Fault Diagnosis of Oil-Immersed Power Transformers Using Kernel Based Extreme Learning Machine

Liwei Zhang

School of Electrical Engineering, Northeast Dianli University, Jilin 132012, China

Abstract

Transformer fault diagnosis based on conventional neural networks (NN) and support vector machines (SVM) face some drawbacks: slow learning speed, trivial human tuned parameters and difficult parameter determination. To overcome these drawbacks, transformer fault diagnosis based on kernel based extreme learning machine (KELM) was proposed in this paper and its performance was studied. The parameters of KELM are optimized by using particle swarm optimization (PSO), and then the optimized KELM is implemented for fault classification of power transformers. To verify its effectiveness, fault diagnosis of power transformers using KELM were compared with the other two ELMs, back-propagation neural network (BPNN) and support vector machines (SVM) on dissolved gas analysis (DGA) samples. Experimental results show that the proposed method is more stable, could achieve better generalization performance, and runs at much faster learning speed.

Key words: POWER TRANSFORMERS, FAULT DIAGNOSIS, DISSOLVED GAS ANALYSIS, EXTREME LEARNING MACHINE, PARTICLE SWARM OPTIMIZATION.

1. Introduction

Power transformers are considered as highly essential equipment of electric power transmission systems and often the most expensive devices in a substation. Failures of large power transformers can cause operational problems to the transmission system [1].

Dissolved gas analysis (DGA) is one of the most widely used tools to diagnose the condition of oil-immersed transformers in service. The ratios of certain dissolved gases in the insulating oil can be used for qualitative determination of fault types. Several criteria have been established to interpret results from laboratory analysis of dissolved gases, such as IEC 60599 [2] and IEEE Std C57.104-2008 [3]. However, analysis of these gases generated in mineral-oil-filled transformers is often complicated for fault interpretation, which is dependent on equipment variables.

With the development of artificial intelligence, various intelligent methods have been applied to improve the DGA reliability for oil-immersed transformers. Based on DGA technique, an expert system was proposed for transformer fault diagnosis and corresponding maintenance actions, and the test results showed it was effective [4]. By using fuzzy information approach, K. Tomsovic et al. [5] developed a framework that combined several transformer diagnostic methods to provide the "best" conclusion. Based on artificial neural network (ANN), Zhang et al. presented a two-step method for fault detection in oil-filled transformer, and the proposed approach achieved good diagnostic accuracy [6]. Wei-Song Lin et al. proposed a novel fault diagnosis method for power transformer based on Cerebellar Model Articulation controller (CMAC), and the new scheme was shown with high accuracy [7]. Yann-Chang Huang et al. presented a fault detection approach of oil-immersed power transformers based on genetic algorithm tuned wavelet networks (GAWNs) demonstrating remarkable diagnosis accuracy [8]. Weigen Chen et al. studied the efficiency of wavelet networks (WNs) for transformer fault detection using gases-in-oil samples, and the diagnostic accuracy and efficiency were proved better than these derived from

BPNN [9]. A fault classifier for power transformer was proposed by Zheng et al. based on multi-class least square support vector machines (LS-SVM) [10]. Xiong Hao et al. developed an artificial immune algorithm for transformer fault detection [11]. These novel diagnosis methods overcome the drawbacks of IEC method and improve the diagnosis accuracy.

However, conventional learning methods on neural networks such as back-propagation (BP) and SVM methods apparently face some drawbacks: (1) slow learning speed, (2) trivial human tuned parameters, and (3) trivial learning variants for different applications [12]. Extreme learning machine (ELM) is an emerging learning technique proposed for generalized single-hidden layer feed forward networks (SLFNs) [13]. ELM overcomes some major constraints faced by conventional learning methods and computational intelligence techniques. Similar to SVM, kernels can be applied in ELM as well [14-15]. Kernel based ELM (KELM) can be implemented in a single learning step, so it runs fast.

Therefore, fault diagnosis of power transformers using KELM is proposed in this paper. Similar to other kernel based methods, the parameters of KELM are usually assigned empirically or obtained by trials [15]. Obviously, it is very time-consuming and the performance achieved with the chosen parameters is suboptimal. The parameters of KELM are optimized by using PSO in this paper. This rest is organized as follows. Section 2 reviews original ELM, equality constrained-optimization-based ELM and KELM. In section 3, the proposed fault diagnosis using KELM is described in detail. Section 4 discusses the comparison results of the proposed method with other approaches. Conclusions are finally drawn in Section 5.

2. Kernel based ELM

2.1 original ELM

Extreme Learning Machine (ELM) was originally developed for the single-hidden layer feedforward networks (SLFNs) and then extended to the "generalized" SLFNs. The hidden layer in ELM need not be tuned. ELM randomly chooses the input weights and the hidden neurons' biases and analytically determines the output weights of SLFNs. Input weights are the weights of the connections between input neurons and hidden neurons and output weights are the weights of the connections between hidden neurons and output neurons.

The output function of ELM for generalized SLFNs (take one output node case as an example) is

$$f_L(\mathbf{x}) = \sum_{i=1}^{L} \beta_i h_i(\mathbf{x}) = \mathbf{h}(\mathbf{x})\boldsymbol{\beta} \qquad (1)$$

where $\boldsymbol{\beta} = [\beta_1, \cdots, \beta_L]^T$ is the vector of the output weights between the hidden layer of L nodes and the output node, and $\mathbf{h}(\mathbf{x}) = [h_1(\mathbf{x}), \cdots, h_L(\mathbf{x})]$ is the output (row) vector of the hidden layer with respect to the input x. $h(x)$ actually maps the data from the d-dimensional input space to the L-dimensional hidden-layer feature space (ELM feature space) **H**, and thus, $h(x)$ is indeed a feature mapping.

Given a set of training data $\{(\boldsymbol{x}_i, \boldsymbol{t}_i) | \boldsymbol{x}_i \in \mathbf{R}^d, \boldsymbol{t}_i \in \mathbf{R}^m, i=1,\ldots,N\}$. Different from traditional learning algorithms, ELM tends to reach not only the smallest training error but also the smallest norm of output weights

Minimize: $\quad \|\mathbf{H}\boldsymbol{\beta} - \mathbf{T}\|^2 \ \text{and} \ \|\boldsymbol{\beta}\|$ (2)

where H is the hidden-layer output matrix

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}(\mathbf{x}_1) \\ \vdots \\ \mathbf{h}(\mathbf{x}_N) \end{bmatrix} = \begin{bmatrix} h_1(\mathbf{x}_1) & \cdots & h_L(\mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ h_1(\mathbf{x}_N) & \cdots & h_L(\mathbf{x}_N) \end{bmatrix} \qquad (3)$$

and T is the expected output matrix

$$\mathbf{T} = \begin{bmatrix} \mathbf{t}_1^T \\ \vdots \\ \mathbf{t}_N^T \end{bmatrix} = \begin{bmatrix} t_{11} & \cdots & t_{1m} \\ \vdots & \ddots & \vdots \\ t_{N1} & \cdots & t_{Nm} \end{bmatrix} \qquad (4)$$

The minimal norm least square method was used in the original implementation of ELM

$$\boldsymbol{\beta} = \mathbf{H}^{\dagger}\mathbf{T} \qquad (5)$$

where $H^{\dagger}$ is the *Moore–Penrose generalized inverse* of matrix H.

The orthogonal projection method can be used to calculate the Moore–Penrose generalized inverse of H in two cases: when H^TH is nonsingular and $H^{\dagger}=(H^TH)^{-1}H^T$, or when HH^T is nonsingular and $H^{\dagger}=H^T(HH^T)^{-1}$.

2.2 KELM

According to the ridge regression theory, one can add a positive value to the diagonal of H^TH or HH^T; the resultant solution is more stable and tends to have better generalization performance.

For multiclass classifier with multi-outputs, classifiers with m-class have m output nodes. If the original class label is p, the expected output vector of the m output nodes is $\mathbf{t}_i = [0, \cdots, 0, \underset{p}{1}, 0, \cdots 0]^T$. In this case, only the pth element of $\mathbf{t}_i = [t_{i1}, \cdots, t_{im}]^T$ is one, while the rest of the elements are set to zero. For the constrained-optimization-based ELM with multi-output

node, the classification problem can be formulated as

Minimize:
$$L_{P_{ELM}} = \frac{1}{2}\|\boldsymbol{\beta}\|^2 + \frac{C}{2}\sum_{i=1}^{N}\|\boldsymbol{\xi}_i\|^2$$

Subject to:
$$\mathbf{h}(\mathbf{x}_i)\boldsymbol{\beta} = \mathbf{t}_i^{\mathrm{T}} - \boldsymbol{\xi}_i^{\mathrm{T}}, \quad i=1,\cdots,N \qquad (6)$$

where $\boldsymbol{\xi}_i = [\xi_{i1},\cdots,\xi_{im}]^{\mathrm{T}}$ is the training error vector of the m output nodes with respect to the training sample $\boldsymbol{x}_i$, C is the cost parameter.

Based on the Karush–Kuhn–Tucker (KKT) theorem, the output function of ELM classifier is

$$\mathbf{f}(\mathbf{x}) = \mathbf{h}(\mathbf{x})\boldsymbol{\beta} = \mathbf{h}(\mathbf{x})\mathbf{H}^{\mathrm{T}}\left(\frac{\mathbf{I}}{C} + \mathbf{H}\mathbf{H}^{\mathrm{T}}\right)^{-1}\mathbf{T} \qquad (7)$$

If a feature mapping $h(\boldsymbol{x})$ is unknown to users, one can apply Mercer's conditions on ELM. We can define a kernel matrix for ELM as follows:

$$\boldsymbol{\Omega}_{\mathrm{ELM}} = \mathbf{H}\mathbf{H}^{\mathrm{T}} : \Omega_{\mathrm{ELM}i,j} = h(\mathbf{x}_i)h(\mathbf{x}_j) = K(\mathbf{x}_i,\mathbf{x}_j) \qquad (8)$$

Then, the output function of ELM classifier (7) can be written compactly as

$$\mathbf{f}(\mathbf{x}) = \mathbf{h}(\mathbf{x})\mathbf{H}^{\mathrm{T}}\left(\frac{\mathbf{I}}{C} + \mathbf{H}\mathbf{H}^{\mathrm{T}}\right)\mathbf{T}$$
$$= \begin{bmatrix} K(\mathbf{x},\mathbf{x}_1) \\ \vdots \\ K(\mathbf{x},\mathbf{x}_N) \end{bmatrix}\left(\frac{\mathbf{I}}{C} + \boldsymbol{\Omega}_{\mathrm{ELM}}\right)^{-1}\mathbf{T} \qquad (9)$$

After ELM was trained, the given testing sample $\boldsymbol{x}$ was taken as the input of the classifier. The index of the output node with the highest output value is considered as the predicted class label of the given testing sample. Let $f_i(\boldsymbol{x})$ denote the output function of the ith output node, the predicted class label of sample $\boldsymbol{x}$ is

$$label(\mathbf{x}) = \arg\max_{i\in\{1,\cdots,m\}} f_i(\mathbf{x}) \qquad (10)$$

In this study, the popular Gaussian kernel function $K(\boldsymbol{u},\boldsymbol{v})=\exp(-\gamma\|\boldsymbol{u}-\boldsymbol{v}\|^2)$ is used as the kernel function in KELM. In order to achieve good generalization performance, the cost parameter C and kernel parameter γ of KELM need to be chosen appropriately. Similar to SVM and LS-SVM, the values of C and γ are assigned empirically or obtained by trying a wide range of C and γ. As suggested in [15], 50 different values of C and 50 different values of γ are used for each data set, resulting in a total of 2500 pairs of (C, γ). The 50 different values of C and γ are $\{2^{-24}, 2^{-23}, \cdots, 2^{24}, 2^{25}\}$. This parameters optimization method is called Grid algorithm.

3. Transformer fault diagnosis using KELM

There are five dissolved gases in oil-immersed transformers: Hydrogen (H_2), ethylene (C_2H_4), methane (CH_4), ethane (C_2H_6), and acetylene (C_2H_2), which are the byproducts caused by internal faults [3]. The technique of dissolved gas analysis (DGA) is effective in detecting incipient fault of power transformers. The attributes of each instance are normalized as {H2/T, CH4/T, C2H6/T, C2H4/T, C2H2/T}, where T represents the total gas.

The two main reasons of oil deterioration in operating transformers are thermal and electrical failures. Detectable faults in IEC Publication 60599 by using DGA are discharges of low energy (D1), discharges of high energy (D2), partial discharges (PD), thermal faults below 300ºC (T1), thermal faults above 300ºC (T2), and thermal faults above 700 ºC (T3).

Similar to SVM and LS-SVM, the generalization performance of KELM with Gaussian kernel are sensitive to the combination of (C, γ) as well when it is used for transformer fault diagnosis, shown as Figure 1
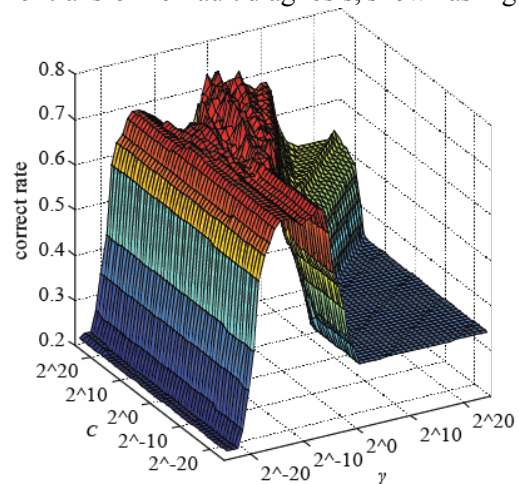

Figure 1. Performances of KELM with Gaussian kernel for transformer fault diagnosis
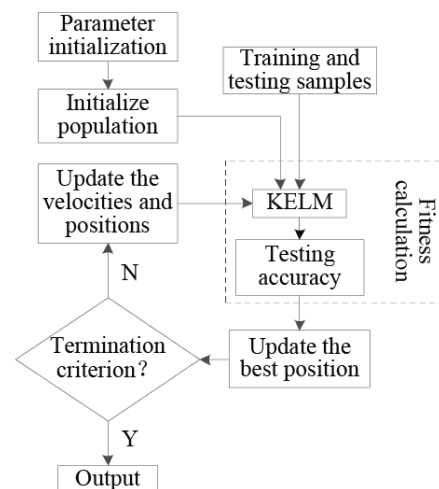

Figure 2. Fault diagnosis procedure using KELM

From practical point of view, it may be time consuming and tedious for users to choose appropriate kernel parameters (C, γ) by using Grid algorithm. In order to reduce time costs and achieve optimal generalization performance, the parameters in KELM with Gaussian kernel were optimized by using particle swarm optimization (PSO) [16] in this paper.

In PSO for parameters optimization, the dimension of searching space is D=2 corresponding to the two parameters (C, γ) of KELM with Gaussian kernel, and the position of each particle represents the parameter values of (C, γ) in Gaussian kernel. The aim of PSO for parameters optimization is to obtain the best generalization performance of KELM. Therefore, the correct rate of fault diagnosis can be taken as the fitness function of PSO.

The process of the proposed fault diagnosis using KELM is illustrated as Figure 2.

4. Experiment results and discussion

In this section, all simulations on DGA data set are carried out in MATLAB 7.6 environment running in Core 2 Duo 1.8GHZ CPU with 2GB RAM. For KELM with PSO in the experiment, the range of cost parameter C and kernel parameter γ were also [2^{-24}, 2^{25}] as mentioned in Section 2; population size was set to 24; maximum number of iterations (epochs) to train was set to 1000; acceleration constants c_1 and c_2 were set to 2; max particle velocity v_{max} was set to 2^{10}; initial inertia weight ω_{max} was set to 0.9 and final inertia weight ω_{min} was set to 0.4; epoch parameter T_e when inertial weight is at final value was set to 750. The training and testing data of all datasets are fixed for all trials of simulations.
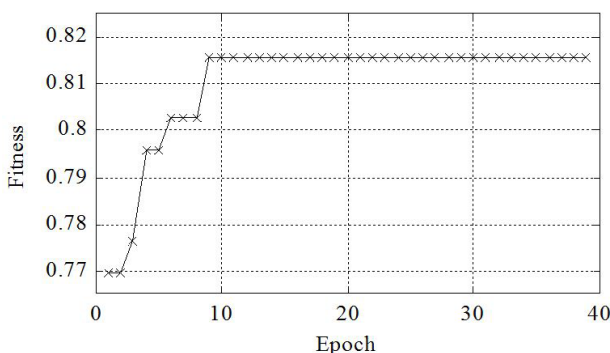

Figure 3. Fitness curve of KELM with PSO on DGA data set

In this study, 387 dissolved gas analysis (DGA) samples from real-world fault transformers are chose as experimental data. These samples were divided into two parts: 235 samples were taken as training data randomly and the rest 152 samples as testing data.

The fitness curve of PSO for KELM parameters optimization is shown in Figure 3.

It can be found from Figure 4 that the best fitness is obtained after 9 iterations. The best testing accuracy was 81.58% and the corresponding parameters (C, γ) were ($2^{6.4301}$, $2^{-4.3483}$). The computational time was 88.078 seconds. In comparison, the best testing accuracy obtained by Grid algorithm was 80.92%, (C, γ) = (2^5, 2^{-5}), and the computational time was 182.33 seconds.

The performance of KELM with PSO on DGA data set is compared with original ELM and ELM with Sigmoid additive node. The parameters (C, γ) of KELM were set to ($2^{6.4301}$, $2^{-4.3483}$) determined by using PSO. In original ELM, the number of hidden nodes L was chosen as 95 through trials with L ranging from 10 to 1000; the hidden nodes used the sigmoid type of activation function. In ELM with Sigmoid additive node, the cost parameter C and the number L of hidden nodes were set as (C, L)=(2^{24}, 700), optimized by using Grid algorithm with C ranging {2^{-24}, 2^{-23}, …, 2^{24}, 2^{25}} and L ranging from 10 to 1000. Testing accuracies of the three different ELMs on DGA data set with 50 trials are shown in Figure 4. The performance comparisons of the three methods are listed in Table 1.
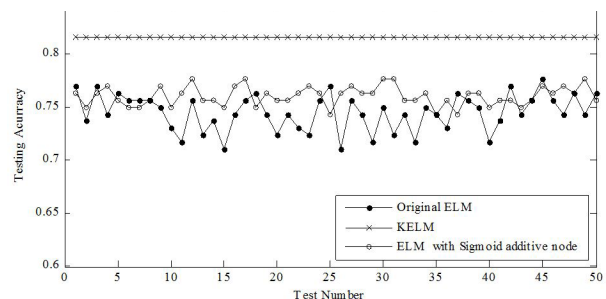

Figure 4. Testing accuracies of different ELMs on DGA data set

From Figure 4, it can be seen that the testing accuracies of original ELM and ELM with Sigmoid additive node are changing in each trial, while the testing accuracy of KELM are constant in all trial and higher than the other two ELMs.

From Table 1, it can be found that KELM requires less training and testing time than the other two ELMs while with the highest training and testing accuracies.

Moreover, the performance of the KELM is compared with other widely used diagnosis methods for power transformers, such as back-propagation neural network (BPNN) and support vector machines (SVM).

Table 1. Performance comparisons of different ELMs on DGA data set

Algorithms	Parameters	Values	CPU time (s)		Accuracy (%)	
			Training	Testing	Training	Testing ± Dev.
original ELM	L	95	0.0625	0.0313	91.91	74.51±1.74
KELM	(C, γ)	$(2^{6.4301}, 2^{-4.3483})$	**0.0097**	**0.0049**	**94.04**	**81.58±0**
ELM with Sigmoid additive node	(C, L)	$(2^{24}, 700)$	0.5156	0.0313	92.34	76.04±0.89

The training and testing samples from DGA data set were the same as Section 5.2 mentioned above. The BPNN used was of single-hidden-layer, provided in the neural networks tools box of MATLAB; the transfer function was tangent sigmoid; the number of hidden layer nodes was chosen by trials. For SVM, the parameters (C, γ) were selected by using Grid algorithm. The performance comparisons are listed in Table 2.

Table 2. Performance comparisons of different methods on DGA data set

Algorithms	Parameters	Values	CPU time (s)		Accuracy (%)	
			Training	Testing	Training ± Dev.	Testing ± Dev.
BPNN	L	28	3.0219	0.0281	78.13±2.34	69.84±2.66
KELM	(C, γ)	$(2^{6.4301}, 2^{-4.3483})$	**0.0097**	**0.0049**	**94.04±0**	**81.58±0**
SVM	(C, γ)	$(2^3, 2^6)$	0.0286	0.0057	91.91±0	79.61±0

From Table 2, it can be seen that the training time of KELM on DGA data set is far less than BPNN and SVM, and the training and testing accuracies are the highest. Obviously, the fault diagnosis approach based on KELM is stable and achieves better generalization performance than that based on BPNN and SVM.

5. Conclusions

In this paper, KELM has been presented for fault diagnosis of power transformers. The parameters of KELM are optimized by using PSO to improve the performance of KELM. Experimental results show that:

(1) Compared with Grid algorithm on nine benchmark classification data sets, KELM optimized by PSO achieves better performance and is less time-consuming.

(2) Compared with original ELM and ELM with Sigmoid additive node, KELM with PSO achieves better and more stable generalization performance in fault classification for power transformers.

(3) Compared with BPNN and SVM on DGA data set, KELM with PSO is able to obtain better diagnosis accuracy and runs faster.

References

1. W. H. Tang, Q. H. Wu (2011) *Condition monitoring and assessment of power transformers using computational intelligence*. Springer-Verlag: London.

2. (2007) *Mineral Oil-Impregnated Electrical Equipment in Service-Guide to the Interpretation of Dissolved and Free Gases Analysis*. IEC Publication 60599.

3. (2009) *Guide for the Interpretation of Gases Generated in Oil-Immersed Transformers*. IEEE Std C57.104-2008.

4. C. F Lin, J. M. Ling, C. L. Huang. (1993) An Expert System for Transformers Fault Diagnosis Using Dissolved Gas Analysis. *IEEE Transactions on Power Delivery*, 8(1), p.p.231-238.

5. K. Tomsovic, M. Tapper, T. Ingvarsson (1993) A Fuzzy Information Approach to Integrate Different Transformer Diagnostic Methods. *IEEE Transactions on Power Systems*, 8(3), p.p.1638-1646.

6. Y. Zhang, X. Ding, Y. Liu (1996) An artificial neural network approach to transformer fault diagnosis. *IEEE Transactions on Power Delivery*, 11(4), p.p.1836-1841.

7. Wei-Song Lin, Chin-Pao Hung, Mang-Hui Wang. (2002) CMAC-based Fault Diagnosis of Power Transformers. *Proceedings of the 2002 International Joint Conference on Neural Networks*, Honolulu, Hawaii, p.p.986-991.

8. Yann-Chang Huang. (2003) A New Data Mining Approach to Dissolved Gas Analysis of Oil-Insulated Power Apparatus. *IEEE Transactions on Power Delivery*, 18(4), p.p.1257-1261.

9. Weigen Chen, Chong Pan, Yuxin Yun, Yilu Liu (2009) Wavelet Network in Power Transformers Diagnosis Using Dissolved Gas Analysis. *IEEE Transactions on Power Delivery*, 24(1), p.p.187-194.

10. H. B. Zheng, R. J. Liao, S. Grzybowski, L. J. Yang (2011) Fault diagnosis of power transformers using multi-class least square support vector machines classifiers with particle swarm optimization. *Electric Power Applications*, 5(9), p.p.691-696.

11. X. Hao, S. Cai-Xin (2007) Artificial immune network classification algorithm for fault diagnosis of power transformer. *IEEE Transactions on Power Delivery*, 22(2), p.p.930-935.

12. G. B. Huang, Q. Y. Zhu, C. K. Siew (2006) Extreme learning machine: theory and applications. *Neurocomputing*, 70, p.p.489-501.

13. G. B. Huang, Q. Y. Zhu, C. K. Siew (2004) Extreme learning machine: a new learning scheme of feedforward neural networks. *Proceedings of the 2004 IEEE International Joint Conference on Neural Networks*, Budapest, Hungary, p.p.985-990.

14. G. B. Huang, C. K. Siew (2005) Extreme learning machine with randomly assigned RBF kernels. *International Journal of Information Technology*, 11(1), p.p.16-24.

15. G. B. Huang, H. Zhou, X. Ding, R. Zhang (2012) Extreme Learning Machine for Regression and Multiclass Classification. *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics*, 42(2), p.p.513-529.

16. Clerc M. Kennedy J. (2002) The particle swarm-explosion, stability, and convergence in a multidimensional complex space. *IEEE Transactions on Evolutionary Computation*, 6(1), p.p.58-73.

www.metaljournal.com.ua