Application of rough sets in audience rating prediction

Meimei Wu[1], Yan Wang[1]*, Xingli Liu[2]

[1]*School of Science, Communication University of China, Beijing 100024, China*

[2]*College of Information Science and Technology, Heilongjiang University of Science and Technology, Harbin, 150022, Heilongjiang, China*

Received 1 June 2014, www.cmnt.lv

Abstract

The Audience Rating Prediction plays a significant role in the increasingly fierce competition in the television industry. This paper proposes an approach of combining the rough sets with the back propagation neural network, which can be used to predict complicated audience ratings with dynamic and non-linear factors. The attribute reduction based on rough sets can remove redundant information, weaken the impacts of noise data and interdependency data to BP neural network and reduce the complexity of the neural network system. Therefore, this approach can improve the accuracy of prediction and reduce the training time. Through the experiments of audience rating data, this paper compares the approach based on Rough sets and BP neural network with that of BP neural network only. These experiments represent that the Audience rating prediction based on Rough sets and BP neural network achieves better results.

Keywords: Audience Rating Prediction, Rough Sets, BP Neural Network

1 Introduction

Audience rating is a very important indicator and now it becomes the import basis of evaluating program, determining the ad price of channel and time and selecting media, channel, time and program for media. With the increasingly fierce competition in the television industry, the traditional survey and evaluation methods cannot meet the requirement. Therefore a scientific and effective method for predicting rating is necessary.

Audience Rating Prediction is to study how to convert various factors that affect the ratings into the relevant index, to analyze the impact of index ratings in order to reduce the deviation of subjective judgments by scientific methods. The methods study on the Audience rating prediction include: predicting through the establishment of multiple linear regression model [1], decision tree theory and its ID3 algorithm, prediction based on neural network model [2], prediction based on Grey GM (1, 1) model [3], and the method based on Bayesian network [4], Markov chain [5], etc. However, the methods mentioned above have their shortcomings [6]. Multiple linear regression model can play better when Problem model variables show a significant linear relationship, but some non-linear factors may cause some errors which could even result in failure; theory of decision tree is prone to over fitting because of the defect itself; the disadvantage of neural networks is that there is often a lot of redundant training set in the progress of training, neural network trained with this training set will cause overmatching problem, caused by the deviation of predicted results; Grey GM (1, 1) model obtained good results in dealing with small sample predicting data and information, but the result in dealing with large sample data has not been verified.

This paper introduces rough sets theory to the BP neural network system that presents a new audience rating prediction method. Rough sets method can reduce the number of the characteristics of representation Information, remove redundant information and reduce the complexity of the BP Neural Network system, therefore this method can improve accuracy of predictions and reduce the training time. In addition, taking the neural network as a post-information recognition system, it will achieve a better fault-tolerance and anti-jamming capability.

2 Research method based on Rough Sets and BP Neural Network

2.1 INTRODUCTION OF ROUGH SETS

Let $X \subseteq U$, R is an equivalence relation on U. When X can be expressed by union of some R basic categories, X is called can be defined by R, otherwise X can not be defined by R.

R definable set is the subset of domain of discourse, it can be precisely defined in the knowledge base K, and R undefined set cannot be defined in this knowledge base. R definable set is called R accurate set, and R undefined set is called R Inexact Sets or R rough sets.

Definition 1:

Let non-empty finite se U, A, in which U is the domain and A is attribute set. For each attribute $a \in A$, if there is a collection of property values $V_a = \{a(x)\ x \in U\}$, $S = (U, A)$ is the information system. If $A = C \bigcup D$ and $C \bigcap D = \varnothing$, where C is the condition attribute set, and D is decision attribute set (usually $D = \{d\}$), we called $S = (U, C \bigcup D)$ decision-making system.

* ***Corresponding author' s*** e-mail: wumm@cuc.edu.cn

Definition 2:

For the information system $S=(U,A)$, based CA, we say the dual equivalence relation:

$IND(C)=\{(\mathrm{x},\mathrm{y})\in U\times U\ \mathrm{a(x)}=\mathrm{a(y)},\mathrm{a}\in C\}$ is the C in distinguish relation in the domain U. In the rough set theory, the collection can be understood as a kind of equivalence relation. According to the attribute set of CA, we divide universe U and get a equivalence family recorded as U/C.

For the domain of U in the decision-making system $S=(U,\mathrm{C}\bigcup\{d\})$, according to the approximation quality of results U/C divided by condition C to:

$U/\{d\}=\{\mathrm{x}_1,\mathrm{x}_1,\cdots,\mathrm{x}_n\}$ divide by decision-making d, we get a decision:

$$\alpha(C,d)=\bigcup X\in U/\{d\}\mathrm{card}[C-(X)],$$

$$card[U]=\sum X\in U/\{d\}card[C-(X)]\mathrm{card}[U].$$

In the decision system $S=(U,C\bigcup\{d\})$, there may be redundancy conditions. We need analysis and deal with the condition-properties for simple rules. We investigate the importance of C_i to decision-making d. This indicator can be signified by the approx-imate quality differentials:

$$\beta=\alpha(C,d)-\alpha(C-\{c_i\},d).$$

If $\beta=0$, it indicates that condition c_i can't affect the decision d and it can be removed in the system. Finally we can get the reduction of decision system.

2.2 INTRODUCTION OF BP NEURAL NETWORK

BP network is a one-way transmission and multilayer feed forward networks. A standard BP neural model consists of three or more layers, including an input layer, one or more hidden layers and an output layer. The theoretical results showed that one hidden layer is sufficient for a BP neural model to approximate any continuous mapping from the input patterns to the output patterns to an arbitrary degree of freedom. In a three-layered m-h-n BP neural model, the parameters m, h and n stand for the total numbers of neurons in the input, hidden and output layers, respectively. The values of m and n are exactly determined according to the dimensions of the input and output vectors in a problem. However, the appropriate number of neurons in the hidden layer(h) is generally selected through trial and error. The training of a BP network involves three stages: the feed-forward of the input training pattern, the calculation and back propagation of the associated error, and the adjustment of the weights. After the network reaches a satisfactory level of performance, it will learn the relationships between input and output patterns and then its weights can be used to recognize new input patterns. Two parameters, learning rate ($0<\alpha<1$) and momentum ($0<\mu<1$) primarily affect the performance of training a BP neural network. The learning rate controls the amount by which weights are changed during training. The momentum avoids a major disruption of the direction of learning when some training data are very different from the majority of the data (and possibly even incorrect). A smaller learning rate and a larger momentum reduce the likelihood that the network will find weights that are a local, but not global, minimum. The detailed algorithm of BP neural network and the guidelines for choosing appropriate training parameters can be found in reference. The structure is shown in Figure 1.

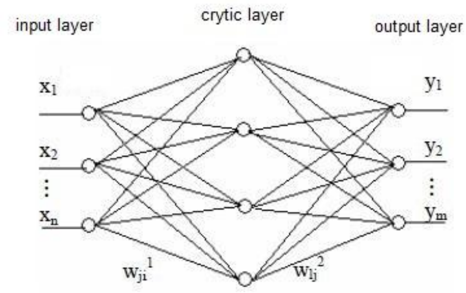

FIGURE 1 Structure figure of BP Neural Network.

2.3 DISCRETIZATION OF CONTINUOUS DATA

Discretization of continuous data is a data pre-processing step of knowledge acquisition. It can not only reduce the amount of calculation, but also suppress noise partly.

Decision table $S=<U,R,V,f>$, $U=\{x_1,x_2,\cdots,x_n\}$ is a limited collection of objects in domain of discourse, $R=C\bigcup\{d\}$ is a collection of attributes, in which $C=$ $=\{C_1,C_2,\cdots,C_k\}$ is the collection of condition attributes, d is the collection of decision attributes. For any $a\in R$, exists $U\to V_a$, V_a is the range of attribute a.

In range $V_a=[l_a\ and\ r_a]$, a breakpoint can be written as (a,c), in which $a\in R$, c is real value.

In $V_a=[l_a,r_a]$, any one of the breakpoint sets

$D_a=\{(a,c_1a),(a,c_2a)\cdots(a,c_{k+1}a)\}$, in which

$k_a\in N\ and\ l_a=c_0a<c_1a<c_2a<\cdots<c_{k+1}a=r_a$,

defines a classification P_a in v_a,

$P_a=\{[c_0a,c_1a]\bigcup[c_1a,c_2a]\bigcup\cdots\bigcup[c_ka,c_{k+1}a]\}$.

Therefore any $P=\bigcup P_a$ defines a new decision table $S_p=[<U,R,V_p,f_p>,f_p(x_a)=iZ_a,\mathrm{c}_{i+1}\,a]$.

After discretization, the original information system is replaced by a new information system, the original decision table is replaced by a new decision table and different breakpoint sets transform the same decision table to different new decision table.

From rough sets point of view, the essence of discretization is keeping the classification ability of decision table unchanged. Under the condition that relative relationship is unchanged, look for the right segmentation point set, divide the space formed by condition attributes.

Algorithm step of discretization of continuous attribute based on entropy is:

Step 1:

Set up mathematical model corresponding to practical problems.

Step 2:

Sort each condition attribute small to large, and determine the candidate breakpoint set of each condition attribute.

$$c_i^a = (v_{i-1}{}^a + v_i{}^a)/2 \ (i = 1, 2, \cdots, n_a) . \quad (1)$$

Step 3:

For each condition attribute, according to each candidate breakpoint, calculates Interval information entropy and total entropy.

$$H(c,L) = H^{Y_1}(c) + H^{Y_2}(c) + \cdots + H^{Y_m}(c) , \quad (2)$$

$$H^x(c_i^a) = \frac{|X_1|}{U} H(X_1) + \frac{|X_r|}{U} H(X_r) . \quad (3)$$

In which $\{Y_1, Y_2, \cdots, Y_m\}$ is equivalence class of $\{X\}$.

Step 4:

For each condition attribute, compares every $H(c,L)$ calculated by each candidate breakpoint, selects the youngest to the selected breakpoints. Each time a new selected breakpoints is added, $\{X\}$ is refactored to equivalence class of $\{Y_1, Y_2, \cdots, Y_m\}$, M = the number of selected points + 1.

Step 5:

Determine whether each of the equivalence class Y_m has the same decision attribute, If the same, the progress will finish, if not, will repeat.

2.4 PROCESS OF ATTRIBUTION REDUCTION

Attribution Reduction is one of the core content of rough set theory [7].

R is a set of equivalence relations, $R \in$ R. If *ind* (R) = *ind* (R-$\{R\}$), it is called that R is unnecessary in R, otherwise is necessary. If each $R \in$ R is necessary in R, it is called that R is independent, otherwise is dependent.

Set Q$\subseteq$P and Q is independent. If *ind* (Q) = *ind* (P), it is called that Q is a reduction of P.

Red (R) means all reductions of R, the set composed by all the necessary relation in P is called the core of P, expressed as *core* (P), and *core* (P)= $\bigcap$*red* (P).

The process of attribution reduction with rough sets is shown as follows:

Step 1:

Data pretreatment.

The data collected from the fact is different from each other in unit and absolute value due to the measure and the criteria, so the difference must be eliminated before the data are used. In this paper, the value of load time series is discrete, so we use the order of spaces which each value of C_j ascends in to replace the original.

Step 2:

Characterization and the foundation of knowledge system.

We draw the characteristic of each condition attribution with the corresponding data after pretreatment so as to make the calculation for the dependence the goal to each condition precise. The common characterization methods are Equal distance quartile, discretion based on Boolean calculation or rough set theory. In this paper, Equal distance quartile is used. Through characterization we can find the knowledge system and equivalence relationship as follows.

$$R_L = \{(u,v) \in U \times U \mid C_i(u) = C_i(v),\ i = 1,2,...,m\}, \quad (4)$$

$$R_{C_j} \begin{Bmatrix} (u,v) \in U \times U \mid C_i(u) = C_j(v), \\ i \neq j, i = 1,2,...,m,\ j = 1,2,...,m \end{Bmatrix}, \quad (5)$$

$$R_d = \{(u,v) \in U \times U \mid d(u) = d(v)\} . \quad (6)$$

It is obvious that R_c, R_{c_j}, R_d, are all equivalence relationship on U, and we mark the set of U as U/R_c, $U/R_{c_j}\ (j = 1,2,...,\mathrm{m})$, U/R_d. The first two are knowledge system on the basis of condition attributes and the last is knowledge system on the basis of decision attributes.

Step 3:

Calculate the dependence degree of d to condition attributes by:

$$H(R_d/R_{C_j}) = \\ = -\sum_{[x]\in U/R_C} p[x] \sum_{y\in U/R_d} p([y]/[x]) \ln(p([y]/[x])) , \quad (7)$$

where the $p[x] = \mathrm{card}[x]/\mathrm{card}[U]$, $x \in U/R$ and x is one of the condition attributes and y is one of the goal condition attributes. And $\mathrm{card}[x]$ is the influence of x (condition attribute) which can be calculated by the number of elements of x, and $\mathrm{card}[U]$ is the number of all records of U. $ind[[y]/[x]] = (card[y]/\mathrm{card}[x]/card[U]$, the numerator is the number of records which has certain values of x but different values of y (goal attribute). The value of $H(R_d/R_c)$ is the dependence of d to C.

Step 4:

Calculate the weight of conditions and reduce the indexes.

The significance of C_j in the set of C can be defined as:

$$\omega(C_j, C, d) = H(R_d/R_{C_j}) - H(R_d/R_C) , \quad (8)$$

$$H(R_d/R_{C_j}) = \\ = -\sum_{[x]\in U/R_C} p[x] \sum_{y\in U/R_d} p([y]/[x]) \ln(p([y]/[x])), \\ j = 1,2,...,16 \quad (9)$$

The higher the value of $x(C_j, C, d)$ is, the more significant C_j is in the set of C. If the value is 0, C_j is thought as a redundant attribute and can be reduced from the set of C, and later we get a new set of C, and then we get a new set of C after reduction, which can be expressed as $C = C - C_j$.

2.5 THE WORKFLOW OF AUDIENCE RATING PREDICTION

The flow chart of audience rating prediction is shown in Figure 2.

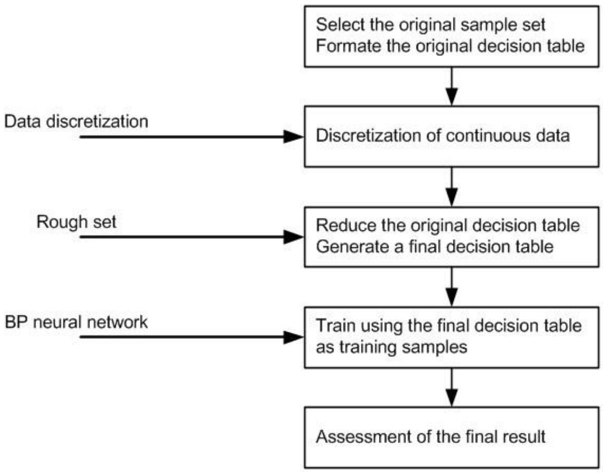

FIGURE 2 The workflow of Audience rating prediction

Step 1: Determine the index set.

Select the most suitable evaluation index.

According to the specific factors that affect the ratings among the many indicators, $U=\{U_1,U_2,U_3,\cdots,U_k\}$.

Step 2: Set the value of construction index.

For the determined set of indicators, we can fill and for the qualitative indicators, by expert advice or by using fuzzy data related approach, transform the qualitative data into quantitative data, in order to facilitate further processsing.

Step 3: Reduction.

Reduce the property values and examples through Rough set theory, look for the core attribute value and remove redundant attributes to reduce the amount of calculation.

Step 4: Self-learning process of BP neural network.

Pre-properties of neural networks are those reduced by rough set. This can reduce the Computation of the neural network. By using self-learning of neural networks we can derive the rules and get relevant results.

Step 5: Assessment of the final result.

Validated the experimental results by using test data and investigated the accuracy and practicality of them.

3 Rough sets and BP neural network for prediction

3.1 ESTABLISH DECISION TABLE

Rating is influenced by many factors. According to actual operating experience, the factors are artistic level, the prevalence of subjects, popular director, audience education level, promotion efforts, star popular level, audiences economic situation conditions, audience dience gender, audience ages, to be broadcast channels such as the overall ratings.

To establish the ratings decision table $S=(U,A), A=C\bigcup D$, where condition attributes $C=(M_1,M_2,\cdots,M_{10})$ is a collection of ten measurable parameters. Decision-making $D=(d_1,d_2)$ corresponds to two different ratings results.

3.2 REDUCTION ACCORDING TO THE THEORY OF THE ROUGH SETS

Now rough set is applied to find the dependency of each goal to each condition and the significance of each condition in the set of C with the corresponding data from training samples. Before this process, first we need to characterize the data with equal distance quartile, in other words, to replace the data with the order of the space it lies in according to the distance. The data after characterization are shown in Table 1.

TABLE 1 Data after characterization

P	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	D
1	1	1	3	2	2	1	1	1	3	1	1
2	2	2	4	3	2	2	1	2	1	1	2
3	2	1	5	4	2	2	2	2	2	2	2
4	1	1	3	2	1	2	2	1	2	1	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
17	1	3	4	3	2	2	1	1	2	2	2

We calculate the dependency of each goal (d) to the set of conditions (D) as Equation (7), which means we calculate the ratios of d that range from 1 to 3 (as Table 1 shows) to the total number of records, as well as the ratio of the combination of d and C to the total number of records. And we multiply them to get the importance of C to each d . Next is the significance of each condition attribute in the set of C according to each goal attribute with Equation (9). The procession is similar to that of Equation (7). Finally, we get the significance of each condition attribute in the set of C with Equation (8). The higher the result is, the more important the condition attribute is. The results are shown in Table 2.

TABLE 2 The significance of each condition attribute

	C_1	C_2	C_3	C_4	C_5
D	0.76	0.81	0.88	0.64	0.57
	C_6	C_7	C_8	C_9	C_{10}
D	0.92	0.32	0.47	0.32	0.83

It is obvious that star popular level (C_6) is the most significant factor for the rating, and the next four conditions are artistic level, the prevalence, popular director and to be broadcast channels such as the overall ratings. In comparison to that, the effect of the rating is inferior, especially audiences economic situation conditions that is the least significant of all ten conditions. So it can be reduced and the corresponding data and values should be deleted.

3.3 TRAIN THE REDUCED DIVISION TABLE BY USING BP NEURAL NETWORK

Topology Model of Assessment and forecast ratings system based on BP neural network, shown as Figure 3, it includes 5 nodes in inputting layer, 7 nodes in implicit layer and 2 nodes in outputting layer. Training result is shown as the Figure 4.

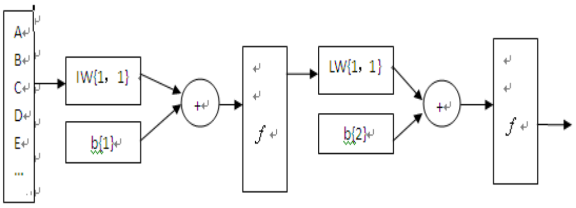

FIGURE 3 Model of the BP Neural Network

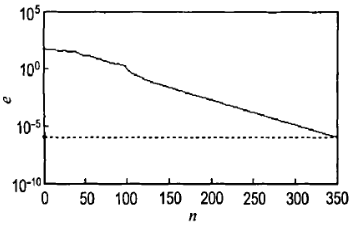

FIGURE 4 Training result

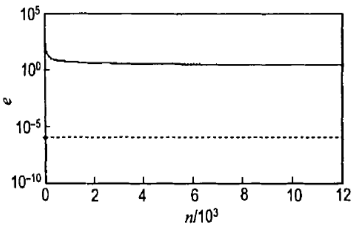

FIGURE 5 Training result that not be reduced

The results above indicate that reduction processing based on rough sets theory products more concise judging rule. It can not only improve forecasting efficiency, reduce forecasting cost, but also simplify the BP network structure using to forecast rating and improve studding efficiency of the network. Pattern Recognition ability of BP network guaranteed the forecasting accuracy. Under the condition of simple data is enough, the forecasting method that combined with the rough sets and BP neural network will achieve better results.

4 Conclusions

Rough sets method can reduce the number of the characteristics of representation Information, remove redundant information and reduce the complexity of the BP neural network system, therefore this method can improve accuracy of predictions and reduce the training time. In addition, taking the neural network as a post-information recognition system, it will achieve a better fault-tolerance and anti-jamming capability.

Acknowledgments

This work was supported by 2013BAH66F02.

References

[1] Xiao Lin, Xie Hong. (2006) Prediction System of TV Audience rating. *Microcomputer Application*, (11), 60-65.
[2] B Zhi Xiao, Shijie Ye, Bo Zhong, Cai Xin. (2009) BP neural network with rough set for short term load forecasting. *Expert Systems with Applications*, (36), 273-279.
[3] Zheng Yilei. (2009) Audience Rating Prediction of New TV Programs Based on GM (1.1) Envelopment Model. *Proceedings of 2009 IEEE International Conference on Grey Systems and Intelligent Services*, (11), 11-12.
[4] Zhang Jing, Bai Bing, Su Yong. (2007) Prediction Research on TV Program Audience rating on the Basis of Bayesian Network. *Science Technology and Engineer*, 19-24.
[5] Yan Yongquan, Zhang Dafang. (2007) Prediction Model on the Basis of Frequent Markov Chain. *Application Research of Computers*, 3-8.
[6] Jianguo tang, Feng zhu. (2010) Survey on combination of rough sets and other soft computing theories. *ApplicationResearch of Computers*, (127), 7-13.
[7] Yang Chuanjian, Ge Hao, Wang Zhisheng. (2012) Overview of attribute reduction based on rough set. *Application Research of Computers*, **29**(1), 16-20.

Authors

Meimei Wu, 16. 01. 1980, China

Current position, grades: works and studies for Ph.D. degree in Communication University of China.
University studies: master's degree in Applied Mathematics from Harbin University of Science and Technology, China in 2005.
Scientific interest: Network new media, Big Data and Database.

Yan Wang, 05. 09. 1979, China

Current position, grades: associate professor of Communication University of China.
University studies: Doctor of Philosophy's degree in Graduate University of Chinese Academy of Sciences, Beijing, China in 2009.
Scientific interest: Big Data, Probality and Statistics.

Xingli liu, 04.27. 1979, China

Current position, grades: works in Heilongjiang University of Science and Technology.
University studies: master's degree in Capital Normal University, Beijing,China in 2011.
Scientific interest: image processing and Software Engineering.