Expert Systems with Applications 36 (2009) 9839–9846

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Evaluating students' answerscripts based on interval-valued fuzzy grade sheets

Shyi-Ming Chen [a,c,*], Hui-Yu Wang [b]

[a] Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan, ROC
[b] Department of Education, National Chengchi University, Taipei, Taiwan, ROC
[c] Department of Computer Science and Information Engineering, Jinwen University of Science and Technology, Taipei County, Taiwan, ROC

ARTICLE INFO

Keywords:
Similarity functions
Students' answerscripts
Interval-valued fuzzy grade sheets
Interval-valued fuzzy sets
Index of optimism

ABSTRACT

In this paper, we propose two new methods for evaluating students' answerscripts based on interval-valued fuzzy grade sheets. The marks awarded to the answers in the students' answerscripts are represented by interval-valued fuzzy sets. Each element in the universe of discourse belonging to an interval-valued fuzzy set is represented by an interval between zero and one. The degree of similarity between an interval-valued fuzzy mark and a standard interval-valued fuzzy set is calculated by a similarity function. An index of optimism λ determined by the evaluator is used to indicate the degree of optimism of the evaluator, where $\lambda \in [0,1]$. The experimental results show that the proposed methods are more stable than Biswas's method for students' answerscripts evaluation. They can evaluate students' answerscripts in a more flexible and more intelligent manner.

1. Introduction

Biswas (1995) pointed out that a chief goal of educational institutions is to provide students with evaluation reports regarding their examination as sufficient as possible and with the unavoidable error as small as possible. In recent years, some methods have been presented for students' evaluation (Bai & Chen, 2008a, 2008b, 2008c; Biswas, 1995; Chang & Sun, 1993; Chen & Lee, 1999; Cheng & Yang, 1998; Chiang & Lin, 1994; Echauz & Vachtsevanos, 1995; Frair, 1995; Hwang, Lin, & Lin, 2006; Kaburlasos, Marinagi, & Tsoukalas, 2004; Law, 1996; Ma & Zhou, 2000; McMartin, Mckenna, & Youssefi, 2000; Pears, Daniels, Berglund, & Erickson, 2001; Wang & Chen, 2007a, 2007b; Weon & Kim, 2001; Wu, 2003). Bai and Chen (2008a) presented a method for evaluating students' learning achievement using fuzzy membership functions and fuzzy rules. Bai and Chen (2008b) presented a method for automatically constructing concept maps based on fuzzy rules for adaptive learning systems. Bai and Chen (2008c) presented a method for automatically constructing grade membership functions of fuzzy rules for students' evaluation. Biswas (1995) presented a fuzzy evaluation method (*fem*) for applying fuzzy sets in students' answerscripts evaluation. He also proposed a generalized fuzzy evaluation method (*gfem*) for students' answerscripts evaluation. Chang and Sun (1993) presented a method for fuzzy assessment of learning performance of junior high school students. Chen and Lee (1999) presented two methods for evaluating students' answerscripts using fuzzy sets. Cheng and Yang (1998) presented a method for using fuzzy sets in educational grading systems. Chiang and Lin (1994) presented a method for applying the fuzzy set theory to teaching assessment. Frair (1995) presented a method for student peer evaluations using the analytic hierarchy process method. Echauz and Vachtsevanos (1995) presented a fuzzy grading system to translate a set of scores into letter grades. Hwang et al. (2006) presented an approach for test-sheet composition with large-scale item banks. Kaburlasos et al. (2004) presented a software tool, called PARES, for computer-based testing and evaluation used in the Greek higher educational system. Law (1996) presented a method for applying fuzzy numbers in educational grading systems. Ma and Zhou (2000) presented a fuzzy set approach for the assessment of student-centered learning. McMartin et al. (2000) used scenario assignments as assessment tools for undergraduate engineering education. Pears et al. (2001) presented a method for student evaluation in an international collaborative project course. Wang and Chen (2007a) presented two methods for students' answerscripts evaluations based on the similarity measure between vague sets (Gorzalczany, 1987, 1989). Weon and Kim (2001) presented a leaning achievement evaluation strategy using fuzzy membership functions. Wu (2003) presented a method for applying the fuzzy set theory and the item response theory to evaluate learning performance.

In Biswas (1995), the fuzzy marks awarded to answers in the students' answerscripts are represented by fuzzy sets (Zadeh, 1965). In a fuzzy set, the grade of membership of an element u_i

* Corresponding author. Address: Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan, ROC. Tel.: +886 2 27376417; fax: +886 2 27301081.
E-mail address: smchen@mail.ntust.edu.tw (S.-M. Chen).

0957-4174/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2009.02.005

in the universe of discourse U belonging to a fuzzy set is represented by a real value between zero and one, however, if we can allow the marks awarded to the questions of the students' answerscripts to be represented by interval-valued fuzzy sets (Gorzalczany, 1987, 1989), then there is room for more flexibility, where the grade of membership of an element in the universe of discourse belonging to an interval-valued fuzzy set is represented by an interval in [0,1].

In this paper, we present two new methods for students' answerscripts evaluation based on interval-valued fuzzy grade sheets. The marks awarded to the answers in the students' answerscripts are represented by interval-valued fuzzy sets. The degree of similarity between an interval-valued fuzzy mark and a standard interval-valued fuzzy set is calculated by a similarity function. An index of optimism λ determined by the evaluator is used to indicate the degree of optimism of the evaluator, where $\lambda \in [0,1]$. If $0 \leqslant \lambda < 0.5$, then the evaluator is a pessimistic evaluator. If $\lambda = 0.5$, then the evaluator is a normal evaluator. If $0.5 < \lambda \leqslant 1.0$, then the evaluator is an optimistic evaluator. We also make an experiment to compare the experimental results of the proposed method with the ones using Biswas's method (1995). The experimental results show that the proposed methods are more stable than Biswas's method for students' answerscripts evaluation. The proposed methods can evaluate students' answerscripts in a more flexible and more intelligent manner.

The rest of this paper is organized as follows. In Section 2, we briefly review similarity measures between interval-valued fuzzy sets from (Chen, 1994). In Section 3, we briefly review Biswas's methods for students' answerscripts evaluation from (Biswas, 1995). In Section 4, we present a new method for students' answerscripts evaluation based on interval-valued fuzzy grade sheets. We also present a generalized fuzzy evaluation method for students' answerscripts evaluation using interval-valued fuzzy grade sheets. In Section 5, we make an experiment to compare the experimental results of the proposed method with Biswas's method (1995). The conclusions are discussed in Section 6.

2. Similarity measures between interval-valued fuzzy sets

Zwick, Carlstein, and Budescu (1987) presented a method for measuring the distance between two real intervals. Let X and Y be two intervals in $[\beta_1,\beta_2]$, where $X = [x_1,x_2]$ and $Y = [y_1,y_2]$. The distance $D(X,Y)$ between the intervals X and Y is calculated as follows:

$$D(X,Y) = \frac{|x_1 - y_1| + |x_2 - y_2|}{2(\beta_2 - \beta_1)}. \quad (1)$$

Then, the degree of similarity $S(X,Y)$ between the intervals X and Y can be calculated as follows (Zwick et al., 1987):

$$S(X,Y) = 1 - D(X,Y). \quad (2)$$

Let X and Y be two intervals in [0,1], where $X = [x_1,x_2]$, $Y = [y_1,y_2]$, $0 \leqslant x_1 \leqslant x_2 \leqslant 1$, and $0 \leqslant y_1 \leqslant y_2 \leqslant 1$. Based on Eqs. (1) and (2), the degree of similarity $S(X,Y)$ between the intervals X and Y can be calculated as follows (Zwick et al., 1987):

$$S(X,Y) = \begin{cases} 1, & \text{if } y_1 \leqslant x_1 \leqslant x_2 \leqslant y_2, \\ 1 - \frac{|x_1 - y_1| + |x_2 - y_2|}{2}, & \text{otherwise}, \end{cases} \quad (3)$$

where $S(X,Y) \in [0,1]$. It is obvious that if X and Y are identical intervals, then $D(A,B) = 0$ and $S(X,Y) = 1$. The larger the value of $S(X,Y)$, the higher the similarity between the intervals X and Y.

Assume that x and y are two real values between zero and one, where $x = [x,x]$ and $y = [y,y]$, then based on Eq. (3), we can see that

$$S(x,y) = S([x,x],[y,y]) = 1 - \frac{|x-y| + |x-y|}{2} = 1 - |x-y|. \quad (4)$$

Let $\widetilde{A}$ and $\widetilde{B}$ be two interval-valued fuzzy sets in the universe of discourse X, where

$$X = \{x_1, x_2, \ldots, x_n\},$$
$$\widetilde{A} = [a_{11}, a_{12}]/x_1 + [a_{21}, a_{22}]/x_2 + \cdots + [a_{n1}, a_{n2}]/x_n,$$
$$\widetilde{B} = [b_{11}, b_{12}]/x_1 + [b_{21}, b_{22}]/x_2 + \cdots + [b_{n1}, b_{n2}]/x_n,$$

$[a_{i1},a_{i2}]$ denotes the grade of membership of x_i belonging to the interval-valued fuzzy set $\widetilde{A}$, $[b_{i1}, b_{i2}]$ denotes the grade of membership of x_i belonging to the interval-valued fuzzy set $\widetilde{B}$, $0 \leqslant a_{i1} \leqslant a_{i2} \leqslant 1$, $0 \leqslant b_{i1} \leqslant b_{i2} \leqslant 1$, and $1 \leqslant i \leqslant n$. Based on the matrix representation method, the interval-valued fuzzy sets $\widetilde{A}$ and $\widetilde{B}$ can be represented by the matrices $\overline{A}$ and $\overline{B}$, respectively, where

$$\overline{A} = \langle [a_{11}, a_{12}], [a_{21}, a_{22}], \ldots, [a_{n1}, a_{n2}] \rangle,$$
$$\overline{B} = \langle [b_{11}, b_{12}], [b_{21}, b_{22}], \ldots, [b_{n1}, b_{n2}] \rangle.$$

If $\widetilde{A}$ and $\widetilde{B}$ are identical interval-valued fuzzy sets (i.e., $\widetilde{A} = \widetilde{B}$), then $a_{ij} = b_{ij}$, $1 \leqslant i \leqslant n$ and $1 \leqslant j \leqslant 2$. In this situation, we can see that $\overline{A} = \overline{B}$.

By applying Eq. (3), the degree of similarity $T(\overline{A}, \overline{B})$ between the interval-valued fuzzy sets $\widetilde{A}$ and $\widetilde{B}$ can be calculated by the similarity function T (Zwick et al., 1987),

$$T(\overline{A}, \overline{B}) = \frac{\sum_{i=1}^{n} S([a_{i1}, a_{i2}], [b_{i1}, b_{i2}])}{n} = \frac{\sum_{i=1}^{n} 1 - \frac{|a_{i1} - b_{i1}| + |a_{i2} - b_{i2}|}{2}}{n}, \quad (5)$$

where $T(\overline{A}, \overline{B}) \in [0,1]$. The larger the value of $T(\overline{A}, \overline{B})$, the higher the similarity between the interval-valued fuzzy sets $\widetilde{A}$ and $\widetilde{B}$. It is obvious that if $\widetilde{A}$ and $\widetilde{B}$ are identical interval-valued fuzzy sets (i.e., $\widetilde{A} = \widetilde{B}$), then $T(\overline{A}, \overline{B}) = 1$.

In Eq. (5), we assume that all elements in the universe of discourse U are of equal importance. However, if we can allow each element in the universe of discourse U to have a different degree of importance, then there is room for more flexibility. Chen (1994) also considers the situation that each element u_i in the universe of discourse U has a different degree of importance. Assume that the degree of importance of each u_i in the universe of discourse U is described by a weighted matrix $\overline{W}$, $\overline{W} = \langle w_1, w_2, \ldots, w_n \rangle$, where w_i denotes the weight of u_i in U, $0 \leqslant w_i \leqslant 1$, and $1 \leqslant i \leqslant n$. Let $\overline{A}$ and $\overline{B}$ be the matrix representations of the interval-valued fuzzy sets $\widetilde{A}$ and $\widetilde{B}$, respectively, where

$$\overline{A} = \langle [a_{11}, a_{12}], [a_{21}, a_{22}], \ldots, [a_{n1}, a_{n2}] \rangle,$$
$$\overline{B} = \langle [b_{11}, b_{12}]), [b_{21}, b_{22}]), \ldots, [b_{n1}, b_{n2}] \rangle.$$

Then, the degree of similarity $G(\overline{A}, \overline{B}, \overline{W})$ between the interval-valued fuzzy sets $\widetilde{A}$ and $\widetilde{B}$ can be calculated by the similarity function G (Chen, 1994),

$$G(\overline{A}, \overline{B}, \overline{W}) = \frac{\sum_{i=1}^{n} [1 - (|a_{i1} - b_{i1}| + |a_{i2} - b_{i2}|)/2] \times w_i}{\sum_{i=1}^{n} w_i}, \quad (6)$$

where $G(\overline{A}, \overline{B}, \overline{W}) \in [0,1]$. The larger the value of $G(\overline{A}, \overline{B}, \overline{W})$, the higher the similarity between the interval-valued fuzzy sets $\widetilde{A}$ and $\widetilde{B}$. It is obvious that if $\widetilde{A}$ and $\widetilde{B}$ are identical interval-valued fuzzy sets, then $G(\overline{A}, \overline{B}, \overline{W}) = 1$. Furthermore, if $w_1 = w_2 = \cdots = w_n$, then $G(\overline{A}, \overline{B}, \overline{W}) = T(\overline{A}, \overline{B})$.

3. A review of Biswas' methods for students' answerscripts evaluation

Biswas (1995) used the matching function S to measure the degree of similarity between fuzzy sets (Chen, 1988; Zadeh, 1965). Let A and B be two fuzzy sets of the universe of discourse X, where

Table 1
A fuzzy grade sheet (Biswas, 1995).

Question no.	Fuzzy mark						Grade
Q.1	x_1	x_2	x_3	x_4	x_5	x_6	
Q.2							
Q.3							
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Q.n							
					Total mark =		

$$X = \{x_1, x_2, \ldots, x_n\},$$
$$A = \mu_A(x_1)/x_1 + \mu_A(x_2)/x_2 + \cdots + \mu_A(x_n)/x_n,$$
$$B = \mu_A(x_1)/x_1 + \mu_A(x_2)/x_2 + \cdots + \mu_A(x_n)/x_n,$$

μ_A denotes the membership function of the fuzzy set A, $\mu_A(x_i) \in [0,1]$, μ_B denotes the membership function of the fuzzy set B, $\mu_B(x_i) \in [0,1]$, and $1 \leqslant i \leqslant n$. The fuzzy sets A and B can be represented by the vectors $\overline{A}$ and$\overline{B}$, respectively, where

$$\overline{A} = \langle \mu_A(x_1), \mu_A(x_2), \ldots, \mu_A(x_n) \rangle,$$
$$\overline{B} = \langle \mu_B(x_1), \mu_B(x_2), \ldots, \mu_B(x_n) \rangle.$$

Then, the degree of similarity $S(\overline{A}, \overline{B})$ between the fuzzy sets A and B is calculated as follows (Chen, 1988; Chen & Wang, 1995):

$$S(\overline{A}, \overline{B}) = \frac{\overline{A} \cdot \overline{B}}{Max(\overline{A} \cdot \overline{A}, \overline{B} \cdot \overline{B})}, \tag{7}$$

where $S(\overline{A}, \overline{B}) \in [0,1]$. The larger the value of $S(\overline{A}, \overline{B})$, the higher the similarity between the fuzzy sets A and B.

Biswas (1995) presented a "fuzzy evaluation method" (*fem*) for evaluating students' answerscripts based on the matching function S. He used five fuzzy linguistic hedges, called standard fuzzy sets (SFS) of the universe of discourse X for students' answerscripts evaluation, i.e., ***E*** (excellent), ***V*** (very good), ***G*** (good), ***S*** (satisfactory) and ***U*** (unsatisfactory). He used the vector representation method to represent the fuzzy sets ***E***, ***V***, ***G***, ***S*** and ***U*** by the vectors $\overline{E}, \overline{V}, \overline{G}, \overline{S}$ and $\overline{U}$, respectively. He pointed out that "***A***", "***B***", "***C***", "***D***" and "***E***" are letter grades, where $90 \leqslant \boldsymbol{A} \leqslant 100$, $70 \leqslant \boldsymbol{B} < 90$, $50 \leqslant \boldsymbol{C} < 70$, $30 \leqslant \boldsymbol{D} < 50$ and $0 \leqslant \boldsymbol{E} < 30$. Furthermore, he presented the concept of "mid-grade-points", where the mid-grade-points of the letter grades ***A***, ***B***, ***C***, ***D*** and ***E*** are $P(\boldsymbol{A})$, $P(\boldsymbol{B})$, $P(\boldsymbol{C})$, $P(\boldsymbol{D})$ and $P(\boldsymbol{E})$, respectively, $P(\boldsymbol{A}) = \mathbf{95}$, $P(\boldsymbol{B}) = 80$, $P(\boldsymbol{C}) = 60$, $P(\boldsymbol{D}) = 40$ and $P(\boldsymbol{E}) = 15$. Assume that an evaluator evaluates the ith question (i.e., *Q.i*) of a student's answerscript using a fuzzy grade sheet as shown in Table 1. In the second row of Table 1, the real values x_1, x_2, x_3, x_4, x_5 and x_6 awarded to the question *Q.1* indicate that the degrees of satisfaction of the evaluator for that answer are 0%, 20%, 40%, 60%, 80% and 100%, respectively, where $0 \leqslant x_i \leqslant 1$ and $1 \leqslant i \leqslant 6$. By using the vector representation method, the fuzzy mark F_1 can be represented by the vector $\overline{F_1}$, where $\overline{F_1} = \langle x_1, x_2, x_3, x_4, x_5, x_6 \rangle$.

In the following, we briefly review Biswas' algorithm for students' answerscript evaluation from (Biswas, 1995) as follows:

Step 1: For each question *Q.i*, in the answerscript, where $1 \leqslant i \leqslant n$, repeatedly perform the following tasks:

(1) The evaluator awards a fuzzy mark F_i to each question *Q.i* by his/her best possible judgment and fills up each cell of the ith row for the first seven columns, where $1 \leqslant i \leqslant n$. Let $\overline{F_i}$ be the vector representation of F_i, where $1 \leqslant i \leqslant n$.

(2) Calculate the degrees of similarity $S(\overline{E}, \overline{F_i}), S(\overline{V}, \overline{F_i}), S(\overline{G}, \overline{F_i}), S(\overline{S}, \overline{F_i})$ and $S(\overline{U}, \overline{F_i})$, respectively, where $\overline{E}, \overline{V}, \overline{G}, \overline{S}$ and $\overline{U}$ are the vector representations of the standard fuzzy sets ***E*** (excellent), ***V*** (very good), ***G*** (good), ***S*** (satisfactory) and ***U*** (unsatisfactory), respectively.

(3) Find the maximum value among the values of $S(\overline{E}, \overline{F_i}), S(\overline{V}, \overline{F_i}), S(\overline{G}, \overline{F_i}), S(\overline{S}, \overline{F_i})$ and $S(\overline{U}, \overline{F_i})$. If $S(\overline{E}, \overline{F_i})$ is the maximum value among the values of $S(\overline{E}, \overline{F_i}), S(\overline{V}, \overline{F_i}), S(\overline{G}, \overline{F_i}), S(\overline{S}, \overline{F_i})$ and $S(\overline{U}, \overline{F_i})$, then award the letter grade "***A***" to the question *Q.i* due to the fact that the letter grade "***A***" corresponds to ***E*** (excellent) of the standard fuzzy set; if $S(\overline{V}, \overline{F_i})$ is the maximum value among the values of $S(\overline{E}, \overline{F_i})$, $S(\overline{V}, \overline{F_i}), S(\overline{G}, \overline{F_i}), S(\overline{S}, \overline{F_i})$ and $S(\overline{U}, \overline{F_i})$, then award the letter grade "***B***" to the question *Q.i* due to the fact that the letter grade "***B***" corresponds to ***V*** (very good) of the standard fuzzy set; if $S(\overline{G}, \overline{F_i})$ is the maximum value among the values of $S(\overline{E}, \overline{F_i}), S(\overline{V}, \overline{F_i}), S(\overline{G}, \overline{F_i})$, $S(\overline{S}, \overline{F_i})$ and $S(\overline{U}, \overline{F_i})$, then award the letter grade "***C***" to the question *Q.i* due to the fact that the letter grade "***C***" corresponds to ***G*** (good) of the standard fuzzy set; if $S(\overline{S}, \overline{F_i})$ is the maximum value among the values of $S(\overline{E}, \overline{F_i}), S(\overline{V}, \overline{F_i}), S(\overline{G}, \overline{F_i}), S(\overline{S}, \overline{F_i})$ and $S(\overline{U}, \overline{F_i})$, then award the letter grade "***D***" to the question *Q.i* due to the fact that the letter grade "***D***" corresponds to ***S*** (satisfactory) of the standard fuzzy set; if $S(\overline{U}, \overline{F_i})$ is the maximum value among the values of $S(\overline{E}, \overline{F_i})$, $S(\overline{V}, \overline{F_i}), S(\overline{G}, \overline{F_i})$, $S(\overline{S}, \overline{F_i})$and $S(\overline{U}, \overline{F_i})$, then award the letter grade "***E***" to the question *Q.i* due to the fact that the letter grade "***E***" corresponds to ***U*** (unsatisfactory) of the standard fuzzy set.

Step 2: Calculate the total mark of the student as follows:

$$\text{Total Mark} = \frac{1}{100} \times \sum_{i=1}^{n} [T(Q.i) \times P(g_i)], \tag{8}$$

where $T(Q.i)$ denotes the mark allotted to question *Q.i* in the test paper, g_i denotes the derived letter grade awarded to *Q.i* by Step 1 of the algorithm, and $P(g_i)$ denotes the mid-grade-point of g_i. Put this total score in the appropriate box at the bottom of the fuzzy grade sheet.

Biswas (1995) also presented a generalized fuzzy evaluation method (*gfem*) for students' answerscripts evaluation, where a generalized fuzzy grade sheet shown in Table 2 is used to evaluate the students' answerscripts. In the generalized fuzzy grade sheet shown in Table 2, for all j = 1,2,3,4 and for all i, g_{ij} denotes the derived letter grade by the fuzzy evaluation method *fem* for the awarded fuzzy mark F_{ij} and m_i denotes the derived mark awarded to the question *Q.i*, where

$$m_i = \frac{1}{400} \times T(Q.i) \times \sum_{j=1}^{4} P(g_{ij}), \tag{9}$$

Table 2
A generalized fuzzy grade sheet (Biswas, 1995).

Question no.	Sub-questions	Fuzzy mark	Derived letter grade	Mark
Q.1	*Q.11*	F_{11}	g_{11}	m_1
	Q.12	F_{12}	g_{12}	
	Q.13	F_{13}	g_{13}	
	Q.14	F_{14}	g_{14}	
Q.2	*Q.21*	F_{21}	g_{21}	m_2
	Q.22	F_{22}	g_{22}	
	Q.23	F_{23}	g_{23}	
	Q.24	F_{24}	g_{24}	
⋮	⋮	⋮	⋮	⋮
Q.n	*Q.n1*	F_{n1}	g_{n1}	m_n
	Q.n2	F_{n2}	g_{n2}	
	Q.n3	F_{n3}	g_{n3}	
	Q.n4	F_{n4}	g_{n4}	
			Total mark =	

T(Q.i) denotes the mark allotted to question *Q.i* in the question paper, g_{ij} denotes the derived letter grade awarded to sub-question *Q.ij*, $P(g_{ij})$ denotes the mid-grade-point of g_{ij}, $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant 4$, and the Total Mark $= \sum_{i=1}^{n} m_i$.

4. A new method for evaluating students' answerscripts based on interval-valued fuzzy grade sheets

In this section, we present a new method for evaluating students' answerscripts based on interval-valued fuzzy grade sheets. Let X be the universe of discourse, where $X = \{0\%, 20\%, 40\%, 60\%, 80\%, 100\%\}$. Biswas (1995) used the five fuzzy linguistic hedges $\boldsymbol{E}$ (excellent), $\boldsymbol{V}$ (very good), $\boldsymbol{G}$ (good), $\boldsymbol{S}$ (satisfactory) and $\boldsymbol{U}$ (unsatisfactory) of the universe of discourse X, called the standard fuzzy sets, for students' answerscripts evaluation, defined as follows:

$$\boldsymbol{E} = 0/0\% + 0/20\% + 0.8/40\% + 0.9/60\% + 1/80\% + 1/100\%,$$
$$\boldsymbol{V} = 0/0\% + 0/20\% + 0.8/40\% + 0.9/60\% + 0.9/80\% + 0.8/100\%,$$
$$\boldsymbol{G} = 0/0\% + 0.1/20\% + 0.8/40\% + 0.9/60\% + 0.4/80\% + 0.2/100\%,$$
$$\boldsymbol{S} = 0.4/0\% + 0.4/20\% + 0.9/40\% + 0.6/60\% + 0.2/80\% + 0/100\%,$$
$$\boldsymbol{U} = 1/0\% + 1/20\% + 0.4/40\% + 0.2/60\% + 0/80\% + 0/100\%.$$

It is obvious that these five standard fuzzy sets $\boldsymbol{E}$ (excellent), $\boldsymbol{V}$ (very good), $\boldsymbol{G}$ (good), $\boldsymbol{S}$ (satisfactory) and $\boldsymbol{U}$ (unsatisfactory) also can equivalently be represented by interval-valued fuzzy sets $\widetilde{E}$, $\widetilde{V}$, $\widetilde{G}$, $\widetilde{S}$ and $\widetilde{U}$, respectively, where

$$\widetilde{E} = [0,0]/0\% + [0,0]/20\% + [0.8,0.8]/40\% + [0.9,0.9]/60\% + [1,1]/80\% + [1,1]/100\%,$$

$$\widetilde{V} = [0,0]/0\% + [0,0]/20\% + [0.8,0.8]/40\% + [0.9,0.9]/60\% + [0.9,0.9]/80\% + [0.8,0.8]/100\%,$$

$$\widetilde{G} = [0,0]/0\% + [0.1,0.1]/20\% + [0.8,0.8]/40\% + [0.9,0.9]/60\% + [0.4,0.4]/80\% + [0.2,0.2]/100\%,$$

$$\widetilde{S} = [0.4,0.4]/0\% + [0.4,0.4]/20\% + [0.9,0.9]/40\% + [0.6,0.6]/60\% + [0.2,0.2]/80\% + [0,0]/100\%,$$

$$\widetilde{U} = [1,1]/0\% + [1,1]/20\% + [0.4,0.4]/40\% + [0.2,0.2]/60\% + [0,0]/80\% + [0,0]/100\%.$$

The standard interval-valued fuzzy sets $\widetilde{E}, \widetilde{V}, \widetilde{G}, \widetilde{S}$ and $\widetilde{U}$ can be represented by matrices $\overline{E}$, $\overline{V}$, $\overline{G}$, $\overline{S}$ and $\overline{U}$, respectively, where

$$\overline{E} = \langle [0,0], [0,0], [0.8,0.8], [0.9,0.9], [1,1], [1,1] \rangle,$$
$$\overline{V} = \langle [0,0], [0,0], [0.8,0.8], [0.9,0.9], [0.9,0.9], [0.8,0.8] \rangle,$$
$$\overline{G} = \langle [0,0], [0.1,0.1], [0.8,0.8], [0.9,0.9], [0.4,0.4], [0.2,0.2] \rangle,$$
$$\overline{S} = \langle [0.4,0.4], [0.4,0.4], [0.9,0.9], [0.6,0.6], [0.2,0.2], [0,0] \rangle,$$
$$\overline{U} = \langle [1,1], [1,1], [0.4,0.4], [0.2,0.2], [0,0], [0,0] \rangle.$$

Assume that "***A***", "***B***", "***C***", "***D***" and "***E***" are letter grades, where $90 \leqslant \boldsymbol{A} \leqslant 100$, $70 \leqslant \boldsymbol{B} < 90$, $50 \leqslant \boldsymbol{C} < 70$, $30 \leqslant \boldsymbol{D} < 50$ and $0 \leqslant \boldsymbol{E} < 30$. Assume that an evaluator evaluates the ith question (i.e., *Q.i*) of a student's answerscript using an interval-valued fuzzy grade sheet as shown in Table 3. In the second row of the interval-valued fuzzy grade sheet shown in Table 3, the interval-valued fuzzy marks [0,0], [0.2,0.3], [0.4,0.5], [0.6,0.7], [0.8,0.9] and [1,1] awarded to the answer to question *Q.1* indicate that the degrees of the evaluator's satisfaction for that answer are 0%, 20%, 40%, 60%, 80% and 100%, respectively. Let this interval-valued fuzzy mark of the answer to question *Q.1* be denoted by $\widetilde{F}_1$. Then, we can see that $\widetilde{F}_1$ is an interval-valued fuzzy set of the universe of discourse X, where

Table 3
An interval-valued fuzzy grade sheet.

Question no.	Interval-valued fuzzy mark						Derived fuzzy letter grade
Q.1	[0,0]	[0.2,0.3]	[0.4,0.5]	[0.6,0.7]	[0.8,0.9]	[1,1]	
Q.2							
Q.3							
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Q.n							
					Total mark =		

$$X = \{0\%, 20\%, 40\%, 60\%, 80\%, 100\%\},$$
$$\widetilde{F}_1 = [0,0]/0\% + [0.2,0.3]/20\% + [0.3,0.4]/40\% + [0.6,0.7]/60\% + [0.8,0.9]/80\% + [1,1]/100\%.$$

In this case, $\widetilde{F}_1$ can be represented by a matrix $\overline{F}_1$, shown as follows:

$$\overline{F}_1 = \langle [0,0], [0.2,0.3], [0.3,0.4], [0.6,0.7], [0.8,0.9], [1,1] \rangle.$$

The proposed interval-valued fuzzy evaluation method (*IVFEM*) for students' answerscripts evaluation is presented as follows:

Step 1: For each question in the answerscript repeatedly perform the following tasks:

(1) The evaluator awards an interval-valued fuzzy mark $\widetilde{F}_i$ represented by an interval-valued fuzzy set to each question *Q.i* by his/her judgment and fills up each cell of the ith row for the first seven columns, where $1 \leqslant i \leqslant n$. Let $\overline{F}_i$ be the matrix representation of the interval-valued fuzzy mark $\widetilde{F}_i$ of question *Q.i*, where $1 \leqslant i \leqslant n$.

(2) Based on Eq. (5), calculate the degrees of similarity $H(\overline{E},\overline{F}_i), H(\overline{V},\overline{F}_i), H(\overline{G},\overline{F}_i), H(\overline{S},\overline{F}_i)$ and $H(\overline{U},\overline{F}_i)$, respectively, where $\overline{E}, \overline{V}, \overline{G}, \overline{S}$ and $\overline{U}$ are matrix representations of the standard fuzzy sets $\widetilde{E}$ (excellent), $\widetilde{V}$ (very good), $\widetilde{G}$ (good), $\widetilde{S}$ (satisfactory) and $\widetilde{U}$ (unsatisfactory), respectively. Assume that $H(\overline{E},\overline{F}_i) = \beta_{i1}, H(\overline{V},\overline{F}_i) = \beta_{i2}, H(\overline{G},\overline{F}_i) = \beta_{i3}, H(\overline{S},\overline{F}_i) = \beta_{i4}$ and $H(\overline{U},\overline{F}_i) = \beta_{i5}$, where $\beta_{ij} \in [0,1]$, $1 \leqslant i \leqslant n$, and $1 \leqslant j \leqslant 5$.

(4) Because the standard fuzzy sets $\widetilde{E}, \widetilde{V}, \widetilde{G}, \widetilde{S}$ and $\widetilde{U}$ correspond to the letter grades "***A***", "***B***", "***C***", "***D***" and "***E***", respectively, the derived fuzzy letter grade $\tilde{g}_i$ of question *Q.i* is represented by a fuzzy set, shown as follows:

$$\tilde{g}_i = \beta_{i1}/\boldsymbol{A} + \beta_{i2}/\boldsymbol{B} + \beta_{i3}/\boldsymbol{C} + \beta_{i4}/\boldsymbol{D} + \beta_{i5}/\boldsymbol{E}, \tag{10}$$

where ***A***, ***B***, ***C***, ***D*** and ***E*** are letter grades, $H(\overline{E},\overline{F}_i) = \beta_{i1}, H(\overline{V},\overline{F}_i) = \beta_{i2}, H(\overline{G},\overline{F}_i) = \beta_{i3}, H(\overline{S},\overline{F}_i) = \beta_{i4}$ and $H(\overline{U},\overline{F}_i) = \beta_{i5}, \beta_{ij} \in [0,1], 1 \leqslant i \leqslant n$, and $1 \leqslant j \leqslant 5$.

Step 2: Calculate the total mark of the student as follows:

$$\text{Total Mark} = \frac{1}{100} \times \sum_{i=1}^{n} [T(Q.i) \times K(\tilde{g}_i)], \tag{11}$$

where *T(Q.i)* denotes the mark allotted to the question *Q.i* in the test paper, $\tilde{g}_i$ denotes the fuzzy letter grade awarded to *Q.i* by Step 1, and $K(\tilde{g}_i)$ denotes the derived grade-point of the derived fuzzy letter grade $\tilde{g}_i$ based on the index of optimism λ determined by the evaluator, where $\lambda \in [0,1]$. If $0 \leqslant \lambda < 0.5$, then the evaluator is a pessimistic evaluator. If $\lambda = 0.5$, then the evaluator is a normal evaluator. If $0.5 < \lambda \leqslant 1.0$, then the evaluator is an optimistic evaluator. Because $90 \leqslant \boldsymbol{A} \leqslant 100$, $70 \leqslant \boldsymbol{B} < 90$, $50 \leqslant \boldsymbol{C} < 70$, $30 \leqslant \boldsymbol{D} < 50$ and $0 \leqslant \boldsymbol{E} < 30$, the derived grade-point $K(g_i)$ shown in Eq. (11) is calculated as follows:

$$
\begin{aligned}
K(\tilde{g}_i) = \{ & \beta_{i1} \times [(1-\lambda) \times 90 + \lambda \times 100] \\
& + \beta_{i2} \times [(1-\lambda) \times 70 + \lambda \times 90] \\
& + \beta_{i3} \times [(1-\lambda) \times 50 + \lambda \times 70] \\
& + \beta_{i4} \times [(1-\lambda) \times 30 + \lambda \times 50] \\
& + \beta_{i5} \times [(1-\lambda) \times 0 + \lambda \times 30]\} \\
& /(\beta_{i1} + \beta_{i2} + \beta_{i3} + \beta_{i4} + \beta_{i5}),
\end{aligned} \quad (12)
$$

where λ is the index of optimism determined by the evaluator, $\lambda \in [0,1]$, $H(\overline{E},\overline{F}_i) = \beta_{i1}, H(\overline{V},\overline{F}_i) = \beta_{i2}, H(\overline{G},\overline{F}_i) = \beta_{i3}$, $H(\overline{S},\overline{F}_i) = \beta_{i4}$ and $H(\overline{U},\overline{F}_i) = \beta_{i5}, \beta_{ij} \in [0,1]$, $1 \leqslant i \leqslant n$, and $1 \leqslant j \leqslant 5$. Put the derived total mark in the appropriate box at the bottom of the interval-valued fuzzy grade sheet.

Example 4.1. Consider a student's answerscript to an examination of 100 marks. Assume that in total there are four questions to be answered:

TOTAL MARKS = 100,
Q.1 carries 30 marks,
Q.2 carries 20 marks,
Q.3 carries 30 marks,
Q.4 carries 20 marks.

Assume that an evaluator awards the student's answerscript using the interval-valued fuzzy grade sheet shown in Table 4, where the index of optimism λ determined by the evaluator is 0.60 (i.e., λ 0.60). Assume that "*A*", "*B*", "*C*", "*D*" and "*E*" are letter grades, where $90 \leqslant A \leqslant 100$, $70 \leqslant B < 90$, $50 \leqslant C < 70$, $30 \leqslant D < 50$ and $0 \leqslant E < 30$. Assume that the five standard fuzzy sets are $\widetilde{E}$ (excellent), $\widetilde{V}$ (very good), $\widetilde{G}$ (good), $\widetilde{S}$ (satisfactory) and $\widetilde{U}$ (unsatisfactory) represented by interval-valued membership values, shown as follows:

$$
\begin{aligned}
\widetilde{E} = {} & [0,0]/0\% + [0,0]/20\% + [0.8,0.8]/40\% + [0.9,0.9]/60\% \\
& + [1,1]/80\% + [1,1]/100\%,
\end{aligned}
$$

$$
\begin{aligned}
\widetilde{V} = {} & [0,0]/0\% + [0,0]/20\% + [0.8,0.8]/40\% + [0.9,0.9]/60\% \\
& + [0.9,0.9]/80\% + [0.8,0.8]/100\%,
\end{aligned}
$$

$$
\begin{aligned}
\widetilde{G} = {} & [0,0]/0\% + [0.1,0.1]/20\% + [0.8,0.8]/40\% + [0.9,0.9]/60\% \\
& + [0.4,0.4]/80\% + [0.2,0.2]/100\%,
\end{aligned}
$$

$$
\begin{aligned}
\widetilde{S} = {} & [0.4,0.4]/0\% + [0.4,0.4]/20\% + [0.9,0.9]/40\% \\
& + [0.6,0.6]/60\% + [0.2,0.2]/80\% + [0,0]/100\%,
\end{aligned}
$$

$$
\begin{aligned}
\widetilde{U} = {} & [1,1]/0\% + [1,1]/20\% + [0.4,0.4]/40\% + [0.2,0.2]/60\% \\
& + [0,0]/80\% + [0,0]/100\%.
\end{aligned}
$$

From Table 4, we can see that the interval-valued fuzzy marks of the questions *Q.1*, *Q.2*, *Q.3* and *Q.4* represented by interval-valued fuzzy sets are $\widetilde{F}_1, \widetilde{F}_2, \widetilde{F}_3$ and $\widetilde{F}_4$, respectively, where

$$
\begin{aligned}
\widetilde{F}_1 = {} & [0,0]/0\% + [0,0]/20\% + [0,0]/40\% + [0.5,0.6]/60\% \\
& + [1,1]/80\% + [0.7,0.8]/100\%,
\end{aligned}
$$

$$
\begin{aligned}
\widetilde{F}_2 = {} & [0,0]/0\% + [0,0]/20\% + [0,0]/40\% + [0.7,0.8]/60\% \\
& + [0.8,0.9]/80\% + [1,1]/100\%,
\end{aligned}
$$

$$
\begin{aligned}
\widetilde{F}_3 = {} & [0,0]/0\% + [0.5,0.6]/20\% + [1,1]/40\% + [0.7,0.8]/60\% \\
& + [0.4,0.5]/80\% + [0,0]/100\%,
\end{aligned}
$$

$$
\begin{aligned}
\widetilde{F}_4 = {} & [1,1]/0\% + [0.7,0.8]/20\% + [0.5,0.6]/40\% + [0,0]/60\% \\
& + [0,0]/80\% + [0,0]/100\%.
\end{aligned}
$$

[Step 1] The standard interval-valued fuzzy sets $\widetilde{E}, \widetilde{V}, \widetilde{G}, \widetilde{S}$ and $\widetilde{U}$ can be represented by the matrices $\overline{E}, \overline{V}, \overline{G}, \overline{S}$ and $\overline{U}$, respectively, where

$$
\begin{aligned}
\overline{E} &= \langle [0,0],[0,0],[0.8,0.8],[0.9,0.9],[1,1],[1,1] \rangle, \\
\overline{V} &= \langle [0,0],[0,0],[0.8,0.8],[0.9,0.9],[0.9,0.9],[0.8,0.8] \rangle, \\
\overline{G} &= \langle [0,0],[0.1,0.1],[0.8,0.8],[0.9,0.9],[0.4,0.4],[0.2,0.2] \rangle, \\
\overline{S} &= \langle [0.4,0.4],[0.4,0.4],[0.9,0.9],[0.6,0.6],[0.2,0.2],[0,0] \rangle, \\
\overline{U} &= \langle [1,1],[1,1],[0.4,0.4],[0.2,0.2],[0,0],[0,0] \rangle.
\end{aligned}
$$

The interval-valued fuzzy marks $\widetilde{F}_1, \widetilde{F}_2, \widetilde{F}_3$ and $\widetilde{F}_4$ also can be represented by matrices $\overline{F}_1$, $\overline{F}_2$, $\overline{F}_3$ and $\overline{F}_4$, respectively, where

$$
\begin{aligned}
\overline{F}_1 &= \langle [0,0],[0,0],[0,0],[0.5,0.6],[1,1],[0.7,0.8] \rangle, \\
\overline{F}_2 &= \langle [0,0],[0,0],[0,0],[0.7,0.8],[0.8,0.9],[1,1] \rangle, \\
\overline{F}_3 &= \langle [0,0],[0.5,0.6],[1,1],[0.7,0.8],[0.4,0.5],[0,0] \rangle, \\
\overline{F}_4 &= \langle [1,1],[0.7,0.8],[0.5,0.6],[0,0],[0,0],[0,0] \rangle.
\end{aligned}
$$

By applying Eq. (5), we can get

$$
\begin{aligned}
H(\overline{E},\overline{F}_1) = {} & \frac{1}{6}\left[\left(1 - \frac{|0-0|+|0-0|}{2}\right) + \left((1 - \frac{|0-0|+|0-0|}{2}\right)\right. \\
& + \left(1 - \frac{|0.8-0|+|0.8-0|}{2}\right) + \left(1 - \frac{|0.9-0.5|+|0.9-0.6|}{2}\right) \\
& \left. + \left(1 - \frac{|1-1|+|1-1|}{2}\right) + \left(1 - \frac{|1-0.7|+|1-0.8|}{2}\right)\right] \\
= {} & \frac{1}{6}(1+1+0.2+0.65+1+0.75) = 0.767,
\end{aligned}
$$

$$
\begin{aligned}
H(\overline{V},\overline{F}_1) = {} & \frac{1}{6}\left[\left(1 - \frac{|0-0|+|0-0|}{2}\right) + \left(1 - \frac{|0-0|+|0-0|}{2}\right)\right. \\
& + \left(1 - \frac{|0.8-0|+|0.8-0|}{2}\right) + \left(1 - \frac{|0.9-0.5|+|0.9-0.6|}{2}\right) \\
& \left. + \left(1 - \frac{|0.9-1|+|0.9-1|}{2}\right) + \left((1 - \frac{|0.8-0.7|+|0.8-0.8|}{2}\right)\right] \\
= {} & \frac{1}{6}(1+1+0.2+0.65+0.9+0.95) = 0.783,
\end{aligned}
$$

Table 4
Interval-valued fuzzy grade sheet of Example 4.1.

Question no.	Interval-valued fuzzy mark						Derived fuzzy letter grade
Q.1	[0,0]	[0,0]	[0,0]	[0.5,0.6]	[1,1]	[0.7,0.8]	
Q.2	[0,0]	[0,0]	[0,0]	[0.7,0.8]	[0.8,0.9]	[1,1]	
Q.3	[0,0]	[0.5,0.6]	[1,1]	[0.7,0.8]	[0.4,0.5]	[0,0]	
Q.4	[1,1]	[0.7,0.8]	[0.5,0.6]	[0,0]	[0,0]	[0,0]	
					Total mark =		

$$H(\overline{G},\overline{F}_1)=\frac{1}{6}\left[\left(1-\frac{|0-0|+|0-0|}{2}\right)+\left(1-\frac{|0.1-0|+|0.1-0|}{2}\right)+\left((1-\frac{|0.8-0|+|0.8-0|}{2}\right)+\left(1-\frac{|0.9-0.5|+|0.9-0.6|}{2}\right)+\left(1-\frac{|0.4-1|+|0.4-1|}{2}\right)+\left(1-\frac{|0.2-0.7|+|0.2-0.8|}{2}\right)\right]$$
$$=\frac{1}{6}(1+0.9+0.2+0.65+0.4+0.45)=0.600,$$

$$H(\overline{S},\overline{F}_1)=\frac{1}{6}\left[\left(1-\frac{|0.4-0|+|0.4-0|}{2}\right)+\left(1-\frac{|0.4-0|+|0.4-0|}{2}\right)+\left(1-\frac{|0.9-0|+|0.9-0|}{2}\right)+\left(1-\frac{|0.6-0.5|+|0.6-0.6|}{2}\right)+\left(1-\frac{|0.2-1|+|0.2-1|}{2}\right)+\left(1-\frac{|0-0.7|+|0-0.8|}{2}\right)\right]$$
$$=\frac{1}{6}(0.6+0.6+0.1+0.95+0.2+0.25)=0.450,$$

$$H(\overline{U},\overline{F}_1)=\frac{1}{6}\left[\left(1-\frac{|1-0|+|1-0|}{2}\right)+\left(1-\frac{|1-0|+|1-0|}{2}\right)+\left(1-\frac{|0.4-0|+|0.4-0|}{2}\right)+\left(1-\frac{|0.2-0.5|+|0.2-0.6|}{2}\right)+\left(1-\frac{|0-1|+|0-1|}{2}\right)+\left(1-\frac{|0-0.7|+|0-0.8|}{2}\right)\right]$$
$$=\frac{1}{6}(0+0+0.6+0.65+0+0.25)=0.250.$$

Because the standard fuzzy sets are $\widetilde{E},\widetilde{V},\widetilde{G},\widetilde{S}$ and $\widetilde{U}$ corresponding to the letter grades "***A***", "***B***", "***C***", "***D***" and "***E***", respectively, the derived fuzzy letter grade $\widetilde{g}_1$ of question *Q.1* is represented by a fuzzy set, shown as follows:

$$\widetilde{g}_1=0.767/\boldsymbol{A}+0.783/\boldsymbol{B}+0.600/\boldsymbol{C}+0.450/\boldsymbol{D}+0.250/\boldsymbol{E}.$$

Because the value of the index of optimism λ given by the evaluator is 0.60 (i.e., λ = 0.60), by applying Eq. (12), we can get

$$\begin{aligned}K(\widetilde{g}_1)=\{&0.767\times[(1-0.60)\times 90+0.60\times 100]\\&+0.783\times[(1-0.60)\times 70+0.60\times 90]\\&+0.600\times[(1-0.60)\times 50+0.60\times 70]\\&+0.450\times[(1-0.60)\times 30+0.60\times 50]\\&+0.250\times[(1-0.60)\times 0+0.60\times 30]\}\\&/(0.767+0.783+0.600+0.450+0.250)\\=&(0.767\times 96+0.783\times 82+0.600\times 62\\&+0.450\times 42+0.250\times 18)/2.85\\=&(73.632+64.206+37.2+18.9+4.5)/2.85=69.627.\end{aligned}$$

In the same way, we can get $K(\widetilde{g}_2)=71.343, K(\widetilde{g}_3)=60.466$ and $K(\widetilde{g}_4)=44.610$.

[Step 2] Because the questions *Q.1*, *Q.2*, *Q.3* and *Q.4* carry 30 marks, 20 marks, 30 marks and 20 marks, respectively (i.e., T(*Q*.1) = 30, T(*Q*.2) = 20, T(*Q*.3) = 30 and T(*Q*.4) = 20), based on Eq. (11), we can get

Total mark

$$\begin{aligned}&=\frac{T(Q.1)\times K(\widetilde{g}_1)+T(Q.2)\times K(\widetilde{g}_2)+T(Q.3)\times K(\widetilde{g}_3)+T(Q.4)\times K(\widetilde{g}_4)}{100}\\&=\frac{30\times 69.627+20\times 71.343+30\times 60.466+20\times 44.610}{100}\\&=\frac{2088.81+1426.86+1813.98+892.2}{100}\\&=62.2185\\&=62\ \text{(assuming that no half mark is given in the total mark).}\end{aligned}$$

In the following, we present a generalized interval-valued fuzzy evaluation method (*GIVFEM*) for students' answerscripts evaluation based on interval-valued fuzzy sets, where a generalized interval-valued fuzzy grade sheet shown in Table 5 is used to evaluate the students' answerscripts. In the generalized interval-valued fuzzy grade sheet shown in Table 5, each question *Q.i* consists of four sub-questions, i.e., *Q.i1*, *Q.i2*, *Q.i3* and Q.*i4*, and $\widetilde{F}_{ij}$ denotes the interval-valued fuzzy mark of sub-question *Q.ij*, where $\widetilde{F}_{ij}$ is represented by an interval-valued fuzzy set, $1 \leqslant i \leqslant n$, and $1 \leqslant j \leqslant 4$. For all j = 1, 2, 3, 4 and for all i, $\widetilde{g}_{ij}$ denotes the derived fuzzy letter grade by the proposed interval-valued fuzzy evaluation method *IVFEM* of the awarded interval-valued fuzzy mark $\widetilde{F}_{ij}$ with respect to the sub-question *Q.ij*, and m_i is the derived mark awarded to the question *Q.i*,

Table 5
A generalized interval-valued fuzzy grade sheet.

Question No.	Sub-questions	Interval-valued fuzzy mark	Derived fuzzy letter grade	Mark
Q.1	*Q.11*	$\widetilde{F}_{11}$	$\widetilde{g}_{11}$	m_1
	Q.12	$\widetilde{F}_{12}$	$\widetilde{g}_{12}$	
	Q.13	$\widetilde{F}_{13}$	$\widetilde{g}_{13}$	
	Q.14	$\widetilde{F}_{14}$	$\widetilde{g}_{14}$	
Q.2	*Q.21*	$\widetilde{F}_{21}$	$\widetilde{g}_{21}$	m_2
	Q.22	$\widetilde{F}_{22}$	$\widetilde{g}_{22}$	
	Q.23	$\widetilde{F}_{23}$	$\widetilde{g}_{23}$	
	Q.24	$\widetilde{F}_{24}$	$\widetilde{g}_{24}$	
⋮	⋮	⋮	⋮	⋮
Q.n	*Q.n1*	$\widetilde{F}_{n1}$	$\widetilde{g}_{n1}$	m_n
	Q.n2	$\widetilde{F}_{n2}$	$\widetilde{g}_{n2}$	
	Q.n3	$\widetilde{F}_{n3}$	$\widetilde{g}_{n3}$	
	Q.n4	$\widetilde{F}_{n4}$	$\widetilde{g}_{n4}$	
			Total mark =	

$$m_i=\frac{1}{400}\times T(Q.i)\times\sum_{j=1}^{4}K(\widetilde{g}_{ij}), \tag{13}$$

and

$$\text{Total Mark}=\sum_{i=1}^{n}m_i,$$

where *T(Q.i)* denotes the mark allotted to *Q.i* in the test paper, g_{ij} denotes the derived letter grade awarded to *Q.ij*, and $K(\widetilde{g}_{ij})$ denotes the derived grade-point of the letter grade $\widetilde{g}_{ij}$ based on the index of optimism λ determined by the evaluator, where $\lambda\in[0,1]$. If $0 \leqslant \lambda < 0.5$, then the evaluator is a pessimistic evaluator. If λ = 0.5, then the evaluator is a normal evaluator. If $0.5 < \lambda \leqslant 1.0$, then the evaluator is an optimistic evaluator. Assume that the derived letter grade with respect to the sub-question Q.*ij* is $\widetilde{g}_{ij}$, where $\widetilde{g}_{ij}$ is represented by a fuzzy set, shown as follows:

$$\widetilde{g}_{ij}=\beta_{(ij)1}/\boldsymbol{A}+\beta_{(ij)2}/\boldsymbol{B}+\beta_{(ij)3}/\boldsymbol{C}+\beta_{(ij)4}/\boldsymbol{D}+\beta_{(ij)5}/\boldsymbol{E}, \tag{14}$$

where ***A***, ***B***, ***C***, ***D*** and ***E*** are letter grades, $H(\overline{E},\overline{F}_{ij})=\beta_{(ij)1}, H(\overline{V},\overline{F}_{ij})=\beta_{(ij)2}, H(\overline{G},\overline{F}_{ij})=\beta_{(ij)3}, H(\overline{S},\overline{F}_{ij})=\beta_{(ij)4}$ and $H(\overline{U},\overline{F}_{ij})=\beta_{(ij)5}, \beta_{(ij)k}\in[0,1], \overline{F}_{ij}$ is the matrix representation of the interval-valued fuzzy mark $\widetilde{F}_{ij}$, $1 \leqslant i \leqslant n$, and $1 \leqslant j \leqslant 4$, and $1 \leqslant k \leqslant 5$. Then, the derived grade-point $K(\widetilde{g}_{ij})$ shown in Eq. (13) is calculated as follows:

$$\begin{aligned}K(\widetilde{g}_{ij})=\{&\beta_{(ij)1}\times[(1-\lambda)\times 90+\lambda\times 100]\\&+\beta_{(ij)2}\times[(1-\lambda)\times 70+\lambda\times 90]\\&+\beta_{(ij)3}\times[(1-\lambda)\times 50+\lambda\times 70]\\&+\beta_{(ij)4}\times[(1-\lambda)\times 30+\lambda\times 50]\\&+\beta_{(ij)5}\times[(1-\lambda)\times 0+\lambda\times 30]\}\\&/(\beta_{(ij)1}+\beta_{(ij)2}+\beta_{(ij)3}+\beta_{(ij)4}+\beta_{(ij)5}),\end{aligned} \tag{15}$$

where λ is the index of optimism determined by the evaluator and $\lambda\in[0,1]$. Put the derived total mark in the appropriate box at the bottom of the generalized interval-valued fuzzy grade sheet.

September 1, 2007

Question No.	Satisfaction levels						Grade
Q.1	0	0	0	0.6	0.9	0.8	
Q.2	0	0	0.6	0.9	0.8	0	
Q.3	0	0	0	0.6	0.8	0.9	
Q.4	0	0.6	0.9	0.8	0.2	0	
					Total mark =		

September 2, 2007

Question No.	Satisfaction levels						Grade
Q.1	0	0	0	0.8	0.9	1	
Q.2	0	0	0.7	0.8	0.9	0	
Q.3	0	0	0	0.7	0.9	0.8	
Q.4	0	0.5	0.8	0.7	0	0	
					Total mark =		

September 3, 2007

Question No.	Satisfaction levels						Grade
Q.1	0	0	0	0.6	0.9	0.7	
Q.2	0	0	0.6	0.8	0.7	0	
Q.3	0	0	0	0.5	0.7	0.9	
Q.4	0	0.5	0.8	0.6	0	0	
					Total mark =		

September 4, 2007

Question No.	Satisfaction levels						Grade
Q.1	0	0	0	0.6	0.8	0.7	
Q.2	0	0	0.5	0.9	0.7	0	
Q.3	0	0	0	0.7	0.9	0.8	
Q.4	0	0.6	0.9	0.7	0	0	
					Total mark =		

Fig. 1. Evaluating the student's answerscript at different days using Biswas's method (1995).

September 1, 2007

Question No.	Interval-valued fuzzy marks						Grade
Q.1	[0, 0]	[0, 0]	[0, 0]	[0.6, 0.7]	[0.8, 0.9]	[0.8, 0.9]	
Q.2	[0, 0]	[0, 0]	[0.6, 0.7]	[0.8, 0.9]	[0.8, 0.9]	[0, 0]	
Q.3	[0, 0]	[0, 0]	[0, 0]	[0.6, 0.7]	[0.8, 0.9]	[0.8, 0.9]	
Q.4	[0, 0]	[0.5, 0.6]	[0.8, 0.9]	[0.7, 0.8]	[0.1, 0.2]	[0, 0]	
					Total mark =		

September 2, 2007

Question No.	Interval-valued fuzzy marks						Grade
Q.1	[0, 0]	[0, 0]	[0, 0]	[0.6, 0.7]	[0.8, 0.9]	[0.8, 0.9]	
Q.2	[0, 0]	[0, 0]	[0.6, 0.7]	[0.8, 0.9]	[0.8, 0.9]	[0, 0]	
Q.3	[0, 0]	[0, 0]	[0, 0]	[0.6, 0.7]	[0.8, 0.9]	[0.8, 0.9]	
Q.4	[0, 0]	[0.5, 0.6]	[0.8, 0.9]	[0.7, 0.8]	[0, 0]	[0, 0]	
					Total mark =		

September 3, 2007

Question No.	Interval-valued fuzzy marks						Grade
Q.1	[0, 0]	[0, 0]	[0, 0]	[0.6, 0.7]	[0.8, 0.9]	[0.7, 0.8]	
Q.2	[0, 0]	[0, 0]	[0.6, 0.7]	[0.8, 0.9]	[0.7, 0.8]	[0, 0]	
Q.3	[0, 0]	[0, 0]	[0, 0]	[0.5, 0.6]	[0.7, 0.8]	[0.8, 0.9]	
Q.4	[0, 0]	[0.5, 0.6]	[0.8, 0.9]	[0.6, 0.7]	[0, 0]	[0, 0]	
					Total mark =		

September 4, 2007

Question No.	Interval-valued fuzzy marks						Grade
Q.1	[0, 0]	[0, 0]	[0, 0]	[0.6, 0.7]	[0.8, 0.9]	[0.7, 0.8]	
Q.2	[0, 0]	[0, 0]	[0.5, 0.6]	[0.8, 0.9]	[0.7, 0.8]	[0, 0]	
Q.3	[0, 0]	[0, 0]	[0, 0]	[0.7, 0.8]	[0.8, 0.9]	[0.8, 0.9]	
Q.4	[0, 0]	[0.6, 0.7]	[0.8, 0.9]	[0.7, 0.8]	[0, 0]	[0, 0]	
					Total mark =		

Fig. 2. Evaluating the student's answerscript at different days using the proposed method.

Table 6
A comparison of the evaluating results for different methods.

Days	Methods	
	Biswas's method	The proposed method
September 1, 2007	69	66
September 2, 2007	72	66
September 3, 2007	55	66
September 4, 2007	55	66

5. Experimental results

We have made an experiment to compare the evaluating results of the proposed method with Biswas's method (1995) for different days. In our experiment, there are four questions to be answered in a student's answerscript, where

TOTAL MARKS = 100,
Q.1 carries 20 marks,
Q.2 carries 25 marks,
Q.3 carries 25 marks,
Q.4 carries 30 marks.

Assume that the optimism index λ of the evaluator is 0.60 (i.e., $\lambda = 0.60$). That is, the evaluator is a slightly optimistic evaluator. The evaluator uses Biswas's method (1995) and the proposed method to evaluate the student's answerscript on different days, respectively. The results are shown in Figs. 1 and 2, respectively. A comparison of the evaluating results of the student's answerscript is shown in Table 6, where the degree of optimism of the total mark for each day evaluated by the proposed method is equal to 0.60. From Table 6, we can see that the proposed method is more stable to evaluate the student's answerscript than Biswas's method (1995). It can evaluate students' answerscripts in a more flexible and more intelligent manner.

6. Conclusions

In this paper, we have presented two new methods for evaluating students' answerscripts based on interval-valued fuzzy grade sheets. The marks awarded to the answers in the students' answerscripts are represented by interval-valued fuzzy sets. The degree of similarity between an interval-valued fuzzy mark and a standard interval-valued fuzzy set is calculated by a similarity function. An index of optimism λ determined by the evaluator is used to indicate the degree of optimism of the evaluator, where $\lambda \in [0, 1]$. From the experimental results shown in Table 6, we can see that the proposed methods can evaluate students' answerscripts more stable than Biswas's method. They can evaluate students' answerscripts in a more flexible and more intelligent manner.

Acknowledgements

The authors would like to thank Professor Jason Chiyu Chan, Department of Education, National Chengchi University, Taipei, Taiwan, Republic of China, for providing very helpful comments and suggestions. This work was supported in part by the National Science Council, Republic of China, under Grant NSC 95-2213-E-011-116-MY2.

References

Bai, S. M., & Chen, S. M. (2008a). Evaluating students' learning achievement using fuzzy membership functions and fuzzy rules. *Expert Systems with Applications, 34*(1), 399–410.

Bai, S. M., & Chen, S. M. (2008b). Automatically constructing concept maps based on fuzzy rules for adaptive learning systems. *Expert Systems with Applications, 35*(1), 41–49.

Bai, S. M., & Chen, S. M. (2008c). Automatically constructing grade membership functions of fuzzy rules for students' evaluation. *Expert Systems with Applications, 35*(3), 1408–1414.

Biswas, R. (1995). An application of fuzzy sets in students' evaluation. *Fuzzy Sets and Systems, 74*(2), 187–194.

Chang, D. F., & Sun, C. M. (1993). Fuzzy assessment of learning performance of junior high school students. In *Proceedings of the 1993 first national symposium on fuzzy theory and applications, Hsinchu, Taiwan, Republic of China* (pp. 10–15).

Chen, S. M. (1988). A new approach for handling fuzzy decisionmaking problems. *IEEE Transactions on Systems, Man, and Cybernetics, 18*(6), 1012–1016.

Chen, S. M. (1994). A new method for handling multicriteria fuzzy decision-making problems. *Cybernetics and Systems, 25*(3), 409–420.

Cheng, C. H., & Yang, K. L. (1998). Using fuzzy sets in education grading system. *Journal of Chinese Fuzzy Systems Association, 4*(2), 81–89.

Chen, S. M., & Lee, C. H. (1999). New methods for students' evaluating using fuzzy sets. *Fuzzy Sets and Systems, 104*(2), 209–218.

Chen, S. M., & Wang, J. Y. (1995). Document retrieval using knowledge-based fuzzy information retrieval techniques. *IEEE Transactions on Systems, Man, and Cybernetics, 25*(5), 793–803.

Chiang, T. T., & Lin, C. M. (1994). Application of fuzzy theory to teaching assessment. In *Proceedings of the 1994 second national conference on fuzzy theory and applications, Taipei, Taiwan, Republic of China* (pp. 92–97).

Echauz, J. R., & Vachtsevanos, G. J. (1995). Fuzzy grading system. *IEEE Transactions on Education, 38*(2), 158–165.

Frair, L. (1995). Student peer evaluations using the analytic hierarchy process method. In *Proceedings of 1995 frontiers in education conference* (Vol. 2, pp. 4c3.1–4c3.5).

Gorzalczany, M. B. (1987). A method of inference in approximate reasoning based on interval-valued fuzzy sets. *Fuzzy Sets and Systems, 21*(1), 1–17.

Gorzalczany, M. B. (1989). An interval-valued fuzzy inference method – Some basic properties. *Fuzzy Sets and Systems, 31*(2), 243–251.

Hwang, G. J., Lin, B. M., & Lin, T. L. (2006). An effective approach for test-sheet composition with large-scale item banks. *Computers & Education, 46*(2), 122–139.

Kaburlasos, V. G., Marinagi, C. C., & Tsoukalas, V. T., (2004). PARES: A software tool for computer-based testing and evaluation used in the Greek higher education system. In *Proceedings of the 2004 IEEE international conference on advanced learning technologies* (pp. 771–773).

Law, C. K. (1996). Using fuzzy numbers in education grading system. *Fuzzy Sets and Systems, 83*(3), 311–323.

Ma, J., & Zhou, D. (2000). Fuzzy set approach to the assessment of student-centered learning. *IEEE Transactions on Education, 43*(2), 237–241.

McMartin, F., Mckenna, A., & Youssefi, K. (2000). Scenario assignments as assessment tools for undergraduate engineering education. *IEEE Transactions on Education, 43*(2), 111–119.

Pears, A., Daniels, M., Berglund, A., & Erickson, C. (2001). Student evaluation in an international collaborative project course. In *Proceedings of the 2001 symposium on applications and the internet workshops* (pp. 74–79).

Wang, H. Y., & Chen, S. M. (2007a). Artificial intelligence approach to evaluate students' answerscripts based on the similarity measure between vague sets. *Educational Technology & Society, 10*(4), 224–241.

Wang, H. Y., & Chen, S. M. (2007b). A new approach for evaluating students' answerscripts based on interval-valued fuzzy sets. In *Proceedings of the 20th international conference on industrial, engineering & other applications of applied intelligent systems, Kyoto, Japan* (pp. 74–83).

Weon, S., & Kim, J. (2001). Learning achievement evaluation strategy using fuzzy membership function. In *Proceedings of the 31st ASEE/IEEE frontier in education conference, Reno, NV* (pp. T3A-19–T3A-24).

Wu, M. H. (2003). *A research on applying fuzzy set theory and item response theory to evaluate learning performance*. Master Thesis, Department of Information Management, Chaoyang University of Technology, Wufeng, Taichung County, Taiwan, Republic of China.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control, 8*, 338–353.

Zwick, R., Carlstein, E., & Budescu, D. V. (1987). Measures of similarity among fuzzy concepts: A comparative analysis. *International Journal of Approximate Reasoning, 1*(2), 221–242.