Computers in Industry 100 (2018) 287–296

Contents lists available at ScienceDirect

Computers in Industry

journal homepage: www.elsevier.com/locate/compind

A deep Boltzmann machine and multi-grained scanning forest ensemble collaborative method and its application to industrial fault diagnosis

Guangzheng Hu[a], Huifang Li[a,*], Yuanqing Xia[a], Lixuan Luo[b]

[a] School of Automation, Key Laboratory of Intelligent Control and Decision of Complex Systems, Beijing Institute of Technology, Beijing 100081, PR China
[b] Department of Computer Science, University of Toronto, Toronto, Ontario, M5S 3G4, Canada

ARTICLE INFO

Keywords:
Deep learning
Multi-grained cascade forest (gcForest)
Deep Boltzmann machine
Fault diagnosis
Industrial big data

ABSTRACT

The essence of big data-based intelligent industrial fault diagnosis lies in the process of machine learning and feature engineering. Deep learning methods can discover the complex relationship between data and potential faults, and outperform the traditional machine learning methods. The gcForest is able to generate a deep forest ensemble, which allows gcForest to do representation learning and fault classification. At each node of the random tree, gcForest selects the one with the best *gini* value from candidates for splitting. However, most of the data acquired from industrial scene is with continuous and unstructured attributes, accordingly the node-splitting procedure will be generally intractable. We present a novel approach with the combination of deep Boltzmann machine and multi-grained scanning forest ensemble, to effectively deal with industrial fault diagnosis based on big data. At first, we use deep Boltzmann machine to turn all features of data to be processed by forests into binary, and then utilize multi-grained scanning forest ensemble to process them in every layer of deep Boltzmann machine. By means of the collaborative method, we can address the aforementioned issues. The experimental results and analysis on industrial fault diagnosis under different experimental conditions, show that the fault classification accuracy of the proposed approach is competitive to other popular deep learning algorithms, but also takes much less time than gcForest.

1. Introduction

Failures frequently occurring in industrial scene may introduce unwanted downtime and productivity losses. Fault diagnosis can identify the reasons for failure and reduces these risks, and has been of greater importance as a performance assessment of the industrial system. However, effective fault diagnosis in industrial scene is challenging due to its inherent complexity and environmental interference. Big data and its analysis has been proposed for describing datasets and as analytical technologies in large-scale complex programs, which need to be analyzed with advanced analytical methods. Big data analysis can provide useful values via judgments, recommendations, supports, or decisions. Recently, big data-based and intelligent fault diagnosis methods are widely applied.

Previous studies have shown that the performance of the fault diagnosis results, to a large extent, depends on whatever of feature extractors and classifiers. The existing diagnostic mode is “feature extraction and classification”. To improve the effectiveness of fault classification and decrease the professional requirement to the operator, a number of classic and typical machine learning methods have been utilized, such as artificial neural networks (ANNs), support vector machine (SVM), and random forest (RF) [1]. These methods in some studies have complete diagnostic function containing feature extraction and classification, because they can do representation learning independently [2]. But sometimes they only act as classifiers, and meanwhile wavelet packet decomposition (WPD), etc., provide good support for fault feature extraction [3].

Before deep learning, some existing methods such as ANNs, SVM and RF etc., which have shallow architectures, have been widely applied in various fields [4]. Yang and Di (2008) showed that applying RF to diagnosing machine faults is feasible and approbatory [5]. Konar and Chattopadhyay (2011) proposed SVM classifier with the wavelet transform to detect bearing faults and obtained the remarkable results [3]. Abad et al. (2013) indicated that the combination of discrete wavelet transform and ANNs is an effective method for gearbox fault detection based on acoustic signals [6]. Jegadeeshwaran and Sugumaran (2015) used statistical features and SVM to deal with the fault diagnosis of automobile hydraulic brake systems and acquired good results [7]. In addition, these methods are also applied to fault diagnosis with some swarm intelligent optimization algorithms. Chen et al. (2013) presented

* Corresponding author.
E-mail addresses: 1020698051@qq.com (G. Hu), huifang@bit.edu.cn (H. Li), xia_yuanqing@bit.edu.cn (Y. Xia), lixuan.luo@mail.utoronto.ca (L. Luo).

https://doi.org/10.1016/j.compind.2018.04.002
Received 31 May 2017; Received in revised form 26 March 2018; Accepted 4 April 2018
Available online 26 May 2018
0166-3615/

a novel model for gearbox fault diagnosis based on wavelet SVM with immune genetic algorithm (IGA) [8]. Zhang et al. (2015) put forward an improved approach that using SVM with ant colony algorithm (ACO) for synchronous feature selection and parameter optimization [9]. Cerrada et al. (2016) proposed a spur gears fault diagnosis method based on genetic algorithm and RF [10]. The most important task in such studies is to effectively learn elemental feature information from complex and heterogeneous data or signals, and classify the features [11]. Shallow-architectural machine learning algorithms have poor representation learning capability, so these algorithms have been used with feature engineering together, which is essentially lossy [12].

Recent developments have demonstrated the capacity of deep learning methods to deal with the problems involving massive and high dimensional data powerfully, such as text, images, video and voice information [13]. Deep belief networks (DBNs) [14], proposed by Hinton et al., and convolutional neural networks (CNNs) [15], presented by LeCun and Bengio, are two typical examples of deep learning. Due to the success in representation learning [16] and its very high accuracy, the application of deep learning algorithms to default diagnosis has been one of the research hotspots in fault diagnosis field in recent years. Tamilselvan and Wang (2013) used DBN to address failure diagnosis and showed very impressive diagnostic results [17]. Tran et al. (2014) discussed the application of Teager–Kaiser energy operator and DBN for fault diagnosis of reciprocating compressor valves [18]. Chen et al. (2015) addressed gearbox fault diagnosis by using DBN model as well [19]. Guo et al. (2016) and Lu et al. (2017) proposed bearing fault-diagnosis methods by means of hierarchical adaptive CNN [20,21]. Besides DBNs and CNNs, some deep level algorithms, such as deep autoencoder and deep Boltzmann machine (DBM), also occupy important positions in intelligent fault diagnosis area. Sun et al. (2016) applied sparse auto-encoder-based (SAE) DNN approach into the induction motor faults classification [22]. Lu et al. (2017) utilized stacked denoising auto-encoder (DAE) to learn useful representations for fault diagnosis of rotary machinery components [23]. However, most of these deep learning algorithms belong to deep neural networks (DNNs). As known, deep neural networks have too much parameters which need to be fine-tuned, and are prone to overfitting when the training data is not sufficient. In reality, it is a frequent occurrence that we do not have sufficient fault training data at hand, particularly labeled data on account of high time-cost and burdensome tasks. Furthermore, supervised DNNs generally require powerful computational facilities in the training process, like distributed computers and GPU facilities, the extensive calculation is prohibitive for many individuals without powerful computing resources. Moreover, the learning performance of DNNs is quite sensitive to hyper-parameters tuning, such as learning rate, regularization parameters, and mini-batch size. Different options of parameters generate different models, corresponding to different performance.

Besides the deep learning, ensemble methods are equally powerful by combining multiple learners together [24,25]. Breiman proposed Bagging and random forest in 1996 and 2001 respectively [26,27]. A good ensemble always has some individual learners with high accuracy and diversity however, there is no well accepted formal definition of diversity [28,29]. Peter et al. (2015) shows ensemble methods facilitated with deep neural network features will perform better than that of simply using DNN or ensemble methods [30]. Zhou and Feng (2017) proposed a different deep learning method called gcForest (multi-grained cascade forest), which generates a deep forest ensemble with a cascade framework and is competent to do representation learning [31]. This method can be perceived as an ensemble of decision trees' ensemble. It is naturally apt to parallel implementation.

Due to many advantages aforementioned, gcForest may open a door towards alternative to DNNs for many tasks. However, there are still a few fatal flaws while using gcForest in some tasks. In the process of multi-grained scanning and cascade forest, random and complete-random forests are utilized to tackle the feature information. Each random forest has many random decision trees. In contrast with complete-random trees generated by randomly choosing a feature for splitting at each node of the tree, the random decision trees choose the feature following the regulation, which randomly selects $\sqrt{d}$ features as candidates (d is the number of the input feature dimensions), and picks the one with the best *gini* values for splitting. Unfortunately, most of data acquired from industrial scene is with continuous attributes, and the number of the available values is infinite, so the node-splitting procedure will be generally intractable. In addition, each level of the cascade forest produces much fewer further features, compared with the transformed features generated from multi-grained scanning. Concatenating the few features with the transformed features as the inputs of the next level may drive performance up poorly. It means that many levels may be required to enhance the performance. This approach requires a great deal of computation at each node-splitting process in many levels, and provides no guarantees for acceptable consuming time.

In this paper, we propose a novel approach that is a combination of deep Boltzmann machine and multi-grained scanning forest ensemble to deal with big data-based industrial fault diagnosis. Based on deep Boltzmann machine, we can turn raw features into some binary vectors. These binary vectors will be processed by multi-grained scanning and forests, which can generate corresponding class vectors, and then final result can be derived by means of forest ensemble. Indeed, turning continuous attributes into these binary ones can abstain from sophisticated calculation. The proposed approach can address the deficiency of the gcForest. In theory, the training time cost of the gcForest is several times than that of the proposed method in classification tasks with continuous attribute features.

Using experiments on industrial fault diagnosis under different experimental conditions, the results show that the classification accuracy of the proposed approach is competitive with deep belief networks, however takes less time than that of gcForest and is much easier for theoretical analysis than DNNs. Section 2 introduces the principle of gcForest and gives a detailed account of the aforementioned deficiencies. Section 3 introduces a novel collaborative approach that is a combination of deep Boltzmann machine and multi-grained scanning forest ensemble, and DBM which is also described briefly. Section 4 shows the configuration, processes, results and comparative analysis of our experiments, followed by conclusion.

2. Multi-grained cascade forest

The gcForest method is composed of multi-grained scanning and cascade forest. Multi-grained scanning is used and found to be powerful in handling feature relationships so that it can act as the feature-transformation part. Cascade forest part, inspired by layer-by-layer processing in deep neural networks, can extract further features and output the final prediction results simultaneously.

2.1. Multi-grained scanning

Multi-grained scanning, inspired by multi-convolution kernels utilized in CNNs, can extract features for the following cascade forests. An illustration of its process is given in Fig. 1. Sliding windows are used to scan the raw input features. Suppose there are a 400-dimensional feature vector and three sliding windows with size of 100, 200 and 300 respectively. If a window size of 100 is used, a 100-dimensional truncated feature vector will be generated by the sliding window at each step hence, a total of 301 feature vectors corresponding to the window will be produced. These vectors will be utilized to train a random forest and a complete-random forest, detailed in Section 2, and then the class vectors will be generated. So do the other windows and all these class vectors are concatenated into a transformed feature vector. As Fig. 1 illustrated, suppose here 4 classes and 3 windows are used to scan, and then forests produce 602, 402, 202 four-dimensional class vectors,