Online monitoring and fault identification of mean shifts in bivariate processes using decision tree learning techniques

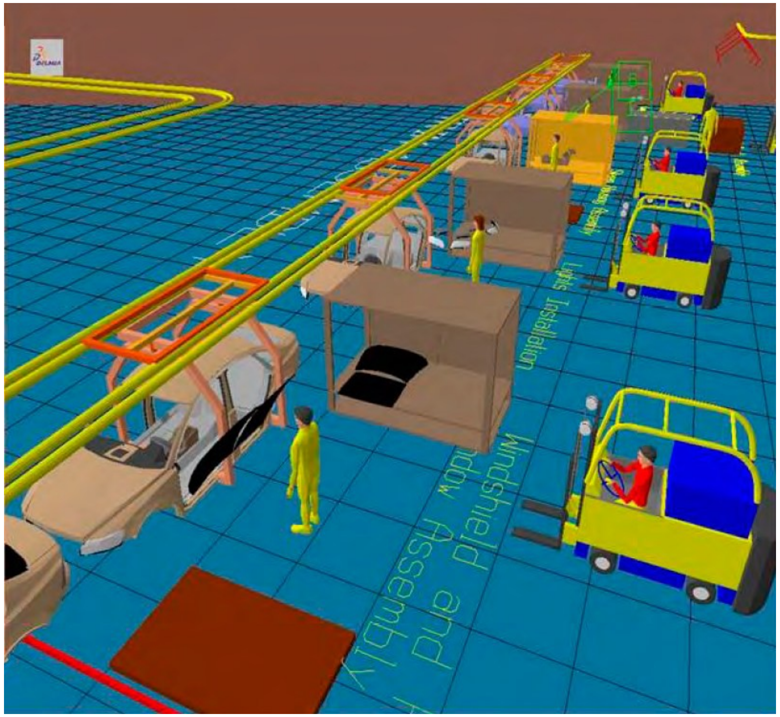
TECHNISCHE UNIVERSITÄT DARMSTADT

Overview

- Introduction
- Modules overview
- Data pre-processing
- Assumptions
- Evaluation
- Comparison of results
- Conclusions & further research

Motivation

- On-line process monitoring in manufacturing processes
- Fault identification in manufacturing processes
- Many correlated process variables simultaneously monitored

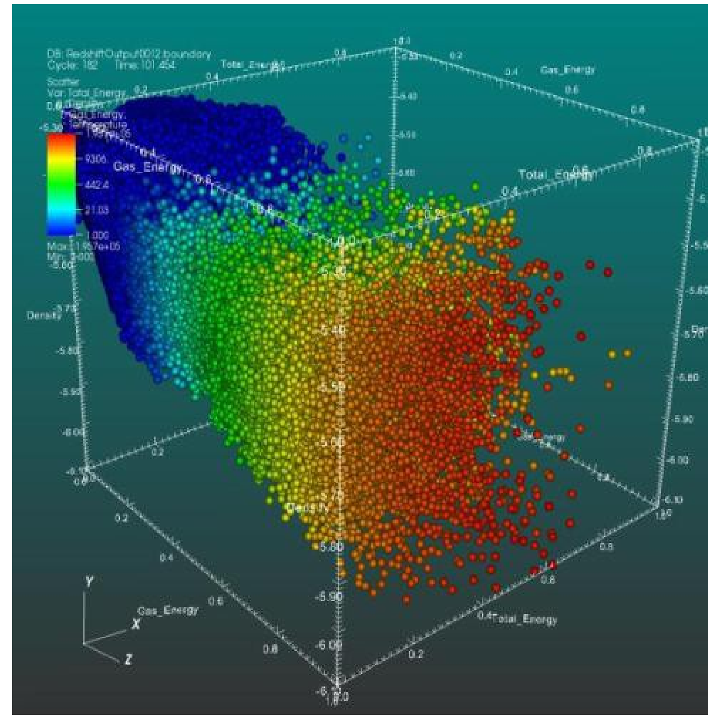


Motivation

TECHNISCHE
UNIVERSITÄT
DARMSTADT

- No direct information from multivariate control charts to which variable or subset of variables caused the out-of-control signal

- Bivariate processes can provide this information



Introduction

- Monitoring vectors $X = [x_1, x_2, . . . , x_p]$
- Determine whether there are shifts in mean vector or variance-covariance matrix
- Many possible control charts can be used

T^2 statistics

- The most widely used

- Manufacturing process has *p* correlated variables:
$X = (X_1 X_2 \ldots X_p)$

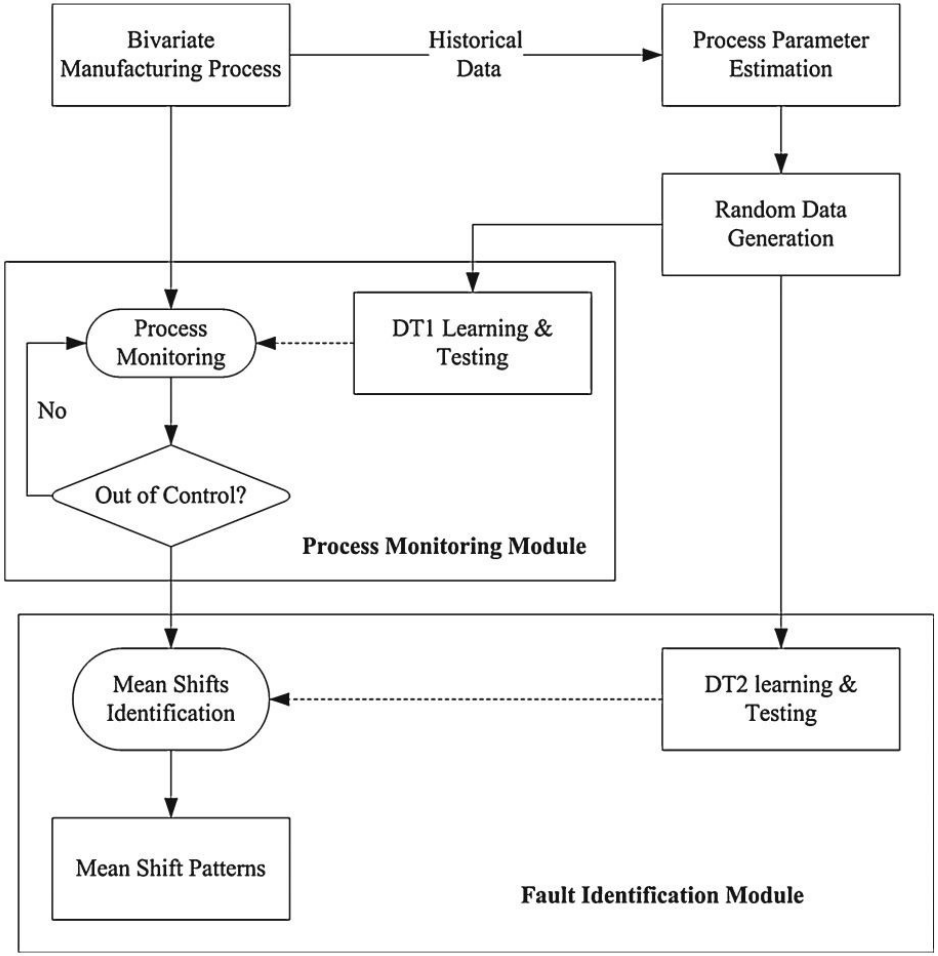
- *N* samples obtained with sample size *m*

$$\mathbf{X} \sim \mathrm{N}_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$$

- $$T_i^2 = m(\bar{\mathbf{X}}_i - \boldsymbol{\mu}_0)\Sigma_0^{-1}(\bar{\mathbf{X}}_i - \boldsymbol{\mu}_0)^T, \quad i = 1, 2, \ldots, N$$

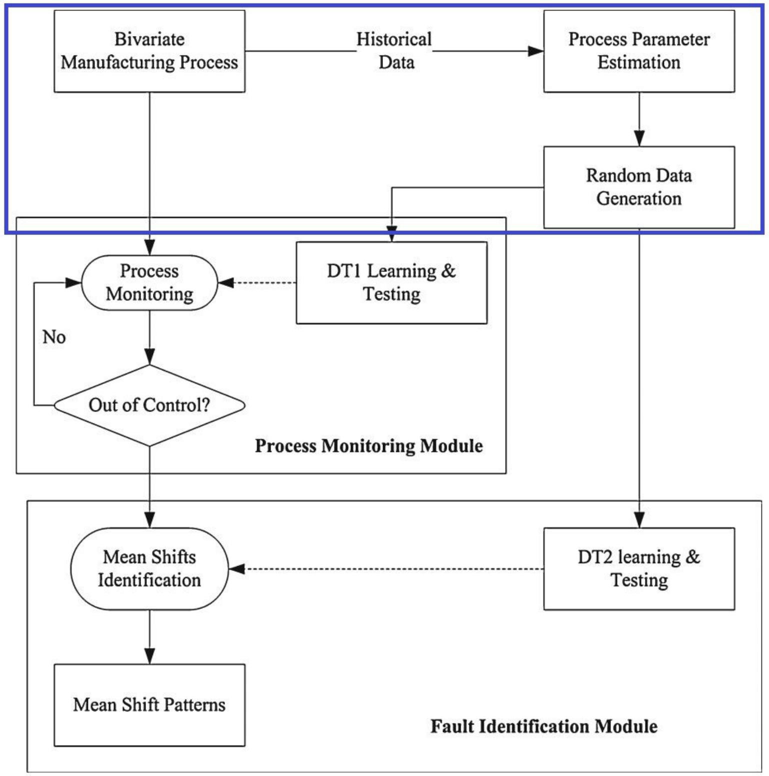
Modules

- Random data generation
- Process monitoring module
- Fault identification module



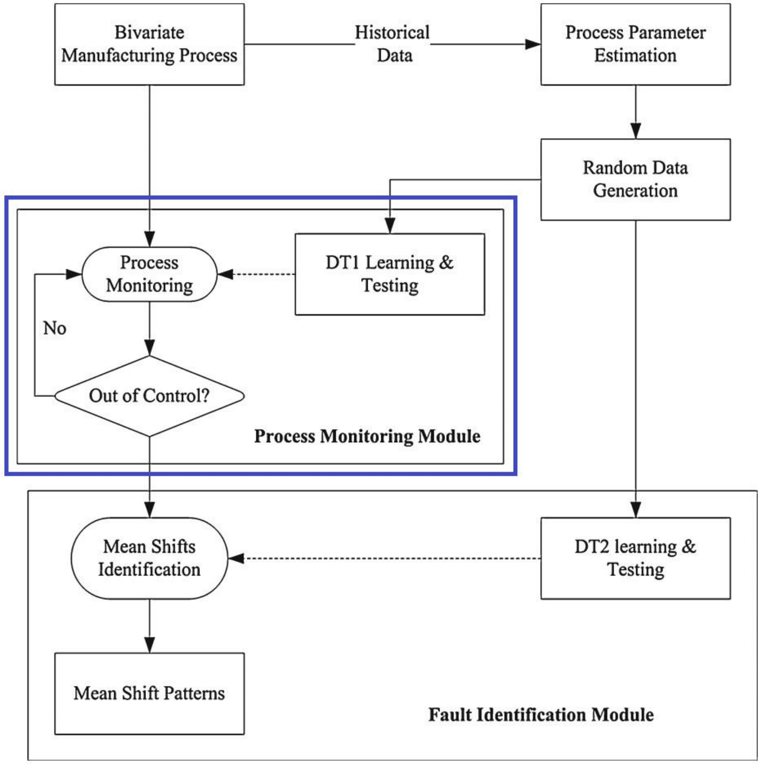
Random data generation

- Required: data with specified mean shift patterns and shift magnitudes
- Data collected from a manufacturing process don't cover it
- Generate random dataset (under the assumption of a bivariate normal distribution)



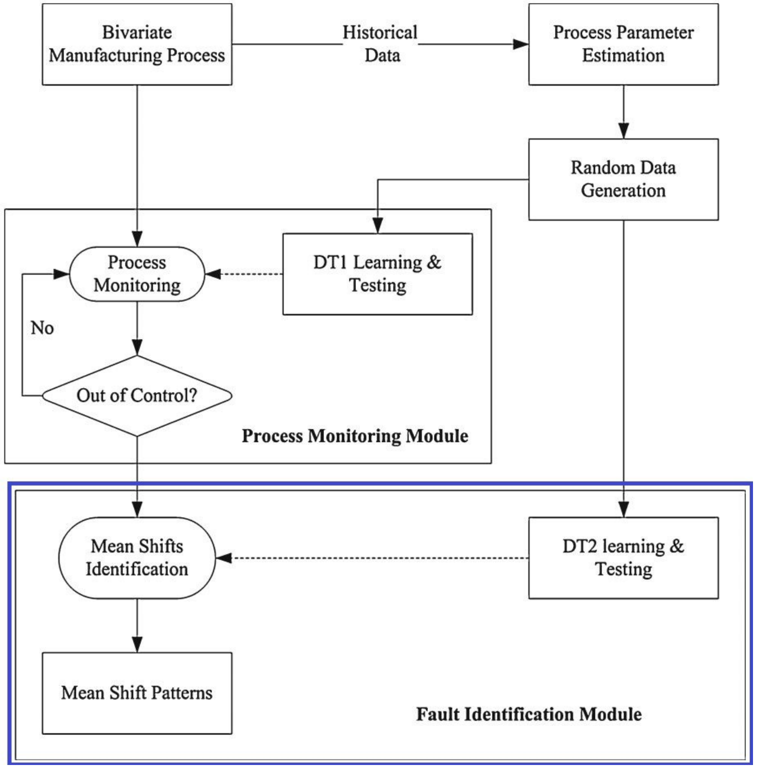
Process monitoring module

- Detects mean shifts in a manufacturing process
- DT1 to differentiate out-of-control data from in-control data
- In-control instances have a class label 0
- Out-of-control instances are labeled with 1
- The trained DT1 classifier is used to monitor the process



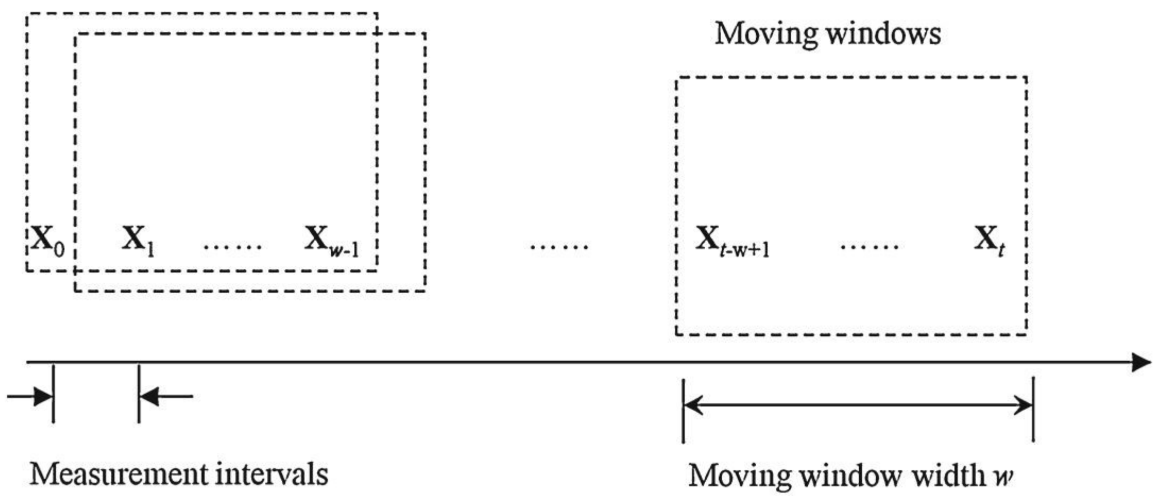
Fault identification module

- Identifies the causes of out-of-control instances
- DT2 classifier is trained with generated out-of-control instances
- The model is used for classifying out-of-control instances into different mean shift patterns



Moving window approach

- When a new observation is valid, it is combined with the foregoing w − 1 vectors
- Make a sample with sample size m(m = w)
- N samples $X_{wi} = [x_{ij1}\ x_{ij2}]$ i = 1,2, . . . , N, j = 1,2, . . . , m



Data pre-processing approach

- If the current time is *t* we get a sample with sample size *m,*

$X_t = [x_{ij1}\ x_{ij2}]$ i = t – w + 1, t – w + 2, ..., t; j = 1, 2, ..., w

- Sample mean vector: $\bar{\mathbf{X}}_{wi} = [\bar{x}_{i1}\ \bar{x}_{i2}] = \frac{1}{m}\sum_{j=1}^{m}\mathbf{X}_{wi}$
- The Mahalanobis distance:

$$d_i = (\bar{\mathbf{X}}_{wi} - \boldsymbol{\mu}_0)\boldsymbol{\Sigma}_0^{-1}(\bar{\mathbf{X}}_{wi} - \boldsymbol{\mu}_0)^T$$

- A vector V_t is made:

$$\mathbf{V}_i = [\bar{x}_{i1}\ \bar{x}_{i2}\ d_i], \quad i = 1, 2, \ldots, N$$

Data pre-processing approach

- The vector V_t is imported into DT1 to determine whether there are shifts in the process

- If the output of DT1 is 1 (an out-of-control signal),
  - the vector V_t is continuously imported into DT2 to classify it into a specific class as the result of fault identification

Assumptions

TECHNISCHE UNIVERSITÄT DARMSTADT

- (1) The process mean vector $\boldsymbol{\mu}_0$ and variance-covariance matrix $\boldsymbol{\Sigma}_0$ are all known when the process is in control

- (2) Only mean shifts are considered in this work for simplifying

- (3) Considered are only abrupt shifts where quality variables before and after a shift can all be modeled reasonably as independently and identically distributed variables

Samples generating

- The DT learning and testing samples are generated using the rules below:
  - When the process is in control, random data are generated following the distribution of $N(0,\boldsymbol{\Sigma}_0)$
  - If there is mean shift occur at time *t* then the data after *t* are generated following the distribution of $N(0+\Delta\boldsymbol{\mu},\boldsymbol{\Sigma}_0)$, where $\Delta\boldsymbol{\mu} = [k_1\ k_2]$, k_1 and k_2 are the mean shift magnitudes

Mean shift patterns coding

- We define the coding of the mean shift patterns as shown in the table

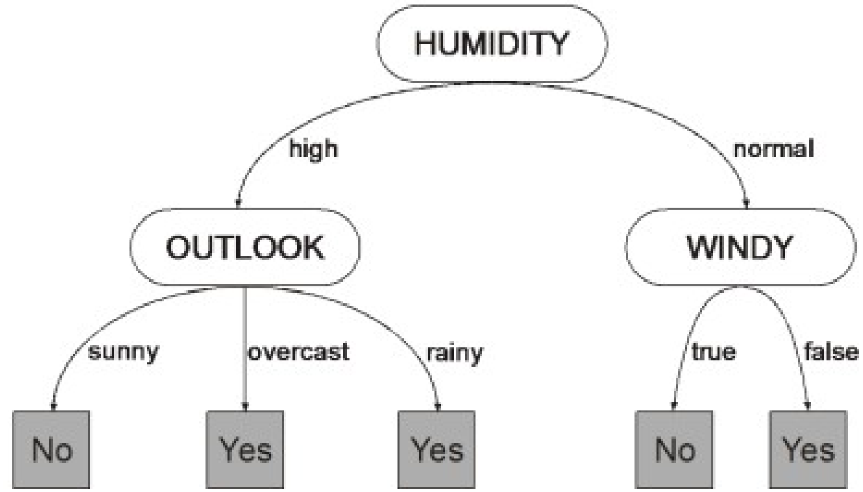
Shift	x_1	0	2	2	2	0	0	1	1	1
	x_2	0	2	0	1	2	1	2	0	1
Coding		T_0	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8

- The mean shifts are encoded as 0 (no mean shifts), 1 (downward mean shifts), or 2 (upward mean shifts)

- The coding of T0 represents that the process is in-control and the coding of T1 ~T8 represent that the process is out-of-control

DT learning algorithm

- The main advantages are its simplicity and efficiency
- It can deal with large amount of high-dimensional data with high computing efficiency
- The classification results are easy to understand and interpret
- DT are able to solve nonlinear classification problems



Evaluation measures

- The ARL (average run length) is used for evaluating the performance of the monitoring procedure

- ARL_0 is the in-control average run length: the average number of samples needed for a control chart to give an out-of-control signal when the process is in control

- ARL_1 is the out-of-control ARL: the average number of samples needed for a control chart to give an out-of-control signal when there are shifts in the process

- A good multivariate process monitoring procedure: large ARL_0 and small ARL_1

Evaluation measures

- The performance of DT1 is evaluated using the metrics ARL and Correct Ratio (CR)
- The CR is the ratio of the number of correctly classified testing samples over the total number of testing samples
- CR is applied to evaluate the performance of both DT1 and DT2

DT classifiers

TECHNISCHE UNIVERSITÄT DARMSTADT

- In this work, two DT classifiers are used
- In the learning process, the two classifiers can be trained independently
- In the model testing, DT1 is applied first. If the output of DT1 is 1, DT2 will be used subsequently
- In DT1 learning process, we define a misclassification matrix as the following to increase ARL_0

		Predicted values	
		0	1
Real values	0	0	3.0
	1	1.0	0

Numerical experiments

TECHNISCHE UNIVERSITÄT DARMSTADT

- The bivariate normal distribution with unit variances was used to generate learning and testing cases for the proposed model

- For presenting the interesting mean shift intervals, the mean shift magnitudes (k_1, k_2) for the two variables are set to take a value in ($-3.00, -2.75, -2.5, \ldots, -1.25, -1.0, 0, 1.0, 1.25, \ldots, 2.5, 2.75, 3.0$)

- For a bivariate process, there are 19*19(361) mean shift combinations including the in-control condition when μ = [0 0] and 360 mean shift combinations when the process is out-of-control

- $\mathbf{N_1}$ in-control cases and $\mathbf{N_2}$ out-of-control cases are generated. Therefore, there are $N = N_1 + 360 \times N_2$ cases generated for model training.

Numerical experiments

- Set $N_1 = 5{,}000+w-1$ and $N_2 = 100$ to generate random data for model training
- Generate the same number of samples for model testing with same mean shift patterns and shift magnitudes
- Analyze the effects of moving window width and correlation coefficients on the performance of the proposed model

The moving window width *w*

- When evaluating the performance of the proposed model, we set w to be the values in (4, 6, 10, 20) respectively and ρ to be 0.5
- ARL_0 increases with the increase of window widths. At the same time ARL_1 decreases and CR increases
- But a large w delays out-of-control signals when mean shifts occur when the moving window approach is used

Table 2 ARL and CR values of DT1 based on different *w*

w	ARL_0	ARL_1	CR(%)
4	96.10	1.0466	95.97
6	113.52	1.0237	97.86
10	332.73	1.0083	99.24
20	4,981.00	1.0034	99.70

The moving window width *w*

Table 3 CR values of DT2 based on different w

	w=4			w=6			w=10			w=20		
	L	*M*	*S*	*L*	*M*	*S*	*L*	*M*	*S*	*L*	*M*	*S*
T_1	**96.28**	**97.28**	**97.00**	**96.81**	**97.06**	**96.33**	**97.72**	**98.11**	**97.89**	**97.44**	**98.00**	**97.33**
T_2	70.67	69.33	66.33	73.67	81.33	78.33	**94.67**	**96.33**	**92.00**	**97.67**	**97.67**	**98.00**
T_3	**99.00**	**96.45**	88.11	**99.30**	**97.52**	**95.00**	**98.98**	**97.41**	**96.00**	**98.09**	**97.45**	**98.33**
T_4	61.33	60.67	49.67	79.33	76.33	71.33	84.33	72.33	78.33	**95.00**	89.67	86.00
T_5	60.00	62.00	61.33	73.67	69.33	76.33	81.33	85.00	**94.67**	**95.33**	**95.33**	**91.00**
T_6	**98.56**	**95.44**	79.00	**99.02**	**97.44**	88.89	**98.89**	**97.81**	**95.22**	**97.71**	**97.37**	**95.44**
T_7	80.00	75.67	74.33	76.33	70.00	77.00	**92.67**	89.00	**95.67**	**96.33**	**96.00**	**95.00**
T_8	**97.02**	**96.67**	**94.89**	**98.40**	**98.26**	**97.44**	**97.98**	**97.48**	**97.22**	**97.60**	**97.33**	**95.22**
Average	82.86	81.69	76.33	87.07	85.91	85.08	**93.32**	**91.69**	**93.38**	**96.90**	**96.10**	**94.54**

The CR values greater than 90% are in bold

- The CR values of DT2 also increase with the increase of w
- A larger *w* will lead to a larger sample size and more accurate estimation on the process parameters can be obtained

Effect of correlation coefificients

TECHNISCHE UNIVERSITÄT DARMSTADT

- Set ρ to be a value in (–0.9, –0.7, –0.5, –0.3, –0.1, 0.1, 0.3, 0.5, 0.7, 0.9) to analyze the effect of correlation coefficients on the performance of the proposed model

- For simplicity presented are only the results based on w = 10

- The performance of both DT1 and DT2 is analyzed

- All different corelation coefficients hold good performance of DT1

Table 4 ARL and CR values of DT1 based on different correlation coefficients

ρ	ARL_0	ARL_1	CR(%)
–0.9	453.73	1.0072	99.35
–0.7	623.88	1.0106	99.06
–0.5	623.88	1.0088	99.21
–0.3	831.83	1.0078	99.31
–0.1	184.85	1.0092	99.13
0.1	713.00	1.0115	98.98
0.3	178.25	1.0082	99.21
0.5	623.88	1.0105	99.07
0.7	184.85	1.0069	99.33
0.9	356.50	1.0065	99.40
Average	**477.46**	**1.0087**	**99.21**

The average values of ARL0, ARL1, CR are in bold

Effect of correlation coefficients

- The minimum average CR value is 88.97%
- The performance of the proposed model is acceptable

Table 5 CR values of DT2 based on different correlation coefficients

		ρ									
		−0.9	−0.7	−0.5	−0.3	−0.1	0.1	0.3	0.5	0.7	0.9
L	T_1	98.86	98.69	98.94	98.31	97.61	97.94	97.22	96.86	98.06	97.67
	T_2	95.00	88.33	91.33	91.00	89.00	88.33	89.67	91.33	95.67	92.67
	T_3	98.07	97.91	98.67	98.02	98.30	98.77	98.72	98.72	98.95	98.86
	T_4	85.00	91.33	80.67	86.00	80.00	78.67	85.67	81.00	88.33	86.33
	T_5	89.00	94.00	82.33	87.00	86.33	75.67	91.33	88.67	85.00	86.67
	T_6	97.71	98.00	97.78	98.31	98.13	97.98	98.73	98.24	98.42	98.89
	T_7	91.00	93.33	84.67	90.67	89.00	77.33	68.67	94.00	84.00	94.00
	T_8	99.00	98.58	98.42	98.56	98.27	98.73	98.33	98.56	98.16	98.24
	Average	**94.21**	**95.02**	**91.60**	**93.48**	**92.08**	**89.18**	**91.04**	**93.42**	**93.32**	**94.17**
M	T_1	98.78	98.92	98.94	97.75	98.25	97.75	98.94	97.92	97.78	99.06
	T_2	88.67	84.67	86.00	85.33	88.67	89.00	92.00	94.33	92.00	92.33
	T_3	98.14	98.07	97.90	98.10	98.41	97.59	97.93	97.07	97.62	98.41
	T_4	83.00	90.00	83.33	82.00	81.33	86.33	84.67	92.00	93.67	90.33
	T_5	88.33	92.00	81.33	89.00	82.67	81.00	87.00	84.00	92.33	93.33
	T_6	97.81	98.07	97.22	97.59	98.04	97.81	97.70	97.56	97.52	98.07
	T_7	93.33	85.00	86.67	90.00	87.67	87.33	80.00	90.00	84.00	91.33
	T_8	97.81	97.59	97.93	97.04	97.33	97.96	98.30	98.41	97.67	97.44
	Average	**93.23**	**93.04**	**91.17**	**92.10**	**91.55**	**91.85**	**92.07**	93.91	94.07	95.04
S	T_1	95.78	95.67	95.44	91.33	93.56	95.11	94.89	98.33	97.78	98.89
	T_2	94.00	91.67	82.00	93.00	87.67	88.00	89.67	84.67	92.00	94.00
	T_3	97.78	98.33	97.11	96.56	96.44	95.56	95.33	94.44	94.78	93.00
	T_4	83.67	85.67	80.67	88.00	75.33	88.00	83.00	89.67	85.67	91.67
	T_5	90.67	87.33	83.00	85.67	83.33	83.67	74.67	86.00	90.67	78.33
	T_6	97.89	97.22	95.44	93.11	95.67	94.11	95.11	94.44	92.89	93.44
	T_7	97.33	90.33	84.00	79.67	90.00	83.33	86.33	80.33	86.00	88.33
	T_8	94.00	91.67	94.11	92.89	90.11	96.56	94.89	97.44	97.00	97.67
	Average	**93.89**	**92.24**	**88.97**	**90.03**	**89.01**	**90.54**	**89.24**	**90.67**	**92.10**	**91.92**

The average CR values are in bold

Evaluation: parameter values

TECHNISCHE UNIVERSITÄT DARMSTADT

- A bivariate process based on $\rho = 0.5$ and specified mean shift magnitudes is studied
- The moving window width is set to 10
- The results of the proposed model are compared to Guh's Model

Comparison of results

- the ARL$_0$ of the proposed model is 201.10 compared to that of 192 in Guh's model

Table 8 A comparison between the proposed model and Guh's model

k_1	k_2	Proposed model		Guh's model	
		ARL	Average	ARL	Average
0	0	201.10	**201.10**	192	**192**
−0.900	−0.828	1.7286	**1.8113**	3.470	**4.14**
−0.800	−0.920	1.6168		3.743	
−0.700	−0.969	1.9455		4.020	
0.000	0.866	2.0725		5.841	
0.700	0.969	1.7652		4.406	
0.800	0.920	1.8450		3.842	
0.900	0.828	1.7050		3.677	
−1.400	−1.167	1.0272	**1.0521**	2.293	**2.61**
−1.200	−1.380	1.0515		2.230	
−1.000	−1.468	1.1044		2.430	
0.000	1.299	1.0565		4.254	
1.000	1.468	1.0293		2.550	
1.200	1.380	1.0417		2.277	
1.400	1.167	1.0543		2.253	
−2.000	−1.015	1.0055	**1.0042**	1.547	**1.80**
−1.500	−1.896	1.0060		1.250	
−1.000	1.000	1.0040		2.853	
0.000	1.732	1.0081		2.517	
1.000	2.000	1.0015		1.673	
1.500	1.896	1.0015		1.260	
2.000	1.015	1.0025		1.470	

Comparison of results

TECHNISCHE UNIVERSITÄT DARMSTADT

Table 8 continued

k_1	k_2	Proposed model		Guh's model	
		ARL	Average	ARL	Average
−2.500	−1.269	1.0005	**1.0018**	1.477	**1.17**
−1.500	−2.482	1.0020		1.327	
−1.000	1.484	1.0025		2.252	
0.000	2.165	1.0020		2.187	
1.000	2.484	1.0020		1.848	
1.500	2.482	1.0015		1.370	
2.500	1.269	1.0020		1.473	
−3.000	−1.523	1.0010	**1.0010**	1.297	**1.59**
−2.000	−2.937	1.0005		1.050	
−1.000	1.949	1.0005		2.300	
0.000	2.598	1.0010		2.073	
1.000	2.950	1.0015		1.983	
2.000	2.937	1.0010		1.140	
3.000	1.523	1.0015		1.277	

The average ARL values of the proposed model and Guh's model are in bold

- When there are mean shifts, the ARL_1 of the proposed model are all smaller than those of the Guh's model. It shows in Table 8 that the proposed model outperforms Guh's model.

Advantages of the proposed model

TECHNISCHE UNIVERSITÄT DARMSTADT

- (1) Guh's model: a single DT classifier was built for both process monitoring and fault identification
  - in our model, two DT classifiers are built respectively for process monitoring and fault identification
  - it leads to a smaller number of classes of the DT classifiers

- (2) The dimension of the input in Guh's model is $(p+1)\times w$: all data in the moving windows are selected as the inputs to the DT classifiers
  - in our model we use the mean vectors of the samples in the moving windows and the Mahalanobis distance as the inputs to the DT classifiers
  - the dimension of the inputs in our model is only $(p+1)$

Conclusions

- A bivariate process monitoring and fault identification model was built using DT learning based techniques

- Two DT classifiers were built, one for process monitoring while the other for fault identification

- Numerical experiments of the proposed model based on different correlation coefficients and different moving window widths were presented

- all the CR values for fault identification were greater than 80% and most of them were greater than 90%

Further research: two directions

- (1) Only a special circumstance of $p = 2$ was studied in this work. The cases where $p > 2$ should be studied in future to test the performance of the proposed DT learning based model

- (2) The assumption of constant variance-covariance matrix was made in this work. Although it is rational in specific situations, in some manufacturing processes the variances may change over time. How to use the proposed model in such situations is another further research topic

Notes

- The proposed model clearly outperforms the Guh's model
  - this models in real manufacturing processes
  - origin of data
- One of the pointed advantages is a smaller number of classes of the DT classifiers
  - the difference between around 370 and 360 labels should be not a crucial factor
- The differentiation of logical parts (monitoring and fault identification) comparing to Guh's model is an advantage

THANKS FOR LISTENING.
Q?

References

- He, Shu-Guang, Zhen He, and Gang A. Wang. "Online monitoring and fault identification of mean shifts in bivariate processes using decision tree learning techniques." Journal of Intelligent Manufacturing 24.1 (2013): 25-34.
- Guh, R. S. (2005). A hybrid learning-based model for on-line detection and analysis of control chart patterns. Computers and Industrial Engineering, 49(1), 35–62.
- Guh, R., & Shiue, Y. (2008). An effective application of decision tree learning for on-line detection of mean shifts in multivariate control charts. Computers and Industrial Engineering, 55(2), 475–493.
- https://en.wikipedia.org/wiki/Covariance_matrix

Source for used images

- http://upload.wikimedia.org/wikipedia/commons/c/c4/Scatter_plot.jpg
- http://www.texample.net/media/tikz/examples/PNG/scatterplot.png
- http://upload.wikimedia.org/wikipedia/commons/a/ac/NIST_Manufacturing_Systems_Integration_Program.jpg
- https://upload.wikimedia.org/wikipedia/commons/c/c0/Gaussian-2d.png
- http://upload.wikimedia.org/wikipedia/en/5/5a/Decision_tree_for_playing_outside.png