Research Article

International Journal of Advanced Technology and Engineering Exploration, Vol 5(42)
ISSN (Print): 2394-5443 ISSN (Online): 2394-7454
http://dx.doi.org/10.19101/IJATEE.2018.542014

A fuzzy based enhancement on prism and J48 classifier prediction of student performance

Sasi Regha. R[1*] and Uma Rani. R[2]
Department of Computer Science, SSM College of Arts and Science, Komarapalayam, Namakkal, India[1]
Department of Computer Science, Sri Sarada College for Women, Salem, India[2]

Abstract

A modified computed aided design of experiments (MCADEX) using Kullback-Leibler divergence and modified principal component analysis (MPCA) was proposed for selecting set of samples to improve the prediction of student's performance in prism and J48 classifiers. The classification accuracy of prism and J48 is required to enhance further for classifying the students dataset which have complex related attributes. Hence, a fuzzy neuro prism (FNP) and fuzzy neuro J48 (FNJ48) are introduced for improving the classification accuracy. A fuzzy system is using fuzzy if then rules in obtaining knowledge from human experts can deal with imprecise problems. These rules are generating for describing the relationship among the input attribute space and classes. In fuzzification, Gaussian membership function is used. In this method, the weight value of each attribute is calculated using neural network. Fuzzy membership function parameters are optimized by using Cuckoo search algorithm. The attribute with maximum weight value and fuzzified value of features are used for constructing tree of prism and J48 classifiers. The experimental results show that the proposed approach is providing better results in terms of accuracy, true positive rate and true negative rate.

Keywords

Modified computed aided design of experiments, Modified principal component analysis, Fuzzy neuro prism, Fuzzy neuro J48, Classifiers.

1.Introduction

The software requirement prioritization is a process in which requirement Data mining (DM) in higher education is forming a new research field named education DM [1, 2]. The application of DM to education allows the educators to find novel and useful knowledge about students [3]. Educational DM develops techniques to explore the kinds of data which come from educational organizations. There are numerous DM approaches, consists of statistics and visualization, outlier detection, clustering and classification. Among these, classification is one of the most frequently studied approaches. Classification is a procedure of supervised learning where data is divided into various classes. The objective of a classification model is to predict the target class for every sample in the dataset. There are different approaches for classification of data, including support vector machine (SVM), artificial neural network (ANN) and Bayesian classifier methods [4]. These methods are extensively applied in educational environments for student's performance prediction.

The ability to predict a student's performance is very important in educational environments. Student's performance is based upon diverse factors like personal, social, psychological and other environmental variables. In existing system, MCADEX method [5] was proposed using Kullback-Leibler divergence for enhancing accuracy of the student's performance prediction. In this method, MPCA was used for deriving the eigenvectors of covariance matrix from various similarity measurements. The sample selected datasets are used to predict student performance by using prism and J48 classifiers.

In this paper, the classification accuracy is improved by using FNP and FNJ48. In this approach, fuzzy rules are generating for defining the relationship between the input feature space and classes. Gaussian membership function is used in fuzzification. Fuzzy membership function parameters are optimized based on Cuckoo search algorithm. By using neural network, the weight value of every attribute is measured. The fuzzified values of features and attribute with maximum weight value are used to construct tree of prism and J48 classifiers.

*Author for correspondence

The remainder of the article is organized as follows: Section 2 describes about the existing fuzzy methods. Section 3 describes about the proposed method. Section 4 demonstrates the overall performance evaluation of the proposed approach. Section 5 concludes the article work.

2.Related work

A novel student classification model [6] was designed by using neuro fuzzy method for predicting student's academic performance. The main step in the model was the collection of data, processing of data, student classification and an evaluation of the classification model. An adaptive fuzzy technique [7] was proposed to monitor a student's facilitate the educational process. This technique was enabled the successful modeling of observed relationships between the basic elements and emotional labels (affective state). This technique was provided for creating generic models. However in collaborative learning task, monitor the groups of student's was required of this approach.

A neuro-fuzzy method and improved scaled conjugate gradient (SCG) [8] was presented for classifying students. The improved SCG technique was used for optimization process. The feature space was divided into multiple fuzzy subspaces using fuzzy if-then rules in the neuro-fuzzy scheme. These rules were represented using a network structure. The classifier was contained multiple inputs and outputs. A fuzzy logic computational model [9] was presented for classifying the performance of academic staff and Sudanese universities. The novel approach was introduced a consistent preference linguistic value as an option to the experts in case of inconsistent judgment in evaluation performance. By using this approach, a novel tool was proposed which was allowed to trace and recognize the roots of inconsistency and choose the relevant consistent option.

A fuzzy decision tree and a cased based data clustering [10] were integrated for data classification in medical field. A fuzzy decision tree was applied to the data in every cluster and genetic algorithms were used to construct a decision-making system by the selected features. A case-based clustering method was utilized to preprocess the dataset and a set of fuzzy rules was generated for every cluster. A new neuro-fuzzy classification approach [11] was presented for extracting the feature. The inputs to the Neuro-fuzzy were fuzzified by using generalized bell-shaped membership function. This approach was used a fuzzification matrix in that the input patterns were related to a degree of membership to various classes. By using the value of degree of membership a pattern would be attributed to a specific class.

An online ensemble of classifiers with Navie Bayes, 1-NN and the WINNOW approaches [12] was proposed for predicting the student's performance in a distance learning system. In predicting student's performance, the application of this approach was provided to be useful for discovering poor performers. It was enabled tutors to take precautionary measures at an earlier stage, even from the beginning of an academic year. A decision tree and fuzzy genetic method [13] were proposed for predicting the student's performance. By using this approach, prediction of student's academic performance was measured in bachelors and master's degree for every subject. Thus, the lectures were classified the student's for enhancing their performance.

3.Proposed methodology

A FNP and FNJ48 are introduced to enhance the classification accuracy. Fuzzy rules are generating for describing the relationship between the input attribute space and classes. Gaussian membership function is used in fuzzification. By using Cuckoo search algorithm, fuzzy membership function parameters are optimized. In this method, the weight value of each attribute is calculated using neural network. The fuzzified value of features and attribute with maximum weight value and are used for constructing tree of prism and J48 classifiers.

3.1Decision trees

(a) Prism

The objective is to induce modular classification rules directly from the training set. The prism algorithm is generated the rules concluding each of the possible classes in turn. Each rule is generated term by term with each term of the form 'attribute = value'. The attribute pair added at every step is chosen to maximize the probability of the target 'result class'. In this algorithm, α_X denotes an attribute-value pair and Ci indicates a specific classification.

The amount of information about occurrence of Ci given α_X as follows,

$$I(Ci, \alpha_X) = \log_2 \left(\frac{P(Ci|\alpha_X)}{P(Ci)}\right) bits \tag{1}$$

In the above equation, $P(Ci|\alpha_X) = \frac{Number\ of\ instances\ labeled\ Ci}{|S_{\alpha_X}|}$, S_{α_X} denotes the subset of instances contain α_X, the probability of occurrence of Ci after knowing α_X in S_{α_X}. $P(Ci) = \frac{Number\ of\ instances\ labeled\ Ci}{|S|}$, the probability of occurrence of Ci before knowing α_X in S.

Prism algorithm steps:
Input: A training dataset with n classes $C_i, i = 1, 2, 3, \ldots, n$

Output: Created rules for the entire classes
1. For every class Ci start with the complete training set every time
2. Calculate the probability of every attribute/ value pair for the class, Ci
3. Choose the pair with the largest probability and generate a subset of the training set comprising the entire instances with the chosen attribute/value combination for every class, Ci
4. Go to steps 2 and 3 for this subset until a subset is reached that have only instances of Ci
5. The rule is induced using the conjunction of all the attribute/value pairs chosen
6. Eliminate the entire instances covered through this rule from the training set S
7. Go to step 2 through 6 until the entire instances of Ci have been eliminated
8. Go to step 1 until the entire classes are examined

(b) J48
J48 classifier is a simple C4.5 decision tree for classification. This classifier creates a binary tree. Decision trees are created in the J48 algorithm based on information entropy on a set of training data. In this approach, a tree is constructed to model the classification process. Data attributes are organized into subsets and the normalized information gain, computed using the difference in entropy, is utilized to calculate these subsets to identify the best attributes used as nodes in the decision tree.
The entropy of $\underset{y}{\rightarrow}$ is,

$$Entropy\left(\underset{y}{\rightarrow}\right) = -\sum_{m=1}^{n} \frac{|y_m|}{\left|\underset{y}{\rightarrow}\right|} \log\left(\frac{|y_m|}{\left|\underset{y}{\rightarrow}\right|}\right) \quad (2)$$

$$Entropy\left(m\middle|\underset{y}{\rightarrow}\right) = \frac{|y_m|}{\left|\underset{y}{\rightarrow}\right|} \log\left(\frac{|y_m|}{\left|\underset{y}{\rightarrow}\right|}\right) \quad (3)$$

Gain is computed as follows,

$$Gain\left(\underset{y}{\rightarrow}, m\right) = Entropy\left(\underset{y}{\rightarrow} - Entropy\left(m\middle|\underset{y}{\rightarrow}\right)\right) \quad (4)$$

The main aim is to increase the gain, splitting through entropy because of separate argument $\underset{y}{\rightarrow}$ using value m. A classification is executed on the instances of the training set and tree is formed. The pruning is executed for reducing classification errors that are being created using specialization in the training set. Pruning is executed for the generalization of the tree.

J48 algorithm steps:
Input: Training data
Output: Decision tree
1. In case the instances belong to the same class the tree represents a leaf so the leaf is returned by labeling with the same class. (leaf node represents class labels)
2. The potential information is computed for every attribute, given by a test on the attribute (internal node). After that, the gain in information is computed in equation (4) that would result from a test on the attribute.
3. The best attribute is found on the basis of the present selection criterion and that attribute chosen for branching (result of the test).

3.2Fuzzy logic
A fuzzy rule has two components: an if-part (antecedent) and a then-part (consequent). A fuzzy classification rule R_i that shows the relation among the classes and input feature space,

$$R_i{:}\ if\ x_1\ is\ A_{i1}\ and \ldots x_j\ is\ A_{ij} \ldots and\ x_n\ is\ A_{in}, then \quad (5)$$

$$f_i = q_{i1}x_1, +q_{i2}x_2 + \cdots + q_{in}x_n + q_{i(n+1)} \quad (6)$$

In the above equation, x_1 symbolizes the jth feature. A_{ij} symbolizes the fuzzy group of the jth feature within the ith rule. A_{ij} represents recognized through the appropriate membership function. The parameters $\left(q_{ij}, where\ 1 \le i \le \#\,(rules)\ and\ 1 \le j \le \#\,(inputs)\right)$ be the coefficients in the consequent part.

A training dataset values is converted into fuzzy dataset values using membership function. A smooth and concise fuzzification is achieved in classification through executing a Gaussian membership function (μ_{ij}) which retains the property of overlapping its membership values within various linguistic classes along every feature axis. Every node i is a fuzzification with a node function, where x_j denotes the input to node i. It is denoted using two parameters $\{c, \sigma\}$,

$$\mu_{ij} = \mu_i(x_j) = exp\left\{-\frac{1}{2}\left(\frac{x_j - c_{ij}}{\sigma_{ij}}\right)^2\right\} \qquad (7)$$

In the above equation, c_{ij} *and* σ_{ij} denotes the center and width respectively of the j^{th} attribute of the i^{th} MF. Any attribute value is defined in terms of some combination of overlapping membership values (MFs) in the linguistic sets Low (L), Medium (M), High (H) and Critical (C). The minimum value of the membership functions is activated in a rule and is called as firing strength.

$$w_i = \min(\mu_{ij}) \qquad (8)$$

The i^{th} node is computed the percentage of the i^{th} rule's firing strength to the sum of the entire rule's firing strength,

$$\overline{w_i} = w_i / \sum_i w_i \qquad (9)$$

The output of the consequent part consists of three factors such as a scaling factor k, a normalized firing strength of the antecedent part $\overline{w}$ and a linear function of input variables with the consequent coefficients. In each attribute, the weight value is computed as follows,

$$F: \overline{w_i} \rightarrow \overline{w_i} k\left(\sum_{j=1}^{\#(inputs)} q_{ij} x_j + q_{i(j+1)}\right) \qquad (10)$$

In defuzzification, the consequent membership function F_i is changed into a crisp consequent $s = d_i$ where d_i denotes the center of gravity of F_i. It is expressed as,

$$s = \frac{\sum_{i=1}^{M} w_i d_i}{\sum_{i=1}^{M} w_i} \qquad (11)$$

In the above equation, w_i denotes the degree to that the i^{th} rule matches the input data. The cuckoo search algorithm is used to optimize the Gaussian membership parameters.

Optimizing the Gaussian membership parameters using cuckoo search algorithm (CSA)

Begin
Initialize a population of n host nests Z_i $(i = 1, 2, 3, \dots, n)$
While (iteration < maxgeneration)
Give Gaussian membership parameters c *and* σ to each cuckoo
Calculate Gaussian membership function for each nest (attribute) using equation (7)
Fitness= max (membership function value)
Remove a fraction probability [0, 1] of the worst nests
Keep the best solutions
Rank the solutions and discover the current best
if (current best nest (fitness) > best nest)
Update best nest = current best nest
End if
End while
Post process results and visualization
End

3.3Fuzzy neuro prism (FNP) algorithm

Input: A training dataset with n classes $C_i, i = 1,2,3, \dots, n$
Output: Fuzzy classifier

1. The given training dataset into fuzzy training set by using fuzzy logic (Gaussian membership function)
2. For every class C_i start with the complete training set each time
3. Calculate the weight value of each attribute pair using equation (10) for the class, C_i
4. Choose the pair with the largest weight value and generate a subset of the training set comprising the entire instances with the selected attribute combination for every class, C_i
5. Go to step 3 and 4 for this subset until a subset contains only instances of C_i or covers the entire attributes
6. The rule is induced through the conjunction of all the attribute pairs selected
7. Eliminate the entire instances covered via this rule from the training set.
8. Go to step 3 through 7 until the entire instances of C_i have been eliminated
9. Go to step 2 until the entire classes are examined
10. Perform the defuzzification process using equation (11) for classical output

3.4Fuzzy neuro J48 (FNJ48) algorithm

Input: Training dataset
Output: Fuzzy Decision tree (Fuzzy classifiers)

1. The given training data set into the fuzzy training set by using fuzzy logic with Gaussian membership function
2. In case the instances belong to the same class the tree denotes a leaf so the leaf is returned by labeling with the same class
3. The weight value is calculated using equation (10) for each attribute, given by a test on the attribute (internal node)

4. The best attribute is found on the basis of the present selection criterion and that attribute selected for branching (result of the test)
5. Execute the defuzzification process using equation (11) for crisp or classical output

4.Experimental results
In this section, the performance of the classification is compared between prism, FNP, J48 and FNJ48 based on optimized hybrid feature selection-modified CADEX-MPCA-sample selection (OHFS-MC-MPCA-SS) in terms of True Positive rate, True Negative rate and classification accuracy.

4.1Data description
In the experimentation, we are considered the students' dataset that have 600 data example, which are gathered from various colleges. In dataset 40 attributes are present that integrates students' name, course, age, gender and nature of college consists of medical/engineering, college type similar to government, self-financed, location feature, family belong to nuclear family or joint family, family factors such that occupation & educational qualification of family members, economic factors, college factors, social factors and spending time in television, mobile, computer, personal factors, academic factors etc.,. For example, location features described as the location in that students' home, school and college placed consists of rural area, urban area and semi-urban area. College features are one of the attributes that offers the information about whether student refer lecturer notes that is known by lecturer or books, techniques of teaching consists of lecturer technique/black board, number of students in class, whether college acceptable mobile phones or not, etc. Social features such as regulation of relatives for studies, number of friends and academic overall performance of friends.

In the data, the student's performance are evaluated consists of good /poor in the academy along with the features present. Data examples with these features are specified in the feature selection method then achieves chosen features. These chosen features are given to the classifiers for overall performance evaluation. In our experimentation, we are used prism and J48 classifier. Prism and J48 are classification algorithms. Prism is used for inducing modular rules and J48 is used for building apruned or unpruned C4.5 decision tree. In our experimentation, the 150 data example is given as training data (with class label) to classifier for learning procedure and remaining data are assumed as test data (without class label) that is given to classifier with the intention of discovering the class label. At last, the output variable or attribute or class is to be determined in the dilemma is the academic status or student overall performance, which has two possible values: PASS (student who pass the course) or FAIL (a student who has to repeat the course).

4.2True positive (TP) rate
In this work, if the result class label from a prediction is PASS and the actual class label is also PASS, then it is named a TP rate. It is also known as sensitivity or recall. It is computed as follows,

$$TP = \frac{TP}{TP + FN}$$

In the above equation, TP denotes the true positive and FN denotes the false negative.

From *Figure 1*, the performance comparison of FNP, J48 and FNJ48 based on OHFS-MC-MPCA-SS is shown in terms of the true positive rate. In the graph, classification techniques such as FP, J48 and FJ48 are symbolized in the X-axis and true positive rate is indicated in the Y-axis. From the recall comparison graph, it is proved which the FJ48 achieves higher recall than the FP and J48 methods.

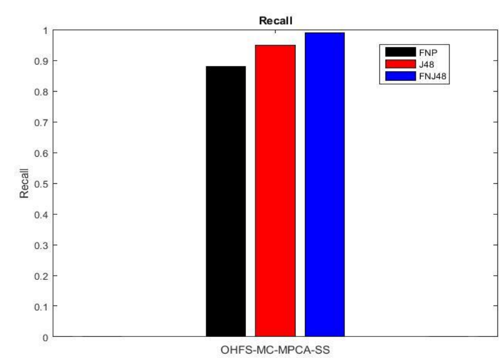

Figure 1 Comparison of true positive rate

4.3True negative (TN) rate
In this work, if the result class label from a prediction is FAIL and the actual class label is also FAIL, then it is called a TN rate. It is also known as specificity. It is computed as follows,

$$TN = \frac{TN}{TN + FP}$$

In the above equation, FP denotes the False Positive.

It is shown by *Figure 2*, comparison of the precision between FNP, J48 and FNJ48 based on OHFS-MC-MPCA-SS. From the precision comparison graph, it is found which the FJ48 technique provides a higher precision for proposed approach than the FP and J48 techniques.

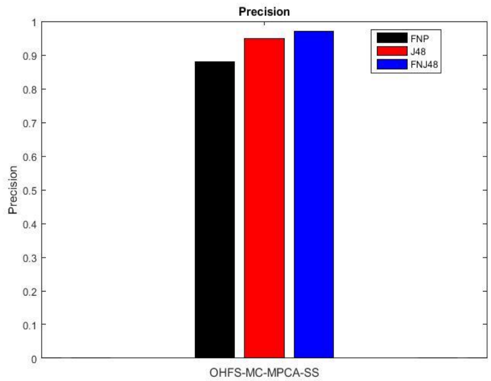

Figure 2 Comparison of true negative rate

4.4Accuracy rate

It represents as the classification accuracy. False Positive rate is explained as the percentage of actual negatives which are predicted to be positive. In this work, if the result class label from a prediction is PASS and the actual class label is FAIL. False negative (FN) rate is indicated as the percentage of actual positives which are predicted to be Negative. In this work, if the result class label from a prediction is FAIL and the actual class label is PASS, then it is called a FN rate. Accuracy is computed as follows,

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

The comparison of FNP, J48 and FNJ48 based on OHFS-MC-MPCA-SS approach for metric accuracy is shown in *Figure 3*. In this graph, the classification techniques such as FP, J48 and FJ48 are symbolized in the X-axis and the accuracy rate is indicated in the Y-axis. From the bar graph, FJ48 has more efficient in accuracy performance.

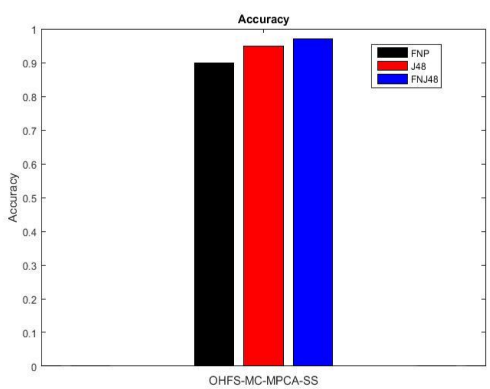

Figure 3 Comparison of accuracy rate

5.Conclusion

In this paper, a FNP and FNJ48 are proposed for improving the classification accuracy. Rules are generated for describing the relationship between the input attribute space and classes. In fuzzification process, Gaussian membership function is used. In the membership function, the weight value of each attribute is calculated by using neural network. Fuzzified value of features and the attribute with maximum weight value are utilized for constructing tree of prism and J48 classifiers. Fuzzy membership function parameters are optimized based on Cuckoo search algorithm. The experimental results show that the proposed approach is providing better results in terms of Accuracy, True positive rate and True negative rate.

Acknowledgment

None.

Conflicts of interest

The authors have no conflicts of interest to declare.

References

[1] Merceron A, Yacef K. Educational data mining: a case study. In AIED 2005 (pp. 467-74).
[2] Romero C, Ventura S. Educational data mining: a survey from 1995 to 2005. Expert Systems with Applications. 2007; 33(1):135-46.
[3] Barker K, Trafalis T, Rhoads TR. Learning from student data. In proceedings of the systems and information engineering design symposium 2004 (pp. 79-86). IEEE.
[4] Sharma A, Kumar R, Varadwaj PK, Ahmad A, Ashraf GM. A comparative study of support vector machine, artificial neural network and Bayesian classifier for mutagenicity prediction. Interdisciplinary Sciences: Computational Life Sciences. 2011; 3:232-9.
[5] Regha RS, Rani RU. A representative standardized sample set selection for improving student's performance prediction. International Journal of Recent Scientific Research. 2018; 9(2), 24612-7.
[6] Hidayah I, Permanasari AE, Ratwastuti N. Student classification for academic performance prediction using neuro fuzzy in a conventional classroom. In international conference on information technology and electrical engineering 2013 (pp. 221-5). IEEE.
[7] Karyotis C, Doctor F, Iqbal R, James A. An intelligent framework for monitoring students affective trajectories using adaptive fuzzy systems. In international conference on fuzzy systems 2015 (pp. 1-8). IEEE.
[8] Do QH, Chen JF. A neuro-fuzzy approach in the classification of students' academic performance. Computational Intelligence and Neuroscience. 2013; 2013:1-7.

[9] Yousif MK, Shaout A. Fuzzy logic computational model for performance evaluation of Sudanese Universities and academic staff. Journal of King Saud University-Computer and Information Sciences. 2018; 30(1):80-119.

[10] Fan CY, Chang PC, Lin JJ, Hsieh JC. A hybrid model combining case-based reasoning and fuzzy decision tree for medical data classification. Applied Soft Computing. 2011; 11(1):632-44.

[11] Ghosh S, Biswas S, Sarkar D, Sarkar PP. A novel Neuro-fuzzy classification technique for data mining. Egyptian Informatics Journal. 2014; 15(3):129-47.

[12] Kotsiantis S, Patriarcheas K, Xenos M. A combinational incremental ensemble of classifiers as a technique for predicting students' performance in distance education. Knowledge-Based Systems. 2010; 23(6):529-35.

[13] Hamsa H, Indiradevi S, Kizhakkethottam JJ. Student academic performance prediction model using decision tree and fuzzy genetic algorithm. Procedia Technology. 2016; 25:326-32.

Sasiregha. R received the M.Sc from JJK College of Arts and Science, Affilated to Periyar University. M.Phil from SNR Son's College, Coimbatore. She is also pursuing part time Ph.D in Bharathiyar University. She is currently an Assitant Professor at the Department of Computer Science in SSM College of Arts and Science, Tamilnadu, India. She has published tweleve research articles in national and international journals. Her research interests include data mining, neural network and swarm intelligence.

Email: sasireghaphd123@gmail.com

Dr. R. Uma Rani completed her M.C.A from Regional Engineering College, Trichy. M.Phil from Mother Teresa University, Kodaikanal. Ph.D., received from Periyar University in 2006. She is currently an Associate Professor at the Department of Computer Science, Sri Sarada College for Women, Tamil Nadu, India. She has published 120 research articles in international journals and authored 3 books. She has acted as resource person in national and international conferences and chaired technical sessions too. Her research interests include network security, data mining, image processing and mobile communications.