Engineering Applications of Artificial Intelligence 65 (2017) 240–251

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Machining vibration states monitoring based on image representation using convolutional neural networks

Yang Fu [a], Yun Zhang [a,*], Yuan Gao [b], Huang Gao [a], Ting Mao [a], Huamin Zhou [a], Dequn Li [a]

[a] *State Key Laboratory of Material Processing and Die & Mold Technology, Huazhong University of Science and Technology, Wuhan, 430074, PR China*
[b] *State Key Laboratory of Digital Manufacturing Equipment and Technology, Huazhong University of Science and Technology, Wuhan, PR China*

A R T I C L E   I N F O

Keywords:
Intelligent manufacturing
State monitoring
Image representation
Convolutional neural networks

A B S T R A C T

Measured signals are usually fed into filters or signal decomposers to extract useful features to assist making identification in state monitoring or fault diagnosis. But what is routinely ignored is that an experienced expert can realize what is happening just by watching the signals presented on the oscilloscope even without the analyzing report. The vision image input and the experience feedback are the two keys in this identification process by the brain. The experience can be easily quantified, like 1 for “good” and 0 for “bad”, and used for identification model construction, while there has been no attempt to use pictured signal as the model input. For closed-loop control system, it is necessary to acquire signal feedback point by point to adjust the system in real time. But for state monitoring and fault diagnosis, the pattern hiding among the signal points is usually more important, which is exactly one of the special fields of image representation to indicate complex interrelationship. Taking machining state monitoring as example, this paper explore the possibility to use the pictured signals as input to construct identification model without traditional feature engineering based on signal analysis. Convolutional neural networks (CNN) is introduced to connect pictured signals to different vibration states with experience feedback. Results validate the proposed method with excellent modeling performance. Time complexity analysis proves this pictured signal image representation based CNN method to be capable to be real-time. Two dimensional image representation is a powerful way to exhibit and fuse information. With high flexibility, the proposed method may be a promising framework for monitoring or fault diagnosis tasks.

1. Introduction

Machining is the most widely used processing method in product manufacturing. Under the pressure of manpower cost and the market demand, machining automation, or called unmanned manufacturing, has become the most promising way to balance the economic cost and the production (Altintas, 2012). As a part of intelligent manufacturing technology, effective automatic machining state monitoring system, instead of the human operators, has become a key component to protect the machining tool, timely reveal equipment failures and reduce the risk of breakdown (Altintas, 2012; Teti et al., 2010; Quintana et al., 2011; Siddhpura and Paurobally, 2012, 2013; Abellan-Nebot and Subirón, 2010). Thanks to the advancement of sensor technology, abundant machining information of the machining state during the machining can be obtained using proper sensors, like accelerometer, dynamometer, microphone, et al., which provides an alternative way to monitor the machining state according to the real-time measured signals (Quintana et al., 2011). And as the market develops, many sensors are becoming essential parts of the machine tool when it is produced (Siddhpura and Paurobally, 2013). Therefore, monitoring the machining states, like different vibration states, based on the real-time measured vibration signals is a very promising way.

To connect the measured vibration signals to the machining states, machine learning is a popular used method. It is usually composed of three steps, *signal collection*, *feature engineering* and *model training*, usually called the “signal-feature-model” way. *Signal collection* aims at collecting signal patterns and its associated machining states as many as possible. High diversity of the dataset is the basis for the model to achieve good generalization performances. *Feature engineering* aims at extracting several key characteristic parameters from the original measured signals to make the relationship between the signal and the machining states clearer. Traditionally, the features are manually defined which needs extensive expertise. This makes the feature engineering accepted to be the most critical part for the monitoring

* Corresponding author.
E-mail address: marblezy@hust.edu.cn (Y. Zhang).

http://dx.doi.org/10.1016/j.engappai.2017.07.024
Received 12 January 2017; Received in revised form 29 June 2017; Accepted 26 July 2017

method (Altintas, 2012; Teti et al., 2010; Quintana et al., 2011; Siddhpura and Paurobally, 2012; Sick, 2002). After all the preparations, *model training* works to build a discrimination model to map the features to the machining states. The most common used methods are logistical regression (Phillips et al., 2015), neural network (Liu and Altintas, 1999; Dimla and Lister, 2000; Prasad and Ramamoorthy, 2001; Jia et al., 2016), Bayes methods (Karandikar et al., 2015; Elangovan et al., 2010), support vector machines (Yao et al., 2010; Çaydaş and Ekici, 2012; Zhang et al., 2015; Liu et al., 2011) etc.

Because of the limitation of the modeling capability, the traditional modeling methods cannot process inputs in very high dimension (LeCun et al., 2015). When the dimension of the model input increases, the model either easily gets stuck in the huge number of local minimums or is not able to get convergent within acceptable time cost. These limits make it necessary to adopt feature engineering to reduce the dimension of the original signal inputs. Statistical method is the most common used technique to build features from the long signal sequence, like average value, standard variation, skewness, kurtosis, etc. or other self-defined statistics (Teti et al., 2010; Ge et al., 2013). The signal may be transformed from time domain into frequency domain using Fourier transform or time/frequency domain using wavelet method before the statistics are calculated (Quintana et al., 2011). These manually defined features are expected to include most of the signal information using a much lower dimensional representation. For the specific machining vibration monitoring problem, the number of features is usually around 2 (Yao et al., 2010; Liu et al., 2011; Kuljanic et al., 2009; Tangjitsitcharoen et al., 2015; Tangjitsitcharoen and Pongsathornwiwat, 2013; Fu et al., 2016; Cao et al., 2013).

Manual feature engineering takes use of human experience to supervise the extraction of the concerned information from the measured signals for the to-be-solved problem. Although it has achieved some success in some problems, its defects are also obvious: (1) the extracted features usually work only for the specific problem. Different problems need different features; (2) for a certain problem, there are lots of feature sets which can achieve similar performance, while there seems to be no rules to tell which feature set is better; (3) the extracted feature set is a biased representation, instead of a perfect representation, of the original signal sequence. The signal usually cannot be recovered from the lower dimensional features. The lost information varies with different extractors; (4) the identification model, trained on the biased features, represents an unbiased relationship between the features and the machining states, rather than the practical relationship between the signal and the machining states. This “signal-feature-model” way greatly depends on whether the features have exactly caught the critical property of the monitored machining states (Quintana et al., 2011). Human experience plays an important role in the extracting process and is the biggest uncertainty factor (Teti et al., 2010; Quintana et al., 2011; Siddhpura and Paurobally, 2012).

However, this manual feature engineering step may be not essential. The main purpose of feature engineering is to reduce the modeling complexity by reducing the input dimension. However, the advances of modeling method have made it possible to directly take high dimensional data as the model input. By introducing new network structure and new network training strategy, deep network method is able to model high-dimensional data distribution with great performance within acceptable time cost. Convolutional neural network (CNN), a kind of deep network method, has practically made great breakthrough on complex problems with two-dimensional data input (LeCun et al., 2015; Mohamed et al., 2012; Zhen-Hua et al., 2015; Goodfellow et al., 2014), like the policy and value network in AlphaGo (Silver et al., 2016) which recently beats the human professional 9p *dan* Go player. The improvement of deep network method makes it possible to take the original high dimensional signal as input to directly estimate the essential relationship between the signal and the machining states, which may release researchers from the troublesome directionless knowledge-intensive manual feature engineering. Theoretically, without the manual feature engineering, the model is mathematically the best relationship representation, rather than an empirical model using the traditional “signal-feature-model” way.

This paper proposes to take the original signal sequence as input to directly build the relationship model. Pictured signal is used as the model input. Convolutional neural network (CNN) method is introduced to model the relationship between the pictured signal and the different machining states. Different signal representation forms and different modeling methods will be compared to demonstrate the performance of this proposed “signal-model” way.

2. Reasons to use image representation for the signal

Image is a two-dimensional data matrix which is usually captured by optical devices, like cameras or human eyes. It can also be created by manually defining every matrix element. As a kind of information carrier, image is capable to represent very complex structure distribution and interrelationship. Although signal sequence is a typical one-dimensional data type, we choose to picture the signal into two-dimensional image format as the model input in this work for the following reasons:

(1) From the intuitive experience, image is an easier way for human to learn the information, which implies that image may be a more appropriate format to represent information for learning process. A recent research on pigeons (Levenson et al., 2015) reinforced this thought by the fact that with enough training on the cancer image, the pigeon can “learnt” to identify different cancers with human level performances. Another practical situation is showed in Fig. 1. An experienced operator can identify the chatter state by only listening the machining sound, while the same judgment can also be made by only observing the sound signal image displayed on the oscilloscope screen. This shows that the signal sequence (sound) and the signal image (oscilloscope) contain the same information and can achieve exactly the same judgment. These two examples indicates that the two-dimensional image representation is a proper data carrier to transfer information and can be used as the learning source to build the identification model. Considering the success of the image application on both biology and engineering, image format may also be a good representation for the signal-based recognition tasks;

(2) Image is a much more powerful information representation compared to the one-dimensional signal. A one-dimensional signal can only represent information in which every independent variable corresponds to only one dependent variable, like time sequence which has only one value at every moment. But image can represent much complex relationship in which every independent variable may have more than one dependent variable. Fig. 2 illustrates some typical data representations used in machining signal processing, like the time/frequency spectrum (Fu et al., 2016) when applying wavelet decomposition method, the synchronously sampled signal for audio signal (Schmitz, 2003), the Nyquist diagram for system stability research (Claesson and Håkansson, 1998). Except the signal sequence, none of the other three spectrums can be expressed using one-dimensional format because every x corresponds to more than one y values. The two-dimensional image matrix is a more universal machining data representation;

(3) For the machining monitoring task, the system response speed is very important. With high dimensional input and deep layer network structure, the computation complexity of the algorithm is also very high. Fortunately, because of the promotion of the widespread image application, there has developed a kind of hardware named graphics processing unit (GPU) specially designed for the scientific calculation of two-dimensional image. By using large scale parallel computing technology and deeply optimizing the calculation process, the GPU acceleration technology can significantly improve the computational efficiency of two-dimensional matrix operation, which makes it possible for the image based identification model to meet the responding speed requirement for machining monitoring task.