Expert Systems with Applications 38 (2011) 11129–11136

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

# A fuzzy inference system-based criterion-referenced assessment model

Kai Meng Tay [a,*], Chee Peng Lim [b]

[a] Electronic Engineering Department, Faculty of Engineering, Universiti Malaysia Sarawak, Malaysia
[b] School of Electrical and Electronic Engineering, University of Science, Malaysia

ARTICLE INFO

*Keywords:*
Criterion-referenced assessment
Fuzzy inference system
Monotonicity property
Sub-additivity property
*Sufficient conditions*
Rule refinement

ABSTRACT

The main aim of criterion-referenced assessment (CRA) is to report students' achievements in accordance with a set of references. In practice, a score is given to each test item (or task). The scores from different test items are added together and then projected or aggregated, usually linearly, to produce a total score. Each component score can be weighted before being added together in order to reflect the relative importance of each test item. In this paper, the use of a fuzzy inference system (FIS) as an alternative to the conventional addition or weighted addition in CRA is investigated. A novel FIS-based CRA model is presented, and two important properties, i.e., the monotonicity and sub-additivity properties, of the FIS-based CRA model are investigated. A case study relating to assessment of laboratory projects in a university is conducted. The results indicate the usefulness of the FIS-based CRA model in comparing and assessing students' performances with human linguistic terms. Implications of the importance of the monotonicity and sub-additivity properties of the FIS-based CRA model in undertaking general assessment problems are discussed.



## 1. Introduction

Assessment in education is defined as a process of forming judgement about quality and extent of students' achievements or performances. Judgement usually is based on information obtained by requiring students to attempt a number of specified test items or tasks, and submit their work for an appraisal of the quality. In criterion-referenced assessment (CRA), students' grades are determined by comparing their achievements with a set of clearly defined criteria of learning outcomes. The main aim of CRA is to evaluate students' achievements with reference to a set of objective reference points, which can be a simple pass–fail grading schema, or a series of key criteria (Burton, 2007; Sadler, 2005). There is a possibility for all students within a particular group to obtain very high or very low grades depending on the individuals' performances against the established criteria. From the literature, the use of CRA in essay writing (Patrick & Phan, 2005; White, 2002) and clinical performance (Patricia, Shelley, & Trisha, 2009) has been reported.

It has been pointed out that scoring usually refers to test items/tasks rather than to the overall achievement (Joughin, 2008; Sadler, 2005). To ease the process of assessment, in general, a score is given to each test item/task. The individual scores are then aggregated to produce a *total score*. Sadler (2005) pointed that the scores from different test items/tasks can be added together and then projected, usually in a linear manner. Each score can also be weighted to reflect the relative importance of each task (Sadler, 2005).

The use of fuzzy set related techniques in education assessment models is not new. Biswas (1995) presented a fuzzy set related method to evaluate students' answer scripts. The work was further enhanced by Chen and Lee (1999). Ma and Zhou (2000) investigated another fuzzy set related method for assessment of student-centered learning tasks. Saliu (2005) used the fuzzy inference system (FIS) in CRA for Constrained Qualitative Assessment (CQA) with a case study. Kwok, Zhau, Zhang, and Ma (2007) proposed a fuzzy group multi-criteria decision making model for CRA of student group projects. Cin and Baba (2008) employed the FIS for English proficiency assessment. Chang and Chen (2009) proposed a fuzzy peer assessment system to satisfy the requirements of cooperative learning in an e-learning environment. Bai and Chen (2008) presented a fuzzy membership function and fuzzy rule-based approach that takes into consideration the difficulty, importance, and complexity of the questions in evaluating answer scripts from students. The proposed approach is able to distinguish the ranking order of students with the same score. The applicability of the FIS to other assessment models, i.e., FIS-based Failure More and Effect Analysis (FMEA) (Bowles & Peláez, 1995; Tay & Lim, 2008a, 2008b), FIS-based groundwater vulnerability assessment (Afshar, Mariño, Ebtehaj, & Moosavi, 2007) has also been reported.

In this paper, a new FIS-based CRA model that utilizes rubric as the scoring tool and that involves subjectivity in learning is

* Corresponding author.
*E-mail address:* kmtay@feng.unimas.my (K.M. Tay).

0957-4174/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.02.158

proposed. The FIS-based CRA model acts as an alternative to aggregate the score of each test item/task, and to produce a *total score*. We examine a number of reasons to support the use of the FIS, instead of the conventional addition or weighted addition, in assessment models. Inspired by the theoretical properties of a *length function* (Inder, 2005), we further propose two properties, i.e., the monotonicity and sub-additivity properties, to be associated with the FIS-based CRA model in order to ensure that a useful and valid comparison among students' performances can be made. In this paper, we investigate and propose the *sufficient conditions* for the FIS-based CRA model to preserve the monotonicity property (Kouikoglou & Phillis, 2009; Tay & Lim, 2008a, 2008b). A rule-refinement method to improve the sub-additivity property of the FIS-based CRA model is also suggested.

The organization of this paper is as follows. Motivations of this work are presented in Section 2. Background of the FIS, the theoretical properties of monotonicity and sub-additivity, the *sufficient conditions*, and other complementary techniques is given in Section 3. Details of the proposed FIS-based CRA model and its application to a laboratory project assessment task are presented in Section 4. Concluding remarks are presented in Section 5.

## 2. Motivations

In this paper, we focus on the CRA model that utilizes rubric as the scoring tool for the test items/tasks, where subjectivity is involved. A rubric provides a list of criteria, and it helps to grade the quality from bad to good for each criterion (Kurt & Izmirli, 2009). It enables students' performances from various aspects to be assessed.

Scoring usually refers to test items/tasks, rather than to the overall achievement (Joughin, 2008; Sadler, 2005). The score from each test item is aggregated to produce a *total score*. The scores from different assessments are then added and projected (Sadler, 2005). Each score can be weighted before being added to reflect the relative importance of each task (Sadler, 2005). As an example, weighted addition in essay writing assessment was employed in White (2002).

In this paper, the FIS is adopted as an alternative to simple addition or weighted addition. The idea of replacing simple addition or weighted addition with a more complicated method is not new. Sadler (2005) pointed out that aggregation of scores can be done by some designed algorithm or mathematical equation. Chen, Cheng, and Liu (2010) employed an adaptive ordered weighted averaging operator and a K-nearest-neighbor classification method to simulate teachers' evaluation behaviors.

The FIS is adopted owing to several reasons. First, the rubric criteria can be qualitative (rather than quantitative), e.g., a rubric score of 4 does not mean two times better than that of 2. Saliu (2005) considered the FIS as a solution to qualitative assessment, with the aim to keep qualitative assessment accountable. Second, the relative importance of each test item can be different. Indeed, the importance of each test item depends on the learning outcome. The FIS is able to customize the relationship between the score of each test item and the aggregated score. Third, various combinations of the scores from different test items can produce the same aggregated score, but the performance of the student can be different. The FIS can be deployed to tackle this problem.

In addition to the above reasons, the FIS has a good function approximation capability, as demonstrated in a variety of problems including control, modeling, and classification (Jang, Sun, & Mizutani, 1997). Another advantage of using the FIS is its capability in incorporating human/expert knowledge, whereby information can be described by using vague and imprecise statements. Besides, the behavior of an FIS can be easily interpreted by humans, whereby the relationship between its input(s) and output(s) is described by a set of fuzzy If-Then rules.

From the literature, FIS-based assessment models are normally initiated as an alternative to the conventional assessment models to allow modeling of nonlinear relationship between the input(s) and output(s) (Afshar et al., 2007; Bowles & Peláez, 1995). The surface plot in Saliu (2005) demonstrates a nonlinear relationship between the score associated with each test item (input) and the total score (output).

Over the years, one of the research directions of FIS-based assessment models is on their theoretical properties. Kouikoglou and Phillis (2009) argued that monotonicity is a *natural requirement* in assessment models. Tay and Lim (2008a,2008b) explained the importance of the monotonicity and output resolution properties for an FIS-based FMEA model. Broekhoven and Baets (2008,2009) further pointed out that monotonicity property is a common property in evaluation and selection procedures. In this paper, we explain the importance of the monotonicity and sub-additivity properties of the FIS-based CRA model from the theory of *length function* (Inder, 2005). Saliu (2005) considered the failure of an FIS-based CRA model to fulfill the monotonicity property as an anomaly, and efforts to construct a monotonicity-preserving FIS-based CRA is essential.

In this work, the main aim is to develop a simple (which can be easily understood and visualized by users), easy-to-use (which can be easily incorporated into an FIS-based assessment model), and yet reliable (which can be interpreted from the theoretical foundation) method to preserve the monotonicity property of the FIS. One of the practical and effective solutions is to apply the *sufficient conditions* to the FIS. In this paper, we investigate the applicability of the *sufficient conditions* to the proposed FIS-based CRA model. A rule refinement method is also adopted to improve the sub-additivity property of the FIS-based CRA model.

## 3. Fuzzy inference systems and the monotonicity and sub-additivity properties

To make this paper self-contained, the background information on the FIS is presented, with monotonicity and sub-additivity defined. The *sufficient conditions* and the rule refinement techniques are also examined.

### 3.1. Fuzzy inference systems

Consider an FIS with $n$ inputs. Let $\bar{x} = (x_1, x_2, \ldots, x_n)$ be the input vector in a rectangular region, $U = U_1 \times U_2 \times \cdots \times U_n$, $U_i = [L_i, H_i]$ for $1 \leqslant i \leqslant n$. Consider $M_i$ terms at the $i$th input space, $A_i^1, A_i^2, \ldots, A_i^{M_i}$, which are represented by fuzzy membership functions $\mu_i^1(x_i)$, $\mu_i^2(x_i), \ldots, \mu_i^{M_i}(x_i)$, respectively. The output of the FIS, $y = f(\bar{x})$, falls within the range of $[L_V, H_V]$. If a full grid partition is used, the total number of fuzzy rules is $\prod_{i=1}^{n} M_i$.

The fuzzy production rules can be represented as follows:

$$\begin{aligned}&\mathbf{R}^{\mathbf{j_1, j_2, \ldots, j_n}}\ (\mathbf{Rule}\#\mathbf{M'}):\\ &\mathbf{IF}(x_1\ \mathbf{is}\ \mathbf{A}_1^{j_1})\ \mathbf{AND}\ (x_2\ \mathbf{is}\ \mathbf{A}_2^{j_2})\cdots\mathbf{AND}\ (x_n\ \mathbf{is}\ \mathbf{A}_n^{j_n})\\ &\mathbf{THEN}\ y\ \mathbf{is}\ B^{j_1, j_2, \ldots, j_n}\end{aligned}$$

where $1 \leqslant \mathbf{j}_i \leqslant M_i$. The product function is the AND operator in the rule antecedent. The compatibility grade, or known as the firing strength, of each fuzzy rule, i.e., $\mathbf{R}^{\mathbf{j_1 j_2 \cdots j_n}}$, is defined as $\mu_1^{j_1}(x_1) \times \mu_2^{j_2}(x_2) \times \cdots \times \mu_n^{j_n}(x_n)$. To simplify the notation, each fuzzy rule is represented by an index, $M'$, where $1 \leqslant M' \leqslant \prod_{i=1}^{n} M_i$. The output is obtained by using the weighted average of a representative real value, $b^{j_1 j_2 \cdots j_n}$, with respect to its compatibility grade, as in Eq. (1)