A Membership-based Multi-dimension Hierarchical Deep Neural Network Approach for Fault Diagnosis

Liangliang Li
School of Computer Science and Technology
Tsinghua University
Beijing, China
Email: lll15@mails.tsinghua.edu.cn

Dai Guilan, Zhang Yong
Department of Computer Science and Technology and
Research Institute of Information Technology
Tsinghua University
Beijing, China
Email: {daigl, zhangyong05}@mail.tsinghua.edu.cn

***Abstract*—Accurate fault prognosis of machine component is important to maintain industry operation system. Faults analysis can be very helpful in fault early warning and reducing maintenance cost. The goal of our work is to design an integrated approach of machine faults analysis. A method widely used is Fuzzy Neural Networks (FNNs), but such method lacks of flexibility. We present a Membership-based Multi-dimension Hierarchical (MMH) neural network model to jointly include new feature selection approaches and generalized membership operators. MMH model is an adaptive model that employs modified KPCA and Back Propagation algorithm respectively. By introducing optimized KPCA we can extract features of higher importance that are appropriate for fault diagnosis. Our prediction model is inspired by the traditional fixed membership. In our approach, an observing value will be segmented into multiple dimensions where each dimension captures deep structural information in the network. The transformation is updated by back propagation. The proposed approach takes advantage of membership thinking and benefits from large learning capacity of deep neural networks (DNNs). This is aiming to take advantage of membership thinking and neural network deep learning abilities. Experimental results on public datasets demonstrate the superiority of our model that has the character of faster convergence, which also improving the accuracy by an average of 5% for fault prediction.**

***Index Terms*—Feature Selection; Modified KPCA; Back Propagation; Multi-dimension Hierarchical Neural Network**

I. Introduction

Growing attentions on resource shortages around the world have led to an increasing number of researches on improving the energy efficiency. At the same time, machine maintenance and repairs have played an indivisible role in energy consuming. It has been reported that faults in machine may increase about 15% of energy consumption [1], which may also result in many other additional costs.

Fault diagnosis and resolution in a system network are essential for clearing faults that manifest in an electrical sensor transmission or distribution network. Many studies have been carried out on the use of intelligent methods for fault diagnosis in an electrical system.

In the case of faults analysis, we usually have different dimensions of fault feature indication data which is represented as $x_1, x_2, \cdots, x_D$ and $F(x)$ that is on behalf of the fault type of comparison given a series of faults feature observations:

$$F(x) = f(\boldsymbol{x_1}, \boldsymbol{x_2}, \cdots, \boldsymbol{x_n})$$

assuming that the fault type ranges from f_1 to f_c, from this point, we hope to get implicit relationships from $\boldsymbol{x_n}$ to $F(x)$ in the real application scenario.

We get used to analyze and mine this set of D-dimensional vectors, but the complexity of many machine learning algorithms is closely related to the dimensionality of the data, so it is necessary for us to reduce the dimensionality of the data first.

Principal component analysis (PCA) [2] and kernel PCA [3] [4] are well-known methods in feature engineering. However it is still not enough for a KPCA algorithm solving faults feature selection. The existence of noise will keep on disturbing eigenvalues. In this paper, we propose a integrated algorithm combining the original eigenvectors' importance with the final faults type. Our main contributions are listed as follows:

1. By calculating the between and within class in a new way, it can minimize the impact of uncertain factors and increase the benifits of reducing dimensions of input features.
2. When processing the faults diagnosis problem, we find that the existing methods are lacking of mining information of each input dimension. Based on a membership algorithm, we change the structure of now existing multiple level proceptron by seperating input layer into several patchs, also we remove the full connected edge to the hidden layer in order to keep locality of each dimension feature contribution.

The remainder of the paper is organized as follows. Section II describes related work on fault diagnosis algorithm. In Section III, we present how to make fault feature extracted and evaluated by putting the faults type information into consideration before modeling. The new idea of MMH modeling for fault diagnosis problem is provided in Section IV. The experiment on public UCI datasets as well as discussions on baseline algorithm is shown in Section V. Finally we conclude our work in Section VI.

[1]DOI reference number: 10.18293/SEKE2017-074

II. Related Work

The study of fault diagnosis and prognosis recently have concentrated on theoretical research, mainly based on fuzzy theory, pattern recognition, bayes rules, logistics regression, neural network algorithms and so on. Others are focusing on building deep learning models to infer the relation between data and fault results or estimate the probability of faults occur. Below we highlight a few and explain what advantages and drawbacks they have.

- **Classical Fuzzy Set Interface Theory.** Previous research [5] has been done extensively concentrated on inference system design. Fuzzy rule based system (FRBS) deals with IF-THEN rules. FRBS consititute an extension to the classical fuzzy rule inference [6] [7].
- **Deep Learning Models.** Deep learning so long has became a point of focus as it is the skilled-expert in the domain of complex problems. In particular of fault diagnosis domain, different neural network (NN) models are proposed to fitting various background [8] [9] [10].
- **Hidden Markov Model.** A classification method [11] for reluctance motors' fault diagnosis using HHM is carryed out and shown that parameter learning need huge a mount of histrorical data.

III. Faults Feature Extraction

A. Principal component analysis (PCA)

Principal Component Analysis (PCA) analysis is an important means of dimension reduction. It applies a linear correlation transformation on original data features which can explain most of the datasets information in new scope.

Given a set of centered input vectors $\boldsymbol{x_t}$ $(t = 1, \cdots, n)$, and each of which is one of m dimension: $\boldsymbol{x_t} = (x_t(1), x_t(2), \cdots, x_t(m))^T$ then we have the input data matrix $X_{n\times m}$ (usually $n > m$), In general, we will select the eigenvector in which the largest eigenvalues are located. The information in these directions is rich, and is generally considered to contain more information of interest.

B. Theory of Kernel Principal Component Analysis (KPCA)

On account of there are some limitations of PCA, there is no way for the existence of high-order correlation, Kernel PCA can be introduced, using a kernel function we can transform the nonlinear correlation into a linear one. Given a set of input data $\Phi(x_i), i = 1, 2, \ldots, n$ for this discussion, the covariance matrix $\bar{C}$ of centralized data: $\bar{C} = \frac{1}{n}\sum_{i=1}^{n}\tilde{\Phi}(x_i)\tilde{\Phi}(x_i)^T$ Now finding the eigenvingue and eigenvector of $\bar{C}$ and kernel matrix is respectively donated as λ_c, $\tilde{\lambda}_k$ and υ_c, α_c

$$\bar{C}\upsilon_k = \lambda_k\upsilon_k(k = 1, 2, \cdots, D) \qquad (*) \tag{1}$$

the final conclusion by reducing both sides of the equation (1): $\lambda_c = \frac{\tilde{\lambda_k}}{N}$, $\alpha_c = \frac{1}{\sqrt{\tilde{\lambda_k}}}\tilde{\Phi}\alpha_k$. If the adoption of the kernel function is Radial Basis Function,

$$k(x_i, x_j) = \exp(-\frac{\|x_i - x_j\|^2}{2\sigma^2}) \tag{2}$$

here σ is the only parameter of function. In general, KPCA realize non-linear transformation between the data space and feature space through the kernel function.

C. Modified Kernel Principal Component Analysis

The traditional KPCA algorithm only considers the maximum information content of the reserved feature space and does not consider whether these information quantities are effective for classification.

Here, we reconsider the degree of dispersion among the intra-classes and inter-classes, which can both retain good dimension reduction and more conducive to the fault pattern classification.

Before we want to balance the degree of aggregation within class and between classes for each feature vector, a important notation firstly is introduced to represente one class center as $\bar{x}_i = \frac{1}{n_i}\sum_{p=1}^{n_i} x_{ip}, i = 1, 2, \cdots, c$, where c stands for the number of fault classes, n_i is the total number of labeled class i and x_i is the principal component after kernel transformation. The within class distance:

$$W_\sigma = \frac{1}{n_i}\sum_{i=1}^{c}\sum_{q=1}^{n_i}\|x_{iq} - \bar{x}_i\| \tag{3}$$

in the equation, we can calculate each W_σ vary from the extracted dimension d, also the inter-class discretization degree of each eigenvector is:

$B_\sigma = \sum_{i=1}^{c}\sum_{j=i+1}^{c}\|m_j - m_i\|$ where $m_i = \sum_{i=1}^{n_i} x_i$. If we get a bigger between class value B_σ and a smaller within class value W_σ, the more it is with the ability to distinguish categories. Intuitively, the definition of χ is $\chi = \frac{W_\sigma}{B_\sigma}$.

IV. Faults Diagnosis Modeling

After selecting the most informative feature in section III. In this part, we present the wide and multiple neural network and compare it with the traditional model we mentioned in related work. we want to highlight a few previous work by applying neural networks in the domain of fault diagnosis.

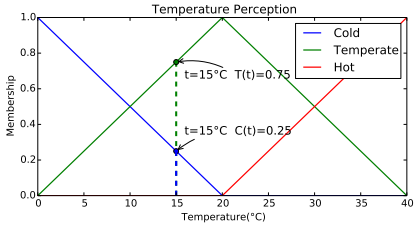

Fig. 1. Example of a tempreture degree membership.

Figure 1 shows a temperature perception in a real scene. In the theory, we assume that any temperature value will correspond to a linguistic description value when a fuzzy rule needs to be applied, or findind the corresponding exact value range. In the figure, when the temperature appears $15°C$, we subjectivity feel that the temperature value is a degree of cold is *0.25*, a moderate degree of *0.75* or when the description of the value of the cold, the corresponding temperature range of $0°C$ and $20°C$.

Here goes our Membership-based Multidimensions model assumptions as follows:

1. Each observed measurable value υ (normalized) will consists by multi-tuples $\mathcal{M}(\upsilon_1, \upsilon_2, \cdots, \upsilon_d)$, d is a multi-dimension parameter that represent the disperse level.
2. The summation of υ equals to a fixed setting: $\sum_{i=1}^{d} \upsilon_i = s$, s here represents the multi-dimension central degree. (e.g., specific $s = 1$, due to the result of normalization, it somehow play as a limitation to d-dimension tuples)
3. Such extended dimensions in a descriptive way (like Fig 1) are independent.
4. Nonlinear relations exist during the learning the faults classes patterns.

Consider the basic structure of a back-propagation network with a single hidden layer, as shown in Figure 2:

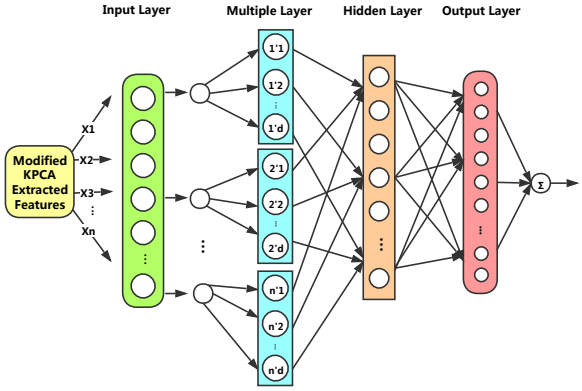

Fig. 2. Model of faults diagnosis based on multiple dimensions neural network.

We divide the input each of the n dimensions into independent description vectors 1-d of the d dimension. If we want our multiple layer outputs are scaled to $[a, b](a, b \in (0, 1])$, the sigmoid function is given as follows: $s(x_i) = \frac{1}{1+e^{-\alpha_i(x_i-\beta_i)}}$ where: $\alpha_i = \frac{(Q_i-1)\ln(\frac{b}{1-b})}{x_{i,max}-x_{i,min}}$, $\beta_i = \frac{Q_i x_{i,min}-x_{i,max}}{Q_i-1}$, $Q_i = \frac{\ln(\frac{1-b}{b})}{\ln(\frac{1-a}{a})}$, and $Q_i \neq 1$. Here α_i *and* β_i are parameters of the active function. The purpose of this approach is to avoid x_i spills out and can be mapped to $(0, a)$ or $(b, 1)$ respectively.

If we have N training examples and C classes label of fault diagnosis then the loss for our prediction $\hat{y}$ with respect to the true labels y is given by:

$$Loss(y, \hat{y}) = -\frac{1}{N}\sum_{n\in N}\sum_{i\in C} y_{n,i} log\hat{y}_{n,i} + l2\_reg\_para * (a_0 + a_1 + \cdots + a_i) \tag{4}$$

Before we doing the back propagation, the difference vector should be compute as : $\delta a = a_2 - y$, then we describe the backpropagation in a functional way, and specially the w_0 matrix update: $\delta a_0 = \delta z1 \cdot w_1^T$, $\delta z_0 = \delta a_0 \cdot S'(z_0)$, $\delta b_0 = \delta z_0$, $\delta w_0 = x_i^T \cdot \delta z_0$, $\sum_{i=1,d,2d,\cdots,nd}^{i+d-1} w_{0_i} = s$. By finding parameters that minimize the loss of our training data, variations such as SGD (stochastic gradient descent) or minibatch gradient descent typically perform better in practice.

$$w_{new} \leftarrow w_{old} - \eta \cdot \delta w_{old} \tag{5}$$

In the process of gradient descent, to prevent local shock, we also introduce decaying learning rate over time: $\eta = \eta_0 \cdot e^{-d_0 t}$.

V. Experiments

In this section, we empirically study the performance of Modified Kernel PCA and MMH model in public date sets.

Datasets: we used two large UCI datasets: *Secom* [12] and *Sensorless Drive Diagnosis* [13]. The first is a binary classification problem where data were taken from a semi-conductor manufacturing process and used to select most relevant signals. The second dataset was extracted from 11 different labels motor current with intact components.

TABLE I
DATESETS FOR KPCA AND MMH MODEL

Dateset	Dimens.	Classes	Instances	Train Prop.
Secom	591	2	1567	-
Sensorless	49	11	58509	0.8

A. Feature Extraction Task

We solve feature extraction task by compareing the PCA and KPCA with *RBF* kernel in the first step.

TABLE II
COMPARATION IN SVM RESULT BY ORIGINAL AND MODIFIED KPCA

Method	Reduced Dimensions	Accumulate Contribution	Accuracy
Original KPCA	6	89.20%	72.24%
	9	90.73%	79.92%
	12	96.21%	82.21%
Modified KPCA	6	90.82%	**85.27%**
	9	92.12%	**89.23%**
	12	97.44%	**92.27%**

In tabel II, we put the Original KPCA, Modified KPCA into the based SVM classificator. After 10-fold cross-validation, chaging the dimensions from 6, 9, 12, our proposed Modified KPCA algorithm can obtain more valuable information form each disparate dimension. It can more likely find the number of selected features with the strongest causal effect relationship.

Using Modified KPCA we proposed, comparing with PCA, KPCA in the same dataset experiment, we can find modified kpca performs well and quickly from the view of aggregate proportion of importance. Although the contribution of the first dimension, PCA is little higher than KPCA and modified KPCA, as dimensions extend bigger, the accumulative proportion line of KPCA goes higher than naive PCA.

B. Faults Event Diagnosis Task

We cast faults event diagnosis tasks upon the UCI sensorless dataset. In this task, basically, we set up multi layer perceptron serve as a contrast to our proposed model with 3 hidden layer of $[100, 300, 100]$ and the learning rate is 0.05. Also, we experiment a few traditional method as baseline.

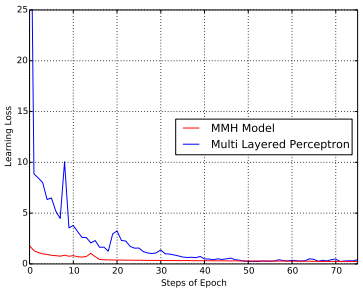

Fig. 3. Comparation of Loss Convergence Case

Since the existing open-source machine learning library can not meet the needs of the model, we build our own MMH model in Tensorflow. Under the premise of multiple dimensions equal to 3, the hidden layer structure is $[50, 300]$, and we set the adaptive learning rate η to 0.5.

To avoid overfitting, we introduced the normalization parameter of 0.005 to reduce overfitting probability as mentioned in the previous model introduction, and randomly discarded some of the neural during each iteration.The idea behind dropout is simple. The drop-out approach stochastically disables a fraction of its neurons. This can prevent neurons from co-adapting and forces them to learn individually useful features. The fraction of neurons we keep enabled is defined by the dropout probability with 0.1 input to our network.

In Figure 3 we can see a more undulating concussion loss convergence in common multi-layer perceptron. Our approach has more competitives in the aspect of convergence speed. Figure 4 show the diagnosis results we obtain through baseline algorithms and our MMH model. The results indicate that our method outperforms others by an average of 5% on the faults diagnosis problem.

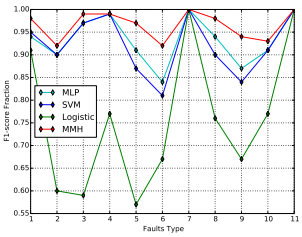

Fig. 4. Comparation of Different Algorithm's F1-score

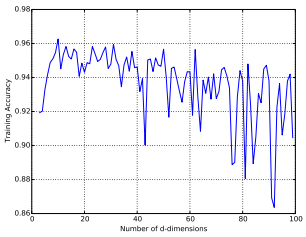

Fig. 5. Different d Dimension ValueS with Accuracy

In particular of Figue 5, when we take the range of d from 1 to 100, we find that the accuracy of the classification of MMH model increases first, and then slowly decline, basically meet the hypothesis, we speculate it is due to the fact that when d is becoming huger, the information is increased, and the contribution of this increment to the accuracy of the model is lower than the expected value of its own noise, which makes the accuracy of the model moves down. The certain range of evaluation is helpful for us to obtain the most appropriate d dimension mapping space according to the actual problem.

VI. Conclusion

The MMH neural network proposed in this paper is an effective and practical model for faults diagnosis and detection from a series of extracted features. MMH model gains the merit of classic neural network and benefits from valuable describing and multi-dimension information. When compared to traditional MLP model, experimental results shows that more latent information obtained by MMH are more convincing and quickly in convergence. Our Modified KPCA approach also provides a better supervised approach considering type labels which can accomplish more meaningful and useful feature extraction work.

Possible future directions for this work include limiting the summation of each multi-dimensions central degree with more certified principle and accelerating the process of Modified KPCA evaluating. As a result of calculating multi-dimensions vector enlarging the set size of multi-variables to estimate, our MMH model is more likely slower than a traditional MLP method. It would be of more importance to improve efficiency and reducing computational expense of vector extending for larger faults data input.

References

[1] D. Westphalen, K. W. Roth, and J. Brodrick, "System & component diagnostics," *ASHRAE journal*, vol. 45, no. 4, pp. 58–59, 2003.
[2] I. Jolliffe, *Principal component analysis*. Wiley Online Library, 2002.
[3] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural computation*, vol. 10, no. 5, pp. 1299–1319, 1998.
[4] R. Rosipal, M. Girolami, L. J. Trejo, and A. Cichocki, "Kernel pca for feature extraction and de-noising in nonlinear regression," *Neural Computing & Applications*, vol. 10, no. 3, pp. 231–243, 2001.
[5] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoningi," *Information sciences*, vol. 8, no. 3, pp. 199–249, 1975.
[6] N. M. Pathade and S. M. Ansari, "Modeling, design and implement of fuzzy logic controlleron fpga robotic platform," 2015.
[7] K. D. Sharma, M. Ayyub, S. Saroha, and A. Faras, "Advanced controllers using fuzzy logic controller (flc) for performance improvement," *International Electrical Engineering Journal (IEEJ) vol*, vol. 5, pp. 1452–1458, 2014.
[8] C. P. Chen, Y.-J. Liu, and G.-X. Wen, "Fuzzy neural network-based adaptive control for a class of uncertain nonlinear stochastic systems," *IEEE Transactions on Cybernetics*, vol. 44, no. 5, pp. 583–593, 2014.
[9] Z. Gao, C. S. Chin, W. L. Woo, J. Jia, and W. Da Toh, "Genetic algorithm based back-propagation neural network approach for fault diagnosis in lithium-ion battery system," in *2015 6th International Conference on Power Electronics Systems and Applications (PESA)*. IEEE, 2015, pp. 1–6.
[10] M. D. Buhmann, "Radial basis functions," *Acta Numerica 2000*, vol. 9, pp. 1–38, 2000.
[11] B. Ilhem, B. Amar, and A. Lebaroud, "Classification method for faults diagnosis in reluctance motors using hidden markov models," in *2014 IEEE 23rd International Symposium on Industrial Electronics (ISIE)*. IEEE, 2014, pp. 984–991.
[12] https://archive.ics.uci.edu/ml/datasets/Dataset+for+Sensorless+Drive+Diagnosis.
[13] https://archive.ics.uci.edu/ml/datasets/SECOM.