Centre for Intelligent Systems and their Applications
School of Informatics
The University of Edinburgh

A Data-Driven Fuzzy Rule-Based Approach for Student Academic Performance Evaluation

Ph.D. Thesis Proposal

November 2002

Khairul Anwar Rasmani

List of contents page

A Data-Driven Fuzzy Rule-Based Approach for Student Academic Performance Evaluation

Khairul Anwar Rasmani

Centre for Intelligent Systems and their Applications
School of Informatics, The University of Edinburgh

1 Introduction

Evaluation of student academic performance usually consists of several components, each involving a number of judgements often based on imprecise data. This imprecision arises from human (teacher/tutor) interpretation of human (students) performance. Arithmetical and statistical methods have been used for aggregating information from these assessment components. These methods have been accepted by many educational institutions around the world although there are limitations with these traditional approaches. In this proposed study, it is argued that the current method of classifying and grading student academic performance using arithmetical and statistical techniques does not necessarily offer the best way to evaluate human acquisition of knowledge and skills. It is expected that reasoning based on fuzzy models will provide an alternative way of handling various kinds of imprecise data, which often reflects the way people think and make judgements.

2 Problem Identification and Research Objectives

This section gives an overview of student academic performance evaluation, identifies the problems needing to be tackled and describes the research objectives.

2.1 Brief Overview of Student Academic Performance Evaluation

Evaluation of student performance is one of the most important parts in educational systems. It has to be done for several important reasons arranged below. It aims to provide scoring or grading scheme that is interpretable by ordinary people, especially students, teachers, parents, employers and policy planners.

Reasons for students academic performance evaluation

There are several important reasons why student performance needs to be assessed. Firstly, the evaluated level of performance can be used as an indication of a student's level of understanding. This is important in providing information for teachers to take further action if necessary, such as planning remedial activities, or planning further instruction. This information is also very useful to enable students to overcome any weaknesses. Besides this, students who succeed in examinations may be motivated to learn more [23].

Secondly, assessment is important for the purpose of making academic decisions about the students now, or in the future. For example, students who do not achieve a certain level have to repeat the course while other students will proceed onto the next stage. Assessment is also important to indicate the level of performance for graduation purposes, and this usually has a permanent effect on the future career of students [3].

Thirdly, assessment is important to provide information about the teacher’s ability to instruct and the ‘system’ being practiced.

Atkins et al. [4] state that the reasons for assessing students also include:

- to give staff feedback on the effectiveness of their teaching,
- to determine the extent to which course aims have been achieved,
- to obtain information on the effectiveness of the learning environment, and
- to monitor standards over time.

Format of academic performance evaluation

Evaluation of academic performance can be done in the format of *formative assessment* or *summative assessment.* Formative assessment is conducted to monitor the progress of instruction. This is important to give feedback to students and teachers [18]. This type of assessment usually involves a wide range of activities on frequent occasions, such as a series of observations, short tests and quizzes, etc. Summative assessment is conducted at the end of each instructional segment through tests and final examinations to provide information on how much the students have achieved [4].

Final evaluation may involve a combination of summative assessment and formative assessment. There are several possible reasons of using this combination, for example:

(a) To make use of different types of assessment which have different aims to measure students' achievement. There are a wide range of assessments that can be chosen, such as classroom observations, essays, homework, open book tests, portfolios, etc. Some methods may be more suitable to assess what cannot be assessed in final written examinations, such as investigation and group work. Written examinations at the end of the course may not provide a complete picture of what the students have learnt [7].

(b) Assessment is part of a learning process. As pointed in [43], assessment is not something to be done after teaching is finished. Continuous assessment in the form of formative assessment will provide feedback in regards to student performance at an early stage of the course. Discussion of students' errors in a series of tests or quizzes will reinforce their knowledge acquisition.

(c) A combination of several methods of assessment is generally viewed as essential to provide full coverage of important learning outcomes [16]. This is quite similar to the argument put forward in [24] that views the use of a multiple strategy of assessment as ‘usually seen as a clear intention to measure and represent broader educational experience’.

(d) To motivate students; for example, higher marks obtained in continuous assessment may motivate students to work harder for a final written examination.

Various different modes of evaluation methods have been used for primary, secondary and tertiary education. A variety of evaluation methods also exist in different countries around the world. For example, results from a survey conducted by Hounsell et al. [26] in Scottish Higher Education in 1996 show that there were at least 137 different strategies of evaluation. This reflects a wide availability of different types of academic performance evaluation method.

Assessment components

Academic performance evaluation usually consists of several assessment components. These components consist of a wide variety of assessment methods such as:

- Series of tests and quizzes
- Portfolios
- Formal written examinations
- Individual Assignments and Coursework
- Group work
- Observation
- Theses and publishable materials
- Posters and oral presentation

Different assessment components reflect different modes of evaluation used to assess student academic performance. The term *assessment components* will be used repeatedly in this report, referring to the assessment methods such as those mentioned above.

Typical existing methods to represent student academic performance

Student academic performance evaluation usually involves awarding numerical values or linguistic labels to a given piece of student work. These values and labels have been used to represent student achievement by reasoning with arithmetical or statistical methods. A combination of different assessment components usually has been used with different allocation of marks. By using arithmetical methods, for examples different scores from each assessment are added up to obtain a single score. Simple statistical methods such as calculating the average from several assessments are also often used. Further analysis of students scores can be made using more complex statistical methods such as calculating the mean, median, mode, range, standard deviation, variance and standard "z-score".

In general, methods represent student academic performance can be classified into several categories:

(a) Single letter-grade (e.g. A, B, C, D, E, F). This letter-grade is usually based on a numerical interval-value that refers to a certain category of achievement.

(b) Nominal score (e.g. 1, 2, …, 10). This numerical-grade might refer to another numerical interval-value that refers to a certain category of achievement.

(c) Single numerical score that usually refers to 100 percent.

(d) Linguistic terms such as "Pass" and "Fail"

(e) Single ‘fine’ grade-points from 0.00 to 4.00 (400 points: 0.00, 0.01, 0.02, … , 3.99, 4.00) based on Grade- Point Average (GPA) and Cumulative Grade-Point Average (CGPA).

Combinations of different methods are also used, for example the use of GPA alongside the single letter-grade. Numerical scores have more commonly been used than letter-grades or linguistic terms because grades using numbers can be used for further arithmetical or statistical analysis.

Hierarchical process in academic performance evaluation

As mentioned earlier, academic performance evaluation usually consists of several assessment components. Although there are different modes of evaluation method, there is one similarity among them. This can be described as follows:

Suppose that X is the evaluation method and X_i are the assessment components, then X_i can be written as:

$X_i = X_1, X_2, X_3, \ldots, X_n$, where n is an integer.

Each of the assessment components, X_i will carry w_i, the weight of the respective component.

In order to make use of the information obtained from each assessment component for further analysis, the assessment components will usually contain numerical data. An inference mechanism will aggregate this data to obtain a single score representing the overall achievement. For example, to obtain a single score to represent student achievement in a final written examination, p different marks from p different questions will be used. This is followed by aggregation of different scores from different evaluation methods (e.g. assignments, tests and written examinations). The same process will then be repeated at a higher level. Thus, academic performance evaluation usually involves a hierarchical process. Figure 1 shows an example of the hierarchical process in student academic performance evaluation.

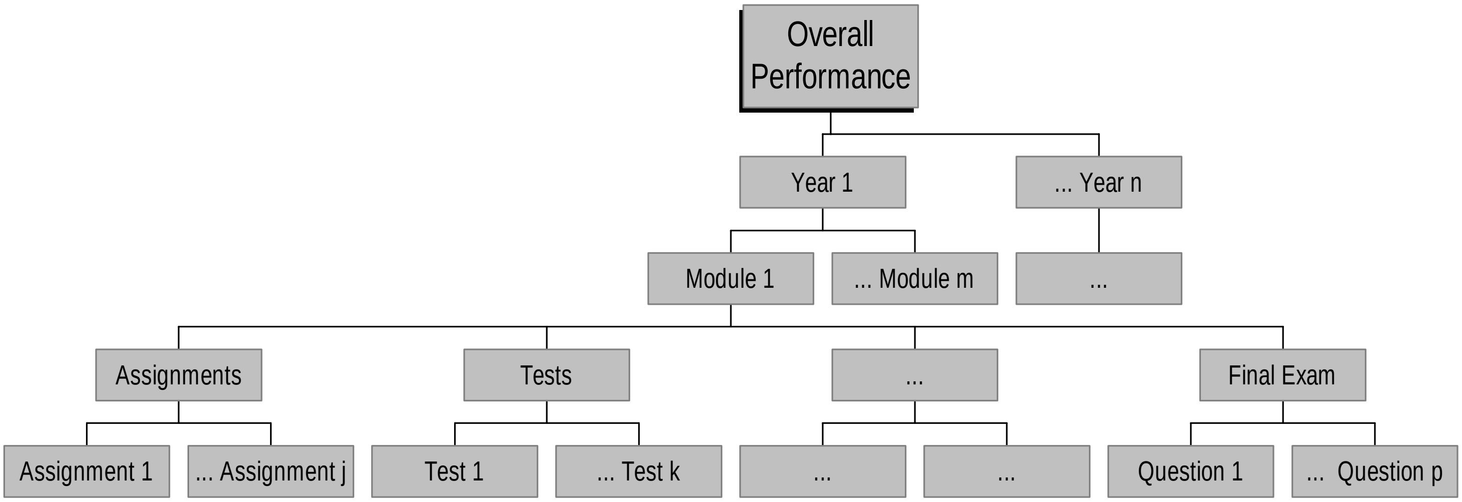

Figure 1: Hierarchical process in student academic performance evaluation

2.2 New Method for Student Academic Performance Evaluation

The Need for a New Method

Important reasons exist for creating a new evaluation method for student academic performance:

a) *To check true student performance.* Existing evaluation methods, especially statistical ones, have been used without exploiting any other alternative method to double check student performance. Sometimes evaluators use ad hoc inference methods, but they lack a formal mechanism to support the inference. The option of a new method is needed and it will be very useful to help the user (students, parents, decision-makers, etc.) to confirm or refute decisions based on traditional evaluation methods. This is particularly important for boundary cases.

b) *To handle uncertain scores.* The primary method of assessment usually involves awarding numerical marks by an evaluator. Such marks are usually given according to a given marking scheme. These marks are usually numerical values that may fluctuate a little as different evaluators may award different marks [31, 55]. Evaluation of a student's work may be affected by the evaluator's experience, sensitivity and the standard used. Thus marks awarded by an evaluator to represent student performance are only an approximation. Although linguistic terms (e.g. bad, good, very good, excellent etc.) have also been widely used to represent the final student's performance [31], their inherent nature of vagueness is often ignored. However, academic performance evaluation involves the measurement of ability, competence and skills. Ability, competence and skills are fuzzy concepts and can be approximately captured in fuzzy terms.

c) *Evaluation of academic performance using natural language.* Instead of using numerical values, linguistic terms can be awarded to represent student achievement in each of the assessment components. The use of natural language such as "good", "very good" and "extremely good" would allow more flexible ways to make judgement on students' performance. Therefore a method that is able to aggregate information given in the form of natural language is needed.

Reasoning based on fuzzy approaches has been successfully applied for inference of multiple attributes containing imprecise data [14, 44, 59]; in particular, fuzzy rule-based systems (FRBS) provide intuitive methods of reasoning based on linguistic models [1]. It is worth mentioning that the applications of FRBS in numerous real world classification problems have been mentioned frequently in most of the literature in the area of fuzzy systems. Typical examples can be found in [44]. Recent developments in this area also show the availability of FRBS, which allow interpretation of their inference results and have high accuracy rates. This is very important in providing a platform for the application of FRBS in student academic performance evaluation.

The Proposed Extended Method

One of the drawbacks of the current academic evaluation methods is the lack of information behind the evaluation methods that have been used and what criteria the 'final result' or 'score' refers to. For example, in a *criterion-reference evaluation* (evaluation that refers to established criteria of performance) [21], student score is referred to a specific criterion of achievement. A 75% score might show that *'a student has been assessed and gets 75 marks out of 100'*. By using fixed grading systems, this score may be translated into a single letter-grade, for example 'B' or a linguistic term such as 'very good'. Again, the comparison is usually being made against 100 percent achievement.

In order to produce a meaningful statement, humans usually make reference to a standard. For example, *'excellent'* may not have any useful meaning to represent student achievement unless a complete statement is created, for example *'excellent compared to other students in the group'*. This is similar with statements containing numerical values. For example *'75 marks'* does not mean a student has performed excellently, unless this mark is compared to other marks in a given population. Thus, instead of using criterion-referenced evaluation, evaluation is often made based on *norm-referenced evaluation* (evaluation that refers to other students' performance in a population) [21]. For example, an evaluation that was based on student achievements in the previous five years may be more meaningful than referring to percentage marks.

Previous and current student performance data in computer systems can be used to provide an evaluation, which refers to a range of evaluation results. Simple statistical methods can be used to make comparisons of individual achievements with the larger student population. Other statistical methods such as measure of central tendency, measure of dispersion and standard 'z-score' methods can also be used. However, they have not been widely adopted for student academic performance evaluation because they produce numerical values that are less meaningful to the user, particularly the z-score method which is usually known as a 'standardized score' that transforms a student score into a new score based on the normal probability distribution.

Although norm-referenced evaluation seems better than criterion-referenced evaluation, this technique also has some drawbacks. Questions usually arise concerning which population should be used as the standard. Thus, instead of using a norm-referenced evaluation or a criterion-referenced evaluation, it is suggested that evaluation of student performance could also be done using a combination of both methods. This will produce a new score that gives new information about the student's achievement. Such information is very useful since it provides additional guidance to people in making decisions.

To do so, a fuzzy approach will be used to perform the proposed extended method of student performance evaluation. It is important to point out that the aim of the proposed method is not to replace the current traditional method of evaluation, instead it will strengthen the present system by providing additional information to be used for decision making by the user. Figure 2 shows the proposed extended method of student academic performance evaluation.

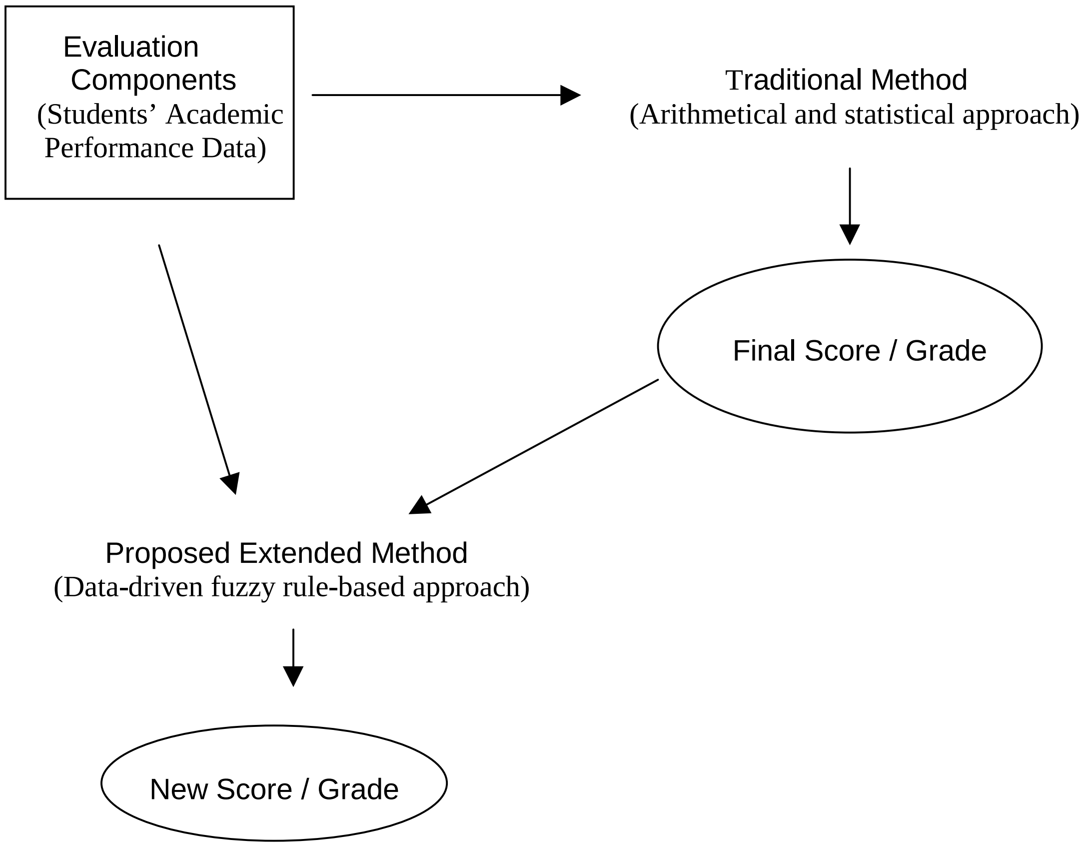

Figure 2: Proposed extended method of student academic performance evaluation

2.3 Research Objectives

The objectives of the proposed study are to:

(a) Study the application of fuzzy modelling based on previous or current data for classifying student academic performance.

(b) Develop data-driven fuzzy rule models for such an application, which arithmetical and statistical methods are unable to offer effectively, and which allows inference to be performed in a more natural way using linguistic variables rather than numerical values.

(c) Implement prototypes based on the data-driven fuzzy rule models, investigating the effectiveness of the prototype systems in handling multiple attributes, containing imprecise data, to perform human-like reasoning.

(d) Compare the work against existing approaches such as statistical and other classifier based methods.

2.4 Scope of Study

This study is focused on the development of fuzzy systems. It aims to builds a method which uses fuzzy rule models and their associated inference mechanisms for student performance evaluation.

Focus of Study

For the purpose of this research, traditional methods of student performance evaluation, which will be repeatedly mentioned in this report, refer to non-fuzzy approaches, mainly the methods that use statistical techniques. Due to the diversity of existing educational evaluation methods, this research will focus on 'high level' of student academic performance. The proposed method will be applied hierarchically for aggregating scores from assessment components (assignments, tests, final exams, etc.) to produce a score for individual modules, aggregating results of different modules to produce a score for yearly performance, and aggregating different years' achievement to produce an overall performance.

The proposed method will use current and/or historical data. It is expected that its performance will depend on the availability and quality of the training datasets, due to the nature of data-driven learning.

Training and Testing Datasets

For testing the accuracy rate of the proposed method, data from the UCI machine learning databases such as the Iris Plant dataset, the Wine Recognition dataset and the Glass Identification dataset [51] will be used. For application of the proposed method, a real student academic performance dataset will be obtained from the School of Informatics, University of Edinburgh. In particular, data for first and second academic year undergraduate students will be used. An application of this method for other datasets will be performed if time permits.

3 Theoretical Foundations: Fuzzy Rule-Based Systems

This section describes the components involved in the development of a Fuzzy Rule-Based System (FRBS). Sub-section 3.1 consists of brief explanations of fuzzy set theory, fuzzy membership functions, fuzzy logical operators and fuzzy IF-THEN rules. This is followed by sub-section 3.2 that describes data-driven FRBS and rule induction for handling classification tasks. Finally, sub-section 3.3 presents an approach of generating fuzzy rules, based on fuzzy subsethood values.

3.1 Fuzzy Set Theory

In classical set theory, a set is normally defined as a collection of elements. Each element can either belong or not belong to the set. If $x \in A$ is true, then $x \notin A$ is false.

For any set A, a function called a characteristic function (membership function) of A is defined by:

$$\mu_A(x) = \begin{cases} 0 & \text{if } x \text{ is not in set } A \\ 1 & \text{if } x \text{ is in set } A \end{cases} \qquad (1)$$

This classical approach shows that everything is precise. However in the real world, there are attributes that are not precise such as old, close, beauty, good, excellence, etc. In a statement such as "He did badly in his final exam", someone could ask ‘how bad?’ "is it really bad?" and so on. Such imprecision leads to the introduction of fuzzy sets.

A fuzzy set A in X, is written in the form of ordered pairs:

$$A = \{(x, \mu_A(x)) \mid x \in X)\} \tag{2}$$

where $\mu_A(x)$ is the degree of membership of x in A and $\mu_A(x) \in [0,1]$.

The characteristic function of a fuzzy set is a mapping to a portion of the real line, allowing a continuum of possible choice [46]. That is, the value of an element in a fuzzy set is defined in terms of the degree in an interval from 0 to 1. If 0 is false and 1 is true, a value approaching 0 means that the value is becoming ‘false’ and a value approaching 1 means that the value is approaching ‘true’. For example, if $\mu_A(x)$ is the membership function of the student performance 'excellent' and x is the mark given to indicate the student performance, then the closer the value of $\mu_A(x)$ is to 1, the more x belongs to 'excellent' and the closer $\mu_A(x)$ is to 0, the less x belongs to 'excellent'.

3.1.1 Fuzzy Membership Functions

A *linguistic variable* is defined as a variable whose values are words or sentences in a natural or synthetic language [59]. For example, 'achievement' of a student can be a *linguistic variable* that take the fuzzy sets 'bad', 'average' and ‘good’ as its *linguistic term*.

Figures 3 shows graphical representations of each fuzzy set, which depend on the definition of 'bad', 'average' and 'good'. The combinations of the fuzzy sets will characterize the membership function $\mu(x)$, where x is the mark/score to represent the achievement of the student. Figure 3 also shows the characteristics of fuzzy sets that representing a linguistic variable where there is overlapping (or no sharp boundary) between each of the fuzzy sets 'bad', 'average' and 'good'.

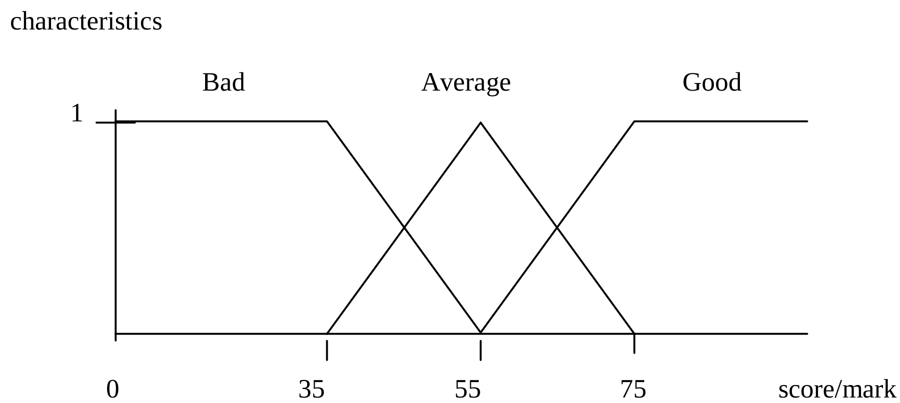

Figure 3: Membership functions for 'achievement'

As there will be more than one word to describe the 'achievement' (i.e. the linguistic variable), then we may use other terms such as 'extremely bad', 'very good', 'more or less average', etc. A quantifier such as 'extremely' 'very' and 'more or less' is a modifier and is termed a *linguistic hedges* [47, 52].

The fuzzy membership functions can be determined by several possible methods. The most popular method is the heuristic method where pre-defined shapes will be chosen to represent certain linguistic terms [28]. This is usually done by an expert. The most popular functions are piecewise linear functions such as triangular and trapezoidal membership functions [28] due to their computational efficiency [27]. There are other functions that can be used, for example, simple linearly decreasing and increasing functions, S shape functions [39], and sigmoid function [15].

Other methods for generating membership functions are also available, for example based on probabilistic curves and the distribution of values presented as a histogram [28]. Neural network-based methods [28, 35, 36] and clustering methods [1, 28] have also been used to generate fuzzy membership functions from historic data or training samples.

Fuzzy membership functions play an important role in the success of fuzzy rule-based systems [28, 34]. Typically, they depend on the type of problem and data. Therefore choosing or generating an appropriate fuzzy membership function to represent a linguistic term is very important. Several techniques for fine tuning the membership function are available, for example the use of linguistic hedges [33, 34].

3.1.2 Fuzzy Logical Operators

In classical set theory, the basic operations of a set involve complement, union and intersection. In logical terms these correspond to negation, disjunction and conjunction, respectively. These operators are unique and therefore carry a unique interpretation. On the other hand in fuzzy set theory, there is more than one operator that can be used for the interpretation of negation, conjunction and disjunction.

Fuzzy Negation

This operation applies to a single fuzzy set. The negation of a fuzzy value is the fuzzy set complement of the original set A:

$$\mu_{NOT\text{-}A}(x) = 1 - \mu_A(x) \tag{3}$$

Apart from (3), there are other fuzzy set complements, for example, Sugeno *class* and *Yager class* [52] and Wu complement [29].

Fuzzy Conjunction and Disjunction

There is a well established class of functions for fuzzy conjunction and disjunction [29]. The logical operator for fuzzy conjunction is also known as a triangular norm (t-norm) and fuzzy disjunction as triangular conorm (t-conorm). The fuzzy conjunction and disjunction shown below ((4) and (5)) are also known as the Min-Max operators and have been used widely probably because of their simplicity.

$$\mu_{A\ AND\ B}(x) = \min\ [\mu_A(x), \mu_B(x)] \tag{4}$$

$$\mu_{A\ OR\ B}(x) = \max\ [\mu_A(x), \mu_B(x)] \tag{5}$$

There are several other operators, for example: Algebraic Product-Sum operators, Bounded Difference-Sum operators, Einstein operators, Drastic *t-norm* and *t-conorm* operators [29, 34]. To be accepted as a t-norm and t-conorm class of functions, an operator must satisfy the four conditions: identity, associativity, commutativity and monotonicity [29].

The availability of a choice of operators for complement, disjunction and conjunction in fuzzy systems provides a rich platform for generating fuzzy rules [34]. This may be a strong point of fuzzy logic systems, which have been successfully applied to various real world problems.

3.1.3 Fuzzy IF-THEN Rules

A rule-based system utilizes a model that represents human knowledge in the form of "IF-THEN" rules. This conventional approach has been adapted to build fuzzy rule-based systems. A simple fuzzy IF-THEN rule can be written in the form of " IF x is A THEN y is B" where A and B are fuzzy sets. This can be extended to more than two fuzzy sets resulting in *compound fuzzy propositions*. In general, fuzzy IF-THEN rules are production rules whose antecedents, consequences or both are fuzzy [17]. Mendel [34] classified fuzzy rules into 6 different types, namely *Incomplete Rules*, *Mixed Rules*, *Fuzzy Statement Rules*, *Comparative Rules*, *Unless Rules and Quantifier Rules*. However there is no agreed classification of fuzzy rule models [17] and a single rule might involve a combination of several different classification types [34].

In general, a linguistic fuzzy model may involve simple or complex structures based on fuzzy IF-THEN rules. The Mamdani-type FRBS and the Takagi-Sugeno-Kang (TSK)-type FRBS are two types of linguistic fuzzy model that have been widely used [2].

Mamdani-type FRBS

The Mamdani-type FRBS is the most widely used linguistic fuzzy model and has been used for various of real-world applications [56]. This model also has high interpretability [2, 56]. The Mamdani-type FRBS has the following structure:

$$IF\ \ X_{1i}\ \ is\ \ A_1\ \ and \ldots and\ \ X_n\ \ is\ \ A_n\ \ THEN\ \ Y_1\ \ is\ \ B_1\ \ and \ldots and\ \ Y_m\ \ is\ \ B_m \qquad (6)$$

where X_{1i} are fuzzy input linguistic variables, Y_j are fuzzy output linguistic variables, and A_{i} and B_j are linguistic terms in the form of fuzzy sets that characterize X_i and Y_j

An extension of Mamdani-type FRBS also exists [2] in the following form:

$$IF\ \ X_{1i}\ \ is\ \ \overline{A}_1\ \ and \ldots and\ \ X_n\ \ is\ \ \overline{A}_n\ \ THEN\ \ Y\ is\ \ B \qquad (7)$$

where $\overline{A}_1 = A_{11}\ \ OR \ldots OR\ \ A_{1k}$ and $\overline{A}_n = A_{n1}\ \ OR \ldots OR\ \ A_{nl}$

For example, if $\overline{A}_1 = \{A_{11}, A_{12}, A_{1k}\}$, then an example of one possible rule for linguistic variable A_1 could be *IF* A_1 *is* $\{A_{11}$ *OR* $A_{13}\}$ *THEN Y is B.*

Takagi-Sugeno-Kang (TSK)-type FRBS

The TSK-type FRBS is a model having linguistic antecedents but the consequences are in the form of a function of linguistic variables [2]. The TSK-type FRBS has the following structure:

$$\begin{aligned}&IF\ \ X_{1i}\ \ is\ \ A_1\ \ and \ldots and\ \ X_n\ \ is\ \ A_n\ \ THEN\\&Y_1 = p_1^1.\mathrm{X}_1 + \ldots + \mathrm{p}_\mathrm{n}^1.\mathrm{X}_\mathrm{n} + \mathrm{p}_0^1\ \ and \ldots and\ \ Y_m\ \ is\ \ p_1^m.\mathrm{X}_1 + \ldots + \mathrm{p}_\mathrm{n}^\mathrm{m}.\mathrm{X}_\mathrm{n} + \mathrm{p}_0^\mathrm{m}\end{aligned} \qquad (8)$$

where X_i are fuzzy input linguistic variables, Y_1 is the fuzzy output value and p are real parameters.

Although the design of this method is easily compared to Mamdani-type FRBS, this model has less interpretability [2] and involves complicated computation [56].

Of particular interest to this research, *Fuzzy Complete Rules* [38, 52] can be defined as follows. Consider a set of fuzzy rules that involve p conditional variables (or the system being modelled has p inputs) $x_i \in X_i$, $i = 1, 2, \ldots, p$ and one conclusion (output) $y \in Y$, a complete rule has the form:

$$Rule^l : IF\ x_1\ is\ F_1^l and\ \ldots\ and\ x_p\ is\ F_p^l\ THEN\ y\ is\ G^l \tag{9}$$

where l = 1,2, …, n, with n standing for the total number of the rules contained within the set. If only a subset of p inputs are being used to describe the rules, then these rules are termed *Fuzzy Incomplete Rules*. Both complete and incomplete types of fuzzy rules are sub-classes of Mamdani-type representation.

In interpreting antecedents for fuzzy rules, the connective "OR" is used for disjunction, "AND" for conjunction and "NOT" for complement. Rules with the quantifier "some" or "all" are also acceptable and these can be categorized as *Quantifier Rules*.

3.2 Data-Driven FRBS and Rule Induction for Handling Classification Tasks

Applications of fuzzy rule-based systems (FRBS) to numerous real world problems have been reported in the literature. Observations from these applications have given rise to much of the recent development of data-driven machine learning techniques, specifically devised for building FRBS. This is largely inspired by the fact that rules generated by such techniques tend to possess essential knowledge embedded in the training samples [49]. Although human experts played, and will still play, an important role in the development of conventional fuzzy systems [25], automatically generating fuzzy rules from data is very helpful when human experts are not available and may even provide information not previously known by experts.

Several approaches have been devised to develop data-driven learning for FRBS. They involve the use of a method that automatically generates membership functions or fuzzy rule structures or both from training data. Decision trees are one of the most popular choices for learning and reasoning from data [27]. This method has been employed for generating fuzzy decision trees, such as in approaches presented in Janikow [27], Crockett et al. [15], Yuan and Shaw [57], and Chiang and Hsu [12]. Janikow showed a method which builds fuzzy decision trees using procedures employed in ID3, a program generating decision trees popularized by Quinlan [27, 57]. Crockett et al. [15] proposed fuzzification of crisp decision trees using non-linear membership functions. This method is also based on ID3. Chiang and Hsu [12] also proposed an approach that integrates fuzzy classifiers with decision trees and that, as claimed by the authors, works well for the classification of data with noise. Additionally, Castro and Zurita [9] presented a method for generating fuzzy rules from training samples based on a Truth Maintenance System (TMS). All methods proposed in [9, 12, 15, 27, 57] were based on the Mamdani-type linguistic model.

Hong and Lee [25] and Nozaki et al. [37] presented methods for automatically generating both fuzzy membership functions and fuzzy IF-THEN rules from training samples. A genetic algorithm [2] approach has also been used by Crockett et al. to optimize a fuzzy region around the decision node of the decision tree. Yuan and Shaw [57] gave a new method of generating fuzzy rules by measuring the degree of truth of a fuzzy rule based on fuzzy subsethood values. So did Yuan and Zhuang [58]. Nauck et al. [36] presented a method of generating fuzzy rules from data using a Neuro-Fuzzy System called NEFCLASS which employs neural-network and fuzzy methods. The fuzzy partitions were created based on the NEFCLASS learning method. All these methods are again based on the use of the Mamdani-type FRBS representation.

Of particular interest to the present work is the method for generating fuzzy rules for classification problems, originally proposed by Chen et al. as reported in [11]. This method uses fuzzy subsethood values for generating fuzzy rules. This will be further detailed below.

In summary, figure 4a shows the general structure of building FRBS with fuzzy rule induction and figure 4b shows the structure of the actual FRBS. In particular, the development of data-driven FRBS for handling classification tasks, which form the major concern of this work, usually consists of the following steps:

(a) Identify the nature of the classification problem. This may include the identification of the type of data: numerical, interval valued, linguistic variables or fuzzy numbers. This task is important to make the decision whether to use pre-defined fuzzy partitions or generate fuzzy partitions from training samples in order to develop the 'fuzzy encoder'.

(b) Define or generate fuzzy partitions for the input variables and output variables according to the type of data and the nature of classification problems. This may involve experts or techniques such as clustering methods, neural networks or those based on classical probability theory.

(c) Transform crisp values (training samples) to fuzzy input values using the fuzzy encoder.

(d) Generate a set of fuzzy rules based on a fuzzy linguistic model. This may involve choosing a suitable fuzzy inference, revision of the fuzzy partitioning method in (b) or fine tuning the linguistic model.

(e) Apply the rule sets for classification (using the training dataset or testing dataset). This data set must be transformed into fuzzy input values using the same 'fuzzy decoder' employed for creating the fuzzy rules.

(f) Defuzzify the fuzzy output values using the 'fuzzy encoder' to generate a crisp value again. The identification of the method used in the 'fuzzy encoder' is dependent on the classification problem.

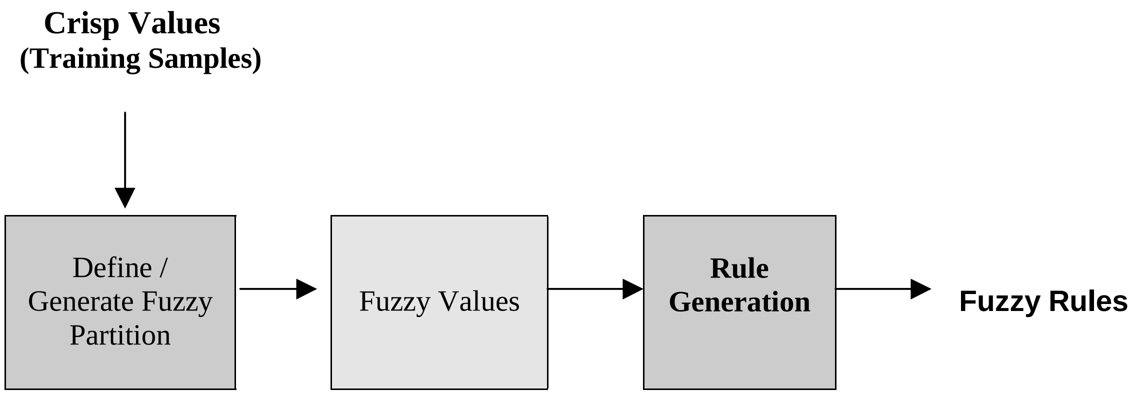

Figure 4a : Structure of fuzzy rule induction

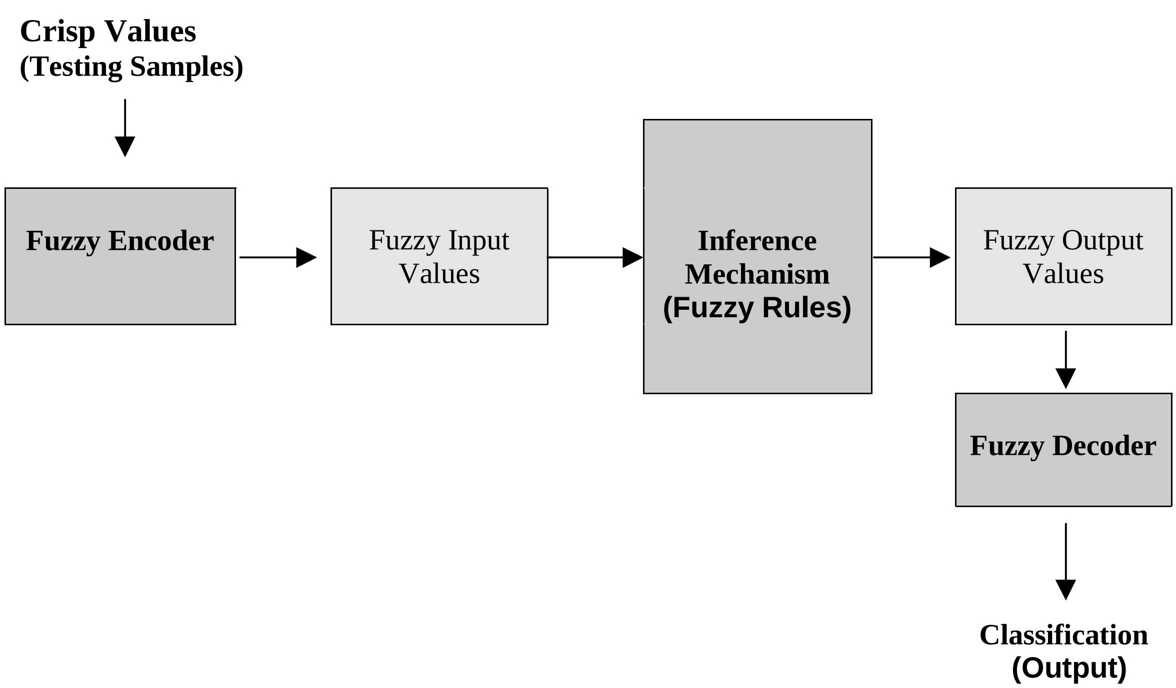

Figure 4b: Structure of the actual FRBS

Finally, it is worth noting that several criteria have been used to compare the performance of data-driven FRBS models. This includes the readability of rule sets [1], simplicity of computation, the number of rules generated and the model accuracy [11].

3.3 Subsethood-Based Rule Generation Algorithm (SBA)

This sub-section describes the method employed in [11], one of the current FRBS models that have been used for handling classification problems. This method has been highlighted for two main reasons: (a) simplicity of the method which is based on fuzzy subsethood value, and (b) possibility of producing high accuracy rates with fewer rules when compared to other methods.

The method for generating fuzzy rule models based on fuzzy subsethood values [11], which are formally defined as follow. Let A and B be two fuzzy sets defined on the universe *U*. The fuzzy subsethood value of A with regard to B, S(B, A) represents the degree to which A is subset of B [9, 49]:

$$S(B,A) = \frac{M(B \wedge A)}{M(B)} = \frac{\sum_{x \in U} \nabla(\mu_B(x), \mu_A(x))}{\sum_{x \in U} \mu_B(x)} \tag{10}$$

where S(B, A) $\in$ [0,1] and ∇ is the t-norm operator.

The purpose of the generated rules is to handle classification problems. SBA involves three main steps: a) classify training data into subgroups according to the underlying classification results, b) calculate fuzzy subsethood values for every variable in each subgroup, and c) create rules. Figure 5 summarises this algorithm.

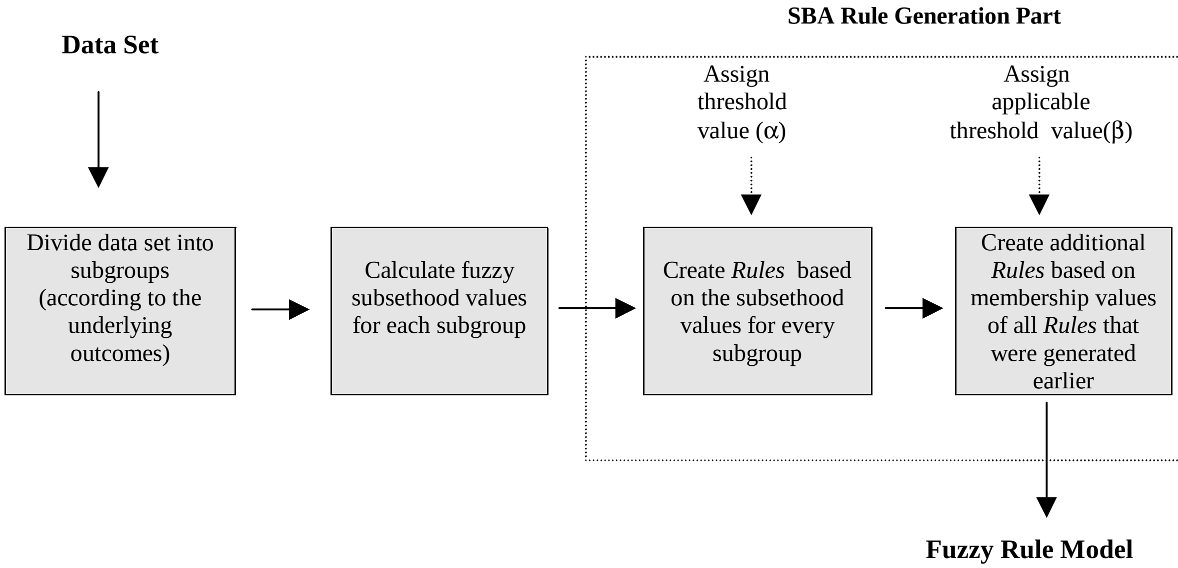

Figure 5: Subsethood-Based Rule Generation Algorithm (SBA)

The generation of fuzzy rules is dependent on the fuzzy subsethood values and a pre-specified threshold value $\alpha \in [0, 1]$. Any variables that have a subsethood value that is greater than or equal to α will automatically be chosen as an antecedent for the fuzzy rules. This includes the negation of subsethood values that are greater than or equal to α. If there are training cases that the generated rules do not cover because none of the subsethood values is greater than α, additional rules will be generated based on membership function values of those rules generated earlier, with regard to another preset threshold value $\beta \in [0, 1]$.

This technique has been extended within this project, the extension will be described in section 5.

4 Existing Fuzzy Approach for Student Academic Performance Evaluation

Student academic performance evaluation usually involves linguistic terms such as good, bad, satisfactory, excellent, etc., which involve a substantial amount of fuzziness [6]. The characteristic functions of students' achievement could be defined, for example as shown in Figure 6.

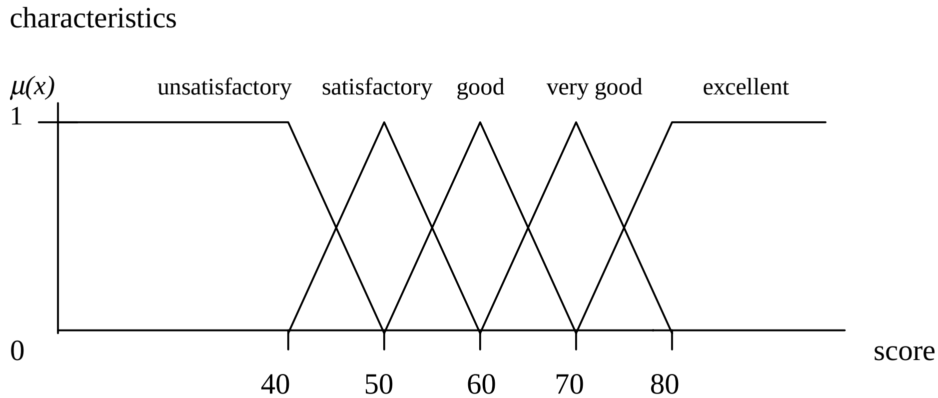

Figure 6: Membership function of student academic performance

As there are different kinds of academic performance evaluation, different types of evaluation components (questions, topics, subjects, courses, etc.) may be presented using different membership functions. Each of the evaluation components could be defined according to the characteristics of the components. For example, 'achievement in Mathematics' could be different from 'achievement in English language'. Thus different kinds of linguistic terms may be used. Furthermore, the characteristic of each linguistic term used also varies according to the relevant standard. For example, the linguistic term 'excellent' might be a score above 90 instead of 80 as shown in figure 6. Thus, the proposed membership function $\mu(x)$ of each evaluation component can be defined according to the criteria of the evaluation components. The availability of choices of fuzzy membership functions such as described in sub-section 3.1.1 will make it easier for the adoption of fuzzy set theory to represent student academic performance. However, the application of fuzzy set theory and FRBS for student academic performance evaluation is rather new. This section gives a description of several previous studies using fuzzy approaches for such applications.

There have been several studies which use a fuzzy approach for student academic performance evaluation. These studies can be categorized into four different approaches, namely a) Fuzzy similarity based, b) Fuzzy expected value based, c) Using fuzzy membership function values alongside statistical theory, and d) Simple fuzzy rule-based.

4.1 Fuzzy Similarity Based Approach

Biswas [6] proposes an application of fuzzy sets to student academic evaluation. The reasons behind the use of the fuzzy approach are that an educational grading system involves substantial amounts of fuzziness and that fuzzy theory can provide a model of subjective judgements.

Biswas employs the theory of the fuzzy similarity approach which was defined as follows. The similarity between two fuzzy sets *F* and *M* is:

$$S(F,M) = \frac{\overline{F}.\overline{M}}{\max(\overline{F}.\overline{F}, \overline{M}.\overline{M})}, \tag{11}$$

where $\overline{F} = (\mu_{\mathrm{F}}(x_1), \mu_{\mathrm{F}}(x_2), \ldots)$, $\overline{M} = (\mu_{\mathrm{M}}(x_1), \mu_{\mathrm{M}}(x_2), \ldots)$ and $S(F,M) \in [0,1]$. The larger the value of $S(F,M)$, the greater the similarity between fuzzy sets *F* and *M*. Biswas' method works as follows:

(a) Create *Standard Fuzzy Sets (SFS)* for linguistic variables Excellent, Very Good, Good, Satisfactory and Unsatisfactory. The *SFS* was defined as fuzzy sets containing membership values which represent the corresponding linguistic variables [6].

(b) Award *fuzzy marks* to Question *i*, Q_i using the *fuzzy grade sheet* (which contain rows for question number and columns for awarding marks in term of fuzzy values).

(c) Calculate the degree of similarities between Q_i and *SFS*.

(d) Find the maximum values and award the final grade.

(e) Calculate the total score based on several questions and the total mark.

$$Score = \frac{1}{100}\left[\sum T(Qi) \times P(gi)\right] \tag{12}$$

where $T(Q_i)$ = mark allocated for Q_i and $P(g_i)$ = score awarded to Q_i according to the grade.

This method can be extended to be used for assessing students' answer using several criteria such as accuracy, adequate coverage, conciseness and clear expression of the answer. Biswas pointed out that the strength of this method was that it would provide an evaluation report in detail, specially pointed form rather than interval valued.

Although this method shows the usefulness of using fuzzy membership values for aggregating marks from different questions, the main drawback is that this method needs an evaluator to award *fuzzy marks* (that containing several fuzzy values), instead of awarding a single score to each question as is done in traditional method. It was also commented in [10] that this method would take a large amount of time to perform the matching operations between *SFS* and the *fuzzy marks*.

4.2 Fuzzy Expected Value Based Approach

Law [31] proposed a student performance evaluation based on the fuzzy expected value approach. The main objective of the work was to build a structural model of the educational grading system using a fuzzy approach. Three reason were given to support this: score/marks will fluctuate a little, meaning that the score given for student performance is not always very precise; examination consists of vague data; and a common method of grading students is the use of a letter grade, implying the need for employing linguistic variables for evaluation.

In order to draw meaning from information collected from several different evaluation methods, Law employed the fuzzy expected value approach, which can be summarised as follows:

(a) Calculate expected value for linguistic variables A, B, C, D and F, $E(\overline{A}),\ E(\overline{B}),\ E(\overline{C}),\ E(\overline{D})$ and $E(\overline{F})$

where the *fuzzy expected value* for $\overline{A}$ in R^n is defined by:

$$E(\overline{A}) = \frac{\int_{R^n} x\mu_A(x) f(x)dx}{\int_{R^n} \mu_A(x) f(x)dx} \tag{13}$$

with *f(x)* being the distribution function of x in $\overline{A}$.

(b) Calculate the new score, *T*, using the centroid defuzzification method

$$T = M \times (E(\overline{A}),\ E(\overline{B}), ...,\ E(\overline{F}))^t \tag{14}$$

where M is a *fuzzy matrix assessment* (contains fuzzy membership values).

(c) Calculate *aggregate score*, *w*

$$w = \sum_{j=1}^{n} w_j T_j \tag{15}$$

where w_j are the weightings of each question.

The drawback of this method is that it involves complex computational processes and cannot integrate different fuzzy environments, as pointed out [55].

4.3 Using Fuzzy Memberships Function Values alongside Statistical Theory

The first research carried out into the use of fuzzy membership function values alongside statistical theory for performance assessment was reported in [19]. The work was proposed for the evaluation of prior learning via a portfolio of evidence. The reasons behind the use of a fuzzy approach are that academic competence are a fuzzy concept and that a decision on evidence (e.g. academic certificate) is fuzzy as different assessors may have different standards.

This method works by calculating the final score, *X*, using the expected value defined as follows:

$$X = \frac{\sum f(x_i).x_i}{\sum f(x_i)} \tag{16}$$

where x_i is a rating value from 0 to 10 which refers to competency (where 0 indicates extremely unsatisfactory and 10 indicates extremely satisfactory), and $f(x_i)$ are ordinate values (frequency of rating x_I in fuzzy values) with respect to the rating x_i. Thus, the method consists of two steps:

(a) Award a rating value x_i for each category of evidence.

(b) Calculate the expected value, *X*, based on ordinate values $f(x_i)$ and the use of balance ratings, if all criteria are equally important, or weighting the criteria using relative weights, if different criteria carry different weights.

Chen and Lee [10] proposed a method for the evaluation of student answerscripts. The purpose of the study was to counter some drawbacks of the method proposed by Biswas. Although not mentioned in the Chen and Lee article, the method proposed by Chen and Lee is similar to the method proposed by Fourali [19]. It applies fuzzy membership function values and probability theory. A brief outline of this work is given below.

(a) Create the satisfaction level and degree of satisfaction according to the satisfaction level: Extremely good, Very-very good, Very good, Good, More or less good, Fair, More or less bad, Bad, Very bad, Very-very bad and Extremely bad.

(b) Award fuzzy marks to Q_i (using the *fuzzy extended grade sheet*).

(c) Calculate the degree of satisfaction for Q_i

$$D(Q_i) = \frac{\sum y_i T(x_i)}{\sum y_i} \tag{17}$$

where y_i = membership values awarded to each question and $T(x_i)$ = degree of satisfaction for x_I .

(d) Calculate the total score based on several questions

$$Totalmark = \left[\sum T(Q_i) x D(Q_i)\right] \tag{18}$$

where $T(Q_i)$ = mark allocated for Q_i and $D(Q_i)$ = degree of satisfaction for Q_i.

This method reduces the complexity and is still able to produce alternative results for student performance evaluation when compared to Biswas' method. It can also be applied to the evaluation of answerscripts based on several criteria such as accuracy, adequate coverage, conciseness and clear expression (which is similar to Biswas' work). The major difference between these two methods rests in that, in Fourali's method, fuzzy membership values are generated from the frequency of rating whereas in Chen and Lee's method, fuzzy membership values are awarded by an expert evaluator.

Weon and Kim [55] also proposed a learning achievement evaluation strategy using a similar approach. The main objective of the research was the use of different types of fuzzy membership functions for student learning achievement evaluations. The authors argued that the common letter-grade system has several drawbacks because of human (teachers) differences in evaluation. This method is also expected to be used effectively for student performance evaluation using media-based Computer Aided Instruction (CAI). In this work, achievement evaluation was defined as a compound and hierarchical process where each element has sub elements. The evaluation is implemented by the following:

(a) Calculate response accuracy,

$$COR(P) = \bigcup_{i=1}^{n} \{P \sum_{j=1}^{n} (\mu P_{ij} \times \mu T_{ij})\} \tag{19}$$

where P is the question domain, μP_{ij} = membership grade for question i and μT_{ij}= membership grade for condition or criteria.

(b) Normalize the accuracies via dividing each resulting response accuracy by the number of questions and the number of sub-questions.

(c) Calculate the evaluation

$$EVAL(P) = \bigcup_{i=1}^{n} (P_i, NORM(COR(P_i)) \tag{20}$$

In particular, by using the maximum defuzzification method:

$$EVAL(P) = \bigcup_{i=1}^{n} (P_i, FUZSET(MAX(NORM(COR(P_i))))$$

This method can be used for more than one criteria of assessment (fuzzy environment), such as importance, complexity and difficulty of question. The drawback of this method is that it involves complex computational processes.

4.4 Simple Fuzzy Rule-Based Approach

Shimizu and Yamashita [50] presented a study of applying fuzzy reasoning for educational evaluation of calligraphy. The reason behind the use of fuzzy approach is that the evaluation of calligraphy and art is affected by the teacher's sensitivity which is fuzzy. The authors employed a fuzzy rule-based approach where the rule set is provided by an expert (teacher). The process of building the evaluation system is as follows:

(a) Construct evaluation tree based on *technical variable, x* and *sensitivity variable, y*.

(b) Generate reasoning tree based on fuzzy partition (high, medium and low) of input variables *x* and *y*, and fuzzy partition (excellent, very good, good, fair, poor) of output variable *z*.

The system-building process is straightforward. It generates rules based on all possible combinations of fuzzy partitions of the input variables. As there are only two input variables and three fuzzy partitions for each variable, it generates 9 rules. Evaluation results are generated by firing these rules followed by defuzzification of the output using *gravity-centre defuzzification*. Although this method is quite simple, very limited work exists in this type of approach.

4.5 Discussion

Arithmetical and statistical methods have been established and used for quite a long time for classifying and grading student academic performance. These methods have been accepted by many educational institutions around the world. Thus, applying fuzzy techniques for academic performance evaluation should meet some of the basic criteria such as simplicity and manageability of the proposed fuzzy method.

Several attempts to use fuzzy techniques for classifying and grading student academic performance have been tried [6, 10, 19, 31, 50, 55] as described in sub-sections 4.1 to 4.4. Considering the use of a fuzzy approach in academic performance evaluation is very new, most of the previous studies [10, 19, 55] employ basic fuzzy theory (fuzzy membership function values) alongside basic statistical approaches (average). This is supported by the fact that academic performance evaluation consists of several components that usually carry different weights, where calculating an average from several scores is the most appropriate and simplest way to obtain a single score or grade to represent student performance.

Methods presented by Fourali [19] and Chen and Lee [10] seem very simple and manageable but these techniques have been employed without the use of a fuzzy inference mechanism. Methods proposed by Law [31] and Biswas [6] seem more demanding and involve more complicated calculations.

However, typically, these methods are only suitable for certain evaluation methods. For example the method proposed by Weon and Kim [55] seems more suitable for the evaluation of student performance using Computer Aided Instruction (CAI) where time for student responses to each question can be recorded. The method proposed by Law [31] was based on an expected value calculated from a predetermined percentage of students who will get certain grades (for example 15% of the students are expected to get a grade A). This predetermined quota may not be suitable for other assessment methods. The methods of [6] and [10] need an evaluator (teacher/tutor) to award fuzzy marks in terms of numerical values from 0 to 1 on a 'fuzzy grade sheet'. This can be very confusing because it is not easy to award fuzzy marks for an attribute that has more than three fuzzy partitions.

One significant aspect that can be pointed out in the studies presented [6, 10, 19, 31, 50, 55] is that, instead of using current traditional methods, student academic performance evaluation can be done using various fuzzy approaches. However, it can be observed that there is no evidence that these methods have produced better results compared to the traditional arithmetical or statistical methods.

Another significant aspect that can be pointed out is that the use of a fuzzy approach may produce different scores compared to arithmetical and statistical methods. Thus, the score obtained using a fuzzy approach may carry different meaning. This is due to the fact that arithmetical and statistical methods use an exact mode of inference based on exact input and output, whereas fuzzy logic uses an approximate mode of inference based on fuzzy inputs and outputs.

The method proposed by Shimizu and Yamashita is quite simple and follows the techniques employed in FRBS. However, it is clear that this method can only be applied for problems, which a) contain small number of attributes and a small number of fuzzy partitions of input values, and b) where the generation of fuzzy rules depends on knowledge given by an expert. Further development of this approach using current, say, data-driven FRBS techniques is therefore needed.

In summary, several aspects should be taken into account in the development of a new method for student performance evaluation:

(a) Use original scores (crisp values) as inputs and transform the values into fuzzy values, instead of asking the evaluator to award scores in terms of fuzzy values, as in the methods presented in [6], [10] and [19].

(b) Use a proper fuzzy inference mechanism, instead of just manipulating fuzzy numbers to be used with a statistical method (average), as presented for example, in [10], [19] and [55].

(c) Not limited to certain evaluation methods, such as the methods in [6] and [10] (for answerscripts), or in [29, 50] (for specific evaluation method).

(d) Convenient to be used hierarchically (different level of performance evaluation), without the need to develop another method.

(e) As there are varieties fuzzy logical operators that can be used for inference, the new method should also make use of this facility to perform the inference.

Besides the characteristics mentioned above, a method which is less complex will certainly have some benefit for the user as it will be understood better and more easily.

5 Theoretical Advances: Weighted Subsethood-based Rule Generation Algorithm

Classifying student academic performance using data-driven FRBS is a new approach. Classifying and grading student academic performance using currently available data-driven FRBS classification techniques seems to be challenging because of the difficulty of generating rules from strictly numerical data. A system that is highly reliable in producing the classification results is clearly desirable.

This section describes the Weighted Subsethood-Based Algorithm (WSBA) developed in the initial phase of this project, which includes a modification of the Subsehood-Base Algorithm (SBA) [11], the use of fuzzy general rules and the use of subsethood values as weights.

5.1 Modification of SBA

The main idea in developing the Weighted Subsethood-Based Algorithm (WSBA) is the use of subsethood values as relative weights over the significance of different conditional attributes which they may have upon the conclusion, in conjunction with the use of default fuzzy general rules. Although it modifies the SBA, it does not make it more complicated than it already is, but simplifies the rule-learning process. Weights are created from the subsethood values to provide a multiplication factor for each linguistic variable in the *compound fuzzy propositions*. The modified algorithm generates one rule for each possible conclusion or decision class. Figure 7 shows the resultant algorithm, with the algorithmic details which differ from the original SBA explained below.

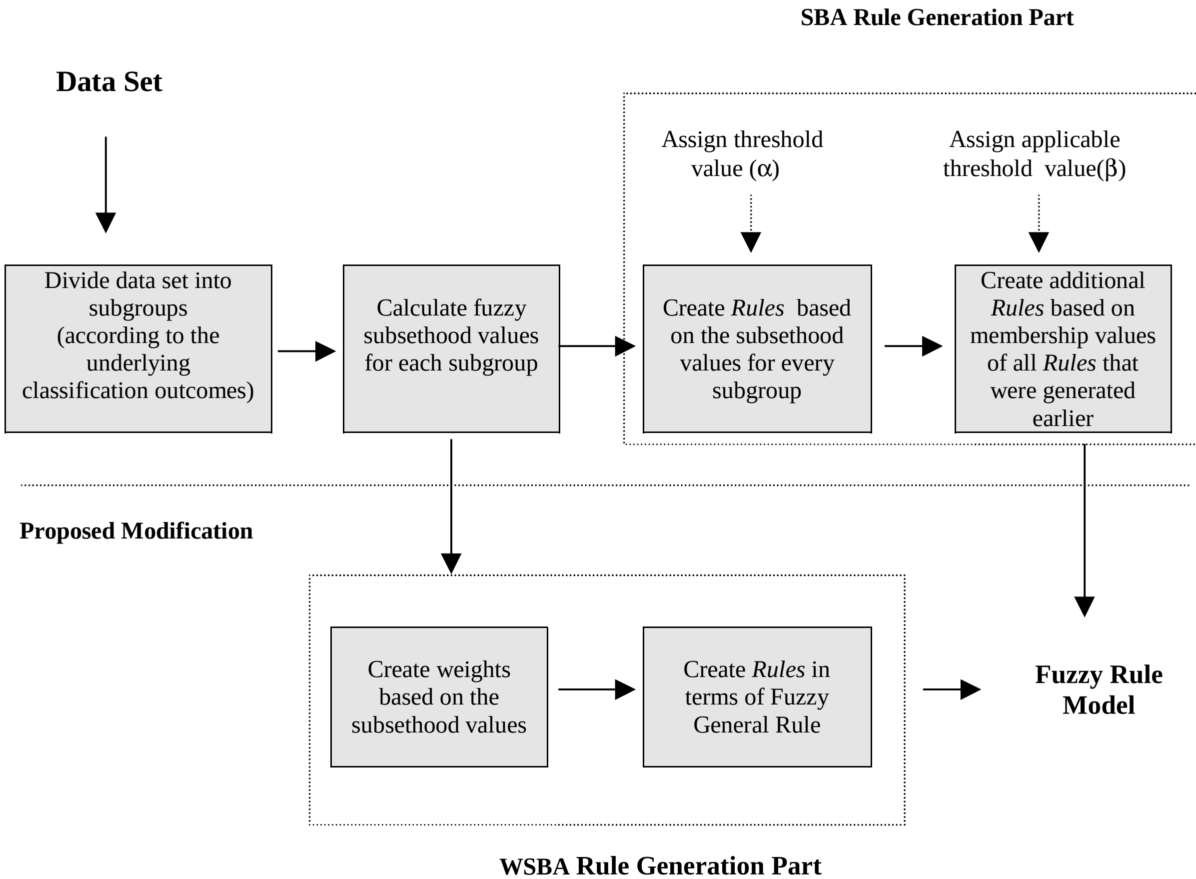

Figure 7: Proposed modification of SBA

5.2 Calculation of Weights from Subsethood Values

As with many existing techniques for representing weights, in this work, measures of weighting are limited to the range of 0 to 1, with 0 representing the lowest weight (or of least importance) and 1 the highest (or of most importance). Such weights can be calculated from fuzzy subsethood values as follows. Note that the meaning of *subsethood* is herein extended to allow fuzzy sets associated with different linguistic variables to be related.

Suppose that the subsethood value for a certain linguistic term A_j of linguistic variable A with regard to classification X is $S(X, A)$, and that the linguistic variable A has the following possible linguistic terms: $A_1, A_2, \ldots, A_l$. Then, the relative weight for linguistic term A_i, with regard to classification X is:

$$\mathrm{w}(X, A_i) = \frac{S(X, A_i)}{\max_{j=1..l} S(X, A_j)} \quad (21)$$

Clearly, $w(X, A_i) \in [0,1]$ and $i = 1, 2, \ldots, l$. This allows the creation of a weight for each linguistic term per condition attribute. Intuitively, the linguistic term with the highest subsethood value will be the most important and that with the lowest will be the least important. The weights generated for each antecedent of fuzzy rules are non-negative numbers and adjustable according the learning datasets.

The resulting weights are attached to the linguistic terms associated with conditional attributes. Therefore, for each conditional attribute A, the compound weight $T(A)$ of the weighted conjunction of linguistic terms associated with it can be calculated such that

$$T(A) = \left(\frac{w_1}{w}(A_1) \nabla \ \ldots \ \nabla \frac{w_m}{w}(A_m) \right) \quad (22)$$

where ∇ is the t-norm, A_i , $i = 1, 2, \ldots, m$, are the linguistic terms of variable A, which are conjuctively combined, and w is the largest amongst the m associated weights: $w(X, A_i)$, $i = 1, 2, \ldots, m$.

Similarly, the compound weight $T(B)$ of the weighted disjunction of linguistic terms associated with variable B is

$$T(B) = \left(\frac{w_1}{w}(B_1) \Delta \ \ldots \ \Delta \frac{w_n}{w}(B_n) \right) \quad (23)$$

where Δ is the *t-conorm*, and A_i , $i = 1, 2, \ldots, n$ are the linguistic terms of variable B, which are disjunctively combined.

5.3 WSBA Rule Generation

In order to make a system that is more readily comprehensible to the user, WSBA employs a rule generation algorithm which is based on *fuzzy general rules* or the extension of a Mamdani-type FRBS (refer to sub-section 3.1.3).

Consider fuzzy rules with multi-inputs and a single output. These rules can be written in the following form:

IF A is (A_1 *OR* A_2 *OR* …*OR* A_i) **AND** B is (B_1 *OR* B_2 *OR*… *OR* B_j) **AND** … **AND** H is (H_1 *OR* H_2 *OR* … *OR* H_k) **THEN** the classification output is (X_1 *OR* X_2 *OR*…*OR* X_n) (24)

This general rule can be re-written in a more specific form with each rule corresponding to one classification output value:

Rule 1 **IF** A is (A_1 *OR* A_2 *OR* …*OR* A_i) **AND** B is (B_1 *OR* B_2 *OR*… *OR* B_j) **AND** … **AND** H is (H_1 *OR* H_2 *OR* … *OR* H_k) **THEN** the classification output is X_1

Rule 2 **IF** A is (A_1 *OR* A_2 *OR* …*OR* A_i) **AND** B is (B_1 *OR* B_2 *OR*… *OR* B_j) **AND** … **AND** H is (H_1 *OR* H_2 *OR* … *OR* H_k) **THEN** the classification output is X_2

.

.

.

Rule n **IF** A is (A_1 *OR* A_2 *OR* …*OR* A_i) **AND** B is (B_1 *OR* B_2 *OR*… *OR* B_j) **AND** … **AND** H is (H_1 *OR* H_2 *OR* … *OR* H_k) **THEN** the classification output is X_n

(25)

Thus, all linguistic terms of each attribute are used to describe the antecedent of each rule initially. This may look tedious, but the reason for keeping this complete form is that every linguistic term may contain important information that should be taken into account. Otherwise, there is no need for adopting the given fuzzy partitions of the underlying domains in the first place. Of course, during training, some of such terms may be omitted due to no evaluated contribution (or with a relative weight of 0) with regard to the training data (see later).

However, the above default rules do not tell any differences between the relative contributions made by the individual linguistic terms of each variable towards the eventual conclusion drawn. It is here that relative weights computed via subsethood values can help. Following this idea, by multiplying each linguistic term by its respective weight, the fuzzy rules to be generated will be of the form:

Rule 1 **IF** A is $(w(X_1,A_1)A_1$ *OR* $(w(X_1,A_2)A_2$ *OR* …*OR* $w(X_1,A_i)A_i)$ **AND** B is $(w(X_1,B_1)B_1$ *OR* $w(X_1,B_2)B_2$ *OR*… *OR* $w(X_1,B_j)B_j)$ **AND** … **AND** H is $(w(X_1,H_1)H_1$ *OR* $w(X_1,H_2)H_2$ *OR* … *OR* $w(X_1,H_k)H_k)$ **THEN** the classification output is X_1

Rule 2 **IF** A is $(w(X_2,A_1)A_1$ *OR* $(w(X_2,A_2)A_2$ *OR* …*OR* $w(X_2,A_i)A_i)$ **AND** B is $(w(X_2,B_1)B_1$ *OR* $w(X_2,B_2)B_2$ *OR*… *OR* $w(X_2,B_j)B_j)$ **AND** … **AND** H is $(w(X_2,H_1)H_1$ *OR* $w(X_2,H_2)H_2$ *OR* … *OR* $w(X_2,H_k)H_k)$ **THEN** the classification output is X_2

.
.
.

Rule n **IF** A is $(w(X_n,A_1)A_1$ *OR* $(w(X_n,A_2)A_2$ *OR* …*OR* $w(X_n,A_i)A_i)$ **AND** B is $(w(X_n,B_1)B_1$ *OR* $w(X_n,B_2)B_2$ *OR*… *OR* $w(X_n,B_j)B_j)$ **AND** … **AND** H is $(w(X_n,H_1)H_1$ *OR* $w(X_n,H_2)H_2$ *OR* … *OR* $w(X_n,H_k)H_k)$ **THEN** the classification output is X_n

(26)

As with the original SBA algorithm, "OR" is interpreted by the t-conorm operator and "AND" by the t-norm operator, of course. In so doing, the proposed WSBA will produce linguistic rules that can be interpreted as a combination of *Fuzzy General Rules* and *Fuzzy Quantifier Rules*. The weights for each linguistic term are considered as a quantifier "some" or "all". If the weight = 1, the quantifier is regarded to be "all", otherwise it is considered to represent "some". The extent to which "some" is interpreted depends on the value of the weights of the respective linguistic terms.

It is interesting to note that, although this model is based on an extension of the Mamdani-type linguistic models, it can be readily extended to suit the TSK-type linguistic models. Note that in running the FRBS that employs such learned rules, the concluding classification will be that of the rule whose overall weight is the highest amongst all.

5.4 How this Method Works.

To demonstrate how this method works, a small dataset will be used. The dataset is an artificial set of Student Academic Performance (SAP) data. The dataset is devided into two subsets: SAP-1which will be used for training, and SAP-2, which will be used for testing. Each dataset contains 15 instances, five instances for each classification outcome. The SAP data comprises three attributes: Assignment, Test and Final Exam. The classification outcomes are one of the final grades: Poor, Average and Good. Table 1 shows the training dataset and Table 2 shows the testing dataset. Table 3 shows the labels used to represent the linguistic terms employed within the whole SAP dataset.

Case	Assignment (20%)	Test (30%)	Final Exam (50%)	Final Marks (100%)	Grade
1	2	12	10	24	Poor
2	16	24	45	85	Good
3	9	11	30	50	Average
4	5	17	17	39	Poor
5	10	20	15	45	Average
6	18	28	48	94	Good
7	15	20	23	58	Average
8	5	5	10	20	Poor
9	7	23	35	65	Average
10	15	17	43	75	Good
11	2	6	15	23	Poor
12	17	23	48	88	Good
13	15	25	50	90	Good
14	7	11	11	28	Poor
15	11	15	15	41	Average

Table1: Training dataset (SAP-1).

Case	Assignment (20%)	Test (30%)	Final Exam (50%)	Final Marks (100%)	Grade
1	2	7	9	18	Poor
2	8	11	20	39	Poor
3	6	10	14	30	Poor
4	7	14	16	37	Poor
5	3	6	19	28	Poor
6	11	10	21	42	Average
7	9	15	31	55	Average
8	8	22	27	57	Average
9	10	21	31	62	Average
10	17	25	24	66	Average
11	17	20	38	75	Good
12	15	24	41	80	Good
13	18	22	44	84	Good
14	16	26	48	90	Good
15	19	27	50	96	Good

Table 2: Testing dataset (SAP-2).

Label	Linguistic terms
A1	Assignment is Poor
A2	Assignment is Average
A3	Assignment is Good
B1	Test is Poor
B2	Test is Average
B3	Test is Good
C1	Final Exam is Poor
C2	Final Exam is Average
C3	Final Exam is Good
X	Final Grade is Poor
Y	Final Grade is Average
Z	Final Grade is Good

Table 3: Labels used for each linguistic term in SAP dataset.

Fuzzy Rule Generation

The procedure to create the required fuzzy model using WSBA is composed of the following 5 steps:

Step 1: Divide dataset into subgroups

The training dataset was divided into three subgroups according to the classification outcomes.

Subgroup	Cases	Outcome
Subgroup 1	1, 4, 8, 11, 14	Poor
Subgroup 2	3, 5, 7, 9, 15	Average
Subgroup 3	2, 6, 10, 12, 13	Good

Table 4: Subgroups of SAP-1 with respect to the classification outcomes.

Step 2: Define fuzzy partition

The fuzzy partition is pre-defined according to the criterion of evaluation. It will be used to transform crisp values (of both conditional attributes and classification results) into fuzzy values.

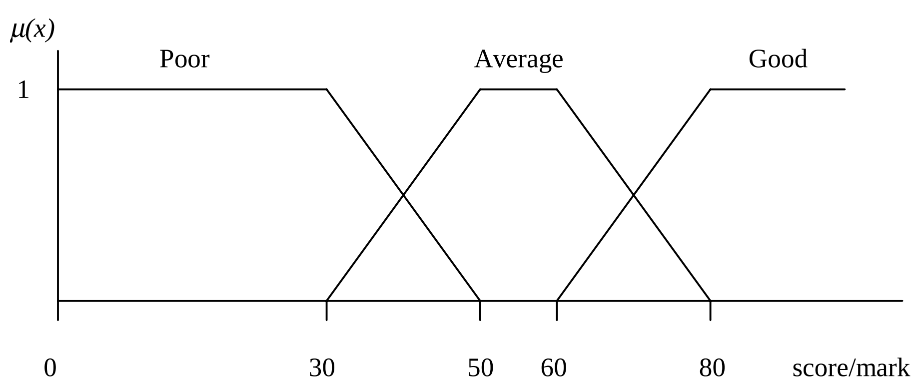

Figure 8: Fuzzy partition for SAP dataset.

Step 3: Calculate fuzzy subsethood values

Calculate fuzzy subsethood values (according to definition (10)) for each linguistic term in each subgroup. The subsethood values which were calculated according to each classification result, are shown in Table 5.

Linguistic Term	Final Score		
	Poor	Average	Good
A1	0.95	0.25	0.00
A2	0.25	0.63	0.11
A3	0.00	0.19	0.95
B1	0.7	0.16	0.00
B2	0.3	0.59	0.19
B3	0.00	0.35	0.81
C1	1.00	0.37	0.00
C2	0.04	0.57	0.00
C3	0.00	0.12	1.00

Table 5: Subsethood values calculated from SAP-1 dataset.

Step 4: Calculate weights for each linguistic term

Calculate weights for each linguistic term (according to definition (21)) using subsethood values calculated in Step 3.

Linguistic Term	Final Score		
	Poor	Average	Good
A1	1	0.40	0
A2	0.05	1	0.12
A3	0	0.30	1
B1	1	0.27	0
B2	0.43	1	0.23
B3	0	0.59	1
C1	1	0.68	0
C2	0.04	1	0
C3	0	0.21	1

Table 6: Weights for each linguistic term.

Step 5: Create rule

The ruleset that has been generated (according to definition (26)), are as follows:

Rule 1: The Final Grade is Poor (X)

> IF Assignment is (A1 OR 0.05A2) AND Test is (B1 OR 0.43B2) AND Final Exam is (C1 OR 0.04C2) THEN the Final Grade is Poor

Rule 2: The Final Grade is Average (Y)

> IF Assignment is (0.4A1 OR A2 OR 0.3A3) AND Test is (0.27B1 OR B2 OR 0.59B3) AND Final Exam is (0.68C1 OR C2 OR 0.21C3) THEN the Final Grade is Average

Rule 3: The Final Grade is Good (Z)

> IF Assignment is (0.12A2 OR A3) AND Test is (0.13B2 OR B3) AND Final Exam is (C3) THEN the Final Grade is Good

<u>Calculate the final classification (Grade)</u>

Once the ruleset is obtained, classification of student performance can be performed if observations over the three conditional attributes are obtained. The following example shows how a Final Grade can be decided.

Given that Assignment: 7 marks, Test: 13.5 marks, and Final Exam: 35 marks, their transformed fuzzy values are: $\mu_{A1}(x)$=0.75, $\mu_{A2}(x)$=0.25, $\mu_{A3}(x)$=0.00, $\mu_{B1}(x)$=0.25, $\mu_{B2}(x)$=0.75, $\mu_{B3}(x)$=0.00, $\mu_{C1}(x)$=0.00, $\mu_{C2}(x)$=0.5 and $\mu_{C3}(x)$=0.5

By using the ruleset generated in Step 5:

Rule 1: X = min[max (0.75, 0.0125) , max(0.25, 0.3225), max(0, 0.02)] = 0.02

Rule 2: Y = min[max (0.3, 0.25, 0) , max(0.0675, 0.75, 0), max(0, 0.5, 0.105)] = 0.3

Rule 3: Z = min[max (0.03, 0) , max(0.0975, 0), max(0.5)] = 0.0975

The classification result is Y (i.e. the final grade is Average) because the highest truth-value is associated with Rule 2. Here the Min-Max Operator is used, other choices of fuzzy logical operator such as mentioned in sub-section 3.1.2 may be utilised as alternatives.

Testing the ruleset for classification tasks

For testing the ruleset trained using SAP-1 for classification of student performance, the SAP-2 dataset is used. Table 7 shows comparisons of the outcomes obtained by using statistical average and those by the present approach. They happen to match very well for this simple example.

Case	Assign -ment (20%)	Test (30%)	Final Exam (50%)	Statistical Average		Fuzzy Approach			
						Fuzzy Membership Values*			New Grade
				Final Marks	Grade	Poor	Average	Good	
1	2	7	9	18	Poor	1	0.28	0	Poor
2	8	11	20	39	Poor	0.5	0.28	0	Poor
3	6	10	14	30	Poor	0.83	0.33	0	Poor
4	7	14	16	37	Poor	0.36	0.29	0	Poor
5	3	6	19	28	Poor	0.6	0.28	0	Poor
6	11	10	21	42	Average	0.06	0.23	0	Average
7	9	15	31	55	Average	0.04	0.75	0.08	Average
8	8	22	27	57	Average	0.04	0.4	0	Average
9	10	21	31	62	Average	0.04	0.5	0.1	Average
10	17	25	24	66	Average	0	0.29	0	Average
11	17	20	38	75	Good	0	0.2	0.33	Good
12	15	24	41	80	Good	0	0.22	0.75	Good
13	18	22	44	84	Good	0	0.22	0.67	Good
14	16	26	48	90	Good	0	0.22	1	Good
15	19	27	50	96	Good	0	0.22	1	Good

** Highest truth-value indicates the classification result*

Table 7: Comparison between statistical and fuzzy approaches over classification results.

Investigation on classification accuracy of the WSBA

To investigate the classification accuracy of the current WSBA, the Iris Plant dataset [51] has been used as a benchmark. A comparative study with regard to the original SBA was carried out.

The Iris Plant dataset contains four linguistic variables: sepal length, sepal width, petal length and petal width. It consists of 150 object instances and 3 classes: Iris-setosa, Iris-versicolor and Iris-virginica. Each class has an equal number of instances. To perform the experiment, the dataset was divided randomly into two sub-sets IP-1 and IP-2, each consisting of 25 instances per class.

All linguistic variables were defuzzified into three fuzzy partitions using a trapezoidal membership function representation. Table 8 shows all of the linguistic terms for each linguistic variable, and Tables 9 and 10 present the subsethood values calculated from training datasets IP-1 and IP-2.

Linguistic Variable	Linguistic Terms
Sepal length	SL1, SL2, SL3
Sepal width	SW1, SW2, SW3
Petal length	PL1, PL2, PL3
Petal width	PW1, PW2, PW3

Table 8: Linguistic labels for Iris Plant datatset.

Linguistic Label	Class of Plant		
	Iris-setosa	Iris-versicolor	Iris-virginica
SL1	0.92	0.13	0.00
SL2	0.08	0.78	0.64
SL3	0.00	0.09	0.35
SW1	0.00	0.34	0.19
SW2	0.84	0.64	0.81
SW3	0.17	0.02	0.00
PL1	1.00	0.04	0.00
PL2	0.00	0.91	0.20
PL3	0.00	0.05	0.80
PW1	1.00	0.05	0.00
PW2	0.00	0.93	0.17
PW3	0.00	0.02	0.83

Table 9: Subsethood values calculated from IP-1 dataset.

Linguistic Label	Class of Plant		
	Iris-setosa	Iris-versicolor	Iris-virginica
SL1	0.73	0.24	0.05
SL2	0.27	0.72	0.68
SL3	0.00	0.04	0.27
SW1	0.05	0.46	0.26
SW2	0.68	0.54	0.66
SW3	0.27	0.00	0.07
PL1	1.00	0.05	0.03
PL2	0.00	0.90	0.35
PL3	0.00	0.05	0.62
PW1	1.00	0.06	0.00
PW2	0.00	0.91	0.27
PW3	0.00	0.03	0.73

Table 10: Subsethood Values Calculated from IP-2 dataset.

Rulesets Generated by SBA

Two sets of rules were generated using SBA, one from each of the training datasets (with $\alpha = 0.9$):

Rules created using IP-1 as the training dataset:

Rule 1: IF SL is (SL1) AND SW is (NOT SW1) AND PL is (PL1) AND PW is (PW1) THEN the class is Iris-setosa.

Rule 2: IF SL is (NOT SL3) AND SW is (NOT SW3) AND PL is (PL2) AND PW is (PW2) THEN the class is Iris-versicolor.

Rule 3: IF SL is (NOT SL1) AND SW is (NOT SW3) AND PL is (NOT PL1) AND PW is NOT (PW1) THEN the class is Iris-virginica.

Rules created using IP-2 as the training dataset:

Rule 1: IF SL is (NOT SL3) AND SW is (NOT SW1) AND PL is (PL1) AND PW is (PW1) THEN the class is Iris-setosa.

Rule 2: IF SL is (NOT SL3) AND SW is (NOT SW3) AND PL is (PL2) AND PW is (PW2) THEN the class is Iris-versicolor.

Rule 3: IF SL is NOT SL1 AND SW is NOT SW3 AND PL is NOT PL1 AND PW is NOT PW1 THEN the class is Iris-virginica.

<u>Rulesets Generated by WSBA</u>

Similarly, two sets of rules are generated using WSBA from the IP-1 and IP-2 datasets:

Rules created using IP-1 as the training dataset:

Rule 1: IF SL is (SL1 OR 0.09SL2) AND SW is (SW2 OR 0.2SW3) AND PL is (PL1) AND PW is (PW1) THEN the class is Iris-setosa.

Rule 2: IF SL is (0.17SL1 OR SL2 OR 0.12SL3) AND SW is (0.53SW1 OR SW2 OR 0.03SW3) AND PL is (0.04PL1 OR PL2 OR 0.05PL3) AND PW is (0.05PW1 OR PW2 OR 0.02PW3) THEN the class is Iris-versicolor.

Rule 3: IF SL is (SL2 OR 0.55SL3) AND SW is (0.23SW1 OR SW2) AND PL is (0.25PL2 OR PL3) AND PW is (0.2PW2 OR PW3) THEN the class is Iris-virginica.

Rules created using IP-2 as the training dataset:

Rule 1: IF SL is (SL1 OR 0.37SL2) AND SW is (0.07SW1 OR SW2 OR 0.40SW3) AND PL is (PL1) AND PW is (PW1) THEN the class is Iris-setosa.

Rule 2: IF SL is (0.33SL1 OR SL2 OR 0.06SL3) AND SW is (0.85SW1 OR SW2) AND PL is (0.06PL1 OR PL2 OR 0.06PL3) AND PW is (0.07PW1 OR PW2 OR 0.03PW3) THEN the class is Iris-versicolor.

Rule 3: IF SL is (0.07SL1 OR SL2 OR 0.4SL3) AND SW is (0.39SW1 OR SW2 OR 0.11SW3) AND PL is (0.09PL1 OR 0.56PL2 OR PL3) AND PW is (0.37PW2 OR PW3) THEN the class is Iris-virginica.

Table 11 shows the classification accuracy of SBA and WSBA obtained by using IP-1 and IP-2 as the training datasets and the *Min-Max* Operators for fuzzy logical interpretation. The testing datasets are listed in this table.

Training Dataset	Testing Dataset	Accuracy Rate (SBA, α=0.9, β=0.6))	Accuracy Rate (WSBA)
IP-1	IP-2 (75 instances)	80% (15)	92% (6)
	Whole dataset (150 instances)	78% (33)	94.67% (8)
IP-2	IP-1 (75 instances)	81.33% (14)	96% (3)
	Whole dataset (150 instances)	78.69% (32)	94% (9)

Figure in brackets shows the number of misclassification

Table 11: Classification Accuracy of Rules Learned Using IP-1 and IP-2 dataset and the Min-Max Operator.

Clearly, in experiments carried out for the Iris-Plant dataset, the classification accuracy of the rules learned by WSBA is much better than that by SBA. Importantly, such results were obtained without the need of additional information for thresholding for WSBA, which SBA requires.

Another sets of experiments were performed to confirm the above results, using the Iris-Plant dataset, with it divided equally into two sub-datasets: IP-3 and IP-4. The entire dataset was labeled from 1 to 150, with IP-3 consisting of the odd numbered objects and IP-4 of the even numbered ones. Once again, the learned rules were employed for testing. The test datasets and the resultant classification accuracy measures are shown in Table 12.

Training Dataset	Testing Dataset	Accuracy Rate (SBA, α=0.9, β=0.6))	Accuracy Rate (WSBA)
IP-3	IP-4 (75 instances)	80% (15)	93.33% (5)
	Whole dataset (150 instances)	78.69% (32)	94.67% (8)
IP-4	IP-3 (75 instances)	78.67% (16)	93.33% (5)
	Whole dataset (150 instances)	78% (33)	93.33% (10)

Figure in brackets shows the number of misclassification

Table 12: Classification Accuracy of Rules Lesrned Using IP-3 and IP-4 dataset and the Min-Max Operator.

These experiments demonstrate that by using either IP-3 or IP-4, the accuracy rate of the system using rules learned by the WSBA is considerably better than that by SBA. The result shows the consistency of WSBA in producing high classification accuracy. With the current performance of WSBA, it is expected that it will produce good results when it is applied to real (student academic performance) datasets.

5.5 What is the New Approach Expected to Offer?

The new approach can be expected to offer several advantages to strengthen the evaluation using traditional statistical methods. These include:

(a) The objective of using the proposed method is to produce new scores (as shown in Table 7) which carry new information, alongside the original scores obtained by statistical or arithmetical methods. This may help to confirm or refute marks devised by humans or by other automated systems.

(b) In the fuzzy approach, there are degrees (membership values) by which a mark belongs to a grade (as shown in Table 7), whereas in traditional methods a crisp value (single numerical value) is usually used. This is very useful to confirm in certain borderline cases, to which grade a score actually belongs. Thus, the new score obtained from the fuzzy approach can be used as an aid to help the decision-maker (evaluator) to reach a decision.

(c) The use of a fuzzy rule-based approach which employs a linguistic fuzzy model will allow inference in the form of linguistic statement such as "If assignment is very poor and exam is average then the final result is poor". This reflects the natural way humans make judgements and decisions.

(d) The use of linguistic hedges in fuzzy approach such as "very", "more or less" and "extremely", allow more flexible judgements compared to numerical values, in particular when comparing student performance. Processing information given in term of linguistic hedges can only be done properly using fuzzy inference [6].

In particular, the WSBA that has been developed offers several significant advantages for application to student academic performance evaluation:

(a) Simplicity of the ruleset generated by WSBA. A fixed number of rules will be generated according to the number of classes. This is very useful to show what rules have been generated and which have been used for each classification. This is particularly an advantage when the method is applied to different levels of evaluation.

(b) Simplicity of the method. It generates default fuzzy rules without the need to use any threshold values because any linguistic term that has a weight greater than zero will be automatically promoted to become part of the antecedents of the resulting fuzzy rules. In so doing, any linguistic term that has a weight equal to 0 will be removed from the fuzzy rules. This is a significant advantage compared to more complicated methods that use threshold values such as the methods proposed in [9], [11] and [57]. The use of threshold values might also create confusion in student performance evaluation because different threshold values will create different rules and may result in different classification outcomes.

(c) By using WSBA, evaluation of student performance can be done hierarchically without the need to use other methods. In WSBA, the multiplication factor (i.e subsethood-based relative weights) will change the rulesets automatically according to the learning datasets. For example, a model that has been used to aggregate scores from assessment components to obtain a single score for a module, can also be used to aggregate scores from several modules to obtain a score for overall achievement.

All the criteria mentioned above are essential for the application of this method to real student academic performance datasets.

5.6 Further Works

The main goal of this research is to develop a data-driven fuzzy rule-based model which is simple but has a high accuracy rate. The model will be applied to classifying student academic performance data to produce a score or grade which is more meaningful to the user (student/teacher). In summary, this research will be carried out in two phases. The first phase is the development of WSBA and the second phase is the application of WSBA.

Some of the initial work has been completed, and was presented in previous subsections. In order to achieve the final outcomes, there are several further tasks that have to be carried out. These includes:

<u>Further improvements of WSBA</u>

(a) *Applicability of WSBA.* The existing WSBA uses three fuzzy partitions representing three basic linguistic terms for academic performance: bad, average and good. As academic achievement involves more than three basic linguistic terms (as above), WSBA need to be further improved. The prototype will then be tested with (i) datasets that have more than three classes (classification output), and (ii) datasets that have more than four attributes.

(b) *Model comprehensibility and computational complexity.* The existing WSBA makes a direct use of relative weights in the learned rules. This seems to reduce the resultant rules' readability. As interpretability plays a significant role in gaining popularity for fuzzy systems, a natural next step is to convert such weights into fuzzy linguistic quantifiers or hedges [33]. Also, to cope with problems that involve high dimensional datasets, supportive methods for attribute selection may be required (to increase the comprehensibility of the resulting models and to decrease the model's run-time computational complexity) [48].

Application of the improved WSBA to real student datasets.

(a) *Applications of WSBA for grading student performance.* WSBA will be trained using artificial student performance datasets to perform classification of student academic performance. The main aim of the application is to produce grades which are valid and reliable based on criterion-referenced evaluation, with the accuracy as high as possible compared to traditional evaluation methods.

(b) *Application of WSBA hierarchically.* WSBA will then be employed to perform the classification at a different level of student evaluation, such as mentioned in sub-section 2.5.

(c) *Application of WSBA for norm-referenced evaluation.* As mentioned in sub-section 2.4, the proposed method will be used to produce new scores which are based on historic or current data taken as the population. As this data has different features, it will produce a model based on the characteristics of the data. This will produce scores (evaluations) which refer to other scores in the population (i.e. norm-referenced evaluation).

(d) *Application of WSBA for combining norm-referenced and criterion evaluations.* The proposed method will also be employed to produce new scores that combine norm-referenced evaluations and criterion-referenced evaluations. The method which will be used, is still under development and requires further research.

A summary of the development and application of WSBA is shown in figures 9a and 9b.

Investigations	Aim
Building the initial model and iteratively revising the models via: (a) Testing the model using different fuzzy inference mechanisms. (b) Testing the model using different numbers of fuzzy partitions of input values. (c) Testing the accuracy rate using a small data set. (d) Testing the accuracy rate of the model using data from the UCI repository, such as the Iris Plant dataset, Wine Recognition dataset and Glass Identification dataset [51]. (e) Testing the accuracy rate using an artificial student performance data set.	Develop a data-driven FRBS which is suitable for student academic performance data and having high accuracy rate.

Figure 9a: Development of WSBA

Investigations	Aim
(f) Evaluation of student academic performance using a dataset obtained from the School of Informatics, University of Edinburgh. (g) Evaluation of student academic performance using a hierarchical fuzzy model. (h) Comparing the results obtained using this method to the results assign by experts, statistical methods and other fuzzy approaches.	Classify student academic performance and produce a score/ grade, which is more meaningful.

Figure 9b: Application of WSBA

6 Research Tasks

This section describes the research schedule, the activities to date and the plan for the rest of the study period.

6.1 Research Schedule and Major Progress So Far

The main tasks for this research are as follows:

(a) *Identification of problems*. Identify the problems in the current practice of traditional methods of student academic performance evaluation and choose specific problems to tackle.

(b) *Background theory*. Understand the background theory of fuzzy approaches and their application to classification problems, and analyze the previous and current study on using the fuzzy approach for educational evaluation.

(c) *Further theoretical investigations*. Investigate specific approaches for data-driven FRBS and Subsethood-based fuzzy rule generation algorithm.

(d) *Further investigations and improvement of the initial prototype.* Perform experiments as stated in sub-section 6.1, in an effort to produce a prototype that can handle the classification of student academic performance effectively.

(e) *Applications*. Apply the prototype to (i) classifying student academic performance using real dataset and (ii) evaluating student performance based on a hierarchical fuzzy model.

(f) *Analysis of results.* Analyze the results obtained from the prototype and compare them with the classification results obtained from (i) educational expert, (ii) traditional statistical method, and (iii) typical existing fuzzy approaches of student academic performance evaluation.

(g) *Thesis writing-up*.

Several important tasks have been largely completed so far, including the following:

(a) Literature Review

A literature review on using fuzzy approaches for student academic performance evaluation has been completed. New developments relating to this issue will be monitored closely.

(b) WSBA Building

The initial version of this has been established [42].

(c) Experiments with SBA and WSBA

A series of experiments has been performed using the SBA and WSBA and three different fuzzy logical interpretations. These experiments were conducted using the Saturday Morning Problem dataset (16 instances) and an artificial student academic performance dataset (60 instances). The details of the experimental results can be found in [41]. Another series of experiments has been performed for WSBA using Iris Plant dataset. The details of the experimental results can be found in [42].

(d) Presentation and publication

(i) Part of the work carried out was published in a paper presented at the 2002 UK Workshop on Computational Intelligence Birmingham (UKCI-02), under the title "Fuzzy Modelling for Student Academic Performance Evaluation".

(ii) Part of the work is currently under review for presentation at the 2003 IEEE International Conference on Fuzzy Systems USA (FUZZ-IEEE 2003), under the title "Weighted Linguistic Modelling Based on Fuzzy Subsethood Values".

6.2 Activities to Date and Plan

<u>**Time Spent for the First Eleven Months (November 2001 - August 2002)**</u>

<u>November - December 2001</u>

- Preliminary and background reading.
- Identifying problems to be tackled.
- Attending lectures which are relevant to the research area.

<u>January - Mac 2002</u>

- Reviewing the literature in the area of data-driven FRBS, in particular the SBA.
- Developing programming skills.
- Implementing the initial prototype based on the SBA.
- Further reading of the literature relevant to the research topic.
- Attending lectures which are relevant to the research area.

<u>April - June 2002</u>

- Analyzing experimental results based on the SBA.
- Further reading.

- Proposing new fuzzy rule generation method, the WSBA.
- Implementing the prototype WSBA.
- Testing the prototype for classification tasks using small dataset.
- Writing up a paper for presentation at UKCI-02.

July - September 2002

- Preparing prototype for more challenging classification tasks.
- Testing prototype using the Iris Plant dataset.
- Writing up the Ph.D. research proposal.
- Attending and presenting a paper.

Plan for the Next Six Months (October 2002 - March 2003)

October - November 2002

- Writing up a paper for the FUZZ-IEEE 2003.
- Preparing prototype for experiment using a dataset that contains more than four attributes. A Wine Recognition dataset [51] that contains three classes and nine attributes will be used for this experiment.
- Research Proposal refinement.
- Preparing for the presentation of the research proposal.

December 2002 - January 2003

- Preparing prototype for test using a dataset that contains more than three classes. A Glass Identification dataset [51] that contains 7 classes will be use for this experiment.
- Writing up a paper for presentation at a selected conference (e.g. UKCI-03) scheduled in the third quarter of 2003.

February - March 2003

- Analyzing results from the experiment conducted using the Wine Recognition dataset and Glass Identification dataset.
- Further reading.
- Improvement of the prototype.
- Writing up publishable material for publication in the area of educational evaluation.

Quarterly Plan for April 2003 - December 2004

April - Jun 2003

- Preparing prototype for applications to real Student Academic Performance (SAP) dataset.
- Gathering SAP dataset for the application of WSBA.
- Attending and presenting a paper at the FUZZ-IEEE 2003.

July - September 2003

- Application of the prototype to SAP dataset.
- Analysis of results.
- Attending and presenting a paper at the UKCI-03.
- Preparing an article for refereed journal publication.

October - December 2003

- Further improvement of the prototype.
- Revision of main tasks.
- Analysis of results and further investigation.

January - March 2004

- Revision of main tasks.
- Writing/composition of existing materials.
- Attending a further conference.

April - June 2004 and July - September 2004

- Thesis writing up.
- Preparing an extended external publication.

October - December 2004

- Thesis refinement.

Quarterly plan for April 2003 - December 2004 can be referred to figure 10 which shows plan for main tasks and ongoing task. Revision of main tasks may introduce new tasks throughout this research.

Tasks													
(g)													
(f)													
(e)													
(d)													
(c)													
(b)													
(a)													
Month	Nov. - Dec	Jan. - Mar.	Apr. - Jun	Jul. - Sep.	Oct. - Dec.	Jan. - Mar.	Apr. - Jun	Jul. - Sep.	Oct. - Dec.	Jan. - Mar.	Apr. - Jun	Jul. - Sep.	Oct. - Dec.
Year	2001	2002				2003				2004			

Legend:

	Main task
	Revision of main task

Figure 10: Summary of quarterly plan (November 2001 - December 2004)

7 Conclusion

This report has presented a proposal for research in the area of data driven FRBS for student academic performance evaluation. The main purpose of this research is to show that this new approach will have several benefits to strengthen the current traditional arithmetical and statistical methods. To achieve this objective, an initial data-driven FRBS called Weighted Subsethood-Based Rule Generation Algorithm (WSBA) has been developed. Although the application of this method the area of educational evaluation is quite new, a series of experiments carried out using small student academic performance dataset and the Iris Plant dataset have show a promising results.

The development of WSBA, which is based on fuzzy subsethood values, offers simplicity by generating default fuzzy rules without the need to use any threshold value. This is quite useful in the area of educational evaluation, which needs a system that is easily understood by many people such as educators, policy planners, parents and students.

This report also describes the previous attempts to use a fuzzy approach in educational evaluation and discusses some advantages and disadvantages of the proposed methods. This is useful particularly in giving direction on what this research should focus.

The work to be carried out in the rest of this project, if successful, will lead to the establishment of a systematic approach to evaluating student academic performance, which will help reinforce decision made by alternative methods and reveal from one more angle the agreeable marks awarded to boundary cases where traditional marking methods are difficult to decide on.

References

[1] Alcala, R., Casillas, J., Cordon, O. and Herera, F. (1999). Approximate Mamdani-type Fuzzy Rule-Based Systems: Features and Taxonomy of Learning Methods, Unpublished Technical Report, Department of Computer Science and Artificial Intelligence, Universidad de Granada.

[2] Alcala, R., Casillas, J., Cordon, O., Herera, F. and Zwir, S.J.I. (1999). Techniques for Learning and Tuning Fuzzy Rule-Based Systems for Liguistic Modeling and Their Application, in Leondes, C.T. (ed.), *Knowledge Engineering, Systems, Techniques and Applications.*

[3] Ashworth, A.E. (1982). *Testing for Continuous Assessment*, Evans Brothers Limited, London.

[4] Atkins, M.J., Beattie, J. and Dockrell, W.B. (1993). *Assessment Issues in Higher Education*, Employment Department Group: United Kingdom.

[5] Bassey, M. (1998). Fuzzy Generalization: An Approach to Building Educational Theory. British Educational Research Association Annual Conference, The Queen's University of Belfast, Northern Ireland.

[6] Biswas, R. (1995). An Application of Fuzzy Sets in Students' Evaluation, *Fuzzy Sets and Systems*, 74 : 187-194.

[7] Capper, J. (1996). *Testing to Learn - Learning to Test*, International Reading Association and the Academy for Educational Development, United States.

[8] Capuano, N., Marcella, M. and Salerno, S. (2000). ABITS: An Agent Based Intelligent Tutoring System for Distance Learning. Available online: http://virtcampus.cl-ki.uni-osnabrueck.de/its-2000/paper/capuano/ws2-paper3.htm. [October 1, 2000].

[9] Castro J.L. and Zurita J.M. (1997). An Inductive Learning Algorithm in Fuzzy Systems, *Fuzzy Sets And Systems*, 89 : 193 - 203.

[10] Chen, S.M. and Lee, C.H. (1999). New Methods for Students' Evaluation Using Fuzzy Sets, *Fuzzy Sets and Systems*, 104 : 209 - 218.

[11] Chen, S.M., Lee, S.H. and Lee, C.H. (2001). A New Method for Generating Fuzzy Rules From Numerical Data for Handling Classification Problems, *Applied Artificial Intelligence*, 15 : 645 - 664.

[12] Chiang, I. J. and Hsu Y. J. (2002). Fuzzy Classification Trees for Data Analysis, *Fuzzy Sets and Systems*, 130 : 87 - 99.

[13] Cordon, O., Herera, F. and Peregrin, A. (1999). Looking for the Best Defuzzification Method Features for each Implication Operator to Design Accurate Fuzzy Models. Unpublished Technical Report, Department of Computer Science and Artificial Intelligence, Universidad de Granada.

[14] Cordon, O., Herera, F. and Zwir, I. (2002). Linguistic Modeling by Hierarchical Systems of Linguistic Rules, *IEEE Transactions on Fuzzy Systems*, 10(1) : 2 - 20.

[15] Crockett, K.A., Bandar, Z. and Al-Attar, A. (2000). Soft Decision Trees: A New Approach Using Non-linear Fuzzification, *Proceedings of the FUZZ-IEEE'2000 Conference*, USA, 209 - 215.

[16] Dressel, P.L. (1976). *Handbook of Academic Evaluation*, Jossey-Bass Limited, London.

[17] Dubois, D. and Prade, H. (1996). What Are Fuzzy Rules and How to Use Them, *Fuzzy Sets and Systems*, 84 : 169 - 185.

[18] Ebel, R.L. and Frisbie, D.A. (1991). *Essentials of Educational Measurement-Fifth Edition*, Prentice Hall, New Jersey.

[19] Fourali, C. (1994). Fuzzy Logic and the Quality of Assessment of Portfolios, *Fuzzy Sets and Systems*, 68 : 123 - 139.

[20] Gipps, C.V. (1994). *Beyond Testing: Towards a Theory of Educational Assessment*, The Falmer Press: London.

[21] Gipps, C. and Stobart, G. (1993). *Assessment: A Teachers' Guide to the Issue*, Hodder and Stoughton Educational, London.

[22] Graesser, A.C., Person, N. and Harter, D. (2001). Teaching Tactics and Dialog in AutoTutor. *International Journal of Artificial Intelligence in Education,* 12(3) : 257-279.

[23] Herera, F. and Herera-Viedma, E. (1995). Aggregation Operators for Weighted Linguistic Weighted Information, Unpublished Technical Report, Department of Computer Science and Artificial Intelligence, Universidad de Granada.

[24] Heywood, J. (1989). *Assessment in Higher Education: Second Edition,* John Wiley and Sons, Chichester.

[25] Hong, T.P. and Lee, C.Y. (1996). Induction of Fuzzy Rules and Membership Functions from Training Examples, *Fuzzy Sets and Systems,* 84(1) : 33 - 47.

[26] Hounsell, D., McCulloch, M and Scott, M. (eds) (1996). *The ASSHE Inventory: Changing Assessment Practices in Scottish Higher Education.* Centre for Teaching, Learning and Assessment, The University of Edinburgh.

[27] Janikow, C.Z. (1998). Fuzzy Decision Trees: Issues and Methods. *IEEE Transactions on Systems , Man and Cybernatics*, 28 (1) : 1-14.

[28] Khrisnapuram, R. (1998). Membership function elicitation and learning, in Ruspini, E. H., Bonissone, P.P and Pedrycz, W. (eds) (1998) *Handbook of Fuzzy Computation*, Institute of Physics Publishing, Dirac House, Temple Bath, Bristol.

[29] Klir, G.J. and Yuan, B. (1998). Operation of fuzzy sets, in Ruspini, E. H., Bonissone, P.P and Pedrycz, W. (eds) (1998) *Handbook of Fuzzy Computation*, Institute of Physics Publishing, Dirac House, Temple Bath, Bristol.

[30] Kosko, B. (1993). *Fuzzy Thinking*, Harper Collins, London.

[31] Law, C-K. (1995), Using Fuzzy Numbers in Educational Grading System, *Fuzzy Sets and Systems,* 83 : 311- 323.

[32] Luo, X. and Zhang, C. (2002). The Weighting Issue in Fuzzy Logic, *Informatica: An International Journal of Computing and Informatica*, 21 : 255 - 262.

[33] Marin-Blazquez, J.G. and Shen, Q. (2002). From Approximate to Descriptive Fuzzy Classifiers, *IEEE Transactions on Fuzzy Systems*, 10 (4) : 484 -509.

[34] Mendel, J.M. (2001), *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions,* Prentice Hall PTR, New Jersey.

[35] Nauck, D. and Kruse, R. (1998). Creating Fuzzy Rules from data, in Ruspini, E. H., Bonissone, P.P and Pedrycz, W. (eds) (1998) *Handbook of Fuzzy Computation*, Institute of Physics Publishing, Dirac House, Temple Bath, Bristol.

[36] Nauck, D., Nauck, U. and Kruse, R. (1996) Generating Classification Rules with the Neuro-Fuzzy System NEFCLASS, *Proceeding of Biennial Conference of the North American Fuzzy Information Processing Society NAFIPS' 96*, Berkeley, CA.

[37] Nozaki, K., Ishibuchi, H. and Tanaka, H. (1997). A Simple but Powerful Heuristic Method for Generating Fuzzy Rules From Numerical Data, *Fuzzy Sets and Systems*, 86(3) : 251- 270.

[38] Pedrycz, W. (1998). *Computational Intelligence: An Introduction,* CRC Press, New York.

[39] Pedrycz, W. and Gomide, F. (1998). *An Introduction to Fuzzy Sets: Analysis and Design*, The MIT Press, Cambridge, Massachusetts.

[40] Popov, D.I. (1999). Intelligent System of Distance Education in Internet With Fuzzy Logic Model of Knowledge Evaluation. Available online: http://www.actr.org/ sep/English/1999Fall/FuzzyPaper/FuzzyPaperPub.html. [August 15, 2002].

[41] Rasmani, K.A. and Shen, Q. (2002). Fuzzy Modelling for Student Academic Performance Evaluation, *Proceeding of the UK Workshop in Computational Intelligence,* University of Birmingham

[42] Rasmani, K.A and Shen, Q. (2002). Weighted Linguistic Modelling Based on Fuzzy Subsethood Values. Paper submitted to the International Conference on Fuzzy Systems, May 2003, St. Louis USA.

[43] Ridgway, J. (1988). *Assessing Mathematical Attainment*, NFER-NELSON, Berkshire.

[44] Ruspini, E. H., Bonissone, P.P and Pedrycz, W. (eds) (1998) *Handbook of Fuzzy Computation*, Institute of Physics Publishing, Dirac House, Temple Bath, Bristol.

[45] Satterly, D. (1981). *Assessment in Schools*, Basil Blackwell, Oxford.

[46] Schmucker, K.J, (1989). *Fuzzy Sets, Natural Language Computations, and Risk Analysis*, Computer Science Press, Maryland.

[47] Schwartz, D.G. (1998). Fuzzy mathematical objects, in Ruspini, E. H., Bonissone, P.P and Pedrycz, W. (eds) (1998) *Handbook of Fuzzy Computation*, Institute of Physics Publishing, Dirac House, Temple Bath, Bristol.

[48] Shen, Q. and Chouchoulas, A. (2000). A Modular Approach to Generating Fuzzy Rules with Reduced Attributes for the Monitoring of Complex Systems. *Engineering Applications of Artificial Intelligence*, 13 (3) : 263-278.

[49] Shen, Q. and Chouchoulas, A. (2002). A Fuzzy-Rough Approach for Generating Classification Rules, *Pattern Recognition*, 35(11) : 341-354.

[50] Shimizu, S. and Yamashita, H. (2000), Educational Evaluation of Calligraphy: Applying Fuzzy Reasoning. Available online: http://www.coe.uh.edu/insite/ elec_pub/HTML1997 /id_shim.htm [October 9, 2000].

[51] UCI Machine Learning Databases. Available online: [http:// ftp.ics.uci.edu/pub/ machine-learning-databases/ [March 10, 2002].

[52] Wang, L.X. (1997). *A Course in Fuzzy Systems and Control*. Prentice-Hall International, New Jersey.

[53] Wang, X.Z., Wang, Y.D., Xu, X.F., Ling, W.D. and Yeung, D.S. (2001). A New Approach to Fuzzy Rule Generation: Fuzzy Extension Matrix, *Fuzzy Sets and Systems,* 123 : 291-306.

[54] Webb, N.L. (1992). Assessment of Students' Knowledge of Mathematics: Steps Toward a Theory, in Grouws, D.G. (ed) (1992). *Handbook of Research in Mathematics Teaching and Learning*, Maxwell Macmillan International, New York.

[55] Weon, S. and Kim, J. (2001). Learning Achievement Evaluation Strategy Using Fuzzy Membership Function, *31st ASEE/IEEE Frontiers in Education Conference*, October 10 - 13, Reno, NV, 19 - 24.

[56] Yen, J. and Wang, L. (1998). Granule-based models, in Ruspini, E. H., Bonissone, P.P and Pedrycz, W. (eds) (1998) *Handbook of Fuzzy Computation*, Institute of Physics Publishing, Dirac House, Temple Bath, Bristol.

[57] Yuan, Y. and Shaw, M.J. (1995). Induction of Fuzzy Decision Trees, *Fuzzy Set and Systems,* 69 (2) : 125 - 139.

[58] Yuan, Y. and Zhuang, H. (1996). A Genetic Algorithm for Generating Fuzzy Classification Rules, *Fuzzy Sets and Systems*, 84 (1) : 1 - 19.

[59] Zadeh, L.A. (1988). Fuzzy Logic, *IEEE - CS Computer*, 21(4) : 83 - 93.

[60] Zhou, D., Ma, J., Turban, E. and Bolloju, N. (2002). A Fuzzy Set Approach to the Evaluation of Journal Grades, *Fuzzy Sets and Systems*, 131 : 63 - 74.