Sains Malaysiana 51(12)(2022): 4153-4160
http://doi.org/10.17576/jsm-2022-5112-22

Performance Analysis and Discrimination Procedure of Two-Group Location Model with Some Continuous and High-Dimensional of Binary Variables

(Analisis Prestasi dan Prosedur Pembezaan Model Lokasi Dua Kumpulan dengan Sebilangan Pemboleh Ubah Selanjar dan Dimensi Tinggi Pemboleh Ubah Binari)

HASHIBAH HAMID[1,*], FRIDAY ZINZENDOFF OKWONU[2], NOR AISHAH AHAD[1] & HASLIZA ABDUL RAHIM[3]

[1]*School of Quantitative Sciences, UUM College of Arts and Sciences, Universiti Utara Malaysia, 06010 UUM Sintok, Kedah, Malaysia*
[2]*Department of Mathematics, Delta State University, Abraka, Delta, P.M.B.1, Nigeria*
[3]*School of Computer and Communication Engineering, Universiti Malaysia Perlis, 02600 UniMAP Arau, Perlis, Malaysia*

Received: 23 April 2022/Accepted: 26 August 2022

ABSTRACT

This research's primary goal was to evaluate the performance analysis of the recently constructed smoothed location models (SLMs) for discrimination purposes by combining two kinds of multiple correspondence analysis (MCA) to handle high dimensionality problems arising from the binary variables. A previous study of SLM, together with MCA as well as principal component analysis (PCA), displayed that the misclassification rate was still very high with respect to a large number of binary variables. Thus, two new SLMs are constructed in this paper to solve this particular problem. The first model results from the combination of SLM with Burt MCA (denoted as SLM+Burt), and the second one is with the joint correspondence analysis (denoted as SLM+JCA). The findings showed that both models performed well for all sample sizes (n) and all binary variables (b) under investigation, except n=60 and b=25 for the SLM+JCA model. Overall, the SLM+JCA model yields a greater performance in contrast to the SLM+Burt model. Moreover, the concept and procedures of the discrimination for the two-group classification conducted in this paper can be extended to multi-class classification as practitioners often deal with many groups and complexities of variables.

Keywords: Discrimination; large binary variables; misclassification rate; multiple correspondence analysis; smoothed location model

ABSTRAK

Matlamat utama penyelidikan ini adalah untuk menilai analisis prestasi model lokasi terlicin (SLMs) yang dibina sebelum ini untuk tujuan pembezaan dengan menggabungkan dua jenis analisis kesepadanan berganda (MCA) bagi menangani masalah dimensi tinggi yang berlaku daripada pemboleh ubah binari. Kajian terdahulu mengenai SLM bersama-sama dengan MCA serta analisis komponen utama (PCA), menunjukkan bahawa kadar salah pengelasan masih sangat tinggi dengan sejumlah besar bilangan pemboleh ubah binari. Oleh itu, dalam kajian ini, dua SLMs baharu dibina untuk menyelesaikan masalah khusus ini. Model pertama terhasil daripada gabungan SLM dengan Burt MCA (ditandakan sebagai SLM+Burt), dan yang kedua adalah dengan analisis kesepadanan bersama (ditandakan sebagai SLM+JCA). Hasil kajian menunjukkan bahawa kedua-dua model menunjukkan prestasi yang baik untuk semua saiz sampel (n) dan semua pemboleh ubah binari (b) di bawah kajian, kecuali untuk kes n=60 dan b=25 bagi model SLM+JCA. Secara keseluruhan, model SLM+JCA menghasilkan prestasi yang lebih baik berbanding model SLM+Burt. Selain itu, konsep dan prosedur pembezaan untuk pengelasan dua kumpulan yang dijalankan dalam kajian ini boleh diperluaskan kepada pengelasan berbilang kumpulan kerana pengamal sering berurusan dengan banyak kumpulan dan kerumitan pemboleh ubah.

Kata kunci: Analisis kesepadanan berganda; diskriminasi; kadar salah pengelasan; model lokasi terlicin; pembezaan; pemboleh ubah binari besar

Introduction

The smoothed location model (SLM) is among the most widely used discrimination methods for data comprising both continuous and binary variables concurrently (Hamid 2010). Although SLM can manage some empty cell issues, it is impractical when handling a large number of binary variables (Hamid 2010; Hamid, Ngu & Alipiah 2018). This is due to the over-parameterization issue, as large binary variables with high computational complexity burden and increases the computational cost and time prevent the computing process. Furthermore, biased estimators are achieved due to the occurrence of numerous empty cells from the inclusion of large binary variables which absolutely have a direct impact on the model designed. In a worst-case scenario, the model cannot even be constructed. In the study by Hamid (2014), the hybrid of Burt multiple correspondence analysis (MCA) and principal component analysis (PCA) have been implemented in the SLM to resolve the high dimensionality problem. However, such a model produces a high misclassification rate, especially when the binary variables considered are large. Burt MCA, Indicator MCA, Adjusted MCA as well as joint correspondence analysis (JCA) are the four variants of MCA developed by Greenacre and Blasius (2006) as well as Nenadic and Greenacre (2007). This paper discusess Burt MCA and JCA as they share similar characteristics. We build and compare the performance of the SLM with JCA and SLM with Burt MCA for research reasons. Thus, the SLMs possess two distinct variables extraction techniques, Burt MCA and JCA, which are developed to solve high dimensionality problems associated with large binary variables.

This research is being carried out as a result of inspirations and evidence from past research, which have typically restricted the number of binary variables to some acceptable value in order to develop a location model. For example, Vlachonikolis and Marriott (1982) employed a classical location model, which condensed the research into five binary variables. If the size of the training set is not large, Krzanowski (1983) only dealt with six binary variables while performing the maximum likelihood location model. Even if a large training set is available, the computational work required to estimate the misclassification rate grows in direct proportion to the number of discrete variables. Current computer capabilities preclude the addition of many variables. Hamid, Ngu and Alipiah (2018) further showed that even with only six binary variables, the SLM's utility is limited and sometimes impossible to implement.

As a result, it would be intriguing to design an alternate method for the location model that permits discrimination incorporating mixed variables, especially when the binary is of high dimension.

Materials and Methods

Smoothed Location Model (SLM)

Smoothed location models (SLM) discriminate new objects by classifying them into one of the two groups (denoted as π_1 and π_2) referring to the data vector comprising both binary and continuous variables. To conduct a discriminant analysis relying on the location model, the multinomial cell is exponentially extended from the binary variables following $s = 2^b$, in which s represents the multinomial cells' number and b denotes the binary variables' number in the research. SLM assumes a multivariate normal distribution having a mean $\boldsymbol{\mu}_{im}$ in cell m of π_i $(i = 1, 2)$ with a common covariance matrix $\boldsymbol{\Sigma}$ across all groups and cells for a vector of continuous variables. If the future object falls inside multinomial cell m and meets the requirement as Equation (1), $\mathbf{z}^T = (\mathbf{x}^T, \ \mathbf{y}^T)$ will be classified to π_1, given by

$$(\boldsymbol{\mu}_{1m} - \boldsymbol{\mu}_{2m})^T \Sigma^{-1}\left\{y - \frac{1}{2}(\boldsymbol{\mu}_{1m} - \boldsymbol{\mu}_{2m})\right\} \geq \log\left(\frac{p_{2m}}{p_{1m}}\right) + \log(a), \quad (1)$$

or else, it will be categorized to π_2 (Krzanowski 1995, 1993, 1980).

It is possible to estimate the smoothed mean vectors of the continuous variables y in cell m of π_i using

$$\hat{\boldsymbol{\mu}}_{imj} = \left\{\sum_{k=1}^{s} n_{ik} w_{ij}(m,k)\right\}^{-1} \sum_{k=1}^{m}\left\{w_{ij}(m,k)\sum_{r=1}^{n_{ik}} y_{rijk}\right\}, \quad (2)$$

under conditions $0 \leq w_j\ (m,k) \leq 1$ as well as $\left\{\sum_{k=1}^{s} n_{ik} w_j\,(m,k)\right\} > 0$, in which m, $k = 1, 2, \ldots, s$; $i = 1, 2$ and $j = 1, 2, \ldots, c$. Note that n_{ik} denotes the number of objects falling in cell k of π_i, y_{rijk} represents the j^{th} continuous variable of r^{th} object falling in cell k of π_i, meanwhile w_{ij} (m, k) resembles a weight with respect to cell m and variable j of all objects of π_i falling in cell k.

Some possible functions of weight w_{ij} (m, k) were given by Asparoukhov and Krzanowski (2000), but we focus only on the exponential function because of its most basic form as

$$w_{ij}(m,k) = \lambda_{\boldsymbol{ij}}^{\boldsymbol{d(m,k)}}, \quad (3)$$

in which $d(m,k)$ denotes the dissimilarity coefficient between the k^{th} and m^{th} cell of the binary vectors evaluated via the distance function as $d(m,k) = d(x_m, x_k) = (x_m - x_k)^T (x_m - x_k)$. In the meantime, the degree of smoothing parameter denoted by λ_{ij} is chosen from the [0, 1] interval, maximizing the leave-one-out pseudo-likelihood function, according to Asparoukhov and Krzanowski (2000) given by

$$PL_{loo}(\mathbf{\Lambda} \mid D) = \prod_{r=1}^{n} p(\mathbf{y}_r \mid D - \mathbf{z}_r, \mathbf{\Lambda}), \quad (4)$$

where $p(\mathbf{y}_r \mid D - \mathbf{z}_r, \mathbf{\Lambda}$ is the probability density of $\mathbf{y}_r$ if object r falls in cell m of π_i and $D - \mathbf{z}_r$ is the training set of π_1 and π_2 with object r excluded.

Next, a smoothed pooled covariance matrix is estimated using the smoothed cell mean vectors via

$$\hat{\Sigma} = \frac{1}{(n_1 + n_2 - g_1 - g_2)} \sum_{i=1}^{2} \sum_{m=1}^{s} \sum_{r=1}^{n_{im}} (y_{rim} - \hat{\boldsymbol{\mu}}_{im})(y_{rim} - \hat{\boldsymbol{\mu}}_{im})^T, \quad (5)$$

in which n_{im} represents the number of objects falling in cell m of π_i, y_{rim} resembles the vector of continuous variables of the r^{th} object in cell m of π_i while g_i denotes the number of non-empty cells of π_i.

Subsequently, the cell probabilities may be calculated using standardization of the exponential smoothing on the cell probability by

$$\hat{p}_{im(std)} = \hat{p}_{im} / \sum_{m=1}^{s} \hat{p}_{im}, \quad (6)$$

where

$$\hat{p}_{im} = \frac{\sum_{k=1}^{s} w(m,k)\, n_{im}}{n_i}. \quad (7)$$

VARIABLES EXTRACTION TECHNIQUES

To handle a large number of mixed variables, Hamid (2014) had integrated SLM with variable extraction techniques like multiple correspondence analysis (MCA) as well as principal component analysis (PCA). PCA was employed to minimize the huge number of continuous variables in the study, whereas MCA was employed to minimize the big number of binary variables. PCA is a prominent variable extraction technique that can decrease and compress data while maintaining as much as possible of the original dataset's most significant information (Hamid, Zainon & Yong 2016; Kemsley 1996). PCA decreases the data dimension by reducing a set of p correlated variables into orthogonal linear combinations of q uncorrelated variables (Jolliffe 1986). According to Massey (1965) and Rencher (2002), principal component scores (PCs) are a linear combination of the original variables that produces the maximum variance. PCs are often selected based on the most common measure, the Guttmann-Kaiser criterion, where the eigenvalue of any PCs greater than 1.0 will be retained in the model (Kaiser 1961).

Conversely, MCA is a widely utilized technique for dealing with categorical variables that have high dimensionality. Burt MCA is a substitute data structure for MCA that is employed to evaluate the entire two-way cross-tabulations set with identical margins in both vertical and horizontal tables. It can be represented as $\mathbf{B} = \mathrm{Z}^T \mathrm{Z}$, in which $\mathbf{B}$ is the Burt Matrix, and $\mathbf{Z}$ represents the indicator matrix (Greenacre 2007). In the meantime, joint correspondence analysis (JCA) is a unique algorithm that can correctly define the cross-tabulation of all variables by neglecting the Burt matrix's diagonal blocks and concentrating on the off-diagonal sub-table optimization. When the amount of explained variance revealed is at least 70%, both techniques are able to keep an optimal number of components intact.

MODEL CONSTRUCTION AND EVALUATION

First and foremost, JCA and Burt MCA are employed to extract a large number of binary variables for the goal of dimension reduction. Next, utilizing the smaller sets of extracted binary components obtained from the Burt MCA and JCA, two SLMs are created. Finally, the misclassification rate is employed to evaluate the newly created models employing the leave-one-out (LOO) approach. By dividing the total number of misclassified objects via the total number of observed samples, the misclassification rate is determined as

$$LOO = \frac{\sum_{k=1}^{n} error}{n}. \quad (8)$$

The strategy of performing variables extraction prior to construction of the smoothed location model was tested on some artificial data generated from a normal population using R software. The data generations create a collection of multivariate data for a few sample sizes (n) and various binary variables (b), with the number of continuous variables (c) was set to 10. Meanwhile, the size for the binary variables was set to 5, 10, 15, 20, and

25, whereas the sample size was set to 60, 120, and 180; for assessing the developed models in different angles. The procedure for discriminating and integrating SLM and Burt MCA is presented in Algorithm 1.

ALGORITHM 1

Step 1: Omit an object k from the sample n, in which the remaining objects are treated as a training dataset.

Step 2: Conduct Burt MCA to extract and decrease the large binary variables using the training dataset.

Step 3: Construct a new SLM using a reduced dataset from Step 2, producing a SLM+Burt model.

Step 4: Predict the group of the omitted object k employing the newly constructed SLM in Step 3, and assign an *error* $(\varepsilon_k) = 0$ if the prediction made is correct; otherwise, assign $\varepsilon_k = 0$.

Step 5: All steps from 1 to 4 are repeated until all objects $(k_1, k_2, ..., k_n)$ take turns successfully.

Step 6: Compute the rate of misclassification employing the LOO procedure.

To develop another new model named SLM+JCA, the steps in Algorithm 1 are repeated by replacing Burt MCA with JCA. Figure 1 demonstrates the study design that integrates SLM along with two types of MCA to perform discrimination tasks.

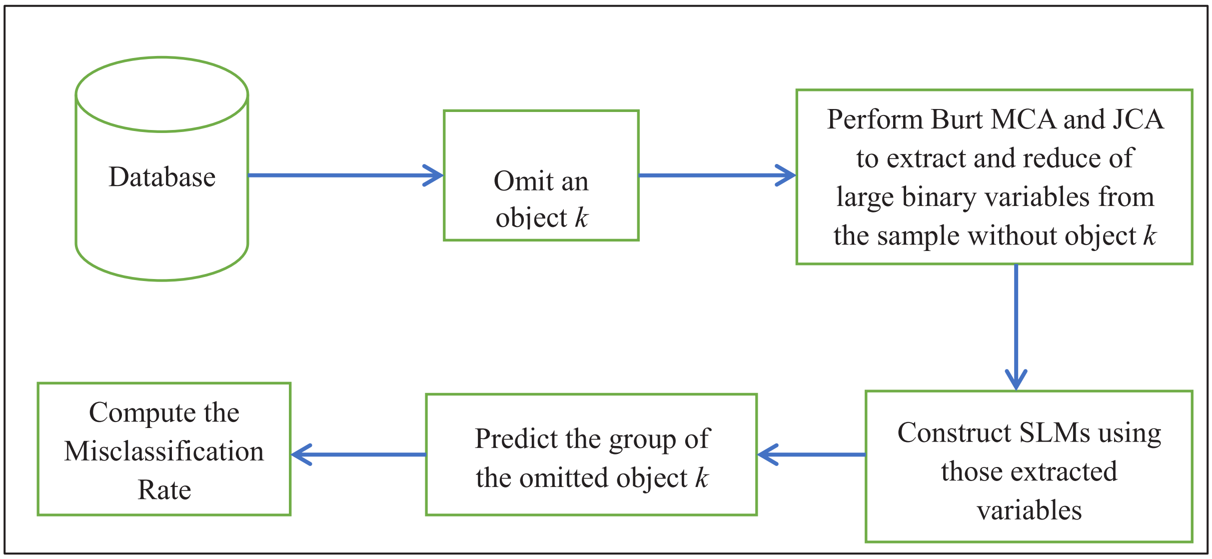

FIGURE 1. Construction of the new SLMs with high dimensional of the binary variables

RESULTS AND DISCUSSION

Three distinct sample sizes (n) and five binary variable settings (b) are used to compare the performance of the newly developed SLM+Burt and SLM+JCA models. Table 1, Table 3 and Table 4 illustrate the performance of the constructed models for n = 60, 120, and 180, accordingly. The overall outcomes demonstrate that the misclassification rate of the location model is strongly related to the three major factors, including (1) the number of binary extracted [the binary used in the model], (2) the separation or the distance between the observed groups [calculated utilizing Kullback-Leibler (KL) distance] as well as (3) the number of empty cells in the group. Based on the findings of n = 60 in Table 1, the misclassification rate is only obtained when b=25 for both models. The SLM+JCA model indicates substantially high misclassified objects as 0.1577 compared to the SLM+Burt model with only 0.0157 for its misclassification rate. This is due to this model extracting a smaller binary, which is five compared to six by the SLM+JCA model.

Although the difference of the extracted binary is only one unit, it has an effect on the separation between the group and the existence of empty cells, which ultimately affect the constructed model's performance. The smaller the binary extracted which was retained in the model, causing the distance between the observed groups farther with 16.01 units in the SLM+Burt model compared to 7.25 units in the SLM+JCA model, thereby making the SLM+Burt model performs better. Meanwhile, the number of empty cells is much higher for a larger binary size used to construct the model, as happens in the SLM+JCA model, leading to poor performance.

This analysis can be seen further in Table 2 as disclosing the percentage of empty cells for the SLM+JCA model is double over the SLM+Burt model, leading to 10 times worse in its misclassification rate. Accordingly, it can be inferred that the larger the binary retained/used in the model, the model's performance deteriorates.

TABLE 1. Performance analysis of the new constructed Smoothed Location Models tested on different binary sizes under n=60

SLM+Burt model	Size of Binary Variables				
	5	10	15	20	25
Misclassification Rate	0	0	0	0	0.0157
Number of Binary Extracted (PC_b)	2	4	4	5	5
Number of Non-empty Cells (π_1, π_2)	(4, 4)	(14, 14)	(14, 15)	(24, 25)	(22, 23)
Number of Empty Cells (π_1, π_2)	(0, 0)	(2, 2)	(2, 1)	(8, 7)	(10, 9)
KL Distance	291.27	271.50	86.55	16.23	16.01
SLM+JCA model	5	10	15	20	25
Misclassification Rate	0	0	0	0	0.1577
Number of Binary Extracted (PC_b)	2	3	3	5	6
Number of Non-empty Cells (π_1, π_2)	(4, 4)	(6, 6)	(8, 8)	(19, 22)	(23, 23)
Number of Empty Cells (π_1, π_2)	(0, 0)	(2, 2)	(0, 0)	(13, 10)	(41, 41)
KL Distance	235.92	96.11	353.51	84.23	7.25

TABLE 2. Performance of the constructed models under n=60 with b=25

Criteria	SLM+Burt model	SLM+JCA model
Misclassification Rate	0.0157	0.1577
Number of Binary Extracted (PC_b)	5	6
Number of Created Cells per Group	32	64
Percentage of Empty Cells (π_1, π_2)	(31.25%, 28.13%)	64.06%, 64.06%)

Table 3 further shows a strong relationship between the misclassification rate and extracted binary number that influences the distance between groups and the occurrence of empty cells indirectly. This relationship is obviously demonstrated that the SLM+JCA model records zero misclassification rate when no more than seven binary variables are extracted. Meanwhile, the SLM+Burt model records misclassified objects if more than seven binary variables are extracted.

TABLE 3. Performance analysis of the new constructed Smoothed Location Models tested on different binary sizes under n=120

SLM+Burt model	Size of Binary Variables				
	5	10	15	20	25
Misclassification Rate	0	0	0	0	0.0156
Number of Binary Extracted (PC_b)	3	5	6	7	8
Number of Non-empty Cells (π_1, π_2)	(6, 6)	(28, 29)	(38, 40)	(46, 50)	(53, 54)
Number of Empty Cells (π_1, π_2)	(2, 2)	(4, 3)	(26, 24)	(82, 78)	(203, 202)
KL Distance	154.02	832.81	158.83	32.53	8.32
SLM+JCA model	5	10	15	20	25
Misclassification Rate	0	0	0	0	0
Number of Binary Extracted (PC_b)	2	2	3	5	7
Number of Non-empty Cells (π_1, π_2)	(4, 4)	(4, 4)	(8, 8)	(27, 28)	(49, 45)
Number of Empty Cells (π_1, π_2)	(0, 0)	(0, 0)	(0, 0)	(5, 4)	(79, 83)
KL Distance	365.12	255.04	802.66	881.26	37.35

Next, the performance of the models against the number of empty cells (no object) is examined. For example, only eight binary components are extracted via Burt MCA (Table 3) which producing 256 cells in each group, from the initial of 25 binary variables which should create 33,554,432 number of cells per group. However, only 53 cells (20.70%) in π_1 and 54 cells (21.09%) in π_2 are filled with objects in this case. This discovery showed low percentages of objects in the appropriate cells, causing the developed SLM+Burt model to perform somewhat poorer than the SLM+JCA. This is because most of the produced cells, 79.30% of π_1 and 78.91% of π_2, are empty, resulting in skewed estimated parameters and further affecting the performance of the constructed model.

Table 4 displays a result for the two constructed models, both of which showed zero misclassification rates for all binary sizes tested. This demonstrated that sample size plays an important role in model performance, which has improved the achievement of the constructed model. The SLM+JCA model still showed consistent results, where no more than seven binary variables were extracted.

TABLE 4. Performance analysis of the new constructed Smoothed Location Models tested on different binary sizes under n=180

SLM+Burt model	Size of Binary Variables				
	5	10	15	20	25
Misclassification Rate	0	0	0	0	0
Number of Binary Extracted (PC_b)	3	5	7	8	9
Number of Non-empty Cells (π_1, π_2)	(6, 6)	(28, 27)	(68, 68)	(74, 77)	(83, 81)
Number of Empty Cells (π_1, π_2)	(2, 2)	(4, 5)	(60, 60)	(182, 179)	(429, 431)
KL Distance	203.59	1556.76	147.94	27.75	6.87
SLM+JCA model	5	10	15	20	25
Misclassification Rate	0	0	0	0	0
Number of Binary Extracted (PC_b)	2	2	4	7	5
Number of Non-empty Cells (π_1, π_2)	(4, 4)	(4, 4)	(16, 16)	(71, 64)	(31, 29)
Number of Empty Cells (π_1, π_2)	(0, 0)	(0, 0)	(0, 0)	(57, 64)	(1, 3)
KL Distance	473.61	623.74	2457.74	124.61	2334.08

For all data situations, except when n=60 and b=25, the developed SLM+JCA model outperforms the SLM+Burt model. This is owing to the fact that when utilizing JCA, there are fewer extracted binaries and considerably greater spacing between the observed groups. The only reason why the SLM+Burt model performs much better than the SLM+JCA model for the case n=60 with b=25 is because Burt MCA creates a much larger distance between the groups. Thus, the SLM+Burt model has classified objects more precisely as its groups are more separated than the SLM+JCA model.

The overall results showed that the sample size had no effect on the SLM+Burt model as its performance was quite similar, although the samples have been increased from 60 to 180. On the contrary, the SLM+JCA model performed better when larger sample sizes were considered in the study. It can be observed that the performance of the SLM+JCA model is far superior when n=120 (Table 3) as well as when n=180 (Table 4) compared to n=60 (Table 1).

Classification methods can be applied to many fields of study to identify unique membership (Hamid, Zainon & Yong 2016; Okwonu et al. 2012). For instance, El Abbassi et al. (2021) applied a univariate classifier to nanoelectronics and spectroscopy to classify relevant information from the dataset. Classification methods also have been implemented to determine ICT knowledge awareness by Dávideková, Michal Greguš and Bureš (2019), while Jimoh, Abisoye and Uthman et al. (2022) used for classifying malaria infection.

Conclusion

In the majority of the underlying investigated conditions, the newly created SLM+JCA model outperformed the SLM+Burt model, according to the study's findings. SLM+JCA model does not misclassify objects in most circumstances except when n=60 and b=25. This is because the SLM+JCA model extracts a smaller number of binary variables than the SLM+Burt model, resulting in a greater distance (separation) between the groups, making this model performs better. This is the key reason that makes the former model works well. It is thus can be concluded that both of the constructed models are feasible and applicable for up to 25 number of the binary variables with the help of either Burt MCA or JCA. This discovery provides new insights for the location model methodology, which may fill the gap of previous studies by Vlachonikolis and Marriott (1982), Krzanowski (1983) as well as Hamid, Ngu and Alipiah (2018). They limit the studies to no more than six binary variables that can be included in the model, otherwise, the achievement of model is severely degraded.

Therefore, both SLM+Burt and SLM+JCA can be considered as efficient classification models due to they are able to perform well even with many empty cells and other investigated conditions. Overall, these two newly built SLMs are good alternative methods for classification and data reduction when encountering mixed variables with a large number of the binary.

Acknowledgements

The authors are grateful for the assistance they received from Universiti Utara Malaysia in order to complete this study. There was no specific grant for this research from any funding source in the non-profit sectors, private, or public.

References

Asparoukhov, O. & Krzanowski, W.J. 2000. Non-parametric smoothing of the location model in mixed variable discrimination. *Statistics and Computing* 10(4): 289-297.

Dávideková, M., Michal Greguš, M.L. & Bureš, V. 2019. Yet another classification of ICT in knowledge management initiatives: Synchronicity and interaction perspective. *Journal of Engineering and Applied Sciences* 14(Special Issue 9): 10549-10554.

El Abbassi, M., Overbeck, J., Braun, O., Calame, M., van der Zant, H.S. & Perrin, M.L. 2021. Benchmark and application of unsupervised classification approaches for univariate data. *Communications Physics* 4(1): 1-9.

Greenacre, M.J. 2007. *Correspondence Analysis in Practice* (2nd ed.). Boca Raton: Chapman & Hall.

Greenacre, M.J. & Blasius, J. 2006. *Multiple Correspondence Analysis and Related Methods*. London: Taylor and Francis Group.

Hamid, H. 2018. New location model based on automatic trimming and smoothing approaches. *Journal of Computational and Theoretical Nanoscience* 15(2): 493-499.

Hamid, H. 2014. Integrated smoothed location model and data reduction approaches for multi variables classification. PhD Dissertation, Universiti Utara Malaysia, Malaysia (Unpublished).

Hamid, H. 2010. A new approach for classifying large number of mixed variables. *International Journal: World Academy of Science, Engineering and Technology* 46: 156-161.

Hamid, H., Zainon, F. & Yong, T.P. 2016. Performance analysis: An integration of principal component analysis and linear discriminant analysis for a very large number of measured variables. *Research Journal of Applied Sciences* 11(11): 1422-1426.

Hamid, H., Ngu, P.A.H. & Alipiah, F.M. 2018. New smoothed location models integrated with PCA and two types of MCA for handling large number of mixed continuous and binary variables. *Pertanika Journal of Science & Technology* 26(1): 247-260.

Jimoh, R.G., Abisoye, O.A. & Uthman, M.M.B. 2022. Ensemble feed-forward neural network and support vector machine for prediction of multiclass malaria infection. *Journal of Information and Communication Technology* 21(1): 117-148.

Jolliffe, I.T. 1986. *Principal Component Analysis*. New York: Springer-Verlag.

Kaiser, H.F. 1961. A note on Guttmann's lower bound for the number of common factors. *British Journal of Mathematical and Statistical Psychology* 14: 1-2.

Kemsley, E.K. 1996. Discriminant analysis of high-dimensional data: A comparison of principal component analysis and partial least squares data reduction methods. *Chemometrics and Intelligent Systems* 33: 47-61.

Krzanowski, W.J. 1995. Selection of variables, and assessment of their performance, in mixed variable discriminant analysis. *Computational Statistics and Data Analysis* 19(4): 419-431.

Krzanowski, W.J. 1993. The location model for mixtures of categorical and continuous variables. *Journal of Classification* 10: 25-49.

Krzanowski, W.J. 1983. Stepwise location model choice in mixed-variable discrimination. *Applied Statistics* 32(3): 260-266.

Krzanowski, W.J. 1980. Mixtures of continuous and categorical variables in discriminant analysis. *Biometrics* 36: 493-499.

Massey, W.F. 1965. Principal components regression in exploratory statistical research. *Journal of American Statistical Association* 60: 234-246.

Nenadic, O. & Greenacre, M.J. 2007. Correspondence analysis in R, with two- and three-dimensional graphics: The ca Package. *Journal of Statistical Software* 20(3): 1-13.

Okwonu, F.Z., Dieng, H., Othman, A.R. & Ooi, S.H. 2012. Classification of aedes adults mosquitoes in two distinct groups based on fisher linear discriminant analysis and FZOARO techniques. *Mathematical Theory and Modeling* 2(6): 22-30.

Rencher, A.C. 2002. *Methods of Multivariate Analysis: Wiley Series in Probability and Statistics*. 2nd ed. New York: John Wiley & Sons, Inc.

Vlachonikolis, I.G. & Marriott, F.H.C. 1982. Discrimination with mixed binary and continuous data. *Applied Statistics* 31(1): 23-31.

*Corresponding author; email: hashibah@uum.edu.my