Models for Ranking Students: Selecting Applicants for a Master of Science Studies

Pavle Milošević[1], Ivan Nešić[1], Ana Poledica[1],
Dragan G. Radojević[2], and Bratislav Petrović[1]

[1] Faculty of Organizational Sciences, University of Belgrade,
Jove Ilića 154, Belgrade, 11000, Serbia
pasha.47@gmail.com
ivan.nesic@ewcom.ch
{ana.poledica,bratislav.petrovic}@fon.bg.ac.rs
[2] Mihajlo Pupin Institute, Volgina 154,Belgrade, 11000, Serbia
dragan.radojevic@automatika.imp.bg.ac.rs

Abstract. In this paper, we present a problem of candidates ranking for Master of Science studies at Faculty of Organizational Sciences, Belgrade (FOS). Current ranking model is based on weighted sum of two factors: average grade gained at undergraduate studies and number of points scored at the entrance exam for master studies. This universal model, although widespread and frequently used, is not entirely appropriate for a number of different Master of Science programs at FOS. Major problems of current model are that the model does not emphasize essential knowledge in accordance to wanted program and that the weighted sum is unable to model connection between variables. This paper presents a series of models, each more complex than the previous one, which are aggregating more relevant factor for ranking. Logical Aggregation (LA) is used as a method for the aggregation of certain variables in the last few models. LA is based on Interpolative Boolean algebra (IBA), a consistent multi-valued realization of Boolean algebra. This paper may be particularly interesting to all of those dealing with any form of students ranking, especially to university departments involved in enrolling students and selection for scholarships.

Keywords: weighted sum, fuzzy logic, Interpolative Boolean algebra, Logical aggregation, ranking students.

1 Introduction

Since 2007 Faculty of Organizational Sciences (FOS) has been enrolling students to Master of Science studies. The criterions of enrollment are entrance exam output and average grade at undergraduate level. Considering the differences between the courses existing at Master of Science studies at FOS, this universal ranking model may not be completely appropriate. The student's average grade is taken as indicator of previously acquired knowledge. That grade, as it is, does not show candidate's specific knowledge. For example, knowledge and grades gained at management domain should be additionally scored for enrolling in management departments. In this way

V.E. Balas et al. (Eds.): Soft Computing Applications, AISC 195, pp. 93–103.
springerlink.com

the advantage at the enrollment would be given to the candidates who showed special interests in certain domain.

The propositions of enrollment and description of the module that will be used as an example are given at Section 2. The current ranking model based on weighted sum and its deficiencies are explained at Section 3. The details are underlined by description of necessary criterions considering the specific module. In Section 4 there is the proposition considering the usage of changed weighted sum, which includes several factors. One of them, the one which underlines required specific knowledge, is aggregated by fuzzy logic. There is also a theoretical introduction to fuzzy logic. The subject of Section 5 is Interpolative Boolean algebra (IBA) and Logical aggregation (LA). IBA threats negation differently from fuzzy logic. There is also a suggestion of pseudo-logical function as criterion function. One of the criterions is aggregated using logical aggregation, which is based on IBA. In Section 6 there is an example of classical student ranking and the ranking using models suggested at previous sections and these results are analyzed and compared. In final section we give our conclusions and guidelines for the future work.

2 Proposition for Enrollment to a Master of Science Program at FOS

Eight different study programs are offered for enrollment at FOS, considering the proposition for student's enrollment to Master of Science studies 2011/2012. Every single program is divided into modules or study groups that candidate enrolls. There are thirty six modules, which are the forms of additional guidance. Study groups are highly specialized in very different areas. Some of fields that are modules focused on are Software engineering, E-business, Ecological Management, Business Intelligence and Decision Making, etc. Students are ranked in one of six rang lists depending on chosen Master of Science program. The criterions of enrollment for all study departments are the entrance exam output and the average grade at undergraduate level. The number of points at the entrance exam is more important than the average grade in 3/2 ratio.

Ranking models presented and analyzed in this paper will be focused on module Operational Researches. This module is analyzed because certain similarities with module Computer Statistics, the other module on the same program, and universal qualification exam, that is used for two more study departments. Module Operational Researches is a part of study department Operational Researches and Computer Statistics. Thirty five students can be enrolled to this module, and twelve of them have full scholarship. The qualification exam itself consists of thirty test questions with five offered answers each, but only one correct. The questions consider operational researches, statistics, software design, database and informational systems. This kind of qualification exam (the same test and the same literature) is used for Informational Systems and Technologies study department, as well as Software Engineering and Computer Sciences study departments. Candidates who enroll this module are ranked on a ranking list for the particular study department.

3 Current Ranking Model Based on Weighted Sum

The weighted sum model is probably the most commonly used approach for modeling this kind of problems. It ranks candidates based on weighted sum of normalized attributes. Weights in the model can be static or moving. The normalizing of the attributes ensures their comparability, as otherwise high numbered attributes would make disproportionate contributions to the overall score [9]. Weighted sum general form for one dimensional problem is following:

$$\sum_{i=1}^{n} w_i * u_i = p \tag{1}$$

According to previous, the conclusion is that weighted sum of criterion is used as a student ranking criterion and have this form:

$$w_1 * u + w_2 * k = p \tag{2}$$

It contains the following variables:

- u – average grade gained at undergraduate level in the interval [0, 100];
- k – number of points on the entrance exam for Master of Science studies in the interval [0, 100];
- p – candidate's total points in the interval [0, 100].

Values w_1 and w_2 are weights in this model, whose values are predefined at the enrolment proposition, so the criterion function has the following form:

$$0.4 * u + 0.6 * k = p \tag{3}$$

This function is modeled by weighted sum considering only the values of variables. This way, it is not possible to model their conditionality or connection between the variables using the logical operators.

The qualification exam to module Operational Researches consists of questions considering different domains: informational systems, computer sciences, operational researches, and statistics. All the questions are scored the same way. This is illogical – it is wanted to have knowledge considering all of these domains, but knowledge considering operational researches should be the most important.

Average grade shows the general picture about candidate's success at undergraduate level. However, the grades in subjects which are focused to operational researches, statistics, mathematics, and system theory should be more important if enrolling this module. Analogous to that, subjects focused on informational systems, programming, database and data structures etc. should be underlined if enrolling the study program Informational Systems and Technologies.

The flaw is also restriction of weighted sum itself – it cannot be used to model logical expressions. This restriction has a certain number of models for multi-criteria decision making, like the weighted product model, AHP, etc.

3.1 New Model Based on Weighted Sum

The first two comments, that the data aggregated using arithmetic mean and that total score at qualification exam does not underline characteristic data important for the program of Master of Science studies, can be solved by modeling new weighted sum:

$$w_1 * u + w_2 * u_2 + w_3 * k + w_4 * k_2 = p \tag{4}$$

Two new components figure in this expression, and they are:

- u_2 – average grade at undergraduate level at subjects which are focused on operational researches, statistics, mathematics, and system theory in the interval [0, 100];
- k_2 – points gained at qualification exam to Master of Science studies, but specifically in questions considering operational researches and statistics, in the interval [0, 100].

Values w_1, w_2, w_3 and w_4 are weights in this model. Values of this factors are 0.3, 0.1, 0.5 and 0,1, respectively, so the criterion function has the following form:

$$0.3 * u + 0.1 * u_2 + 0.5 * k + 0.1 * k_2 = p \tag{5}$$

The significance of specific knowledge, important for the enrolling Master of Science studies, is underlined by new variables. Other requirements can be emphasized by introducing new variables and setting the corresponding weights. However, the criterion function defined this way does not solve the problem, because the subject of interests is qualification exam achievement and appropriate subjects accomplishment, and not these two variables separately.

4 Ranking Using Fuzzy Logic

The reason for introducing logic into consideration is the need of modeling conditions like following: “Candidate’s ranking on the list should depend on number of points on the entrance exam for Master of Science studies obtained on matters related to operational researches and statistics and the average grade at undergraduate level at subjects that are focused on operational researches, statistics, mathematics, and systems theory“. This condition implies that predefined variables u_2 and k_2 should be connected by conjunction. Weighted sum considers variables separately and can’t model interaction between them, so it is not sufficient to model this problem. It is necessary to introduce logic and logical operators, which can provide more options for aggregation.

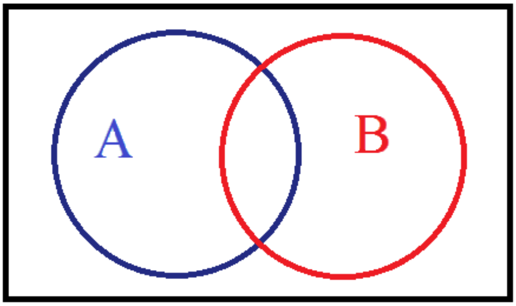

Fig. 1. What can be modeled using weighted sum, representation using sets

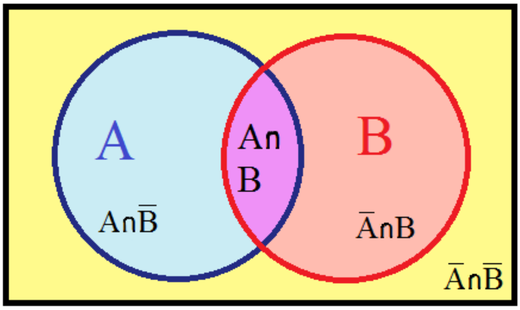

Fig. 2. What can be modeled using fuzzy logic, representation using sets

Classical logic deals with variables that are either true or false, that are either 0 or 1. Values in described problem are decimal, so it is not appropriate. That is the reason for interduction and usage of fuzzy logic.

4.1 Fuzzy Logic

Fuzzy logic is generalization of classical logic – it can process all values in interval from 0 to 1. Fuzzy logic is not fuzzy. Basically, fuzzy logic is a precise logic of imprecision and approximate reasoning [11]. Fuzzy logic is developed on fuzzy sets theory. It is particularly suitable for working with linguistic variables like *hot*, *very good*, *around twenty*, *not that busy*. Compared with conventional approaches, fuzzy control utilizes more information from domain experts and relies less on mathematical modeling about a physical system.

Functions that qualify as fuzzy intersections and fuzzy unions are usually referred as t-norms and t-conorms (or s-norms), respectively. The standard fuzzy intersection is min operator, and it produces the smallest membership value of all the t-norms [8]. On top of that, algebraic product is used often as fuzzy intersection. The standard fuzzy union is max operator, and it produces the smallest membership value of all the t-conorms. Probabilistic sum is another operator often used as s-norm.

Conventional fuzzy logic is based on principle of truth functionality - the truth valued of a complex formula is uniquely determined by the truth values of its subformulas. In the general, it doesn't follow the law of excluded middle, one of the Boolean laws. Some authors say that fuzzy set theory simply does not happen to have an axiom of the excluded middle – it does not need, constrained by, such an axiom [8]. Others see this as inconsistency and problem.

4.2 New Model Based on Fuzzy Logic

The modified weighted sum is used to rank students. The third element of weighted sum is changed, and it is defined as: "Conjunction of average grade at undergraduate level at subjects that are focused on operational researches, statistics, mathematics, and systems theory and number of points on the entrance exam for Master of Science studies obtained on matters related to operational researches and statistics". In accordance to that, weighted sum is:

$$w_1 * u + w_2 * k + w_3 * 100 * (u_2 \wedge k_2) = p \tag{6}$$

Variable u_2 are normalized in the interval [0, 1]. Variable k_2 is linguistic variable, with suggested values: *very low, low, medium, high, and very high*. These linguistic terms first can be converted to fuzzy numbers using a conversion scale. Then a fuzzy scoring method is used to convert each fuzzy number to a corresponding crisp value. According to Chen and Hwang's five-scale fuzzy linguistic scaling [1], listed variables are converted to crisp values - 0.091, 0.283, 0.5, 0.717, 0.909 respectively. Weights w_1, w_2 and w_3 have following values - 0.3, 0.5 and 0,2. Product is used as t-norm. The third element in weighted sum is multiplied by 100, so all elements would have the same order of magnitude. In accordance to that, weighted sum has the following form:

$$0.3 * u + 0.5 * k + 0.2 * 100 * u_2 * k_2 = p \tag{7}$$

5 Ranking Using Pseudo-logical Aggregation

5.1 Interpolative Boolean Algebra and Logical Aggregation

Interpolative Boolean algebra is a consistent multi-valued realization of Boolean algebra in the sense that it preserves all the laws which Boolean algebra relies on. It has two levels – symbolic and valued. On symbolic level all laws of Boolean algebra are valued indifferent. IBA element represents the analyzed object. IBA element on symbolic level is treated independently of its realization. All elements consist of one or more atomic element. Expressions are calculated based on the principle of structural functionality: Structure of any IBA element can be directly calculated on the basis of structures of its components [4, 5]. This principle treats negation differently and that allows preservation of the law of excluded middle.

IBA is technically based on generalized Boolean polynomials (GBP). GBP uniquely corresponds to any element of Boolean algebra and any Boolean function can be transformed into corresponding GBP. Operators allowed in GBP are plus, minus and generalized product. Generalized product (GP) is any function ⊗: [0,1]⊗[0,1]→[0,1] that satisfied all four conditions of t-norms: Commutativity, Associativity, Monotonicity, Boundary condition and Non-negativity condition [6]:

$$\sum_{K\in P(\Omega/S)} (-1)^{|K|} \otimes A_i(x) \geq 0, S \in P(\Omega), A_i(x) \in [0,1], A_i \in \Omega \tag{8}$$

On a valued level the IBA is value realized – elements take values from an interval [0, 1] and suitable operator for GP is chosen.

Consistent and transparent procedure based on IBA for aggregating factors is called Logical aggregation. Logical aggregation is the fusion of the primary attributes of quality in a crisp value. The task of LA is the fusion of primary attributes' values into one resulting globally representative value using logical tools [3]. It has two steps:

- Normalization of attributes' values

$$||\bullet||: \Omega \to [0,1] \tag{9}$$

- Aggregation of normalized values of features into one resulting value by logical or pseudo-logical function as a LA operator [7]

$$Aggr[0,1]^n \to [0,1] \tag{10}$$

Pseudo-logical function, called pseudo GBP, is a linear convex combination of generalized Boolean polynomials. In this paper it is the proposed approach for aggregation.

A logical aggregation depends on the measure of an aggregation operator as well as of generalized product. Aggregation measure is a structural function of pseudo-logical function.

5.2 New Model Based on Pseudo Logical Aggregation

Two different functions are proposed as pseudo-logical function for ranking. Both are similar to the weighted sum presented in Section 4.2: first two elements and weights in sum are the same (0.3, 0.5 and 0,2 respectively). The third element in sum is the difference.

Case 1: In this model, candidates who have successfully mastered the material on operational research have the advantage. If they haven't, the important thing is that they learned operational researches and statistics for entrance exam, as well as they have good prior knowledge in mathematics.

The third element of weighted sum is defined this way: "If candidate's average grade at undergraduate level at subjects that are focused on operational researches and statistics is well, we are interested only in it. If it is not well, we are interested in the average grade at undergraduate level at subjects that are focused on mathematics and systems theory and number of points at the entrance exam for Master of Science studies obtained on matters related to operational researches and statistics."

Pseudo-logical function for ranking with this condition has the following form:

$$w_1 * u + w_2 * k + w_3 * 100 * (a \vee (\neg a \wedge d \wedge x)) = p \tag{11}$$

The third element is transformed on symbolic level:

$$a \vee (\neg a \wedge d \wedge x) = a + d \otimes x - a \otimes d \otimes x \tag{12}$$

After that, expression can be calculated. The operator of generalized product ⊗ is product (⊗:=*):

$$w_1 * u + w_2 * k + w_3 * 100 * (a + d * x - a * d * x) = p \tag{13}$$

Case 2: In this model, candidates who have successfully mastered the material on operational research and statistics at undergraduate studies have the advantage. If they haven't mastered operational researches, the important thing is that they learned operational researches and statistics for entrance exam, as well as they have good prior knowledge in mathematics. This model is interesting because results considering operational researches and statistics are separated and interpreted it in different ways.

The third variable of weighted sum is defined like this: "If candidate's average grade at undergraduate level at subjects that are focused on operational researches is well, we are interested in the average grade at undergraduate level at subjects that are focused on statistics, too. If it is not well, we are interested in the average grade at undergraduate level at subjects that are focused on mathematics and systems theory and number of points at the entrance exam for Master of Science studies obtained on matters related to operational researches and statistics."

Pseudo-logical function for ranking with this condition has the following form:

$$w_1 * u + w_2 * k + w_3 * 100 * ((b \wedge c) \vee (\neg b \wedge d \wedge x)) = p \tag{14}$$

The third element is transformed on symbolic level:

$$(b \wedge c) \vee (\neg b \wedge d \wedge x) = b \otimes c + d \otimes x - b \otimes d \otimes x \tag{15}$$

After that, expression can be calculated. The operator of generalized products ⊗ is product (⊗:=*):

$$w_1 * u + w_2 * k + w_3 * 100 * (b * c + d * x - b * d * x) = p \tag{16}$$

Variables in presented expressions are:

- u – average grade gained at undergraduate level in the interval [0, 100];
- k – number of points at the entrance exam for Master of Science studies in the interval [0, 100];

- *a* – average grade gained at undergraduate level at subjects that are focused on operational researches and statistics, normalized in the interval [0, 1];
- *b* – average grade gained at undergraduate level at subjects that are focused on operational researches, normalized in the interval [0, 1];
- *c* – average grade gained at undergraduate level at subjects that are focused on statistics, normalized in the interval [0, 1];
- *d* – average grade gained at undergraduate level at subjects that are focused on mathematics and systems theory, normalized in the interval [0, 1];
- *x* – number of points at the entrance exam for Master of Science studies obtained on matters related to operational researches and statistics, normalized in the interval [0, 1];
- *p* – candidate's total points in the interval [0, 100].

The third element in weighted sum is multiplied by 100, so all elements would have the same order of magnitude.

6 Analysis and Comparison of Proposed Models

In this section suggested models are simulated and compared. Thirty students' data are used for simulation. Data were collected via internet survey. Twenty four students have scored maximum on theirs qualification exam. Data on the exact number of points at the entrance exam for Master of Science studies obtained on matters related to operational researches and statistics do not exist, because it is not important for current ranking model. Students were asked to write their hypothesis about scored points at that study fields (fuzzy number) which is converted in according to Chen and Hwang's conversion scales with five different linguistic terms. Data on the eight students is shown in following table. These students are chosen because their characteristics are suitable to present differences in the rankings using different models.

Table 1. Input data

Student	u	k	a	b	c	d	x
A	100	100	1.00	1.00	1.00	1.00	1.00
B	99.0	100	0.94	0.90	1.00	0.92	1.00
C	98.6	100	1.00	1.00	1.00	1.00	1.00
D	90.7	100	0.88	1.00	0.70	0.70	1.00
E	89.0	100	0.90	0.90	0.90	0.83	1.00
F	77.0	98.0	0.62	0.60	0.65	0.68	1.00
G	88.0	85.0	0.90	0.90	0.90	0.90	0.95
H	77.0	80.0	0.72	0.70	0.75	0.80	0.90

Table 2. Output data

Student	Current model	Model based on weighted sum	Model based on fuzzy logic	Model 1 based on pseudo LA	Model 2 based on pseudo LA
A	100.00	100.00	100.00	100.00	100.00
B	99.60	99.10	98.50	99.60	99.54
C	99.44	99.58	99.58	99.58	99.58
D	96.28	95.21	93.21	96.49	91.21
E	95.60	95.47	94.23	96.36	94.56
F	89.60	88.53	84.97	89.67	85.34
G	86.20	87.40	86.00	88.61	86.81
H	78.80	79.60	76.60	81.53	77.92

Table 3. Students‘ rankings

Student	Current model	Model based on weighted sum	Model based on fuzzy logic	Model 1 based on pseudo LA	Model 2 based on pseudo LA
A	1	1	1	1	1
B	2	3	3	2	3
C	3	2	2	3	2
D	4	5	5	4	5
E	5	4	4	5	4
F	6	6	7	6	7
G	7	7	6	7	6
H	8	8	8	8	8

Rankings given by different models differ. The first model is an initial model, which is applying in practice. The second and the third model are transitional. Last two models are the final proposed models. Differences in students' ranking are caused by emphasizing specific factors and different method of aggregation. Users can model their own pseudo-logical function according to their needs.

Student A and student H are the first and the last at every ranking list. It is noticeable that students B and C, students D and E, and students F and G have different rankings when different model is in use. Student B has higher average grade than student C, so he is better than student C in accordance to initial model. Student B has much lower grade at subjects that are focused on operational researches than student C, so other models favour student C. In first model based on pseudo LA that variable doesn't figure (that model has arithmetic mean of grades at subjects that are focused on operational researches and also statistics, as one variable). It is similar for students G and F. The main difference between students D and E is that student E's grades are balanced, opposite of student D's. This explains the large difference between these students when the last model is in use (it has conjunction of attributes that have the most different values).

7 Conclusion

The Master of Science programs at Faculty of Organizational Sciences specialize students for different occupations – managers, programmers, analytics, database specialists, etc. The advantage at ranking should be given to students who showed great results at subjects which are related to chosen domains. The ranking model must acknowledge the requirements that are expressed using the logical connective and should process fuzzy values, so the usage of fuzzy logic is necessary. It is important to find and point out small differences between inputs and to treat negation in a proper way. That can be achieved by using the suggested models based on LA. LA is a technique which gives the user the most options in modeling. Ranking students with pseudo-logical function gives results similar to current, but different enough to introduce these proposals and test it practically. Users can model their own pseudo-logical function in accordance to their requests and needs.

References

1. Chen, S.J., Hwang, C.L., Hwang, F.P.: Fuzzy Multiple Attribute Decision Making. Methods and Applications. Springer, New York (1992)
2. Hsieh, C.S., Wu, C.H., Huang, T., Chen, Y.W.: Analyzing the Characteristics of Fuzzy Synthetic Decision Methods on Evaluating Student's Academic Achievement. In: International Conference on Artificial Intelligence and Computational Intelligence, pp. 491–495. Conference Publications (2009)
3. Mirković, M., Hodolič, J., Radojević, D.: Aggregation for Quality Management. Yugoslav Journal for Operational Research 16(2), 177–188 (2006), doi:10.2298/YJOR0602177M
4. Radojević, D.: Interpolative Realization of Boolean Algebra. In: The 8th Symposium on Neural Network Application in Electrical Engineering, pp. 201–206. Conference Publications (2006), doi:10.1109/NEUREL.2006.341214
5. Radojević, D.: Interpolative realization of Boolean algebra as a consistent frame for gradation and/or fuzziness. In: Nikravesh, M., Kacprzyk, J., Zadeh, L.A. (eds.) Forging New Frontiers: Fuzzy Pioneers II. STUDFUZZ, vol. 218, pp. 326–351. Springer, Heidelberg (2007)
6. Radojević, D.: Fuzzy Set Theory in Boolean Frame. International Journal of Computers Communications and Control 3(S), 121–131 (2008)
7. Radojević, D.: Logical Aggregation Based on Interpolative Boolean Algebra. Mathware & Soft Computing 15, 125–141 (2008)
8. Ross, T.: Fuzzy Logic With Engineering Application, 3rd edn. John Wiley & Sons, Ltd., Chichester (2010)
9. Triantaphyllou, E.: Multi-Criteria Decision Making Methods: A Comparative Study. Kluwer Acad. Publ., Dordrecht (2000)
10. Yeh, C.: The Selection of Multiattribute Decision Making Methods for Scholarship Student Selection. International Journal of Selection and Assessment 11(4), 289–296 (2003), doi:10.1111/j.0965-075X.2003.00252.x
11. Zadeh, A.L.: Is there a need for fuzzy logic? Information Sciences: an International Journal 178(13), 2751–2779 (2008), doi:10.1016/j.ins.2008.02.012