A Novel Expertise-Guided Machine Learning Model for Internal Fault State Diagnosis of Power Transformers

Qunli Wu 1,2 and Hongjie Zhang 1,*

1 Department of Economics and Management, North China Electric Power University, 689 Huadian Road, Baoding 071000, China; 51851220@ncepu.edu.cn
2 Beijing Key Laboratory of New Energy and Low-Carbon Development, North China Electric Power University, Changping, Beijing 102206, China
* Correspondence: 2172218037@ncepu.edu.cn

Received: 20 January 2019; Accepted: 6 March 2019; Published: 14 March 2019

Abstract: The fault diagnosis of power transformers is of great significance to improve the reliability of power systems. This paper proposes a novel fault diagnosis method called the expertise-guided machine learning (EGML) model where a genetic algorithm (GA) and a mind evolutionary algorithm (MEA) are used as optimization algorithms. Thereby, two types of EGML models are generated, that is, the GA-EGML model and the MEA-EGML model. In the EGML model, knowledge function replaces the cost function of traditional artificial intelligence algorithms, which can provide additional information for each individual and bring some corrections to the prediction results. To investigate the application potentials of the proposed models in power transformer fault diagnosis, real dissolved gases data are utilized to evaluate the diagnosis performance of the proposed models. Results indicate that the performance of the EGML model outperforms the traditional back propagation neural network (BPNN) model and all other models participating in the comparison. Both the GA-EGML model and MEA-EGML model can be used to diagnose the faults of a power transformer, and the latter is better. In addition, to further investigate the robustness of the proposed models for different data, four scenarios are simulated. Empirical results show that the accuracies of all models decrease in the other three scenarios compared to the baseline scenario, especially in scenario 2. However, the proposed models decline less than the traditional models in scenario 2 and scenario 4, and obtain satisfactory accuracy in all scenarios.

Keywords: Expertise-Guided Machine Learning model; fault diagnosis; power transformer; genetic algorithm; mind evolutionary algorithm; robustness test

1. Introduction

With the drastic increase in the power system capacity, the requirements for the reliability of power transmission and supply are higher [1]. As important equipment for power transmission in power systems, the power transformer is significant for safe and stable operation of the entire power network. In the event of a fault in a transformer, generating capacity will be harmed. The worst faults may even lead to the collapse of the entire power system, greatly hindering the development of the whole national economy. Therefore, it is important to study fault diagnosis technology pertaining to power transformers [2].

Power transformer faults generally arise from electrical and thermal stresses, and these faults differ only in their energy, location, and time of occurrence. The oil temperature will rise and some gases will be produced when the fault appears. There are five common dissolved gases in transformer oil, namely, hydrogen (H₂), methane (CH₄), ethane (C₂H₆), ethylene (C₂H₄), and acetylene (C₂H₂) [3,4].
They are considered as fault indicators of power transformers since the patterns and quantities they generate depend on the fault category. Therefore, the character and quantity of dissolved gases have a notable role in evaluating the fault type and operational reliability of power transformers [5,6].

In light of the corresponding relationship between power transformer faults and the dissolved gases, many traditional methods have been proposed to diagnose transformer faults combining with gas chromatography. These approaches for fault diagnosis are generally classified into three types, namely, the characteristic gas method [7,8], the gas production rate method [9] and the three-ratio method [10–12]. In China, more than 50% of the power transformer faults in the power system are found using dissolved gas analysis (DGA)-based diagnosis methods which diagnose transformer fault types and their severity based on the content, ratio to each other, and gas production rate of the dissolved gases in the transformer oil [13]. Therefore, in addition to the above three main traditional approaches, some improved approaches have emerged, such as the Doernenburg method, the Rogers ratio method, the Duval triangle method, the International Electrotechnical Commission (IEC) ratio method, and the Key gas method [4,14–16]. These methods generally utilize several gas ratios or compare gas concentrations with the designated criteria to diagnose the state of a power transformer [17]. However, most of these traditional diagnosis methods only make a limited contribution to a transformer’s fault diagnosis, which cannot accurately reflect its real fault type [18]. Particularly, it is more difficult to accurately judge the fault state with few dissolved gases; high probability of misdiagnosis will happen when the measured and calculated gas ratio is close to the critical value [19]. In addition, the more detailed the classifications of fault types are, the lower the accuracy rate of fault diagnosis is, and vice versa. However, too rough classification is not conducive to the fault diagnosis of a power transformer, and it is difficult to meet the requirements of engineering application.

In response to the deficiencies of the above traditional approaches, artificial intelligence (AI) techniques about power transformer fault diagnosis have attracted considerable attention owing to their high flexibility and powerful fault diagnosis performance (e.g., expert system (EPS), fuzzy theory, support vector machine (SVM), extreme learning machine (ELM), and artificial neural network (ANN)). EPS is a smart computer program system combined with expert experience, which can diagnose faults more comprehensively, accurately, and quickly. For example, Ma et al. [20] developed an EPS for power transformer insulation fault diagnosis, which took DGA as the characteristic parameter. The diagnosis results showed that this designed EPS can comprehensively analyze the insulation status of a transformer and identify the type of fault correctly. Mani and Jerome [21] presented an intuitive fuzzy EPS to diagnose power transformer faults, such that the estimation of key–gas ratio in the transformer oil can become simpler. Fuzzy theory mainly studies the interrelationship among fuzzy things, so it can deal well with these issues with fuzziness and uncertainty. For example, Huang and Sun [22] used fuzzy logic combined with dissolved gas of mineral oil for power transformer fault diagnosis. Empirical results demonstrated that the most effective fault diagnosis technique was to combine outputs from various DGA diagnostic methods and to aggregate them into an overall evaluation. Velásquez and Lara [23] established the intelligent diagnosis system based on principal component analysis (PCA) and adaptive decision system based on fuzzy logic permits to predict incipient fault diagnosis of power transformers. SVM is an AI technique based on statistical learning theory which possesses great advantages in nonlinear problems. Bacha et al. [24] investigated a novel extension method in which a SVM was applied to diagnose the power transformers faults and to choose the most appropriate gas signature between the DGA traditional methods and a novel extension method. The test results indicated that the novel extension method and the SVM approach can significantly improve the diagnosis accuracies for power transformer fault classification. Fei and Zhang [25] proposed an optimized model combining SVM with a genetic algorithm (SVMG) to diagnose power transformer faults. The experimental results indicated that the SVMG method can achieve higher diagnostic accuracy than IEC three ratios, normal SVM classifier, and artificial neural network. ELM is an emerging learning algorithm which has been introduced to transformer fault
diagnosis in recent years. Malik and Mishra [26] applied ELM combined with PCA to classify the incipient faults of power transformers and compared its performance with fuzzy-logic and ANN. The compared results showed that ELM can provide better diagnosis results. Yuan et al. [27] put forward an integrated PSO and ELM method to diagnose power transformer faults.

However, these diagnosis methods discussed above have their inherent drawbacks as follows: (1) For EPS, a complete knowledge base is the paramount factor to ensure the accuracy of diagnosis. However, it is difficult to obtain a complete knowledge base. In addition, EPS generally has a poor learning ability. (2) Fuzzy theory is difficult to determine appropriate membership function between the input and output variables [28]. (3) SVM is essentially a two-classification algorithm, which makes it difficult to construct a learning machine, select kernel functions, and determine parameters in multi-classification problems. As a consequence, SVM has the intrinsic deficiency of low classification efficiency [29,30]. (4) The performance of ELM is not stable since its hidden layer parameter is randomly chosen. Relative to the above fault diagnosis methods, the neural network has more extensive application in fault diagnosis of power transformers due to its simplicity, strong nonlinear-fitting ability, and high precision. For example, Rigatos and Siano [31] proposed the neural-fuzzy network for the detection of incipient faults in power transformers and the performance of the proposed methodology was tested through simulation experiments. Souahlia et al. [32] presented a comparative study for the choice of the most appropriate multi-layer perceptron (MLP) neural network model by comparing two output data types and three hidden layer types. The test results suggested that MLP neural network ratios combination can generalize better than other MLP neural network models. Yang et al. [33] presented a machine learning-based approach to power transformer fault diagnosis based on dissolved gas analysis (DGA), a bat algorithm (BA), and optimizing the probabilistic neural network (PNN). The performance of an ANN-based DGA method had been compared with the Rogers ratios method, and it was found that the proposed ANN-based method detected more accurately. The BPNN model is the most popular one among various neural network algorithms and has been widely employed in different fields of fault diagnosis, such as power electronic system [34], transformer [35], battery [36,37], photovoltaic systems [38,39], etc. However, the BPNN model still has some intrinsic defects, for example, slow convergence speed and over-fitting problem [40–43]. Fortunately, a large collection of optimization algorithms have been developed to optimize the BPNN model, such as GA [44,45], MEA [46], particle swarm optimization (PSO) [47,48], simulated annealing (SA) [49], bat algorithm (BA) [50,51], etc. Among them, evolutionary algorithms, such as GA and MEA, have recently been widely used as optimization algorithms searching the optimal weights and thresholds of neural networks. Thus, this paper employs GA and MEA to optimize the weights and thresholds of the BPNN model. The BPNN model optimized by GA or MEA still has some defects to overcome. For instance, its accuracy rate in terms of transformer fault diagnosis has not reached a satisfactory level. Furthermore, it does not handle the issues of noise samples and small sample data well.

In light of the above analysis, this paper proposes a novel method called the EGML model that employs BPNN as the benchmark learning model, with GA and MEA as the optimization algorithms. Its core lies in employing the modular and mathematical expertise to guide specific learning algorithms to improve learning ability. In this model, a learning machine will be seized of a little basic expertise before mining the data rules so that it can possess a better learning environment and have a higher possibility to learn along the right path. Specifically, the introduction of expertise can provide additional information for each individual, diversify the movement of each individual and enhance each individual’s searching and exploration capability to avoid falling into local optimum. Thereby, a higher level of intelligence technique may be realized. In addition, the expertise embedded in this model will bring some corrections to the prediction results. Thus, the EGML can handle these issues with noise sample.

The contributions and innovations in this paper include:

(1) A novel model called the EGML model, which aims to embed expertise into AI techniques, where the BPNN model is selected as the benchmark learning model, while GA and MEA are used as
optimization algorithm. Although BPNN and its improved models have been widely built to
diagnose the faults of power transformers in the past, works on employing the BPNN model
optimized by MEA are less likely to be found in literatures.

(2) The introduction of a knowledge function to provide additional information for each individual,
diversify the movement of each individual and enhance each individual’s searching and
exploration capability to avoid falling into local optimum. In addition, the expertise embedded
in this model will bring some corrections to the prediction results. It can provide scholars with a
new perspective when diagnosing the state of power transformers.

(3) Considering the difference of the data collected in reality, to further investigate the robustness of
the proposed models for different sample data, four scenarios are simulated in this paper. Then,
the proposed models are compared with the BPNN model and its optimization models in the
diagnosis performance under four scenarios. This EGML model is not sensitive to noise data and
training sample size, indicating that the model can handle the problems of noise samples and
small sample size well. In addition, this model proves to be a powerful tool in the fault diagnosis
of power transformers through comparative experiment among other diagnosis models. The
proposed model employs the advantages of each single model and overcomes the deficits of the
single model.

The paper is organized as follows: Section 2 outlines the methodology. Section 3 details the
performance of the fault diagnosis model. Finally, Section 4 draws the conclusion of this study.

2. Methodology

This section aims to introduce the structure of the EGML model, in which the BPNN model is the
benchmark learning model and GA and MEA are selected as the optimization algorithm.

2.1. Benchmark Learning Model: The BPNN Model

A three-layer BPNN model is developed in this paper due to its powerful self-learning ability
and generalization ability. It is assumed that there are \( m \) input layer nodes, \( n \) output layer nodes and \( l \)
hidden layer nodes. \( a \) and \( b \) respectively refer to the hidden layer threshold and output layer threshold.
\( w_{ij} \) and \( w_{jk} \) are the connection weights between the input layer and the hidden layer and between the
hidden layer and the output layer. On this basis, the output \( H_j \) of the \( j \)th neuron in hidden layer is
as follows:

\[
H_j = f\left(\sum_{i=1}^{m} w_{ij} x_i - a_j\right) j = 1, 2, \cdots, l
\]

where \( f \) is the activation function of the hidden layer. The function selected in this paper is as follows:

\[
f(x) = \frac{1}{1 + e^{-x}}
\]

The prediction value of the BPNN model is calculated according to the output \( H_j \) of hidden layer,
the connection weight \( w_{jk} \) and the threshold \( b \):

\[
O_k = f\left(\sum_{j=1}^{l} H_j w_{jk} - b_k\right) k = 1, 2, \cdots, n
\]

Finally, the prediction error of network is calculated:

\[
E_k = \sum_{k=1}^{n} (Y_k - O_k)^2 k = 1, 2, \cdots, n
\]

where \( O \) refers to the prediction value and \( Y \) is the expected value.
2.2. Parameter Optimization Algorithm

2.2.1. Genetic Algorithm

GA is a parallel stochastic search algorithm inspired by the genetic mechanism which imitates nature and biological evolution theory. There are three main operations in GA, that is, selection, crossover, and mutation.

Selection operation refers to the process in which several individuals are selected from the original group into the new group with a designate probability. The probability of individuals being selected is related to the fitness value. The better the fitness value is, the greater the probability of individuals being selected.

Crossover operation is to select two chromosomes from a group and randomly select one or more chromosome locations for transformation. Supposing \( c_1 \) and \( c_2 \) are two original chromosomes and \( \lambda \) refers to an independently distributed random variance between \([0, 1]\), two new chromosomes will be produced as follows:

\[
\begin{align*}
    c_{1,\text{new}} &= \lambda \times c_1 + (1 - \lambda) \times c_2 \\
    c_{2,\text{new}} &= \lambda \times c_2 + (1 - \lambda) \times c_1
\end{align*}
\]  

(5)

The mutation operation refers to selecting an individual from a group and then selecting a gene from the chromosome in this individual for mutation to produce a better individual. The mutation operation of the gene \( g_{ij} \) in a chromosome is performed as:

\[
 g_{ij,\text{new}} = \begin{cases} 
    g_{ij} + (g_{ij} - g_{\text{max}}) \times f(N), & r > 0.5 \\
    g_{ij} + (g_{\text{min}} - g_{ij}) \times f(N), & r \leq 0.5 \\
    f(N) = \delta \times (1 - N / N_{\text{max}})^2 
\end{cases}
\]

(6)

where \( g_{\text{min}} \) and \( g_{\text{max}} \) refer to lower limit and the upper limit of the gene \( g_{ij} \), respectively, \( N \) is the current iterative step while \( N_{\text{max}} \) refers to the whole iteration steps, and both \( r \) and \( \delta \) are random values with the range \([0, 1]\).

According to the selected fitness function and the operations of selection, crossover, and mutation in genetic mechanism, the individuals with a good fitness value are retained and the other individuals are eliminated. The new individuals not only inherit the information of the previous generation, but are also superior to the previous generation. This cycle is repeated until the conditions are met.

2.2.2. Mind Evolutionary Algorithm

MEA is put forward by Sun et al. [52] in order to overcome the defects of the GA. The former generally outperforms the latter in the convergent speed and the diagnosis performance. In the MEA, the similar-taxis and the dissimilation operations are proposed to displace the crossover and mutation operations in the GA. The structure of MEA can be seen in Figure 1. On this basis, the specific contents are explained as follows:

(a) MEA is an iteration optimization method. In the evolution process, all individuals of each generation constitute a group and one group can be divided into several subgroups. The group has a global bulletin board while each subgroup has a local bulletin board. The bulletin board is equivalent to an information platform, providing a place for information exchange between individuals and subgroups. Each subgroup is composed of several individuals, and each individual corresponds to one score for guiding the subgroup evolution process.

(b) Determining the constituent elements of an individual is extremely significant before generating the individual. In the individual \( X = [x_1, x_2, \ldots, x_{N-1}, x_N] \), \( x_1, \ldots, x_{N-1} \) represents the weights and thresholds of the BPNN, and \( x_N \) represents the score of the individual. The number of
the elements of an individual is determined by the number of the nodes in each layer. The details are as follows:

\[ N = S_1 \times S_2 + S_2 \times S_3 + S_2 + S_3 + 1 \]  

(7)

where \( S_1, S_2 \) and \( S_3 \), respectively, are the numbers of input layer neurons, of hidden layer neurons and of output layer neurons in BPNN.

After determining the number of elements of the individual, the random numbers between \([-1, 1]\) generated by the random function in MATLAB are assigned to the individual \( X \). Afterwards, the individual production process is basically completed. Then several superior and temporary individuals will be searched for according to individual scores.

(c) Some new individuals will be created around these superior and temporary individuals, respectively, and these individuals constitute several subgroups. Then these subgroups are divided into superior and temporary subgroups in light of their scores. Finally, superior subgroups will be posted to the global bulletin board.

(d) At the stage of local competition, the similar-taxis operation will be performed. The similar-taxis operation refers to the process in which individuals compete with each other and then superior individuals are produced. The individual with the highest score is selected as the superior individual by comparing the scores of each individual in the subgroup continuously. Subsequently, this superior individual is employed as a center to generate new subgroups. The similar-taxis operation ends when a new optimal individual is no longer generated. Here, the score of the subgroup is equal to that of the optimal individual. The aim of a similar-taxis operation is to mature the subgroup.

(e) At the stage of global competition, the dissimilation operation will be implemented. If the score of a mature superior subgroup is lower than the score of a temporary subgroup, it will be replaced by the winning temporary subgroup and its individuals are also released. Contrarily, if the score of a temporary subgroup is less than the scores of all superior subgroups, the temporary subgroup will be discarded and its individuals also released. In the meantime, a new temporary subgroup will be produced to displace the original one to ensure the total numbers of subgroups.

(f) The optimal individual is outputted until the number of iterations is met or the global optimal individual is determined. If the above conditions are not met, the algorithm will return to step (a).

---

**Figure 1.** The mechanism of the mind evolutionary algorithm (MEA) algorithm.
2.3. Expertise-Guided Machine Learning Model

The EGML model refers to the model which uses modular and mathematical expertise to guide the learning machine to train the sample data, thereby improving learning performance. The expertise put forward here refers to the basic knowledge, rules, or experience of the problem being studied, rather than the assumptions given in the traditional machine learning algorithm (e.g., the distribution of data, the prior probability, posterior probability, maximum likelihood and least square error, etc.). The former guides the sample to train in the learning process with no factitious hypotheses, achieving the transformation from hypothesis and statistical reasoning to the generalization of statistical reasoning based on knowledge guidance, and integrating knowledge analysis with data mining. It takes into account the robustness of the generalization performance of deterministic expertise and the flexibility of using training data to mine new uncertain rules. However, the latter is a hypothetical condition added subjectively by people to expand the information of the sample set.

In the EGML model proposed in this paper, expertise should possess at least the following five characteristics:

(1) Guidance. Although it does not solve the problem directly, it can partially guide the data or information;
(2) Application. The rule of expertise proposed in this paper can be applied to solve practical problems;
(3) Promotion. The knowledge obtained in practice can be reinforced or revised in the future to make it more precise;
(4) Expression. Expertise can be modular and mathematical;
(5) Openness. It is able to accept information about changes in the outside world.

Knowledge with the above characteristics is the prerequisite for applying the EGML model. For expertise, the proposed model does not need to impose strict limits on its completeness since it is only required to possess the features of universality and promotion to guide the directional mining of the sample rules.

Figure 2 depicts the framework of the EGML model. From it, the main steps of the EGML model are as follows:

Step 1. Employ the knowledge container $c$ to model the expertise $k$ so that it can be called directly in the loop learning iteration.

Step 2. Construct the knowledge function $KF(x)$.

The parameter optimization of traditional machine learning refers to the process in which the learning algorithm $A$ is utilized to obtain a specific hypothesis $h$ for simulating the training set $D$. Hypothesis $h$ will be called the optimal hypothesis when it can fit $D$ most appropriately. That is, $h$ is selected essentially from the hypothesis set $H$ in the light of the criterion of the lowest cost function. Besides, different combinations of learning algorithms and corresponding hypothesis sets constitute various machine learning algorithms. However, in the EGML model, cost function will be replaced with knowledge function ($KF(x)$) for the purpose of realizing the guidance of expertise on sample data.

Supposing $x = (x_1, x_2, \cdots, x_m, \cdots x_n)$, $m \leq n$:

$$KF(x) = \frac{1}{n} \sum \alpha_i e_i + \sum_{m=1}^{M} \sum_{i=1}^{n} \beta_{m,i} d(i)$$

$$d(i) = \begin{cases} 
0, & t(i) \in T(i) \\
1, & t(i) \notin T(i) 
\end{cases}$$

where $m$ refers to the total number of modular knowledge and $n$ is the number of training samples; $\alpha_i$ is the weight corresponding to the $i$th sample; $e_i$ expresses the diagnosis error of the $i$th sample. Furthermore, $\beta_{m,i}$ is the weight of the $i$th sample corresponding to the $m$th knowledge, $d_i$ is the logical
judgment value of the \(i\)th sample, and \(T(i)\) is the feasible domain given by the \(i\)th sample based on knowledge.

\[
\beta_{m,i}d(i) = \sum_{m=1}^{M} \sum_{i=1}^{n} \beta_{m,i}d(i)
\]

According to Equation (8), this paper constructs a specific knowledge function introducing 
\[
\sum_{m=1}^{M} \sum_{i=1}^{n} \beta_{m,i}d(i)
\]
on the basis of the traditional cost function, so that the qualitative knowledge will be modeled and used to guide the learning algorithm. The introduction of 
\[
\sum_{m=1}^{M} \sum_{i=1}^{n} \beta_{m,i}d(i)
\]in Equation (8) aims to improve the robustness of the proposed model via providing additional information for each individual; it will diversify the movement of each individual and enhance each individual’s searching and exploration capability, avoiding premature convergence.

Step 3. Construct the EGML model and experiment repeatedly in order to determine its structure.

Step 4. Determine the fitness function according to the knowledge function.

\[
F(x) = 1/(KF(x))
\]

Step 5. Repeat the selection, crossover and mutation operation of the GA as well as the similar-taxis and the dissimilation operation of MEA to respectively select the optimal individuals for the GA and MEA. Then output the optimal individuals until the number of iterations is met or the global optimal individuals are searched.

Step 6. Obtain the optimal weights and thresholds of the EGML model via decoding the above optimal individual and employ them to train the EGML model.

Step 7. Apply the trained learning model to test samples.
3. Case Study

3.1. Experimental Design

3.1.1. Data Description

Faults of power transformers can be divided into two types, namely, internal faults and external faults. The former has been the emphasis of research for its high probability of occurrence and high degree of damage. There are three main types of internal faults, that is, mechanical faults, overheat faults, and discharge faults, and the latter two are dominant. Once there is an overheating or discharge fault inside the transformer, the insulating oil will be decomposed into several dissolved gases. Different transformer faults will produce different dissolved gas components. Thus, these dissolved gases can be considered vital indicators to diagnose power transformer faults. This paper collected 310 sets of real dissolved gas data from [53] for research after eliminating redundant samples and some singular value. Among them, H\textsubscript{2}, CH\textsubscript{4}, C\textsubscript{2}H\textsubscript{6}, C\textsubscript{2}H\textsubscript{4}, and C\textsubscript{2}H\textsubscript{2} were selected as diagnosis indicators and their gas concentration ratios illustrated in Equation (11) were used as inputs. From Table 1, it can be seen that there are six common fault types occurring in the transformer oil, which are considered as outputs in this paper. Here, the size of the power transformer used for research is 120 MVA/220 kV.

$$GCR = \left( \frac{C_2H_2}{C_2H_4}, \frac{CH_4}{H_2}, \frac{C_2H_6}{C_2H_4} \right)$$ (11)

where GCR refers to three gas concentration ratios.

Table 1. Fault type and sample size of power transformer.

<table>
<thead>
<tr>
<th>Fault Type</th>
<th>Abbreviation</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>High thermal fault</td>
<td>HT</td>
<td>75</td>
</tr>
<tr>
<td>Medium thermal fault</td>
<td>MT</td>
<td>49</td>
</tr>
<tr>
<td>Low thermal fault</td>
<td>LT</td>
<td>17</td>
</tr>
<tr>
<td>Partial discharge</td>
<td>PD</td>
<td>20</td>
</tr>
<tr>
<td>High energy discharge</td>
<td>HD</td>
<td>75</td>
</tr>
<tr>
<td>Low energy discharge</td>
<td>LD</td>
<td>74</td>
</tr>
</tbody>
</table>

3.1.2. Knowledge Representation

Considering the five characters of the knowledge and the actual fault condition of power transformer, this paper extracted six pieces of knowledge:

$$k : \Leftrightarrow \{ \text{if } v_1 < \delta_1 \parallel v_3 \geq \delta_3 \text{ then } P(HT) \geq \alpha_1 \}$$ (12)

This piece of knowledge means the probability of high thermal fault will be not less than \( \alpha_1 \) when \( v_1 \) is below its threshold while \( v_3 \) is not less than its threshold.

$$k : \Leftrightarrow \{ \text{if } v_1 < \delta_1 \parallel v_2 \geq \delta_3 \parallel \delta_2 \leq v_3 < \delta_3 \text{ then } P(MT) \geq \alpha_2 \}$$ (13)

This piece of knowledge indicates the probability of medium thermal fault will be not less than \( \alpha_2 \) when \( v_1 \) is below its threshold, and \( v_2 \) is not less than its threshold while \( v_3 \) is between its thresholds.

$$k : \Leftrightarrow \{ \text{if } v_1 < \delta_1 \parallel \delta_1 \leq v_2 < \delta_3 \parallel \delta_1 \leq v_3 < \delta_3 \text{ then } P(LT) \geq \alpha_3 \}$$ (14)

In this piece of knowledge, the probability of low thermal fault will be not less than \( \alpha_3 \) when \( v_1 \) is below its threshold, and \( v_2 \) and \( v_3 \) are between their respective thresholds.

$$k : \Leftrightarrow \{ \text{if } v_1 < \delta_1 \parallel v_2 < \delta_1 \text{ then } P(PD) \geq \alpha_4 \}$$ (15)
This piece of knowledge suggests the probability of partial discharge will be not less than $\alpha_4$ when $v_1$ and $v_2$ are lower than their respective thresholds.

$$k \iff \{\text{if } \delta_1 \leq v_1 < \delta_3 \text{ then } P(\text{HD}) \geq \alpha_5\}$$ (16)

In this piece of knowledge, the probability of high energy discharge will be not less than $\alpha_5$ when $v_1$ is between its thresholds.

$$k \iff \{\text{if } v_1 > \delta_3 \text{ then } P(\text{LD}) \geq \alpha_6\}$$ (17)

In this piece of knowledge, the probability of low energy discharge will be not less than $\alpha_6$ when $v_1$ exceeds its threshold. where $v_1, v_2$ and $v_3$ are the ratio of the volume fraction of $C_2H_2$ to $C_2H_4$, $CH_4$ to $H_2$, and $C_2H_4$ to $C_2H_6$, respectively; $\delta_1, \delta_2$ and $\delta_3$ are 0.1, 1 and 3, respectively; and $\alpha_1 - \alpha_6$ represent the possibility with the range of $[0, 1]$.

The implied meaning of the above knowledge is that the three gas concentration ratios contribute differently to the fault diagnosis of a power transformer, namely, $v_1 > v_2 > v_3$, which is in line with the actual situation of the transformer. In addition, all the above formal representations of knowledge contain threshold $\delta$th. Previously, an accurate value would be assigned to the threshold. However, in many practical problems, the assigned threshold is subjective, uncertain, and even controversial to a certain extent. The method proposed in this paper does not need to impose strict limits on the threshold since the knowledge the EGML model requires possesses the features of universality and promotion to guide the directional mining of the sample rules.

### 3.1.3. Parameter Setting and Model Performance Evaluation

For demonstrating the performance of the proposed model in transformer fault diagnosis, we collected 310 real fault data, among which 230 samples were used as the training set and the remaining 80 samples were used as the test set. Three gas concentration ratios were employed as inputs of the proposed model, and the corresponding input layers were three layers. In addition, the output layer corresponds to the fault type layer, which includes six layers. In this paper, a computer equipped with an Intel® Core™ i3-2350M processor CPU @ 2.30 GHz, 2 GB RAM and a 32 bit Windows 7 operating system (OS) was utilized. Additionally, this paper used MATLAB R2015a to write all programs. The specific parameters selected for the EGML model can be seen in Table 2.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance function</td>
<td>$KF(x)$</td>
<td>Probability of mutation</td>
<td>0.01</td>
</tr>
<tr>
<td>The number of iterations</td>
<td>1000</td>
<td>Size of population</td>
<td>500</td>
</tr>
<tr>
<td>Learning rate</td>
<td>0.1</td>
<td>Number of superior groups</td>
<td>5</td>
</tr>
<tr>
<td>Size of chromosomes</td>
<td>200</td>
<td>Number of temporary groups</td>
<td>5</td>
</tr>
<tr>
<td>Probability of crossover</td>
<td>0.25</td>
<td>Number of hidden layer neurons</td>
<td>10</td>
</tr>
</tbody>
</table>

In order to verify the comprehensive performance of the proposed model and facilitate the comparison with other models, this paper employed accuracy rate (AR) to evaluate the model, expressed by the following specific equation:

$$AR = \frac{SN}{TN} \times 100\%$$ (18)

where $SN$ represents the number of test samples where the prediction fault type is the same as actual fault type and $TN$ refers to the total number of test samples.
3.2. Results and Discussion

3.2.1. Diagnosis Results Analysis

To evaluate the diagnosis performance of the proposed model, we constructed three groups of comparative experiments. The specific results of the three groups are illustrated in Figures 3–5, where the horizontal axis represents the sample number of the test set, and the vertical axis represents the diagnosis error (namely, the difference between the prediction value and the actual value). The corresponding comparisons of AR for various diagnosis models are shown in Table 3. Specifically, Figure 3 shows the results of the fault diagnosis experiments conducted with the BPNN model, the GA-BPNN model, and the GA-EGML model. From Figure 3, it can be seen that the GA-EGML model produced the least number of misdiagnosis, followed by the GA-BPNN model and the BPNN model, illustrating that the former was better than the latter two. Correspondingly, Figure 4 shows the results of fault diagnosis experiments conducted with the BPNN model, the MEA-BPNN model, and the MEA-EGML model, in which the comparison result was similar to the circumstances in Figure 3. The performance of the MEA-EGML model was superior to the BPNN model and the MEA-BPNN model. Figure 5 shows the results of fault diagnosis experiments conducted between the GA-EGML model and the MEA-EGML model. It can be seen from Figure 5 that the hybrid MEA-EGML model outperformed GA-EGML in diagnosis accuracy. In addition, from the AR given in Table 3, it can be seen that all hybrid models had higher diagnosis accuracy compared to the single BPNN model and identified transformer faults more effectively. However, the accuracy of the EGML models were higher than that of the other hybrid models since the self-learning ability and robustness of the BPNN model had been greatly improved. Moreover, the MEA-EGML model had a higher accuracy rate than the GA-EGML model. The specific analyses of the experimental results are as follows.

![Figure 3](image-url)  
**Figure 3.** Fault diagnosis errors of the back propagation neural network (BPNN) model, the GA-BPNN model and the GA-EGML model.
For a more intuitive view of the diagnosis performance of different model, the AR of each model was calculated and is displayed in Table 3. From Table 3, some analyses have been illustrated as follows:

(a) When comparing the hybrid GA-EGML model with the hybrid MEA-EGML model, the performance of the latter was superior to the former. The MEA-EGML model improved AR by 2.70% on the basis of the GA-EGML model, which means the MEA is more suitable to optimize the EGML model than GA. The reason for this phenomenon is that the MEA does not have problems of local optimization and prematurity which sometimes appear in the GA. Furthermore, the MEA has a significant advantage in that it can remember the optimal value of each iteration so that the global optimum can be acquired quickly.

(b) When comparing the hybrid GA-EGML model with the single BPNN model, the former improved the performance of the latter by 8.82%.

(c) When comparing the hybrid MEA-EGML model with the single BPNN model, the former significantly improved the performance of the latter; the improvement percentage in terms of AR was 11.76%.

(d) When comparing the proposed GA-EGML model with the hybrid GA-BPNN model, the former improved the performance of the latter obviously; the improvement percentage in terms of AR was 5.71%.

(e) When comparing the proposed MEA-EGML model with the hybrid MEA-BPNN model, the former improved AR by 7.04% on the basis of the latter.

The results given in (b)–(e) indicate that both the GA-EGML model and the MEA-EGML model have improved the fault diagnosis performance of the single BPNN model and its optimization models considerably. In the EGML models, knowledge function is employed to replace the cost function of the BPNN model, providing additional information for each individual, enhancing each individual’s searching behavior and exploration capability, and avoiding premature convergence. In addition, it can also bring some corrections to the fault diagnosis results. Thus, a higher AR can be obtained by the EGML model.

**Table 3. Performances of fault diagnosis of various models.**

<table>
<thead>
<tr>
<th>Indexes</th>
<th>BPNN</th>
<th>GA-BPNN</th>
<th>GA-EGML</th>
<th>MEA-BPNN</th>
<th>MEA-EGML</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR (%)</td>
<td>85.00</td>
<td>87.50</td>
<td>92.5</td>
<td>88.75</td>
<td>95.00</td>
</tr>
</tbody>
</table>

**Figure 4.** Fault diagnosis errors of the BPNN model, the MEA-BPNN model and the MEA-EGML model.

**Figure 5.** Fault diagnosis errors of the GA-EGML and the MEA-EGML model.
For a more intuitive view of the diagnosis performance of different models, the AR of each model was calculated and is displayed in Table 3. From Table 3, some analyses have been illustrated as follows:

(a) When comparing the hybrid GA-EGML model with the hybrid MEA-EGML model, the performance of the latter was superior to the former. The MEA-EGML model improved AR by 2.70% on the basis of the GA-EGML model, which means the MEA is more suitable to optimize the EGML model than GA. The reason for this phenomenon is that the MEA does not have problems of local optimization and prematurity which sometimes appear in the GA. Furthermore, the MEA has a significant advantage in that it can remember the optimal value of each iteration so that the global optimum can be acquired quickly.

(b) When comparing the hybrid GA-EGML model with the single BPNN model, the former improved the performance of the latter by 8.82%.

(c) When comparing the hybrid MEA-EGML model with the single BPNN model, the former significantly improved the performance of the latter; the improvement percentage in terms of AR was 11.76%.

(d) When comparing the proposed GA-EGML model with the hybrid GA-BPNN model, the former improved the performance of the latter obviously; the improvement percentage in terms of AR was 5.71%.

(e) When comparing the proposed MEA-EGML model with the hybrid MEA-BPNN model, the former improved AR by 7.04% on the basis of the latter.

The results given in (b)–(e) indicate that both the GA-EGML model and the MEA-EGML model have improved the fault diagnosis performance of the single BPNN model and its optimization models considerably. In the EGML models, knowledge function is employed to replace the cost function of the BPNN model, providing additional information for each individual, enhancing each individual’s searching behavior and exploration capability, and avoiding premature convergence. In addition, it can also bring some corrections to the fault diagnosis results. Thus, a higher AR can be obtained by the EGML model.

(f) When comparing the hybrid GA/MEA-BPNN model with the single BPNN model, both the hybrid GA-BPNN model and the hybrid MEA-BPNN model improved the fault diagnosis accuracy of the single BPNN model to some extent. However, MEA outperformed GA in promoting the diagnosis effect of the BPNN model since their respective improvement percentages were 2.94% and 4.41%. The reason for this phenomenon is that GA and MEA can select the optimal weights and thresholds for the single BPNN, which strengthens the global search ability of the single BPNN mode and prevents it from falling into local optimum. However, the MEA-BPNN model had a higher AR than the GA-BPNN model since the MEA has advantages over GA, as discussed in section (a).

The above results suggest that the proposed GA-EGML model and MEA-EGML model have better fault diagnosis performance than the single BPNN model and its optimization models. In addition, MEA is superior to GA whether in the EGML model or in a separate BPNN.

3.2.2. Robustness Test Analysis

Considering the difference of the data collected in reality, several groups of simulation experiments were conducted using the related data of power transformers for testing the application potential of the EGML model under different scenarios. In addition, the BPNN model, GA-BPNN model and MEA-BPNN model were also employed in simulation experiments for comparison. According to whether they contained noise data and whether they were equalized samples, etc., four typical application scenarios were constructed.

Scenario 1. The balanced sample set participated in training. There were 310 samples; 230 samples were randomly selected as the training set, and the rest were used as the test set.

Scenario 2. Noise samples were added for training. There were a total of 310 samples; 230 samples were randomly selected as the training set, containing 60 noise samples (i.e., the samples marked in error). In the training set, the numbers of noise samples assigned to the six fault types (i.e., HT, MT, LT,
PD, HD, and LD) were 14, 9, 2, 5, 14, and 16, respectively, in light of the number of each fault type in the training samples. The remaining 80 samples were used as the test set. Considering that there may have been distortion or error in the process of data collection, transmission, and disposing, testing the robustness of the EGML model in this scenario was necessary.

Scenario 3. The unbalanced sample set participated in training. There were 310 samples, among which 230 were training samples and 80 were test samples. The number of samples of various fault types in the training set was different, with the maximum discrepancy of six times. In the training set, the numbers of samples for various fault types (i.e., HT, MT, LT, PD, HD, and LD) were 60, 40, 10, 10, 50, and 60, respectively.

Scenario 4. Adjusted the number of training samples and test samples in the total sample. There were 310 samples; 160 samples were randomly selected as training set, and the remaining 150 were used as test set.

Scenario 1 was discussed in the previous section. To evaluate the fault diagnosis performances of different models under four scenarios, scenario 1 was considered as a control group. The comparisons of AR for different diagnosis are illustrated in Figure 6. As can be seen from Figure 6, the AR of all models in other scenarios declined compared to scenario 1, especially in scenario 2. In scenario 2, the traditional models, namely, the single BPNN model, the GA-BPNN model, and the MEA-BPNN model, decreased by 8.82%, 8.57%, and 7.04% in terms of AR, respectively. Among them, the AR of the BPNN model in scenario 2 only reached 77.5%, suffering the greatest decline, which had not met the accuracy requirements of power transformer diagnosis. Compared to them, the accuracy of the proposed GA-EGML model and MEA-EGML model was decreased slighter, with decrease rates of 4.05% and 3.95%, respectively. Different from scenario 2, the AR of all models in scenario 3 only decreased slightly and there was no obvious difference in terms of decrease amplitude between the traditional models and the EGML models. In scenario 4, the AR of each model also had a slight decline but the decline amplitude of the models proposed in this paper was lower compared to the traditional models. The reason for these phenomena is that the mechanism of knowledge function in the EGML model is different from that of the traditional cost function. The cost function only establishes a rule excavating mechanism for training data while the knowledge function extends a single learning mechanism based on data mining to the learning mechanism integrating training samples with expertise. Through knowledge function, the movement of the individual is diversified, and the exploration capability and searching behavior of each individual are strengthened, which can avoid premature convergence. Additionally, the EGML model depends on not only the mining of data regulars, but also the guidance of expertise, which brings some corrections to the prediction results. Thus, the impact of noise samples and small sample size on it is smaller than that on the traditional models. In the above four scenarios, both the GA-EGML model and MEA-EGML model obtained higher diagnosis accuracy, and the latter was better. Especially in scenario 2 and scenario 4, the proposed models were less affected by the sample data than the traditional model, which suggests the EGML model is more favorable compared to the traditional models when the number of sample data is limited or noise sample may exist. In engineering practice, it is difficult to obtain a large collection of sample data in many vertical areas, or the cost of data collection is too high. Even if there are ample samples, it is also difficult to ensure that no noise samples exist. Thus the EGML model is particularly beneficial to the fault diagnosis due to its powerful ability to deal with these problems of less sample data and noise samples in engineering practice.

3.2.3. Comparative Analysis of Different Diagnosis Models

To illustrate the diagnosis performance of the proposed model, this paper used scenario 1 for comparative experiments among different diagnosis models. In this section, the single IEC, softmax regression (SR), SVM, and ELM were constructed to diagnose the power transformer faults. In addition, BA and PSO were employed to optimize the single BPNN model for diagnosis. Fault diagnosis results of different models are shown in Figure 7. From Figure 7, it can be seen that the proposed models,
including both GA-EGML and MEA-EGML, yielded higher AR than other traditional models, which were 92.5% and 95%, respectively. However, the IEC has the worst diagnosis performance, and its AR only reached 72.5%, which does not meet the accuracy requirements of power transformer fault diagnosis. In addition, the AR of other models, except the IEC and the proposed model, are also not satisfactory, ranging from 81.25% to 87.5%.

Moreover, to compare performance discrepancy between the proposed model and other models, the improvement percentages were calculated. When comparing the IEC with the proposed model, the latter significantly improved the diagnosis performance of the former. For example, the AR improvement percentages of the GA-EGML model and the MEA-EGML model were 5.71% and 8.57%, respectively. The AR of the BA-BPNN model was the highest in other models apart from the proposed model. When comparing the BA-BPNN model with the proposed model, the latter also improved the diagnosis performance of the former to some extent. Correspondingly, the AR improvement percentages of the GA-EGML model and the MEA-EGML model were 5.71% and 8.57%,
respectively. It can be seen from the above discussion that the AR improvement percentages of the GA-EGML model based on other traditional diagnosis models were between 5.71% and 27.59%. Similarly, the AR improvement percentages of the MEA-EGML model were between 8.57% and 31.03%. The GA-EGML model or the MEA-EGML model both greatly improved the diagnosis performance of the traditional models. Therefore, the proposed method proves to be a satisfactory model for fault diagnosis of power transformer in the light of dissolved gas data.

4. Conclusions

In order to diagnose power transformer faults effectively, a novel diagnosis model called the EGML model was introduced, where the BPNN model was used as the benchmark learning model and the GA and the MEA were used as optimization algorithm. In this model, the knowledge function replaces traditional cost function, which provides additional information for each individual and brings some corrections to the prediction results. Then, the proposed models were compared with other models via real dissolved gas data. The following conclusions can be drawn from the present study:

(a) The proposed EGML model is superior to the traditional BPNN model and its optimization models, such as the GA-BPNN model and the MEA-BPNN model in terms of accuracy. Thus, the proposed EGML model can be a feasible tool for fault diagnosis. The empirical results are also expected to provide scholars with a new perspective when diagnosing the fault of a power transformer.

(b) Both the GA-EGML model and the MEA-EGML model are feasible for the fault diagnosis issue since they have favorable performance in accuracy and reliability. In terms of AR, the two types of EGML models are different, and the MEA-EGML model outperforms the GA-EGML model. The reason for this phenomenon is that the MEA does not have problems of local optimization and premature which sometimes appear in the GA.

(c) Considering the difference of the data collected in reality, several sets of simulation experiments were conducted for testing the performance of the EGML model under different scenarios. Empirical results indicate the accuracy of all models in other scenarios declined compared to scenario 1, especially in scenario 2. However, the decline amplitude of each model in terms of AR was different and the proposed EGML model decreased less than the traditional model in scenario 2 and in scenario 4. Thus, the proposed model can well handle practical problems of noise data and small sample size.

(d) Comparative analysis indicated that the proposed EGML model outperformed traditional models, such as IEC, SR, SVM, ELM, BPNN, and its optimization model in terms of accuracy and reliability. Specifically, the AR improvement percentages of the GA-EGML model based on other traditional diagnosis models were between 5.71% and 27.59%. Similarly, the AR improvement percentages of the MEA-EGML model were between 8.57% and 31.03%.

Although the proposed model has good performance in transformer fault diagnosis, there are still some problems to be solved. The fault data of power transformers is difficult to collect. Therefore, this paper only studies one size of power transformer. In future studies, power transformers with different sizes can be employed to verify the performance of the proposed model if the data is available. In addition, this paper only studies one application of the EGML model where the BPNN model is used as the benchmark learning model and more applications combining with other AI techniques will be studied in the next work.

Author Contributions: Q.W. designed this paper and provided overall guidance; H.Z. wrote the manuscript.

Funding: This research was funded by the National Social Science Fund, grant number 17BGL252 and the Humanities and Social Sciences Planning Foundation of the Ministry of Education of China, grant number 16YJA790052.

Acknowledgments: Chenyang Peng provided technical support for the experiment.

Conflicts of Interest: The authors declare no conflict of interest.
References


47. You, L.; Tan, Q.; Kang, Y.; Xu, C.; Chong, L. Reconstruction and prediction of capillary pressure curve based on Particle Swarm Optimization-Back Propagation Neural Network method. Petroleum 2018, 4, 268–280. [CrossRef]
53. Yao, H. *Power Transformer Fault Diagnosis Based on Improved Artificial Fish Algorithm-Rbf Network*; Guangxi University: Nanning, China, 2016.

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).