

Hierarchical Genetic Algorithm and Fuzzy Radial Basis Function Networks for Factors Influencing Hospital Length of Stay Outliers

Ahmed Belderrar, MSc, Abdeldjebar Hazzab, PhD

Laboratory of Control, Analysis and Optimization of Electro-Energetic Systems, University Tahri Mohamed Bechar, Bechar, Algeria

Objectives: Controlling hospital high length of stay outliers can provide significant benefits to hospital management resources and lead to cost reduction. The strongest predictive factors influencing high length of stay outliers should be identified to build a high-performance prediction model for hospital outliers. **Methods:** We highlight the application of the hierarchical genetic algorithm to provide the main predictive factors and to define the optimal structure of the prediction model fuzzy radial basis function neural network. To establish the prediction model, we used a data set of 26,897 admissions from five different intensive care units with discharges between 2001 and 2012. We selected and analyzed the high length of stay outliers using the trimming method geometric mean plus two standard deviations. A total of 28 predictive factors were extracted from the collected data set and investigated. **Results:** High length of stay outliers comprised 5.07% of the collected data set. The results indicate that the prediction model can provide effective forecasting. We found 10 common predictive factors within the studied intensive care units. The obtained main predictive factors include patient demographic characteristics, hospital characteristics, medical events, and comorbidities. **Conclusions:** The main initial predictive factors available at the time of admission are useful in evaluating high length of stay outliers. The proposed approach can provide a practical tool for healthcare providers, and its application can be extended to other hospital predictions, such as readmissions and cost.

Keywords: Data Mining, Intensive Care Units, Length of Stay, Machine Learning, Medical Informatics

Submitted: January 16, 2017

Revised: 1st, April 30, 2017; 2nd, May 13, 2017; 3rd, June 10, 2017

Accepted: July 2, 2017

Corresponding Author

Ahmed Belderrar, MSc

Laboratory of Control, Analysis and Optimization of Electro-Energetic Systems, University Tahri Mohamed Bechar, BP 417, Bechar 08000, Algeria. Tel: +213-663-086848, E-mail: belahd@hotmail.com

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

© 2017 The Korean Society of Medical Informatics

1. Introduction

Nowadays, critical intensive care units (CICU) are often dangerously overcrowded, and the acceptance of patients is becoming an increasingly significant public health system problem [1]. Waiting time [2,3], limited resources [1,4], and hospital costs [5-8], are difficulties facing hospital policy-makers and healthcare authorities. Addressing hospital high length of stay outliers (HHSOs) is a major task that continues to preoccupy healthcare providers while maintaining and improving the quality of healthcare. One solution is to have a better allocation of resources by reducing [9] hospital stays and expanding patient acceptance rates. More specifically, reducing HHSOs can have a significant impact on improv-

ing flow and reducing in-patient services. However, finding a way to predict HHSOs earlier could be advantageous for managers in that it could help control hospital costs, could support optimal admissions scheduling, staff planning, and management of resources [10].

To this end, both statistical analysis methods and machine learning have been investigated. Typically, HHSO analysis methods follow a statistical, deviation-based, and distance-based approach [11]. Machine learning techniques, have been proposed to build prediction models for length of stay, and they are applied in some care services. These techniques include logistic regression [12], artificial neural network [12-15], decision tree [12,15], ensemble model [12], and support vector machines [15].

Various predictive factors have been considered in statistics and data mining. These factors include age [5,12,16-18], gender [5,16,17,19], marital status [19], ethnicity [19], ambulance use [16], admission type [5,19], admission status [16], department type [5], discharging reason [5], length of stay [16], comorbidity [16,19], in-hospital mortality [16], diagnosis [19], and costs [19].

Up to now, there is no universal agreement about the predictive factors. To this end, we propose a novel approach to identify the main predictive factors influencing HHSOs using data collected from various types of CICUs. We also designed a high-performance hybrid prediction model (HPM).

II. Case Description

1. Data Collection

Data was extracted from MMIC III [20,21]. The following five types of CICUs were investigated: neonatal intensive care units (NICUs), medical intensive care units (MICUs), coronary care units (CCUs), cardiac surgery recovery units (CSRUs), and surgical intensive care units (SICUs). We excluded admissions with missing predictive factors and patients older

than 89 years because their ages are replaced by 300 years. We obtained HHSO data using the geometric mean plus two standard deviations [22,23]. The extracted predictive factors studied in this work are summarized in Table 1.

We encoded the input binary data following a $-1/+1$ scheme, and for the input categorical data, we followed 1-of-C dummy-coding. We rescaled the input data using both z-score standardization, in which inputs have a mean of 0 and a standard deviation of 1, and min-max normalization, in which inputs are in the range of $[-1,1]$. We did not encode or rescale the HHSOs. We split the data into two sets: 80% of admissions for the training phase and 20% for the testing phase.

2. HPM Design

The proposed HPM is based on the hierarchical genetic algorithm (HGA) and fuzzy radial basis function networks (FRBFN). The purpose of the HGA is to provide the main predictive factors and the structure of the FRBFN. The purpose of the FRBFN is to accurately predict HHSOs of new admissions.

1) HGA chromosome representation

Our chromosome is composed of two parametric genes and two control genes. The first parametric gene is a vector of seven bits defining the optimal structure of the FRBFN. The second parametric gene is a vector of 28 bits identifying the main predictive factors. When a bit 1 is indicated in the control gene, the corresponding parametric gene is activated. A chromosome is defined by $\{f, V, k, C, \Sigma, W\}$, where F is the fitness function of the HGA, V is the set of predictive factors, and the set $\{k, C, \Sigma, W\}$ defines the parameters of the FRBFN, where k is the number of hidden neurons, C and Σ are respectively the set of centers and widths of the transfer Gaussian functions, and W is the set of weights.

Table 1. Overview of the predictive factors used to predict hospital high length of stay outliers

Group	Predictive factors
Demographic characteristics	Age, sex, language, marital status, ethnicity, insurance, religion, height, weight
Hospital characteristics	Admission type, the first location within the hospital prior to CICU, the first CICU, The next CICU in the transfer process, the first and the next service that will be provided to the patient the first and the next ward in which the patient will stay, The discharge location
Potential medical factors	Laboratory events, microbiology events, medication-related order entries, input CareVue, input MetaVision, output fluids, chart events, current procedural terminology for patient as billed through the CICU or the respiratory cost center, diagnosis, comorbidities

CICU: critical intensive care units.

2) HPM algorithm

Our algorithm takes as input the training set and its HHSOs, and it returns an optimal solution (chromosome). First, a population of many chromosomes is generated, where V and k are generated randomly. Next, in each evolutionary cycle, a new population of potential solutions is generated based on the application of the genetic operations. In each cycle, the parameters of each produced chromosome are defined as follows: V and k are defined by the application of crossover and mutation on the two control genes separately. The set C is determined using the fuzzy C-means algorithm [24]. The set Σ is defined using the fuzzy K-nearest neighbors algorithm. The set W is defined by the singular value decomposition algorithm. Next, f is evaluated in terms of the mean absolute error (MAE) as shown in Equation (1). The cycle stops when an optimal solution is good enough or the maximum number of generations is expected.

$$f = MAE = \frac{1}{N} \sum_{i=1}^N |Predicted_i^{HHSO} - Desired_i^{HHSO}| \quad (1)$$

A predicted HHSO is calculated by Equations (2) and (3), where $\omega_k \in W$, β_0 is the bias, $c_k \in C$, $\sigma_k \in \Sigma$, and φ_k is the k^{th} Gaussian function:

$$HHSO = \beta_0 + \sum_{k=1}^l \omega_k \times \varphi_k(admission_i), \quad (2)$$

$$\varphi_k(admission_i) = e^{-\frac{1}{2} \frac{\|admission_i - c_k\|^2}{\sigma_k^2}}. \quad (3)$$

3. Performance Evaluation

We evaluated the performance of the FRBFN in terms of the mean magnitude relative error MMRE and Pred(q) [25] prediction at level q as shown by Equations (4) and (5). Conte et al. [26] maintained that a good estimation model should have $MMRE \leq 0.25$ and $Pred(q) \geq 0.75$. The parameter m denotes the number of computed HHSOs for which the MRE is less than or equal q :

$$MMRE = \frac{1}{N} \sum_{i=1}^N MRE_i = \frac{1}{N} \sum_{i=1}^N \left| \frac{Desired_i^{HHSO} - Predicted_i^{HHSO}}{Desired_i^{HHSO}} \right|, \quad (4)$$

$$Pred(q) = \frac{m}{N}. \quad (5)$$

4. Results

Among the 26,897 admissions considered in this study (Table 2), we found 1,365 (5.07%) patients with HHSO; 44.69% of them were hospitalized in an MICU. The mean age of the patients was 59.96 ± 0.43 years (range, 0–89 years) with most subjects between 40–64 years, while 57.58% of them were men and 42.42% were women. Most of the discharges (62.93%) required other healthcare facilities. Also, 62.86% of the patients were covered by Medicare or Medicaid. In addition, a notable 88.27% of patients showed multiple comorbidities. Table 3 summarizes the statistical results of common factors and some other important factors.

The population was composed of 150 chromosomes, and the number of iterations was 15,000. The predicted outputs were rounded to the zero decimal place because an HHSO is calculated based on the number of days. We used the technique in which the closest respected output is selected as the predicted output. The results shown in Table 4 indicate that the FRBFN performed better for all CICUs. The outcomes obtained by the z-score standardization were more effective and efficient in terms of MRME and Pred(0.25) than those obtained by min-max normalization.

Table 5 summarizes the numbers of predictive factors and the numbers of neurons. There are three key observations to be made. First, the structure of the prediction model differed according to the studied CICUs. Second, the results suggest that each CICU has its own main predictive factors. Third, the obtained optimal configurations depend on the scaling data. The predictive factors found in the two scaling techniques were almost the same. We note that we could not verify any relationship between the FRBFN structure and the main predictive factors.

Table 6 summarizes the common main predictive factors

Table 2. Distribution of hospital high length of stay outliers

CICU	All admissions	Outliers	Min	Max	Mean
NICU	148	9	62	169	103.78
MICU	12,308	610	29	169	44.71
CCU	4,336	232	25	124	37.91
CSRU	4,933	241	25	172	37.13
SICU	5,172	273	31	133	47.78

CICU: critical intensive care units, NICU: neonatal intensive care units, MICU: medical intensive care units, CCU: coronary care units, CSRU: cardiac surgery recovery units, SICU: surgical intensive care units.

Table 3. Statistical result (in % except mean) of common factors and some other important factors

Factor	Distribution	NICU	MICU	CCU	CSRU	SICU
Age ^a (yr)	Mean	13.89 ± 7.30	57.86 ± 0.63	64.10 ± 0.94	65.17 ± 0.97	58.54 ± 0.89
	<2	100	-	-	-	-
	2–19	-	0.49	-	0.41	-
	20–39	-	13.44	6.90	6.22	7.69
	40–64	-	51.97	41.81	39.83	58.61
	65–79	-	26.39	34.91	37.76	28.21
	80+	-	7.70	16.38	15.77	5.49
Sex	Male	44.44	55.25	60.34	56.43	61.90
	Female	55.56	44.75	39.66	43.57	38.10
Marital status	Married	-	47.87	49.57	51.45	48.72
	Single	100	35.57	28.02	21.16	33.33
	Other	-	16.56	22.41	27.39	17.95
Insurance	Private	55.56	36.72	27.16	24.07	37.73
	Government	11.11	2.62	3.02	4.56	4.40
	Medicaid/Medicare	33.33	60.00	69.40	71.37	57.14
	Self-pay	-	0.66	0.43	-	0.73
Ethnicity	White	22.22	69.67	75.43	72.61	75.82
	Latino	-	4.43	5.17	5.81	5.86
	Black	33.33	13.44	8.62	8.30	9.52
	Asian	11.11	3.44	3.02	3.32	3.30
	Arabic	-	0.33	0.43	-	0.37
	Other	33.33	8.69	7.33	9.96	5.13
Admission type	Elective	-	5.08	5.17	17.01	12.82
	Emergency	44.44	94.92	94.83	82.99	87.18
	Other	55.56	-	-	-	0.00
Care service	1	100	44.59	34.48	56.43	64.84
	2+	-	55.41	65.52	43.57	35.16
Chart events	Yes	100	99.51	99.14	100	100
	No	-	0.49	0.86	-	-
Lab events	Yes	100	99.67	99.14	100	100
	No	-	0.33	0.86	-	-
Microbiology	Yes	55.56	91.31	85.78	89.63	92.31
	No	44.44	8.69	14.22	10.37	7.69
Input MetaVision	Yes	100	30.66	47.84	53.11	46.52
	No	-	69.34	52.16	46.89	53.48
Discharge location ^c	Home	88.89	23.93	21.55	14.11	23.08
	Death	-	20.00	15.52	9.54	8.79
	Other	11.11	56.07	62.93	76.35	68.13
Comorbidity	0	88.89	2.30	3.45	4.56	2.56
	1	-	7.54	6.03	7.88	12.09
	2+	11.11	90.16	90.52	87.55	85.35

NICU: neonatal intensive care units, MICU: medical intensive care units, CCU: coronary care units, CSRU: cardiac surgery recovery units, SICU: surgical intensive care units.

^aExcept NICU (day).

between the studied CICUs. Medical comorbidities were not identified in HHSO prediction of NICUs.

III. Discussion

HHSOs are closely related to hospital costs and should be controlled and justified. We examined the predictive factors associated with HHSOs using both the HGA and FRBFN. We have presented the necessary steps to implement the HPM and have illustrated its applicability using data collected from various CICUs. We obtained the main factors of the prediction model FRBFN and its optimal structure using the HGA.

A comprehensive and accurate determination of the underlying factors associated with HHSOs is a critical requirement as this can help to obtain a lighter and more efficient prediction model. The optimal configurations obtained for HPM are related to the studied data; therefore, the scalability of our approach to other hospital predictions, such as re-admissions or costs may be problematic, and it would require the adoption of pruning heuristics, determination of the effective optimal parameters of FRBFN, or merging with a statistical analysis. Therefore, we cannot claim that similar results can be obtained with other hospital predictions, but it would be an interesting area for future research.

The HHSO distribution is highly skewed as shown in Table 2. This has inspired us to delve deeper into investigating for

Table 4. Performance of the hybrid prediction model in terms of MMRE and Pred(q)

CICU	z-score standardization		Min-max normalization	
	MMRE (%)	Pred(q) (%)	MMRE (%)	Pred(q) (%)
NICU	0	100	0	100
MICU	2.13	94.69	3.78	92.77
CCU	6.31	90.23	6.33	89.78
CSRU	3.26	92.33	4.25	91.02
SICU	4.98	91.30	5.04	90.73

CICU: critical intensive care units, NICU: neonatal intensive care units, MICU: medical intensive care units, CCU: coronary care units, CSRU: cardiac surgery recovery units, SICU: surgical intensive care units.

Table 5. Configuration of the proposed hybrid prediction model

CICU	z-score standardization		Min-max normalization	
	Factors	Neurons	Factors	Neurons
NICU	9	8	9	8
MICU	16	51	16	68
CCU	19	32	18	41
CSRU	15	45	16	37
SICU	16	28	17	52

CICU: critical intensive care units, NICU: neonatal intensive care units, MICU: medical intensive care units, CCU: coronary care units, CSRU: cardiac surgery recovery units, SICU: surgical intensive care units.

Table 6. Overview of the common main predictive factors

Factor	Distribution
Age	0–89 years
Insurance	Private, Government, Medicaid, self-pay
Ethnicity	White, Latino, Black, Multi, Arabic, Asian, and other
Admission type	Elective, emergency, and urgent
Current service	CMED, CSURG, GU, MED, OMED, PSURG, SURG, VSURG, NMED, NSURG, ORTHO, TSURG
Discharge location	Home, left against medical adv., hospice medical facility, long term care hospital, short term hospital, discharge/transfer, dead/expired, rehab./distinct part hospital, SNF
Has chart events	Yes, no
Has lab events	Yes, no
Microbiology events	Yes, no
Comorbidities	Congestive heart failure, cardiac arrhythmias, valvular heart disease, pulmonary circulation, Peripheral vascular, hypertension, paralysis, other neurological, chronic pulmonary, diabetes uncomplicated, diabetes complicated, hypothyroidism, renal failure, liver disease, peptic ulcer, aids, lymphoma, metastatic cancer, solid tumor, rheumatoid arthritis, coagulopathy, obesity, weight loss, fluid electrolyte, blood loss anemia, deficiency anemias, alcohol abuse, drug abuse, psychoses, and depression

which fraction of the admissions, our hybrid model can predict HHSOs with error margins that are reasonably bounded.

Regarding the race factor in MICUs, the admissions rate of White and Asian races represents 69.67% (average of 44.67 days) and 3.44% (average of 47.1 days) respectively. The hospital stays of most Asian patients were longer than those of White patients. This information cannot be interpreted by our approach; therefore, other potential analysis should be investigated.

To improve the performance of HPM, the following suggestions should be considered: (1) Is there a relationship between the factors and the HPM structure? (2) Explore the sample to identify other factors. (3) The performance of HGA depends on the genetic operators that should be carefully selected. (4) To speed up the HPM, the first control genes can be used without corresponding parametric genes.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgments

We wish to thank the group of general practitioners GGP U.M. for their contributions to discussions about the predictive factors used in this study.

References

1. Di Somma S, Paladino L, Vaughan L, Lalle I, Magrini L, Magnanti M. Overcrowding in emergency department: an international issue. *Intern Emerg Med* 2015;10(2):171-5.
2. Siciliani L, Moran V, Borowitz M. Measuring and comparing health care waiting times in OECD countries. *Health Policy* 2014;118(3):292-303.
3. Barua B, Esmail N, Jackson T. The effect of wait times on mortality in Canada. Vancouver: Fraser Institute; 2014.
4. Richards JR, van der Linden MC, Derlet RW. Providing care in emergency department hallways: demands, dangers, and deaths. *Adv Emerg Med* 2014;2014:495219.
5. Cyganska M. The impact factors on the hospital high length of stay outliers. *Procedia Econ Financ* 2016;39:251-5.
6. Ithman MH, Goplarkrishna G, Beck NC, Das J, Petroski G. Predictors of length of stay in an acute psychiatric hospital. *J Biosaf Health Educ* 2014;2(2):1000119.
7. Pakzad H, Thevendran G, Penner MJ, Qian H, Younger A. Factors associated with longer length of hospital stay after primary elective ankle surgery for end-stage ankle arthritis. *J Bone Joint Surg Am* 2014;96(1):32-9.
8. Gruenberg DA, Shelton W, Rose SL, Rutter AE, Socaris S, McGee G. Factors influencing length of stay in the intensive care unit. *Am J Crit Care* 2006;15(5):502-9.
9. Clarke A, Rosen R. Length of stay. How short should hospital care be? *Eur J Public Health* 2001;11(2):166-70.
10. Mak G, Grant WD, McKenzie JC, McCabe JB. Physicians' ability to predict hospital length of stay for patients admitted to the hospital from the emergency department. *Emerg Med Int* 2012;2012:824674.
11. Han J, Pei J, Kamber M. Data mining: concepts and techniques. Amsterdam: Morgan Kaufmann; 2012.
12. Jiang X, Qu X, Davis L. Using data mining to analyze patient discharge data for an urban hospital. *Proceedings of the 2010 International Conference on Data Mining*; 2010 Jul 12-15; Las Vegas, NV. pp. 139-44.
13. Rowan M, Ryan T, Hegarty F, O'Hare N. The use of artificial neural networks to stratify the length of stay of cardiac patients based on preoperative and initial post-operative factors. *Artif Intell Med* 2007;40(3):211-21.
14. Tu JV, Guerriere MR. Use of a neural network as a predictive instrument for length of stay in the intensive care unit following cardiac surgery. *Comput Biomed Res* 1993;26(3):220-9.
15. Hachesu PR, Ahmadi M, Alizadeh S, Sadoughi F. Use of data mining techniques to determine and predict length of stay of cardiac patients. *Healthc Inform Res* 2013;19(2):121-9.
16. Miyata H, Hashimoto H, Horiguchi H, Matsuda S, Motomura N, Takamoto S. Performance of in-hospital mortality prediction models for acute hospitalization: hospital standardized mortality ratio in Japan. *BMC Health Serv Res* 2008;8:229.
17. Son YJ, Kim HG, Kim EH, Choi S, Lee SK. Application of support vector machine for prediction of medication adherence in heart failure patients. *Healthc Inform Res* 2010;16(4):253-9.
18. Ahn M, Choi M, Kim Y. Factors associated with the timeliness of electronic nursing documentation. *Healthc Inform Res* 2016;22(4):270-6.
19. Sushmita S, Khulbe G, Hasan A, Newman S, Ravindra P, Basu Roy S, et al. Predicting 30-day risk and cost of "All-Cause" hospital readmissions. The Workshops at the 30th AAAI Conference on Artificial Intelligence:

- Expanding the Boundaries of Health Informatics Using AI; 2016 Feb 12-17; Phoenix, AZ.
20. Johnson AE, Pollard TJ, Shen L, Lehman LW, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. *Sci Data* 2016;3:160035.
 21. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* 2000;101(23):E215-20.
 22. Cots F, Elvira D, Castells X, Saez M. Relevance of outlier cases in case mix systems and evaluation of trimming methods. *Health Care Manag Sci* 2003;6(1):27-35.
 23. Freitas A, Silva-Costa T, Lopes F, Garcia-Lema I, Teixeira-Pinto A, Brazdil P, et al. Factors influencing hospital high length of stay outliers. *BMC Health Serv Res* 2012; 12:265.
 24. Jain AK, Murty MN, Flynn PJ. Data clustering: a review. *ACM Comput Surv* 1999;31(3):264-323.
 25. Briand LC, Wieczorek I. Resource estimation in software engineering. In: Marciniak JJ, editor. *Encyclopedia of software engineering*. New York (NY): John Wiley & Sons Inc.; 2002.
 26. Conte SD, Dunsmore HE, Shen VY. *Software engineering metrics and models*. Redwood City (CA): Benjamin/Cummings Publishing; 1986.