>
<td>0.858</td>
<td>0.122</td>
<td>0.02</td>
</tr>
<tr>
<td>6</td>
<td>0.013</td>
<td>0.984</td>
<td>0.004</td>
</tr>
<tr>
<td>7</td>
<td>0.129</td>
<td>0.829</td>
<td>0.042</td>
</tr>
<tr>
<td>8</td>
<td>0.197</td>
<td>0.76</td>
<td>0.042</td>
</tr>
<tr>
<td>9</td>
<td>0.1</td>
<td>0.835</td>
<td>0.066</td>
</tr>
<tr>
<td>10</td>
<td>0.046</td>
<td>0.943</td>
<td>0.012</td>
</tr>
<tr>
<td>11</td>
<td>0.012</td>
<td>0.002</td>
<td>0.966</td>
</tr>
<tr>
<td>12</td>
<td>0.03</td>
<td>0.052</td>
<td>0.918</td>
</tr>
<tr>
<td>13</td>
<td>0.002</td>
<td>0.005</td>
<td>0.993</td>
</tr>
<tr>
<td>14</td>
<td>0.004</td>
<td>0.008</td>
<td>0.988</td>
</tr>
<tr>
<td>15</td>
<td>0.02</td>
<td>0.036</td>
<td>0.944</td>
</tr>
</tbody>
</table>

Table 7: Shadowed sets and corresponding partition matrix.

<table>
<thead>
<tr>
<th>No. of sample</th>
<th>Class 1</th>
<th>Class 2</th>
<th>Class 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>[0, 1]</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>[0, 1]</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>[0, 1]</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>[0, 1]</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>[0, 1]</td>
<td>[0, 1]</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>[0, 1]</td>
<td>[0, 1]</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
<td>[0, 1]</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>0</td>
<td>[0, 1]</td>
</tr>
</tbody>
</table>

Method based on shadowed sets to refine the training data set. According to the characteristics of ENN and related research works, core data and boundary data are selected from original training data set for training ENN. Core data are usually representative and valuable for supervised learning and boundary data are important data to achieve generalization ability of ENN. Compared with traditional ENN, our method does not require much computational effort. Moreover, IENN is independent with respect to the supervised learning algorithms and initial labeled data. Experimental results on synthetic data set, benchmark data set, and practical industry data set indicate that using selected data to train ENN can improve effectively supervised learning performance.

Generally, a lot of research works on classification problem of ANN consider all training data points equally significant. In this case, the importance of each training data point is the same. In this work, the data points of core data and boundary data are simply selected without considering the weight of each data point. There is a possibility that the performance is improved by differentiating the importance of each training data point. Therefore, it is interesting to see whether the weighted training data could further enhance the performance of ENN.
Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors gratefully acknowledge the project supported by the Research Foundation of Education Bureau of Henan Province, China, under Grant no. 14B120002, the project supported by the Scientific Research Foundation of Henan Province for Returned Chinese Scholars, China, and the project supported by the Open Project Foundation of Key Laboratory of Innovation Method and Decision Management System of Guangdong Province under Grant no. 2011A060901001-I2D. The authors are grateful for the anonymous reviewers who made constructive comments.

References


