

Incorporating Prior Knowledge into Extension Neural Network and its Application to Recognition of Safety Status Pattern of Coal Mines

Yu Zhou^{1, 2*} and Lian Tan²

^{1*}Key Laboratory of Innovation Method and Decision Management System of Guangdong Province, Guangzhou Guangdong, 510641, China,

¹zhouyu_beijing@126.com

²School of Electric Power, North China University of Water Resources and Electric Power, Zhengzhou 450011, China

Abstract

Incorporating prior knowledge (PK) into learning methods is an effective means to improve learning performance. On the bases of requirements of engineering practice and the characteristics of knowledge representation of extension neural network (ENN), with the purpose of further improving the performance of ENN in engineering practice, a prior-knowledge-based ENN (PKENN) recognition method is proposed and applied in the application of safety status pattern recognition of coal mines in this paper. The PKENN recognition method effectively combines domain knowledge with training data set. The prior knowledge can provide additional information about the classical domain of characteristic vector that may compensate for the low quality of training data in a complex application environment. This method can set the initial weights of ENN, guide the learning of ENN and alleviate the learning burden. To demonstrate the validity and effectiveness of the proposed method, a real-world application on the geological safety status pattern recognition of coal mines is tested. Comparative experiments with existing methods and other ANN-based methods are conducted. The experimental results show that the proposed PKENN recognition method has a better performance.

Keywords: Extension neural network, Prior knowledge, Neural network, Coal mines, Safety status

1. Introduction

A neural network is a massively parallel distributed processor made up of simple processing units that has a natural propensity for storing experiential knowledge and making it available for use. It resembles the brain in two respects: (1) knowledge is acquired by the network from its environment through a learning process and (2) interneuron connection strengths, known as synaptic weights, are used to store the acquired knowledge [1]. A major task for a neural network is to learn a model of the environment in which it is embedded, and to maintain the model sufficiently consistently with the real world so as to achieve the specified goals of the application of interest.

At present, nearly all neural networks are formulated by learning only from examples or pattern (training data). In real-world applications, the prior knowledge which is not in the form of pattern or examples always exists. In the present of prior knowledge, it would be inefficient for a neural network to make use of only the training patterns or examples without taking advantage of all known information about the problem. Nonetheless, in real world applications, a certain amount of information on the problem is usually known beforehand. The incorporation of this prior knowledge into the training is the key element that will allow an increase of performance in many applications.

Prior knowledge represented by facts about what is and what has been known refers to all information about the problem available in addition to the training data [2]. However, in this most general form, determining a neural network model from a finite set of samples without prior knowledge is an ill-posed problem, in the sense that a unique model may not exist.

From the point of view of information theory, prior knowledge provides information that is additional to the information contained in the training patterns. This information may not even exist in the training patterns but could be crucial to the neural network which is being developed. This is very important, if the network is required to extrapolate into regions of the input space, where it has not seen any training pattern. In this case, the neural network can rely on the prior knowledge if it is available in these regions [13].

Therefore, the incorporation of various forms of prior knowledge into machine learning is of much interest to the researchers, and many exciting results have been reported [3-8]. In this paper, we will discuss how to incorporate prior domain knowledge into the extension neural network (ENN) [9] with purpose of improving the performance of ENN and a prior-knowledge-based ENN (PKENN) recognition method is proposed. The basic idea of our approach is to determine the initial weights of ENN by using the domain prior knowledge. Further the paper is organized as follows. Section 2 introduces relevant theories, including the extension theory and the ENN technology. The mechanism of our proposed prior-knowledge-based ENN (PKENN) is presented in detail in Section 3. Some experiments will be tested to show the effective of PKENN in Section 4. Section 5 contains general discussions and conclusions.

2. Theoretical Background

2.1. Outline of Extension Theory

Extension theory was originally created by Cai to solve contradictions and incompatibility problems in 1983 [10]. Now, extension theory has been used in the research field of artificial intelligence (AI) and its relevant sciences and extension-theory-based methods have been applied to some engineering fields [11-16].

The hard core of extension theory is two theoretical pillars that include matter-element theory and the theory of extension set. The former studies matter-elements and their transformations; it can be easy to represent the nature of the matter. The latter is the quantitative tool of extension theory; it can represent the dependent degree of two matter-elements in the designed correlation function.

The extension set extends the crisp set and the fuzzy set. In the crisp set, an element either belongs to or does not belong to a set, so the range of the truth-values is $\{0, 1\}$, which can be used to solve a two-valued problem. In contrast to the crisp set, the fuzzy set allows for the description of concepts in which the boundary is not explicit. It concerns not only whether an element belongs to the set but also to what degree it belongs. The range of membership function is $[0, 1]$ in fuzzy set. The extension set extends the fuzzy set from $[0, 1]$ to $(-\infty, +\infty)$. This means that an element belongs to any extension set to a different degree. Define the membership function by $K(x)$ to represent the degree to which an element belongs to a set. A degree between zero and one corresponds to the normal fuzzy set. When $K(x) < 0$, it describes the degree to which x does not belong to a set, which is not defined in a fuzzy set. When $-1 < K(x) < 0$, this means that the element x still has a better chance to be included into the set if the set is adjusted. $K(x) < -1$ implies that the element x has no chance to belong to the set. It is also to represent the degree of an element not belonging to a set. The extension theory tries to solve incompatibility or contradiction problems by the transformation of the matter element. Comparisons of crisp sets, fuzzy sets, and extension sets are shown in Table 1.

Table 1. Comparisons among Crisp Set, Fuzzy Set and Extension Set

Compared Item	Crisp set	Fuzzy set	Extension set
Research objects	Certainty problems	Fuzzy problems	Contradictory problems
Model	Mathematics model	Fuzzy Mathematics model	Matter-element model
Descriptive function	Characteristic function	Membership function	Correlation function
Property	Precision	Ambiguity	Extension
Range of set	$C_{A(x)} \in \{0,1\}$	$\mu_{A(x)} \in [0,1]$	$K_{A(x)} \in (-\infty, +\infty)$
Distance	Classical distance	Classical distance	Extension distance
Logic system	Mathematics logic	Fuzzy logic	Extension logic

2.1.1. Matter-element Theory

1) *Definition of matter-element*: Defining the name of a matter by N , one of the characteristics of the matter by c and the value of c by v , a matter-element in extension theory can be described as follows:

$$R = (N, c, v) \tag{1}$$

where N , c and v are called three fundamental elements of the matter-element. For example, $R = (\text{Refrigerator}, \text{Rated voltage}, 220\text{V})$ can be used to state that the rated voltage of a refrigerator is 220V. If the value of the characteristic has a classical domain or a range, we define the matter-element for the classical domain as follows:

$$R = (N, c, v) = (N, c, < w^L, w^U >) \tag{2}$$

where w^L and w^U are the lower bound and upper bound of classical domain, respectively. For example, $R = (N, c, v) = (\text{Refrigerator}, \text{Operating voltage}, < 187\text{V}, 242\text{V} >)$ can be used to state that the operating voltage of a refrigerator is ranging from 187V to 242V.

2) *Multidimensional matter-element*: Assuming $R = (N, C, V)$, a multidimensional matter-element, $C = [c_1, c_2, \dots, c_n]$, a characteristic vector and $V = [v_1, v_2, \dots, v_n]$, a value vector of C , then a multidimensional matter-element is defined as :

$$R = (N, C, V) = \begin{bmatrix} N & c_1 & v_1 \\ & c_2 & v_2 \\ & \vdots & \vdots \\ & c_n & v_n \end{bmatrix} = \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_n \end{bmatrix} \tag{3}$$

where $R_i = (N, c_i, v_i), i = 1, 2, \dots, n$ is defined as the sub-matter-element of R . For example

$$R = (N, C, V) = \begin{bmatrix} N_3, & \text{average fault throw } (c_1), & < 1 \sim 1.3 > \\ & \text{fault number per unit area } (c_2), & < 1 \sim 1.5 > \\ & \text{variation coefficients of coal seam thickness } (c_3), & < 0.3 \sim 0.4 > \\ & \text{the difficulty degree of management of roof and floor } (c_4), & < 15 \sim 20 > \end{bmatrix}$$

is a four-dimensional matter element. Where N_3 stands for the pattern of geological safety status of coal mine – moderately safe (III).

2.1.2. Extension Set Theory

1) *Definition of extension set*: Let U be the universe of discourse. Then an extension set A on U is defined as a set of ordered pairs as follows:

$$A = \{(x, y) \mid x \in U, y = K(x) \in (-\infty, \infty)\} \tag{4}$$

where $y = K(x)$ is called the membership function for extension set A . $K(x)$ maps each element of U to a membership grade between $-\infty$ and ∞ . The higher the degree, the more the element belongs to the set. Under a special condition, when $0 \leq K(x) \leq 1$, it corresponds to a normal fuzzy set. $K(x) \leq -1$ implies that the element x has no chance to

belong to the set. When $-1 < K(x) < 1$, it is called an extension domain, which means that the element x still has a chance to become part of the set.

2) *Definition of correlation function:* If $X_0 = \langle a, b \rangle$ and $X = \langle c, d \rangle$ are two intervals in the real number field, and $X_0 \subset X$, then the correlation function in the extension theory can be defined as follows:

$$K(x) = \begin{cases} -\rho(x, X_0) & x \in X_0 \\ \frac{\rho(x, X_0)}{\rho(x, X) - \rho(x, X_0)} & x \notin X_0 \end{cases} \quad (5)$$

where $\rho(x, X_0)$ is defined as the extension distance between x and X_0 and written as

$$\rho(x, X_0) = \left| x - \frac{a+b}{2} \right| - \frac{b-a}{2} \quad (6)$$

$$\rho(x, X) = \left| x - \frac{c+d}{2} \right| - \frac{d-c}{2} \quad (7)$$

The correlation function $K(x)$ can be used to calculate the membership grade between x and X_0 as shown in Figure 1. When $K(x) \geq 0$, it indicates the degree to which x belongs to X_0 . When $K(x) < 0$, it describes the degree to which x does not belong to X_0 , which is not defined in fuzzy set theory. When $-1 < K(x) < 0$, it is called the extension domain, which means that the element x still has a chance to become part of the set if conditions change.

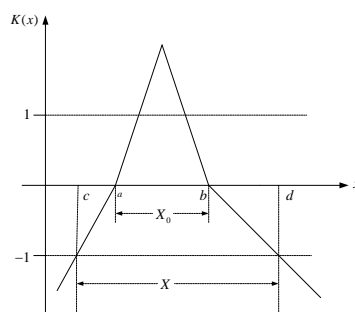


Figure 1. The Extension Membership Function

2.2. Extension Neural Network

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 120V. 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 M. H. Wang [9] to solve these problems. This new neural network is a combination of extension theory 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 [9, 17-23].

The structure of ENN is depicted in Figure 2. 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 j -th input node and k -th output node are w_{kj}^L and w_{kj}^U , respectively. ENN use ED (as shown in Eqn. (8)) 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 3.

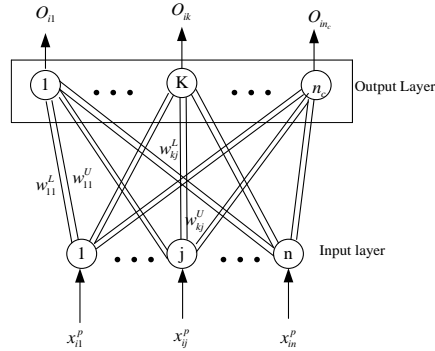


Figure 2. Structure of ENN

$$ED = \frac{|x - (w^U + w^L) / 2| - (w^U - w^L) / 2}{|(w^U - w^L) / 2|} + 1 \quad (8)$$

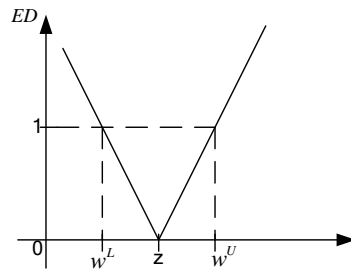


Figure 3. Extension Distance

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

Let training data set be $X = \{X_1, X_2, \dots, X_{N_p}\}$, where N_p the size of training data set is. The i -th instance in training data set is $X_i^p = \{x_{i1}^p, x_{i2}^p, \dots, x_{in}^p\}$, where n is the total number of the feature of instances, and the label of the i -th instance is p . To evaluate the ENN performance, the error rate function E_T is defined below:

$$E_T = N_m / N_p \quad (9)$$

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 i -th ($i = 1, 2, \dots, N_p$) training instance is read. Use the ED to calculate the distance between the training instance X_i^p and the k -th class as follows:

$$\begin{cases} ED_{ik} = \sum_{j=1}^n \left[\frac{|x_{ij}^p - z_{kj}| - (w_{kj}^U - w_{kj}^L) / 2}{|(w_{kj}^U - w_{kj}^L) / 2|} + 1 \right], k = 1, 2, \dots, n_c \\ z_{kj} = (w_{kj}^U + w_{kj}^L) / 2 \end{cases} \quad (10)$$

Find the k^* , such that $ED_{ik^*} = \text{Min}\{ED_{ik}\}$. 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 Eqn.(11)-(14). Repeat 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:

$$z_{pj}^{new} = z_{pj}^{old} + \eta(x_{ij}^p - z_{pj}^{old}) \quad (11)$$

$$z_{k^*j}^{new} = z_{k^*j}^{old} - \eta(x_{ij}^p - z_{k^*j}^{old}) \quad (12)$$

(b) Update the weights of the p -th and the k^* -th classes:

$$\begin{cases} W_{pj}^{L(new)} = W_{pj}^{L(old)} + \eta(x_{ij}^p - z_{pj}^{old}) \\ W_{pj}^{U(new)} = W_{pj}^{U(old)} + \eta(x_{ij}^p - z_{pj}^{old}) \end{cases} \quad (13)$$

$$\begin{cases} W_{k^*j}^{L(new)} = W_{k^*j}^{L(old)} - \eta(x_{ij}^p - z_{k^*j}^{old}) \\ W_{k^*j}^{U(new)} = W_{k^*j}^{U(old)} - \eta(x_{ij}^p - z_{k^*j}^{old}) \end{cases} \quad (14)$$

where η is a learning rate.

Because the ENN just adjusts the weights of the p -th and the k^* -th class, the learning of ENN has a speed advantage over the other supervised learning algorithms, and can quickly adapt to new and important information.

3. The Mechanisms of PKENN-based Pattern Recognition Method

3.1. Basic Design Idea of PKENN

When dealing with real-world problems, there is considerable amount of prior domain knowledge that can provide insights on various aspect of the problem. However, in many real-world applications, many machine learning methods rely solely on the data sets for their learning phase and do not take into account any explicitly expressed domain knowledge. These data-driven neural networks rely heavily on the sample data and ignore most of existing domain knowledge. Figure 4 (a) is ENN-based training model driven by training data. In fact, in the engineering practice, especially in the industries such as coal mine, the historical databases for training data set are very limited. Beside, these databases are inherently noisy, being subject to errors due to sensor noise and system imperfections. So, the training data set unavoidably contain some noise and uncertainties. If the training samples are not informative enough, the learning of ENN may fail to reveal the inherent relationship between inputs and outputs. Hence, the ENN model only driven by training data may cause performance degradation. On the other side, there exist accumulation of knowledge and experience making the product. The ability of a neural network mode to incorporate domain knowledge is particularly valuable when it is applied to real-world problems, where valuable prior domain knowledge is available. If successfully incorporated such prior domain knowledge will improve both the quality of the result produced by the learning process of neural network and the efficiency of the learning process itself. Figure 4 (b) is PKENN-based training model driven by combination of training data set and prior knowledge.

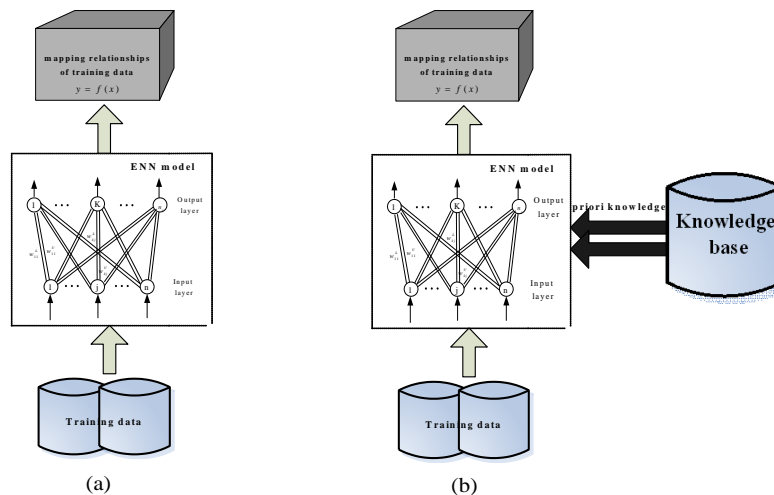


Figure 4. ENN-based Training Model (a) Data-driven ENN-based Training Model (b) Prior-knowledge-based ENN Training Model

3.2. Knowledge Representation of ENN

Generally, the performance of a neural network classifier model is evaluated from two aspects, that is: (1) the fitting precision of training data set and prediction and recognition ability of unknown data (*i.e.*, the generalization ability) and (2) the embodying degree of knowledge to environment of applications. Knowledge refers to stored information or models used by a person or machine to interpret, predict, and appropriately respond to the outside world [1]. If the knowledge can be used in engineering applications easily, knowledge should be in a form (*i.e.*, the representational form of knowledge) that is understandable by both engineering experts and academic researchers. Knowledge representation is one of the principal elements of artificial intelligence (AI). The primary characteristics of knowledge representation are twofold: (1) what information is actually made explicit; and (2) how the information is physically encoded for subsequent use. Therefore, knowledge representation is goal directed. In real-world applications of neural networks, a good solution depends on a good representation of knowledge [24]. Typically, however, the possible forms of representation from the inputs to internal network parameters are highly diverse, which tends to make the development of a satisfactory solution by means of a neural network a real design challenge.

In a neural network of specified architecture, knowledge representation of the surrounding environment is defined by the values taken on by the free parameters (*i.e.*, weights and biases) of the network. The form of this knowledge representation constitutes the very design of the neural network, and therefore holds the key to its performance. The subject of how knowledge is actually represented inside an artificial network is every complicated. An important issue that has to be addressed, of course, is how to develop a specialized structure by building prior information into its design. Unfortunately, there are currently no well-defined rules for doing this [1].

Knowledge representation of traditional neural networks is implicit representation that is stored in the form of internal encoding information. Training data set represents knowledge about application environment that a neural network can learning through training. Through training process, the weights and biases of the network are modified for storing the knowledge of environment application. However, for traditional neural networks, such as BP network and LVQ network, the knowledge representation that stored in weight matrix and bias matrix is hard to understand and interpret. Hence, they can hardly incorporate their knowledge into the learning algorithms, even if the architecture of learning system allows manual tuning of a set of control

parameters [25]. For ENN, the acquired knowledge after a learning process also is stored in the weight matrix. But different with traditional neural networks, for ENN, this problem can be solved easily because the ENN has its own characteristics. The distinctive characteristic of ENN is that the ENN can effectively solve the classification and recognition problems whose features are defined over an interval. For this point, the maximum value and minimum value of the interval are very meaningful for the training of ENN. The physical meaning of knowledge reflected in the weight matrix is explicit and easy to understand and interpret. The two connection weights between input neuron node and output node of ENN represent the maximum value and minimum value for the classical domain of the feature, respectively. Hence, for ENN, representation of domain knowledge has a single standardized format makes easy to code this representation in computer system. We can incorporate the knowledge into the weights of the ENN and the initial connection weights can be determined by prior knowledge, as shown in Figure 5.

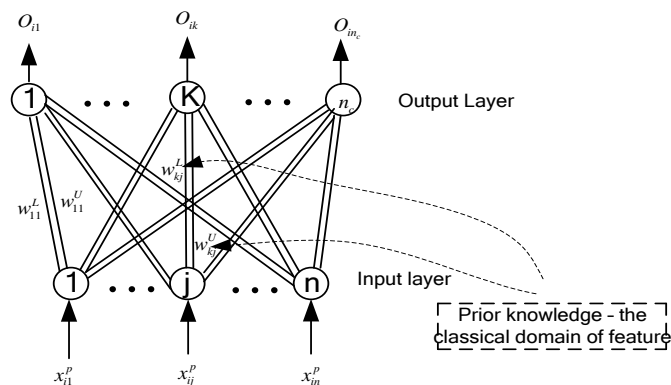


Figure 5. The Structure of PKENN

3.3. Advantages of PKENN-based Recognition Method

Generally, the historical database about coal mines safety status is very limited. The traditional-ANN-based methods are not appropriate for application in this field due to needing a large amount of training data that are both correct and high-quality. Moreover, the problem of recognition and evaluation of coal mines safety status is a problem whose features are defined in a range. So it is hard to implement an appropriate recognition method using traditional neural networks. At the same time, the structure design of traditional neural networks is sometimes not an easy problem, especially the determination of the number of neurons in the hidden layer and some control parameters. In contrast, the proposed ENN-based recognition method in [21] is no special demands on training data and has a better performance than other traditional-ANN-based methods in application of recognition and evaluation of safety status pattern of coal mines. However, the ENN-based method does not consider the function of prior domain knowledge. With purpose of further improving the performance of ENN in practical engineering applications, prior domain knowledge is incorporated into ENN. Many research works point out that prior knowledge about the problem domain at hand may significantly improve the performance of classifiers in many applications. Here, the prior domain knowledge is the classical domain of each feature. This kind of prior knowledge can be obtained from accumulated years of experience, which are convenient and easy relatively. More importantly, the knowledge about classical domain of each feature is easy to embed the weights of ENN. With constraint weights, prior domain knowledge could provide important information to eliminate noise within the features and examples for selection and sanitizing and direct ENN to lean more effectively. Therefore, PKENN-based method is used to design the pattern recognition system for coal mines safety status. Compared

with other methods and traditional ENN-based method, the PKENN-based pattern recognition method has following advantages:

(1). Be not strict with historical database. This is because that domain knowledge contains much useful information. This additional information may compensate for the lack of training data. Hence, even in the case of complicated environment and small size of training data set, the PKENN-based classifier still has excellent performance. For traditional ANN, the training data must be sufficient and compatible to ensure proper training.

(2). The required domain knowledge can be obtained expediently and easy to understand for both researchers and domain experts.

(3). The structure design of PKENN-based model is simple and easy to operate. The design of PKENN-based model consists of two aspects: (1) the structure design of ENN. We just need to know the number of features of input pattern and the number of status patterns, and then the structure can be determined. (2) The incorporating domain knowledge into ENN.

(4). The initial weights of ENN are certain, which can be directly determined from training data and domain knowledge according to the upper bound and lower bound of classical domain of input features.

(5). The proposed method has a speed advantage because of the learning algorithm. The proposed method permits fast adaptive processes for the new data. The domain knowledge could delete error and noisy data, guide the learning of ENN and alleviate the learning burden.

3.4. Structure Design of ENN

The environment and working condition underground coal mines are very complicated, which involve many hazards. All hazards constitute the dangerous sources to the operating workers and production system in the coal mines. Each hazard in coal mines includes many factors influencing safety status of coal mines. Different kind of hazard has the different number of factors. These factors of each hazard are chosen as the features and resulting feature vector can be denoted by $C = [c_1, c_2, \dots, c_n]$. Here, n stands for the number of the features. Each hazard of coal mines has several safety status patterns. According to the engineering practice, past studies [12, 21, 26] and the opinion of relevant experts, the coal mines safety status is classified into 5 status patterns, i.e. pattern I: very safe (N_1); pattern II: safe (N_2); pattern III: moderately safe (N_3); patternIV: unsafe (N_4) and pattern V: very unsafe (N_5).

Compared with traditional ANN, the problem of structure design of ENN can be easier to solve. Based on the above discussion, the structure of the ENN consists of the input layer and the output layer. There are five output neuron nodes which correspond to five patterns of safety status of coal mines in the output layer, while there are n input neuron nodes which correspond to n features in input layer. Only one output neuron node in the output layer remains active (the output value is 1) and the output values of other neuron nodes are zero to indicate a safety status pattern of the input instance. The safety status patterns of coal mines and corresponding output encoding of ENN are shown in Table 2.

Table 2. Types of Safety Status and Corresponding Output Encoding of ENN

Types of safety statue	Corresponding output encoding of ENN
1 Very safe (level I)	10000
2 Safe (Level II)	01000
3 Moderately safe (Level III)	00100
4 Unsafe (Level IV)	00010
5 Very unsafe (Level V)	00001

3.5. PKENN-based Recognition Algorithm

We apply the PKENN-based recognition method to the problem of safety status pattern recognition of coal mines mainly because of the advantages of PKENN and the actual engineering needs. In other words, in the application of safety status pattern recognition of coal mines, the historical database about coal mines safety status is very limited, and the database unavoidably contain some noise and uncertainties. Moreover, the prior knowledge about this engineering application is convenient to access and understand. Prior knowledge may be acquired from several source, *e.g.*, human judgment, application guidelines, and manuals, world knowledge, and in-domain websites. Furthermore, the domain knowledge could be incorporated into ENN easily and effectively guide the learning of ENN. The detailed PKENN-based recognition algorithm can be described as following:

Step 1: Prepare for training data set and knowledge base. Based on the characteristics of knowledge representation of ENN, the domain knowledge required here is about the quantitative range of each feature about each safety status pattern, that is $L_{kj} = \langle b_{kj}^L, b_{kj}^U \rangle$. L_{kj} stands for the quantitative range of the j -th feature regarding to the k -th safety status pattern. For example, in the application of evaluation and recognition of geological safety status pattern in coal mines, average fault throw (m) is an important factor. Based on the domain knowledge, we can know that when the geological safety status pattern is safe (pattern II), then the quantitative range of this feature is $\langle 0.8, 1 \rangle$.

Step 2: Set up the structure of the ENN. In practical terms, the number of neurons in input layer and the output layer should be determined.

Step 3: From training data set and domain knowledge, build matter-models of each safety status pattern:

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

In the extension theory, c_j is the j -th feature of N_k and $V_{kj} = \langle w_{kj}^L, w_{kj}^U \rangle$ are the classical domains of the k -th class (N_k) about the j -th feature c_j . The range of classical domains is jointly determined by training data set and domain knowledge:

$$\begin{cases} w_{kj}^L = \text{Min}\{\text{Min}_{i \in N_p}\{x_{ij}^k\}, b_{kj}^L\}, & \text{if } b_{kj}^L \text{ exist} \\ w_{kj}^L = \text{Min}\{x_{ij}^k\}, & \text{if } b_{kj}^L \text{ does not exist} \end{cases} \quad (16)$$

$$\begin{cases} w_{kj}^U = \text{Max}\{\text{Max}_{i \in N_p}\{x_{ij}^k\}, b_{kj}^U\}, & \text{if } b_{kj}^U \text{ exist} \\ w_{kj}^U = \text{Max}\{x_{ij}^k\}, & \text{if } b_{kj}^U \text{ does not exist} \end{cases} \quad (17)$$

Step 4: Set the connection weights between input nodes and output nodes according to the matter-element models.

Step 5: Calculate the initial center of every safety status pattern:

$$Z_k = \{z_{k1}, z_{k2}, \dots, z_{kn}\} \quad (18)$$

$$Z_{kj} = (w_{kj}^L + w_{kj}^U) / 2, k = 1, 2, \dots, n_c; j = 1, 2, \dots, n \quad (19)$$

Step 6: Read the i -th training instance and its label of safety status pattern p

Step 7: Using Eqn. (10) to calculate extension distance between the training instance x_i^p and the k-th safety status pattern.

Step 8: Find the k^* , such that $ED_{ik^*} = \text{Min}\{ED_{ik}\}$. If $k^* = p$ then go to Step 10; otherwise Step 9.

Step 9: Update the weights of the p-th and the k^* -th classes according to Eqn. (11) ~ (14)

Step 10: Repeat Step 6 to Step 9, if all instances have been classified, then a learning epoch is finished.

Step 11: Stop, if the learning process has converged, or the total error rate E_T has arrived at a preset value, otherwise, return to Step 6.

Step 12: Use the trained ENN to recognize the safety status pattern of coal mines. Use extension distance to measure the similarity between the tested object and every safety status pattern. Find the k^* , such that $ED_{ik^*} = \text{Min}\{ED_{ik}\}$, and set the $o_{ik^*} = 1$ to indicate that the safety status pattern of the tested object is k^* .

Step 13: Stop, if all the tested objects have been recognized, otherwise go to Step 12.

4. Experimental Results and Discussion

In this section, a real-world engineering application is taken as an example to illustrate the advantages of PKENN recognition method.

4.1. Background of the Application

The background of the application is the geological safety status pattern recognition in coal mines. The historical data sets came from Ref. [21, 26]. The training data set contains 20 instances (as listed in Table 3) and the testing data set contains five instances (as listed in Table 4). According to the degree of influence of the various factors upon the geological safety status of coal mines and in agreement with past studies [21, 26], the following four main factors are chosen as the features for the recognition of the geological safety condition of coal mines. These features of input patterns would be most useful for recognition. These features include average fault throw c_1 , fault number per unit area c_2 , variation coefficients of coal seam thickness c_3 and the difficulty degree of management of roof and floor c_4 . So, the feature vector of the geological safety status of coal mines could be denoted by $C = [c_1, c_2, c_3, c_4]$. Based on Section 3.4, the structure of ENN for the application of geological safety status recognition comprises both input layer and the output layer. The numbers of input neuron nodes and output neuron nodes are 4 and 5, respectively.

Prior knowledge may be acquired from several sources, *e.g.*, human judgment and experience, application guidelines, and manuals, world knowledge, and in-domain websites. In this application, the prior knowledge comes from the experience of domain-experts. The prior knowledge here is the quantitative range of each factor about the geological safety condition of coal mines, as shown in Table 5-8 [26]. In order to prove the efficiency and validity of the proposed PKENN recognition method, existing methods [21, 26] are compared with proposed method. At the same time, some experiments are conducted to compare proposed method with several other ANN-based methods.

Table 3. Training Data Set of Coal Mines Geological Safety Status

No.	c_1	c_2	c_3	c_4	Safety status pattern
1	0.5	0	0.1	6	I
2	0.75	0.12	0.15	7	I
3	0.61	0.17	0.17	8	I
4	0.57	0.45	0.18	7.5	I
5	0.8	0.5	0.2	10	II
6	0.82	0.55	0.29	13	II
7	0.91	0.65	0.21	11	II
8	0.92	0.78	0.28	14	II
9	1	1.2	0.3	15	III
10	1.2	1.2	0.37	17	III
11	1.15	1.1	0.25	16	III
12	1.24	1.4	0.33	18.5	III
13	1.3	1.5	0.4	20	IV
14	1.35	1.7	0.46	23	IV
15	1.48	1.65	0.49	21	IV
16	1.41	1.85	0.42	23.5	IV
17	1.5	2	0.5	25	V
18	1.6	2.5	0.6	26	V
19	1.65	2.9	0.7	27	V
20	1.78	3	0.8	28	V

Table 4. Testing Data Set of Coal Mines Geological Safety Statue

No.	c_1	c_2	c_3	c_4	Expecting result
1	0.75	0.12	0.14	6	I
2	0.82	0.55	0.27	11	II
3	1.18	1.2	0.37	17	III
4	1.48	1.65	0.49	21	IV
5	1.65	2.9	0.7	27	V

Table 5. Average Fault Throw c_1 (m)

Quantitative range	<0.8	0.8-1	1-1.3	1.3-1.5	≥ 1.5
Safety status pattern	Very safe (I)	Safe (II)	Moderately safe (III)	Unsafe (IV)	Very unsafe (V)

Table 6. Fault Number Per Unit Area (m^2)

Quantitative range	<0.5	0.5-1	1-1.5	1.5-2	≥ 2
Safety status pattern	Very safe (I)	Safe (II)	Moderately safe (III)	Unsafe (IV)	Very unsafe (V)

Table 7. Variation Coefficients of Coal Seam Thickness

Quantitative range	0.1-0.2	0.2-0.3	0.3-0.4	0.4--0.5	≥ 0.5
Safety status pattern	Very safe (I)	Safe (II)	Moderately safe (III)	Unsafe (IV)	Very unsafe (V)

Table 8. The Difficulty Degree of Management of Roof and Floor

Quantitative range	<10	10-15	15-20	20-25	≥ 25
Safety status pattern	Very safe (I)	Safe (II)	Moderately safe (III)	Unsafe (IV)	Very unsafe (V)

4.2. Comparison with Existing Methods

The purpose of these experiments is to compare with BP-based method in Ref. [26] and ENN-based method in Ref. [21]. The experiment environment is the same as Ref. [21] and Ref. [26]. Based on the training data set (shown in Table 3) and the prior domain knowledge (shown in Table 5-8), according to the Eqs. (15-17), the element-models of each safety status pattern of coal mines geology can be built (as listed in Table 9). According to the experience, the learning rate η of ENN is set to be 0.1 in the following experiments. Training data set is used to train ENN, and then the test data set (shown in Table 4) is used to test this recognition method. The simulation results indicate prior-knowledge-based ENN method can accurately distinguish all five test instances.

Table 9. Matter-models of Coal Mine Geological Safety Status

Safety status pattern	Matter-element model
1 Very safe (I)	$\mathbf{R}_1 = (N_1, \mathbf{C}, \mathbf{V}_1) = \begin{bmatrix} N_1, & c_1, & < 0.5-0.8 > \\ & c_2, & < 0-0.5 > \\ & c_3, & < 0.1-0.2 > \\ & c_4, & < 6-10 > \end{bmatrix}$
2 Safe (II)	$\mathbf{R}_2 = (N_2, \mathbf{C}, \mathbf{V}_2) = \begin{bmatrix} N_2, & c_1, & < 0.8-1 > \\ & c_2, & < 0.5-1 > \\ & c_3, & < 0.2-0.3 > \\ & c_4, & < 10-15 > \end{bmatrix}$
3 Moderately safe (III)	$\mathbf{R}_3 = (N_3, \mathbf{C}, \mathbf{V}_3) = \begin{bmatrix} N_3, & c_1, & < 1-1.3 > \\ & c_2, & < 1-1.5 > \\ & c_3, & < 0.3-0.4 > \\ & c_4, & < 15-20 > \end{bmatrix}$
4 Unsafe (IV)	$\mathbf{R}_4 = (N_4, \mathbf{C}, \mathbf{V}_4) = \begin{bmatrix} N_4, & c_1, & < 1.3-1.5 > \\ & c_2, & < 1.5-2 > \\ & c_3, & < 0.4-0.5 > \\ & c_4, & < 20-25 > \end{bmatrix}$
5 Very unsafe (V)	$\mathbf{R}_5 = (N_5, \mathbf{C}, \mathbf{V}_5) = \begin{bmatrix} N_5, & c_1, & < 1.5-1.78 > \\ & c_2, & < 2-3 > \\ & c_3, & < 0.5-0.8 > \\ & c_4, & < 25-28 > \end{bmatrix}$

In order to test the generalization ability, the leave-one-out cross-validation (LOOCV) experiment is carried out. LOOCV involves using a single instance from the original data as the validation data, and the remaining instances as the training data. It is repeated such that each observation in the sample is used once as the validation data. In this experiment, 25 times training processes are repeated. At each time, one instance is used as the testing data and other 24 instances are used as the training. For BP BP-based method, the total error number is 3 and the average error recognition rate is 12%. For ENN-based method, the total error number is 2 and the average error recognition rate is 8%. The same experiment is conducted using PKENN-based recognition method and the experiment results are perfect, and the average error recognition rate is 0. It is obvious that the proposed PKENN recognition method in this paper has stronger generalization ability.

In Table 10, the performance of the proposed PKENN-based method is compared with the existing BP-based method [26] and ENN-based method [21]. From Table 10, performance analysis shows the PKENN-based method is an optimal method. Simulation results show the new method possesses excellent properties, and is suitable for the application of safety status pattern recognition of coal mines. The reason for this conclusion this that the ENN embedded the domain knowledge, and the domain knowledge can constrain the learning of ENN. Loading of constraints plays a key role in the learning processing of ENN, which can guide the learning of ENN effectively and alleviate the learning burden.

Table 10. Comparison Results of Existing Methods and PKENN-based Method

Item	BP-based method ^[26]	ENN-based method ^[21]	PKENN-based method
Structure	4-7-4-1	4-5	4-5
Number of connections	60	40	40
Learning rate	Variable learning rate	0.1	0.1
Initial weights	random values	determined by training data set	determined by training data set and prior knowledge
Learning speed (epochs)	980	1	1
Learning error	0.01	0	0
Accuracy of test data set	100%	100%	100%
Average accuracy of LOOCV	88%	92%	100%

4.3. Comparison with Other ANN-based Methods

In this experiment, the training data set contained 25 training instances (*i.e.*, the full data from Table 3 and Table 4) and the testing data set was equal to the training data set, containing 25 instances. Several ANN-based methods are used to compare the performance. To be more objective, each method is done 10 times and the average results are taken. In Table 9, the performance of the proposed PKENN-based method is compared with the other recognition methods in the same data set. It can be seen from Table 11 that the proposed method is at least as good as the other methods in structure, learning time, recognition accuracy and learning error, but it has stronger generalization ability. Moreover, the proposed method can take expert experiences (prior knowledge) before learning, and can also produce meaningful output after learning. On the other hand, the proposed prior-knowledge-based based ENN method permits fast adaptive process for significant and new information, and it is easy to acquire the knowledge and maintain the classification database. From all above, the PKENN-based method has the best performance, it can promote the whole performance of recognition system effectively in engineering.

Table 11. Comparison of the Recognition Performance of Various ANN-based Methods

Method	Structure	Number of Weights	Average learning time (epochs)	Learning error	Average classification accuracy rate (%)	Average accuracy of LOOCV (%)
Elman	4-15-5	135	>1000	>0.01	100%	84%
LVQ	4-15-5	135	25.7	0	100%	100%
BP	4-9-5	81	216.6	0.0133	96%	88%
ENN	4-5	40	1	0	100%	92%
PKENN	4-5	40	1	0	100%	100%

Elman: Simple recurrent neural network; LVQ: Learning vector quantization; BP: Back propagation network; ENN: Extension neural network

4.4. Tests of Error-containing Data

The environment of coal mines is so complex and the parameters are changeable and uncertain, so the input data of the geological safety status pattern recognition system would unavoidably contain some noise and uncertainties. The sources of error include environmental noise, sensors, human mistakes, *etc.*, which could lead to data uncertainties

and errors. To take into account the noise and uncertainties, all data (25 instances) are created by adding 0% to $\pm 40\%$ of random, uniformly distributed error to the training data to evaluate the fault-tolerant ability of the proposed PKENN-based recognition method. The testing results using different amounts of errors added are given in Table 12 for the different recognition methods. Usually, the error containing data indeed degrades the recognition capabilities in proportion to the amounts of error added. This Table shows that PKENN-based method can bear remarkable tolerance to the errors contained in the data.

It should be noted that the proposed PKENN-based recognition method has a significantly higher recognition accuracy of 100% with $\pm 35\%$ errors added. Moreover, the proposed method shows good tolerance to added errors, and has a high accuracy of 96% in extreme error of $\pm 40\%$.

From the Table 12, we can find that the fault-tolerant ability of LVQ-based method is not very good. Under the same conditions, the average accuracy rate is less than 60% with $\pm 30\%$ errors added. At the same time, the Elman-based method, BP-based and ENN-based method all have a good fault-tolerant ability, but Elman-based method and BP-based methods need much time to train network in actual application. For ENN-based method, comparing with PKENN-based method, the performance of PKENN-based method is further improved. This experiment proved that the PKENN-based recognition method has the best excellent fault-tolerant ability. The key reason why the PKENN-based method has excellent performance is that certain prior domain knowledge provides important information to eliminate noise. In theory, if we understand all the knowledge about this application and successfully incorporate full knowledge into ENN, there is no doubt that the recognition rate is 100%.

Table 12. Recognition Performances of Different Methods with Different Percentages of Noise Added

Error percentage	Average accuracy rate of different recognized methods				
	Elman	LVQ	BP	ENN	PKENN
$\pm 0\%$	100%	100%	100%	100%	100%
$\pm 5\%$	100%	96.8%	100%	100%	100%
$\pm 10\%$	100%	91.6%	98%	100%	100%
$\pm 20\%$	96%	76.4%	90%	100%	100%
$\pm 30\%$	92%	<60%	84.8%	100%	100%
$\pm 35\%$	80.4%	<60%	71.2%	92%	100%
$\pm 40\%$	<60%	<60%	<60%	84%	96%

5. Conclusions

On the bases of requirements of engineering practice and the characteristics of knowledge representation of ENN, with the purpose of further improving the performance of ENN in engineering practice, a novel prior-knowledge-based ENN (PKENN) recognition method is proposed and applied in the application of safety status pattern recognition of coal mines in this paper. The results of experiments indicate that, compared with the results of other ANN-based methods and ENN-based recognition method, the PKENN-based recognition method has the advantage of less learning time, lower learning error and higher accuracy. Moreover, the proposed recognition method has strong generalization ability and excellent fault tolerant ability, which are the most important aspects for overall evaluation of the safety status pattern recognition system of core mines. Hence, our proposed PKENN-based method is more efficient in engineering applications of safety of coal mines.

Acknowledgements

The authors gratefully acknowledge the project supported by the Research Foundation from Ministry of Education of China under grant No 107021, the project supported by the Scientific Research Starting Foundation for High-level Talents, North China Institute of Water Conservancy and Hydroelectric Power under grant No.201117 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. 2011A060901001-12D. The authors would like to thank the anonymous reviewers and the editors for their valuable comments and suggestions

References

- [1] S. Haykin, "Neural networks and learning machines (Third edition)", Prentice Hall, Upper Saddle River, New Jersey, (2008).
- [2] B. Scholkopf and A. Smola, "Learning with Kernels", MIT Press, Cambridge, Massachusetts, (2002).
- [3] S. Zhao and T. S. Dillon, "Incorporating prior knowledge in the form of production rules into neural network using Boolean-like neurons," *Applied intelligence*, vol. 7, (1997), pp. 275-285.
- [4] F. Han and D.-S. Huang, "A new constrained learning algorithm for function approximation by encoding a priori information into feedforward neural network", *Neural Comput & Applic*, vol. 17, (2008), pp. 433-439.
- [5] S. Milanič, S. Strmčnik, D. Šel, N. Hvala and R. Karba, "Incorporating prior knowledge into artificial neural networks—an industrial case study", *Neurocomputing*, vol. 62, (2004), pp. 131–151.
- [6] T. Yu, S. Simoff and T. Jan, "VQSVM: A case study for incorporating prior domain knowledge into inductive machine learning", *Neurocomputing*, vol. 73, (2010), pp. 2614–2623.
- [7] L. V. Utkin and Y. A. Zhuk, "Imprecise prior knowledge incorporating into one-class classification," *Knowledge and Information Systems*, (2013), doi: 10.1007/s10115-013-0661-7.
- [8] L. H. Song and Y. Jia, "Incorporation of prior knowledge in adaptive learning for modeling the robotic profile grinding", *Robot*, vol. 33, no. 6, (2011), pp. 641-648.
- [9] M. H. Wang and C. P. Hung, "Extension neural network and its applications," *Neural Networks*, vol. 16, no. 5-6, (2003), pp. 779-784.
- [10] W. Cai, "Extension theory and its application," *Chinese Science Bulletin*, vol. 44, no. 17, (1999), pp. 1538–1548.
- [11] Y. Zhou, X. Qian and T. Jian, "Research on the extension assessment model of underground coalmine safety behavior", *Journal of China University of Mining & Technology*, vol. 38, no. 4, (2009).
- [12] Z. Yu, Q. Xu and Z. Jin-Hui, "Application of extension theory to the evaluation of coal mines geological safety," *Proceedings of the 21st annual international conference on Chinese control and decision conference*, Piscataway: IEEE Press, (2009), pp. 4937 – 4941.
- [13] M. H. Wang and Y. F. Tseng, "A novel clustering algorithm based on the extension theory and genetic algorithm", *Expert Systems with Applications*, vol. 36, no. 4, (2009), pp. 8269-8276.
- [14] D. Hong-Gui, C. Jian and L. An, "Application of extension method to fault diagnosis of transformer", *Journal of central south university of technology*, vol. 14, no. 1, (2007), pp. 88-93.
- [15] M. H. Wang, "Application of extension theory to PD pattern recognition in high-voltage current transformers", *IEEE Trans on Power Delivery*, vol. 20, no. 3, (2005), pp. 1939-1946.
- [16] J. Ye, "Application of extension theory in misfire fault diagnosis of gasoline engines", *Expert Systems with Applications*, vol. 36, no. 2, (2009), pp. 1217-1221.
- [17] Y. Zhou, L. Tian and L. Liu, "Improved Extension Neural Network and Its Applications," *Mathematical Problems in Engineering*, vol. 2014, Article ID 593021, (2014), pp. 14.
- [18] M. H. Wang, "Partial Discharge Pattern Recognition of Current Transformers Using an ENN", *IEEE Transactions on Power Delivery*, vol. 20, no. 3, (2005), pp. 1984-1990.
- [19] H.-C. Chen, F.-C. Gu and M.-H. Wang, "A novel extension neural network based partial discharge pattern recognition method for high-voltage power apparatus", *Expert Systems with Applications*, vol. 39, no. 3, (2012), pp. 3423-3431.
- [20] M.-H. Wang, K.-H. Chao, W.-T. Sung and G.-J. Huang, "Using ENN-1 for fault recognition of automotive engine", *Expert Systems with Applications*, vol. 37, no. 4, (2010), pp. 2943-2947.
- [21] Y. Zhou, W. Pedrycz and X. Qian, "Application of extension neural network to safety status pattern recognition of coal mines", *Journal of Central South University of Technology*, vol. 18, no. 3, (2011), pp. 633-641.
- [22] M. H. Wang, K. H. Chao, W. T. Sung and G. J. Huang, "Using ENN-1 for Fault Recognition of Automotive Engine", *Expert Systems with Applications*, vol. 37, no. 4, (2010), pp. 2943-2947.
- [23] M. H. Wang, "Application of Extension Neural Network Type-1 to Fault Diagnosis of Electronic Circuits," *Mathematical Problems in Engineering*, vol. 2012, Article ID 352749, (2012), pp. 12.

- [24] W. A. Wood, "Important issues in knowledge representation," Proceedings of the Institute of Electrical and Electronics Engineers, vol. 74, (1986), pp. 1322-1334.
- [25] T. Yu, S. Simoff and T. Jan, "VQSVM: A case study for incorporating prior domain knowledge into inductive machine learning", Neurocomputing, vol. 73, (2010), pp. 2614-2623.
- [26] S. Shi-Liang and W. Hai-Qiao, "Non-linear dynamic safety assessment in coal mines", Coal Industry Press, Beijing, (2001).

Authors



Yu Zhou, was born in China. He received his B.E. in Automation from Henan University in 2004, China, his M.E. in Detection Technology and Automatic Equipment from Guangdong University of Technology in 2007, China, and his Ph.D. in Control Theory and Control Engineering from China University of Mining and Technology (Beijing) in 2010. He has been a visiting Ph.D. student at Department of Electrical and Computer Engineering, University of Alberta, Canada from Sep. 2009 to Sep. 2010. He is now working at North China University of Water Conservancy and Hydroelectric Power as a lecturer. As a researcher, he also works at Key Laboratory of Innovation Method and Decision Management System of Guangdong Province, Guangzhou. His current research interests include the following areas: artificial neural network, pattern recognition, intelligence computing and intelligent control.



Lian Tan, was born in China. He received his B.E. in Automation from Zhengzhou University in 2001, China, his M.E. in Control Theory and Control Engineering from Xi'an Jiaotong University in 2005, China. He is now working at North China University of Water Conservancy and Hydroelectric Power as a lecturer. His current research interests include the following areas: artificial neural network, pattern recognition, intelligence computing.

