

Article

Application of Entropy-Based Attribute Reduction and an Artificial Neural Network in Medicine: A Case Study of Estimating Medical Care Costs Associated with Myocardial Infarction

Qingyun Du ^{1,2}, Ke Nie ^{1,2} and Zhensheng Wang ^{1,2,*}

¹ School of Resource and Environmental Sciences, Wuhan University, 129 Luoyu Road, Wuhan 430079, China; E-Mails: qydu@whu.edu.cn (Q.D.); nieke@whu.edu.cn (K.N.)

² Key Laboratory of GIS, Ministry of Education, Wuhan University, 129 Luoyu Road, Wuhan 430079, China

* Author to whom correspondence should be addressed; E-Mail: wangzhens@whu.edu.cn; Tel.: +86-27-6877-8842; Fax: +86-27-6877-8893.

Received: 6 July 2014; in revised form: 19 August 2014 / Accepted: 25 August 2014 /

Published: 29 August 2014

Abstract: In medicine, artificial neural networks (ANN) have been extensively applied in many fields to model the nonlinear relationship of multivariate data. Due to the difficulty of selecting input variables, attribute reduction techniques were widely used to reduce data to get a smaller set of attributes. However, to compute reductions from heterogeneous data, a discretizing algorithm was often introduced in dimensionality reduction methods, which may cause information loss. In this study, we developed an integrated method for estimating the medical care costs, obtained from 798 cases, associated with myocardial infarction disease. The subset of attributes was selected as the input variables of ANN by using an entropy-based information measure, fuzzy information entropy, which can deal with both categorical attributes and numerical attributes without discretization. Then, we applied a correction for the Akaike information criterion (AIC_C) to compare the networks. The results revealed that fuzzy information entropy was capable of selecting input variables from heterogeneous data for ANN, and the proposed procedure of this study provided a reasonable estimation of medical care costs, which can be adopted in other fields of medical science.

Keywords: artificial neural network; fuzzy information entropy; medical costs estimation; myocardial infarction disease; attribute reduction

1. Introduction

Reliable estimates of medical care costs for myocardial infarction (MI)-related patients can provide an alternative to cost-effectiveness evaluations of MI prevention, screening and treatment policies [1,2]. Besides, obtaining accurate estimates of this outcome will allow the administration to properly manage the available medical resources for inpatient hospitalizations. Previous works have demonstrated that demographic factors, percutaneous coronary intervention, coronary artery bypass graft surgery and length of stay were significantly associated with higher healthcare costs [3,4].

Artificial neural networks (ANNs) provide a rich, powerful and robust nonparametric modeling framework currently being used in a variety of applications in medicine, such as diagnosis, electronic signal analysis, medical image analysis, radiology and clinical outcome prediction. In [5], the authors made use of ANN analysis to assess the accuracy of real-time endoscopic ultrasound elastography in focal pancreatic lesions. In [6], the authors developed random forests, support vector machines and ANN models to diagnose acute appendicitis. Shi *et al.* validated the use of ANN models for predicting quality of life after breast cancer surgery [7]. In [8], the author developed a biomedical-based decision support system for the classification of heart sound signals by using principal component analysis (PCA) and ANN. In [9], an ANN model was developed to predict survival in patients with pancreatic ductal adenocarcinoma. In [10], the authors applied the ANN method to model the sample entropy.

ANNs offer several advantages, including requiring less formal statistical training, the ability to implicitly detect complex nonlinear relationships between dependent and independent variables, the ability to detect all possible interactions between predictor variables and the availability of multiple training algorithms [11]. However, the “black box” nature, heavy computational burden, proneness to overtraining and the empirical nature of model selection are the disadvantages of ANN models. Hybrid methods, like ANN and genetic algorithms (GAs) [12], ANN and PCA [8], ANN and the artificial bee colony algorithm (ABC) [13], ANN and autoregressive integrated moving average (ARIMA) models [14] and ANN and rough sets [15] were developed to overcome the above problems. In this study, an integrated method based on fuzzy information entropy and ANN was developed to estimate medical care costs for admissions of MI disease.

To handle the multidimensional data efficiently, dimensionality reduction should be performed to map the data to a lower dimensional space. A typical method is attribute reduction based on Pawlak’s rough set model, which has been successfully used in feature subset selection and attribute reduction [16,17]. Pawlak’s rough set model works in circumstances where only nominal attributes exist in an information system and is limited in dealing with numerical variable directly unless applying a discretizing algorithm, which may lead to some information loss [18,19]. To address this problem, rough-fuzzy and fuzzy-rough sets were proposed in [20] and analyzed in detail in [21], which were successfully used in attribute reduction. In [22,23], the authors proposed an integrated use of fuzzy and rough set theories to reduce the data redundancy based on the fuzzy dependency function.

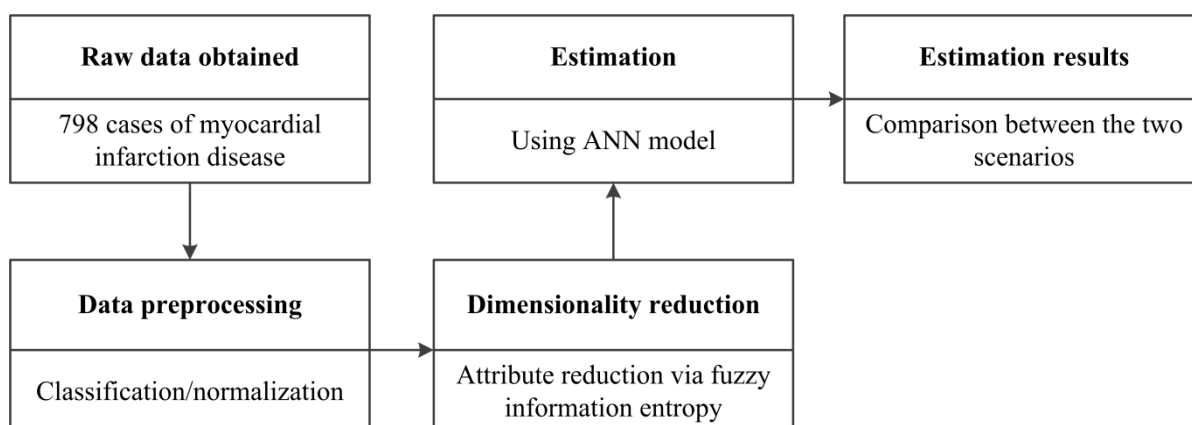
Then, the authors introduced an information measure, fuzzy information entropy, to measure the discernibility power of a fuzzy equivalence relation [24]. The significance of categorical, numeric and fuzzy attributes can be defined in a general form with this measure. Thus, we applied fuzzy information entropy to compute attribute reduction in this research.

In this study, a method was introduced to estimate the medical care costs, obtained from 798 cases of MI-related patients. Then, the data set was mapped to a lower dimensional space using fuzzy information entropy, being applied as an attribute reduction technique. Therefore, the problem of overtraining and heavy computational burden of ANNs could be avoided by eliminating superfluous input attributes. The result showed that the proposed method was efficient in estimating medical care costs and could be adopted in various medical applications.

2. Materials and Methods

Figure 1 shows the procedure proposed in this research. It consists of five parts: (1) raw data obtainment; (2) data preprocessing, including classification and normalization; (3) dimensionality reduction via fuzzy information entropy; (4) estimating medical costs through an ANN; and (5) comparing the results.

Figure 1. The procedure of the proposed method.



2.1. Raw Data

For this study, we obtained 798 cases of inpatients with myocardial infarction (ICD-10 code I21, International Classification of Diseases, 10th revision) from three comprehensive hospitals in Wuhan, China. The raw data contain the demographic factors of patients (age in years, gender), characteristics of patients (history of diabetes, blood pressure, history of smoking, cholesterol, physically active or not, obesity, history of angina, history of MI), medical treatment process (prescribed nitroglycerin, taking anti-clotting drugs, electrocardiogram result, creatine phosphate kinase blood result, troponin T blood result, taking clot-dissolving drugs, time of hospitalization, hemorrhaging, magnesium, digitalis, beta blockers, surgical treatment, surgical complications and length of stay) and the medical care costs for each patient.

2.2. Fuzzy Information Entropy

2.2.1. Fuzzy Rough Set Model

The equivalence relations are introduced in Pawlak’s rough set-based methodology to partition the universe and generate mutually exclusive equivalence classes as elemental concepts for categorical variables. In [22], the authors suggested that the fuzzy equivalence relation should be generated for numeric and fuzzy attributes, instead of crisp equivalence relations.

The heterogeneous data can be depicted as an information system $S = (U, C, D)$ and $B \subseteq C$, where $U = \{x_1, x_2, \dots, x_n\}$, and R is a binary relation on B , denoted by a relation matrix $M(R) = (r_{ij})$, where $r_{ij} \in [0,1]$ is the relation of x_i and x_j .

Given that U and R represent a non-empty finite set and a binary fuzzy relation on this set, for $\forall x, y, z \in U$, R satisfies:

- (1) Reflectivity: $R(x, x) = 1$;
- (2) Symmetry: $R(x, y) = R(y, x)$;
- (3) Transitivity: $\min_y(R(x, y), R(y, z)) \leq R(x, z)$

Definition 1. For $\forall x_i \in U$, the fuzzy equivalence class generated by x_i and R is defined as:

$$[x_i]_R = \frac{r_{i1}}{x_1} + \frac{r_{i2}}{x_2} + \dots + \frac{r_{in}}{x_n} \tag{1}$$

Definition 2. Given a fuzzy information system, $S = \{U, C \cup D, V, f\}$, where C is the set of condition attributes and D is the decision attribute. For $B \subseteq C$, $\gamma_B(D) = |POS_B(D)|/|U|$ represents the dependency of B on D , where $POS_B(D)$ is a lower approximation of the decision and also called a positive region of the decision.

Definition 3. For $\forall b \in B$, b is superfluous in B on D if $\gamma_{B-b}(D) = \gamma_B(D)$; otherwise b is indispensable. If $\forall b \in B$ is indispensable, B is indispensable. B is a reduct of C if B satisfies:

- (1) $\gamma_B(D) = \gamma_C(D)$;
- (2) $\gamma_{B-b}(D) < \gamma_B(D)$, for $\forall b \in B$

2.2.2. Entropy-Based Information Measure

The fuzzy equivalence class $[x_i]_R$ has been defined in Definition 1. Then, the pertinent points of the fuzzy information entropy are introduced as follows.

Definition 4. The cardinality of $[x_i]_R$ is defined as:

$$|[x_i]_R| = \sum_{j=1}^n r_{ij} \tag{2}$$

for $\forall x_i \in U$ and $r_{ij} \leq 1$, $|[x_i]_R| \leq n$.

Definition 5. The information quantity of the fuzzy equivalence relation is introduced as:

$$H(R) = - \sum_{i=1}^n \frac{1}{n} \log_2 \frac{|[x_i]_R|}{n} \tag{3}$$

when the relation R is a crisp equivalence relation, this information quantity is defined as Shannon's entropy [24].

Definition 6. Given a fuzzy decision system $S = \{U, C \cup D, V, f\}$, B_1, B_2 are two subsets of C and the fuzzy equivalence classes on B_1 and B_2 are $[x_i]_{B_1}$ and $[x_i]_{B_2}$; the joint entropy and the conditional entropy are defined as:

$$H(B_1 \cup B_2) = H(R_{B_1} \cap R_{B_2}) = - \sum_{i=1}^n \log_2 \frac{|[x_i]_{B_1} \cap [x_i]_{B_2}|}{n} \quad (4)$$

$$H(B_2|B_1) = - \sum_{i=1}^n \log_2 \frac{|[x_i]_{B_1} \cap [x_i]_{B_2}|}{|[x_i]_{B_1}|} \quad (5)$$

Definition 7. Given a fuzzy decision system $S = \{U, C \cup D, V, f\}$, $B \subseteq C$, $\forall b \in B$, b is superfluous if $H(B) = H(B - b)$; and B is independent if $H(B) > H(B - b)$. B is a reduct if $H(B) = H(C)$.

2.3. Artificial Neural Network (ANN)

An ANN is composed of a number of interconnected neurons (referred to as "nodes", "processing units" or "processing elements"), organized hierarchically in layers. In an ANN, knowledge about the problem is modeled by using learning algorithms and saved in weighted connections. The feed-forward neural network of multi-layer perception (MLP) with an error back-propagation (BP)-type of learning algorithm is the most popular ANN model used in estimation and regression problems. An MLP consists of an input layer, hidden layers and an output layer, each of which is composed of a set of neurons. The MLP with the back-propagation algorithm is trained using a dataset of associated input and target values, and the network is updated by rearranging the weights of neurons in every epoch upon calculating the error in the network's output, until the performance of the network is satisfactory.

The number of neurons in the input layer depends on the result of the dimensionality reduction based on the fuzzy information entropy, and the problems of overtraining and heavy computational burden could be avoided with the lower number of input variables. For the neurons in the hidden and output layer, their inputs are processed by multiplying each input by a corresponding weight and summing the products and then transmitting this sum to the output by using a nonlinear transfer function. In each epoch, the weights among the neurons of the network are adjusted based on the errors between the actual outputs and the target outputs. When the training is complete, the network should be able to provide an accurate estimation for a given input.

A three-layered MLP feed-forward neural network with the back-propagation learning algorithm was applied in our study. For this research, there is only one neuron in the output layer representing medical care costs; thus, the most critical problem is defining the size of the input layer, which is directly related to the network's performance. The entropy-based information measure for the fuzzy equivalence relation can analyze the significance of various factors to remove the superfluous attributes from the heterogeneous data. Therefore, the combined method presented in this paper can overcome the disadvantages of ANNs in estimating medical costs.

2.4. Model Selection and Comparison Methods

2.4.1. Akaike Information Criterion (AIC)

Information-based criteria, such as the Akaike information criterion (AIC), which are measures of the relative quality of an estimated statistical model, are widely used as the model selection approach. The underlying idea of the information-based criteria is to identify an optimal trade-off between an unbiased approximation of the model and the complexity of the model. In [25,26], the authors use AIC in the neural network model selection to determine the optimal parameters of the ANN model. In this study, the comparison of ANN models is conducted using AIC. In the general sense, by using the likelihood L , the AIC is calculated as:

$$AIC = -2 \ln(L) + 2k \quad (6)$$

or using residual sum of squares (RSS):

$$AIC = N \ln(RSS/N) + 2k \quad (7)$$

where k denotes the total number of estimable parameters in the model and N is the sample size. For a small sample size, when N/k is less than 40, a correction for AIC (AIC_C) is recommended in [27] as follows.

$$AIC_C = N \ln(RSS/N) + 2k + 2k(k+1)/(N-k-1) \quad (8)$$

The model exhibiting the smallest AIC value is selected as the best fit model in this study.

2.4.2. Evaluation of Performance

We use two different criteria to evaluate the performance of the integrated method based on fuzzy information entropy and ANNs: root mean square error (RMSE) and the mean absolute percentage error (MAPE). The first criterion is calculated by:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (C_i - T_i)^2} \quad (9)$$

where C_i and T_i denote the calculated and target value of the medical costs, respectively, and n is the number of training data. The reason for applying the training error in the criteria is that it is directly related to the ANN's memorization ability. The second criterion is MAPE, defined as:

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{C_i - T_i}{T_i} \right| \quad (10)$$

where n , C_i and T_i are defined as in (9). The second criterion is the average percentage of the absolute values of relative error, which can be used directly for experimental data without data normalization.

3. Experimental Results

3.1. Data Preprocessing

For this study, the data preprocessing consists of two parts: (1) classification of categorical variables; and (2) data normalization for numerical data. Categorical variables should be converted to quantitative data through the use of classification based on the categories of each variable, and then, they can be used as input variables of an ANN (Table 1). The classification of categorical variables is demonstrated in Table 1.

Table 1. Classification of 21 categorical variables used in the analysis.

Variables	Classification
Gender	Male; female
History of diabetes	No; Yes
Blood pressure	Hypotension; normal; hypertension
Smoker	No; Yes
Cholesterol	Normal; high
Physically active	No; Yes
Obesity	No; Yes
History of angina	No; Yes
History of MI	No; Yes
Prescribed nitroglycerin	No; Yes
Anti-clotting drugs	None; aspirin; heparin; warfarin
EKG ^a result	No ST elevation; ST elevation
CPK ^b blood result	Normal CPK; high CPK
Troponin T blood result	Normal troponin T; high troponin T
Clot-dissolving drugs	None; streptokinase; reteplase; alteplase
Hemorrhaging	No; Yes
Magnesium	No; Yes
Digitalis	No; Yes
Beta blockers	No; Yes
Surgical treatment	None; PTCA ^c ; CABG ^d
Surgical complications	No surgery performed; No; Yes

^a EKG (electrocardiogram); ^b CPK (creatinine phosphokinase); ^c PTCA (percutaneous transluminal coronary angioplasty); ^d CABG (coronary artery bypass grafting)

The numerical variables include age in years, time to hospital, length of stay and medical treatment costs. Before computing attribute reduction based on fuzzy information entropy, numerical variables should be normalized into [0, 1]. Given a numerical variable X , the formula of min-max normalization is given as:

$$X' = \frac{X - \min(X)}{\max(X) - \min(X)} \quad (11)$$

where X' is the normalized value of X ranging from zero to one, and $\max(X)$ and $\min(X)$ represent the maximum and minimum value of X , respectively. Table 2 demonstrates the descriptive statistics of four numerical variables.

Table 2. Statistical properties of four numerical variables.

Variables	Max	Min	Mean	Median	SD
Age in years	87	45	61.75	61.00	8.835
Medical visits (times)	10	1	2.97	3.00	1.432
Length of stay (day)	11	1	3.54	4.00	2.656
Medical treatment costs (10 ³ CNY)	61.80	1.71	19.92	25.77	17.164

3.2. Dimension Reduction via Fuzzy Information Entropy

A three-layered MLP feed-forward ANN structure has been made use in this work. However, one of the problems related to ANN modeling is that sample data usually contains superfluous, irrelevant, and noisy input variables, which may obstruct knowledge acquisition and affect the network’s generalization ability. Therefore, due to the multiple input variables, the choice of ANN input variables could overcome these drawbacks.

Definition 8. Given $S = \{U, C \cup D, V, f\}$ represents a fuzzy information system, for $B \subseteq C$, $\forall b \in B$, the significance of attribute b in B relative to D is defined as:

$$SIG(b, B, D) = H(D|B - b) - H(D|B) \tag{12}$$

The fuzzy information entropy is introduced in Section 2.2. The greater the entropy value is, the stronger the discernibility is and the more significant the attribute is. If the significance equals zero, then we consider that the attribute b is superfluous; otherwise, b is indispensable. The aim of attribute selection is to search a subset of attributes that has the same approximating power as the original data and that does not have any redundant attribute. We apply a forward search algorithm, which has been discussed in detail in [24].

Table 3. Increment of the entropy and the significance of each attribute.

i	$Reduct_i$	$H(D B_i)$	$SIG(b_i, B_i, D)$
1	$B_1 = \{\text{Length of stay}\}$	1.4379	1.4379
2	$B_2 = B_1 \cup \{\text{Surgical treatment}\}$	1.5734	0.1355
3	$B_3 = B_2 \cup \{\text{Age in years}\}$	1.6572	0.0838
4	$B_4 = B_3 \cup \{\text{Surgical complications}\}$	1.6682	0.011
5	$B_5 = B_4 \cup \{\text{Time to hospital}\}$	1.6768	0.0086
6	$B_6 = B_5 \cup \{\text{Clot-dissolving drugs}\}$	1.6786	0.0018
7	$B_7 = B_6 \cup \{\text{Taking anti-clotting drugs}\}$	1.6813	0.0027
8	$B_8 = B_7 \cup \{\text{Magnesium}\}$	1.6824	0.0011

The parameter ε ($\varepsilon = 0.001$) is a tiny positive real number that controls the convergence. Attribute reduction starts with an empty set of attributes, and in each iteration, one attribute is added into this empty set to produce the maximal increment of fuzzy information entropy; the process stops when the increment of entropy is less than ε in one round by adding any attribute into this subset. Then, the superfluous attributes were eliminated, and eight key attributes were selected as the input variables of the ANN. Let $B_i = \{b_1, b_2, \dots, b_i\}$ be a reduct, where b_i ($i = 1, 2, \dots, 8$) denotes the eight selected attributes, respectively; the maximal increment of entropy is fulfilled during this process, and the

significance of the selected attribute is summarized in Table 3. The subset of attributes B_8 provides the maximal entropy as 1.6824, which is the result of reduction.

3.3. Estimation Using the Artificial Neural Network

Two scenarios are simulated and compared in this step, namely estimation of the medical costs using the ANN only (T_1) and using the proposed method (T_2). Table 4 demonstrates the structure and training parameters, which consists of the number of neurons in the layers, initial learning rate, momentum constant value and activation functions used in different layers.

The stopping criteria in Table 4 determine when to stop training the neural network. Criterion 1 defines the maximum number of minutes for the algorithm to run; Criterion 2 depicts the maximum number of epochs allowed, and the training stops if the maximum number of epochs is exceeded; Criterion 3 describes an threshold value in the training error, and the training discontinues if the relative change in the training error is less than the threshold value compared to the previous epoch. In each epoch, whether training is proceeding or not is based on the stopping criteria, which is checked in the given order. In this study, 496 samples were selected as the training set to update the weights of the network; 153 samples were selected as the validation set to prevent the overtraining; and 149 samples were selected as the test set to measure the performance of the network. The results of network with the best estimation performance of the two scenarios are described in Table 5. For each network, the training process has been replicated 20 times to acquire stable performance with random initially values of weight and bias.

Table 4. ANN architecture and training parameters.

ANN architecture	
The number of layers	3
The number of neurons in the layers	Input: 24 (T_1)/8 (T_2) Hidden: ≤ 10 Output: 1
The initial weights and bias	The Nguyen–Widow method
Activation functions	Hidden: Hyperbolic tangent Output: Identity
ANN parameters	
Learning algorithm	Back-propagation
Optimization algorithm	Gradient descent
Initial learning rate	0.4
Momentum	0.9
Stopping criteria	
Maximum training time	10 min
Maximum training epochs	1000
Minimum relative change in training error	0.0001

Table 5. Results of the selected networks. RSS, residual sum of squares.

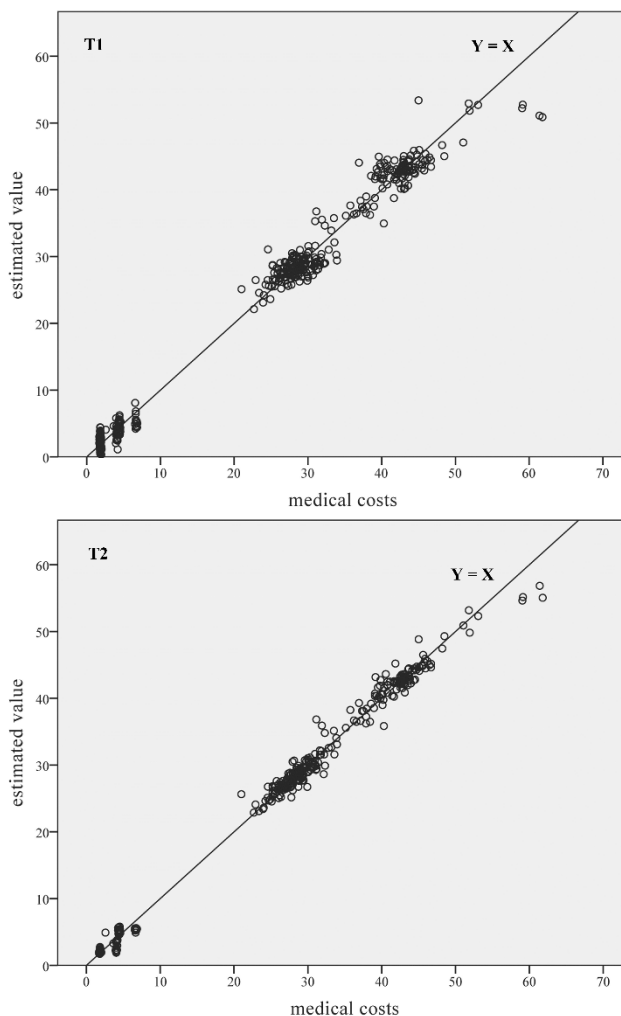
Scenario	<i>H.U.</i>	<i>k</i>	<i>N</i>	<i>RSS</i>	<i>AIC_C/n</i>	<i>RMSE</i>	<i>MAPE</i>
<i>T</i> ₁	9	235	496	1,593.81	2.97	1.79	17.46%
<i>T</i> ₂	8	81	496	586.42	0.56	1.09	6.57%

* *H.U.* denotes the number of neurons in the hidden layer; *k* denotes the total number of estimable parameters.

Table 5 contains comparison between the chosen networks of scenario *T*₁ and *T*₂ selected by *AIC_C*. From Table 5, a network with nine hidden neurons was selected in *T*₁, while the chosen network in *T*₂ has eight hidden units. The results of *AIC_C* reveal that the network complexity is simplified by eliminating the superfluous input attributes via fuzzy information entropy, which will prevent the problem of overtraining and the heavy computational burden of ANN. At the same time, the results of *RMSE* and *MAPE* show that the proposed method provides a better estimation performance.

Figure 2 displays two scatterplots of estimated value on the *y*-axis by observed medical costs on the *x*-axis for all training samples based on the two selected networks, which demonstrates that the selected network of *T*₂ has a relatively good ability of estimating medical cost than the chosen network of *T*₁. The results indicate that the reduct can capture the content in the original dataset while maintaining good approximating ability.

Figure 2. Estimated-by-observed charts for medical cost.



4. Conclusions

In this paper, we introduced an integrated method to address the overtraining and heavy computational burden of ANN modeling, and the proposed approach provides a relatively good approximating ability of estimating medical costs related to myocardial infarction. The results showed that attribute reduction based on fuzzy information entropy can be applied for selecting the input variables of an ANN.

The choice of input variables is a fundamental and crucial consideration in identifying the optimal functional form of statistical models [28]. However, ANN is a typical data-driven statistical modeling approach, and there is no prior assumption made regarding the structure of the model. Selecting input variables is complicated by the fact that the data is multidimensional and heterogeneous and there is interdependence between available input variables and redundant variables with little predictive power, which results in the usage of dimensionality reduction techniques, often accompanied by discretizing algorithms. Shannon's entropy has been widely used as an information measure in machine learning. In this research, we applied fuzzy information entropy, which is a generalization of Shannon's entropy, to deal with both categorical attributes and numerical variables simultaneously and to measure the fuzzy equivalence relation.

In this study, the results of comparison indicate that AIC_C can be used in model selection and multi-model inference of ANN for a small sample size. The problem of model selection is considerably important for acquiring a higher level of performance in an ANN. In the field of ecology, AIC is widely used to compare and rank multiple statistical models and to estimate which of them best approximates the "true" process underlying the biological phenomenon under study [29], and it has been applied to select the optimal structure of a neural network in many works.

Although the proposed approach generates a reasonable result in estimating the medical costs related to myocardial infarction, further work should address more detailed individual-level data collection, such as income level, education level, behavioral factors, residential environment, socioeconomic factors, neighborhood characteristics, *etc.* Moreover, the subjective factors of patients should be considered in further research. One major contribution of our work is to introduce an integrated method to model the nonlinear relationship in issues involving multivariate data and to overcome the disadvantages of ANN modeling, which can be applied in other fields of medical science.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (Grant Number: 41371427/D0108) and the Fundamental Research Funds for the Central Universities of China (Grant Numbers: 2012205020214 and 2012205020215).

Author Contributions

Ke Nie and Qingyun Du conceived of and designed the study. Ke Nie and Zhensheng Wang analyzed the data and performed the experiments. Ke Nie, Qingyun Du and Zhensheng Wang wrote and revised the paper together. All of the authors have read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Mark, D.B.; Hlatky, M.A.; Califf, R.M.; Naylor, C.D.; Lee, K.L.; Armstrong, P.W.; Barbash, G.; White, H.; Simoons, M.L.; Nelson, C.L.; *et al.* Cost effectiveness of thrombolytic therapy with tissue plasminogen activator as compared with streptokinase for acute myocardial infarction. *New Engl. J. Med.* **1995**, *332*, 1418–1424.
2. Frasure-Smith, N.; Lespérance, F.; Gravel, G.; Masson, A.; Juneau, M.; Talajic, M.; Bourassa, M.G. Depression and health-care costs during the first year following myocardial infarction. *J. Psychosom. Res.* **2000**, *48*, 471–478.
3. Chaikledkaew, U.; Pongchareonsuk, P.; Chaiyakunapruk, N.; Ongphiphadanakul, B. Factors affecting health-care costs and hospitalizations among diabetic patients in Thai Public hospitals. *Value Health* **2008**, *11*, 69–74.
4. Wang, G.; Zhang, Z.; Ayala, C.; Dunet, D.; Fang, J. Costs of Hospitalizations with a Primary Diagnosis of Acute Myocardial Infarction Among Patients Aged 18–64 Years in the United States. In *Ischemic Heart Disease*, 1st ed.; Gaze, D.C., Ed.; InTech: Rijeka, Croatia, 2013.
5. Săftoiu, A.; Vilmann, P.; Gorunescu, F.; Janssen, J.; Hocke, M.; Larsen, M.; Iglesias-Garcia, J.; Arcidiacono, P.; Will, U.; Giovannini, M.; *et al.* Efficacy of an artificial neural network-based approach to endoscopic ultrasound elastography in diagnosis of focal pancreatic masses. *Clin. Gastroenterol. Hepatol.* **2012**, *10*, 84–90.
6. Hsieh, C.H.; Lu, R.H.; Lee, N.H.; Chiu, W.T.; Hsu, M.H.; Li, Y.C.J. Novel solutions for an old disease: Diagnosis of acute appendicitis with random forest, support vector machines, and artificial neural networks. *Surgery* **2011**, *149*, 87–93.
7. Shi, H.Y.; Tsai, J.T.; Chen, Y.M.; Culbertson, R.; Chang, H.T.; Hou, M.F. Predicting two-year quality of life after breast cancer surgery using artificial neural network and linear regression models. *Breast Cancer Res. Treat.* **2012**, *135*, 221–229.
8. Uğuz, H. A biomedical system based on artificial neural network and principal component analysis for diagnosis of the heart valve diseases. *J. Med. Syst.* **2012**, *36*, 61–72.
9. Ansari, D.; Nilsson, J.; Andersson, R.; Regnér, S.; Tingstedt, B.; Andersson, B. Artificial neural networks predict survival from pancreatic cancer after radical surgery. *Am. J. Surg.* **2013**, *205*, 1–7.
10. Huang, J.R.; Fan, S.Z.; Abbod, M.F.; Jen, K.K.; Wu, J.F.; Shieh, J.S. Application of multivariate empirical mode decomposition and sample entropy in EEG signals via artificial neural networks for interpreting depth of anesthesia. *Entropy* **2013**, *15*, 3325–3339.
11. Tu, J.V. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *J. Clin. Epidemiol.* **1996**, *49*, 1225–1231.
12. Mantzaris, D.; Anastassopoulos, G.; Adamopoulos, A. Genetic algorithm pruning of probabilistic neural networks in medical disease estimation. *Neural Netw.* **2011**, *24*, 831–835.
13. Yeh, W.C.; Hsieh, T.J. Artificial bee colony algorithm-neural networks for S-system models of biochemical networks approximation. *Neural Comput. Appl.* **2012**, *21*, 365–375.

14. Ömer Faruk, D. A hybrid neural network and ARIMA model for water quality time series prediction. *Eng. Appl. Artif. Intell.* **2010**, *23*, 586–594.
15. Azadeh, A.; Saberi, M.; Moghaddam, R.T.; Javanmardi, L. An integrated data envelopment analysis-artificial neural network-rough set algorithm for assessment of personnel efficiency. *Expert Syst. Appl.* **2011**, *38*, 1364–1373.
16. Dai, J.; Xu, Q. Attribute selection based on information gain ratio in fuzzy rough set theory with application to tumor classification. *Appl. Soft Comput.* **2013**, *13*, 211–221.
17. Mac Parthaláin, N.; Jensen, R. Unsupervised fuzzy-rough set-based dimensionality reduction. *Inf. Sci.* **2013**, *229*, 106–121.
18. Garcia, S.; Luengo, J.; Sáez, J.A.; López, V.; Herrera, F. A survey of discretization techniques: Taxonomy and empirical analysis in supervised learning. *IEEE Trans. Knowl. Data Eng.* **2013**, *25*, 734–750.
19. Kotsiantis, S.; Kanellopoulos, D. Discretization techniques: A recent survey. *GESTS Int. Trans. Comput. Sci. Eng.* **2006**, *32*, 47–58.
20. Dubois, D.; Prade, H. Putting rough sets and fuzzy sets together. In *Intelligent Decision Support*, 1st ed.; Slowinski, R., Ed.; Kluwer Academic: Dordrecht, The Netherlands, 1992; pp. 203–232.
21. Morsi, N.N.; Yakout, M.M. Axiomatics for fuzzy-rough sets. *Fuzzy Sets Syst.* **1998**, *100*, 327–342.
22. Jensen, R.; Shen, Q. Fuzzy-rough attribute reduction with application to web categorization. *Fuzzy Sets Syst.* **2004**, *141*, 469–485.
23. Jensen, R.; Shen, Q. Semantics-preserving dimensionality reduction: Rough and fuzzy-rough-based approaches. *IEEE Trans. Knowl. Data Eng.* **2004**, *16*, 1457–1471.
24. Hu, Q.; Yu, D.; Xie, Z. Information-preserving hybrid data reduction based on fuzzy-rough techniques. *Pattern Recogn. Lett.* **2006**, *27*, 414–423.
25. Tortum, A.; Yayla, N.; Çelik, C.; Gökdağ, M. The investigation of model selection criteria in artificial neural networks by the Taguchi method. *Physica A* **2007**, *386*, 446–468.
26. Arifovic, J.; Gencay, R. Using genetic algorithms to select architecture of a feedforward artificial neural network. *Physica A* **2001**, *289*, 574–594.
27. Hurvich, C.M.; Tsai, C.L. A corrected Akaike information criterion for vector autoregressive model selection. *J. Time Ser. Anal.* **1993**, *14*, 271–279.
28. May, R.; Dandy, G.; Maier, H. Review of input variable selection methods for artificial neural networks. In *Artificial neural networks—Methodological advances and biomedical applications*; Suzuki, K., Ed.; InTech: Rijeka, Croatia, 2011; pp. 19–44.
29. Symonds, M.R.; Moussalli, A. A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike’s information criterion. *Behav. Ecol. Sociobiol.* **2011**, *65*, 13–21.