Journal of Mathematics and System Science 2 (2012) 399-403

Modification of Intensive Care Unit Data Using Analytical Hierarchy Process and Fuzzy C-Means Model

Mohd Saifullah Rusiman[1], Efendi Nasibov[2], Kavikumar Jacob[1] and Robiah Adnan[3]

1. Faculty of Science,Technology and Human Development, Universiti Tun Hussein Onn Malaysia, Batu Pahat 86400, Malaysia

2. Faculty of Science, Dokuz Eylul University, Buca 35678, Turkey

3. Faculty of Science, Universiti Teknologi Malaysia, Skudai 81300, Malaysia

Received: May 13, 2012 / Accepted: June 12, 2012 / Published: July 25, 2012.

Abstract: This paper proposes a proper methodology in data modification by using AHP (analytical hierarchy process) technique and FCM (fuzzy c-mean) model in the ICU (intensive care unit). The binary data were created from continuous data using FCM model, while the continuous data were constructed from binary data using AHP technique. The models used in this study are FCRM (fuzzy c-regression model). A case study in scale of health at the ICU ward using the AHP, FCM model and FCRM models was conducted. There are six independent variables in this study. There are four cases which are considered as the result of using AHP technique and FCM model against independent data. After comparing the four cases, it was found that case 4 appeared to be the best model, because it has the lowest MSE (mean square error) value. The original data have the MSE value of 97.33, while the data in case 4 have the MSE value of 82.75. This means that the use of AHP technique can reduce the MSE value, while the use of FCM model can not reduce the MSE value. In other words, it can be proved that the AHP technique can increase the accuracy of prediction in modeling scale of health which is associated with the mortality rate in the ICU.

Key words: Analytical hierarchy process, fuzzy c-means model, fuzzy c-regression models, mean square error.

1. Introduction

Nowadays, fuzzy modeling has become popular since it describes complex systems better. It has been broadly used in economics, computer sciences, engineering, social sciences and other fields. The FCM (fuzzy c-mean) model introduced by Bezdek [1] in 1981 creates hyper-spherical-shaped clusters. In contrast, the FCRM (fuzzy c-regression models) proposed by Hathaway and Bezdek [2] in 1993, create hyper-plane-shaped clusters [1-4]. AHP (analytical hierarchy process) has been introduced by Saaty [5] in 1977 in dealing with the factor weights due to short of information for relevant variables. It has been widely used in decision making, because it includes natural feelings of human nature. Many researchers use AHP technique in dealing with data mining problem [5-6].

The ICU (intensive care unit) contributes a vital function in the medical care sector not only for the seriously ill, who represents 5% of inpatients, but also makes a significant contribution of health care funds. The United States hospital health care contributes 15%-20% to the total hospital cost. In 1968, the first ICU in Malaysia was established. Intensive care has then expanded rapidly and is now available in all tertiary care hospitals and selected secondary care hospitals. The National Audit on Adult Intensive Care Units in Malaysia in 2002 is modeled on the UK experience in 1994 and coordinated by a national committee consisting of senior intensive care specialists in the Ministry of Health. This audit unit develops a national database to achieve fundamental aspects of intensive care functions within a hospital.

Corresponding author: Mohd Saifullah Rusiman, Ph.D., lecturer, research field: fuzzy statistics. E-mail: saifulah@uthm.edu.my.

The clinical indicators developed by ACHS (The Australian Council on Healthcare Standard) are useful tools for clinicians to indicate potential problems and areas for improvement [7].

Presently, the common method used in the ICU involves the use of logistic regression such as done by Colpan et al. [2] and others. Only Pilz and Engelmann [8] studied a basic fuzzy rule to make the medical decision in ICU. This work inspires the authors to study fuzzy model into ICU area that could give a challenge to this study. The first research on mortality rates in Malaysian ICU has been done at a general hospital in Ipoh, involving only a logit model [7]. The second research is continued by Rusiman et al. [9] on the analysis of logit, probit and linear probability models. Upon comparing the three models, the logit model has emerged to be the best model.

The main objective of this study is to explore data modification using AHP technique and FCM model in the scale of health at the ICU using FCRM models. The other objective is to compare the beginning data (without any method), the use of AHP technique, the use of FCM model or any combination of methods which are applied to data in order to find the best FCRM models. So, the authors can make better suggestions based on this data mining method in predicting the scale of health in the ICU.

2. Material

The data were obtained from the ICU of a general hospital in Johor. The data obtained were classified as a cluster sampling. It involves 1,311 patients in the ICU within the interval of January, 2001 to August, 2002. The dependent variable is scales of health or score of SAPS II (simplified acute physiology score) discharge from hospital (s2sdisc). Scales of health variable is associated with patient status (alive or dead) since the correlation between these two variables is so strong with the pearson correlation value of 0.87. There are six independent variables considered in this study which are sex, race, organ failure (orgfail), comorbid diseases (comorbid), mechanical ventilator (mecvent) and score of SAPS II admit (s2sadm). The s2sdisc and s2sadm scores are 15 accumulated values for heart rate, blood pressure, age, body temperature, oxygen pressure, urine result, urea serum level, white blood count, potassium serum level, sodium serum level, bicarbonate serum level, bilirubin level, Glasgow coma score, chronic illness and type of admittance as introduced by Le Gall et al. [10].

3. Methods

3.1 AHP Technique

The AHP technique is a complete decision making process that allows more complete consideration of multi-factors/criteria. The AHP procedure involves three steps as:

Step 1: Create the decision hierarchy

The decision maker must identify the overall decision, the factors that must be weighted or used to make the decision and the alternative choices from which a decision will be made. Once these are identified they are placed in a decision hierarchy;

Step 2: Calculate the weighted of alternatives

The decision maker or expert must compare each alternative with all other alternatives at one time. The rating measure scale used to evaluate the alternatives is from the range of 1 to 9 as it relate to each factor;

Step 3: Calculate the weighted of factors

The decision maker uses the previously determined comparison ratings to compute a set of priorities for the individual factors. This involves several small computation sub-steps where the probabilities or weighted obtained from a paired comparison matrix have a total probability of one [11].

3.2 FCM Model

Based on the Dunn [4] and Bezdek [1] algorithm in FCM clustering, the authors have to minimise the function of F in Eq. (1),

$$F = \sum_{q=1}^{c}\sum_{p=1}^{N} u_{pq}^{w} d_{pq}^{2}, \; w > 1 \tag{1}$$

where u_{pq} is the membership values, d_{pq}^2 is $\left\| x_p - \sum_{p=1}^{N} u_{pq}^w x_p / \sum_{p=1}^{N} u_{pq}^w \right\|^2$ or the Euclidean distances, N is the number of data, c is the number of clusters and w is the weight or fuzzifier. In order to minimize function of F in Eq. (1), the authors have to:

(a) Determine the value of clusters, c, membership values, U or u_{pq} and the termination tolerance, δ where $\delta > 0$;

(b) For each cluster, update Euclidean distances, d_{pq} for given U by computing the weighted averages;

(c) Update membership values as in Eq. (2):

$$u_{pq} = \frac{1}{\sum_{r=1}^{C} \left(\frac{d_{pq}}{d_{pr}} \right)^{\frac{2}{w-1}}}, \quad \text{for } w > 1$$

$$u_{pq} = \begin{cases} 1 & \text{if } d_{pq} = \min(d_{pr}) \\ 0 & \text{otherwise} \end{cases} \quad \text{for } w = 1 \tag{2}$$

(d) Compute the function of F in Eq. (1) and check for convergence. If $|F_{old} - F_{new}| < \delta$ stop the iteration, else go to Step (b).

3.3 FCRM Models

Based on the FCRM algorithm in Hathaway & Bezdek [2], Wolkenhauer [12] and Kung & Su [13], the authors have to:

(a) Determine the number of clusters c, $c \geq 2$, the termination tolerance δ, $\delta > 0$ and the weight w, $w > 1$ ($w = 2$ is used as a common choice in practice) and the initial value for membership function matrix, $\mathbf{U}^{(0)}$ satisfying Eq. (4);

(b) Estimate $\theta_1, \ldots, \theta_c$ simultaneously by modifying the FCM algorithm. If the regression functions $f_p(x;\theta_p)$ are linear in the parameters θ_p, the parameters can be computed as a solution of the weighted least squares as in Eq. (3) where y:

$$\theta_p = [\mathbf{X}_b^T \mathbf{W}_p \mathbf{X}_b]^{-1} \mathbf{X}_b^T \mathbf{W}_p \mathbf{Y} \tag{3}$$

(c) Compute the objective function:

$$E_w[\mathrm{U}, \{\theta_p\}] = \sum_{p=1}^{c} \sum_{q=1}^{N} u_{pq}^w E_{pq}[\theta_p] \tag{4}$$

where

(i) u_{pq} is membership degree (p = 1, …, c , q = 1, …, N);

(ii) $E_{pq}[\theta_p]$ is the measure of error with $E_{pq}[\theta_p] = \| Y_q - f_p(X_q;\theta_p) \|^2$. The most commonly used is the squared vector Euclidean norm for $Y_q - f_p(X_q;\theta_p)$;

(d) In order to minimize the objective function in Eq. (4), do iterations for m = 1, 2, …, ∞ until $\| \mathbf{U}^{(m)} - \mathbf{U}^{(m-1)} \| < \delta$ as the steps below:

Step 1: Compute the model parameters $\theta_p^{(m)}$ to globally minimize Eq. (4);

Step 2: Update U with $E_{pq} = E_{pq}[\theta_p^{(m-1)}]$, to satisfy:

$$u_{pq}^{(m)} = \begin{cases} \dfrac{1}{\sum_{r=1}^{c} \left(\frac{E_{pq}}{E_{qr}} \right)^{\frac{2}{w-1}}} & , \text{for } I_q = \phi \\ 0 & , \text{ for } I_q \neq \phi, p \notin I_q \end{cases} \tag{5}$$

where $I_q = \{ p \mid 1 \leq p \leq c \text{ and } E_{pq} = 0 \}$ until $\| \mathbf{U}^{(m)} - \mathbf{U}^{(m-1)} \| < \delta$.

The MSE value is used as follow:

$$MSE = \frac{1}{N} \sum \left(Y_p - \hat{Y}_p \right)^2 \tag{6}$$

where Y_p denotes the real data, $\hat{y}$ represents the predicted value of Y_p and N is the number of data.

4. Results and Discussion

4.1 AHP Technique

The AHP technique is used to organ failures variable (orgfail) where this independent variable has only two binary data, that is, patients who have or do not have organ failures. This technique will fuzzify the binary data of organ failures into a continuous data within the interval [0, 1]. Organ failures are divided into 6 types which are respiratory failure (M), cardiovascular failure (N), neurological failure (O), renal failure (P), hepatic failure (Q) and haematological failure (R). Based on the specialists' observation in the general hospital, they claim that N and P types have a twice higher probability of contributing to high mortality rate if compared to the

Table 1 The paired-comparison matrix and weighted for organ failures.

	N	P	M	R	O	Q	Total	Weighted
N	1	1	2	2	4	4	14	0.2857
P	1	1	2	2	4	4	14	0.2857
M	½	½	1	1	2	2	7	0.1429
R	½	½	1	1	2	2	7	0.1429
O	¼	¼	½	½	1	1	3.5	0.0714
Q	¼	¼	½	½	1	1	3.5	0.0714

M and R types. In fact, the M and R types have a twice higher probability if compared to the O and Q types. However, N and P types have the same weightage. The same weightage are also given to the M and R types. O and Q types also receive the same weightage. Table 1 shows the paired comparison matrix and probabilities (weighted) for organ failures.

4.2 FCM Model

In this study, to categorize s2sadm data with "1" and "2" coded, the authors have to cluster s2sadm data based on FCM clustering algorithm. The data for cluster 1 with 860 data ranges from 0 to 43 whereas the data for cluster 2 with 443 data ranges from 44 to 126. This is the same as the cluster given by the doctors who indicated that the s2sadm score over 43 is classified as a bad condition. Graph of the membership function for s2sadm variable is shown in Fig. 1.

4.3 FCRM models

There are four cases considered as shown in Table 2 as a result of using AHP and FCM model against independent data. The four cases involve six variables with different combination of variable types in each case were considered in order to find the best model using FCRM models. The variables involved are sex (x_1 is binary), race (x_2 is category), orgfail (x_3 is binary or continuous), comorbid (x_4 is binary), mecvent (x_5 is binary) and s2sadm (x_6 is binary or continuous). Case 3 is the beginning data without any modification being carried out.

In order to find the optimal solution, four cases are considered as a result of combination cases with/ without using AHP technique and/or FCM model

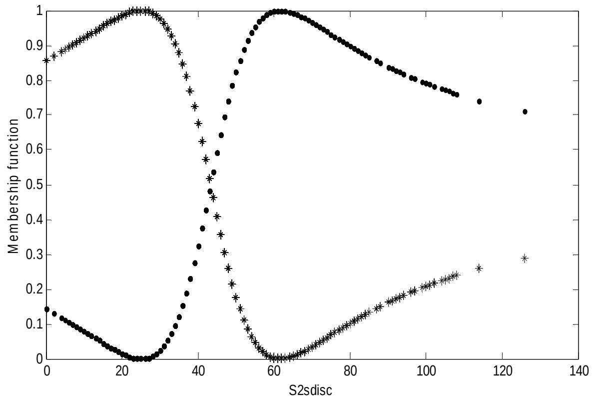

Fig. 1 Plot for s2sadm membership function (FCM clustering).

against independent data. Table 2 shows that case 4 is the best case with the lowest MSE (mean squared error) value, that is, when x_1 is binary, x_2 is category, x_3 is continuous, x_4 is binary, x_5 is binary and x_6 is continuous. The MSE value for FCRM models for case 4 is 83.55 (all variables) and 82.75 (significant variables—x_1, x_3, x_4, x_6). The MSE value for significant variables shows better result. The MSE value for case 3 (original data) is 97.33, while the MSE value for case 4 is 82.75. The MSE values for the other cases are 97.29 and 121.92. In conclusion, case 4 is the best case where data modification involves only the orgfail variable. These selected models (y vs. x_1, x_3, x_4, x_6) are represented in Eq. (7) with two clusters.

5. Conclusions

Data modifications have been made to the case study in the ICU where the binary data (s2sadm variable) were constructed from continuous data using FCM model, whereas the continuous data (orgfail variable) were created from binary data using AHP technique. There are four cases considered to find the optimal solution for FCRM models. After comparing the four cases, it was found that case 4 appeared to be the best model, having the lowest MSE value of 82.75, while the original data have the MSE value of 97.33. This means that the use of AHP technique can reduce the MSE value, while the use of FCM model can not reduce the MSE value in modeling scale of health in

Table 2 Different cases of multivariate data (*y* vs. x_1, x_2, x_3, x_4, x_5, x_6).

Case	1	2	3	4
x_1	B	B	B	B
x_2	Ca	Ca	Ca	Ca
x_3	B	Co	B	Co
x_4	B	B	B	B
x_5	B	B	B	B
x_6	B	B	Co	Co
MSE for MLR(SV)	632.14 (VA)	526.40 (VA)	498.29 (VB)	462.21 (VA)
MSE for FCRM(AV)	116.05	114.71	98.28	83.55
MSE for FCRM(SV)	121.92 (VA)	97.29 (VA)	97.33 (VB)	82.75 (VA)

B: Binary; Ca: Category, Co: Continuous; AV:All variables; SV: Significant variables (VA, VB); MLR: multiple linear regression
VA: 4 Variables chosen— x_1, x_3, x_4 & x_6; VB: 5 Variables chosen — x_1, x_3, x_4 & x_6.

$$\begin{aligned}
&\text{Cluster 1:}\quad R^1: \text{ IF } x_1 \text{ is } A_1^1 \text{ and } x_3 \text{ is } A_3^1 \text{ and } x_4 \text{ is } A_4^1 \text{ and } x_6 \text{ is } A_6^1\\
&\qquad\qquad \text{THEN } y^1 = 2.3332\ x_1 + 12.5753\ x_3 + 4.5334\ x_4 + 0.1223\ x_6 + 59.4365\\
&\text{Cluster 2:}\quad R^2: \text{ IF } x_1 \text{ is } A_1^2 \text{ and } x_3 \text{ is } A_3^2 \text{ and } x_4 \text{ is } A_4^2 \text{ and } x_6 \text{ is } A_6^2\\
&\qquad\qquad \text{THEN } y^2 = 1.1143\ x_1 - 1.5543\ x_3 + 3.8553\ x_4 + 0.4233\ x_6 - 4.5569
\end{aligned} \tag{7}$$

the ICU. In other words, it can be announced that the AHP technique can improve the accuracy of modeling prediction and should be used as a reference in hospitals to improve data accuracy in modeling scale of health in the ICU which is associated with the mortality rate. Therefore, the mortality rate or scale of health can be monitored better based on the six independent variables.

Acknowledgments

This research work is supported by the FRGS (fundamental research grant scheme) grant (Vot. No. 0821), Ministry of Higher Education, Malaysia.

References

[1] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Kluwer Academic Publishers, USA, 1981.

[2] R.J. Hathaway, J.C. Bezdek, Switching regression models and fuzzy clustering, IEEE Transactions on Fuzzy Systems 1 (3) (1993) 195-204.

[3] A. Colpan, E. Akinci, A. Erbay, N. Balaban, H. Bodur, Evaluation of risk factors for mortality in ICUs: A prospective study from a referral hospital in turkey, American Journal of Infection Control 33 (1) (2005) 42-47.

[4] J.C. Dunn, A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters, Journal of Cybernetics 3 (1973) 32-57.

[5] T.L. Saaty, A scaling method for priorities in hierarchical structures, Journal of Mathematical Psychology 15 (3) (1977) 234-281.

[6] T.L. Saaty, K. Peniwati & J. S. Shang, The analytic hierarchy process and human resource allocation: Half the story. Mathematical and Computer Modelling, 46 (2007) 1041–1053.

[7] The Committee for National Audit on Adult Intensive Care Units, Protocol : National audit on adult intensive care units. Malaysia: Ministry of Health Malaysia, 2002.

[8] U. Pilz, L. Engelmann, Integration of medical knowledge in an expert system for use in intensive care medicine, Fuzzy and Neuro-Fuzzy Systems in Medicine 2 (1998) 290-315.

[9] M.S. Rusiman, Z. Mohd Daud and I. Mohamad, The comparison between logit, probit and linear probability model toward mortality rate at ICU general hospital, Jurnal Statistika Universitas Islam Bandung 4 (2) (2004) 129-138.

[10] J-R. Le. Gall, S. Lemeshow, F. Saulnier, A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study, The Journal of the American Medical Association 270 (24) (1993) 2957-2963.

[11] M.J. Schniederjans, Information Technology Investment: Decision-Making Methodology, World Scientific Publishing Company, Singapore, 2004.

[12] O. Wolkenhauer. Fuzzy regression models, fuzzy clustering and switching regression [Online], May 3, 2008, http://www.sbi.uni-rostock.de/uploads /tx_templavoila/SBI_Materials_Fuzzy-Regression.pdf.

[13] C.C. Kung, J.Y. Su, Affine Takagi-Sugeno fuzzy modelling algorithm by fuzzy c-regression models clustering with a novel cluster validity criterion, IET Control Theory Application 1 (5) (2007) 1255-1266.