Computers & Industrial Engineering 136 (2019) 494–503

Contents lists available at ScienceDirect

Computers & Industrial Engineering

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

# Real-time quality monitoring and diagnosis for manufacturing process profiles based on deep belief networks

Yumin Liu[a], Haofei Zhou[b], Fugee Tsung[c], Shuai Zhang[a,*]

[a] Dept. of Management, Zhengzhou University, Zhengzhou, China
[b] Dept. of Management Science, Zhengzhou University of Aeronautics, Zhengzhou, China
[c] Dept. of Industrial Engineering and Logistics Management, Hong Kong University of Science and Technology, Hong Kong

ARTICLE INFO

*Keywords:*
Manufacturing process profile
Monitoring and diagnosis
Quality spectrum
Deep belief network

ABSTRACT

A large number of real-time quality data are collected through various sensors in the manufacturing process. However, most process data are high-dimension, nonlinear and high-correlated, so that it is difficult to model the process profiles, which restricts the application of conventional statistical process control technique. Motivated by the powerful ability of deep belief network (DBN) to extract the essential features of input data, this paper develops a real-time quality monitoring and diagnosis scheme for manufacturing process profiles based on DBN. The profiles collected from a manufacturing process are mapped into quality spectra. A novel DBN recognition model for quality spectra is established in the off-line learning phase, which can be applied to monitor and diagnose the process profiles in the on-line phase. The effectiveness of DBN recognition model for manufacturing process profiles is demonstrated by simulation experiment, and a real injection molding process example is applied to analyze the performance. The results show that the proposed DBN model outperforms alternative methods.

## 1. Introduction

Many sensors are applied in a modern production line to collect huge amounts of process data. In examples, more than million pieces of tonnage signal data per minute are acquired from an progressive stamping process (Zhou, Liu, Zhang, Zhang, & Shi, 2016), and in more than 1034 oven runs, 160 temperature data points in each run are obtained within 500-second data collection period (Jensen, Grimshaw, & Espen, 2016). With the rapid development of data acquisition technology, massive process data, which reflects the operating status of a process, can be collected to monitor and improve the process performance and production quality (Ge, Xu, & Du, 2008; Grasso, Menafoglio, Colosimo, & Secchi, 2016; Pacella, 2018; Qin, 2014; Wang, Kim, Huo, Hur, & Wilson, 2015). The quality of a process is better characterized by the relationship between a response variable and one or more explanatory variables rather than the distribution of variables (Woodall, Spitzner, Montgomery, & Gupta, 2004). Monitoring such a relationship is usually called profile monitoring problem. How to monitor these profiles by using effective methods has become a hot topic, which could significantly improve the product quality of a manufacturing process.

In the existing literatures, the research methods of monitoring profile can be divided into two categories: linear and nonlinear. Linear profiles are widely employed in some manufacturing processes, such as calibration process and stamping processes. In order to detect a mean shift or other abnormal changes in linear profiles, control charts are applied to monitor some estimated parameters, such as intercepts and slopes (Kang & Albin, 2000; Noorossana, Eyyazian, Amiri, & Mahmoud, 2010; Zou, Tsung, & Wang, 2007). However, the linear assumption is not always valid in practice. In the recent years, various methods have been proposed to address the nonlinear profile monitoring problem for some manufacturing process (Kazemzadeh, Noorossana, & Amiri, 2008; Chen & Nembhard, 2011; Grasso et al., 2016). Chou, Chang, and Tsai (2014) developed a proper process monitoring strategy for monitoring multiple correlated nonlinear profiles. Zhang, Ren, Yao, Zou, and Wang (2015) studied multivariate profiles based on the regression adjustment method. Besides, Jensen et al. (2016) proposed a nonlinear model for reducing the oven temperature profiles in a manufacturing process to a smaller set of parameters estimated.

Implementation of the above monitoring profile schemes usually assumes a baseline parametric model of the specific distribution, against which new observations are compared in order to identify significant departures from the baseline model. The baseline model is often unknown and estimated from a limited phase I sample in practice. In order to warrant the accurate of the baseline model, a sufficiently

* Corresponding author.
*E-mail address:* zhang1227zzu@163.com (S. Zhang).

https://doi.org/10.1016/j.cie.2019.07.042
Received 29 May 2018; Received in revised form 24 March 2019; Accepted 23 July 2019
Available online 24 July 2019
0360-8352/ 

large reference sample of in-control data is needed. However, in many cases, the reference samples may be small and, therefore, it may not be possible to accurately estimate the in-control parameters, which strongly deteriorate the performance of the monitoring methods (He, Jiang, & Deng, 2016).

When the reference sample collection is limited, one has to apply change point model to monitor the process (Hawkins & Qiu, 2003), which isn't necessary to construct a baseline model. Nevertheless, most change-point control charts assume that the data follow a Gaussian distribution. Due to the complexity of the production process and the different types of sensors used, the normal distribution assumption may be violated in these process applications (Fan, Han, & Liu, 2014).

It would be a challenging work to monitor a process without a baseline model, especially when the targeting process doesn't follow a normal distribution. Fortunately, intelligent approaches based on artificial neural network (ANN) (Zhou et al., 2016) and support vector machine (SVM) (Liu & Zhou, 2015, 2016) have been widely applied for manufacturing process monitoring and fault diagnosis due to their advantages in handling complex data. Such techniques can effectively overcome the limitations of conventional statistical process control. These intelligent methods can be naturally extended to monitoring profiles. Pacella and Semeraro (2011) proposed a roundness profiles monitoring model based on unsupervised neural network algorithm. Hosseinifard, Abdollahian, and Zeephongsekul (2011) monitored profiles via an ANN approach. Atashgar, Amiri, and Nejad (2015) monitored Allan variance nonlinear profile using artificial neural network approach.

The existing researches transform the process monitoring problem to a supervised learning problem so that any classifiers can be used. However, to our best knowledge, the deep belief network (DBN) model has not been applied in monitoring manufacturing process profiles.

Considering the situation that we have both in-control and out-control dataset, which don't follow any distributions, and have no baseline model, we propose a nonparametric process monitoring scheme for manufacturing process profiles based on DBN. The pattern between a sample covered with moving windows from the input stream and the reference dataset is compared by the DBN. Data experiments show that the proposed DBN-based monitoring scheme has some advantages and outperforms than the existing methods.

The key contributions of this work are as follows. Firstly, the proposed DBN model is a nonparametric process monitoring procedure, which can be used to monitor any processes without distribution assumption. Secondly, the DBN model is different from conventional distance-based process monitoring scheme such as the $T^2$ chart, since the monitoring metric defined in the DBN model is compared the patterns between a sample covered with moving windows from the input stream and the reference, whereas the conventional method measures the distance between the centers of the reference dataset and the moving window dataset. Thirdly, considering the limited and imbalanced reference dataset in real applications, the structure of proposed DBN model is dynamic and can be adjusted based on the sample size of the reference.

The remainder of this paper is organized as follows. In Section 2, a motivating example of injection molding process is introduced. In Section 3, the architecture of the Gauss-Bernoulli restricted Boltzmann machine (GBRBM) and a novel DBN model for recognizing the quality spectrum are proposed. Besides, the DBN framework for monitoring and diagnosing the manufacturing process profiles is developed, which includes off-line learning phase and the on-line monitoring and diagnosis phase. In Section 4, the effectiveness analysis of proposed DBN model is demonstrated by the simulated data. The optimal structure of DBN model and the effect of weighting parameter are tested. In Section 5, a real example of monitoring and diagnosing profiles from the injection molding process is applied and the performance of proposed method is compared with alternative methods. Suggestions and directions for further research are discussed towards the end of the paper.

## 2. Process profiles and quality spectra

### 2.1. A motivating example

The injection molding process is a typical multivariate manufacturing process and has three major characteristics: rapid data variation, high sampling rate and multiple monitoring variables. Usually, the injection molding process has four stages: padding, packing, cooling and demolding. In the padding stage, the shift in temperature and pressure in the mold cavity may generate weld marks, which are the common defects. Weld marks not only affect the product appearance, but also break the microstructure. The concentrated stress causes the injection product to rupture easily. Thus, the operational conditions of temperature and pressure should be monitored in the padding stage. In order to prevent the shrinkage of injection products, the pressure in the mold cavity is increased sequentially to improve the product density in the packing stage. However, if the product is too densely packed, defects will appear, such as flash and overflow. In this stage, the pressure is a critical monitoring variable. In the cooling stage, the injection product is solidified to improve the product quality further. Usually, cooling imbalance will give rise to warpage. Hence, the temperature is also an important variable that needs to be monitored. In the last demolding stage, the mold is opened using a proper method to ensure quality of production.

One reciprocating screw injection molding process from a Chinese manufacturing company motivated us to perform this study, and the screw injection molding machine is shown in Fig. 1. The temperature of the charging barrel and mold, and the pressure in the mold can be measured by the corresponding sensors over time. The locations of these sensors are shown in Fig. 1.

There are 10 quality monitoring variables in the injection process, denoted by $x_i$, (i = 1, 2, …10), which are measured by seven temperature sensors, denoted by $ST_j$, (j = 1, 2, …7), and three pressure sensors, denoted by $SP_j$, (j = 1, 2, 3), shown in Table 1.

In reciprocating injection molding, the total processing time is 20 s. The quality of production is characterized by the temperature or pressure where the main challenge lies in how to monitor these variables during the injection process.

### 2.2. Manufacturing process profile

Usually, the profile is well known as a set of measurements with a response variable and some explanatory variables (Woodall, 2007) and is formed by the data sequence of quality characteristics (Fan, Jen, & Lee, 2016; Jensen et al., 2016). Such profiles include the thickness measurements of semiconductor wafers from a slicing process (Liu, Jin, & Kong, 2018), and radius measurements as a function of a turning process (Colosimo, Semeraro, & Pacella, 2008). In this paper, the process profile can be defined as a set of measurements from a manufacturing process with multiple quality monitoring variables over time. The process profile monitoring requires more attention to be paid to the sequential data streams from the manufacturing process than to the

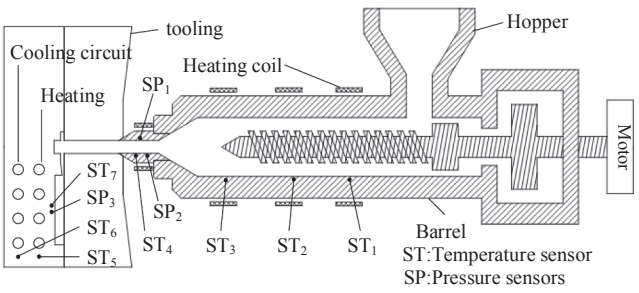


**Fig. 1.** A reciprocating screw injection molding process.

**Table 1**
Quality monitoring variables.

| No. | Variable name | Sensor | Target value |
|---|---|---|---|
| $x_1$ | Barrel temperature 1 | $ST_3$ | 160 °C |
| $x_2$ | Nozzle temperature | $ST_4$ | 120 °C |
| $x_3$ | Barrel temperature 2 | $ST_2$ | 200 °C |
| $x_4$ | Heating water temperature | $ST_5$ | 140 °C |
| $x_5$ | Barrel temperature 3 | $ST_1$ | 200 °C |
| $x_6$ | Cooling water temperature | $ST_6$ | 25 °C |
| $x_7$ | Internal mold temperature | $ST_7$ | 30 ~ 130 °C |
| $x_8$ | Injection pressure | $SP_1$ | 90 MPa |
| $x_9$ | Holding pressure | $SP_2$ | 60 MPa |
| $x_{10}$ | Internal mold pressure | $SP_3$ | 0 ~ 60 MPa |

quality characteristics.

With the sensors installed in the production line, a large number of observations can be obtained during the manufacturing process. The manufacturing process profile can be represented by the observation matrix $X$, which indicates the running status during the process. The process profile is shown in Eq. (1),

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ x_{m1} & \cdots & \cdots & x_{mn} \end{bmatrix} \tag{1}$$

where $x_{ij}$ (i = 1, 2, …, m, j = 1, 2, …, n) is the jth measurement of the ith monitoring variable during the process.

The observations from the injection molding process form a typical manufacturing process profile. Examples of injection molding process profiles are shown in Fig. 2. In these examples, there are 10 monitoring variables during the process so $m = 10$. The whole process lasts 20 s and the sampling frequency is 20 ms, which means that 1000 measurements of each monitoring variable can be observed from the normal process and the abnormal process with flash defect respectively, so $n = 1000$.

The monitoring window size is 20 s in the injection example. The

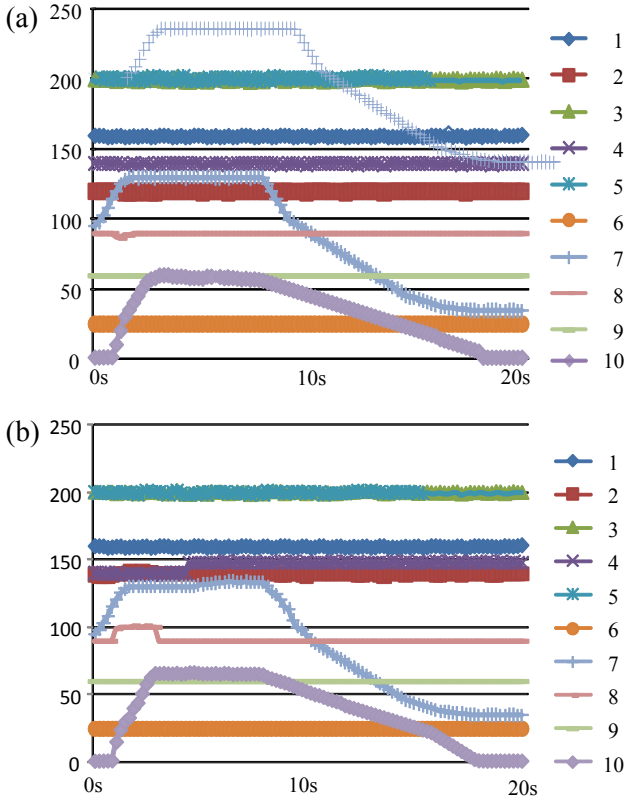


**Fig. 2.** Process profiles for 10 monitoring variables.

choice of window size is a critical parameter in the manufacturing process, because the window size is related to the shape of abnormal patterns that we want to detect. However, all kinds of abnormal patterns can be seen in a whole process at least if the defects happened. Thus, we suggest that the length of the manufacturing process be taken as the window size, such as the 20 s in the injection example.

The 10 quality monitoring variables in Fig. 2 are nonlinear and correlated. Due to their complex relationship, it is difficult to observe the difference between the profiles without defect and the ones with flash directly in Fig. 2. Fortunately, a gray-scale image is a useful tool to represent information from a process profile, and it will be introduced in the following.

### 2.3. Quality spectrum of the process profile

For reasons mentioned above, it is difficult to detect a shift from a process profile directly. Even though the signal of a shift can be detected by quality techniques, the root cause of an abnormal status cannot be traced without visualization, which poses a great challenge to multivariate process diagnosis. In the literatures, some equipment failure data, which are highly correlated and nonlinear, can be visualized using a digital image (Amar, Gondal, & Wilson, 2015; Sun, Gao, Gao, Gao, & Wang, 2015).

A grayscale image is formed by the grayscale values between black and white, and can represent data much more accurately than a binary image. In this paper, the process profiles can be mapped into grayscale images based on the operating conditions of the manufacturing process. Such grayscale images can present the normal or abnormal running status of the process, as well as show the boundary or the edge if the observations in the data matrix change, so that it is easy to detect edges or determine abnormal causes using a DBN algorithm. The grayscale images can intuitively present the quality information implied by the process profiles, so the image can be defined as a quality spectrum. In a quality spectrum, the current process is stable if the gray stripe is uniform and the grayscale changes smoothly, which is called a normal quality spectrum. In an abnormal quality spectrum, the gray stripe is mutated and some significant boundaries can be found, which indicates that an abnormal change has occurred in the process.

For generating the grayscale image of a process profile, the observations in a observation matrix, $X$, should be standardized first, which is given by

$$x'_{ij} = \frac{x_{ij} - \min\limits_{j}(x_{ij})}{\max\limits_{j}(x_{ij}) - \min\limits_{j}(x_{ij})} \tag{2}$$

$x'_{ij}$ can be mapped into a grayscale using Eq. (3), which means a grayscale image can be formed by mapping all of the observations.

$$g_x = \mathrm{INT}[x'_{ij} \times 255] \tag{3}$$

In Eq. (3), $g_x$ is the grayscale vaule of $x'_{ij}$, and $g_x \in [0, 255]$. The grayscale image is black when $g_x = 0$ and white when $g_x = 255$. So the process profile has been represented as a grayscale image to visualize the operating status of the manufacturing process.

In the motivating example, two common defects, flash and warpage, often occurred in the process. According to historical data, these defects indicate six abnormal statuses of the process. Under a different abnormal process, the quality spectrum of a process profile will be different. The corresponding relationships between the process status and the spectrum are shown in Table 2.

In order to monitor the normal profile, 10 monitoring variables are continuously sampled by every 20 ms in 20 s from the normal process of injection molding. In the same way, six abnormal profiles can be collected from the corresponding process status.

According to Eqs. (2) and (3), a normal and six abnormal process profiles of the injection molding can be mapped into their quality spectra respectively. The normal and two abnormal quality spectra, as

**Table 2**
The corresponding relationships between the process and spectrum.

| Defect | Process status | Quality spectrum |
|---|---|---|
| None | Normal | $F_0$ |
| Flash | Upward shift in injection pressure | $F_{11}$ |
| | Upward shift in nozzle temperature | $F_{12}$ |
| | Upward shift in injection pressure Upward shift in nozzle temperature | $F_{13}$ |
| Warpage | Upward shift in heating water temperature | $F_{21}$ |
| | Upward shift in cooing water temperature | $F_{22}$ |
| | Upward shift in heating water temperature | $F_{23}$ |
| | Upward shift in cooling water temperature | |

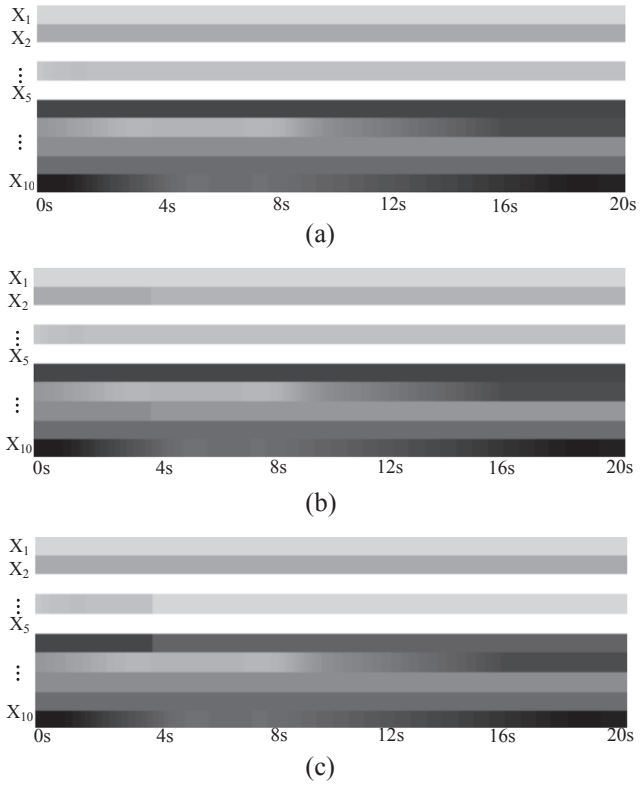


**Fig. 3.** Quality spectra of injection process profiles.

seen in Fig. 3, are chosen for illustration purpose.

In the normal quality spectrum, the grayscales of each monitoring variable change naturally and smoothly, and there are no obvious changes in brightness in Fig. 3(a). In Fig. 3(b), the grayscales of $X_2$ and $X_8$ change significantly in the quality spectrum. Thus, the grayscale value of the image becomes smaller after a shift, which can generate a distinct boundary.

The variation in the grayscale from dark to light indicates that the observations of $X_2$ and $X_8$ are shifted by a special cause. In other words, the nozzle temperature and injection pressure of the current process are in a state of upward shift. If the injection pressure shifts suddenly in the process, the capacity of melting materials in the mold cavity will increase, so that the material will overflow which usually leads to flash. If the nozzle temperature rises abruptly during the injection process, the melting degree of the material in the mold cavity will be increased, also leading to a flash.

The process profiles collected from injection modeling can be represented precisely by the quality spectrum, $F_{13}$, which describes how flash occurs when the nozzle temperature and injection pressure are shifted.

In Fig. 3(c), the grayscales of $X_4$ and $X_6$ change obviously in the quality spectrum $F_{23}$. The variation in the grayscale from dark to light indicates that the observations of $X_4$ and $X_6$ are shifted by a special cause. In other words, both heating water temperature and cooling water temperature of the current process are in a state of upward shift. In the injection molding process, if the control of heating and cooling is mismatched, the stress of the injection product will be uneven, which will lead to warpage. The process profiles collected from injection modeling can be represented precisely by their quality spectra which describe how warpage occurs when the heating water temperature and cooling water temperature are shifted.

From the above discussion, the normal and abnormal quality profiles in a manufacturing process can be visualized clearly and accurately by their quality spectra. How to identify the abnormal quality spectra is the next problem in the profile monitoring and diagnosis.

## 3. The DBN recognition model for quality spectra

### 3.1. The structure of the DBN model

The DBN is an advanced network based on deep learning and can be applied to analyze high-dimensional data sequences, such as image or motion capture data (Hinton, Osindero, Welling, & Teh, 2006; Najafabadi et al., 2015). DBN is a generative model with multiple layers, each of which is made up of an RBM. An RBM is a two-layer network, which only has good recognition ability for binary images. However, the grayscale image of a process profile is not a binary image, so that a novel DBN model should be developed for recognizing the quality spectra of a process profile.

The GBRBM is a two-layer network and has good recognition ability for non-binary images (He, Wang, Li, & Zhou, 2016). Therefore, the GBRBM is first applied to recognize the quality spectrum in the DBN model instead of the RBM in this paper. The architecture of the GBRBM is shown in Fig. 4.

As stated above, the novel DBN used in this paper is composed of many GBRBMs and an output layer. The weight matrices and bias vectors of GBRBMs can be determined by training GBRBMs sequentially. A GBRBM is made up of a visible layer $\mathbf{v}$ and a hidden layer $\mathbf{h}$, where $v_i$ is the ith visible unit (i = 1,2,…,p) of $\mathbf{v}$, $h_j$ is the jth hidden unit (j = 1,2,…,q) of $\mathbf{h}$, $a_i$ and $b_j$ are their biases and $w_{ij}$ is the weight between them. All visible units are connected to all hidden units, and there are no connections between any two units in the same layer. The visible units of GBRBM are linear units with independent Gaussian noise, whereas the hidden units are binary stochastic units. GBRBM is an energy-based stochastic neural network. The joint probability distribution over visible layer $\mathbf{v}$ and hidden layer $\mathbf{h}$ is defined by an energy function.

The energy function of the GBRBM is given by

$$E(v, h; \varphi) = \sum_{i=1}^{m} \frac{(v_i - a_i)^2}{2\sigma_i^2} - \sum_{j=1}^{n} b_j h_j - \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{v_i}{\sigma_i} w_{ij} h_j \quad (4)$$

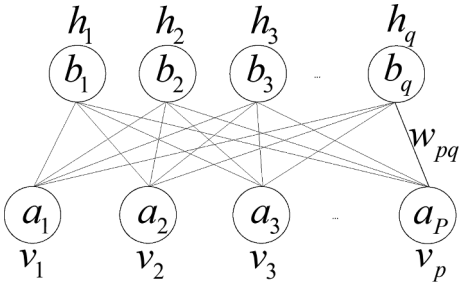


**Fig. 4.** Architecture of GBRBM.

where $\varphi = \{w_{ij}, a_i, b_j\}$ is the parameter set of the GBRBM. $\sigma_i$ is the standard deviation of the Gaussian noise for the visible unit i. The joint probability density function of the visible layer and a hidden layer, (**v**, **h**), can be obtained by Eq. (5).

$$p(v, h; \varphi) = \frac{1}{Z(\varphi)} \exp(-E(v, h; \varphi)) \tag{5}$$

$Z(\varphi) = \int \sum_h \exp(-E(v, h; \varphi))dv$ is the partition function. The marginal density function of the visible layer **v** is given by summing over all possible hidden units.

$$p(v; \varphi) = \sum_h \frac{\exp(-E(v, h; \varphi))}{\int \sum_h \exp(-E(v, h; \varphi))dv} \tag{6}$$

Because there are no direct connections between hidden units in the GBRBM, they are independent of each other. So, it is very easy to obtain the probability of $h_j$ for a given v when $h_j$ is set to 1.

$$p(h_j = 1 \mid v) = \frac{1}{1 + \exp\left(-\sum_i w_{ij} \frac{v_i}{\sigma_i} - a_j\right)} \tag{7}$$

For the same reason, it is also very easy to obtain the probability of $v_i$ for a given h when $v_i$ is set to x.

$$p(v_i = x \mid h) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{(x - b_i - \sigma_i \sum_j h_j \omega_{ij})^2}{2\sigma_i^2}\right] \tag{8}$$

Because of $p(v \mid h) = \prod_i p(v_i \mid h)$, the new visible units can be re-structured.

The parameters of the GBRBM should be updated according to the following equations:

$$\Delta\omega_{ij} = \eta(\langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_{recon}) \tag{9}$$

$$\Delta a_i = \eta(\langle v_i \rangle_{data} - \langle v_i \rangle_{recon}) \tag{10}$$

$$\Delta b_j = \eta(\langle h_j \rangle_{data} - \langle h_j \rangle_{recon}) \tag{11}$$

where $\eta$, $\langle \bullet \rangle_{data}$ and $\langle \bullet \rangle_{recon}$ are the learning rate, the expectation of the original data and the expectation of the restructured data respectively.

All GBRBMs of DBN are trained by the unsupervised algorithm presented above. But the performance provided by the unsupervised training is limited and classification cannot be achieved without output layer. Thus the BP algorithm can be applied to fine-tune the weight matrices and bias vector parameters of GBRBMs and output layer to minimize the loss function, and then the DBN recognition model of quality spectra can be constructed. The architecture of the proposed DBN model for manufacturing process profiles is shown in Fig. 5. The input and output of the DBN recognition model are the visible layer of $GBRBM_1$ and the output layer $\theta$ respectively. There are n hidden layers in the DBN model, where the visible layer of one GBRBM is the hidden layer of the previous GBRBM. For example, in $GBRBM_1$, $h_1$ is the hidden layer; meanwhile, it is also the visible layer of $GBRBM_2$. The tuning parameters, such as *p*, *q* and *n* in the DBN model, would determine the structure of the recognition model.

### 3.2. A DBN recognition model for imbalance data

The fine-tuning process aims at minimizing the loss function, that is, the error of mean square (MSE) between expected output and actual output. Usually, the loss function of DBN can be expressed as

$$E_P = \frac{1}{m} \sum_{i=1}^{m} (y_{pi} - \hat{y}_{pi})^2 \tag{12}$$

where $\hat{y}_{pi}$ and $y_{pi}$ denote actual value and expected value of *i*th output neuron respectively, m is the number of output nodes. When using MSE, there is a hypothesis that the predicted errors of each output node are of equal importance. However, this assumption will hold only when the number of samples in each normal or abnormal class is approximately the same. In other words, the performance of DBN recognition model will degrade when the training samples for the normal class greatly outnumbers the ones for the abnormal class. In fact, a large number of abnormal data is unavailable to collect in a real manufacturing process. Thus the original loss function needs to be modified in order to address this imbalance problem. We add a weighting parameter $1 - \frac{N_i}{N_{total}}$ in loss function that assigns different importance to each abnormal class according to their specific sample size. Meanwhile, a penalty term $\frac{\lambda}{N_w} \sum_w |w|$ is considered to overcome the overfitting problem caused by small sample size of abnormal class, where *w* denotes the weights in the DBN recognition model, and $N_w$ is the number of weights in DBN. The loss value of the *p*th process quality spectra is defined as

$$E_P = \sum_{i=1}^{m} \left(1 - \frac{N_i}{N_{total}}\right)(y_{pi} - \hat{y}_{pi})^2 + \frac{\lambda}{N_w} \sum_w |w| \tag{13}$$

where $N_{total}$ represents the total number of training samples. $N_i$ is the sample size of training set that belongs to *i*th class of quality spectra. λ is the tuning parameter for the sparsity of weights between neuron. The proposed loss function has two properties. First, the weighting parameter provides the minority class with more importance in loss function. In this way, the misclassification of the samples from minority class can be punished more severely than that from majority class. Second, the added penalty term can reduce the complexity of DBN recognition model, so that the overfitting problem can be solved.

### 3.3. The framework for monitoring and diagnosing process profiles

A framework for monitoring and diagnosing manufacturing process profiles using the DBN recognition model (DBNRM) is proposed in this subsection. This framework can be divided into phase I and phase II. In phase I, it aims to train the DBNRM off-line from the collected quality spectrum data and obtain an appropriate DBN model for recognizing quality spectra. After learning in phase I, this DBNRM can be applied to monitor and diagnose the manufacturing process profile on-line in phase II. The framework is shown in Fig. 6, and the details are introduced in the following.

The goal in phase I is to generate a DBN recognition model for abnormal quality spectrum using learning algorithms. The off-line learning procedure includes two steps as follows: The first step is to train each GBRBM in DBN recognition model. Unlabeled process quality spectrum is used to train GBRBMs layer by layer and the parameters can be updated. The second step is to fine-tune the whole DBN network using BP algorithm. A few labeled process quality spectra are used to train output layer and further update GBRBMs.

After training in the phase I, the DBN recognition model can be applied to monitor and diagnose the process profiles in the phase II.

The purpose of phase II is to monitor and diagnose the quality spectra on-line with the DBN recognition model trained in phase I. The key problem in this phase is to identify the abnormal quality spectra and determine the root causes of abnormal process profiles simultaneously. There are three main steps in this phase.

The first step is to collect the profiles from a real-time process. The profiles are collected through the monitoring window. The size of the monitoring window should be determined by the product processing time, so that the profiles can be mapped into a suitable spectrum.

The second step is to monitor the process profiles. The quality spectra in the current monitoring window will be recognized by the DBNRM trained in phase I. If the predicted value of the first output node in this model is close to 1, the quality spectrum is the normal one and thus the process is in control. Then the sliding window will move forward to collect new observations until the DBNRM triggers an alarm. If any predicted value of other output nodes is larger than first node, the spectrum is abnormal, which means the process is out of control. Therefore, it is necessary to diagnose the potential abnormal variables

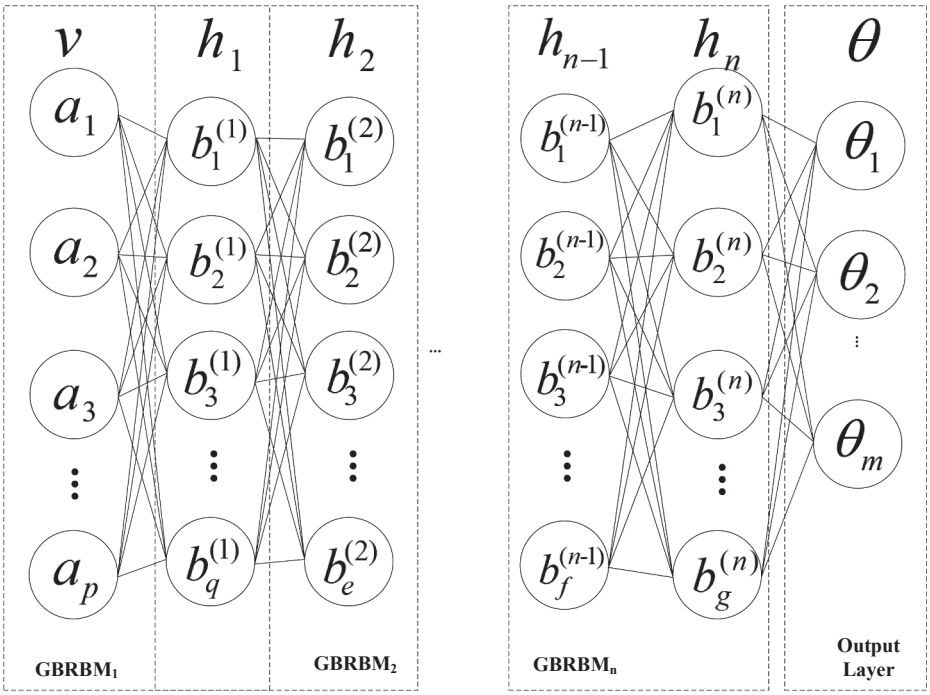


**Fig. 5.** Architecture of the proposed DBN model.

which can be determined from the abnormal quality spectrum.

The last step is to diagnose the abnormal quality spectra. The abnormal class is recognized by the DBNRM and the root cause is found.

According to the above phase I and phase II, the manufacturing process profiles can be monitored and diagnosed simultaneously in real time. It benefits from the collection of real-time quality profiles and the powerful learning ability of the DBN recognition model.

## 4. Effectiveness analysis

The effectiveness of the proposed method depends on the structure of DBNRM and the setting of important parameters, which will be investigated by the simulated experiments in this section. To evaluate the ability of detecting the abnormal pattern, the accuracy of test data (ATD) index is applied, which ATD is a ratio of the number of samples

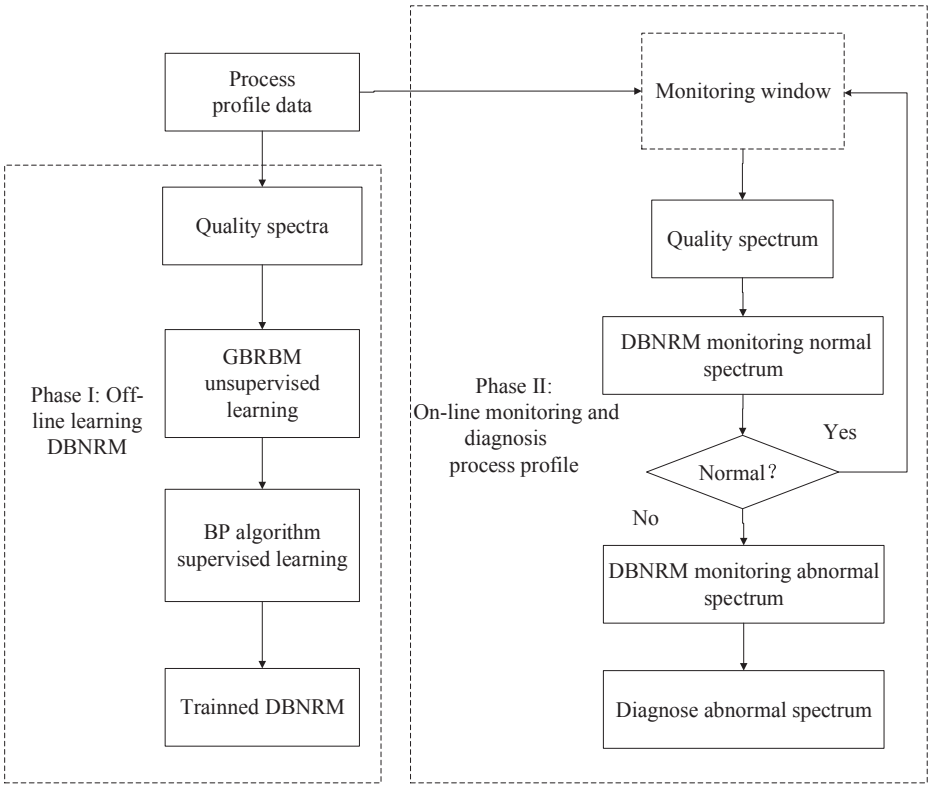


**Fig. 6.** Monitoring and diagnosis framework of DBNRM.

recognized correctly and the total number of samples. When a process is out-of-control, the DBNRM will be expected to detect the abnormal pattern as precisely as possible. Meanwhile, when the process is in control, the DBNRM also wants to recognize the normal pattern more accurately, so that the higher accuracy indicates the better effectiveness. We compare the accuracy of test data for a given training error = 0.1. So, larger ATD indicates better effectiveness. The following experiments consist of different scenarios to show the effectiveness under the different DBNRM structures. In all the experiments, the training data and test data are generated from an injection molding process of display panel in ModeFlow 2015. These data not only do not follow any distribution but also are difficult to found a baseline model.

### 4.1. The structure of DBNRM

The structure of DBNRM is determined by the numbers of hidden layer and hidden layer nodes. A lot of studies indicate that the DBN model with more hidden layers outperforms than one. However, too many hidden layers may result in inefficiency and overfitting. Thus the structure of DBNRM will be determined layer by layer as below.

We first consider the initialization of DBNRM with one hidden layer. Twenty scenarios are considered during the initialization, in which the hidden layer nodes ranging from 100 to 2000. The best recognition accuracy can be inferred with one hidden layer. Then, DBN models with two hidden layers are initialized. Since the best accuracy of 1-hidden layer DBNRM occurs before, the number of first layer nodes is set to find the optimal structure in the 2-hidden layer DBNRM. In addition, the number of the second hidden layer nodes ranging from 100 to 2000 is checked in the initialization. In the similar way, the different structures of DBNRM would be tested layer by layer. The accuracy we expected will increase until the best structure finding out. If the accuracy is decreasing, the former structure of DBNRM would be selected as the best structure. Besides, 15,000 data sets of normal pattern, abnormal pattern are generated respectively. Among that, 5000 sets are training data, and the left are test data. The result is shown in Table 3.

Table 3 shows the variation of accuracy as changing the hidden-layer numbers, where the boldface values represent the highest accuracy in each layer. It can be inferred that the prediction accuracy obviously improves with the increase in hidden-layer numbers at early stage. The recognition accuracy is optimal when hidden-layers number is 4. Therefore, four hidden layers are best chosen for the DBNRM. The optimal structure of the DBNRM is determined as 6000-1400-1500-700-200-2, in other words, the structure with six thousand input nodes, four hidden layers with 1400, 1500, 700, and 200 hidden nodes, and two output node, is the optimal one by training. Then, the DBNRM with this optimal structure will be applied to determine its important parameter in the following.

**Table 3**
Effects of the structure on recognition accuracy.

| The number of nodes in hidden layer | Recognition accuracy of test data (%) | | | | | |
|---|---|---|---|---|---|---|
| | 1st hidden layer | 2nd hidden layer | 3rd hidden layer | 4th hidden layer | 5th hidden layer | 6th hidden layer |
| 100 | 91.34 | 97.33 | 98.55 | 99.56 | 99.58 | 99.54 |
| 200 | 92.78 | 97.64 | 98.67 | **99.61** | 99.58 | **99.56** |
| 300 | 93.62 | 97.83 | 98.75 | 99.58 | **99.6** | 99.48 |
| 400 | 94.21 | 97.98 | 98.82 | 99.58 | 99.54 | 99.36 |
| 500 | 94.88 | 98.09 | 98.89 | 99.53 | 99.52 | 99.31 |
| 600 | 95.27 | 98.21 | 98.93 | 99.48 | 99.48 | 99.24 |
| 700 | 95.66 | 98.28 | **98.97** | 99.45 | 99.49 | 99.37 |
| 800 | 95.92 | 98.35 | 98.95 | 99.4 | 99.36 | 99.25 |
| 900 | 96.28 | 98.41 | 98.89 | 99.36 | 99.41 | 99.23 |
| 1000 | 96.51 | 98.45 | 98.86 | 99.21 | 99.28 | 99.14 |
| 1100 | 96.69 | 98.48 | 98.73 | 99.08 | 99.06 | 98.97 |
| 1200 | 96.81 | 98.51 | 98.65 | 98.98 | 98.92 | 98.62 |
| 1300 | 96.9 | 98.5 | 98.56 | 98.85 | 98.86 | 98.59 |
| 1400 | **96.99** | 98.51 | 98.49 | 98.71 | 98.84 | 98.21 |
| 1500 | 96.92 | **98.52** | 98.32 | 98.51 | 98.75 | 98.14 |
| 1600 | 96.89 | 98.37 | 98.17 | 98.32 | 98.61 | 97.79 |
| 1700 | 96.84 | 98.34 | 98.03 | 98.03 | 98.47 | 97.52 |
| 1800 | 96.81 | 98.21 | 97.91 | 98.06 | 98.11 | 97.21 |
| 1900 | 96.73 | 98.07 | 97.56 | 98.04 | 97.84 | 96.96 |
| 2000 | 96.71 | 98.02 | 97.49 | 97.82 | 97.41 | 96.53 |

### 4.2. The parameter of DBNRN

The $\lambda$ is a critical parameter for the DBNRM to overcome the overfitting problem. In an extreme case, if $\lambda = 0$, the DBNRM will reduce to the primary DBN model with loss function as Eq. (12), thus this DBNRM will face the overfitting problem easily. Conversely, if $\lambda \to +\infty$, the weight of DBNRM will close to zero so that the weights of DBN model will be sparse, which indicate that the recognition ability is very poor. So the relationship between the recognition accuracy of DBNRM and parameter $\lambda$ ranging from 0 to 1 would be checked by ten scenarios. The result is shown in Fig. 7.

From Fig. 7, it can be seen that the recognition accuracy of proposed DBN model is dynamic with the parameter change. The accuracy increases when the parameter $\lambda$ ranges from 0 to 0.3, because the weight in the DBNRM is penalized. The accuracy comes down, when the parameter $\lambda$ is from 0.4 to 1. It indicates that the penalization of weight is too heavily, thus all the weights in the DBNRM closed to 0 so that the recognition accuracy is disappointed when $\lambda = 1$. In one word, $\lambda = 0.3$ is the best parameter of proposed DBN model found out by the simulation experiments.

### 4.3. The robust of DBNRM for imbalance data

In the real manufacturing process, the abnormal pattern data are difficult to collect comparing with normal pattern data. Thus, the sample size between normal and abnormal pattern is imbalance, which will weak the performance of DBN model. Considering such problem, the loss function is improved by adding a weighting parameter which can remove the imbalance effect in the proposed DBN model. To check the robust of DBNRM for the imbalance data, 9 scenarios are design as shown in Table 4.

In this experiment, we compare the performance of DBNRM with original DBN model. The result is shown in Fig. 8.

From Fig. 8, it can be inferred that the DBNRM outperforms than original DBN model by using imbalance data. The accuracy of proposed model is more stable in all the scenarios, however the accuracy of original DBN model is poor when the training data is imbalance significantly. Moreover, the proposed DBN model still outperforms than alternative model when the training data is balance. It indicates the proposed DBNRM is more robust after adding a weighting parameter which controls the balance of training data.

## 5. Performance comparisons

In order to demonstrate the performance of the proposed method, the injection molding process profiles in the motivating example will be monitored and diagnosed in this section.

For constructing a DBN model that can recognize the molding process profile, 2000 samples are randomly extracted from a normal and six abnormal quality spectra respectively. 1000 of them are used as training samples and the others as test samples. Among these training samples based on the phase I procedure for training a DBN model, the DBN for the normal quality spectrum and the six abnormal quality spectra should be formed simultaneously. The parameters of the BP algorithm should then be initialized, including the activation function $\theta$, leaning rate, setting error and the number of training steps, which are set as shown in Table 5.

In order to determine the optimal structure of DBN model, the

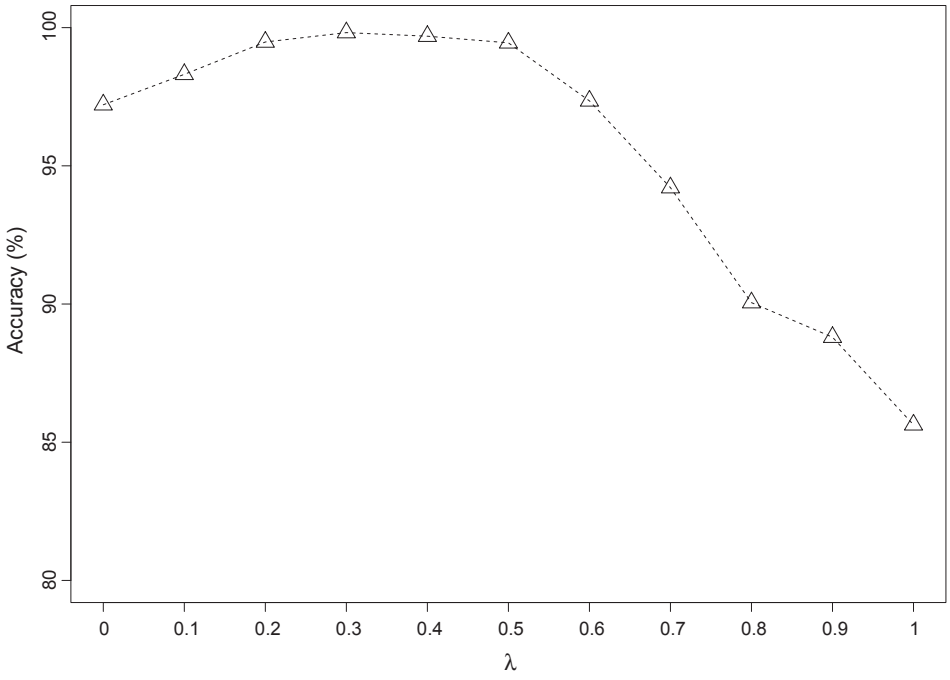


**Fig. 7.** The accuracy of DBNRM with different $\lambda$.

**Table 4**
Scenarios for imbalance data.

| Scenario | Sample size of normal pattern | Sample size of abnormal pattern one |
|---|---|---|
| 1 | 9000 | 1000 |
| 2 | 8000 | 2000 |
| 3 | 7000 | 3000 |
| 4 | 6000 | 4000 |
| 5 | 5000 | 5000 |
| 6 | 4000 | 6000 |
| 7 | 3000 | 7000 |
| 8 | 2000 | 8000 |
| 9 | 1000 | 9000 |

**Table 5**
Parameters of the BP algorithm.

| Activation function | Leaning rate | Setting error | Training steps |
|---|---|---|---|
| Sigmoid | 0.01 | 0.1 | 1000 |

similar way as simulation experiment is applied. The initialization of DBNRM with one hidden layer, and the number of nodes ranging from 100 to 2000 is considered. Then, the different structures of DBNRM would be tested layer by layer until the best structure finding out. Finally, the structure of DBN model is found out with 4000, 1500, and 400 hidden nodes respectively. The practicality of the DBNRM

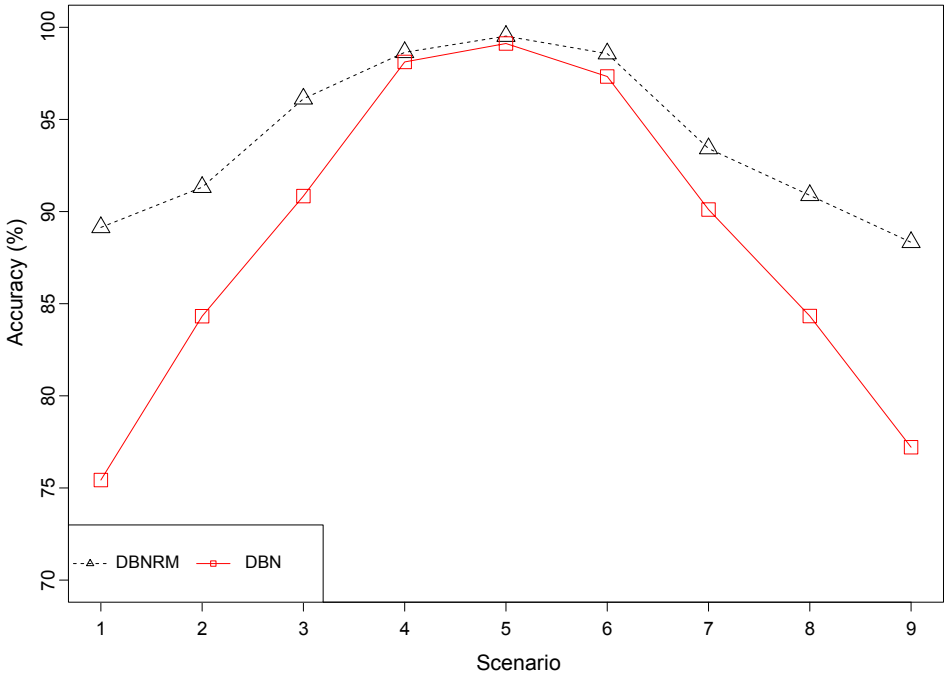


**Fig. 8.** Accuracy of DBNRM for imbalance data.

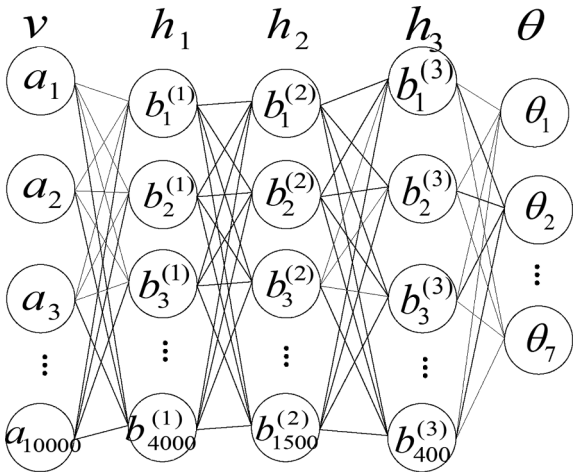


**Fig. 9.** DBNRM architecture of the injection molding process.

architecture in real application is confirmed below (see Fig. 9).

The injection molding process profiles are monitored and diagnosed on-line using the trained DBN. Because the period of injection process is 20 s, all kinds of abnormal patterns can be seen in this period as long as the defects happened. Thus the length of the moving window is taken as 20 s. The profile in the current monitoring window should be mapped into their quality spectra before monitoring. An example of real-time monitoring and diagnosis for quality spectra is shown in Fig. 10.

The change in the first two quality spectra is smooth and steady, so the model does not trigger an alarm. Once the window moves to the third spectrum, the $F_{12}$ quality spectrum is recognized by proposed DBN model, which indicates that the current injection molding process is abnormal, and flash is found. When the root causes are determined from the abnormal spectrum, we can see that the injection pressure $X_8$ and nozzle temperature $X_2$ are shifted upward simultaneously. Hence, the parameters for controlling the pressure and temperature can be directly adjusted based on the abnormal spectrum. Then the injection molding process returns to normal. After the process has been running for 120 s, the $F_{23}$ quality spectrum is recognized by proposed DBN model, which means that the current process is not normal and warpage is detected. After checking the injection molding process, we find that the temperatures of heating and cooling water both exceeded their thresholds, which led to uneven cooling of the mold. After adjusting the temperature of heating and cooling water, the process returns to normal. Thus the DBN model proposed in this paper for process profiles can be applied to real application.

To further investigate the performance of the DBN, two related methods are considered for comparison. One is the BP neural network with a single hidden layer integrated with wavelet decomposition (WD-B) and the other is the support vector machine based on wavelet decomposition (WD-S). The comparison results are shown in Table 6.

From Table 6, it can be seen that the recognition accuracy (A) of the alternative models are significantly lower than that of the DBN method. The average recognition accuracies of the WD-B model, the WD-S model, and the DBN for all quality spectra are 88.57%, 89.86%, and 96.57% respectively. Thus the DBN model proposed in this paper can recognize quality spectra more accurately than the two alternative methods.

**Table 6**
Performance comparison of proposed method with alternative methods.

| | $F_0$ | $F_{11}$ | $F_{12}$ | $F_{13}$ | $F_{21}$ | $F_{22}$ | $F_{22}$ |
|---|---|---|---|---|---|---|---|
| | A(%) | A% | A% | A% | A% | A% | A% |
| WD-B | 90 | 89 | 89 | 88 | 89 | 88 | 87 |
| WD-S | 92 | 91 | 90 | 89 | 89 | 90 | 88 |
| DBNRM | 98 | 96 | 97 | 97 | 96 | 96 | 96 |

## 6. Conclusions

A novel monitoring and diagnosis method based on DBN model for manufacturing process profiles is proposed in this paper. The proposed method not only focuses on the monitoring of some particular manufacturing processes, but also can be applied to shift detection and fault diagnosis in common manufacturing processes. The DBN algorithm can be used to monitor and diagnose process profiles without any assumption of the distribution, so it is widely applicable. The manufacturing process profiles contain rich quality information, but it is difficult to visualize the normal or abnormal status of the process in such profile data due to the monitoring variables are high-dimensional, nonlinear and highly correlated. In this paper, process profiles are mapped into their quality spectra which are digital images, so that deep belief networks can be utilized to capture and visualize information hidden in process profiles. Moreover, the proposed DBN recognition models are developed in phase I and used to monitor and diagnose the manufacturing process profiles on-line in phase II. In the simulation experiments, the way to finding out optimal structure of proposed DBN model is applied and the effect of weighting parameter is tested. The simulation shows that the performance of proposed DBN model is better than alternative DBN model, especially when the training data is imbalance. Besides, the reciprocating injection molding example shows the excellent performance of the DBN in detecting process shifts accurately. The proposed recognition model outperforms the methods of wavelet decomposition, BP neural network and support vector machine.

In our research, the proposed DBN model would be better trained by the data with labels. It indicates that the common faults patterns should be detected and represented in the dictionary list type before monitoring and diagnosis. It would be a limitation of proposed DBN model for real applications. Generally speaking, the patterns of different manufacturing processes are quite different, so it is a challenging work to define some patterns in a specific process. Moreover, the data cleaning and data labeling also are the important part of monitoring and diagnosing process profiles. It is the precondition of DBN model for monitoring the process, which should be investigated in the further research.

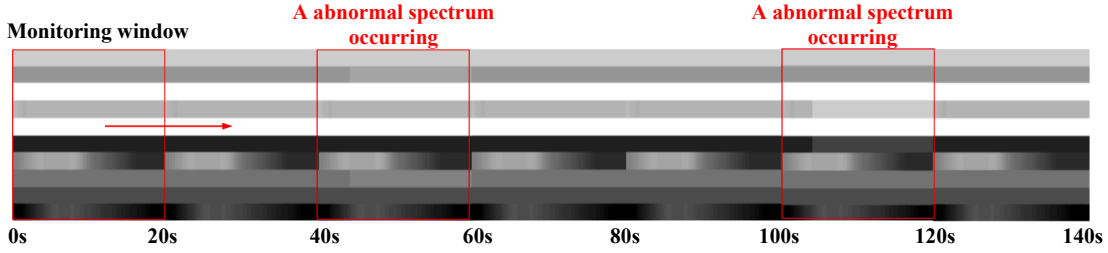


**Fig. 10.** Real-time monitoring for the injection molding process profile.

## Acknowledgement

We thank the editor and referees for their constructive comments and suggestions, which greatly improved the quality of the paper. This work was supported by National Natural Science Foundation of China Grant (No. 71672182, No. 71711540309 and No. U1604262).

## References

Amar, M., Gondal, I., & Wilson, C. (2015). Vibration spectrum imaging: A novel bearing fault classification approach. *IEEE Transactions on Industrial Electronics, 62*(99), 494–502.

Atashgar, K., Amiri, A., & Nejad, M. K. (2015). Monitoring Allan variance nonlinear profile using artificial neural network approach. *International Journal of Quality Engineering and Technology, 5*(2), 162–177.

Chen, S. H., & Nembhard, H. B. (2011). A high-dimensional control chart for profile monitoring. *Quality & Reliability Engineering International, 27*(4), 451–464.

Chou, S. H., Chang, S. I., & Tsai, T. R. (2014). On monitoring of multiple non-linear profiles. *International Journal of Production Research, 52*(11), 3209–3224.

Colosimo, B. M., Semeraro, Q., & Pacella, M. (2008). Statistical process control for geometric specifications: On the monitoring of roundness profiles. *Journal of Quality Technology, 40*(1), 1–18.

Fan, J., Han, F., & Liu, H. (2014). Challenges of big data analysis. *National Science Review, 1*(2), 293–314.

Fan, S. K. S., Jen, C. H., & Lee, T. Y. (2016). Modeling and monitoring the nonlinear profile of heat treatment process data by using an approach based on a hyperbolic tangent function. *Quality Engineering, 29*(2), 1–19.

Ge, M., Xu, Y. S., & Du, R. X. (2008). An intelligent online monitoring and diagnostic system for manufacturing automation. *IEEE Transactions on Automation Science and Engineering, 5*(1), 127–139.

Grasso, M., Menafoglio, A., Colosimo, B. M., & Secchi, P. (2016). Using curve-registration information for profile monitoring. *Journal of Quality Technology, 48*(2), 512–523.

Hawkins, D. M., & Qiu, P. (2003). The changepoint model for statistical process control. *Journal of Quality Technology, 35*(4), 355–366.

He, S. G., Jiang, W., & Deng, H. T. (2016). A distance-based control chart for monitoring multivariate process using support vector machines. *Annal of Operations Research, 2018*(263), 191–207.

He, X. H., Wang, D., Li, Y. F., & Zhou, C. H. (2016). A novel bearing fault diagnosis method based on gaussian restricted boltzmann machine. *Mathematical Problems in Engineering, 2016*(3), 1–8.

Hinton, G., Osindero, S., Welling, M., & Teh, Y. W. (2006). Unsupervised discovery of nonlinear structure using contrastive backpropagation. *Cognitive Science, 30*(4), 725–731.

Hosseinifard, S. Z., Abdollahian, M., & Zeephongsekul, P. (2011). Application of artificial neural networks in linear profile monitoring. *Expert Systems with Applications, 38*(5), 4920–4928.

Jensen, W. A., Grimshaw, S. D., & Espen, B. (2016). Nonlinear profile monitoring for oven-temperature data. *Journal of Quality Technology, 48*(1), 84–97.

Kang, L., & Albin, S. L. (2000). On-line monitoring when the process yields a linear profile. *Journal of Quality Technology, 32*(4), 418–426.

Kazemzadeh, R. B., Noorossana, R., & Amiri, A. (2008). Phase I monitoring of polynomial profiles. *Communications in Statistics - Theory and Methods, 37*(10), 1671–1686.

Liu, J. P., Jin, R., & Kong, Z. J. (2018). Wafer quality monitoring using spatial Dirichlet process based mixed-effect profile modeling scheme. *Journal of Manufacturing Systems, 48*, 21–32.

Liu, Y. M., & Zhou, H. F. (2015). Dynamic process abnormal detection based on multi-features hybrid with support vector machine. *Computer Integrated Manufacturing Systems, 21*(10), 2637–2643.

Liu, Y. M., & Zhou, H. F. (2016). Online intelligent monitoring for dynamic process based on wavelet reconstruction and SVM-BPNN. *Systems Engineering-Theory & Practice, 36*, 1890–1897.

Najafabadi, M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Waldl, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data, 2*(1), 1–21.

Noorossana, R., Eyyazian, M., Amiri, A., & Mahmoud, M. A. (2010). Statistical monitoring of multivariate multiple linear regression profiles in phase I with calibration application. *Quality & Reliability Engineering International, 26*(3), 291–303.

Pacella, M. (2018). Unsupervised classification of multichannel profile data using PCA: An application to an emission control system. *Computers & Industrial Engineering, 122*, 161–169.

Pacella, M., & Semeraro, Q. (2011). Monitoring roundness profiles based on an unsupervised neural network algorithm. *Computers & Industrial Engineering, 60*(4), 677–689.

Qin, S. J. (2014). Process data analytics in the era of big data. *AIChE Journal, 60*(9), 3092–3100.

Sun, K., Gao, J. M., Gao, Z. Y., Gao, X., & Wang, Z. (2015). Plant-wide fault pattern recognition method based on fault-spectrum. *Computer Integrated Manufacturing Systems, 21*(2), 519–527.

Wang, H., Kim, S. H., Huo, X., Hur, Y., & Wilson, J. R. (2015). Monitoring nonlinear profiles adaptively with a wavelet-based distribution-free CUSUM chart. *International Journal of Production Research, 53*(15), 4648–4667.

Woodall, W. H. (2007). Current research on profile monitoring. *Production, 17*(3), 420–425.

Woodall, W. H., Spitzner, D. J., Montgomery, D. C., & Gupta, S. (2004). Using control charts to monitor process and product quality profiles. *Journal of Quality Technology, 36*, 309–320.

Zhang, J., Ren, H., Yao, R., Zou, C., & Wang, Z. (2015). Phase I analysis of multivariate profiles based on regression adjustment. *Computers & Industrial Engineering, 85*(C), 132–144.

Zhou, C., Liu, K., Zhang, X., Zhang, W., & Shi, J. (2016). An automatic process monitoring method using recurrence plot in progressive stamping processes. *IEEE Transactions on Automation Science & Engineering, 13*(2), 1102–1111.

Zou, C., Tsung, F., & Wang, Z. (2007). Monitoring general linear profiles using multivariate EWMA schemes. *Technometrics, 49*(4), 395–408.