Fuzzy-Clustering Embedded Regression for Predicting Student Academic Performance

Zhenpeng Li, Changjing Shang and Qiang Shen
Department of Computer Science
Institute of Mathematics, Physics and Computer Science
Aberystwyth University, UK
Email: {zhl6, cns, qqs}@aber.ac.uk

***Abstract*—The prediction of student academic performance is important to both educational institutions and students themselves for a variety of reasons. However, previous techniques often consider only past numeric data for prediction, whereas others overuse different types of indicative attribute, leading to the creation of complicated predicting methods whose results are difficult to interpret. This paper proposes a novel approach to predicting student final period grade, using attributes related to student past academic records and attributes of normal study behaviour, which are readily obtainable and easily interpretable. The proposed approach works by employing fuzzy clustering and multi-variable regression within an integrated framework, which also includes an offset value mechanism to support the use of attributes that are related to normal student study behaviour. Comparative experimental investigations are carried out, demonstrating the potential of the proposed work in producing more accurate results.**

***Index Terms*—Fuzzy clustering, regression, prediction, student performance.**

I. Introduction

The prediction of student academic performance plays an important role in educational institutions, especially for higher educational institutions such as universities and colleges. For example, when reviewing applications from prospective students, the prediction may help higher educational institutions to find candidates who are eligible for a particular academic program and identify those applicants who are likely to perform well in future study [1]. In the process of dealing with student application, the final period grades are usually regarded as the benchmark for evaluation and unfortunately, are always not available at the time when receiving student applications. Thus, methods of predicting the likelihood of student performance in their final period exam are necessary. This is also the case for individual students; accurate prediction for the final period grade would help them to choose institution and subject discipline when applying for higher education.

More generally speaking, predicted results of academic performance for students already at a certain educational institution are also very useful, enabling the educational institution to provide them with appropriate additional support such as customised personal assistance and tutoring resources. Also, predicting student academic grade offers a feasible means to handle emergencies where a student fails to attend an exam due to inevitable reasons, such as physical injury or other medical situations. There may even be cases where a number of students in the same course fail to attend the exam because of unavoidable weather conditions or a natural disaster. Additionally, the results of prediction can be used by lecturers to specify the most appropriate teaching materials and actions for each group of students to meet their needs, and by students themselves in making informed decision for seeking suitable employment. For any of these cases, successfully predicting student exam scores and using them as evidence to evaluate student academic performance offers great potential benefits. Thus, developing a prediction tool is very important for educational institutions.

Over the past years, methods for data mining and machine learning have been applied in the area of education, although at a rather coarse level compared their usage in other academic fields. For instance, artificial neural networks have been employed to predicting student academic performance in an engineering course [2]. An alternative for training neural networks in an effort to predict student performance was introduced in [1]. A neuro-fuzzy approach for classifying students through academic performance prediction in a conventional classroom context has been reported in [3], and an attempt to predict student exam scores by analysing social network data between students has also been made [4].

In a more broad sense of addressing the problem of predicting student academic performance, a number of proposed methods work based on the use of large quantities of previous exam results. For example, student performance in prior academic courses is used to predict their performance in a subsequent course [5]. It has been shown that previous success in high school mathematics and science has a positive correlation with the study of computer science at universities [6]. Also, high school performance and background in mathematics is utilised to predict final exam grades in an introductory computer science course [7]. Apart from previous academic records, different types of other attributes, including age and gender [8], educational level of the parents [9], emotional factors [10], social relationships [4], and even the complexity measure of lecture notes [11] have been taken into consideration in the existing work.

Whilst researching into relationships between student academic performance and a wide variety of individual attributes is meaningful and worthwhile, the overuse of different types of indicative attribute has led to the creation of complicated

978-1-5090-0626-7/16/$31.00 ©2016 IEEE

score predicting methods which may be difficult to implement and whose results may be difficult to interpret. Certain types of attribute may not be easy to obtain during the normal teaching process. Moreover, previous methods for prediction may be excessively focused on the relationship between student academic performance and a particular type of attribute, ignoring the fact that such performance is a synthesised consequence of many reasons. Having taken notice of this, a novel approach to predicting student academic performance is proposed here, based on the synthesis of just basic attributes that are related to the academic course and the students' normal study behaviour.

The rest of this paper is arranged as follows. Section II introduces the proposed architecture for building an intelligent system to predict students academic performance, describes the functionality of each component within the system, and analyses their complexity. Section III shows the experimental results, supported by comparative studies with the real grades and other methods of prediction. Section IV concludes the paper with suggestions for further development.

II. Predicting System

This section presents the proposed general framework to predict student final period grade, including the description of its component subsystems and their associated algorithm complexity analyses.

A. System Structure

The structure of the proposed predicting system is shown in Fig. 1. It comprises four distinct component subsystems, each of which implements the following functionalities, respectively: partition, regression, offset value generation and estimation. These activities are integrated together to form the overall student score-predicting mechanism, whose implementation involves a 4-step computational algorithm:

1) Partition data of sample students (typically from previous years on the same course, whose final period grades are available) into different categories based on the similarities of their existing academic records (not including their final period grade), and obtain the fuzzy membership values for each of the sample students with regard to different partitions.
2) Determine for each partition, the relationship between the final period grade and the previous records in the academic module concerned.
3) Generate the offset value of the predicted final period grade for the target student according to the similarity of the student's own normal study behaviour and the behaviour of other students with the same or similar previous academic records.
4) Estimate the predicted final period grade of the target student based on the fuzzy membership values obtained in step 1, the relationship determined in step 2, and the offset value acquired in step 3.

The working details of these subsystems are explained below.

B. Partition Subsystem

Let an academic record list of n students be the input of the partition subsystem. In general, each record in the list, describing a student with their previous grade and final period grade (the grade values are in numeric form) for a given academic module, forms an instance of the training dataset.

The aim of the partition subsystem is to divide the instances regarding a certain module into different categories according to a formulaic synthesis of student academic records over previous periods. Those students with similar previous academic records are partitioned into the same group. Hence, the outputs of the partition subsystem are denoted as groups of instances with similar academic records. However, it may be difficult to distinguish exact groups to which the instances may belong in accordance to their formulaic synthesis of previous academic records, due to their relevance to different groups. Thus, fuzzy c-means clustering, which has natural appeal to handling such uncertainty, is used to implement this subsystem. The resulting membership values for an instance to each group will play an important role in later steps.

The initial centroid of each cluster can be preset by lecturers (as domain experts), or generated through the following steps from the training samples:

1) Calculate the arithmetic average avg_i $(1 \leqslant i \leqslant n)$ of the previous grades for each instance in the training dataset.
2) Sort the instances in the training dataset according to their arithmetic averages.
3) Divide all the instances into K clusters evenly based on their arithmetic averages after sorting, and name the results C_j $(1 \leqslant j \leqslant K)$.
4) Calculate for each cluster C_j, the arithmetic average of the academic grades over each of the previous period, and assign the results as the initial centroid of the corresponding cluster.

In the process of clustering, the Euclidean distance between a newly given instance and the centroid of each cluster is calculated in order to determine which cluster the instance belongs to. The pseudo-code for the implementation of fuzzy c-means student clustering algorithm is shown in Alg. 1.

The time complexity of this algorithm is $O(ndKl)$ [12], where n represents the number of instances in the training dataset, d represents the dimension of the instance in the training dataset, K represents the number of clusters, and l represents the number of iterations taken by the algorithm to converge. Typically, K and l are fixed in advance and are usually not too large. Therefore, the algorithm has the time complexity of the size of the training dataset times the dimensionality of each training instance, namely $O(nd)$.

C. Regression Subsystem

Regression analysis is a popular statistical process for estimating the relationships among variables, which has been widely applied [13]. It is utilised here as a segment to predict student final period grade. In particular, multi-variable linear regression, a highly flexible mechanism for examining the

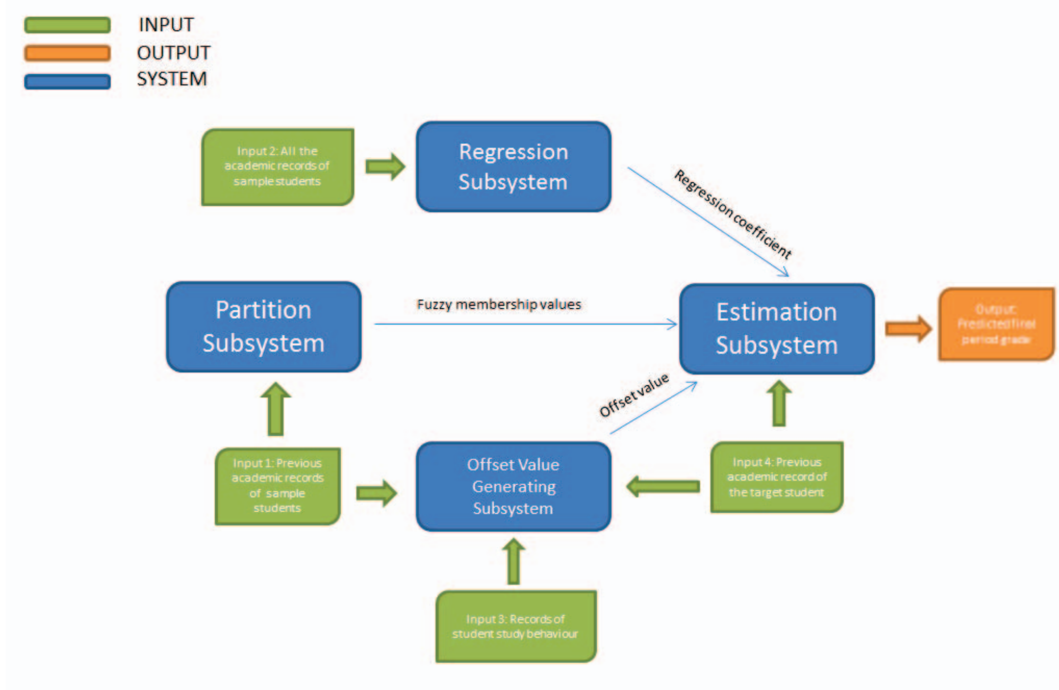

Fig. 1. Architecture of Final Period Grade Predicting System

Algorithm 1 Fuzzy C-means Clustering of Students

K: number of clusters
$C = \{c_1, c_2, ..., c_K\}$, centroids of K clusters
m: fuzzy partition matrix exponent, $m > 1$
n: number of instances in the training dataset
$S = \{s_1, s_2, ...s_n\}$, training dataset of n student records
μ_{ij}: degree of membership of s_i $(1 \leq i \leq n)$ in j^{th} cluster $(1 \leq j \leq K)$
$J = \sum_{i=1}^{n}\sum_{j=1}^{K} \mu_{ij}^m ||s_i - c_j||^2$, objective function
ϵ: specified minimum threshold between iterations

1: Calculate initial cluster membership values μ_{ij} by
$$\mu_{ij} = \frac{1}{\sum_{l=1}^{K}(\frac{||s_i - c_j||}{||s_i - c_l||})^{\frac{2}{m-1}}}$$
where $||*||$ stands for Euclidean distance
2: **repeat**
3: Calculate cluster centers:
$$c_j = \frac{\sum_{i=1}^{n}\mu_{ij}^m s_i}{\sum_{i=1}^{n}\mu_{ij}^m}$$
4: Update μ_{ij} according to:
$$\mu_{ij} = \frac{1}{\sum_{l=1}^{K}(\frac{||s_i - c_j||}{||s_i - c_l||})^{\frac{2}{m-1}}}$$
5: Calculate objective function J
6: **until** J improves by less than ϵ

relationship of a collection of independent variables with a single dependent variable [14], is an appropriate choice to perform the prediction. This is because different intuitive academic attributes are required to be taken into consideration to form the required regression model, with each of them being regarded as an independent factor in the evaluation of student academic performance.

The basic idea of the linear regression model is: given a dataset $\{y_i, x_{i1}, x_{i2}, ...x_{im}\}, (i = 1, ..., n)$ of n instances, the relationship between the dependent variable y_i and the m independent variables x_{ij} $(1 \leqslant j \leqslant m)$ is assumed to be linear. That is, the underlying relationship amongst all the variables takes the form of

$$y_i = \alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \cdots + \alpha_m x_{im} \quad (1)$$

where $\alpha_0, \alpha_1, \alpha_2, \cdots, \alpha_m$ are termed regression coefficients.

To implement the algorithm of multi-variable linear regression for predicting student's final period grade, assume that there are m independent variables $G_1, G_2, ...G_m$: their values as given in the training samples are denoted as vectors. Denote these vectors as $V_{G_1} = [G_1^1, G_1^2, ..., G_1^n]^T$, $V_{G_2} = [G_2^1, G_2^2, ..., G_2^n]^T$,..., $V_{G_m} = [G_m^1, G_m^2, ..., G_m^n]^T$, respectively, where n stands for the number of instances in the training dataset. Denote the dependent variable by V_{G_p}, with its value set encoded as the vector $V_{G_p} = [G_p^1, G_p^2, ..., G_p^n]^T$.

Let $\hat{1}$ denote the unit value vector (of an n dimensionality) and

$$X = \begin{bmatrix} \hat{1} & V_{G_1} & V_{G_2} & ... & V_{G_m} \end{bmatrix} = \begin{pmatrix} 1 & G_1^1 & G_2^1 & ... & G_m^1 \\ 1 & G_1^2 & G_2^2 & ... & G_m^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & G_1^n & G_2^n & ... & G_m^n \end{pmatrix} \quad (2)$$

Then, the multi-variable linear regression to estimate the exam marks is of the following form:

$$y = X\beta \quad (3)$$

where β is an $m+1$ dimensional regression coefficients vector, containing $\beta_0, \beta_1, \beta_2, ..., \beta_m$ as sequenced elements.

In order to determine the element values of β, conventional least squares (LS) estimator is adopted here owing to its computational simplicity. The LS method minimises the sum of squared residuals, and leads to a closed-form expression for estimating the unknown vector β:

$$\beta = (X^T X)^{-1} X^T y \quad (4)$$

Note that for each cluster generated by the partition subsystem, a calculation on the corresponding regression coefficient vector $\beta^i (1 \leq i \leq K)$ is required, where K is the number of clusters generated by the partition subsystem. Thus, the complexity of the LS algorithm for multi-variable linear regression is $O(m^2 n)$, where m is the number of independent variables and n is the number of instances in the training dataset. Since m is usually fixed and known in advance, and m is typically much smaller than n for the present application, the asymptotic time complexity of the regression subsystem can be approximated by $O(n)$.

D. Offset Value Generating Subsystem

In practice, predicting student academic performance by considering previous records only is not always sufficient. It is not surprising that students may achieve quite different results in their final period even if they have had the same or similar achievements at previous stages. This reality makes the task of reaching highly accurate prediction a challenge. Having taken notice of this, aspects other than just the student previous academic records need to be taken into account in order to generate better predicting results. Nowadays, it is commonly recognised that the study behaviour has a significant impact upon student academic achievement [15], making it an interesting factor worth investigating.

The present subsystem is developed in an effort to optimise predicted final period grade, by generating an offset value to the interim predicted final period grade for a given (target) student. Suppose that there are m previous academic records available for the target (student) instance. The computational process of this subsystem involves the following steps:

1) Calculate the Euclidean distances ed_i $(1 \leq i \leq n)$ between the target instance and the instances in the training dataset:

$$ed_i = \sqrt{\sum_{j=1}^{m} (G_j^t - G_j^i)^2} \quad (5)$$

where n is the number of instances in the training dataset and G_j^t is the academic record of the target instance regarding the j^{th} period.

2) Find the nearest instances to the target by sorting the Euclidean distances returned by step 1, and put them into a vector (of a varying dimensionality), named $V_{nearest}$.
3) If the Euclidean distance between the target and a certain instance in $V_{nearest}$ equals 0, copy the fuzzy membership values associated with that instance in the $V_{nearest}$ as the corresponding membership values of the target. If the Euclidean distance between the target instance and an instance in $V_{nearest}$ does not equal to 0, calculate its fuzzy membership values to each cluster in the same way as done by the fuzzy c-means method.
4) Preprocess the given data in response to student normal study behaviour as follows. For a datum presented in boolean form, transform it into "0" or "1", where "0" represents "NO" and "1" represents "YES". For a datum given in numeric form, normalise it to fall within the interval of $[0, 1]$.
5) Without losing generality, suppose that there are p attributes reflecting certain aspects of student normal study behaviour in the dataset, denoted by A_i $(1 \leq i \leq p)$, and that there are N instances in $V_{nearest}$. For each instance in $V_{nearest}$, calculate its difference $\textit{Diff}_k$ $(1 \leq k \leq N)$ to the target by the following:

$$\textit{Diff}_k = \sum_{i=1}^{p} A_i^t - A_i^s \quad (6)$$

where A_i^s represents the i^{th} attribute of the study behaviour involved in the instances within $V_{nearest}$, and A_i^t represents the i^{th} attribute of the study behaviour concerning the target instance.

6) For each instance in $V_{nearest}$, calculate its similarity to the target instance by the following [16]:

$$Sim_k = \frac{\sum_{i=1}^{p} [1 - |A_i^s - A_i^t|]}{p} \quad (7)$$

7) Calculate the differences in the final period grades between each pair of the instances in $V_{nearest}$, and find the maximum difference, denoting it by MAX_d. If $V_{nearest}$ contains only one element then the difference is set to 0.
8) Calculate the offset value of the predicted final period grade for the target instance by:

$$\textit{offset\_value} = MAX_d \cdot \sum_{k=1}^{N} (\textit{offset}_k \cdot Sim_k) \quad (8)$$

Note that several steps in the above are not necessary to be carried out in the present order; for instance, step

7 may be done right after step 2 if desired. Although the implementation of the offset value generating subsystem includes many steps, its time complexity is acceptable, with $O(n^2)$ to form the $V_{nearest}$, $O(N)$ to find the MAX_d, $O(p)$ to calculate Diff_k, $O(p)$ to compute Sim_k, and $O(N)$ to calculate the *offset_value*. Hence, the total time complexity is $O(n^2 + 2N + 2p)$. Since N and p are usually not large numbers for the present problem, the time complexity can be approximated by $O(n^2)$.

E. Estimation Subsystem

Given the regression coefficient vectors $\beta^1, \beta^2, ..., \beta^K$ (where K is the number of clusters given by the partition subsystem), and a set of fuzzy membership values $V_\mu = [\mu_1, \mu_2, ..., \mu_K]$ for a target student, the estimation subsystem implements a straightforward and final step of the entire computation process. Suppose that the vector of previous grades is denoted by V_{G_p} and that the *offset_value* of the target instance has been obtained (see the preceding sub-section), the predicted final period grade $G_{predicted}$ can be calculated as

$$G_{predicted} = (\sum_{i=1}^{K} \mu_i \cdot V_{G_p}^T \beta^i) + \mathit{offset\_value} \tag{9}$$

where $V_{G_p}^T$ denotes the transpose of V_{G_p}.

Note that the task of predicting exam scores for a number of students can be implemented through the recursive application of this method. The predicted exam scores of target students together with information regarding their assignment scores and class test scores can be used to construct new training instances to enlarge the training dataset. Simply, the time complexity of the estimation method is $O(K)$ for one student, and $O(Kn)$ for n students.

III. Experimental Evaluation

This section presents experimental studies of the proposed approach. The work both illustrates the implemented system in action and demonstrates its efficacy.

A. Experimental Setup

1) Dataset preparation: Four datasets, each containing hundreds of instances collected from two Portuguese schools (GP and MS) about their students in Maths and Portuguese language study, are used as examples [17]. As with many European countries (e.g. France), a 20-point grading scale is used to evaluate student academic performance, where 0 is the lowest grade and 20 is the perfect score. During the school year, students are evaluated in three periods and the last evaluation (G3 in Table I) corresponds to the final period grade. In the complete dataset, more than 30 attributes with related data to each attribute are collected. For simplicity and clarity, attributes which are closely related to the student academic performance and the study behaviour based on expert's opinion are selected to conduct the experiment. The selected attributes are shown in Table I.

Before implementing the system, data preprocessing is carried out. In particular, attributes related to the normal student study behaviour, such as "study-time", "failure-count" and "absence" are normalised into the range between 0 and 1. For instance, for a sample dataset with n instances, the normalised value of the data with regard to the attribute "study-time" for the i^{th} $(1 \leq i \leq n)$ instance, denoted by $std\_study\text{-}time_i$, is defined as follows:

$$std\_study\text{-}time_i = \frac{study\text{-}time_i - study\text{-}time_{min}}{study\text{-}time_{max} - study\text{-}time_{min}} \tag{10}$$

where $study\text{-}time_{min}$ and $study\text{-}time_{min}$ are the maximum and minimum value of the attribute "study-time", respectively.

2) Experimental method: In the experiments, each dataset is split into subsets for 10-fold cross validation [18]. The reported results are based on an average of 10 times of the 10-fold cross validation. Since the ground truth of the students final period grades are in the form of integer, whereas the predicted final period grade are in the form of floating-point number, the predicted data need to be transformed back to integers to support interpretability. Without losing fairness, when conducting the experiments and comparing the proposed work with other techniques, both truncation and rounding methods are investigated, each mapping the resulting predicted data onto an integer.

Despite the fact that the main use of the proposed system is to predict the numeric grade (in terms of integer scores) for a given target student, the method can also be applied as a classification model to categorise a target into a specific class based on the predicted numeric grade. According to the Erasmus grade conversion system – a European programme that enables students exchange in 31 countries [17], the grades can be transformed into European Credit Transfer System (ECTS) Grades. The details of the transforming rule are listred in Table II:

TABLE II
ECTS Grades with Corresponding Portugal/France Grades

ECTS Grades	Portugal/France Grades
Excellent (A)	16-20
Good (B)	14-15
Satisfactoty (C)	12-13
Sufficient (D)	10-11
Fail (E)	0-9

B. Results and Discussions

1) Prediction of numeric grades: For the analysis of the proposed approach, the predicted final period grades are compared with the corresponding underlying ground truth. The work is also compared with the standard multi-variable linear regression method (SMLR) [13] and the simple clustered linear-regression method (SCLR) [19]. Both of these methods are widely adopted in the field of numeric prediction, especially when there is little knowledge of the non-linear relations between the end result and the attributes, as it is the case for

TABLE I
PREPROCESSED STUDENT ACADEMIC PERFORMANCE RELATED ATTRIBUTES

Attribute	Description
study-time	weekly study time (numeric: 1: less than 2 hours, 2: 2 to 5 hours, 3: 5 to 10 hours, or 4: more than 10 hours)
failure-count	number of failures in the past for this academic module (numeric: integer)
support	extra support from educational school or family or other sources (binary: 1 for "yes", 0 for "no")
study-aim	whether or not to take higher education (binary: 1 for "yes", 0 for "no")
activities	extra curricular activities (binary: 1 for "yes", 0 for "no")
absence	days of school absence (numeric)
G1	first period grade (numeric: from 0 to 20)
G2	second period grade (numeric: from 0 to 20)
G3	final period grade (numeric: from 0 to 20)

the present application. In performing the step of partition, the number of clusters K is set to 5 in accordance with the common practice of classifying students, m is set to 2 and ϵ is set to 10^{-6}. All 4 available datasets are used here for training and testing. The resulting statistical indicators are listed in Table III, IV and V. The best performance on each dataset is highlighted in boldface.

From Table III, it seems difficult to distinguish the best predicting method by considering the statistical factor of absolute mean error. The values of this statistical factor generated by different methods are rather similar to each other with little variance. However, the experimental results given in both Table IV and Table V show that the proposed method outperforms SMLR and SCLR, regarding predicting accuracy and within 1-grade deviation. These results jointly show that predicted final period grades are closer to the ground truth, demonstrating the significant potential of the proposed framework.

For most of the cases, clustering linear regression outperforms the standard linear regression model [19]. The proposed approach generally performs even better than SCLR because for a large number of sample students, their academic grades are distributed normally and continuously. It is relatively more difficult for SCLR to find clear boundaries amongst the clusters. The proposed system has fuzzy-clustering embedded, avoiding the need of stating exactly to which category a target may belong. Instead, the academic records of each student that are considered belonging to a certain category are associated with membership values. Such membership values are used in computing the weight of each regression model contributing to the predicted grade of the target student. Another reason for the proposed to outperform both SMLR and SCLR is that it makes better use of the attributes about the student normal study behaviour. Although these attributes are also taken into account by the other two predicting models, their values on the 0-1 scale are too small to make significant contributions in these models. Also, the proposed approach possesses an interesting ability thanks to the introduction of an offset value. From a list of sample students with the same or similar previous academic records, given their final period grades, the proposed system can generate a predicted result exceeding the limits of these sample students. This may have helped further improve its performance.

2) Prediction of 5-level grades: To further analyse the results achievable by the proposed work, advanced classification techniques such as neural network, support vector machines (SVM), decision trees and random forest are also employed to classify the predicted final period grates. The Multilayer-Perceptron, SMO, J48 and RandomForest algorithms released with the Weka software [20] are used to represent these classification approaches, with the polynomial kernel selected to implement SMO.

The resultant accuracies are shown in Table VI. Again, for comparison, the highest classification accuracy is denoted in boldface. Clearly, the proposed approach performs better than other listed methods in predicting accuracy. In particular, the proposed system that uses the rounding method generates the most accurate results.

IV. CONCLUSION

This paper has proposed a novel approach to predicting student performance in academic courses. Unlike simple clustering regression analysis which takes part of the precise sample data into consideration, the proposed approach processes the universal data with an embedded step of fuzzy clustering. This has an intuitive appeal in handling a large number of student academic records which are typically normally and continuously distributed. The work makes use of attributes that are related to observed student study behaviour, by introducing an offset value in the predicting model. The implementation of the embedded fuzzy clustering approach, supported by the offset value mechanism, generates better results than the existing methods. With fuzzy representation, the approach synthesises the use of intuitive attributes from an academic course and from student normal study behaviour. This helps make the predicted results more readily interpretable, while involving simple computation.

Whilst promising, the proposed work opens up an avenue for further investigation. The present approach focusses on attributes that are closely related to the academic course and the student study behaviour. However, other performance indicators directly related to the teaching of the course, say,

TABLE III
COMPARISON OF APPROACHES IN TERMS OF ABSOLUTE MEAN ERROR

Dataset	**Proposed method**		**SMLR**		**SCLR**	
	Trunction	Rounding	Trunction	Rounding	Trunction	Rounding
Maths (GP)	0.6362	0.5941	0.6573	0.6023	0.5639	**0.5122**
Portuguese (GP)	0.4654	**0.4451**	0.5255	0.4841	0.4931	0.4624
Maths (MS)	0.6872	**0.6033**	0.7064	0.6334	0.7206	0.6218
Portuguese (MS)	0.5725	0.5244	0.5233	**0.4974**	0.5324	0.5077

TABLE IV
COMPARISON OF APPROACHES IN TERMS OF PREDICTING ACCURACY (%)

Dataset	**Proposed method**		**SMLR**		**SCLR**	
	Trunction	Rounding	Trunction	Rounding	Trunction	Rounding
Maths (GP)	63.4 ± 1.4	**72.1**±1.1	57.1 ± 1.3	67.2 ± 1.1	59.8 ± 0.8	67.4 ± 0.9
Portuguese (GP)	64.2 ± 1.2	**72.8** ±0.9	58.6 ± 1.1	69.6 ± 0.7	59.1 ± 0.6	69.4 ± 0.7
Maths (MS)	61.2 ± 1.0	71.4 ± 1.6	54.9 ± 1.1	66.8 ± 1.3	60.8 ± 0.9	**71.4** ±1.0
Portuguese (MS)	68.4 ± 1.0	**79.1** ±1.1	64.1 ± 0.9	74.2 ± 1.1	68.4 ± 0.8	77.8 ± 0.9

TABLE V
COMPARISON OF APPROACHES IN TERMS OF PREDICTING RESULT WITHIN 1-GRADE DEVIATION (%)

Dataset	**Proposed method**		**SMLR**		**SCLR**	
	Trunction	Rounding	Trunction	Rounding	Trunction	Rounding
Maths (GP)	81.8 ± 0.9	**88.4** ±1.3	75.2 ± 1.7	82.4 ± 0.9	79.9 ± 1.1	88.1 ± 0.9
Portuguese (GP)	83.1 ± 1.8	**90.2** ±1.0	82.6 ± 1.4	88.2 ± 1.3	83.3 ± 1.6	89.7 ± 1.7
Maths (MS)	79.1 ± 1.5	**87.4** ±1.2	74.9 ± 1.0	83.8 ± 1.1	80.1 ± 1.9	86.6 ± 1.8
Portuguese (MS)	82.6 ± 1.4	**91.1** ±1.3	80.6 ± 1.2	88.2 ± 1.9	80.4 ± 1.3	88.7 ± 0.9

TABLE VI
COMPARISON OF CLASSIFICATION ACCURACY (%)

Dataset	**Proposed method**		**MultilayerPerceptron**	**SMO**	**J48**	**RandomForest**
	Trunction	Rounding				
Maths (GP)	74.3 ± 0.5	**76.8** ±0.6	60.3 ± 1.6	59.6 ± 1.2	72.7 ± 0.4	68.62 ± 0.6
Portuguese (GP)	77.2 ± 0.4	**79.6** ±0.4	65.1 ± 0.9	64.5 ± 0.6	74.1 ± 0.3	69.66 ± 0.8
Maths (MS)	72.3 ± 0.4	**73.4** ±0.5	60.1 ± 0.8	61.5 ± 0.7	71.7 ± 0.6	66.71 ± 0.7
Portuguese (MS)	**76.7** ±0.7	75.1 ± 0.6	63.4 ± 1.1	61.9 ± 0.9	73.4 ± 0.5	71.21 ± 0.7

practical results, oral presentation scores and tutor evaluation outcomes, could also be taken into consideration. To integrate these factors, improved fuzzy clustering offset value generating techniques need to be developed. Also, the attributes in the whole dataset are currently chosen by human experts. An additional step to include feature selection [21] techniques in order to make an informed, automated choice of which attributes to use would be very beneficial.

REFERENCES

[1] J.-F. Chen and Q. H. Do, "Training neural networks to predict student academic performance: A comparison of cuckoo search and gravitational search algorithms," *International Journal of Computational Intelligence and Applications*, vol. 13, no. 01, p. 1450005, 2014.

[2] V. Oladokun, A. Adebanjo, and O. Charles-Owaba, "Predicting students academic performance using artificial neural network: A case study of an engineering course," *The Pacific Journal of Science and Technology*, vol. 9, no. 1, pp. 72–79, 2008.

[3] I. Hidayah, A. E. Permanasari, and N. Ratwastuti, "Student classification for academic performance prediction using neuro fuzzy in a conventional classroom," in *Information Technology and Electrical Engineering (ICITEE), 2013 International Conference on*. IEEE, 2013, pp. 221–225.

[4] M. Fire, G. Katz, Y. Elovici, B. Shapira, and L. Rokach, "Predicting student exams scores by analyzing social network data," in *Active Media Technology*. Springer, 2012, pp. 584–595.

[5] A. Chamillard, "Using student performance predictions in a computer science curriculum," in *ACM SIGCSE Bulletin*, vol. 38, no. 3. ACM, 2006, pp. 260–264.

[6] C. G. Petersen and T. G. Howe, "Predicting academic success in introduction to computers," *AEDS Journal*, vol. 12, no. 4, pp. 182–191, 1979.
[7] T. P. Cronan, P. R. Embry, and S. D. White, "Identifying factors that influence performance of non-computing majors in the business computer information systems course," *Journal of Research on Computing in Education*, vol. 21, no. 4, pp. 431–446, 1989.
[8] R. F. Deckro and H. W. Woundenberg, "Mba admission criteria and academic success," *Decision Sciences*, vol. 8, no. 4, pp. 765–769, 1977.
[9] S.-M. R. Ting and T. L. Robinson, "First-year academic success: A prediction combining cognitive and psychosocial variables for caucasian and african american students." *Journal of college student development*, 1998.
[10] J. Bennedsen and M. E. Caspersen, "Optimists have more fun, but do they learn better? on the influence of emotional and social factors on learning introductory computer science," *Computer Science Education*, vol. 18, no. 1, pp. 1–16, 2008.
[11] K. J. Keen and L. Etzkorn, "Predicting students' grades in computer science courses based on complexity measures of teacher's lecture notes," *Journal of Computing Sciences in Colleges*, vol. 24, no. 5, pp. 44–48, 2009.
[12] J. F. Kolen and T. Hutcheson, "Reducing the time complexity of the fuzzy c-means algorithm," *Fuzzy Systems, IEEE Transactions on*, vol. 10, no. 2, pp. 263–267, 2002.
[13] G. A. Seber and A. J. Lee, *Linear regression analysis*. John Wiley & Sons, 2012, vol. 936.
[14] L. S. Aiken, S. G. West, and S. C. Pitts, "Multiple linear regression," *Handbook of psychology*, 2003.
[15] C. Desforges and A. Abouchaar, "The impact of parental involvement, parental support and family education on pupil achievements and adjustment: A literature review," Research report, Tech. Rep., 2003.
[16] W.-J. Wang, "New similarity measures on fuzzy sets and on elements," *Fuzzy sets and systems*, vol. 85, no. 3, pp. 305–309, 1997.
[17] P. Cortez and A. M. G. Silva, "Using data mining to predict secondary school student performance," 2008.
[18] Y. Bengio and Y. Grandvalet, "Bias in estimating the variance of k-fold cross-validation," in *Statistical modeling and analysis for complex data problems*. Springer, 2005, pp. 75–95.
[19] B. Ari and H. A. Güvenir, "Clustered linear regression," *Knowledge-Based Systems*, vol. 15, no. 3, pp. 169–175, 2002.
[20] I. H. Witten and E. Frank, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2005.
[21] R. Diao, F. Chao, T. Peng, N. Snooke, and Q. Shen, "Feature selection inspired classifier ensemble reduction," *Cybernetics, IEEE Transactions on*, vol. 44, no. 8, pp. 1259–1268, 2014.