University of Banjaluka, Faculty of Agriculture
Agro-knowledge Journal

Classical and Fuzzy Logic Evaluation of Students' Master Theses in Matlab Fuzzy Logic Toolbox Software: Dealing with Subjectivity in Human Reasoning

Đurađ Hajder[1], Nikola Mićić[1]

[1]*University of Banja Luka, Faculty of Agriculture, Republic of Srpska, BiH*

Abstract

The two−level evaluation of defined objectives, presented materials and methods and interpretation of results in master theses was done, in order to estimate their scientific contribution and statistical relevance. First level of evaluation was performed using classical methods and consisted of three steps: defining criteria of evaluation, analyzing their fulfilment and positioning 26 master theses into the Likert−type scale in the range from 0 to 1. Second level of evaluation was based on fuzzy logic methodology, conducted mostly in Matlab Fuzzy Logic Toolbox software and consisted of definition of variables, fuzzification, fuzzy inference, defuzzification and interpretation. Obtained marks from two levels of evaluation were than compared. Results indicate that fulfilment of defined criteria of evaluation is moderate. Common mistakes made by authors are accentuated, and clear advices for improving scientific contribution of theses were pointed out here. Classical evaluation marks were higher in 96.15% cases (or 25 out of 26 theses). However, fuzzy approach has advantages, which is also discussed. The interpretation of research results, defined as logical−mathematical argumentation, was found to have the leading role in forming mark in both levels of evaluation.

Key words: expert knowledge, descriptive biometrics, fuzzy set, evaluation criteria, article structure

Introduction

For educational purposes, evaluation can generally be defined in term of measuring the scientific contribution of an individual or an institution, or as a process by which something is measured by comparing it with defined standards and criteria (Pavlović, 2016). Evaluation also concerns merit and worth of the data as applied to a specific use or context (McMillan, 2000). In past, classical methods of evaluation were grounded on traditional logic and binary mathematics but new approaches emerged, like fuzzy logic.

Concepts of fuzzy logic and fuzzy sets were firstly introduced by Zadeh (1965). Fuzzy logic in a narrow sense is a logical system that aims at a formalization of approximate reasoning and is rooted in multivalued logic but in a broad sense is almost synonymous with fuzzy set theory, while the fuzzy set is a class with a continuum of grades of membership and consists of objects in which the transition from membership to non−membership is gradual rather than abrupt. A fuzzy set A in X is characterized by a membership function $fA(x)$ which associates with each point in X a real number in the interval from 0 to 1 with the value of $fA(x)$ at x representing the "grade of membership" of x in A (Zadeh, 1965, 1994; Zimmermann, 2010). Important concept in a fuzzy set theory is a linguistic variable, defined as a variable whose values are sentences in a natural or artificial language in place of or in addition to numerical variables (Zadeh, 1973). In practice, this variable (e.g. success, height etc.) is segmented by fuzzy labels (e.g. good, tall etc.) defined by a range and a specific membership function (*mf* henceforth). Triangular and trapezoidal *mf* are often used in practice, due to their simplicity, but based on a specific problem.

The development of fuzzy methodology enabled studying possibilities of application of fuzzy logic in education. The fuzzy educational grading systems were investigated (Law, 1996). Multifactorial fuzzy clustering (Wang & Bell, 1996) was used to evaluate college and high school students. Student portfolios were assessed by a fuzzy logic (Fourali, 1997). An expert fuzzy classification scoring system is used for evaluation of students' writing samples (Nolan, 1998). Fuzzy models for classification of students (Nykänen, 2006) or a two−level personnel selection (Petrović−Lazarević, 2001) were used. Fuzzy−based grading system in evaluation of postgraduates' research work (McLoone, 2012) and students' performance in oral presentations (Daud et al, 2011) was presented. The fuzzy evaluation of teachers' academic performance (Chaudhari et al., 2012) and faculty performance as well is also investigated (Guruprasad, et al., 2016; Jyothi et al., 2014). New studies are still oriented to evaluation of students' performance by a fuzzy logic (Kharola et al., 2015; Surya et al., 2016; Varghese et al., 2017).

Induced hesitant fuzzy uncertain linguistic correlated averaging was proposed in evaluation of scientific publications (Xu et al., 2015). Some other possibilities for the evaluation of faculty teachers' work are present (Pavlović, 2016). Inspirational study on how to evaluate the quality of scientific research was done by Du Prel et al. (2009).

Research objectives are closely related to the scientific field, current issues in science, methodology and the utility of key findings and can serve as the universal guide in all research phases. If well−formed, with actuality and importance of main findings defined, they can lead a researcher to a valuable conclusions. Descriptive statistics comprise description, analysis, classification and comparing so the important prerequisite for researchers is to properly define utilized methods and biometrical unit(s) of observation. In biometrics, the correct interpretation of experimental results can be defined as logical−mathematical argumentation (Mićić et al., 2014b). Master and doctoral degree programs are also connected to the process of publishing scientific articles, whether it concerns traditional master and PhD theses or an alternative culminating projects such as multiple article format (Thomas et al., 2016). Statistical methods represent a base for performing most scientific experiments, but it is indicative that authors sometimes use incorrect or misleading methodology and fail to define research objectives or to interpret results properly. Therefore, it will be valuable to somehow estimate the scientific contribution of these publications.

However, necessity for a thorough evaluation of scientific publications in education seems reasonable. Here, it was assumed that by evaluation, the level of a scientific contribution and the relevance of descriptive statistics presented in master theses can be estimated. In addition, fuzzy logic was used to overcome disadvantages of a classical evaluation in education (lack of criteria, subjectivity, sharp boundaries between marks etc.). Therefore, the aim of this study is the evaluation of defined objectives, presented materials and methods and interpretation of results from analyzed master theses.

Material and Methods

Master theses defended at the Faculty of Agriculture in the period 1994–2015 were main objects of this investigation. Specific group for further analyses was selected from all master theses from this period, and consists of 26 master theses in which mainly descriptive statistical approach was used. Most experiments in agriculture combine both descriptive and inferential statistics and descriptive data often present a base for inferential methods.

Nevertheless, in this research, a descriptive statistical approach was defined in term of investigations concerning statistical population. Somewhere, it refers to a research with only descriptive methods or measures presented (description, arithmetic mean and standard deviation calculated etc.). Identified 26 theses with descriptive statistics were subjected to thorough analysis. There were two levels of evaluation in our research: classical evaluation (the first level) and fuzzy evaluation (the second level).

Classical evaluation

Classical evaluation consisted of three steps. The first step was the assessment of following sections from individual theses: 1) defined objective(s) and hypotheses (OB); 2) materials and methods (MM) and 3) logical−mathematical argumentation (LMA). To evaluate these sections, specific criteria were defined for OB (1.1. clarity of objectives, their consistency and conciseness, 1.2. necessity and motives for investigation, 1.3. actuality, importance and application of main findings), MM (2.1. definition and relevance of used materials, 2.2. suitability of planned methods) and LMA section (3.1. compatibility with defined objectives, 3.2. control of variation, 3.3. interpretation of statistical tests). Second step was detecting whether pre−defined criteria were fulfilled in these 26 theses. Based on fulfilment of criteria, it is possible to quantify the scientific contribution and the relevance of descriptive statistics for individual thesis to some extent, by setting each thesis into a specific range on a scale. Hence, in a third step a Likert−type or university grading scale was used to distribute all thesis into the range from 0 to 5 with values graded ascending. At this step, an arithmetic mean of experts' marks for three sections from analysed master theses was used as a value on a scale. These values represent a basis for performing the fuzzy logic methodology in the second level of evaluation.

Fuzzy evaluation

Fuzzy evaluation was carried out in Matlab Fuzzy Logic Toolbox software (R2016a 9.0 version), due to its speed and simplicity. It was also assumed that this approach could deal with subjectivity in human reasoning (or expert marks in this case), so the fuzzy output can provide less subjective marks and/or marks with no sharp boundaries between them (i.e. master thesis can be both *good* and *very good* to some extent, in %). For analyses, *min* method was used for the AND operator, *min* method for implication, *max* method for aggregation and centroid method for defuzzification, according to the Mamdani fuzzy inference system.

The fuzzy methodology included definition of variables, fuzzification (design of *mf* and linguistic variables), fuzzy inference (fuzzy if−then rules formation, aggregation, activation and accumulation), defuzzification (transforming inputs to crisp values) and interpretation.

Fuzzy inputs in our research were 1. objectives (OB), 2. material and methods (MM) and 3. logical−mathematical argumentation (LMA). Single fuzzy output was defined as fuzzy evaluation value (FEV). Fuzzification started by forming a same linguistic variable for all fuzzy inputs and fuzzy output, named "master thesis quality". Linguistic variable was then segmented to fuzzy labels *sufficient* (S), *desirable* (D) and *outstanding* (O) for three fuzzy inputs and *adequate* (A), *good* (G), *very good* (VG), *excellent* (E) and *remarkable* (R) for a single fuzzy output. All fuzzy labels were graded ascending and positioned on a scale in the range from 0 to 5. The next step was forming *mf* for inputs and the output. Fuzzy *mf* consists of a support S (a crisp set containing all elements x with $\mu_A (x) > 0$), the core C (a crisp set containing all elements x with $\mu_A (x) = 1$), the height *hgt* (the supremum of all $\mu_A (x)$ of A), the singleton (fuzzy set whose support is a single point with $\mu_A (x) = 1$), and the α−cut Aα (crisp subset of A with elements whose $\mu_A (x) > \alpha$ or $\mu_A (x) \geq \alpha$), where $\mu_A (x)$ is a fuzzy degree of membership and $A_{\alpha=0.5}$ is a crossover point.

For all inputs a trapezoidal *mf* was used. This function is determined by four scalar parameters (*a*, *b*, *c* and *d*) where *a* and *d* locate the "feet" of the trapezoid while *b* and *c* locate the "shoulders" (MatWorks, 2015). The first fuzzy input (OB) had a function defined as f_S (0, 0, 1.5, 2.5), f_D (1.5, 2.5, 3.5, 4.0) and f_O (3.5, 4.0, 5.0, 5.0). MM had a function defined as f_S (0, 0, 1.0, 2.0), f_D (1.0, 2.0, 3.0, 3.5) and f_O (3.0, 3.5, 5.0, 5.0). LMA had a function defined as f_S (0, 0, 2.0, 3.0), f_D (2.0, 3.0, 4.0, 4.5) and f_O (4.0, 4.5, 5.0, 5.0).

FEV had a combination of triangular and trapezoidal *mf*. Triangular *mf* is defined by three scalar parameters (*a*, *b*, *c*) where *a* and *c* locate the "feet" of the triangle while *b* locates the peak (MatWorks, 2015). FEV had a function defined as f_A (0, 0, 1.0, 1.5), f_G (1.0, 1.5, 2.0), f_{VG} (1.5, 2.0, 3.0, 3.5), f_E (3.0, 3.5, 4.0) and f_R (3.5, 4.0, 5.0, 5.0). The $\mu_A (x)$ to fuzzy labels (ranging from 0 to 1) can be calculated according to formulas for specific *mf* (MatWorks, 2015).

Considering that there are three fuzzy inputs (OB, MM and LMA) and that every input can have three different labels (S, D, O) there are 27 rules in the fuzzy rule base.

After defining this range, different combinations of fuzzy inputs resulting in specific fuzzy output should be determined. This can be done only by a thorough understanding of the given scientific field of research and consulting the expert knowledge. A list of fuzzy rules is determined as follows:

A list of fuzzy rules is determined as follows:

1. IF OB is S and MM is S and LMA is S THEN FEV is A;
2. IF OB is S and MM is D and LMA is S THEN FEV is A;
3. IF OB is S and MM is O and LMA is S THEN FEV is A;
4. IF OB is D and MM is S and LMA is S THEN FEV is G
5. IF OB is D and MM is D and LMA is S THEN FEV is G
6. IF OB is D and MM is O and LMA is S THEN FEV is G
7. IF OB is O and MM is S and LMA is S THEN FEV is G
8. IF OB is O and MM is D and LMA is S THEN FEV is G
9. IF OB is O and MM is O and LMA is S THEN FEV is G
10. IF OB is S and MM is S and LMA is D THEN FEV is VG
11. IF OB is S and MM is D and LMA is D THEN FEV is VG
12. IF OB is S and MM is O and LMA is D THEN FEV is VG
13. IF OB is D and MM is S and LMA is D THEN FEV is VG
14. IF OB is D and MM is D and LMA is D THEN FEV is VG
15. IF OB is D and MM is O and LMA is D THEN FEV is VG
16. IF OB is S and MM is S and LMA is O THEN FEV is VG
17. IF OB is S and MM is D and LMA is O THEN FEV is VG
18. IF OB is S and MM is O and LMA is O THEN FEV is VG
19. IF OB is O and MM is S and LMA is D THEN FEV is E
20. IF OB is O and MM is D and LMA is D THEN FEV is E
21. IF OB is O and MM is O and LMA is D THEN FEV is E
22. IF OB is D and MM is S and LMA is O THEN FEV is E
23. IF OB is D and MM is D and LMA is O THEN FEV is R
24. IF OB is D and MM is O and LMA is O THEN FEV is R
25. IF OB is O and MM is S and LMA is O THEN FEV is R
26. IF OB is O and MM is D and LMA is O THEN FEV is R
27. IF OB is O and MM is O and LMA is O THEN FEV is R

Inputs for centroid defuzzification were triplets of arithmetic means (average experts' marks) for OB, MM and LMA. This triplets were defuzzified to 26 numbers, for which the μ_A (x) to fuzzy output labels was calculated. Comparison between fuzzy and classical marks was presented and a general attitude toward all theses was discussed. Expert knowledge accentuated the effect of criteria fulfilment for OB and LMA so their effect on FEV was presented in Fuzzy Logic Toolbox 3D Surface Viewer (MatWorks, 2015).

Results and Discussion

Experts' evaluative marks from classical evaluation

The distribution of 26 master theses into the range from 0 to 5, based on the average experts' marks from classical evaluation as a value on a scale (as their quantified scientific contribution) is presented in Table 1.

Tab. 1. The experts' evaluative marks for individual sections from 26 master theses
Евалуативне оцјене експерата за поглавља из 26 мастер радова

Thesis code	Evaluation mark			Thesis code	Evaluation mark		
	OB*	MM	LMA		OB	MM	LMA
1	3.83	4.43	4.43	*14*	3.63	3.40	3.13
2	2.43	3.00	2.23	*15*	2.83	2.37	2.77
3	2.40	2.77	2.43	*16*	2.00	2.73	2.70
4	4.23	4.40	4.50	*17*	2.70	3.57	3.13
5	3.43	3.87	3.93	*18*	2.17	2.57	2.17
6	2.07	2.90	2.67	*19*	1.80	1.90	1.37
7	2.90	3.17	3.53	*20*	3.77	3.87	3.07
8	3.53	4.23	3.97	*21*	3.33	3.03	2.87
9	2.03	3.40	3.03	*22*	4.23	4.70	4.47
10	3.93	4.47	4.37	*23*	3.07	3.03	2.27
11	1.50	0.93	1.13	*24*	3.67	3.77	3.13
12	2.37	2.50	1.83	*25*	4.37	4.43	4.03
13	4.10	4.57	4.40	*26*	4.30	4.03	3.87

Note: OB = objectives, MM = material and methods, LMA = logical−mathematical argumentation (average)

The lowest mark for OB section was 1.50 and the highest was 4.37, with average of 3.10 ± 0.17 and coefficient of variation (CV henceforth) of 28.49%. The lowest mark for MM was 0.93 and the highest was 4.70, with average of 3.39 ± 0.18 and CV of 27.39%. The lowest mark for LMA was 1.13 and the highest was 4.50, with average of 3.08 ± 0.19 and CV of 31.07%.

Considering that defined Likert−type scale is graded ascending, this is a relatively good result. It seems that fulfilment of three OB criteria (1.1., 1.2. and 1.3.) is satisfactory in average. However, authors of theses with low evaluation mark for OB, defined their objectives too theoretically and some OB criteria (1.2. and 1.3.) were partly fulfilled. A descriptive statistical approach is based on description and argumentation of biometrical units of observation in finite and countable sets (Mićić et al., 2014a) so the objectives should refer to population and its characteristics. Clear connection between objectives and interpretation of results is here a prerequisite. Researchers should take care of essential measurement evidence skills regarding the ability to understand and interpret the meaning of descriptive statistical procedures (McMillan, 2000).

Two MM criteria (2.1. and 2.2.) were moderately fulfilled. A main obstacle for authors of theses with low MM mark was the incorrect usage of different statistical software.

As it is assumed (McMillan, 2000), there is a danger that technology will contribute to the mindless use of new resources, such as using items on−line developed by some companies without adequate evidence of reliability, validity and fairness, and crunching numbers within software without sufficient thought about weighting, error, and averaging. According to our research, a trend in misunderstanding these concepts is still present. That is the reason why an increased technological potential does not automatically mean better applications (Nykänen, 2006). Although software tools facilitate calculation and interpretation of large databases, there are many cases of incomprehension of basic assumptions for performing statistical tests so software tools' inference suggestions are uncritically accepted (Mićić et al., 2014b).

Second obstacle in MM section for the authors was missing to define a fundamental statistical concepts like biometrical unit of research, population size etc. This led to some confusions while interpreting the results. Nevertheless, planned statistical methods in theses with low MM value did not match defined objectives, presented data and argumentation of results.

The lowest general evaluation mark was obtained for the LMA, with an average of 3.08 ± 0.19. Considering that experts' opinion was that LMA section have the greatest impact on evaluative value of individual thesis, this result should be much better in order to improve the scientific contribution and the relevance of statistics in each thesis. For the fulfilment of the first LMA criterion (3.1.) a clear connection between defined objectives and presented results should be established. It is important to discuss what does presented results imply and whether they are accomplished. In case of ill−defined objectives this is impossible to achieve, resulting in low LMA values here.

Fulfilment of the second LMA criterion (3.2.) is also very important. However, authors often missed to detect and explain high coefficients of variation. It is critical that all educators understand concepts like standard error of measurement, reliability coefficients, confidence intervals, and standard setting (McMillan, 2000). For example, performing t–test statistics with very low or very high coefficients of variation can lead a researcher to fallacious conclusions (Mićić et al., 2014a). The third LMA criterion (3.3.) was also partly fulfilled. Descriptive statistics comprise different statistical procedures, whereas operating with large databases or complicated software can sometimes result in mistakes. Common pitfalls occur when authors fail to detect data regularity, interpret data not in accordance with theoretical values, overlook unusual values or collect data carelessly so the LMA in descriptive statistics is related to drawing conclusions based on the manifestation of a given phenomenon or different states of biometrical units of observation (Mićić & Bosančić, 2013; Mićić et al., 2014a).

Common mistakes in analysed theses occurred when authors underestimated mentioned criteria, in particular the compatibility of OB with LMA and MM. Even if this is fulfilled, there is a possibility of misunderstanding the statistical software outcome so a researcher should always have a major role regarding methodology and interpretation of results.

Fuzzification of classical evaluation marks

The fuzzy degree of membership $\mu_A(x)$ to different fuzzy output labels is presented in Table 2.

Tab. 2. The fuzzy degree of membership to fuzzy output labels for 26 master theses
Степен припадности епитетима фази излаза за 26 мастер радова

Thesis code	the fuzzy degree of membership $\mu_A(x)$				
	adequate (A)	*good* (G)	*very good* (VG)	*excellent* (E)	*remarkable* (R)
11	1	0	0	0	0
19	1	0	0	0	0
12	0.44	0.56	0	0	0
18	0.08	0.92	0	0	0
2	0	0.24	0.76	0	0
16	0	0.22	0.78	0	0
6	0	0.22	0.78	0	0
3	0	0	1	0	0
5	0	0	1	0	0
7	0	0	1	0	0
17	0	0	1	0	0
8	0	0	1	0	0
9	0	0	1	0	0
14	0	0	1	0	0
15	0	0	1	0	0
20	0	0	1	0	0
21	0	0	1	0	0
23	0	0	1	0	0
24	0	0	1	0	0
26	0	0	0	1	0
25	0	0	0	0.78	0.22
1	0	0	0	0.18	0.82
10	0	0	0	0.16	0.84
4	0	0	0	0	1
22	0	0	0	0	1
13	0	0	0	0	1

The μ_A (x) was calculated by a subjection of classical marks to fuzzy methodology. As a product of defuzzification, 26 numeric values were obtained, named as fuzzy evaluation value (FEV). According to Table 2, fuzzy degrees of membership are grouping in the centre of the scale. The μ_A (x) = 1 for the fuzzy output label *very good* (VG) was achieved in 46.15% of cases (or 12 out of 26 master theses). Other theses had different μ_A (x) ranging from 0 to 1 and almost uniformly distributed on both sides of a scale. The μ_A (x) = 1 was achieved in 7.69% of cases for the fuzzy output label *adequate* (A) (2 theses), 3.85% of cases for the label *excellent* (E) (1 thesis) and 11.54% of cases for the label *remarkable* (R) (3 theses), respectively. Remaining 8 theses had μ_A (x) ranging from 0 to 1 and belonging to two contiguous fuzzy output labels. For example, thesis 1 was 18% *excellent* and 82% *remarkable*. Grouping of FEV values in the centre of the scale and uniform distribution of remaining values on both sides of a scale is a consequence of classical marks, as well as fuzzy methodology, especially the fuzzy scale, the specific range, type and shape of fuzzy *mf* and fuzzy rules formation.

Comparison between classical and fuzzy evaluation methodology

After averaging OB, MM and LMA marks for each thesis individually, it is possible to compare obtained classical and defuzzification values (Table 3).

Tab. 3. Comparison between classical and defuzzification marks for 26 master theses
Компарација оцјена класичне и фази евалуације за 26 мастер радова

Thesis code	EM*	FEV	Δ	Thesis code	EM	FEV	Δ
1	4.23	3.91	0.32	*14*	3.39	2.62	0.77
2	2.55	1.88	0.67	*15*	2.66	2.40	0.26
3	2.53	2.05	0.48	*16*	2.48	1.89	0.59
4	4.38	4.38	0.00	*17*	3.13	2.50	0.63
5	3.74	2.50	1.24	*18*	2.30	1.46	0.84
6	2.55	1.89	0.66	*19*	1.69	0.79	0.90
7	3.20	2.50	0.70	*20*	3.57	2.78	0.79
8	3.91	2.53	1.38	*21*	3.08	2.44	0.64
9	2.82	2.50	0.32	*22*	4.47	4.35	0.12
10	4.26	3.92	0.34	*23*	2.79	2.05	0.74
11	1.19	0.62	0.57	*24*	3.52	2.66	0.86
12	2.23	1.28	0.95	*25*	4.28	3.61	0.67
13	4.36	4.27	0.09	*26*	4.07	3.50	0.57

*Note: EM = experts' marks (average for sections), FEV = fuzzy evaluation value (defuzzification value), Δ = EM – FEV

Although the average marks from the classical evaluation were inputs to defuzzification, presented data indicate how specific fuzzy methodology (the design of fuzzy *mf*, formation of fuzzy rules etc.) can affect defuzzification process. It also influences its products i.e. fuzzy evaluation value (FEV). Based on data in Table 3, the average EM marks are in all cases (except thesis 4) higher that FEV values. This difference varies from 0.09 (thesis 13) to 1.38 (thesis 8).

The comparison between classical and fuzzy evaluation indicate that 96.15% of theses (or 25 out of 26 theses) obtained higher classical mark. This is similar to some other studies. In three out of five experiments conducted (Kharola et al., 2015) classical mark was higher than fuzzy mark. Guruprasad et al. (2016) obtained higher classical (statistical) value in 90.91% of cases (or 10 out of 11 cases) in faculty performance evaluation. Based on theory, laboratory and project data (Surya et al., 2016) student performance was higher in all nine cases, by using classical methodology.

The effect of OB and LMA fuzzy inputs on FEV fuzzy output

Given that arithmetic mean, once established, is unchangeable except new experts evaluate theses or new criteria are established, the advantage of fuzzy approach is a possibility of modelling the level of severity of evaluation criteria by changing the fuzzy methodology. Fuzzy grading system was also found to be more equitable in Law (1996). Fuzzy evaluation from our research obviously had more strictly criteria, with LMA section affecting the FEV value predominantly. In evaluation of students' portfolios (McLoone, 2012), fuzzy−based grading provided higher values in 73.8% of cases (or 31 out of 42 cases). In the same study it was concluded that the fuzzy−based system eases the assessment process for the assessor as it is simpler to choose a linguistic variable than it is to assign a specific numerical value but none of these two approaches adequately deal with other aspects of subjectivity. This fact was previously claimed (Nolan, 1998), i.e. there is no way of ensuring that scoring rules and standards are being applied the same way by different raters or in the same exact way by a particular teacher. The disadvantage of a peer review, as a common academic evaluation method, was also pointed out (Xu et al., 2015).

However, there is some additional evidence regarding advantages of a fuzzy mark. In evaluation of oral presentations (Daud et al., 2011) fuzzy marks earned by students were higher in all 10 cases. In evaluation of academic performance of teachers (Chaudhari et al., 2012) obtained fuzzy marks were higher in 70.97% of cases (or 20 out of 31 cases).

In student performance evaluation (Sakthivel et al., 2013) two fuzzy−based approaches provided higher marks in 65.0% (13 out of 20 cases) and 55.0% of cases (11 out of 20 cases). Fuzzy approach in Jyothi et al. (2014) resulted in higher marks in all 10 cases in evaluation of faculty performance.

The effect of OB and LMA fuzzy inputs on fuzzy output (FEV)

Defined rules in this research enabled studying the influence of fuzzy inputs on fuzzy output. This is important in order to predict this influence or to quantify expert knowledge. As a read−only tool, FLT 3D Surface Viewer (MatWorks, 2015) can provide valuable data considering previous research and fuzzy rule base forming as well. Figure 1 presents the influence of two fuzzy inputs (OB and LMA) on a fuzzy output (FEV) while the remaining fuzzy input (MM) is held constant. There are several plateaus, representing a stagnation of FEV by increasing the evaluative values of OB and LMA. Between these plateaus, FEV increases by increasing the evaluative values of OB and LMA.
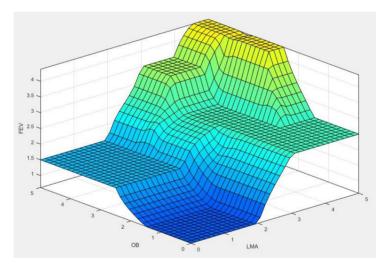

Fig.1. The effect of objectives (OB) and logical−mathematical argumentation (LMA) on a fuzzy evaluation value (FEV) of analysed theses
Утицај циља истраживања (ОВ) и логичко−математичке аргументације (LMA) на фази евалуативну вриједност (FEV) анализираних радова

The first plateau is positioned in interval $0 \rightarrow 1.5$ for OB and $0 \rightarrow 2$ for LMA, which refers to the $\mu_A(x) = 1$ for the fuzzy input label *sufficient* (S). The FEV than rises, in the interval $1.5 \rightarrow 2.5$ for OB and $2 \rightarrow 3$ for LMA.

This growth is followed by a lower $\mu_A(x)$ of OB and LMA to fuzzy input label *sufficient* (S) and higher to the fuzzy input label *desirable* (D), where $\mu_A(x) = 0 \rightarrow 1$. The second plateau is positioned in interval $2.5 \rightarrow 3.5$ for OB and $3 \rightarrow 4$ for LMA, which refers to the $\mu_A(x) = 1$ for the fuzzy input label *desirable* (D). In the interval $3.5 \rightarrow 4$ for OB and $4 \rightarrow 4.5$ for LMA, the FEV again raises, which is followed by a lower $\mu_A(x)$ of OB and LMA to fuzzy input label *desirable* (D) and higher $\mu_A(x)$ to the fuzzy input label *outstanding* (O), where $\mu_A(x) = 0 \rightarrow 1$. The third plateau is positioned in interval $4 \rightarrow 5$ for OB and $4.5 \rightarrow 5$ for LMA, where FEV remains the same regardless of increasing evaluative values of OB and LMA, and $\mu_A(x) = 1$ for the fuzzy input label *outstanding* (O).

It is indicative that defined fuzzy methodology provided a smooth transition between fuzzy output labels. Nevertheless, increasing of FEV is predominantly affected by evaluative marks for LMA section. Evaluative marks for OB section had similar tendency, though less active.

Conclusion

In order to estimate the scientific contribution and the relevance of descriptive statistics presented in students' master theses, a two−level evaluation of defined objectives, presented materials and methods and interpretation of results in individual theses was done. The two−level evaluation indicates that scientific contribution of master theses is mediocre. Descriptive statistics presented also lacks expert knowledge, in order to widen the scientific relevance and practical application of results. Classical evaluation resulted in higher average marks for 25 out of 26 analysed theses.

Based on these results, we can define implications of a paramount importance in further research. In the first level of evaluation, which was consisted of setting criteria, analysis of their fulfilment in individual theses and distribution of all theses into Likert−type scale, the fulfilment of criteria for OB, MM and LMA sections was moderate. A great gap between these sections exists (a gap which can also be present in other scientific publications as well) i.e. these three sections must always be well inter−connected in scientific publications. Though this seems as a well−known fact, it is important to note that similar gaps often occur in students' writings. With regard to research and development in education, here lies the importance of constant evaluation of second as well as third cycle degree programs. Special emphasis should be placed on master and PhD theses. However, there is an opportunity to overcome disadvantages of a traditional evaluation (like uncertainty, subjectivity and sharp boundaries between classes or assigned marks).

Fuzzy logic can therefore facilitate evaluation process. The flexibility of a fuzzy logic approach reflects in potential for changing the severity of established evaluation criteria by adapting fuzzy methodology, instead of substitution of evaluators or criteria. This is not possible in classical approach. Nevertheless, these two levels of evaluation actually complement each other, because classical marks serve as inputs for defuzzification. These findings suggest that fuzzy approach should rather improve traditional methods of evaluation instead of being their alternative. This approach could also help teachers evaluate different students' writings without fear of being too subjective and straightforward while forming marks.

Development of new technologies can significantly contribute to this scope, because software solutions can facilitate evaluation processes in education. Fuzzy Logic Toolbox software (FLT) used in this study showed speed, simplicity and potential to enhance evaluation. Questions regarding adjustments of fuzzy inputs and outputs, modelling fuzzy *mf*, forming fuzzy rule base as well as analyses of effect of fuzzy inputs on fuzzy output remain open, but after setting fuzzy methodology, different calculations are relatively simple. FLT supports marks from the first level of evaluation as inputs, and can help teachers evaluate large number of different publications after specific methodology is set. Nevertheless, defuzzification of fuzzy values to a crisp value is automatic at this step. Defuzzification values can further be analysed, compared and distributed into different scales, and eventually generalized for a specific field of research.

Specific group for a two−level evaluation in this study consisted of 26 master theses in which mainly descriptive statistical approach was used. Therefore, possible generalizations can be done only for master theses with descriptive statistical measures or tests and in accordance with presented evaluation criteria and fuzzy rules. This study was mainly based on the evaluation of OB, MM and LMA section with criteria defined in order to avoid subjectivity and biases of evaluators while assessing theses, but including other sections or choosing different and/or additional evaluation criteria could have helped address some more questions regarding scientific contribution and relevance of descriptive statistics in master theses. Additional limitations also emerge in the design of fuzzy rules and fuzzy *mf* which are quite specific.

The lack of similar research data and study methodologies in a two−level evaluation in education can also be a constraint. Some more comparisons could have been made in that case, with discussion in this field expanded. This could influence traditional as well as fuzzy methodology in domain of educational research and development, particularly the evaluation of different students' writings (oral and poster presentations, master and PhD theses, scientific articles etc.).

Analysed master theses generally lack expert knowledge in domains of scientific methodology, the usage of different statistical software, control of research variables and their variation as well as the argumentation of obtained results. Therefore, if a student aims at future academic engagement it would be useful to discuss these issues in different courses, particularly regarding descriptive statistics.

Advices for improving statistical relevance and scientific contribution of students' writings were previously discussed. This advices are very significant because they could help students in various ways (in setting experimental designs, in writing scientific publications, in preparation for PhD studies etc.). They could also help teachers and evaluators (in advancing knowledge about evaluation process in education, in forming students' marks and in approaching educational technology i.e. evaluation software and specific methodologies).

Future research is also needed in context of two levels of evaluation. In the first level of evaluation, it is necessary to reassess general structure of a publication and defined evaluation criteria. In the second level of evaluation, adjustments of a fuzzy methodology, especially the formation of fuzzy rule base and the design of fuzzy *mf* could provide better insight into the evaluation outcome i.e. fuzzy marks. The fuzzy evaluation value (FEV) involves less subjectivity, because boundaries between fuzzy outputs, as specific marks, are less sharp and straightforward, so the difference between *good* (G) and *very good* (VG) is relativized. This value is found to be predominantly affected by the logical−mathematical argumentation (LMA) so researchers should always consult an expert knowledge in context of classical and fuzzy methodology in evaluation as well as study designs and interpretation of research results.

Testing some other fuzzy *mf* (like Gaussian or sigmoidal) or the adjustment of triangular and trapezoidal *mf* used here is a question that remains open. In this respect, the support S of triangular and trapezoidal *mf* used, as well as the core C of trapezoidal *mf* could be shortened so the "safe zone" (with all elements x having the degree of membership $\mu_A (x) = 1$) decreases. In this case we are less certain which thesis is 100% *good* or 100% *very good*. This could change the distribution of theses into defined scale, but could also reduce evaluators' biases in the assessment of students' writings.

Not less important, studies should be undertaken to see if valid and promissory results are obtained in the evaluation of other scientific aspects of existing master theses, or certain aspects in master and PhD theses from different field of research and/or institution.

References

Chaudhari, O. K., Khot, P. G., & Deshmukh, K. C. (2012). Soft computing model for academic performance of teachers using fuzzy logic. *British Journal of Applied Science & Technology*, *2*(2), 213–226.

Daud, W. S. W., Aziz, K. A. A., & Sakib, E. (2011). An evaluation of students' performance in oral presentation using fuzzy approach. *Proceedings of the 10th International Annual Symposium (UMTAS 2011), Terengganu, Malaysia*, from 11–13 of July 2011, pp. 157–162. Retrieved from https://www.researchgate.net/publication/277776194_An_Evaluation_of_S tudents'_Performance_in_Oral_Presentation_Using_Fuzzy_Approach

Du Prel, J. B., Röhrig, B., & Blettner, M. (2009). Critical appraisal of scientific articles. *Deutsches Arzteblatt International*, *106*(7), 100–105.

Fourali, C. (1997). Using fuzzy logic in educational measurement: the case of portfolio assessment. *Evaluation and Research in Education*, *11*(3), 129–148.

Guruprasad, M., Sridhar, R., & Balasubramanian, S. (2016). Fuzzy logic as a tool for evaluation of performance appraisal of faculty in higher education institutions. *SHS Web of Conferences*, *26*(2016), 1–7.

Jyothi, G., Parvathi, C., Srinivas, P., & Rahaman, S. A. (2014). Fuzzy expert model for evaluation of faculty performance in technical educational institutions. *International Journal of Engineering Research and Applications*, *4*(5), 41–50.

Kharola, A., Kunwar, S., & Choudhury, G. B. (2015). Students performance evaluation: a fuzzy logic reasoning approach. *PM World Journal, 9*(10), 1–11.

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

McLoone, S. C. (2012). *On using fuzzy logic for grading highly subjective assessment material – a case study*. Paper presented at the IET Irish Signals and Systems Conference, 28-29 June 2012, Maynooth, Ireland. Retrieved from https://ieeexplore.ieee.org/document/6621150/

McMillan, J. H. (2000). Fundamental assessment principles for teachers and school administrators. *Practical Assessment, Research & Evaluation*, *7*(8), 1–5.

Mićić N., & Bosančić, B. (2013). Pitfalls of descriptive and inferential statistical approach in biological and agricultural sciences. *Agroznanje*, *14*(4), 617–630.

Mićić, N., Đurić, G., Kurtović, M., & Knezović, Z. (2014a). Biometrics as a method of scientific research in biological and agricultural sciences, *Radovi Poljoprivredno–prehrambenog fakulteta Univerziteta u Sarajevu*, *59*(64/2), 169–177.

Mićić, N., Kurtović, M., Knezović, Z., & Bosančić, B. (2014b). Research goal and logical−mathematical argumentation of biometrical analysis results, *Radovi Poljoprivredno–prehrambenog fakulteta Univerziteta u Sarajevu*, *59*(64/2), 151–160.

Nolan, J. R. (1998). A prototype application of fuzzy logic and expert systems in education assessment. *Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI−98), Madison, Wisconsin,* 1134–1139. https://www.aaai.org/Papers/IAAI/1998/IAAI98-018.pdf

Nykänen, O. (2006). Inducing fuzzy models for student classification. *Educational Technology & Society*, *9*(2), 223–234.

Pavlović, P. (2016). Evaluation of teachers' work in high education institutions of Serbia. *Ekonomika, 62*(2) (2016), 85–96.

Petrović–Lazarević, S. (2001). Personnel selection fuzzy model. *International Transactions in Operational Research*, *8*(1), 89–105.

Sakthivel, E., Senthamarai Kannan, K., & Arumugam, S. (2013). Optimized evaluation of students' performances using fuzzy logic. *International Journal of Scientific & Engineering Research*, *4*(9), 1128–1133.

Surya, A. A. Kurian, M. K., & Varghese, S. M. (2016). Overall performance evaluation of engineering students using fuzzy logic. *International Journal on Cybernetics & Informatics (IJCI)*, *5*(2), 71–78.

MatWorks (2015). *Fuzzy Logic Toolbox user's guide* (R2015b version). Natick, MA: The MathWorks Inc.

Thomas, R. A., West, R. E., & Rich, P. (2016). Benefits, challenges, and perceptions of the multiple article dissertation format in instructional technology. *Australasian Journal of Educational Technology*, *32*(2), 82–98.

Varghese, A., Kolamban S., Shreedhar, J. P., & Nayaki, S. (2017). Outcome based assessment using fuzzy logic. *International Journal of Advanced Computer Science and Applications (IJACSA)*, *8*(1), 103–106.

Wang, H., & McCauley Bell, P. (1996). Fuzzy clustering analysis and multifactorial evaluation for students' imaginative power in physics problem solving. *Fuzzy Sets and Systems*, *78*(1), 95–105.

Xu, H., Xing, L., & Huang, L. (2015). Evaluation of scientific publications with hesitant fuzzy uncertain linguistic and semantic information. *Journal of Intelligent & Fuzzy Systems*, *29*(6), 2737–2742.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, *8*(3), 338–353.

Zadeh, L. A. (1973). Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Transaction on Systems, Man and Cybernetics*, *3*(1), 28–44.

Zadeh, L. A. (1994). Fuzzy logic, neural networks and soft computing. *Communications of the ACM*, *37*(3), 77–84.

Zimmermann, H. J. (2010). Fuzzy set theory. *WIREs Computational Statistics*, *2*(3), 317–332.

Класична и фази евалуација студентских мастер теза у *Matlab Fuzzy Logic Toolbox* софтверском пакету: минимизирање субјективности у закључивању

Ђурађ Хајдер[1], Никола Мићић[1]

[1]*Универзитет у Бањој Луци, Пољопривредни факултет, Република Српска, БиХ*

Сажетак

У овом раду спроведена су два нивоа евалуације дефинисаних циљева истраживања, кориштених материјала и метода, као и интерпретације добијених резултата у магистарским и мастер тезама, у циљу процјене њиховог научног доприноса, као и релевантности кориштених биометричких метода. Први ниво евалуације реализован је употребом класичних метода кроз три кључна корака: дефинисање критеријума евалуације, анализу реализације критеријума и дистрибуцију 26 мастер теза на прилагођену Ликертову скалу у распону од 0 до 1. Други ниво евалуације се базирао на методолошком оквиру фази логике а спроведен је у *Matlab Fuzzy Logic Toolbox* софтверу кроз: дефинисање варијабли истраживања, фазификацију, фази закључивање, дефазификацију и интерпретацију. Потом је извршена компарација оцјена из два нивоа евалуације. Резултати указују на осредњу испуњеност дефинисаних критеријума евалуације. У овом раду су наглашене честе грешке у методолошком и биометричком приступу аутора анализираних теза, уз конкретне савјете за побољшање научног доприноса истих. Оцјене добијене класичним приступом су биле више у 96.15% случајева (25 од 26 теза). Међутим, фази приступ носи одређене предности, што је такође дискутовано. Кључну улогу у формирању оцјена у оба нивоа евалуације имала је интерпретација резултата научних истраживања, дефинисана као логичко−математичка аргументација.

Кључне ријечи: експертско знање, дескриптивна биометрика, фази скуп, критеријум евалуације, структура научног рада.