Proceedings of NAMRI/SME, Vol. 41, 2013

Feature Selection for Manufacturing Process Monitoring Using Cross-Validation

Chenhui Shao[a], Kamran Paynabar[b], Tae Hyung Kim[a], Jionghua (Judy) Jin[c], S. Jack Hu[a]

J. Patrick Spicer[d], Hui Wang[d], and Jeffrey A. Abell[d]

[a] *Department of Mechanical Engineering, The University of Michigan, Ann Arbor, MI, USA*
[b] *H. Milton Stewart School of Industrial and Systems Engineering, The Georgia Institute of Technology, Atlanta, Georgia, USA*
[c] *Department of Industrial and Operations Engineering, The University of Michigan, Ann Arbor, MI, USA*
[d] *Manufacturing Systems Research Lab, General Motors Technical Center, Warren, MI, USA*

ABSTRACT

A novel algorithm is developed for feature selection and parameter tuning in quality monitoring of manufacturing processes using cross-validation. Due to the recent development in sensing technology, many on-line signals are collected for manufacturing process monitoring and feature extraction is then performed to extract critical features related to product/process quality. However, lack of precise process knowledge may result in many irrelevant or redundant features. Therefore, a systematic procedure is needed to select a parsimonious set of features which provide sufficient information for process monitoring. In this study, a new method for selecting features and tuning SPC limits is proposed by applying k-fold cross-validation to simultaneously select important features and set the monitoring limits using Type I and Type II errors obtained from cross-validation. The monitoring performance for production data collected from ultrasonic metal welding of batteries demonstrates that the proposed algorithm is able to select the most efficient features and control limits and thus leading to satisfactory monitoring performance.

KEYWORDS

Feature selection, parameter tuning, cross-validation, SPC monitoring, ultrasonic metal welding.

INTRODUCTION

On-line process monitoring is crucial for product quality and process stability in manufacturing [1]. For example, in electric vehicle battery manufacturing, quality monitoring for battery joining is of great importance because any low-quality joints may result in a failure of the entire battery pack, causing high production loss. Thus, on-line process monitoring has received great attention over the past several decades.

Among various monitoring methods, the classical statistical process control (SPC) method has been widely used in monitoring manufacturing processes [2]. Control charts are the main SPC tools to determine whether a manufacturing process is in a state of statistical control. Two of the most popular types of control charts are the univariate Shewhart control chart and Hotelling T^2 control chart [3]. Exemplary applications of control charts in manufacturing process monitoring can be found in [4] and [5].

In order to monitor manufacturing processes, various sensor signals, such as force, acceleration, temperature, pressure and acoustic emission, are collected on-line to gather process information. Due to the large volume of data, feature extraction is often carried out to reduce the dimensionality of data. Efficient application-dependent features are constructed when expert knowledge about manufacturing processes is available. Whereas, if a lack of expert knowledge is encountered, some general data-driven dimensionality reduction techniques can help. Examples of such techniques include Principal Component Analysis (PCA) [6], kernel PCA [7], semidefinite embedding [8], and wavelets analysis [9].

In manufacturing, when a new process is initially implemented for production, it often occurs that a thorough physical understanding of the process is not available. For example, ultrasonic metal welding is recently utilized to join lithium-ion batteries, but there is insufficient expert knowledge about this process. Thus, signal features without good physical understanding may be irrelevant or redundant. Under this circumstance, feature selection is commonly applied to pick a minimally sized subset of features for monitoring. By removing a large number of irrelevant and redundant features, feature selection is able to help avoid overfitting, improve model performance, provide more efficient and cost-effective process monitoring, and acquire better insights into the underlying processes that generated the data.

Generally speaking, feature selection techniques can be divided into three categories in terms of means of combining feature subset selection search with the classification model construction: filter methods, wrapper methods and embedded methods [10]. Filter techniques determine the relevance of features by looking only at the intrinsic properties of the data. In wrapper methods, the model hypothesis search is embedded within the feature subset search. Embedded techniques build the feature subset search into the classifier construction. A summary of the advantages and disadvantages of each type of method and some examples of these methods can be found in [10].

In this study, a new feature selection algorithm based on cross-validation is developed for quality monitoring of manufacturing processes. The method belongs to the category of wrapper methods. Cross-validation is a common statistical technique for evaluating how the results of a statistical analysis will generalize to an independent data set [11]. It is mainly used to evaluate how accurately a predictive model will perform independent of the training dataset. In this paper, cross-validation is applied to selecting significant features and setting monitoring limits simultaneously, based on Type I (α) and Type II (β) error rates calculated from validation tests.

The rest of this paper is organized as follows. We start by presenting the details of the proposed feature selection algorithm. Then the proposed scheme is utilized for feature selection and control limits tuning for monitoring of ultrasonic metal welding in battery assembly processes. Finally conclusions are presented.

FEATURE SELECTION AND PARAMETER TUNING BASED ON CROSS-VALIDATION

In the proposed feature selection and parameter tuning algorithm, we adopt the stepwise forward feature selection to select the optimal feature subset from candidate features. Forward selection is a greedy search strategy and is particularly computationally advantageous and robust against overfitting [12]. In some cases, this search strategy may alleviate the problem of overfitting, as illustrated in [13]. Forward selection was first utilized in [14] for measurement/feature selection to determine the best subset of measurements/features for pattern classification, and it is still widely used as a feature selection scheme [15]. Forward feature selection starts the search with an empty feature subset. First, all the features are considered for possible selection, and the one feature that performs the classification the best is included in a subset. Then a new step is started, and the remaining features are considered for inclusion. This is repeated until a prespecified number of features have been included in the subset. Usually the search is repeated until all features are included for comparison purpose.

Cross-validation is a statistical technique for evaluating and comparing learning algorithms by partitioning data into two sets: one used for model training and the other used for model validation. This method is applicable for the performance comparison of different predictive modeling procedures [16], as well as for variable selection [17].

In this study, the k-fold cross-validation is employed for simultaneous feature selection and SPC parameter tuning. The original sample is randomly partitioned into k mutually exclusive subsamples/folds of equal (or approximately equal) size. Then k iterations of training and validation are performed such that within each iteration one different subsample is held-out for validation while the remaining $k-1$ subsamples are used for training. After the k iterations are finished, the k results can be averaged (or otherwise combined) to give a single estimation. In this method, all observations are used for both training and validation, and each observation is used for validation exactly once. In practice, 10-fold cross-validation is widely used.

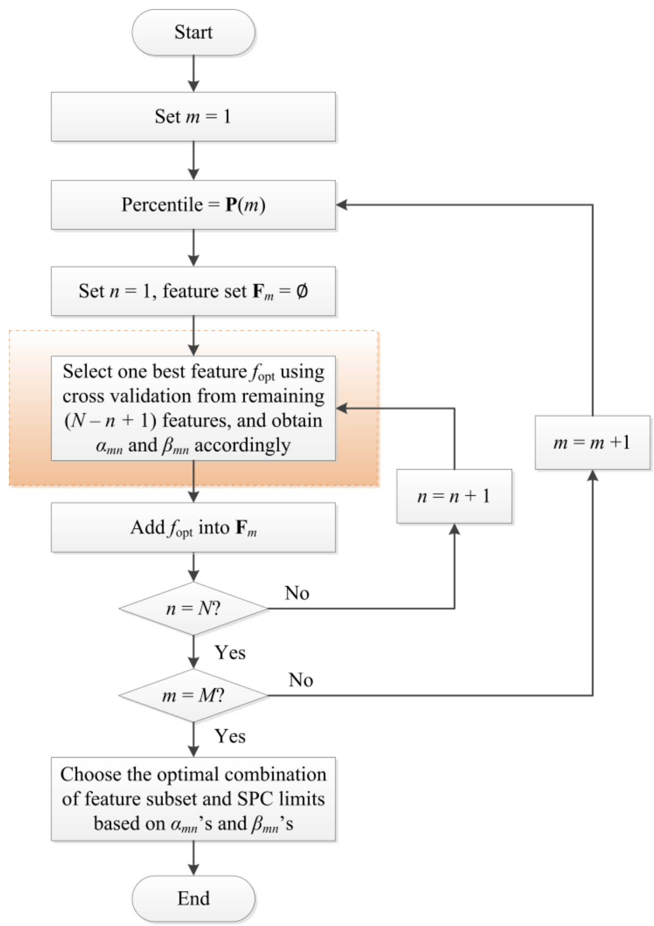

Figure 1. Feature selection and SPC limits tuning.

In the algorithm, candidate features are denoted by f_1, f_2, … , f_N, and the total number of features is N. The percentile limits are used as control limits. It is assumed that the total number of candidate percentile limit sets is M, and the mth set is denoted by $\mathbf{P}(m)$, where m = 1, 2, … , M.

Each percentile limit set includes a lower limit and an upper limit, namely,

$$\mathbf{P}(m) = \left[p_{ml} \quad p_{mu} \right], \tag{1}$$

where p_{ml} and p_{mu} are lower and upper percentile limits, respectively.

Figure 1 shows the proposed algorithm for forward feature selection and SPC limits tuning, and each forward feature selection step is performed using cross-validation. Figure 2 illustrates how to use cross-validation to select the nth feature from remaining $N - n + 1$ features in forward feature selection for the mth percentile limit set.

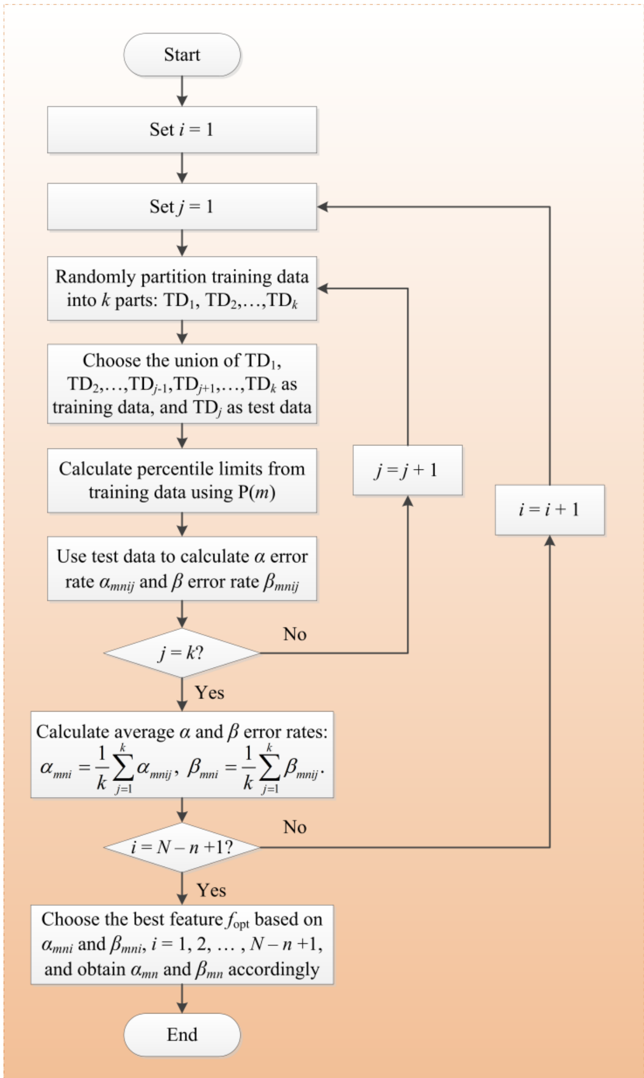

Figure 2. Forward feature selection based on cross-validation.

For each set of percentile limits $\mathbf{P}(m)$, we perform forward feature selection using cross-validation, as illustrated in Figure 2. The forward selection criterion is given by

$$\min\ R_{mn} = A\alpha_{mn} + B\beta_{mn}, \tag{2}$$

where m = 1, 2, … , M; n = 1, 2, … , N; A and B are penalty coefficients for α error rate and β error rate, and they can be tuned according to different monitoring schemes. For example, if β error rate is of higher concern, and then B can be set higher correspondingly.

For each limit set, an arrangement of candidate features is obtained, as given by Eq. (3).

$$\mathbf{F}(m) = \left[\ f_{m(1)},\ f_{m(2)}\ ,\ \cdots\ ,\ f_{m(N)} \right]. \tag{3}$$

Meanwhile, corresponding α error rates as well as β error rates are also calculated stepwise, and we record them in vectors, as shown by Eqs. (4) and (5).

$$\left[\alpha_{m1},\ \alpha_{m2},\ \cdots\ ,\ \alpha_{mN}\right]. \tag{4}$$

$$\left[\beta_{m1},\ \beta_{m2},\ \cdots\ ,\ \beta_{mN}\right]. \tag{5}$$

After performing forward feature selection for all percentile limit sets, we can select from 1 to N features for each set, and therefore there are N available choices per set. Since we have M candidate percentile limit sets, hence there are in total MN combinations of feature subset and SPC limits. Based on α_{mn}'s and β_{mn}'s, the optimal combination of feature subset and SPC limits is then selected for monitoring.

APPLICATION

In this section, the proposed algorithm is applied to select features and tune SPC limits for quality monitoring of ultrasonic metal welding of batteries. Ultrasonic metal welding process is introduced first. Then 81 candidate features are extracted. Fisher's discriminant ratio is applied for feature screening. Finally training and test results are presented.

Ultrasonic Metal Welding

Ultrasonic metal welding is a solid-state bonding process which uses a high frequency ultrasonic vibration energy to generate oscillating shears between metal sheets clamped under pressure. A typical ultrasonic metal welding system is shown by Figure 3. The advantages of using ultrasonic welding for joining dissimilar and conductive materials are discussed in [18]. Recently, ultrasonic metal welding has been adopted for battery tab joining in vehicle battery pack manufacturing. In battery pack manufacturing, creating reliable joints between battery tabs is critical because one single low-quality connection may cause performance degradation or failure of an entire battery pack. Therefore, it is important to evaluate the quality of all joints prior to connecting the modules and assembling the battery pack.

Initially, 100% manual inspection was employed for offline quality check at the beginning of production launch.

This is undesirable for several reasons. First, this checking method is expensive when production rate is high. Second, manual inspection is unreliable since the quality decisions are made largely based on inspectors' feelings, and tired inspectors can easily make mistakes. Finally, the offline inspection is not able to report process changes in a timely manner. When a significant process change occurs, it may take a long time before the battery assembly plant's attention is drawn, and during this period maybe more defective modules may be produced, which will increase the production cost significantly.

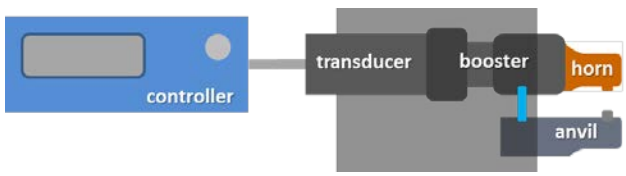

Figure 3. A typical ultrasonic metal welding system.

In order to ensure high quality joints and reduce production cost, an on-line quality monitoring system is necessary. In this research, two sensors, i.e., watt meter and microphone, are used to collect on-line process information. Due to the short duration of welding processes, usually shorter than 0.8s, a high sampling rate, such as 100kHz, is needed, and therefore high density data are collected during welding. For on-line monitoring, quality decisions are required to be made within a very short time, which is challenging in the presence of high volume data. Thus feature extraction and feature subset selection are of great significance for computational efficiency.

In this case study, due to the high cost brought by misdetection, the monitoring of battery joining processes requires a near zero β error rate.

Feature Extraction

Watt meter and microphone signals are employed for process monitoring of ultrasonic metal welding. Figure 4 and Figure 5 show typical signals from these two sensors. In addition, several process data such as the total weld time, total energy, maximum power, tool displacement before vibration, and tool displacement after vibration, are recorded through the welding system without external sensors. These data actually indicate the process conditions and thus are also included in the candidate feature set. Table 1 gives indices and names of all candidate features.

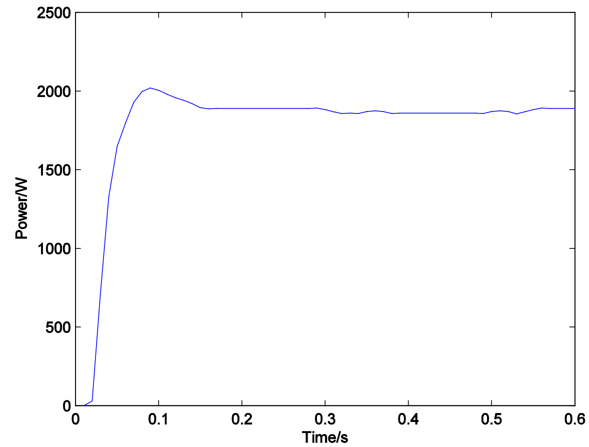

Figure 4. A typical power signal and corresponding features.

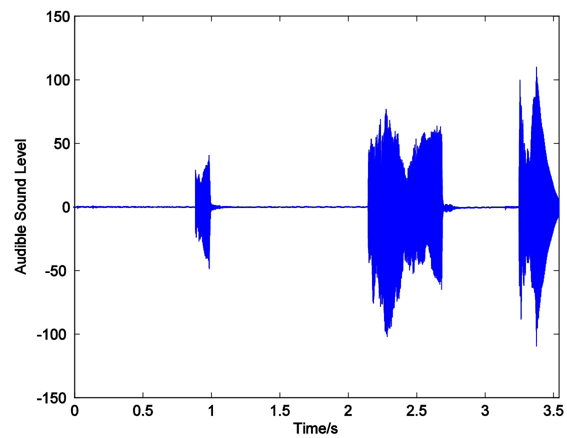

Figure 5. A typical microphone signal.

Table 1. Features and their indices.

Index	1	2	3	4	5
Feature	E	H1	H2	PP	T
Index	6	7	8	9	10
Feature	H	H*T/2	riP	riDur	riSurge
Index	11	12	13	14	15
Feature	riSlope	riE	dDur	dDepth	dSlope
Index	16	17	18	19	20
Feature	dE	raDur	raPar1	raPar2	raRSE
Index	21	22	23	24	25
Feature	raE	P_20L	P_20C	P_20R	P_40L
Index	26	27	28	29	30
Feature	P_40C	P_40R	P_F20	P_F40	P_E

Index	31	32	33	34	35
Feature	W_20L	W_20C	W_20R	W_40L	W_40C
Index	36	37	38	39	40
Feature	W_40R	W_F20	W_F40	W_E	dP
Index	41	42	…	81	
Feature	P(1)	P(2)	…	P(41)	

Feature Screening Based on Fisher's Discriminant Ratio

With the limited engineering knowledge about monitoring signals used in the ultrasonic welding operation, some previously defined features may contain little information about welding quality, so it is needed to carry out feature screening prior to feature selection using cross-validation in order to reduce the extensive computations required in the next step of feature selection.

In this case study, Fisher's discriminant ratio is applied to perform initial feature screening in a computationally simple and fast manner. Fisher's discriminant ratio was first introduced in [19], and it provides a separability measure for feature selection [20]. A larger ratio indicates more significant difference between two classes, and thus a better feature.

The Fisher's discriminant ratio for a feature is defined as

$$J = \frac{|\mu_1 - \mu_2|^2}{s_1^2 + s_2^2}, \tag{6}$$

where μ_1 and μ_2 are means of two classes, i.e., good welds and bad welds, s_1^2 and s_2^2 represent variances of two classes.

Fisher's discriminant ratios are calculated for all features, and then the ratios are ranked from largest to smallest. We select 40 features with ratios ranking from 1 to 40 as a pool of feature candidates. Table 2 lists features selected by Fisher's discriminant ratio and corresponding ratios in descending order.

Table 2. Features selected by Fisher's discriminant ratio.

Index	2	6	7	3	30	43	29	44
Ratio	8.83	7.14	5.93	2.61	2.28	1.23	1.19	1.17
Index	10	27	31	47	45	46	25	48
Ratio	1.16	1.14	1.07	1.06	1.06	1.05	1.04	1.04
Index	24	5	49	28	50	37	80	79
Ratio	1.01	0.94	0.94	0.84	0.82	0.79	0.75	0.74
Index	76	78	38	75	77	81	74	51
Ratio	0.74	0.74	0.73	0.73	0.73	0.73	0.70	0.69
Index	73	71	72	34	70	22	69	52
Ratio	0.68	0.66	0.66	0.66	0.64	0.62	0.62	0.58

Training Results

The training data set is collected from production in a battery assembly plant. The data size is 4500 with 4445 good welds (98.78%) and 55 bad welds (1.22%).

This data set is highly imbalanced in class size, i.e., the bad weld class is only 1.22% of all welds. In addition, β error rate is of higher concern, and near zero β error rate is desirable. Hence, we modify the partition method such that a case where no bad welds are included in training data set can be avoided when cross-validation is performed. In this case study, 10-fold cross-validation is applied. We partition the good weld data from training set into 10 folds. While 9 folds are used for learning, the union of the remaining fold and all bad weld data is used as validation data.

Candidate features approximately follow normal distributions, and therefore, symmetric control limits are used, namely in Eq. (1), we have $p_{mu} + p_{ml} = 1$, and Eq. (1) then becomes

$$\mathbf{P}(m) = [p_m \quad 1 - p_m], \tag{7}$$

where p_m is 0.025, 0.026, … , 0.05.

In addition, in order to ensure a near zero β error rate, we set A = 0, and B = 1 in Eq. (2) for forward feature selection, and it becomes

$$\min\ R_{mn} = \beta_{mn}. \tag{8}$$

Consequently in this case, the performance criterion used in forward selection is β error rate. In each selection step, the feature which can reduce β error rate the most is added into the current feature set. For features with the same β error rate, the feature resulting in the lowest α error rate is selected.

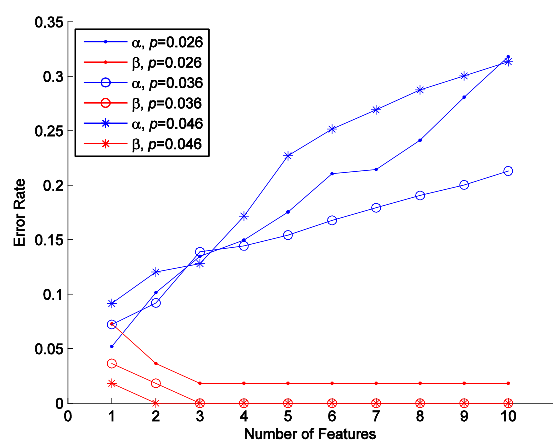

Figure 6. Illustration of training results.

The training results are illustrated by Figure 6. When a small percentile, e.g., p_m = 0.026, is applied, a zero β error is not achievable no matter how many features are used for monitoring. When a medium percentile, such as p_m = 0.036, is used, zero β errors can be achieved with relatively more features, and for p_m = 0.036, three features are needed. When a large percentile, e.g., p_m = 0.046 is applied, all bad welds can be detected with a small number of features, and for p_m = 0.046, two features are sufficient to achieve zero β errors.

Among all combinations of feature subsets and percentiles which can achieve zero β errors, one with lowest α error rate, namely, H*T/2 (Feature 7) and H1 (Feature 2) with percentiles 0.046 and 0.944, is selected for monitoring. The control limits are shown in Table 3, and the corresponding α error rate calculated by cross-validation is 12.04%.

Table 3. Summary of training results.

Feature	Percentiles	Limits	α	β
H*T/2	0.046 & 0.954	-0.0191 & 0.0215	12.04%	0
H1	0.046 & 0.954	-0.0390 & 0.0470		

Test Results

A total number of 500 welds from plant production are used for monitoring performance evaluation. In the test data, there are 497 good welds (99.4%) and 3 bad welds (0.6%).

A performance comparison between training and test data is given in Table 4. It is shown that a zero β error rate is achieved with an α error rate of 10.68% for test data. Also, the α error rate for production monitoring is comparable to that obtained from cross-validation, so the cross-validation is able to give a good estimation of α and β error rates.

Table 4. Performance comparison between training and test.

	α Error Rate	β Error Rate
Training	12.04%	0
Test	10.68%	0

Based on the results presented in this section, it can be concluded that our feature selection algorithm combining Fisher's discriminant ratio and forward selection with k-fold cross-validation is effective in selecting most appropriate features and SPC limits.

CONCLUSION

In this study, a new feature selection and control limit tuning algorithm is developed based on cross-validation for manufacturing processes monitoring. With this algorithm, the best feature subset and SPC limits can be automatically determined simultaneously. A real-world application to on-line monitoring of ultrasonic metal welding demonstrates the effectiveness of the proposed method.

The proposed algorithm is advantageous in the following several aspects. Firstly, a new algorithm is developed for feature selection and SPC limits tuning based on α and β error rates obtained via cross-validation. Therefore, the selected optimal features and their corresponding control limits as well as the predicted monitoring performance are all less sensitive to the training dataset. Secondly, this method does not require a probability distribution assumption on the candidate features, thus it is applicable to non-normally distributed measurements. Finally, this algorithm can be easily incorporated with other control charts, such as multivariate Hotelling T^2 control chart.

In addition, it should be pointed out that the proposed algorithm may encounter the computational challenge when the number of candidate features or candidate percentile limits is too large, since the feature selection is done by calculating α and β error rates for every possible combination. Thus, a more computationally efficient search strategy is needed to improve the algorithm efficiency, and our ongoing research has been focused on this topic.

REFERENCES

[1] Du, R. Elbestawi, M. A., and Wu, S. M., 1995. Automated monitoring of manufacturing processes, part 1: monitoring methods. Journal of Engineering for Industry, 117/2: 121-132.

[2] Montgomery, D. C., 2009, Introduction to Statistical Quality Control, 6th ed. Wiley, New York.

[3] Hotelling, H., 1947. Multivariate quality control illustrated by air testing of sample bombsights. In C. Eisenhart, MW Hastay, and WA Wallis (Eds.), Techniques of Statistical Analysis: 111-184.

[4] Wu, Z., and Shamsuzzaman, M., 2005. Design and application of integrated control charts for monitoring process mean and variance. Journal of Manufacturing Systems, 24/4: 302-314.

[5] Suriano, S., Wang, H., and Hu, S. J., 2012. Sequential monitoring of surface spatial variation in automotive machining processes based on high definition metrology. Journal of Manufacturing Systems, 31/1: 8-14.

[6] Hu, S. J., and Wu, S. M., 1992. Identifying sources of variation in automobile body assembly using principal component analysis. Transactions of NAMRI/SME, 20: 311-316.

[7] Ge, Z., Yang, C., and Song, Z., 2009. Improved kernel PCA-based monitoring approach for nonlinear processes. Chemical Engineering Science, 64/9: 2245-2255.

[8] Ren, C. X., Dai, D. Q., and Yan, H., 2012. Coupled Kernel Embedding for Low-Resolution Face Image Recognition. IEEE Transactions on Image Processing, 21/8: 3770-3783.
[9] Guo, H., Paynabar, K., and Jin, J., 2012. Multiscale monitoring of autocorrelated processes using wavelets analysis. IIE Transactions, 44/4: 312-326.
[10] Blum, A. L., and Langley, P., 1997. Selection of relevant features and examples in machine learning. Artificial intelligence, 97/1: 245-271.
[11] Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. Proceedings of the International Joint Conference on Artificial Intelligence, 14: 1137-1143.
[12] Guyon, I., and André E., 2003. An introduction to variable and feature selection. The Journal of Machine Learning Research, 3: 1157-1182.
[13] Reunanen, J., 2003. Overfitting in making comparisons between variable selection methods. The Journal of Machine Learning Research, 3, 1371-1382.
[14] Whitney, A. W., 1971. A direct method of nonparametric measurement selection. IEEE Transactions on Computers, C-20/9: 1100-1103.
[15] Liao, T. W., 2010. Feature extraction and selection from acoustic emission signals with an application in grinding wheel condition monitoring. Engineering Applications of Artificial Intelligence, 23/1: 74-84.
[16] Feng, C. X., Yu, Z., Kingi, U., and Baig, M. P., 2005. Threefold vs. fivefold cross validation in one-hidden-layer and two-hidden-layer predictive neural network modeling of machining surface roughness data. Journal of Manufacturing Systems, 24/2: 93-107.
[17] Picard, R., and Cook, D., 1984. Cross-validation of regression models. Journal of the American Statistical Association, 79/387: 575-583.
[18] Kim, T. H., Yum, J., Hu, S. J., Spicer, J. P., and Abell, J. A., 2011. Process robustness of single lap ultrasonic welding of thin, dissimilar materials. CIRP Annals - Manufacturing Technology, 60/1: 17-20.
[19] Fisher, R. A., 1936. The use of multiple measurements in taxonomic problems. Annals of Human Genetics, 7/2: 179-188.
[20] Duda, R. O., Hart, P. E., and Stork, D. G., 2001. Pattern Classification, 2nd ed. Wiley, New York.