Neural Comput & Applic (2011) 20:659–669
DOI 10.1007/s00521-009-0280-0

WCCI2008

A new e-learning achievement evaluation model based on RBF-NN and similarity filter

Ching-Hsue Cheng · Tai-Liang Chen · Liang-Ying Wei · Jr-Shian Chen

Received: 29 October 2008 / Accepted: 20 April 2009 / Published online: 8 May 2009

Abstract As Internet rises fast in recent decades, teaching and learning tools based on Internet technology are rapidly applied in education. Learning through Internet can make learners absorb knowledge without the limitations on learning time and distance. Therefore, in academy, e-learning is one of the popular learning assistant instruments. Recently, "student-centered" instruction has become one of the primary approaches in education, and the e-learning system, which can provide the learning environment of personalization and adaptability, is more and more popular. By using e-learning system, teachers can adjust the learning schedule instantly for learners according to their learning achievements, and build more adaptive learning environments. However, in some cases, bias assessments are given for student achievements under specific uncontrollable conditions (i.e. tiredness, preference). In dire need of overcoming this predicament, a new model based on radial basis function neural networks (RBF-NN) and similarity filter to evaluate learning achievements is proposed. The proposed model includes three phases to reduce bias assessments: (1) preprocess: select important features (attributes) to enhance classification performance by feature selection methods and utilize minimal entropy principle approach (MEPA) to fuzzify the quantitative data, (2) similarity filter: select linguistic values for each feature and delete inconsistent data by the similarity threshold (similarity filter) and (3) construct classification model and accuracy evaluation: build the proposed model based on RBF-NN and evaluate model performance. To verify the proposed model, a practical achievement dataset, collected from e-learning online examination system in a university of Taiwan, is used as experiment dataset, and the performance of the proposed model is compared with the listing models in this paper. From the empirical study, it is shown that the proposed model provided more proper achievement evaluations than the listing models.

Keywords RBF-NN · MEPA · Similarity filter · e-Learning · Learning achievement evaluation

C.-H. Cheng (✉) · L.-Y. Wei
Department of Information Management, National Yunlin University of Science and Technology, 123, Sect. 3, University Road, Touliu, Yunlin 640, Taiwan
e-mail: chcheng@yuntech.edu.tw

L.-Y. Wei
e-mail: g9523808@yuntech.edu.tw

T.-L. Chen
Department of Information Management and Communication, Wenzao Ursuline College of Languages, 900 Mintsu 1st Road, Kaohsiung 807, Taiwan
e-mail: 97007@mail.wtuc.edu.tw

L.-Y. Wei
Institute of Biomedical Sciences, Academia Sinica, 128, Sect. 2, Academia Road, Nankang, Taipei 115, Taiwan

J.-S. Chen
Department of Computer Science and Information Management, Hungkuang University, No. 34, Chung-Chie Road, Shalu, Taichung 433, Taiwan
e-mail: jschen@sunrise.hk.edu.tw

1 Introduction

As Internet rises fast in recent decades, e-learning system has been used as teaching and learning tool, and has one of the most popular learning environments in academy. In general, e-learning has the following characteristics: (1) e-learning is effective: e-learning lets the learner to interact with the material for maximum retention of obtained knowledge; (2) e-learning is individual: each e-learner selects activities from a personal menu of learning opportunities relevant to his or her background; (3) e-learning operates in real time: e-learners can get what they need, when they need it; (4) e-learning is quick: e-learners can learn fast by e-learning. Further, both the asynchronous and synchronous features of e-learning can overcome the limitation of time and space and make learning more flexible. In asynchronous e-learning system, to make learners achieve learning goals based on their paces and construct a teaching mechanism with personalization and adaptability, online examination system is applied to evaluate learning process. By online examination system, an objective learning evaluation is given for each stage of learning process, but in general cases the final learning achievement for each learner is rated and given by teachers. However, in some specific uncontrollable conditions (i.e. tiredness and preference) bias evaluations or improper rates are assigned. Therefore, an appropriate achievement-evaluation assistant method is essential for teacher or e-learning system to evaluate learner objectively.

In most cases, statistical methods are utilized to produce achievement evaluations by comparing individual achievement with the achievements in the norm group (e.g. larger student population). However, such approaches have not been widely adopted, possibly because numerical evaluations produced by these methods are less meaningful to the user [1]. In addition, the assumptions of probability density function are made when using statistical methods to give achievement evaluations, and the assumptions are not proper for the observations in each type of achievements dataset.

In recent decades, fuzzy techniques have been applied to many research areas such as supplier selection, stock markets forecasting and innovation diffusion of products [2–4]. Besides these applications above, in the area of education research, fuzzy set theory has also been utilized in many achievement evaluation methods to produce learning achievements for learners as follows. An achievement evaluation method using fuzzy techniques was proposed by Biswas [5] to produce evaluations for student answer scripts. Echauz and Vachtsevanos [6] presented a fuzzy grading system. Law [7] proposed an alternative approach, based on the notion of fuzzy expected values, to give performance evaluation. Cheng and Yang [8] presented a method using fuzzy sets to apply in education grading systems. Chen and Lee [9] presented new methods based on fuzzy sets to give achievement evaluations. Ma and Zhou [10] presented an approach using fuzzy set to provide assessments for student-centered learning process. Weon and Kim [11] pointed out that the chief goal of educational grading system should employ linguistic assessments such as “difficulty”, “importance” and “complexity”, as evaluations for student answer scripts, and therefore proposed an evaluation method using fuzzy membership function to evaluate learning achievement. Rasmani and Shen [1] proposed a new method, based on data-driven fuzzy rule induction, for evaluating student academic performance. In Rasmani and Shen’s research [1], fuzzy membership functions for achievement evaluations are assigned by teachers, tutors or experts. Bai and Chen [12] presented a new method, using fuzzy membership functions and fuzzy rules, to generate learning achievement evaluations for students. From the literature above, several approaches using fuzzy techniques have been provided to evaluate student academic performance in practice. However, these approaches employ expert opinions mostly to produce evaluations, and therefore it is difficult to explore and utilize valuable information embedded in collected data [1].

To address the deficiencies of previous achievement-evaluation methods, this paper proposes a new model based on two advanced methods, minimize entropy principle approach (MEPA) and radial basis function neural network (RBF-NN), to produce objective achievement evaluations without utilizing subjective opinions of teachers or experts, and there are three main phases contained in the model: (1) preprocess; (2) similarity filter; and (3) construct classification model and accuracy evaluation.

Entropy-based discretization [13] method, one of the popular methods in preprocess of data mining, uses class information to discretize features, and the entropy-based discretization of quantitative features is a valuable aspect of data mining, particularly in classification problems. MEPA, one of the entropy-based discretization methods, can discretize the observations contained in target dataset based on data distribution characteristics to construct membership functions objectively instead of subjective judgments (expert’s opinions) used by previous works. In addition, due to the fact that similarity threshold can screen inconsistent data (similarity filter) and feature selection can reduce the number of features, the proposed model incorporates feature selection method and similarity threshold in the phase of preprocess and similarity filter to enhance model performance.

RBF-NN, one type of ANN, applies locally tuned neurons to perform function mappings, and is constructed for several different purposes such as function approximation or curve fitting and pattern recognition [5, 14–20]. The RBF-NN is a simple and fast learning network, and its nonlinear structures can build model for target dataset without the limitations of statistical methods (statistical distribution assumptions are necessarily made for observations). Therefore, in this paper, RBF-NN is employed as a classification model in the last phase of the proposed model to improve classification efficiency.

To verify the proposed model, this paper provides an empirical case study, which employs a practical learning achievement dataset as experimental dataset. The practical dataset is collected from an e-learning online examination system used by a university of Taiwan. From the results of this study, the proposed model may be of useful to the educators in academy who apply e-learning system to teach and evaluate the learning performance of learners. To detail the proposed model, the rest of this paper is organized as follows. Literature review is arranged in Sect. 2. The proposed achievement evaluation model is introduced in Sect. 3. An empirical case study is provided in Sect. 4. In Sect. 5, findings and conclusions along with recommendations for future research are given.

2 Literature review

This section reviews related works of feature selection, MEPA, similarity based on fuzzy rule and RBF-NN.

2.1 Feature selection

Data mining and knowledge discovery has been applied to many fields, and feature selection is a critical aspect of knowledge discovery process. Redundant features duplicate much or all of the information contained in one or more other features. Irrelevant features contain almost no useful information for the data mining task at hand. Redundant and irrelevant features can reduce classification accuracy and quality of the clusters that found. Feature selection is to choose a subset of input features by eliminating redundant and irrelevant features, and has been an important research area in statistics, pattern recognition and data mining [21].

According to a study by Witten and Frank [22], the eight feature selection methods, Cfs, Chi-square, Consistency, Gain Ratio, InfoGain, OneR**,** Relief and Symmetrical uncertainty, are well known and widely used in academic studies. The first two methods are usually used for statistical analysis, and the rest are generally used for data mining. Thus, this study introduces the eight evaluator methods for selecting features as follows:

Cfs (*Correlation-based feature selection*) [23] The method evaluates feature subsets that are highly correlated with the class, preferably with low intercorrelation.
Chi-squared [24] The method determines the value of a feature by computing the value of the chi-squared statistic with respect to the class.
Consistency [25] This method values a feature subset by using the level of consistency in the class values, when the training instances are projected onto the feature subset.
Gain ratio [26] This method values a feature by measuring gain ratio with respect to the class.
InfoGain [24] This method values a feature by measuring the information gain with respect to the class.
OneR [27] This method evaluates the value of a feature by using the OneR classifier.
Relief [28] This method evaluates the value of a feature by repeatedly sampling an instance and considering the value of the given feature for the nearest instance of the same and different class.
Symmetrical uncertainty [29] This method evaluates the value of a feature by measuring the symmetrical uncertainty with respect to the class.

2.2 Minimize entropy principle approach (MEPA)

A key objective of entropy minimization analysis is to determine the quantity of information in a given data set. The entropy of a probability distribution is a measure of the uncertainty of the distribution [30, 31]. Assume that a threshold value is being seeking for a sample in the range between x_1 and x_2. An entropy equation is written for the regions $[x_1, x]$ and $[x, x_2]$, and denote the first region p and the second region q. Entropy with each value of x are expressed as [13]:

$$S(x) = p(x)S_p(x) + q(x)S_q(x) \quad (1)$$

where

$$\begin{aligned} S_p(x) &= -[p_1(x)\ln p_1(x) + p_2(x)\ln p_2(x)] \\ S_q(x) &= -[q_1(x)\ln q_1(x) + q_2(x)\ln q_2(x)] \end{aligned} \quad (2)$$

and $p_k(x)$ and $q_k(x)$ = conditional probabilities that the class k sample is in the region $[x_1, x_1 + x]$ and $[x_1 + x, x_2]$, respectively. $p(x)$ and $q(x)$ = probabilities that all samples are in the region $[x_1, x_1 + x]$ and $[x_1 + x, x_2]$, respectively

$$p(x) + q(x) = 1 \quad (3)$$

A value of x that gives the minimum entropy is the optimum threshold value. The entropy estimates of $p_k(x)$ and $q_k(x)$, $p(x)$ and $q(x)$, are calculated as follows [32]:

$$p_k(x) = \frac{n_k(x) + 1}{n(x) + 1} \quad (4)$$

$$q_k(x) = \frac{N_k(x) + 1}{N(x) + 1} \quad (5)$$

$$p(x) = \frac{n(x)}{n} \quad (6)$$

$$q(x) = 1 - p(x) \quad (7)$$

where

$n_k(x)$ number of class k samples located in$[x_1, x_1 + x]$
$n(x)$ the total number of samples located in $[x_1, x_1 + x]$
$N_k(x)$ number of class k samples located in $[x_1+x, x_2]$
$N(x)$ the total number of samples located in $[x_1+x, x_2]$
n total number of samples in $[x_1, x_2]$

2.3 Similarity based on fuzzy rule

Fuzzy classification is an important application in fuzzy logic research areas [33]. Yuan and Shaw [34] proposed an induced fuzzy decision trees approach to expand the approach of decision trees induction, and the approach consists of the following steps: (1) fuzzification of the dataset, (2) induction of the fuzzy decision trees, (3) conversion (convert the decision trees into rule set), and (4) classification (classify data by the rule set).

Chen et al. [35] present a new approach based on the filtering the fuzzy subsethood values to reduce the number of fuzzy rules and to enhance the accuracy rate. In each subgroup, the fuzzy subsethoods [34, 36] are calculated between the decisions of the subgroup and terms of each feature. The subsethood value is defined as following:

$$S(A,B) = \frac{M(A \cap B)}{M(A)} = \frac{\sum_{u \in U} \min(\mu_A(u), \mu_B(u))}{\sum_{u \in U} \mu_A(u)} \tag{8}$$

In (8), A and B are two fuzzy sets defined on the universe of discourse U. The μ_A and μ_B be the respectively membership functions.

2.4 Radial basis function neural networks (RBF-NN)

Moody et al. [14] proposed radial basis function neural networks based on supervised learning. The RBF-NN has a feed forward structure consisting of a single hidden layer of j locally tuned units, which are fully interconnected to an output layer of linear units. In general, we have assumed a Gaussian basis function for the hidden units given as $Z_j(x)$ [see (9)].

$$Z_j(x) = \exp\left(\frac{-||x-\mu_j||}{2\sigma_j^2}\right) \tag{9}$$

where μ_j and σ_j are mean and the standard deviation, respectively, of the jth unit receptive field.

And the summation of all $Z_j(x)$ weighted with corresponding weight, W_j, is the hidden layer to produce an output value as $\hat{Y}$ shown in (10).

$$\hat{Y} = \sum_{j=1}^{J} W_j Z_j(x) \tag{10}$$

RBF-NN takes the sum of square error criterion function as an error function E to be minimized over the given training set. The process of the RBF-NN algorithm is described as follows:

Step 1 Set the number J of hidden nodes and the stop criterion ε. Initialize all the parameters μ_j, W_j, σ_j, $j = 1,\ldots, J$.
Step 2 Sequentially input a data point x and calculate its network output $\hat{Y}$by (10).
Step 3 Update the parameters μ_j, W_j, σ_j, $j = 1,\ldots, J$ by (11–13).

$$\Delta\mu_j = -\rho_\mu \nabla_{\mu_j} E \tag{11}$$

$$\Delta\sigma_j = -\rho_\sigma \frac{\partial E}{\partial \sigma_j} \tag{12}$$

$$\Delta w_j = -\rho_w \frac{\partial E}{\partial w_j} \tag{13}$$

where ρ_μ, ρ_σ, and ρ_w are small positive constants.
Step 4 Update the parameters μ_j, W_j, σ_j until the algorithm are trained by all the training data $(x_1,\ldots, x_n)$.
Step 5 Evaluate $\hat{Y}(x_1),\ldots,\hat{Y}(x_n)$by final estimated parameters and calculate the RMSE by (14).

$$RMSE\sqrt{\frac{\sum_{i=1}^{n}(Y(x_i)-\hat{Y}(x_i))^2}{n}} \tag{14}$$

Step 6 If RMSE > ε then goes to step 2.

3 Proposed e-learning achievement evaluation model

From the literature reviews in Sect. 1, there are two drawbacks in past achievement evaluation methods:(1) specific statistic assumptions about data distributions is made in statistical methods, and they can not be applied to each type of dataset; and (2) the evaluation methods employing fuzzy set theory produce evaluations mostly based on expert opinions, and it is difficult to explore and utilize the valuable information embedded in collected data. To improve the past methods, two advanced data mining techniques are proposed in achievement evaluation: (1) use MEPA as discretization method; and (2) use RBF-NN as classification model.

There are many advantages to employ MEPA in data preprocess. For example, complex data dimensions in datasets can be reduced and simplified, and the size of dataset with discrete features is usually more compact and smaller than the one with continuous features [37]. In addition, MEPA discretization method is a more objective way to construct membership functions than the past methods utilizing expert opinions to build membership functions. RBF-NN is a simple and fast learning network to build model for datasets, and the nonlinear structures of the neural networks can process data without any statistic assumption for the observations in target dataset.

To provide more objective and proper achievement evaluations than past methods for e-learning system, this paper proposes a new evaluation model, which incorporates MEPA and RBF-NN in evaluation processes. The overall framework of the proposed model (shown as Fig. 1) contains three main phases of achievement evaluation processes as follows. (1) Preprocess: use feature selection to reduce dataset dimension and use MEPA discretization method to construct a more objective and reasonable membership functions of features. (2) Similarity filter: apply similarity threshold to screen inconsistent data and make our achievements evaluation model more compact. (3) Construct classification model and accuracy evaluation: use RBF-NN to construct classification model and evaluate accuracy. To introduce the proposed model, each step contained in three main phases is described in the following.

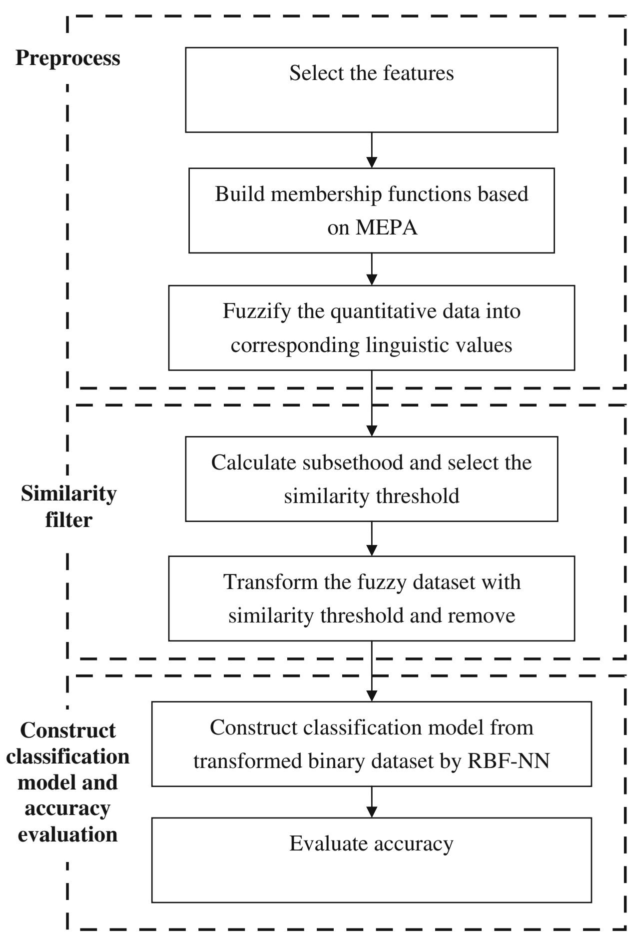

Fig. 1 Framework of the proposed model

Step 1 Select the features

In this step, well-known feature selection methods such as Cfs, Chi-square, Consistency, Gain Ratio, InfoGain, OneR, Relief and Symmetrical uncertainty are employed, and reduction features with consistency are produced by analyzing different results from the feature selection methods. Because irrelevant and redundant features may degrade the performance of data mining algorithms, this step selects relevant features by evaluating usefulness of features and eliminate irrelevant features to simplify classification rules.

Step 2 Build membership functions based on MEPA

This step utilizes MEPA to build membership function for each selected feature from step 1. The entropy value of each quantitative data is computed by (1). By repeating this procedure to subdivide the all features of the data set, the thresholds can be obtained. If $n(x)$ (the total number of samples located in $[x_1, x_1+x]$) and $N(x)$ (the total number of samples located in $[x_1+x, x_2]$) equal to zero, then data within the range are not further subdivided. The thresholds are used as the midpoint of membership function. If the feature value is lower than SEC1 (secondary threshold value), then the membership degree equals to one. Further, the membership degree equals to one when feature value exceeds SEC2 (secondary threshold value). The membership function of minimal entropy approach can be established.

Step 3 Fuzzify the quantitative data into corresponding linguistic values

Use the membership functions generated in step 2, the membership degree for each observation is calculated to determine its corresponding linguistic value. A quantitative dataset is discretized and converted into a corresponding linguistic dataset by MEPA.

Step 4 Calculate subsethood and select the similarity threshold

There are two sub-steps in this step. First, in order to find the closeness between the outcome feature and each term of the selected features, the subsethood values are calculated for them. Based on (8), in this step, the subsethood value is defined by (15):

$$S(DD, A_i) = \frac{M(\mu_{DD} \cap \mu_{A_i})}{M(\mu_{DD})} \tag{15}$$

where $S(DD, A_i)$ denotes subsethood value, DD is the outcome label, A_i is the linguistic value of feature, $M(\mu_{DD})$ is the sum of membership degree in the same outcome label, $M(\mu_{DD} \cap \mu_{A_i})$ denotes the sum of min (μ_{DD}, μ_{A_i}), μ_{A_i} is feature membership degree, and $\mu_{DD}$is outcome membership degree for each record.

Second, select the maximal subsethood value as similarity threshold (S_{T_i}). The similarity threshold is defined as (16).

$$S_{T_i} = \mathrm{Max}\{S(DD, A_i)\} \quad (16)$$

Step 5 Transform the fuzzy dataset with similarity threshold and remove inconsistent data

There are two sub-steps in this step. First, the fuzzified dataset (from step 3) is transformed into binary values with similarity threshold (from step 4). The transformational method is defined by (17):

$$\phi_i = \begin{cases} 1 & I_i \geq S_{Ti} \\ 0 & I_i < S_{Ti} \end{cases} \quad (17)$$

where I_i denotes membership grade of fuzzified dataset, ϕ_i is the transformed binary value, and if the membership grade of fuzzified value is great than or equal to the similarity threshold, then this step transforms the membership grade to ‘1’, otherwise ‘0’.

Second, if this step detects the inconsistent records (binary values in the same feature are all zero. For example, in Table 7, the binary values in each linguistic value of assignment feature are all zero for student No. 4 and student No. 5), then removes these records to make achievements evaluation more compact.

Step 6 Construct classification model from transformed binary dataset by RBF-NN

In this step, use RBF-NN to construct a classification model with the transformed binary dataset derived in step 5. The classification model is used to assign student grade.

Step 7 Evaluate accuracy

In the last step, the classification model derived from step 6 is evaluated. The accuracy of the proposed model is taken to compare with some existing models to evaluate model performance.

4 Empirical case study

An empirical case study is provided in the section to illustrate the proposed model, and a collected e-learning achievement dataset is used as experimental dataset to demonstrate the practicability of the proposed model.

4.1 Experimental dataset

The practical e-learning achievement dataset (see Table 1) is collected from an e-learning online examination system of a university in Taiwan. There are five conditional features contained in the dataset: (1) the frequency of online e-learning (the frequency of reviewing the content of online e-learning over 30 min); (2) the frequency of discussion by e-learning (the frequency of discussing the content of course in online discussion forum); (3) assignment; (4) test; and (5) final exam. Besides, there is one decision feature in the dataset with five grades of student achievement: Excellent (A), Good (B), Average(C), Satisfactory (D), and Unsatisfactory (E). The dataset consists of 115 records (students), and the course name is “system analysis and design”.

4.2 Demonstration of the proposed model

The steps of the achievement evaluation processes (such as select the features, screen inconsistent data and calculate accuracy) by using practical collected e-learning achievement dataset are as follows:

Step 1 Select the features

This step applies eight well-known feature selection methods, Cfs, Chi-square, Consistency, Gain Ratio, InfoGain, OneR, Relief and Symmetrical uncertainty to select features from the achievement dataset. The feature selection results for the eight feature selection methods are shown in Table 2 and items 1–5 denote the ranking of features according to each selection method from 1 (first ordering) to 5 (last ordering). From Table 2, we can see that the first three important features derived by the eight above methods are “assignment”, “test” and “final exam”. Therefore, from the results of feature selection methods, the three key features are “assignment”, “test” and “final exam” as input features for next step.

Step 2 Build membership functions based on MEPA

From step 1, the selected features are assignment, test and final exam. Then this step utilizes MEPA to build membership function for each feature. The entropy values of each data are computed using (1) proposed by MEPA. By repeating this procedure to subdivide the

Table 1 The partial of student achievement dataset

Student No.	The frequency of online e-learning	The frequency of discussing by e-learning	Assignment	Test	Final exam	Grade
1	20	1	5.00	37.00	18.00	E
2	88	18	10.00	23.00	16.00	E
3	46	5	15.00	13.00	6.00	E
4	30	3	40.00	13.00	20.00	E
5	12	2	25.00	31.00	14.00	E
115	35	10	90	90	98	A

Table 2 A comparison of the ranking of features by eight feature selection methods

Feature method	Cfs	Chi-square	Consistency	Gain Ratio	InfoGain	OneR	Relief	Symmetrical uncertainty
The frequency of online e-learning	NA	4	NA	4	4	4	4	4
The frequency of discussing by e-learning	NA	5	NA	5	5	5	5	5
Assignment	3	3	1	2	3	2	3	2
Test	2	2	2	3	2	3	2	3
Final exam	1	1	3	1	1	1	1	1

1–5 denote feature selection ranking by eight different selection methods, and 1 refers to the most important. NA denotes no given answer in this method

Table 3 Threshold of features

	MIN	SEC1	PRI	SEC2	MAX
Assignment	1	15.5	56	60.5	100
Test	10	25.5	46	77.5	98
Final exam	1	32	47	73.5	99

data, the thresholds can be obtained. As Table 3 shows, if $n(x)$ and $N(x)$ equal to zero, then stop subdividing the data in the range.

The thresholds (see Table 3) are used as the midpoint of membership function. When the feature value is lower than SEC1, then the membership degree equals 1. Similarly, the membership degree equals 1 when feature value exceeds SEC2. The membership function of MEPA can be established which is shown in Fig. 2

Step 3 Fuzzify the quantitative data into unique corresponding linguistic values

According to the membership function in step 2, the degree of membership for each datum is calculated. Table 4 shows the partial result of the student achievement dataset.

Step 4 Calculate subsethood and select the similarity threshold

First, to find the closeness between the outcome feature and each term of the selected features, we calculate the subsethood of them. From (15), the subsethood values can be computed from fuzzified student achievement dataset (see Table 5).

Second, the selected similarity thresholds for each linguistic value are obtained by (16) and are listed in Table 6. For example, the similarity threshold value of assignment in 'Bad' linguistic value can be selected as following.

$$\begin{aligned}&\mathrm{Max}\{\mathrm{S(A;Bad)},\mathrm{S(B;Bad)},\mathrm{S(C;Bad)},\mathrm{S(D;Bad)},\\&\quad\mathrm{S(E;Bad)}\}=\mathrm{Max}\{0,0.19,0.47,0.59,0.91\}=0.91\end{aligned}$$

Step 5 Transform the fuzzy dataset by similarity threshold and remove inconsistent data

According to the similarity threshold values in step 4, this step transforms the fuzzified dataset (from step 3) into binary values, shown in Table 7.

For example, the membership grades of fuzzified dataset for feature 'Test' in student No. 1 are

$$I_{\mathrm{Bad}}=0.44, I_{\mathrm{Normal}}=0.56, I_{\mathrm{Good}}=0.$$

The similarity thresholds for these three linguistic values are

$$S_{T\mathrm{Bad}}=0.76, S_{T\mathrm{Normal}}=0.5, S_{T\mathrm{Good}}=0.98.$$

The membership grade of fuzzified value

$$(I_{\mathrm{Bad}}=0.44)<(S_{T\mathrm{Bad}}=0.76).$$

So $\phi_{\mathrm{Bad}}=0$ [from (17)].

In the same way, $\phi_{\mathrm{Normal}}=1$ and $\phi_{\mathrm{Good}}=0$.

Next, this step detects inconsistent data (the binary values in the same feature are all zero) and removes these data to make achievements evaluation more compact. From Table 7, we can see that the binary

Table 4 The partial of fuzzied student achievement dataset

Student No.	Assignment			Test			Final exam			Grade				
	Bad	Normal	Good	Bad	Normal	Good	Bad	Normal	Good	A	B	C	D	E
1	1	0	0	0.44	0.56	0	1	0	0	0	0	0	0	1
2	1	0	0	1	0	0	1	0	0	0	0	0	0	1
3	1	0	0	1	0	0	1	0	0	0	0	0	0	1
4	0.4	0.6	0	1	0	0	1	0	0	0	0	0	0	1
5	0.77	0.23	0	0.73	0.27	0	1	0	0	0	0	0	0	1
115	0	0	1	0	0	1	0	0	1	1	0	0	0	0

Table 5 The subsethood values

Assignment	Test	Final exam
Subgroup1 (A)		
S(A; Bad) = 0	S(A; Bad) = 0	S(A; Bad) = 0
S(A; Normal) = 0.04	S(A; Normal) = 0.02	S(A; Normal) = 0
S(A; Good) = 0.96	S(A; Good) = 0.98	S(A; Good) = 1
Subgroup2 (B)		
S(B; Bad) = 0.19	S(B; Bad) = 0	S(B; Bad) = 0
S(B; Normal) = 0.48	S(B; Normal) = 0.29	S(B; Normal) = 0.28
S(B; Good) = 0.33	S(B; Good) = 0.71	S(B; Good) = 0.72
Subgroup3 (C)		
S(C; Bad) = 0.47	S(C; Bad) = 0.11	S(C; Bad) = 0.07
S(C; Normal) = 0.53	S(C; Normal) = 0.4	S(C; Normal) = 0.57
S(C; Good) = 0	S(C; Good) = 0.5	S(C; Good) = 0.35
Subgroup4 (D)		
S(D; Bad) = 0.59	S(D; Bad) = 0.48	S(D; Bad) = 0.56
S(D; Normal) = 0.37	S(D; Normal) = 0.38	S(D; Normal) = 0.29
S(D; Good) = 0.04	S(D; Good) = 0.14	S(D; Good) = 0.15
Subgroup5 (E)		
S(E; Bad) = 0.91	S(E; Bad) = 0.76	S(E; Bad) = 1
S(E; Normal) = 0.09	S(E; Normal) = 0.24	S(E; Normal) = 0
S(E; Good) = 0	S(E; Good) = 0	S(E; Good) = 0

Table 6 The similarity thresholds

	Assignment			Test			Final exam		
	Bad	Normal	Good	Bad	Normal	Good	Bad	Normal	Good
Similarity threshold	0.91	0.53	0.96	0.76	0.5	0.98	1	0.57	1

values for student No. 4 and student No. 5 in each linguistic value of assignment feature are all zero; therefore theses two records are removed.

Step 6 Construct classification model from transformed binary dataset by RBF-NN

This step utilizes RBF-NN to construct classification model from transformed binary dataset (see Table 7). Then students' grade can be assigned by the produced classification model.

4.3 Performance evaluation of the proposed model

To evaluate the proposed model in detail, the experimental dataset (115 records) is separated into three sub-datasets: sub-dataset A (50 records), sub-dataset B (50 records) and sub-dataset C (15 records), and three types of experimentations are provided: (1) sub-dataset A and sub-dataset B are used for training, and sub-dataset C is used for testing; (2) sub-dataset A is used for training, and sub-dataset B is

Table 7 The transformed dataset based on similarity threshold

Student No.	Assignment			Test			Final exam			Grade				
	Bad	Normal	Good	Bad	Normal	Good	Bad	Normal	Good	A	B	C	D	E
1	1	0	0	0	1	0	1	0	0	0	0	0	0	1
2	1	0	0	1	0	0	1	0	0	0	0	0	0	1
3	1	0	0	1	0	0	1	0	0	0	0	0	0	1
4	0	0	0	1	0	0	1	0	0	0	0	0	0	1
5	0	0	0	0	0	0	1	0	0	0	0	0	0	1
115	0	0	1	0	0	1	0	0	1	1	0	0	0	0

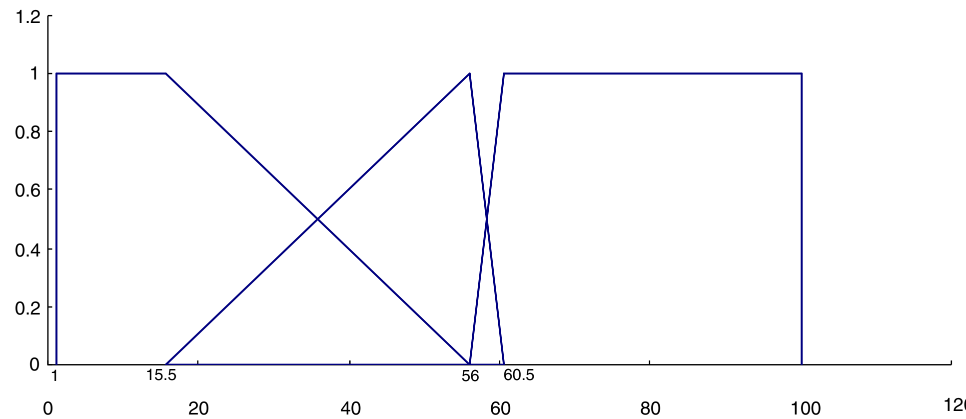

Fig. 2 Membership function of assignment

used for testing; and (3) sub-dataset B is used for training, and sub-dataset A is used for testing.

To verify the performance carefully, the classification accuracy produced by the proposed model is compared with the data from some other classification methods such as RBF-NN [14], rough set theory (RST) [38], decision tree (C4.5) [39] and BayesNet method [40]. The different classification accuracies for the proposed model and comparison methods in three experimentations are listed in Table 8.

From the experimental results, it is shown that the classification accuracies in different experiments under the same classification method are not the same, and the accuracy order for five method (excluding RST) is listed as follows: accuracy in experiment 1 > accuracy in experiment 3 > accuracy in experiment 2. It tells that using different ratios of training to testing for experimental datasets influences the model accuracy. Further, among five classification methods, the proposed model performs best in the three types of experimentations including average accuracy (74.9% for Decision Tree; 71.6% for BayesNet; 69.4% for RST; 80.4% for RBF-NN; and 90.2% for the proposed model). From the overall performance evaluation, it is concluded that the proposed model surpasses the listing models in classification accuracy, and provides more proper achievement evaluations for the learners contained in experimental dataset.

5 Findings and conclusions

In asynchronous e-learning system for self-pace learning, objective and proper achievement evaluations for learners are especially necessary. In the system, teachers have to upload their e-learning contents to system and students can receive their learning content based on their learning paces at any time and any place. Therefore, learning achievement evaluation for each learning stage, contained in a whole course designed by teachers, is the key part to decide the learning pace for students, and objective and proper achievement evaluation model is the critical subsystem to examine whether students have reached the learning objectives each learning stage. Besides, automatic achievement evaluation model can reduce the evaluation works for teachers and avoid the individual preference from teachers. Based on these important points in e-learning system above, in this paper, we have proposed a new machine-learning model to apply in e-learning system to produce objective and proper achievement evaluations for learners. From the empirical case study, four conclusions are given in this paper as follows.

1. Using machine-learning based methods to assign achievements can avoid individual preference in achievement evaluation process. We argue that

Table 8 Results of different classification methods

Experiment	Classification method				
	Decision Tree (C4.5) (%)	BayesNet (%)	RST (%)	RBF-NN (%)	The proposed (RBF-NN with MEPA) (%)
Experiment 1[a]	86.7	86.7	83.3	93.3	100
Experiment 2[b]	66	60	75	64	80
Experiment 3[c]	72	68	50	84	90.5
Average	74.9	71.6	69.4	80.4	90.2

[a] Sub-datasets-A and sub-datasets-B for training, and sub-datasets-C for testing

[b] Sub-datasets-A for training and sub-datasets-B for testing

[c] Sub-datasets-B for training and sub-datasets-A for testing

individual preference from teacher can be avoided if teacher judgments are not involved in achievement evaluation process. Therefore, in the experiment for the empirical case, a machine-learning based method (RBF-NN) is employed in the processes of the proposed model to produce achievement evaluations, and each process of the model deal with data only.

2. Refining classification model with proper preprocess methods such as MEPA and similarity filter can provide more proper achievement evaluations (higher classification accuracy) for learners than common classification models. From the performance evaluation section (see Table 8), it is clear that the proposed model surpasses the listing models, Decision Tree, BayesNet, RST, and RBF-NN, in classification accuracy. Besides, the classifications for the learners' achievements from the proposed model are approved as reasonable achievement evaluations by the instructor in the case study.
3. Employing MEPA to build membership function can improve accuracy. The proposed model utilizes MEPA to fuzzify quantitative data and build membership function without employing expert's opinions. From Table 8, the RBF-NN using MEPA discretization method (the proposed model) has surpassed the same method without using discretization method (average accuracy of RBF-NN is 80.4%; average accuracy of RBF-NN with MEPA is 90.2%). From the evidence, it is shown that using discretization method in preprocess can improve accuracy for classification model.
4. Applying similarity threshold in preprocess can filter inconsistent data to make data mining process more efficient. Using similarity threshold to remove inconsistent data can make the size of dataset more compact, and therefore make RBF-NN learning method to more efficient in building classification model.

In the future works, there are two directions to improve the proposed model to get better accuracy rate: (1) employ other artificial intelligence techniques classification methods such as support vector machine [41] in the proposed model; and (2) employ different data discretization methods such as global discretization method [42] in preprocessing step to partition features.

References

1. Rasmani KA, Shen Q (2006) Data-driven fuzzy rule generation and its application for student academic performance evaluation. Appl Intell 25:305–319. doi:10.1007/s10489-006-0109-9
2. Wang JW, Cheng CH, Huang KC (2009) Fuzzy hierarchical TOPSIS for supplier selection. Appl Soft Comput 9(1):377–386. doi:10.1016/j.asoc.2008.04.014
3. Teoh HJ, Cheng CH, Chu HH, Chen JS (2008) Fuzzy time series model based on probabilistic approach and rough set rule induction for empirical research in stock markets. Data Knowl Eng 67(1):103–117. doi:10.1016/j.datak.2008.06.002
4. Cheng CH, Chen YS, Wu YL (2009) Forecasting innovation diffusion of products using trend-weighted fuzzy time-series model. Expert Syst Appl 36(2):1826–1832. doi:10.1016/j.eswa.2007.12.041
5. Biswas R (1995) An application of fuzzy sets in students' evaluation. Fuzzy Sets Syst 74(2):187–194. doi:10.1016/0165-0114(95)00063-Q
6. Echauz JR, Vachtsevanos GJ (1995) Fuzzy grading system. IEEE Trans Educ 38(2):158–165. doi:10.1109/13.387218
7. Law CK (1996) Using fuzzy numbers in educational grading system. Fuzzy Sets Syst 83:311–323. doi:10.1016/0165-0114(95)00298-7
8. Cheng CH, Yang KL (1998) Using fuzzy sets in education grading system. J Chin Fuzzy Syst Assoc 4(2):81–89
9. Chen SM, Lee CH (1999) New methods for students' evaluating using fuzzy sets. Fuzzy Sets Syst 104(2):209–218. doi:10.1016/S0165-0114(97)00208-X
10. Ma J, Zhou D (2000) Fuzzy set approach to the assessment of student-centered learning. IEEE Trans Educ 43(2):237–241. doi:10.1109/13.848079
11. Weon S, Kim J (2001) Learning achievement evaluation strategy using fuzzy membership function. In: Proceedings of the 31st ASEE/IEEE frontiers in education conference, Reno, NV, vol 1, pp 19–24
12. Bai SM, Chen SM (2008) Evaluating students' learning achievement using fuzzy membership functions and fuzzy rules. Expert Syst Appl 34:399–410. doi:10.1016/j.eswa.2006.09.010
13. Christensen R (1980) Entropy minimax sourcebook, general description. Entropy Limited, Lincoln, MA
14. Moody J, Darken CJ (1989) Fast learning in networks of locally tuned processing units. Neural Comput 1:281–294. doi:10.1162/neco.1989.1.2.281
15. Bishop CM (1995) Neural network for pattern recognition. Oxford University Press, New York
16. Guillén A, Pomares H, Rojas I, González J, Herrera LJ, Rojas F, Valenzuela O (2008) Studying possibility in a clustering algorithm for RBFNN design for function approximation. Neural Comput Appl 17:75–89. doi:10.1007/s00521-007-0134-6
17. Haddadnia J, Faez K, Ahmadi M (2003) A fuzzy hybrid learning algorithm for radial basis function neural network with application in human face recognition. Pattern Recognit 36:1187–1202. doi:10.1016/S0031-3203(02)00231-5
18. Lee SJ, Hou CL (2002) An ART-based construction of RBF networks. IEEE Trans Neural Netw 13:1308–1321. doi:10.1109/TNN.2002.804308
19. Sharma AK, Sharma RK, Kasana HS (2006) Empirical comparisons of feed-forward connectionist and conventional regression models for prediction of first lactation 305-day milk yield in Karan Fries dairy cows. Neural Comput Appl 15:359–365. doi:10.1007/s00521-006-0037-y
20. Zhang D, Deng LF, Cai KY, So A (2005) Fuzzy nonlinear regression with fuzzified radial basis function network. IEEE Trans Fuzzy Syst 13(6):742–760. doi:10.1109/TFUZZ.2005.859307
21. Tan PN, Steinbach M, Kumar V (2006) Introduction to data mining. Addison-Wesley, New York
22. Witten IH, Frank E (2005) Data mining: practical machine learning tools and techniques, 2nd edn. Morgan Kaufmann Publishers, USA
23. Hall MA (1998) Correlation-based Feature Subset Selection for Machine Learning. PhD thesis, University of Waikato

24. Kononenko I (1994) Estimating attributes: analysis and extensions of RELIEF. In: Proceedings of the European conference on machine learning, pp 171–182
25. Liu H, Setiono R (1996) A probabilistic approach to feature selection—a filter solution. In: Proceedings of the 13th international conference on machine learning, pp 319–327
26. Duda RO, Hart PE, Stork DG (2001) Pattern classification. Wiley, New York
27. Holte RC (1993) Very simple classification rules perform well on most commonly used datasets. Mach Learn 11(1):63–90. doi: 10.1023/A:1022631118932
28. Kira K, Rendell LA (1992) A practical approach to feature selection. In: Sleeman D, Edwards P (eds) Proceedings of the international conference on machine learning, Morgan Kaufmann, San Mateo, pp 249–256
29. Press WH, Flannery BP, Teukolski SA, Vetterling WT (1998) Numerical recipes in C. Cambridge University Press, Cambridge
30. Kim CJ, Russell BD (1993) Automatic generation of membership function and fuzzy rule using inductive reasoning. In: Proceedings of the industrial fuzzy control and intelligent systems, Houston
31. Yager R, Filev D (1994) Template-based fuzzy system modeling. Intell Fuzzy Syst 2:39–54
32. Ross TJ (2004) Fuzzy logic with engineering applications. Wiley, USA
33. Zadeh LA (1965) Fuzzy sets. Inf Contr 8:338–353. doi:10.1016/S0019-9958(65)90241-X
34. Yuan Y, Shaw MJ (1995) Induction of fuzzy decision trees. Fuzzy Sets Syst 69(2):125–139. doi:10.1016/0165-0114(94)00229-Z
35. Chen SM, Lee SH, Lee CH (2001) A new method for generating fuzzy rules from numerical data for handling classification problems. Appl Artif Intell 15(6):645–664
36. Kosko B (1986) Fuzzy entropy and conditioning. Inf Sci 40(2):165–174. doi:10.1016/0020-0255(86)90006-X
37. Liu H, Hussain F, Tan C, Dash M (2002) Discretization: an enabling technique. Data Min Knowl Discov 6(4):393–423. doi: 10.1023/A:1016304305535
38. Pawlak Z (1982) Rough Sets. Inf J Comput Inf Sci 11(5):341–356. doi:10.1007/BF01001956
39. Quinlan JR (1993) C4.5: programs for machine learning. Morgan Kaufmann Publishers, San Mateo
40. Murphy KP (2002) Bayes Net ToolBox, Technical Report, MIT Artificial Intelligence Laboratory, http://www.ai.mit.edu/~murphyk/
41. Cortes C, Vapnik V (1995) Support vector network. Mach Learn 20:273–297
42. Polkowski L, Tsumoto S, Lin T (2000) Rough set methods and applications. Physica-Verlag, Heidelberg, pp 49–88