Original Research

Development of Clustering Technique and Genetic Algorithm to Monitor Multivariate Descriptive Processes based on Large-scale Nominal Contingency Tables (Case Study: Renewable Energy Process)

Yaser, Vahedi Geshniani[1]. Bijan, Rahmani[2]. Kamranrad, Reza[3*]

Received: 15 Feb 2021/ Accepted: 02 January 2023/ Published online 02 January 2023

*Corresponding Author, r.kamranrad@semnan.ac.ir

1- Department of Industrial Engineering, Central Tehran Branch, Islamic Azad University, Tehran, Iran.
2- Department of Industrial Engineering, Faculty of Technology and Engineering, Central Tehran Branch, Tehran, Iran.
3- Department of industrial engineering, faculty of engineering, Semnan university, Semnan, Iran.

Abstract

Many real-world issues are based on multivariate processes with descriptive characteristics that are represented by contingency tables. A contingency table is a tool for showing the simultaneous relationship of two or more descriptive variables that is modeled by the log-linear communication function and monitored over time. In some statistical process monitoring (SPM) applications, we are faced with the multiplicity of variables and, of course, the number of nominal classifications of the response variable. To model them, a log-linear model based on large-scale contingency tables is used that are called nominal large-scale descriptive multivariate processes. In monitoring this type of process, we face the negative impact of large dimensions of contingency tables on the performance of control charts. For this purpose, a new approach based on the clustering approach in correspondence analysis have been developed to reduce the effect of large dimensions and improvement performance of the control charts in diagnosing out of control status. It is noted that, the main contribution of this paper is to develop some approaches to monitor the large dimension based multivariate nominal processes which is not considered by previous researches. The performance of control charts has been evaluated using simulated studies and the results indicate the appropriate efficiency of the proposed approach in reducing the impact of the contingency table dimensions on the performance of the control charts. In addition, to demonstrate the performance efficiency of the proposed methods, a real case study in the field of renewable energy has been used, the results of which indicate the proper performance of the proposed control charts in diagnosing out of control status.

Keywords - Large-scale contingency table; Genetic algorithm; log-linear model; correspondence analysis; Statistical process monitoring

INTRODUCTION

Nowadays, control charts are utilized as a useful tool to monitor process parameters with different statistical distributions. One of the best opportunities for increasing productivity is the scheduling of preventive maintenance and the use of control charts (Farahani et al., 2020). One of the types of process monitoring that has been considered by researchers in recent years is the monitoring of multivariate categorical processes based on polynomial statistical distribution that first developed by Li et al. (2012) and Yashchin (2012). These processes contain at least two correlated descriptive attributes that are displayed using a tool called a contingency table. But what has been categorized by various researchers in the field of multivariate process monitoring, including Li et al. (2013), Li et al. (2014), Kamranrad et al. (2017a and 2017b) and Kamranrad et al. (2019) is to present control charts just for the purpose of statistical monitoring of these types of processes.

In this study, we seek to design control charts for monitoring multivariate descriptive processes based on large-scale contingency tables. Therefore, in this section, we will examine the research conducted in the field of descriptive multivariate process monitoring. Yashchin (2012) presented the regenerative likelihood ratio control chart to diagnose changes in the possible behavior of particles in a semiconductor component plant (contingency-based table) over specific time periods in Phase 2. To monitor, he considered the two main parameters of the contingency tables, i.e. λ (the effect of each variable) and pij (the estimated value of the observation likelihood in cell (i, j). In this research, the variable of the type of particles in the air with four levels of metal, organic, inorganic and others has been considered as the first categorical response variable. Also, the variable of the level of particles in the environment, which were divided into three levels, is the second response variable. In another study by Li et al. (2012), multivariate binomial and polynomial processes in phase-2 were investigated using the log-linear model. They presented the error diagnosis process to determine the shifts in the parameters of the model in question. They also used the Generalized Likelihood Ratio Test (GLRT) to diagnose changes in the model parameter and determine if the process was under control. They state that their proposed method for diagnosing small shifts in the model parameters is not very efficient and therefore to solve this problem, they combined the GLRT approach with the EWMA approach and presented it as EWMA-GLRT. Tsung and Zou (2013) proposed a binary spatial model for modeling multistage processes with binary responses. They also applied the monitoring and diagnosis approach using hierarchical likelihood and direct information based on the binary spatial model. Their proposed approach not only considers the correlation in and between stages, but also eliminates the consequences of the complexity of this method in monitoring and diagnosing the signal factor from one stage to another.

Li et al. (2013) presented a multivariate nonparametric control chart for shape parameter monitoring. This chart is based on a combination of a multivariate signal test chart and a EWMA chart for continuous on-line monitoring. They eventually concluded that the proposed nonparametric chart could only be used for the main observations and was not very efficient at diagnosing very large shifts, a feature common to all nonparametric charts. Li et al. (2014) also presented a multivariate binomial / polynomial control chart for monitoring multivariate processes categorical in Phase-2. In their research, they considered multivariate processes categorical with the assumption of correlation among variables and using the log-linear model, improved the relationship among existing categorical variables in the form of multivariate binomial and multinomial distributions. Kamranrad (2017a) presented a control chart based on generalized linear test (GLT) statistics for monitoring multivariate processes categorized in phase 2. They also combined the GLT statistic with the EWMA statistic to improve the performance of their proposed method in small and medium shifts, and proposed a new control chart called EWMA-GLT. In addition, they developed a signal factor parameter diagnosis method based on GLT statistics. Kamranrad et al. (2017b) offered two new control charts based on Wald and Rao score test statistics to monitor processes based on contingency tables in phase 1. To improve the performance of their charts, they combined the mentioned statistics with the EWMA statistics and compared the results with the two EWMA-GLRT and Multinomial/Binomial control charts. In addition, they provided a new method to diagnose the signal factor cell that has led to an out-of-control status.

The same authors (2019) developed F and SLRT control charts for monitoring multivariate processes based on phase 1 contingency tables and evaluated the performance of the proposed control charts with three types of step, drift and outlier changes. In their research, they also developed an SLRT-based approach to estimate the real-time change point in this type of process. In order to demonstrate the efficacy of the proposed methods in practical applications, they evaluated the performance of the methods in a real case study in the field of health care focusing on kidney patients. The results of the calculations have showed the appropriate efficiency of these methods. The research that has been studied by researchers in this field so far has been mainly on nominal multivariate descriptive processes. But there are also researchers who have worked in the field of multivariate process monitoring based on sequential contingency tables. For example, Zafar et al. (2013) used a sequential log-linear model with correspondence analysis in the pharmaceutical industry to predict the amount of drugs in medicinal diagnosis processes. Farahani et al. (2019) have presented an integrated model for optimizing statistical process control policies (sampling interval, sample size and control limit) and preventive maintenance (the preventive maintenance interval).Yamamoto and Murakami (2014) presented a model for evaluating square contingency-based tables processes with sequential variables. In this study, their hypothesis on the unbalanced normal distribution in the field of tooth decay was used. Brzezinska (2016) also proposed a model for sequential contingency tables considering linear effects, rows, columns, and concurrent effects. In addition to the above research, Wang et al. (2017) presented multivariate process monitoring based on sequential log-linear model. In their research, they provided a Generalized Likelihood Ratio Test (GLRT) control chart for monitoring these types of processes. Their proposed control chart has good efficiency for diagnosing spatial shifts and dependency shifts in hidden continuous variables with sequentially categorical features based on multilevel values. In addition, Hakimi et al. (2019) developed a new approach to monitor multivariate processes based on ordinal contingency tables in Phase 2. In their study, they presented the generalized-p control chart provided by Li et al. (2014) for monitoring sequential uni-variate processes in order to monitor multivariate sequential processes and, using simulated studies, they evaluated the efficiency of their proposed method. The results of the simulations show the proper performance of the proposed method in diagnosing various changes in the parameters of the sequential log-linear model.

Hakimi et al. (2021) proposed two control charts including simple ordinal categorical and Generalized-p to monitor the multivariate ordinal processes in Phase II. Performance of the proposed charts has been evaluated through simulation studies and a real numerical example. Results showed that, the simple ordinal categorical based control chart had better performance than the other chart under most shifts in ordinal log-linear model parameters. Hakimi et al. (2022) proposed new statistics for Phase I monitoring the multivariate categorical processes based on ordinal contingency tables. To this aim, two control charts called LRT and MR have been developed and results showed that MR control chart had better performance than LRT under small and moderate shifts in model parameters. In addition, they used a real case study in drug industry to monitor the drug dissolution process and results also confirmed the simulation outputs. In all of the above investigations, multivariate processes categorized according to small-scale contingency tables have been monitored. In this research, we intend to present a new approach to monitor nominal multivariate processes based on large-scale contingency tables. The main research question is what effect will the size of the contingency tables have on the performance of control charts in process monitoring, and how can the effects of large-scale be controlled?

In this research, first, the correspondence analysis approach is proposed to reduce the large scale of the contingency table. Then, control charts F and T^2 based on the proposed approach for monitoring processes of contingency-based tables in phase 2 are presented.The structure of the paper is as follows: In the next section, a review of multivariate categorical processes based on small-scale contingency tables is given. In the third section, the technique of reducing the size of large-scale contingency tables based on the clustering technique is presented in order to reduce the negative effects on process monitoring. In the fourth section, proposed control charts for monitoring large-scale descriptive processes in Phase 2 are presented. The performance of the proposed methods is evaluated using simulated studies in the fifth section. Also, to show the efficiency of the proposed methods, a real data set in the field of renewable energy has been used, the results of which are examined and analyzed in the sixth section. Finally, conclusion and future suggestions will be presented in the seventh section.

A REVIEW OF MULTIVARIATE DESCRIPTIVE PROCESSES BASED ON NOMINAL CONTINGENCY TABLES

Nominal categorical multivariate processes are processes in which at least two categorical variables with a nominal characteristic are simultaneously correlated, and contingency tables are used to illustrate this correlation. In order to model such contingency tables, a communication function called the nominal log-linear model is used as the following equation (Kamran Rad et al., 2019).

If $\boldsymbol{n} = (n_1, n_2, \ldots, n_N)^T$ and $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_N)^T$ represent the column vector of observations and the expected values for the cells of the contingency table, respectively, then the matrix form of the log-linear model is defined as follows.

$$Log\mu = \boldsymbol{X}\hat{\beta} \quad (1)$$

or equivalent to

$$Log\mu = \mathbf{1}\hat{\beta}_0 + \sum_{i=1}^{2^p-1} \boldsymbol{X}\hat{\beta}_i$$

Where p is the number of variables in the contingency table. In addition, X and β are design matrix and column vector of the model parameters, respectively. Note that, in Phase 2 of Monitoring Processes, it is assumed that the process parameters are known and do not need to be estimated. For review and more information on monitoring nominal contingency tables in phase 2, refer to (Kamran-Rad et al., 2017b).

DEVELOPMENT OF CORRESPONDENCE ANALYSIS BASED ON CLUSTERING TECHNIQUE TO REDUCE THE CONTINGENCY TABLES SCALE

One of the important issues in monitoring processes based on large-scale contingency tables is the low efficiency of control charts in diagnosing out-of-control status despite the large amount of data. In fact, large dimensions of contingency tables will have a negative impact on the performance of control charts. The main component analysis technique is used to solve the problem of large data volumes in continuous space. However, since contingency tables contain discontinuous categorical datasets, it is not possible to use this method. According to the studies, the efficient method in analyzing and reducing the dimensions of contingency tables is the correspondence analysis technique. In this research, the correspondence analysis has been developed based on the clustering technique.

I. Clustering in correspondence analysis

Correspondence analysis is one of the types of data analysis methods, the most important part of which is data grouping. Greencre (2007) has proposed two methods for grouping rows and columns of a contingency table in the CA method. In the first method, the integration is based on the logical proximity of rows or columns, and in other words, the rows and columns are integrated and collected without any special calculation. The second method considered in this research is the clustering method. In this method, rows or columns are compared based on proximity criteria based on total inertia, and the most similar rows or columns of the contingency table are selected for integration. Total inertia is calculated using the following equation.

$$\sum_i r_i d_i^2 = \sum_g \bar{r}_g \bar{d}_g^2 + \sum_g \sum_{i \in g} r_i d_{ig}^2 \qquad (2)$$

Where r_i is the mass of i row, d_i the chi-square distance between the row i profile with the mean of the row profiles, $\bar{r}g$ the mass of the members of group g, $\bar{d}g$ the chi-square distance of the group with the mean of the row profiles, and the intergroup inertia can be obtained from the total inertia. At the beginning of the clustering process, all rows and columns are separated and the intergroup inertia is equal to the total inertia. With each integration between categories, the inertia between groups decreases. So the first step is to identify the two rows (columns) that need to be integrated. Equation (3) is used to achieve the best rows for integration. According to the following equation, the smaller the distance for two rows (two columns), the two rows (columns) will integrate with each other (Greencre, 2007).

$$\frac{\bar{r}_g \bar{r}_h}{\bar{r}_g + \bar{r}_h} \left\| \bar{a}_g - \bar{a}_h \right\| \qquad (3)$$

Where

$$\|a_i - a_{i'}\| = \sqrt{\sum_j (a_{ij} - a_{i'j})^2 / c_j} \qquad (4)$$

It should be noted that the above equations show the distance between the two rows of the contingency table and can be generalized to the column distance by changing the row parameters to a column. The distance between the two rows is calculated using equation (3). In this method, first the similarity criterion of two rows (or two columns) is obtained and then two rows (or columns) with the shortest distance are merged with each other. This continues until the number of rows and columns reaches the desired number. It should be noted that the magnitude of the intergroup inertia index indicates the proper integration of the merged groups. The following figure illustrates well the clustering schematic in the contingency tables based on correspondence analysis for a four-sided table.

X1		1	2	3	…	*h*1
			X3			
X2	X4	1	2	3	…	*h*3
1	1	*O*1111	*O*1121	*O*1131	…	*O h*1 1 *h*3 1
2	2	*O*1212	*O*2222	*O*3232	…	*O h*1 2 *h*3 2
3	3	*O*1323	*O*2323	*O*3333	…	*O h*1 3 *h*3 3

⋮	⋮	⋮	⋮	⋮	⋱	⋮
*h*2	*h*4	*O*1 *h*2 1 *h*4	*O*2 *h*2 2 *h*4	*O*3 *h*2 3 *h*4	…	*O h*1 *h*2 *h*3 *h*4

The figure above shows that rows 1, 2 and 3 and columns 1 and 2 are merged; In this case, the number of remaining rows will be equal to h_1-2 and the number of remaining columns equal to h_2-1. Also, O_{ijkl} for i=1,2,…,h_1 ،j=1,2,…,h_2 ،k=1,2,…,h_3 and l = 1,2,…, h_4 represent the observed frequency of cells (i, j, k, l) from the contingency table.

II. Calculation of Eigenvalues

Suppose there is a contingency table with row I and column J. In this case, the expected frequency values of cells (i, j) of the contingency table are calculated as

$$e_{ij} = \frac{n_{i.}n_{.j}}{N_{..}}; i = 1,2,\dots, I \text{ and } j = 1,2,\dots, J. \qquad (5)$$

where $n_{i.}$ And $n_{.j}$ are the sum of the observations values of row i and column j, respectively, and $N_{..}$ is the sum of all observations. In this regard, the matrix K with elements k_{ij} is calculated as follows:

$$k_{ij} = \frac{(n_{ij} - e_{ij})}{\sqrt{e_{ij}}} \qquad (6)$$

Then the SVD matrix for the K matrix can be defined as follows:

$$\boldsymbol{K} = \boldsymbol{U}\,\boldsymbol{D}\,\boldsymbol{V}^{\boldsymbol{T}} \qquad (7)$$

In this case, the Eigenvalues are calculated as follows:

$$\lambda_k = eig(KK^T) \qquad (8)$$

where k=min{I-1,J-1}.

DEVELOPMENT OF GENETIC META-HEURISTIC ALGORITHM TO REDUCE THE DIMENSIONS OF CONTINGENCY TABLES

In this section, a meta-heuristic method based on genetic algorithm is developed to reduce the dimensions of contingency tables. As mentioned earlier, the large size of the contingency table has a negative effect on, firstly, the estimation of the parameters of the log-linear model in phase 1 and secondly, in phase 2, the performance of control charts to detect out-of-control conditions is weakened. For this purpose and to reduce the effect of large dimensions of the contingency table, a genetic algorithm has been used.

In this algorithm, the design space is transformed into a genetic reproduction space that works with a series of coded variables. Depending on the randomization nature of genetic algorithms, the produced responses can be good, bad, or possibly. Therefore, determining the appropriate parameters plays an important role in obtaining the correct response over a period of time. This algorithm includes initial population selection, proportionality function calculation, parent selection, cross-child selection, and replacement of children with parents (Kamran Rad and Bashiri, 2015).The initial pre code of genetic algorithm to reduce the dimensions of contingency tables in this study is given in the figure 1.

It should be noted that the fit index of the bilateral contingency table in order to determine the best combination of children or in other words the best dimensions of the contingency table is obtained from the following equation.

Initial population

1. Design the initial large contingency table based on the collected observations

2- Form the initial matrix cells to form contingency tables with different dimensions

3- Randomly form the initial matrix after aggregating the rows of the initial contingency table

4- Randomly form the initial matrix after aggregation the columns of the initial contingency table

Determine the fit function

5. Obtain the fit index of the contingency table (Equation 9) after aggregating the rows and columns.

repeat

6- Continue until the fitness index is not less than a small desired number

7- Randomly select two children including two aggregated contingency tables by the minimum mean squared error.

8- Do clause 7, provided that the dimensions of the two selected tables are equal and also the sum of the dimensions of the two tables is not larger than the dimensions of the initial contingency table.

9- Obtain the fit of two new children

10. If the fitness of the new contingency table is less than the value specified in clause 6, stop the algorithm and otherwise go to clause 7.

11- If the fitness index of the new children of clause 10 is less than that of the two children of clause 9, save the children of clause 10, otherwise save the previous children and go to clause 7 again.

12- Continue clause 11 until the condition of clause 6 is fulfilled.

13- Display the best contingency table with reduced dimensions.

FIGURE 1
PRE-CODE OF GENETIC ALGORITHM TO REDUCE THE DIMENSIONS OF LARGE-SCALE CONTINGENCY TABLE

$$MSE = \sum_{i=1}^{h_1} \sum_{j=1}^{h_2} (O_{ij} - E_{ij})^2 \qquad (9)$$

As O_{ij} and E_{ij} are observed frequency and the expected frequency of cells (i, j), respectively. Also, the values of expected cell observations (i, j) can be calculated as follows.

$$E_{ij} = \frac{O_{i.} \times O_{.j}}{O_{..}} \quad (10)$$

As *Oi.* And *O.j* is the sum of the observations of row *i* and column *j*, respectively, and *O ..* is the sum of all observations.

MULTIVARIATE DESCRIPTIVE PROCESS MONITORING APPROACHES BASED ON LARGE-SCALE CONTINGENCY TABLES IN PHASE 2

Multivariate descriptive process monitoring approaches based on large-scale contingency tables in phase 2.

I. T^2 approach for monitoring non-metric processes based on large-scale contingency tables

In this section, the T^2 control chart for monitoring processes based on large-scale contingency tables in phase 2 based on the statistics introduced by Yeh et al. (2009) has been developed. The null hypothesis of this statistic is based on the equality of process parameters with the mean parameters over time and is written as follows.

$$H_0\colon \boldsymbol{\beta} = \boldsymbol{\beta}_0$$

$$H_1\colon \boldsymbol{\beta} \neq \boldsymbol{\beta_0} \quad (11)$$

To test the above hypothesis, the following statistic is developed.

$$T^2 = (\beta_i - \hat{\beta})^T \Sigma_{\beta}^{-1} (\beta_i - \hat{\beta}) \quad (12)$$

Where

$$\hat{\beta} = \frac{\sum_{i=1}^{m} \beta_i}{m} \quad (13)$$

$$cov(\beta) = \{\boldsymbol{X}'[diag(\mu) - \mu\mu'/N]\boldsymbol{X}\}^{-1}. \quad (14)$$

where $\hat{\boldsymbol{\mu}}$ is a vector of expected values and is obtained from the equation (10). In addition, in the variance-covariance matrix, the above relation is N number of the total data. It should be noted that the upper limit of the T^2 statistic is obtained by simulation so that the mean value of the controlled sequence length is equal to a certain value.

I. CMH approach for monitoring non-metric processes based on large-scale contingency tables

Suppose $\boldsymbol{n}_k$ is the observed frequency vector of a large-scale contingency table and $\mu_k = E(\boldsymbol{n}_k)$ the expected observation vector corresponds to the cells of the same contingency table. Let is the covariance matrix of the observations (which can be calculated from Equation 14). Now if we have

$$\boldsymbol{n} = \sum \boldsymbol{n}_k, \mu = \sum \mu_k, \boldsymbol{V} = \sum \boldsymbol{V}_k$$

Then the *CMH* test statistics will be as follows:

$$CMH = (\boldsymbol{n} - \mu)^T \boldsymbol{V}^{-1} (\boldsymbol{n} - \mu). \quad (15)$$

It should be noted that the upper limit of the CMH statistic is obtained by simulation so that the mean value of the controlled sequence length is equal to a certain value. To better understand of presented methodology, following points should be considered. First, we develop two algorithms including clustering-CA and GA to reduce dimension of the large contingency table. Then, two statistics, named *CMH* and T^2 have been developed to monitor the multivariate categorical processes in Phase II.

EVALUATING THE PERFORMANCE OF THE PROPOSED APPROACHES

In this section, the performance of two proposed approaches for monitoring classified multivariate processes based on large-scale contingency tables is evaluated. For this purpose, first, a bilateral large-scale contingency table with dimensions of 16 x 21 is considered as follows. Then, with each of the two clustering algorithms in the correspondence analysis and genetic meta-innovation, the dimensions of this table are reduced. We then evaluate and compare the performance of each of the two control charts in detecting the out-of-control status of multivariate processes based on large-scale contingency tables.

TABLE I
LARGE-SCALE INITIAL CONTINGENCY TABLE

X1 \ X2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	24	30	19	9	1	30	8	12	15	30	12	22	4	15	20	6
2	29	9	23	20	15	11	30	16	20	3	10	22	7	19	28	15
3	3	11	17	26	19	26	21	24	3	20	12	23	0	20	7	16
4	7	28	22	7	1	23	15	18	0	12	17	24	30	20	24	3
5	0	17	11	20	4	8	8	6	15	21	22	1	6	5	23	7
6	30	16	0	26	25	22	27	13	18	30	17	11	22	1	21	1
7	22	13	4	27	3	12	24	6	1	26	3	11	13	12	8	4
8	9	10	6	6	20	10	17	6	3	21	22	4	10	12	28	17
9	26	20	20	12	5	10	2	8	7	26	0	7	27	15	22	0
10	22	0	14	30	17	1	24	11	16	18	9	27	23	5	21	9
11	9	12	25	8	27	12	29	6	16	24	10	8	30	27	23	12
12	12	22	18	2	9	13	4	30	18	26	30	5	26	14	6	30
13	7	28	18	1	6	7	26	20	0	1	17	15	8	25	15	30
14	0	2	3	3	0	2	20	25	29	25	25	23	13	13	10	15
15	29	26	14	12	30	6	24	23	26	17	30	13	16	17	23	29
16	20	3	16	14	24	9	10	14	16	21	10	8	6	16	23	13
17	5	21	17	27	25	8	27	26	24	2	14	3	4	23	21	5
18	3	1	7	18	14	15	19	5	22	14	29	23	2	22	6	1
19	24	14	3	11	15	11	20	19	7	8	22	2	15	9	11	29
20	22	26	27	1	27	8	22	29	20	26	3	28	30	26	12	13
21	10	8	21	5	23	17	11	20	11	12	18	4	23	29	10	28

I. Evaluating the performance of the proposed control charts based on the correspondence analysis approach

In this section, the performance of the two proposed control charts for monitoring large-scale multivariate processes is evaluated and compared using the ARL index-based correspondence analysis approach. For this purpose, first, the dimensions of the large-scale contingency table are reduced using the correspondence analysis approach, and then a classified multivariate process is formed based on the small-scale table. Finally, the proposed control charts will be used to monitor this process. It should be noted that the upper limit of control of these two control graphs is calculated so that the mean value of the sequence length in the controlled state is 200. Accordingly, the upper limit values of the two proposed control graphs equal to 222.12 and 12.88 have been calculated for CMH and T^2, respectively, which will be used in the simulations.

Using the correspondence analysis approach, the initial contingency table is reduced to a 9 x 5 table. Based on this table, the values of the parameters of the bivariate log-linear model should be estimated using the Newton-Raphson recursive algorithm (Kamran Rad et al., 2019). The final relation of this algorithm is given below.

$$\hat{\beta}^{(t+1)} = \hat{\beta}^{(t)} + \left[\boldsymbol{X}^T diag\left(\hat{\mu}^{(t)}\right)\boldsymbol{X}\right]^{-1}\boldsymbol{X}^T\left(\boldsymbol{n} - \hat{\mu}^{(t)}\right), \quad (16)$$

where $\hat{\boldsymbol{\beta}}^{(t+1)}$ is (T + 1)th of the estimated vector of the parameters of the log-linear model in phase 1 and $\hat{\boldsymbol{\beta}}^{(0)}$ also the initial estimate of **β** which can be calculated based on the estimation of Ordinary Least Square (Yeh et al., 2009). **X** is the design matrix and **n** the observation vector of the contingency table so that $\boldsymbol{n}^T\boldsymbol{1} = \boldsymbol{N}$. Also, $\hat{\boldsymbol{\mu}}^{(t)}$ is the estimated vector of the expected observations vector in the contingency table that can be calculated as follows:

$$\hat{\mu}^{(t)} = exp(\boldsymbol{X}\hat{\beta}^{(t)})$$

If $\|\widehat{\boldsymbol{\beta}}^{(t)} - \widehat{\boldsymbol{\beta}}^{(t-1)}\| / \|\widehat{\boldsymbol{\beta}}^{(t-1)}\| \leq \varepsilon$, where $\|\widehat{\boldsymbol{\beta}}^{(t)}\|$ is the Euclidean norm of an estimated parameter at time t and ε also a very small constant (for example ε = 10^{-5}), then $\hat{\beta} = \hat{\beta}^{(t)}$ will be desirable estimated parameters. Based on the reduced table, the parameters of the log-linear model based on the bilateral reduced contingency table are estimated as, $\beta_0 = 3.35$, $\beta_1 = 0.02$, $\beta_2 = 0.13$, and $\beta_{12} = 0.05$. It should be noted that the results of the simulation studies are presented in the next section along with the results of using the genetic algorithm in tables 2 to 6.

II. Evaluating the performance of the proposed control charts based on genetic algorithm

In this section, the performance of two proposed control charts for monitoring large-scale multivariate processes using an ARL index-based genetic algorithm is evaluated and compared. For this purpose, the dimensions of the large-scale contingency table are first reduced using a genetic algorithm, and then a classified multivariate process is formed based on the small-scale table. Finally, the proposed control charts will be used to monitor this process. It should be noted that the upper limit of control of these two control graphs is calculated so that the mean value of the sequence length in the controlled state is 200. Accordingly, the upper limit values of the two proposed control graphs equal to 50.70 and 12.38 have been calculated for CMH and T^2, respectively, which will be used in the simulations. Also, using the correspondence analysis approach, the initial contingency table is reduced to a 5 x 6 table. Based on the reduced table, the parameters of the log-linear model based on the bilateral reduced contingency table are estimated as $\beta_0 = 9.38$, $\beta_1 = -1.2$, $\beta_2 = -0.70$, and $\beta_{12} = 0.02$. The results of the simulated studies in tables 2 to 4 are presented in order to evaluate the performance of the proposed control charts under different positive and negative individual shifts in parameters β1, β2 and β12. It is worth noting that the performance of the two proposed control charts *CMH* and T^2 has been investigated based on using the two genetic algorithms and correspondence analysis.

TABLE 2

ARL VALUES UNDER DIFFERENT SHIFTS IN β_1

	Shift										
Control chart/Method	-2.0	-1.5	-1.0	-0.5	-0.2	0.0	0.2	0.5	1.0	1.5	2.0
CMH/GA	1.00 (0.00)	1.00 (0.00)	1.02 (0.00)	9.35 (0.29)	142.75 (4.39)	206.71 (6.63)	151.20 (5.03)	11.33 (0.32)	1.39 (0.01)	1.00 (0.00)	1.00 (0.00)
CMH/CA	1.00 (0.00)	1.00 (0.00)	1.26 (0.01)	10.57 (0.27)	140.32 (5.01)	203.21 (6.92)	153.97 (5.52)	12.52 (0.36)	1.79 (0.01)	1.00 (0.00)	1.00 (0.00)
T^2 /GA	1.00 (0.00)	1.00 (0.00)	1.23 (0.01)	10.29 (0.18)	146.25 (4.99)	200.77 (6.91)	152.33 (5.21)	12.67 (0.29)	1.92 (0.01)	1.00 (0.00)	1.00 (0.00)
T^2 /CA	1.00 (0.00)	1.00 (0.00)	1.19 (0.01)	10.82 (0.22)	141.33 (4.61)	202.89 (6.18)	152.82 (5.02)	12.29 (0.27)	1.55 (0.01)	1.00 (0.00)	1.00 (0.00)

TABLE 3

ARL VALUES UNDER DIFFERENT SHIFTS IN β_2

	Shift										
Control chart/Method	-2.0	-1.5	-1.0	-0.5	-0.2	0.0	0.2	0.5	1.0	1.5	2.0
CMH/GA	1.00 (0.00)	1.32 (0.02)	6.67 (0.21)	95.56 (3.02)	181.86 (5.86)	206.71 (6.63)	182.32 (5.77)	99.37 (3.11)	7.92 (0.33)	1.52 (0.02)	1.00 (0.00)
CMH/CA	1.00 (0.00)	1.67 (0.03)	7.75 (0.35)	99.35 (3.20)	183.44 (5.92)	203.21 (6.92)	185.25 (5.99)	100.08 (3.91)	8.93 (0.41)	1.69 (0.03)	1.00 (0.00)

T^2/GA	1.00 (0.00)	1.71 (0.03)	7.02 (0.32)	96.32 (3.17)	180.72 (5.62)	200.77 (6.91)	181.01 (5.39)	99.02 (3.08)	7.99 (0.31)	1.50 (0.02)	1.00 (0.00)
T^2/CA	1.00 (0.00)	1.28 (0.02)	6.92 (0.22)	96.22 (3.01)	181.95 (5.79)	202.89 (6.18)	183.20 (5.59)	99.82 (3.43)	8.02 (0.37)	1.54 (0.02)	1.00 (0.00)

TABLE 4

ARL VALUES UNDER DIFFERENT SHIFTS IN β_{12}

	Shift										
Control chart/ Method	-0.2	-0.15	-0.1	-0.05	-0.02	0.0	0.02	0.05	0.1	0.15	0.2
CMH/GA	1.00 (0.00)	1.00 (0.00)	1.21 (0.02)	10.77 (0.32)	20.92 (0.82)	206.71 (6.63)	19.55 (0.59)	1.07 (0.01)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
CMH/CA	1.00 (0.00)	1.00 (0.00)	1.56 (0.03)	10.77 (0.32)	21.45 (0.93)	203.21 (6.92)	20.24 (0.81)	1.13 (0.01)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
T^2/GA	1.00 (0.00)	1.00 (0.00)	1.45 (0.02)	10.03 (0.30)	20.35 (0.77)	200.77 (6.91)	19.01 (0.51)	1.27 (0.01)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
T^2/CA	1.00 (0.00)	1.00 (0.00)	1.31 (0.02)	10.15 (0.29)	21.03 (0.79)	202.89 (6.18)	19.92 (0.63)	1.19 (0.01)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)

As can be seen from Tables 2 to 4, the performance of the two proposed control charts under different shifts has been evaluated in the parameters of the log-linear model. It should be noted that in order to more accurately evaluate the performance of the two control charts, negative shifts from the parameters of the applied model to the effect of such shifts have been observed. The results of the above tables show that in the negative individual shifts in the parameters of the log-linear model, the performance of the CMH control graph along with the genetic algorithm outperforms the same control graph along with the correspondence analysis algorithm. In addition, based on the negative shifts in the parameters of the log-linear model, the T^2 control chart along with the genetic algorithm performs better under small shifts than the same control chart along with the correspondence analysis algorithm under medium and large shifts.

Looking at the results of Tables 2 to 4 based on positive shifts in the model parameters, such results can be observed and repeated. On the other hand, the comparison of ARL values in the above tables under positive and negative shifts indicates a greater impact of negative shifts than positive shifts in individual impact parameters because the mean sequence length values in negative shifts are less than the corresponding small and medium positive shifts. But in large shifts in model parameters, similar performance can be seen in both positive and negative shifts. In what follows, the performance of the proposed control diagrams under simultaneous shifts in the parameters of the log-linear model is evaluated. Thus, first the performance of *CMH* control chart along with two genetic algorithms and correspondence analysis are given in Table 5 and then in Table 6, the results of T^2 control chart performance are presented and analyzed.

As can be seen in Table 5, different shifts with positive and negative values have been applied to the parameters of the log-linear model simultaneously. In order to evaluate and analyze the results, several different modes must be considered, which we will discuss in this section. First, simultaneous negative shifts (first state) in two parameters are examined. In this mode, as in the mode of individual shifts, the performance of the *CMH* control chart alongside the genetic algorithm has caused a better performance for this control chart than the performance alongside the correspondence analysis algorithm. In the mode of positive simultaneous shifts in the first slope parameter and negative in the second slope parameter, the results of the first mode are repeated. These results are true for comparing the performance of two proposed control charts under positive simultaneous shifts in the slope parameters of the log-linear model.

In addition, the results of the above table show the better performance of the *CMH* control chart along with the genetic algorithm under negative simultaneous shifts in the slope parameters of the log-linear model, compared to the same control chart with correspondence algorithm. These results have been repeated for positive simultaneous shifts in model slope parameters.

In what follows, the results of the simulations to evaluate the performance of the proposed T^2 control chart along with the two genetic algorithms and the correspondence analysis under simultaneous shifts in the model slope parameters are reported in Table 6.

TABLE 5
ARL VALUES OF CMH CONTROL CHART UNDER DIFFERENT SIMULTANEOUS SHIFTS IN β_1 AND β_2

Control chart/ Method	Shift in β_2 / β_1	-2.0	-1.5	-1.0	-0.5	-0.2	0.2	0.5	1.0	1.5	2.0
CMH/GA	-2.0	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.18 (0.01)	1.89 (0.02)	1.91 (0.02)	1.22 (0.02)	1.02 (0.00)	1.00 (0.00)	1.00 (0.00)
	-1.5	1.00 (0.00)	1.00 (0.00)	1.09 (0.00)	1.31 (0.02)	2.08 (0.04)	2.12 (0.05)	1.39 (0.02)	1.21 (0.02)	1.00 (0.00)	1.00 (0.00)
	-1.0	1.00 (0.00)	1.00 (0.00)	1.22 (0.01)	3.20 (0.12)	6.26 (0.19)	7.18 (0.21)	4.23 (0.22)	3.43 (0.19)	1.78 (0.13)	1.00 (0.00)
	-0.5	1.00 (0.00)	1.00 (0.00)	1.72 (0.00)	7.16 (0.21)	38.62 (1.18)	51.32 (1.91)	12.76 (1.09)	2.79 (0.26)	1.66 (0.08)	1.04 (0.01)
	-0.2	1.00 (0.00)	1.00 (0.00)	1.91 (0.00)	7.81 (0.22)	88.62 (2.73)	91.89 (2.03)	18.98 (1.93)	11.25 (1.21)	1.64 (0.48)	1.05 (0.01)
	0.2	1.00 (0.00)	1.02 (0.00)	9.23 (0.11)	19.33 (0.27)	89.95 (2.94)	97.90 (2.54)	21.45 (1.79)	12.76 (1.66)	1.80 (0.71)	1.08 (0.01)
	0.5	1.00 (0.00)	1.00 (0.00)	1.02 (0.00)	10.29 (0.31)	73.09 (2.21)	80.34 (2.32)	16.36 (1.03)	5.07 (0.56)	1.84 (0.09)	1.04 (0.01)
	1.0	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	10.05 (0.13)	29.46 (2.03)	29.35 (2.14)	21.50 (1.80)	4.09 (0.99)	1.78 (1.02)	1.00 (0.00)
	1.5	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	5.59 (0.28)	12.08 (1.02)	15.93 (1.93)	10.32 (0.83)	2.68 (0.08)	1.16 (0.93)	1.00 (0.00)
	2.0	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	2.67 (0.19)	6.57 (0.44)	9.22 (0.89)	2.93 (0.23)	1.09 (0.02)	1.00 (0.00)	1.00 (0.00)
CMH/CA	-2.0	1.00 (0.00)	1.00 (0.00)	1.01 (0.00)	1.22 (0.01)	1.93 (0.03)	1.95 (0.03)	1.24 (0.02)	1.02 (0.00)	1.00 (0.00)	1.00 (0.00)
	-1.5	1.00 (0.00)	1.00 (0.00)	1.21 (0.02)	1.38 (0.03)	2.32 (0.05)	2.44 (0.06)	1.42 (0.03)	1.24 (0.02)	1.01 (0.00)	1.00 (0.00)
	-1.0	1.00 (0.00)	1.01 (0.00)	1.32 (0.01)	3.44 (0.17)	6.97 (0.29)	7.82 (0.28)	5.11 (0.21)	3.91 (0.14)	1.71 (0.11)	1.00 (0.00)
	-0.5	1.00 (0.00)	1.02 (0.00)	1.04 (0.01)	7.72 (0.23)	39.91 (1.33)	53.01 (1.88)	13.21 (1.07)	5.13 (0.20)	1.79 (0.08)	1.04 (0.01)
	-0.2	1.00 (0.00)	1.00 (0.00)	1.02 (0.00)	7.99 (0.26)	90.04 (2.89)	92.19 (2.24)	20.03 (1.94)	10.99 (1.29)	1.69 (0.47)	1.07 (0.01)
	0.2	1.00 (0.00)	1.04 (0.00)	2.00 (0.09)	19.58 (0.22)	95.45 (2.91)	96.24 (2.67)	25.75 (1.88)	12.98 (1.67)	1.99 (0.62)	1.10 (0.01)
	0.5	1.00 (0.00)	1.00 (0.00)	1.08 (0.00)	10.41 (0.48)	75.29 (2.33)	82.24 (2.41)	17.02 (1.00)	5.44 (0.59)	1.92 (0.08)	1.03 (0.01)
	1.0	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	9.94 (0.12)	29.79 (2.11)	28.73 (1.78)	13.29 (1.32)	4.39 (0.97)	1.39 (0.89)	1.00 (0.00)
	1.5	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	5.92 (0.26)	12.44 (0.77)	18.33 (1.72)	9.59 (0.99)	2.77 (0.09)	1.13 (0.82)	1.00 (0.00)
	2.0	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	2.99 (0.14)	5.49 (0.00)	9.69 (0.81)	3.28 (0.21)	1.11 (0.02)	1.00 (0.00)	1.00 (0.00)

TABLE 6
ARL VALUES OF T^2 CONTROL CHART UNDER DIFFERENT SIMULTANEOUS SHIFTS IN β_1 AND β_2

Control chart/ Method	Shift in β_2 / β_1	-2.0	-1.5	-1.0	-0.5	-0.2	0.2	0.5	1.0	1.5	2.0
T^2/GA	-2.0	1.00 (0.00)	1.00 (0.00)	1.02 (0.00)	1.68 (0.04)	1.92 (0.08)	1.99 (0.07)	1.28 (0.02)	1.03 (0.01)	1.00 (0.00)	1.00 (0.00)
	-1.5	1.00 (0.00)	1.00 (0.00)	1.10 (0.02)	1.86 (0.04)	2.32 (0.06)	2.71 (0.11)	1.41 (0.04)	1.39 (0.03)	1.02 (0.00)	1.00 (0.00)
	-1.0	1.00 (0.00)	1.01 (0.00)	1.36 (0.05)	3.02 (0.12)	6.88 (0.28)	8.09 (0.29)	4.97 (0.21)	3.82 (0.18)	1.71 (0.12)	1.00 (0.00)
	-0.5	1.00 (0.00)	1.01 (0.00)	1.82 (0.09)	7.00 (0.27)	35.77 (1.02)	50.03 (1.89)	12.25 (1.07)	5.57 (0.29)	1.89 (0.09)	1.04 (0.01)
	-0.2	1.00 (0.00)	1.03 (0.00)	1.99 (0.18)	7.97 (0.32)	86.32 (2.28)	90.21 (2.19)	19.45 (1.72)	10.31 (1.02)	1.99 (0.16)	1.06 (0.01)
	0.2	1.00 (0.00)	1.04 (0.00)	9.02 (0.10)	11.75 (0.23)	88.25 (2.74)	93.29 (2.42)	20.05 (1.66)	11.19 (1.07)	2.01 (0.42)	1.07 (0.01)
	0.5	1.00 (0.00)	1.01 (0.00)	1.08 (0.01)	9.93 (0.22)	69.59 (2.03)	72.55 (2.21)	15.02 (0.99)	7.78 (0.51)	1.97 (0.07)	1.03 (0.01)
	1.0	1.00 (0.00)	1.00 (0.00)	1.01 (0.00)	7.49 (0.19)	28.21 (1.80)	27.79 (2.00)	13.32 (0.97)	5.02 (0.96)	1.55 (0.35)	1.00 (0.00)
	1.5	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	5.21 (0.18)	12.02 (1.00)	14.03 (1.28)	9.78 (0.64)	2.08 (0.09)	1.06 (0.03)	1.00 (0.00)
	2.0	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	2.14 (0.11)	6.33 (0.39)	9.00 (0.75)	3.01 (0.21)	1.05 (0.01)	1.00 (0.00)	1.00 (0.00)
T^2/CA	-2.0	1.00 (0.00)	1.00 (0.00)	1.02 (0.00)	1.33 (0.05)	1.99 (0.08)	2.00 (0.09)	1.67 (0.04)	1.03 (0.00)	1.00 (0.00)	1.00 (0.00)
	-1.5	1.00 (0.00)	1.01 (0.00)	1.09 (0.01)	1.52 (0.03)	2.13 (0.07)	2.51 (0.08)	1.94 (0.03)	1.42 (0.02)	1.00 (0.00)	1.00 (0.00)
	-1.0	1.00 (0.00)	1.01 (0.00)	1.51 (0.12)	3.89 (0.18)	7.76 (0.24)	8.93 (0.49)	4.79 (0.33)	4.01 (0.15)	1.65 (0.11)	1.00 (0.00)
	-0.5	1.00 (0.00)	1.02 (0.00)	1.92 (0.21)	8.29 (0.33)	40.14 (1.26)	47.09 (1.77)	14.46 (1.12)	5.50 (0.38)	2.21 (0.13)	1.13 (0.03)
	-0.2	1.00 (0.00)	1.03 (0.00)	3.46 (0.28)	11.56 (0.49)	90.36 (2.99)	92.94 (3.00)	20.11 (1.87)	12.33 (1.25)	3.64 (0.48)	1.15 (0.05)
	0.2	1.00 (0.00)	1.04 (0.00)	7.29 (0.29)	19.33 (0.45)	91.71 (2.92)	98.33 (2.49)	24.35 (1.71)	14.02 (1.68)	3.22 (0.88)	1.16 (0.09)
	0.5	1.00 (0.00)	1.01 (0.00)	1.28 (0.09)	11.91 (0.28)	75.82 (2.02)	81.38 (2.11)	19.00 (1.47)	6.61 (1.02)	1.93 (0.47)	1.05 (0.02)
	1.0	1.00 (0.00)	1.01 (0.00)	1.03 (0.01)	9.95 (0.17)	27.37 (1.90)	30.04 (2.08)	18.22 (1.07)	4.35 (0.93)	1.09 (0.77)	1.01 (0.00)
	1.5	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	5.73 (0.21)	15.63 (1.13)	16.67 (1.72)	10.97 (0.74)	3.26 (0.11)	1.16 (0.53)	1.00 (0.00)
	2.0	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	3.04 (0.03)	6.90 (0.51)	9.79 (0.66)	3.38 (0.28)	1.11 (0.04)	1.00 (0.00)	1.00 (0.00)

The results of table 6 can be examined in several modes (as in Table 5). First, simultaneous negative shifts (first mode) in two parameters are examined. In this mode, as in the mode of individual shifts, the performance of the T^2 control chart along with the genetic algorithm has created better performance for this control chart than the performance alongside the correspondence analysis algorithm under small shifts and the better performance of the T^2 control chart along with the correspondence analysis algorithm can be observed under simultaneous medium and large negatives shifts. In the mode of simultaneous positive shifts in the first slope parameter and negative in the second slope parameter, the results of the first mode are repeated. But the results of the above table to compare the performance of the two proposed control charts under most of the simultaneous positive shifts indicate better performance of the control chart along with the genetic algorithm in the slope parameters of the logarithm model. In addition, the results of the above table represent the better performance of the T^2 control chart along with the genetic algorithm under simultaneous negative shifts in the slope parameters of the log-linear model, compared to the same control chart along with algorithm. The results of which have been also repeated for simultaneous positive shifts in model slope parameters. Based on the general results observed from Tables 2 to 6, it is clear that the proposed control charts along with the genetic algorithm have a greater ability to detect out-of-control conditions in most different shifts in process parameters, which is important in that in this study, we were able to develop the performance of the existing correspondence analysis approach by improving a meta-heuristic algorithm in order to improve the efficiency of reducing the dimensions of the contingency table, as well as improving the performance of control charts. In addition, general studies show that the CMH control chart outperformed the T^2 control chart under various shifts in process parameters based on the log-linear model.

A PRACTICAL EXAMPLE IN THE RENEWABLE ENERGY INDUSTRY

One of the most important issues in the renewable energy industry is the amount of energy sales, especially electricity, which our country has suffered from frequent failure circuit in the summer due to the lack of power plants. As you know, energy consumption takes place in different areas ($X1$) such as domestic, public, agricultural, industrial, commercial and street lighting. One of the classified variables in this case study is the energy consumption variable which is classified into the six above categories. Also in this study, there are two other classified variables ($X2$) including peak load and city consuming energy. Peak load is divided into two categories: synchronous peak load and asynchronous peak load. In addition, the city of consuming ($X3$) is also divided into 14 cities or categories, which their names along with the relevant data in a tripartite contingency table with dimensions of 14 x 2 x 12 are listed in the following table.

TABLE 7
LARGE-SCALE CONTINGENCY TABLE IN THE RENEWABLE ENERGY INDUSTRY

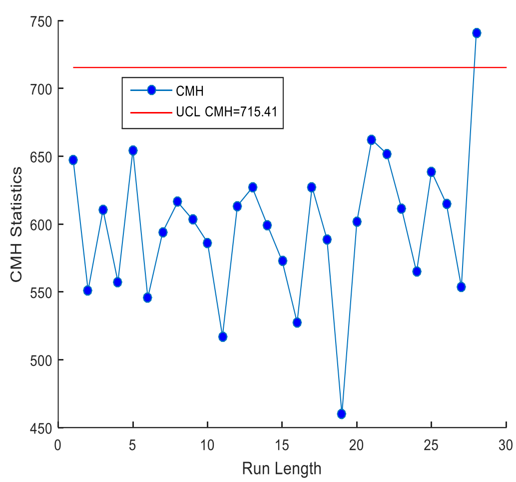
X1						*X2*						
X3_6	X3_5	X3_4	X3_3	X3_2	X3_1	X3_6	X3_5	X3_4	X3_3	X3_2	X3_1	City (X3)
85499.85	18899	20186.79	41943.23	17487.01	6655.496	72846.15	16102	17199.21	35735.77	14898.99	5670.504	1
99515.54	19067.38	26563.87	22492.22	20138.78	10223.79	80957.46	15511.62	21610.13	18297.78	16383.22	8317.212	2
119726.9	22263.96	13071.09	18331.1	26309.94	10253.47	97238.14	18082.04	10615.91	14887.9	21368.06	8327.527	3
88246.03	23402.26	8202.588	19952.02	25139.97	8822.612	62477.97	16568.74	5807.412	14125.98	17799.03	6246.388	4
19467.85	1619.023	1191.367	1849.383	3630.471	2755.104	16449.15	1367.977	1006.633	1562.617	3067.529	2327.896	5
55078.81	9839.911	14973.75	9640.388	19525.52	4699.151	48441.19	8654.089	13169.25	8478.612	17172.48	4132.849	6
109922.8	19339.84	32257.25	10902.5	24330.74	9269.062	91492.24	16097.16	26848.75	9074.5	20251.26	7714.938	7
53366.53	10277.68	6421.124	15780.35	13425.99	6301.466	45643.47	8790.321	5491.876	13496.65	11483.01	5389.534	8
95627.67	19790.36	20591.36	45696.81	22203.53	10175.23	82616.33	17097.64	17789.64	39479.19	19182.47	8790.766	9
89241.15	19321.97	21077.98	44333.17	17333.15	9164.314	77500.85	16780.03	18305.02	38500.83	15052.85	7958.686	10

103067.8	18742.79	21868.59	52688.63	23358.95	10359.67	85176.19	15489.21	18072.41	43542.37	19304.05	8561.329	11
91996.93	19740.81	21534.88	53411.52	23648.84	10567.58	76247.07	16361.19	17848.12	44267.48	19600.16	8758.416	12
99469.88	21194.5	20963.04	42205.51	23640.17	9662.766	78876.12	16806.5	16622.96	33467.49	18745.83	7662.234	13
88609.21	18641.11	21569.26	46750.57	17467.48	8803.318	75678.79	15920.89	18421.74	39928.43	14918.52	7518.682	14

As can be seen from the above table, we are faced with a large-scale table which, based on previous theories, the analysis of which and consequently evaluation the performance of the proposed control charts will be difficult. Therefore, first we reduce its dimensions using the better method specified in the previous section. Based on the previous results, it has been determined that the use of genetic algorithm has a better effect on the performance of control charts, so in this section, this algorithm is used to reduce the dimensions, in which the table below is a reduced 5 x 3 one. In other words, the initial table, which has 14 rows and 12 columns, has become a small table with 5 rows and 3 columns.

TABLE 8

REDUCED CONTINGENCY TABLE OF RENEWABLE ENERGY INDUSTRY USING GENETIC ALGORITHM



*X*2			*X*3
*X*1_2	*X*1_2	*X*1_1	
812000	663800	1563400	X3_1
109900	91500	176100	X3_2
95600	82600	220800	X3_3
89200	77500	207800	X3_4
92000	76200	235700	X3_5

Now, in order to evaluate the performance of the proposed control charts in detecting the out-of-control status in the real world, the data of Table (8) are used and the results of the case studies can be observed in Figures 2 and 3.

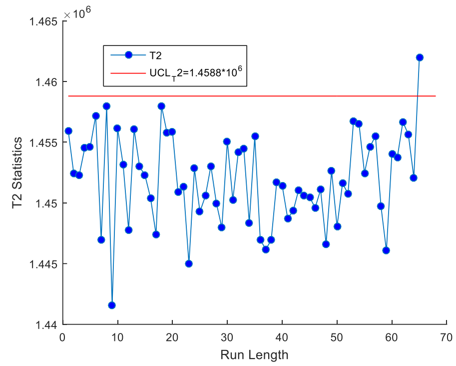
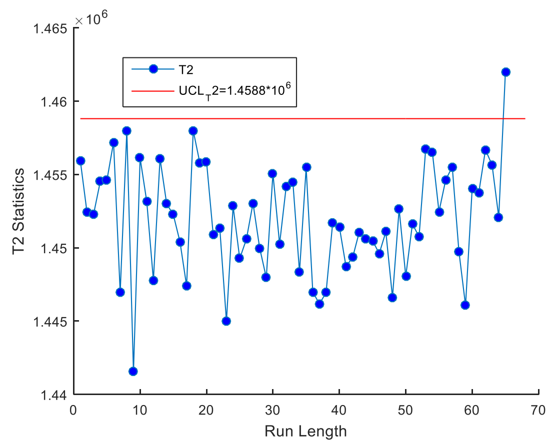


FIGURE 2
PERFORMANCE OF THE CMH CONTROL CHART IN DETECTING THE STATUS OF THE RENEWABLE ENERGY PROCESS

FIGURE 3
PERFORMANCE OF T^2 CONTROL CHART IN DETECTING THE PROCESS STATUS OF RENEWABLE ENERGY

As can be seen from the above two figures, the CMH control chart shows the out-of-control status at point 28th and the T^2 control diagram shows the out-of-control status at point 65th. The results show that the CMH control chart performed better than the other control charts, which is also the case in the simulated studies.



CONCLUSIONS AND FURTHER SUGGESTIONS

In this study, two proposed control charts, CMH and T^2 were developed to monitor classified multivariate processes based on large-scale contingency tables in phase 2. As stated in this study, the large size of the contingency tables has an adverse effect on the performance of control charts in order to detect the out-of-control status in a certain period of time, in which case the results of control charts will face significant computational error. One of the best ways to reduce the negative effects is to reduce the dimensions of the contingency table. In this study, an existing method called correspondence analysis was used. In addition, in order to improve the performance of the proposed control charts, a powerful meta-heuristic algorithm called the genetic algorithm was used to reduce the dimensions of the contingency table to minimize the distance between the observed values and the expected values. Then, the performance of the two proposed control charts was evaluated based on the reduced contingency tables according to the two reduction methods, and the simulation results represent better performance of the CMH control graph along with the genetic algorithm than the other control graph. In addition, in order to efficiently demonstrate the proposed control charts in the real world, a case study in the field of renewable energy was used and the results also confirmed the simulation results. Thus, in this field, effective measures can be taken in order to reduce energy losses, increase reliability, modernize distribution networks, reduce energy theft, and also identify wasteful routes, which will undoubtedly lead to improvement of services using the results obtained. Besides, the present study will be a fundamental and significant factor in the management of sustainable livable cities.

Another point is about the proposed methods limitation. As, these approaches have been developed to monitor the multivariate categorical processes, therefore, the use of these methods for metric processes may reduce the effectiveness of control charts in detecting the process status. Among the suggestions that can be made for future research is the development of other meta-heuristic algorithms to reduce the data that the more the dimensions of the contingency table are reduced with higher efficiency, the more positive impact will have on the performance of the proposed control charts. Furthermore, the development of proposed methods for monitoring classified multivariate processes based on large-scale contingency tables taking into account the correlations between neighboring cells will be a suggestion for future research.

REFERENCES

[1] Amiri, A., Hakimi, A., Farughi, H., & Arkat, J. (2022). Phase I Monitoring of Multivariate Ordinal based Processes: The MR and LRT Approaches (A real case study in drug dissolution process). International Journal of Industrial Engineering and Production Research, 33(1).

[2] Farahani, A., Tohidi, H., Shoja, A. (2019). An integrated optimization of quality control chart parameters and preventive maintenance using Markov chain. Advances in Production Engineering & Management. 14(1), pp 5–14. https://doi.org/10.14743/apem2019.1.307

[3] Farahani, A., Tohidi, H., Shoja, A. (2020). Optimization of Overall Equipment Effectiveness with Integrated Modeling of Maintenance and Quality. Engineering Letters, 28(2),

[4] Greenacre ،M. (2007). Correspondence analysis in practice. CRC press.

[5] Hakimi, A., Farughi, H., Amiri, A., & Arkat, J. (2019). New phase II control chart for monitoring ordinal contingency table based processes. Journal of Industrial and Systems Engineering, 12(Statistical Processes and Statistical Modeling).

[6] Hakimi, A., Farughi, H., Amiri, A., & Arkat, J. (2021). Phase II Monitoring of the Ordinal Multivariate Categorical Processes. Advances in Industrial Engineering, 55(3), 249-267.

[7] Kamranrad, R., & Bashiri, M. (2015). A novel approach in multi response optimization for correlated categorical data., scientia Iranica, 1117-1129.

[8] Kamranrad, R., Amiri, A., & Niaki, S.T.A. (2017a). New Approaches in Monitoring Multivariate Categorical Processes based on Contingency Tables in Phase II. Quality and Reliability Engineering International, 33(5), 1105-1129.

[9] Kamranrad, R., Amiri, A., & Niaki, S.T.A. (2017b). Phase II monitoring and diagnosing of multivariate categorical processes using generalized linear test-based control charts, Communications in Statistics-Simulation and Computation, 46(8), 5951-5980.

[10] Kamranrad, R., Amiri, A., Niaki, S.T.A. (2019). "Phase-I monitoring of log-linear model-based processes (a case study in health care: Kidney patients)," Quality and Reliability Engineering International. 35(6), 1766-1788.

[11] Li J., Tsung F., Zou C. (2012). Directional control schemes for multivariate categorical processes. Journal of Quality Technology. 44: 136-155.

[12] Li J., Zou C., Wang Z., Huwang L. (2013). A multivariate sign chart for monitoring process shape parameters. Journal of Quality Technology. 45: 149-165.

[13] Li, J., Tsung, F., & Zou, C. (2014a). Multivariate binomial/multinomial control chart. IIE Transactions, 46(5), 526-542.

[14] Li, J., Tsung, F., & Zou, C. (2014b). A simple categorical chart for detecting location shifts with ordinal information. International Journal of Production Research, 52(2), 550-562.

[15] Tsung F., Zou C. (2013). Statistical process control for multistage processes with binary outputs. IIE Transactions. 45: 1008- 1023.

[16] Wang, Y., Snee, R.D., Keyvan, G., Muzzio, F.J. (2016). "Statistical comparison of dissolution profiles," Drug Development and Industrial Pharmacy. 42(5), pp. 796–807.

[17] Yamamoto, K., Murakami, H. Model based on skew normal distribution for square contingency tables with ordinal categories. Comput Stat Data Anal. (2014).

[18] Yashchin E. (2012). On detection of changes in categorical data. Quality Technology & Quantitative Management. 9: 79-96.

[19] Yeh A.B., Huwang L. and Li Y.M. (2009). Profile monitoring for a binary response. IIE Transactions, 41: 931-941.

[20] Zafar, S. (2017). Non-iterative Estimation Methods for Ordinal Log-linear Models. Doctoral dissertation, The University of Newcastle.