A Multi-Sensing Collaborative Diagnosis System for the Reliability of Industrial IoT

Haozhen Liu[1], Long Liu[1], Weiguo Wang[1], Qilong Zhao[1], Meng Jin[1], Ziqiang Zhou[2], Zhoubin Liu[2]
[1]School of Software and BNRist, Tsinghua University
[2]State Grid Zhejiang Electric Power Research Institute, Hangzhou, Zhejiang, 310000, China
{liu-hz16, liulong16, wwg18, zhaoql17}@mails.tsinghua.edu.cn,
mengj@mail.tsinghua.edu.cn, jx_zzq@sina.com, jxliuzb@qq.com

Abstract

Accurate and continuous monitoring of large scale machinery is important for modern industries. Existing solutions are often unsuitable for large-scale and complex scenarios where a huge number of data flows that generated by hundreds of heterogeneous sensors should be considered in combination and processed simultaneously to finally judge the status of the machinery. In this paper, we propose a multi-sensing collaborative diagnosis system for accurate and real-time monitoring of large-scale machinery. Our proposed approach tries to capture and model the underlying temporal and spatial structure in sequential data, and use this model for more efficient prediction of machinery. Such prediction model is built on long short-term memory (LSTM) neural networks. A series of data preprocessing methods are also proposed to align the asynchronous data streams and reduce the dimension of the heterogeneous data, which improves the efficiency of the status prediction process. We implement and evaluate our system in a real-world convertor station where many large scale machines are applied for current converting. The results show that the proposed system can achieve less than 2% mean square error, which outperforms the state-of-the-art model-based and ML-based machine fault diagnostics methods.

1 Introduction

Large scale machinery plays a significant role in modern industries and modern technologies in IoT (such as cloud computing and Blockchain). Normal operating of the machines concerns many aspects of the modern industries, such as its safety, efficiency, economy, etc. For example, in the electric-power industry, halt of the generator will not only cause huge economic losses but also affect the normal functioning of all the social sectors that rely on the electric power. Therefore, it is necessary and crucial to ensure the mechanical health and normal operation of the machines.

Under this circumstance, many machine fault diagnosis (MFD) systems have been proposed. In the MFD systems, various types of sensors are implanted on different parts of machines to measure and collect the factors (such as the vibration, temperature, shaft current, etc.) that are considered highly related to the operation of the machines. The collected data is then feed to a data analysis component for further diagnosis. In most existing systems, however, diagnosis typically relies on an explicit relationship between the measured factors and the health conditions of the machines, making them a poor fit for many applications, including monitoring of the large scale machinery in our case. Specifically, the health condition of the machines are related to hundreds of factors and therefore finding a explicit relationship between them is particularly difficult.

As such, machine learning (ML)-based methods have attracted a lot of research efforts recently. ML-based methods typically rely on a trained ML model, whose inputs are the extracted features of the collected data, and the output is the health condition of the machine. Typical ML model that used in diagnosis includes Support Vector Machine (SVM), Bayesian network, etc. These models however perform poor in practice since they process each measurement in the data sequence isolation, ignoring the dependency among them. In fact, the collected data sequence is longitudinal, so it is important to capture the dependencies between the elements of the sequence in order to learn more effective and robust representations, which can then be used in machine diagnosis.

Based on the above analysis, we in this paper exploit Long-Short Term Memory (LSTM), a powerful approach which can capture underlying structure in sequential data, for machine fault diagnosis. LTSM has been applied to many areas such as image caption [1] [2], human action recognition [3], medical diagnosis [4] [5], handwriting recognition [6], etc. However, applying LSTM in large scale machine monitoring still faces the following challenges:

- For many machines (e.g., rotating machinery), it is impossible to embed extra hardware after they are manufactured. Therefore, the sensors should be deployed at some distance to the machines, and thus the collected data is noisy and difficult to accurately indicate the s-

International Conference on Embedded Wireless Systems and Networks (EWSN) 2019
25–27 February, Beijing, China

ISBN: 978-0-9949886-3-8

tate of the machine. How to accurately extract features from a series of noisy data is the first challenge we face.

- Data from different sensors may be redundant and even contradictory due to the noise. Besides, the data are asynchronous since the different sampling rates of the sensors. Extracting valid features of the asynchronous multi-source data is challenging.
- Third, there exist different kinds of factors we need to monitor, including vibration, temperature and shaft current. The data collected from these factors are heterogeneous, where both digital and analog data exist. And the data can be either structured or unstructured. How to deal with such large-scale heterogeneous data in real time is a big challenge for us.

In this paper, we address the above challenges and build a digital twin system for a convertor station in power industry that performs fault diagnosis and real-time condition monitoring on the rotating compensator [7] with LSTM and advanced IoT techniques, such as low-power wireless sensor networks [8] [9] [10], battery-free communication and sensing, RFID [11] and etc. The proposed system targets at the diagnosis of rotating machines, which is a widely used part in industrial equipments, ranging from small motor to massive generator. Note that although the focus of this paper is on the diagnosis of rotating machines, our system can be tailored to other instances. The core design of our diagnosis system involves two main components: data processing and fault detection. In the data processing component, we design a compressive sensing based data filling method to align the data collected by different sensors. In addition, since hundreds of sensor are deployed in a machine and the collected data are interrelated, we then reduce the dimension of multiple time series using Principal component analysis (PCA), which helps to improve the efficiency and effective of fault diagnosis. The processed data series are then feed to the fault detection component, where a LSTM based diagnosis method is designed to continuously trace the state of the target machine, and trigger an alarm once an abnormal behavior is detected.

In summary, the main contributions of this paper are as follows.

- We disclose the time dependencies between the elements in the collected data series and propose a LSTM based model to continuously track state of the target machine, which can detect abnormal operation of the machine with high reliability and accuracy.
- Based on this model, we propose and implement an automatic, accurate and continuous diagnosis system for large scale machinery in industry. The proposed system addresses several unique practical challenges in diagnosis of large scale machines: i) a compressive sensing based scheme is proposed to align the data collected by different sensors; ii) a PCA based method is designed to reduce the dimension of multiple time series.
- We implement and evaluate our method in a real-world convertor station where many large scale machines are applied for current converting. The results show that the proposed system can achieve less than 2% mean square error, which outperforms the state-of-the-art model-based and ML-based machine fault diagnostics methods.

The rest of this paper is organized as follows. We present the related work in Section 2 and then the detailed system design in Section 3. In Section 4, we describe the implementation and experimental results of our system. Finally, we conclude our work and discuss the future work in Section 5.

2 Related Work

2.1 Model-based approaches

Traditionally, there are many model-based mechanical fault diagnosis approaches. The model-based approaches generally use physical or mathematical models of the machines to monitor. Currently, different kinds of model-based diagnostic approaches, including analytical method, finite element (FE) method, and combined analytical FE approach, have been applied to conduct fault diagnosis of a variety of rotating machineries, as is shown in [12]. Ma et al. [13] developed an FE model of a geared rotor system considering the effects of the extended tooth contact and tooth root crack on the time-varying mesh stiffness with tooth root crack. Hu et al. [14] proposed a FE node dynamic model for the gear-rotor-bearing system with different lengths of crack. But these traditional model-based approaches using physical or mathematical models require a lot of professional expertise and knowledge, and are not applicable to multiple types of time series data analysis which we used in our work.

2.2 Signal processing approaches

Machine condition signals analyzed by signal processing methods can generate fault-related characteristic feature for decision making. Many signal processing methods are widely used in mechanical fault diagnosis such as wavelet transform (WT) [15], empirical mode decomposition (EMD) [16], spectral kurtosis (SK) [17]. A kurtosis-guided adaptive demodulation technique for bearing fault detection based on tunable Q-factor WT has been proposed in Ref. [18]. Li et al. [19] used EMD to decompose adaptively angle-domain stationary signals to detect gear faults under different speeds. Wang et al. [20] proposed an adaptive spectral kurtosis (SK) technique with adaptive determination of the bandwidth and center frequency for the fault detection of rolling element bearings. Although the signal processing approaches perform well in machinery fault diagnosis, it is not suitable when using heterogeneous data in machinery fault diagnosis.

2.3 ML-based approaches

The ML-based approaches don't require much expertise in the specifics of diagnostic application. Numerous intelligent system approaches for mechanical fault diagnosis have been proposed, such as k-NN [21] , SVM [22], ANN [23] and LSTM [24] [25] [26]. Zhou et al. [27] provide a contribution analysis-based fault isolation method by decomposing the k-NN distance used as the detection index. Liu et al. [28] propose an intelligent method based on a short-time matching atom decomposition method and SVM for bearing fault diagnosis. Sadeghian et al. [29] present an algorithm for induction motors online detection of rotor bar breakage, based on the combination of wavelet packet decomposition

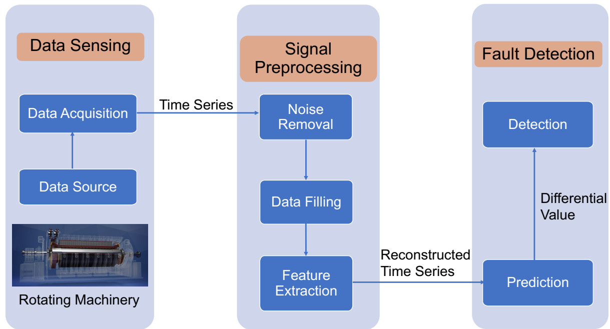


Figure 1. System Overview.

(WPD) and ANN. These approaches are not suitable for time series data. In addition, the k-NN approaches need a lot of storage space and much time to compute , and the efficiency of SVM for big data is low. Zhao et al. [30] explored the direct application of LSTMs on raw time series data to predict the tool wear condition. The LSTM approach is suitable for time series data but this method proposed by [30] is not suitable for diagnosis of the machinical fault data which is collected by multiple sensors.

2.4 Compressive sensing

Compressive sensing [31] is a signal processing technique using to rebuild signals which is sampled randomly by non-linear reconstruction algorithm. Tang et al. [32] developed a sparse classification strategy based on Compressive sensing theory for mechanical fault diagnosis, which helped construct a learning dictionary to represent the vibration signal. Chen et al. [33] proposed a method based on Compressive sensing in order to extracting impulse components in the fault gearbox. Chen et al. [34] also proposed a sparsity-enabled signal decomposition method in order to diagnose the fault localization of automatic tool changers. In our work, we use the compressive sensing to fill the mechanical condition data which sampled by multiple sensors at different time in order to make different data has same timestamp.

3 System Design

In this section, we firstly introduce the design of our system. Then, we will illustrate the main components in detail. The proposed system in this paper targets at the diagnosis of rotating machines, which is a widely used part in industrial equipments, ranging from small motor to massive generator. Note that although the focus of this paper is on the diagnosis of rotating machines, our system can be tailored to other instances.

3.1 System Overview

As shown in Fig. 1, our system consists of three main components: Data Sensing, Signal Preprocessing and Fault Detection.

- *Data Sensing:* This component is composed of two modules: Data Source and Data Acquisition. Data Source module specifies the type, characteristic and sampling frequency of sensor data. Data Acquisition module explains the approach which is used by our system to acquire the sensor data.

- *Signal Preprocessing:* After obtaining sensor data, we preprocess the data to provide high-quality data for the further diagnosis. The preprocess component includes three parts: Noise Removal, Data Filling and Feature Extraction. To achieve these goals, we leverage several general yet effective methods, which are median filter, compressive sensing, and PCA (Principal Component Analysis) respectively.

- *Fault Detection:* In this component, we design a LSTM-based model to predict sensor readings with the preprocessed data. We compute the differential value of predicted data by LSTM model and original data obtained from sensors. And then we infer the machine status by the differential values of the specific sensor and the related sensors. According to the detection results, we give some advice and comments on machine health state.

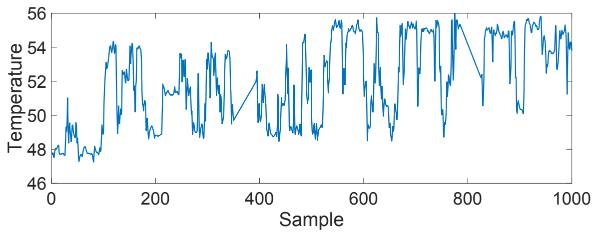
3.2 Data Sensing

1) Data Source: To monitor the machines' running state in real time, various types of sensors are deployed. For example, temperature sensor, electric sensor (current sensor, voltage sensor, and etc.) and rotating speed sensor are the most common sensors on the rotating machinery. Due to the complexity of machine structure, sensors such as temperature sensors can be divided into more fine-grained types. For example, rotating machinery is composed of stator, rotor, bearing, collector and cooling system. And each part is equipped with its own temperature sensors. Besides the sensors mentioned above, vibration signals are also acquired by vibration sensors. From domain knowledge, we know that temperature is a very important physical quantity to measure the operating status of rotating machinery. Therefore, in this paper, we focus on temperature sensors and their related quantities. By analysing the temperature data, we can do a comprehensive diagnosis of the rotating machinery health condition.

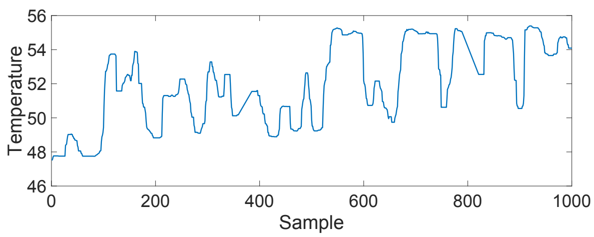
2) Data Acquisition: To gather the data produced by sensors, we develop a acquisition program. Considering the different types of sensors, we utilize two methodologies to implement this program: DCS(Distributed Control System)-based methodology and FTP-based methodology. DCS is a data collection program provided by a third party company. The sensor data of rotating machinery and auxiliary machines are mainly transmitted through the DCS methodology. DCS program first stores the data collected from each sensor into the DCS database. Next, we use the application interfaces provided by DCS to fetch sensor data and put them into our own database. FTP (File Transfer Protocol)-based methodology is designed for transferring vibration data. Because the vibration frequency of the rotating machinery is very fast, it will produce a large amount of data, which makes the transmission of these data through DCS time-consuming and costly. Therefore, we propose a FTP-based method to receive the vibration data directly.

3.3 Signal Preprocess

1) Noise Removal: There are a large number of sensors deployed on the rotating machinery and auxiliary machines. Due to the inherent characteristics of the sensor and the in-



(a) Original time series



(b) Filtered time series

Figure 2. Original and filtered time series data.

fluence of the monitoring environment, the perceived data is very noisy. First, the sensor's communication ability, battery performance and storage capacity are limited, which makes the stability of the sensor threatened. Second, the high-speed rotation of the rotor makes it difficult to attach the sensor to the rotating machinery. In addition, the complex electromagnetic environment inside the rotating machinery also interferes with the data measured by the sensor. Hence, it is inevitable to perceive noise problems in sensor data. In data-driven machine diagnostic systems, these noises will affect the reliability and accuracy of diagnostic results, and even affect the important decisions.

To use sensor data for machine diagnosis, such noise must first be removed. We leverage a Median Filter method to remove the noise from the time series data. Median Filter is a nonlinear digital filtering technique that is often used to remove noise from an image or signal. The main idea of the Median Filter is to run through the signal entry by entry, replacing each entry with the median of neighboring entries. In our scenario, sensor data often shows local spikes. These spikes are usually some noise and should be filtered out. Fig. 2(a) shows the original sensor data and Fig. 2(b) shows the resultant from Median Filter. We observe that Median Filter successfully removes most of the noises from the sensor data.

2) Data Filling: In *Data Source* part, we mention that temperature data is an important indicator of operation state of the rotating machinery. According to domain knowledge, there is a strong correlation between temperature and other data, such as current, voltage, active power, etc. To fully make use of temperature data and the correlation data, We apply the temperature data as a benchmark to schedule the temperature data and its correlation data in time series. Through this operation, we will get a matrix composed of time series data. Because each type of sensors has a different sampling frequency. For example, the sampling frequency of the temperature sensor is one data per ten seconds, while the rotating speed sensor is one data per one minute. Besides, sensor data may also be missing due to line quality, sensor failure or battery instability during transmission. These will cause the matrix to be vacant at certain positions.

In this condition, we use a widely used method for matrix filling: compressive sensing. Compressive sensing theory is an effective signal processing theory developed in the field of signal processing in recent years. It was proposed by D. Donoho, E. Candes and Chinese scientists Tao T. et al. Contrary to traditional Shannon-Nyquist Sampling Theorem, compressive sensing can reconstruct the signals from far fewer samples if these signals are sparsely representable [35]. Signal sparsity is simply understood as the number of non-zero elements in the signal is much smaller than the total number of signals. The real signals that exist in nature are generally not sparsely, but are approximately sparse in a certain transform domain. In other words, any signal has compressibility. As long as we can find its corresponding sparse representation space, it can effectively perform compression sampling. Signal sparsity or compressibility is an important theoretical basis for compressive sensing. Before introducing how we apply compressive sensing for data filling, we need to understand how compressive sensing works.

Given an original signal x with size $N \times 1$ and its measurement signal y with size $M \times 1 (M \ll N)$, they satisfy the relationship shown in Equation (1).

$$y = \Phi x \tag{1}$$

In Equation (1), Φ is a $M \times N$ matrix and it is usually called *measurement matrix*. It represents the distribution of M samples of the original signal. The goal of compressive sensing is reconstructing x by y and Φ. Obviously, since the dimension of y is much lower than the dimension of x, Equation (1) has infinitely many solutions, that is, the problem is underdetermined and it is difficult to reconstruct the original signal. As we mentioned above, if signal x is sufficiently sparse, we can recover signal x from y. In real world, x is usually non-sparse. But we can sparsely represent it to another domain.

$$x = \Psi s \tag{2}$$

As shown in Equation (2), signal x can be divided into Ψ with size $N \times N$ and s with size $N \times 1$. The matrix Ψ is known as the *sparse representation basis*. While s is the sparse representation of signal x in the domain Ψ. s satisfy $\|s\|_0 = K$, where $K \ll N$. Combine Equation (1) and Equation (2), we get Equation (3).

$$y = \Phi \Psi s \tag{3}$$

From Equation (3), we conclude that the signal reconstruction problem can be converted to compute s for given measurement y and known matrices Φ and Ψ. If we know s, we can obtain original signal x by $x = \Psi s$. How to calculate s is an optimization problem. We may acquire s by the smallest l_0 norm:

$$arg\ min \|s\|_0 \ st.\ y = \Phi \Psi s \tag{4}$$

Unfortunately, this optimization problem is NP-hard, we are not able to get a solution in polynomial time [36]. Donoho D L et al. prove that the l_0 norm problem is equivalent to the l_1 norm problem:

$$arg\ min \|s\|_1\ st.\ y = \Phi\Psi s \tag{5}$$

Equation (5) is a convex optimization problem and can be easily solved using linear programming(LP) methods. This convex optimization problem is known as Basis Pursuit(BP). Besides BP, there is another type of methods to recover x which is called Matching Pursuit(MP). Orthogonal Matching Pursuit(OMP) [37] and Regularized Orthogonal Matching Pursuit(ROMP) [38] are typical algorithms of MP. These two methods are faster than BP but the reconstruction quality are worse than BP, especially when the signal is not sparse enough.

The description above is the core idea of compressive sensing. Through the analysis, we can see that the design of *measurement matrix* Φ and choosing proper *sparse representation basis* Ψ are two very important issues in compressive sensing. In order to accurately reconstruct sparse signals, a certain relationship needs to be satisfied between Φ and Ψ. That is Φ and Ψ are not correlated [39]. In other words, the row ϕ_j of Φ cannot be sparsely represented by the column ψ_i of Ψ, and the column ψ_i of Ψ cannot be sparsely represented by the row ϕ_j of Φ. It is difficult to construct a *measurement matrix* directly so that it satisfies the above conditions. Tao T. et al prove that when the measurement matrix is a Gaussian random matrix [31], the condition will be satisfied. For *sparse representation basis*, there are more choices. Commonly used sparse basis include Discrete Cosine Transform basis(DCT), Fast Fourier Transform basis(FFT), Discrete Wavelet Transform(DWT) basis, Cervelat basis, Gabor basis, and Redundant Dictionaries.

In our scenario, different type of sensors have different sampling frequencies. The low-frequency sensors exist due to two reasons. First, the physical quantity monitored by some sensors is basically stable for a certain period of time without significant changes. There is no need to sample the data with a high frequency. For example, when the rotating machinery works normally, the rotating speed will stabilize at around 3,000 revolutions per second. Second, due to limited sensor battery capacity and financial cost, the data collector has to reduce the sampling frequency of some sensors. To our knowledge, the vast majority of low-frequency sensors are produced for the second reason. For these low-frequency sensors, we consider they take a small amount of sampling but retain the key information of the original signal. This is consistent with the idea of compressed sensing.

Next we will explain how we use compressive sensing to solve our problem. Every time when the new temperature data comes, we find the data of sensors that has a relationship with the temperature sensor at this time, and compose all the data into a row vector v with size $1 \times H$ (H denotes the number of sensors, including the temperature sensor and its related sensors). If the related sensors have no data at this time point, it is set zero in the corresponding position of v. Then, we select a slide window w with size $N \times H$ (N denotes the number of temperature sensor data), w contains the latest N data of the temperature and its related sensors. There is no doubt v is included in w. There are many zero elements in w, then we employ compressive sensing to fill w so that we are able to reconstruct the sensor data.

The compressive sensing method we use is BP, And we design the *measurement matrix* by this rule: the $M \times N$ Φ specifies a measurement scheduling policy: it contains a 1 in the (m,n) position ($1 \leq m \leq M$, $1 \leq n \leq N$) if the m-th measurement is taken at time n. In our scenario, only a single measurement is taken in every time. This implies, Φ contains one and only one 1 element in any row, and at most one 1 in any column, and 0 everywhere else. This is different from Gaussian random matrix we talk above which is very dense with no 0-entries. Although this *measurement matrix* does not guarantee completely irrelevance with *sparse representation basis*, the experimental results show that the matrix can still achieve a good filling effect.

3) Feature Extraction: As we mention above, there is strong correlation between rotating machinery data, each sensor may be associated with several other sensors. In order to ensure the real-time nature of our system and reduce the complexity of the prediction algorithm, it is necessary to perform dimension reduction operation which is also called feature extraction. There are two schemes for feature extraction, one is feature selection based on domain knowledge. For example, the strong and weak correlation in rotating machinery data is given by the expert according to the physical structure of the machine, and the other is the data-driven feature extraction method. Researchers use mathematical methods to mine possible relationships in data and extract features.

In this paper, we use a combination of two schemas to extract features. Firstly, based on the domain knowledge, we obtain a single sensor and the sensors which have strong or weak relationship with it, and then extract the target features based on the data-driven method. Here, we use the PCA method to extract features. Our goal is to reduce the data dimension without losing too much information. Since we do not know what dimensions we should reduce to. Therefore, we investigate the relationship between data dimensions and the accuracy of prediction results. The experiment results show that when the dimension is set to 4, it has better prediction accuracy. Meanwhile, there is little loss of information compared to the original data.

3.4 Fault Detection

1) Prediction: The ultimate goal of this paper is to make predictive analysis of rotating machinery operating status based on sensor data, and to provide assistant decisions based on the analysis results. More specifically, we want to predict the current temperature value based on the previous time data and historical data of the temperature sensor and its related sensors. By comparing the difference between the predicted value and the actual value, we can make an assessment of the operating status of the rotating machinery. This is a time series data prediction problem which has been investigated for many years. Traditional methods such as k-NN, Hidden Markov Model(HMM), SVR(Support Vector Regression) and some model-based approaches can effectively handle small-scale time series data, however, they are incapable of handling multi-source, heterogeneous, large-scale rotating machinery data. The main reason lies in that these methods are not able to keep long-term dependencies of the historical data. The LSTM as a deep network architecture is capable of handling long-term dependencies with a gated structure. LSTM network has cell unit which consists of three main gates: an input gate, a forget gate, and an output gate. These gates consist of a Sigmoid neural layer and a dot multiplication operation. The output of the Sigmoid layer is a value between 0 and 1, which is used to control the passing of information. When the Sigmoid layer output is 0, it means that the gate is closed and no information is passed. When it is 1, it means that the door is open, allowing all the information to pass through. Specifically, the input gate will selectively record new information in the cell state. The forget gate allows the cell to remember or forget its previous state. Finally, the output gate is used to control what information is output from the cell state.

As shown in Fig. 3, LSTM as an neural network contains at least three layers, they are the input layer, hidden layer and output layer respectively. Mathematically, at time t, we use x_t, h_t, o_t represent the layers mentioned above. In our case, x_t is the combination of time $t-1$ temperature sensor data and its related sensor data at time t. h_t is the hidden layer state of time t. While o_t is the predicted temperature sensor data at time t. In (6), we have a clear understanding of the relationship among the input layer, hidden layer, output layer, and various gates. In this equation, σ(Sigmoid layer) and *tanh* are nonlinear activation functions, W and U are the weighted matrix, while b is the bias vector.

Our training algorithm adopts stochastic gradient descent for optimizing the objective function. The loss function is the mean square error between true data and predicted data. We also add a dropout layer to avoid overfitting. Finally, we apply the LSTM method in real rotating machinery dataset and achieve promising prediction results.

$$
\begin{aligned}
f_t &= \sigma(W_f x_t + U_f h_{t-1} + b_f) \\
i_t &= \sigma(W_i x_t + U_i h_{t-1} + b_i) \\
o_t &= \sigma(W_o x_t + U_o h_{t-1} + b_o) \\
\widetilde{c_t} &= tanh(W_c x_t + U_c h_{t-1} + b_c) \\
c_t &= \widetilde{c_t} \cdot i_t + c_{t-1} \cdot f_t \\
h_t &= tanh(c_t) \cdot o_t
\end{aligned}
\tag{6}
$$

2) Detection: After obtaining the predicted value, we compare the predicted value with the true value and compute the difference, according to domain knowledge, if the difference is large than 1, our system will alter alarm to the rotating machinery maintenance personnel and indicate the location of failure may occur.

4 Evaluation

The main objective of our dataset-based evaluation is to study the performance of the prediction model and data alignment efficiency based on real world datasets. Specifically, the evaluation has three goals:

- Assess the effectiveness and efficiency of LSTM.

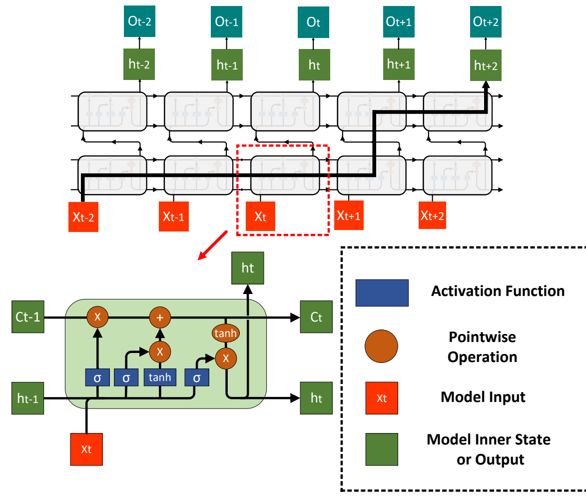


Figure 3. LSTM network structure

- Investigate the contribution of compressive sensing to prediction performance.
- Characterize the relationship between input size and model dimension and find out the optimal parameter collection for predicting rotating machinery system operating status.

4.1 Datasets and performance metrics considered

We evaluate our approach on two real-world rotating machinery datasets: a two-year generator system operating sensor dataset(GSD) and a two-day compensator system operating sensor dataset(CSD). The GSD consists of $1,393,704$ sensor readings of 26 sensors, including temperature, rotating speed, current and active power of one generator. The sensor readings of GSD have been aligned with moving mean filter before we have access to the dataset and the window size is 10 minutes. Table 1 lists the sensor type distribution of GSD. For GSD, the ground truth in terms of the operating status is known, and we use mean square error(MSE) to measure efficacy of our prediction results.

Correlation between data series. We assume the time series of rotating machinery are correlative. Fig. 4 plots the absolute value of Pearson correlation coefficient between 26 series of GSD. Pearson correlation coefficient $\rho_{X,Y}$ is a measure of the linear correlation between two variables X, Y derived as:

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \tag{7}$$

where Cov is the covariance of X and Y, σ_X and σ_Y are the standard deviation of X and Y. $\rho_{X,Y}$ has a value between $+1$ and -1, where 1 is total positive linear correlation, 0 is no linear correlation and -1 is total negative linear correlation. Let us note that column 7–9, 11–13 and 17–23 are strong interrelated, which means the existence of information redundancy within these 26 series. This motivates us to perform dimension reduction on GSD for estimation improvement.

We train our PCA transform matrix and LSTM model leveraging on 80% of GSD and perform validation on the

Table 1. Sensor Type Distribution of GSD

Sensor Type	Number	Description
Temperature	12	coil water temp(°C), hydrogen temp(°C)
Electrical parameters	9	active power(MW), current(A), voltage(kV), generator freq(Hz)
Rotating speed	1	rotating speed(r/min)
Auxiliary parameters	4	water flow(t/h), water pressure(kPa), hydrogen purity(%), hydrogen pressure(kPa)

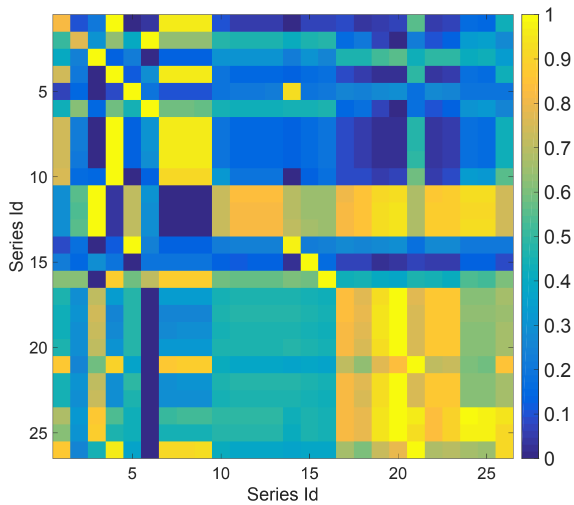


Figure 4. Correlation coefficient of 26 series in GSD.

other 20%. The CSD includes two-days data of 1,148 sensors, including temperature, rotating speed, current, vibration and water pressure. The sampling frequency distribution of CSD is as shown in Table 2. As the compensator system has been deployed for only half a year when this paper is completed, there is no fault operation condition since the system itself has high reliability. Therefore, we estimate our method by calculating the prediction error instead of fault classification accuracy.

4.2 Prediction Accuracy of LSTM

a) Results and Observation: We leverage compressive sensing on GSD and reduce the dimensions of correlation variable of stator outlet temperature T_{out} from 13 to 4. The derived 4 features and the raw T_{out} are fed into LSTM as the training data. Figure 5 shows the results on the T_{out} time series. As shown, the mean square error(MSE) on one single sensor time series is less than 0.2K. Considering our model is expected to be able to fit the pattern of normal behavior of sensors, less than 1K error of temperature means the predicted time series is a good approximation. Let us note that the

Table 2. Sensor Source Distribution of CSD

Sensor Source	Number
Compensator	466
Online monitor	104
Cooling system of stator	74
Cooling system of rotor	68
Lubricant system	168
External cooling system	166
Demineralized water system	102

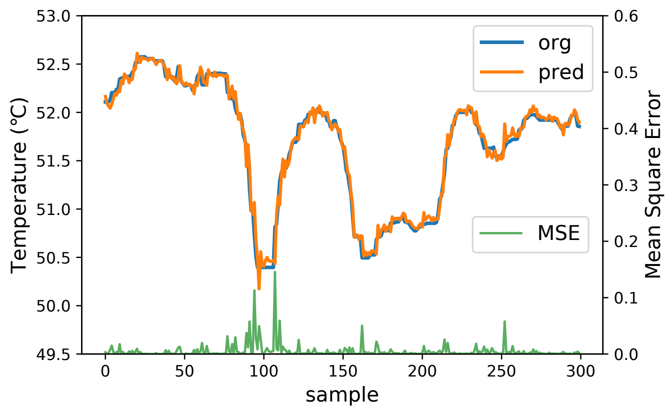
error become relative large when data fluctuations are severe. We consider this is caused by the delayed response of LSTM to jitter in time series. The delayed response lead to no more than 2.48%. Fig. 5(b) shows the CDF of prediction error of T_{out}. It is observed that the system manages to estimate the temperature with less than 0.1K error in 90% time and less than 0.4K error in 99% time.

Here we set hidden layer size $H = 8$, input data size $I = 4$, time steps $T = 5$, and number of cells in each time step $N = 1$. We will describe the parameter choosing principle in 4.4.

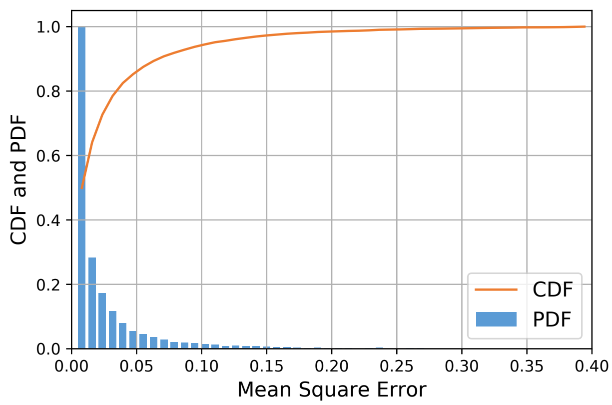
b) Performance Comparison: There have been related studies on estimating value of time series. We compare SVR and LR(Linear Regression) with our LSTM network and Fig. 7 shows the performance of the three models on fitting the pattern of stator outlet temperature. We observe that LSTM model outperform the other two models in terms of estimation error. The MES of LSTM is 0.1265K while SVR get 0.4767K and LR get 0.1742K. However, we find that the maximum error of LR is slightly less than that of LSTM. We consider this is because the temperature fluctuates more violently during certain periods while our LSTM tends to fit it to more smooth pattern. This infers that our LSTM model has stronger noise immunity.

4.3 Performance of Compressive Sensing

Since our LSTM model requires history readings of one sensor and its related variables and sampling frequencies of different sensors varies, it is necessary to perform data alignment on asynchronous time series. Fig. 8 plots the reconstruction performance of compressive sensing on four sensor reading series(stator outlet temperature, cold water flow, stator inlet temperature, active power). We reconstruct the series with 50% of 10,000 samples. It is observed that relative error(i.e. the absolute error divided by the magnitude of the exact value) of stator outlet temperature, cold water flow and stator inlet temperature is below 3%. The reconstructed time series precisely fit the raw data. However, we notice that the average relative error of reconstructed Active power is 12.5%. We can infer from Fig. 8 that stator outlet temperature and active power have a strong positive correlation. When the generator set is in normal operation, the active power can be fluctuate due to changes in the power load. And the auxiliary water temperature varies in a similar pattern with generator load fluctuations. This observation confirms our previous assumptions about data redundancy and we can perform dimension reduction on multiple sensor data and achieve precise estimation and reduce the computing power demand.



(a) Error of 300 samples.



(b) CDF and PDF of mean square error

Figure 5. LSTM prediction result of the stator outlet temperature.

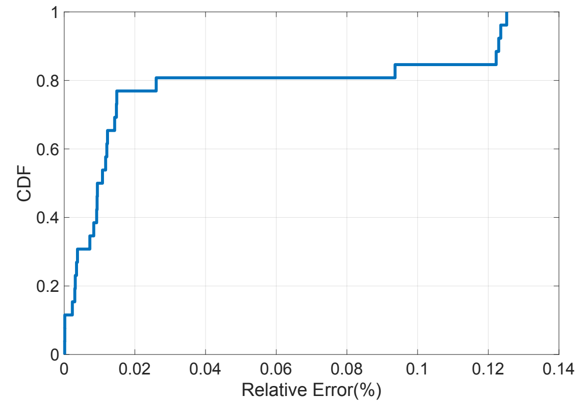


Figure 6. CDF of data filling error.

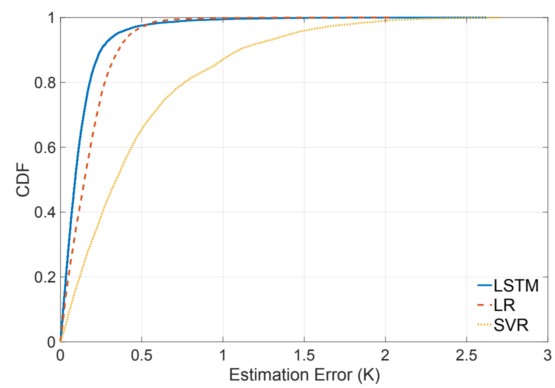


Figure 7. CDF of estimation error while using LSTM, LR and SVR.

Fig. 6 shows the CDF of data filling error on 26 sensor time series. 80% relative error of all sensors are less than 2% and other 20% of sensors have less than 13% relative error. The polarization of reconstruct error is due to the large range of electrical parameters e.g. active power and generator current. The output power of generator changes along with power load, which is hard to represent using a set of sparse basis.

4.4 Hyper-Parameter Selection

LSTM model typically contains various hyper-parameters whose value is set before training phase. Examples of hyper-parameters are hidden layer size, input data size, time steps, and number of cells in each time step. A combination of these hyper-parameters values is a model configuration. LSTM models with different configuration tends to have different performances(i.g. Mean Square Error) and complexities. The more complex LSTM model is, the more resources(i.g. Memory, compute cost) model requires. Notice that configuration space increases exponentially with the number of hyper-parameters. Moreover, we can't estimate each configuration's performance until the end of training phase which requires extremely huge resource.

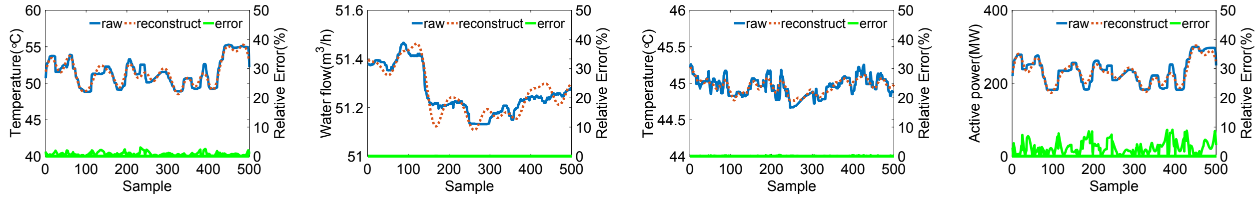
As most in-use rotating machinery data process servers are equipped with only multi-core CPU and no GPU, we consider Complexity-Performance trade-off with multi-dimensional configurations. Our model is expected to maximize the performance and minimize the model complexity. So we conduct comprehensive experiments to explore the relation between complexity and performance.

In this experiment, we focus on four hyper-parameters as mentioned before: hidden layer size(H), input data size(I), time steps(T), and number of cells in each time step(N). we use Mean Square Error as the metric of Performance. According to the diagram of LSTM, we compute the complexity of each configuration i using the following formula:

$$C_i \sim O(H_i \times (H_i + I_i)^2 \times N_i \times T_i) \qquad (8)$$

where $H_i \times (H_i + I_i)^2$ means computation complexity of one cell in LSTM model. As mentioned before, in each time step, LSTM typically stacks N cells to improve its' fitting ability. And only After T time steps, the prediction can be generated.

Fig. 9 shows a scatter plot of Complexity vs. Performance of several hundreds configurations. We use Stator outlet temperature data to perform this experiment. There



(a) Stator outlet temperature (b) Cold water flow (c) Stator inlet temperature (d) Active power

Figure 8. Performacne of compressive sensing data filling on GSD.

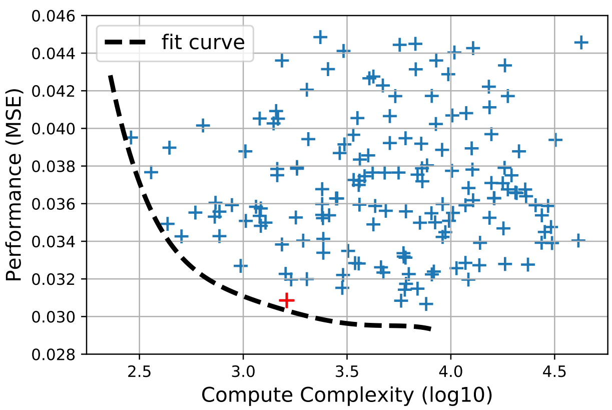


Figure 9. Complexity-Performance.

is two orders of magnitude of difference in Complexity. The black dashed line is the Pareto Boundary. Given certain performance requirement, the configurations near this boundary are considered as more efficient ones. For example, if we require Mean Square Error of prediction of Stator outlet temperature is less than 0.031, the configuration, corresponding to the point in red in Fig. 9, is the best choice. In other words, only configurations that near the Pareto Boundary of Complexity-Performance space will be considered.

5 Conclusion and Future work

Conclusion. In this paper, we present a multi-sensing collaborative analysis system for rotating machinery time-series data. We estimate the value of sensors and the differences between predicted and real value imply the operational condition of machine. We propose a data filling method based on compressive sensing for processing heterogeneous and asynchronous sensor readings of generator and compensator. Processed time series are used to extract principle features for value estimation. We implement our system and evaluate its performance on real world generator sensor data. Experimental results demonstrate that our system outperform traditional model-based methods and achieve less than 2% mean square error.

Future work. In our expriment, we suppose the sampling frequency of each sensor is fixed. In other words, the time interval between two consecutive records is equal. In fact, the sampling frequency of each sensor is not always the same, it will have some fluctuations. We will take dynamic change of time interval into consideration in our LSTM network by adding a time decay factor in it.

We select parameter configuration of LSTM experimentally. In GSD and CSD, the parameter selection problem can be formalized as follows.

$$\min\ z = \sum_{i=1}^{N} P_i(r_i)$$
$$s.t.\quad \begin{cases} \sum\limits_{i=1}^{N} r_t \leq R, \\ P_i(r_i) \leq \varepsilon_i \quad (i = 1, 2, 3, \cdots, n). \end{cases} \tag{9}$$

where R is the CPU computing resource, N is total sensors to be estimated, each estimator consumes $r_i, i = 1, 2, \cdots, N$, and its Pareto boundary is P_i, ε_i are given optimization error cap. We expect the solution of (9) will produce a better result.

6 Acknowledgements

This work was supported by the State Grid of China Science and Technology Fund No.52110417000G.

7 References

[1] Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. Show and tell: A neural image caption generator. In *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*. IEEE, 2015.

[2] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. Show, attend and tell: Neural image caption generation with visual attention. In *International Conference on Machine Learning*, 2015.

[3] Jeffrey Donahue, Lisa Anne Hendricks, Sergio Guadarrama, Marcus Rohrbach, Subhashini Venugopalan, Kate Saenko, and Trevor Darrell. Long-term recurrent convolutional networks for visual recognition and description. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015.

[4] Inci M Baytas, Cao Xiao, Xi Zhang, Fei Wang, Anil K Jain, and Jiayu Zhou. Patient subtyping via time-aware lstm networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2017.

[5] Trang Pham, Truyen Tran, Dinh Phung, and Svetha Venkatesh. Deepcare: A deep dynamic memory model for predictive medicine. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 2016.

[6] Marcus Liwicki, Alex Graves, Santiago Fernàndez, Horst Bunke, and Jürgen Schmidhuber. A novel approach to on-line handwriting recognition based on bidirectional long short-term memory networks. In *Proceedings of the 9th International Conference on Document Analysis and Recognition, ICDAR 2007*, 2007.

[7] Yuan He, Junchen Guo, and Xiaolong Zheng. From surveillance to digital twin: Challenges and recent advances of signal processing for industrial internet of things. *IEEE Signal Processing Magazine*, 35(5):120–129, 2018.

[8] Junchen Guo, Yuan He, and Xiaolong Zheng. Pangu: Towards a software-defined architecture for multi-function wireless sensor net-

works. In *Parallel and Distributed Systems (ICPADS), 2017 IEEE 23rd International Conference on*, pages 730–737. IEEE, 2017.

[9] Xin Miao, Kebin Liu, Yuan He, Dimitris Papadias, Qiang Ma, and Yunhao Liu. Agnostic diagnosis: Discovering silent failures in wireless sensor networks. *IEEE transactions on wireless communications*, 12(12):6067–6075, 2013.

[10] Rui Li, Ke-Bin Liu, Xiangyang Li, Yuan He, Wei Xi, Zhi Wang, Ji-Zhong Zhao, and Meng Wan. Assessing diagnosis approaches for wireless sensor networks: Concepts and analysis. *Journal of Computer Science and Technology*, 29(5):887–900, 2014.

[11] Yilun Zheng, Yuan He, Meng Jin, Xiaolong Zheng, and Yunhao Liu. Red: Rfid-based eccentricity detection for high-speed rotating machinery.

[12] Xuefeng Chen, Shibin Wang, Baijie Qiao, and Qiang Chen. Basic research on machinery fault diagnostics: Past, present, and future trends. *Frontiers of Mechanical Engineering*, 2017.

[13] Hui Ma, Xu Pang, Ranjiao Feng, Rongze Song, and Bangchun Wen. Fault features analysis of cracked gear considering the effects of the extended tooth contact. *Engineering Failure Analysis*, 2015.

[14] Zehua Hu, Jinyuan Tang, Jue Zhong, and Siyu Chen. Frequency spectrum and vibration analysis of high speed gear-rotor system with tooth root crack considering transmission error excitation. *Engineering Failure Analysis*, 2016.

[15] RB Randall. Applications of spectral kurtosis in machine diagnostics and prognostics. In *Key Engineering Materials*, 2005.

[16] Yaguo Lei, Jing Lin, Zhengjia He, and Ming J Zuo. A review on empirical mode decomposition in fault diagnosis of rotating machinery. *Mechanical Systems and Signal Processing*, 2013.

[17] Robert B Randall and Jerome Antoni. Rolling element bearing diagnosticsa tutorial. *Mechanical systems and signal processing*, 2011.

[18] Jiesi Luo, Dejie Yu, and Ming Liang. A kurtosis-guided adaptive demodulation technique for bearing fault detection based on tunable-q wavelet transform. *Measurement Science and Technology*, 2013.

[19] Hui Li, Lihui Fu, and Zhentao Li. Gear fault detection using angle domain average and hilbert-huang transform phase map. In *Image and Signal Processing, 2009. CISP'09. 2nd International Congress on*, 2009.

[20] Yanxue Wang and Ming Liang. An adaptive sk technique and its application for fault detection of rolling element bearings. *Mechanical Systems and Signal Processing*, 2011.

[21] Thomas Cover and Peter Hart. Nearest neighbor pattern classification. *IEEE transactions on information theory*, 1967.

[22] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 1995.

[23] M van Gerven and S Bohte. Editorial: Artificial neural networks as models of neural information processing. *Artificial Neural Networks as Models of Neural Information Processing*, 2018.

[24] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 1997.

[25] Yuting Wu, Mei Yuan, Shaopeng Dong, Li Lin, and Yingqi Liu. Remaining useful life estimation of engineered systems using vanilla lstm neural networks. *Neurocomputing*, 275:167–179, 2018.

[26] Haitao Zhao, Shaoyuan Sun, and Bo Jin. Sequential fault diagnosis based on lstm neural network. *IEEE ACCESS*, 6:12929–12939, 2018.

[27] Zhe Zhou, Chenglin Wen, and Chunjie Yang. Fault isolation based on k-nearest neighbor rule for industrial processes. *IEEE Transactions on Industrial Electronics*, 2016.

[28] Ruonan Liu, Boyuan Yang, Xiaoli Zhang, Shibin Wang, and Xuefeng Chen. Time-frequency atoms-driven support vector machine method for bearings incipient fault diagnosis. *Mechanical Systems and Signal Processing*, 2016.

[29] Alireza Sadeghian, Zhongming Ye, and Bin Wu. Online detection of broken rotor bars in induction motors by wavelet packet decomposition and artificial neural networks. *IEEE Transactions on Instrumentation and Measurement*, 2009.

[30] Rui Zhao, Jinjiang Wang, Ruqiang Yan, and Kezhi Mao. Machine health monitoring with lstm networks. In *Sensing Technology (ICST), 2016 10th International Conference on*, 2016.

[31] Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory*, 52(2):489–509, 2006.

[32] Gang Tang, Qin Yang, Hua-Qing Wang, Gang-gang Luo, and Jianwei Ma. Sparse classification of rotating machinery faults based on compressive sensing strategy. *Mechatronics*, 2015.

[33] Xuefeng Chen, Zhaohui Du, Jimeng Li, Xiang Li, and Han Zhang. Compressed sensing based on dictionary learning for extracting impulse components. *Signal Processing*, 2014.

[34] Xuefeng Chen, Gaigai Cai, Hongrui Cao, and Wei Xin. Condition assessment for automatic tool changer based on sparsity-enabled signal decomposition method. *Mechatronics*, 2015.

[35] Emmanuel J Candès et al. Compressive sampling. In *Proceedings of the international congress of mathematicians*, volume 3, pages 1433–1452. Madrid, Spain, 2006.

[36] Emmanuel J Candes and Terence Tao. Decoding by linear programming. *IEEE transactions on information theory*, 51(12):4203–4215, 2005.

[37] Joel A Tropp and Anna C Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on information theory*, 53(12):4655–4666, 2007.

[38] Deanna Needell and Roman Vershynin. Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit. *IEEE Journal of selected topics in signal processing*, 4(2):310–316, 2010.

[39] Richard G Baraniuk. Compressive sensing [lecture notes]. *IEEE signal processing magazine*, 24(4):118–121, 2007.