Degree-Based (Fuzzy) Techniques in Math and Science Education

Olga Kosheleva

***Abstract*— In education, evaluations of the student's knowledge, skills, and abilities are often subjective. Teachers and experts usually make these evaluations by using words from natural language like "good", "excellent". Traditionally, in order to be able to process the evaluation results, these evaluations are first transformed into exact numbers. This transformation, however, ignores the fuzziness of the original estimates. To get a more adequate picture of the education process and education results, it is therefore desirable to transform these evaluations into intervals – or, more generally, fuzzy numbers. We show that this more adequate transformation can help on many important stages of the education process: planning education, teaching itself, and assessing the education results.**

I. Introduction

In education, evaluations of the student's knowledge, skills, and abilities are often subjective. Teachers and experts often make these evaluations by using words from natural language like "good", "excellent". Traditionally, in order to be able to process the evaluation results, these evaluations are first transformed into exact numbers. This transformation, however, ignores the fuzziness of the original estimates. To get a more adequate picture of the education process and education results, it is therefore desirable to transform these evaluations into intervals – or, more generally, fuzzy numbers.

We show that this more adequate transformation, pioneered in [2], can help on all the stages of the education process: in planning education, in teaching itself, and in assessing the education results.

Specifically, in planning education and in teaching itself, fuzzy techniques help us:

- better plan the order in which the material is presented and the amount of time allocated for each topic;
- fuzzy techniques help us find the most efficient way of teaching inter-disciplinary topics;
- these techniques also help to stimulate students by explaining historical (usually informal) motivations – often paradox-related motivations – behind different concepts and ideas of mathematics and science.

In assessment, fuzzy techniques help:

- to design a better grading scheme for test and assignments, a scheme that stimulates more effective learning,
- to provide a more adequate individual grading of contributions to group projects – by taking into account subjective estimates of different student distributions (and the uncertainty of these estimates), and
- to provide a more adequate description of the student knowledge and of the overall teaching effectiveness.

Olga Kosheleva is with the Department of Teacher Education, University of Texas at El Paso, El Paso, Texas 79968, USA (email: olgak@utep.edu).
The author is thankful to the anonymous referees for valuable suggestions.

This paper summarized, combines, and expands on the ideas and results, some of which published in proceedings of several fuzzy conferences. These published papers also contain additional technical details and practical examples of using these ideas.

II. Use of Fuzzy Techniques in Planning Education and in Teaching Itself

A. Planning the Order in Which the Material is Presented and the Amount of Time Allocated for Each Topic

Many empirical studies have shown that a change in the order in which different parts of the material are presented often drastically changes the learning efficiency. This is not only about using common sense: sometimes, the empirical results are counter-intuitive. For example, it is usually assumed that most students learn mathematical concepts better if they are first presented with concrete examples of these concepts, and they only learn abstract ideas later on. However, it turns out that empirically, the abstract-first approach for presenting the material often enhances learning; see, e.g., [5], [6], [17], [27]. In this section, we explain why a change in presentation order can drastically change the efficiency of learning, and we show how this explanation can be used to enhance the student learning.

To facilitate reasoning about learning, let us start with a simple geometric representation of learning. The process of learning means that we change the state of a student from a state s_0 in which the student did not know the material (or does not have the required skill) to a state in which the student has (some) knowledge of the required material (or has the required skill).

Let S denote the set of all the states corresponding to the required knowledge or skill. We start with a state which is not in the set S ($s_0 \notin S$), and we end up in a state s which is in the set S. On the set of all possible states, it is natural to define a metric $d(s, s')$ as the difficulty (time, effort, etc.) needed to go from state s to state s'. Our objective is to help the students learn in the easiest (fastest, etc.) way. In terms of the metric d, this means that we want to go from the original state $s_0 \notin S$ to the state $s \in S$ for which the effort $d(s_0, s)$ is the smallest possible.

In geometric terms, the smallest possible effort means the shortest possible distance. Thus, our objective is to find the state $s \in S$ which is the closest to s_0. Such closest state is called the *projection* of the original state s_0 on the set S.

The above geometric description of learning as a transition from the original state s_0 to its projection on the desired set S describes learning *as a whole*. Our objective is to find out which order of presenting information is the best. Thus, our objective is to analyze the *process* of learning, i.e., learning as a multi-stage phenomenon. For this analysis, we must explicitly take into account that the material to be learned consists of several pieces.

Let S_i, $1 \le i \le n$, denote the set of states in which a student has learned the i-th part of the material. Our ultimate objective is to make sure that the student learns all the parts of the material. In terms of states, learning the i-th part of the material means belonging to the set S_i. Thus, in terms of states, our objective means that the student should end up in a state which belongs to all the sets $S_1, \ldots, S_n$ – i.e., in other words, in a state which belongs to the intersection $S \stackrel{\text{def}}{=} S_1 \cap \ldots \cap S_n$ of the corresponding sets S_i.

In these terms, if we present the material in the order S_1, S_2, ..., S_n, this means that:

- we first project the original state s_0 onto the set S_1, resulting is a state $s_1 \in S_1$ which is the closest to s_0;
- then, we project the state s_1 onto the set S_2, resulting is a state $s_2 \in S_2$ which is the closest to s_1;
- ...
- at the last stage of the cycle, we project the state s_{n-1} onto the set S_n, resulting is a state $s_n \in S_n$ which is the closest to s_{n-1}.

In some cases, we end up learning all the material – i.e., in a state $s_n \in S_1 \cap \ldots \cap S_n$. However, often, by the time the students have learned S_n, they have somewhat forgotten the material that they learned in the beginning. So, it is necessary to repeat this material again (and again). Thus, starting from the state s_n, we again sequentially project onto the sets S_1, S_2, etc.

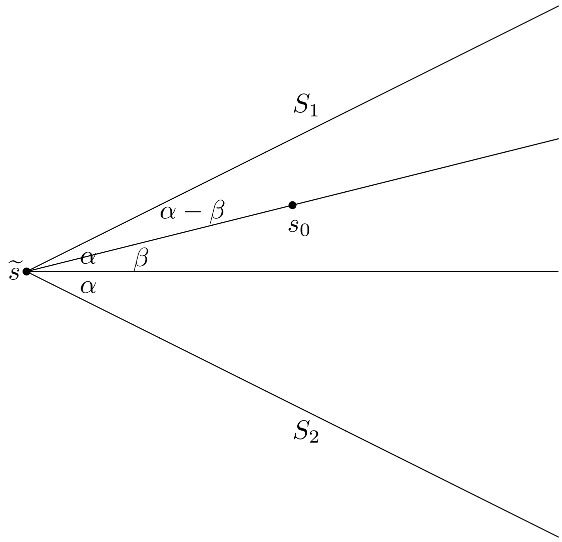
Let us start with the simplest case when knowledge consists of two parts. In this simplest case, there are only two options: The first option is that we begin by studying S_1; then, we study S_2, then, if needed, we study S_1 again, etc. The second option is that we begin by studying S_2; then, we study S_1, then, if needed, we study S_2 again, etc. We want to get from the original state s_0 to the state $\widetilde{s} \in S_1 \cap S_2$ which is the closest to s_0. The effectiveness of learning is determined by how close we get to the desired set $S = S_2 \cap S_2$ in a given number of iterations.

In the case of two-part knowledge, it is natural to conclude that the amount of this knowledge is reasonably small – otherwise, we would have divided into a larger number of easier-to-learn pieces.

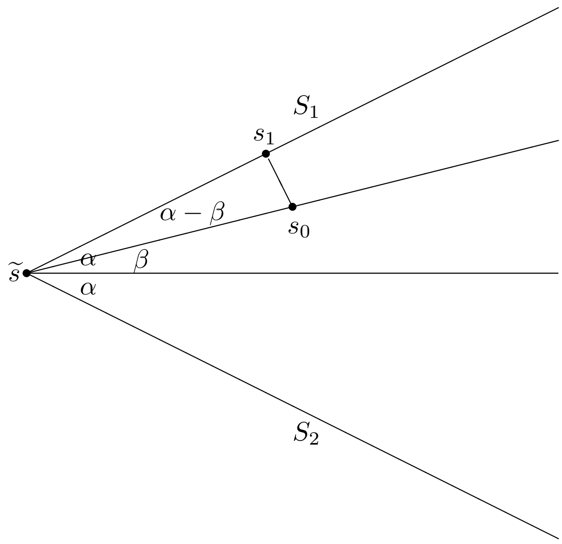
In geometric terms, this means that the original state s_0 is close to the desired intersection set $S_1 \cap S_2$, i.e., that the distance $d_0 \stackrel{\text{def}}{=} d(s_0, \widetilde{s})$ is reasonably small.

Since all the states are close to each other, in the vicinity of the state $\widetilde{s}$, we can therefore expand the formulas describing the borders of the sets S_i into Taylor series and keep only terms which are linear in the (coordinates of the) difference $s - \widetilde{s}$. Thus, it is reasonable to assume that the border of each of the two sets S_i is described by a linear equation – and is hence a (hyper-)plane: a line in 2-D space, a plane in 3-D space, etc.

As a result, we arrive at the following configuration. Let 2α denote the angle between the borders of the sets S_1 and S_2, so that the angles between each of these borders and the midline is exactly α. Let β denote the angle between the direction from $\widetilde{s}$ to s_0 and the midline. In this case, the angle between the border of S_1 and the midline is equal to $\alpha - \beta$. So, we arrive at the following configuration:



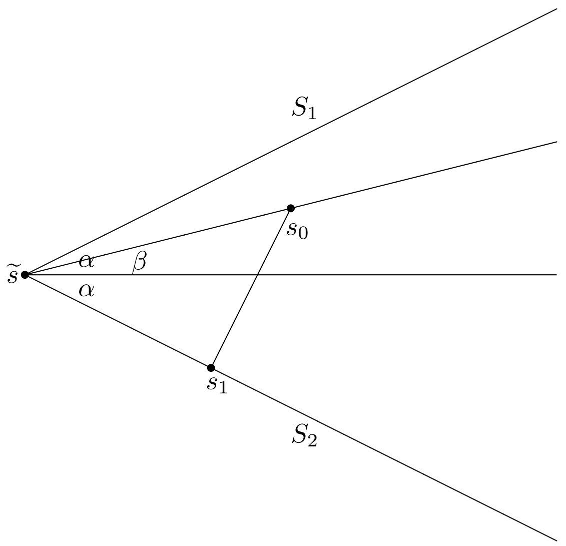
In the first option, we first project s_0 onto the set S_1. As a result, we get the following configuration:



Here, the projection line $s_0 s_1$ is orthogonal to the border of S_1. From the right triangle $\triangle\, \widetilde{s} s_0 s_1$, we therefore conclude that the distance $d_1 \stackrel{\text{def}}{=} d(\widetilde{s}, s_1)$ from the projection point s_1 to the desired point $\widetilde{s}$ is equal to $d_1 = d_0 \cdot \cos(\alpha - \beta)$. On the next step, we project the point s_1 from S_1 onto the line S_2 which is located at the angle 2α from S_1. Thus, for the projection result s_2, we will have $d_2 = d(s_2, \widetilde{s}) = d_1 \cdot \cos(2\alpha) = d_0 \cdot \cos(\alpha - \beta) \cdot \cos(2\alpha)$. After this, we may again project onto S_2, then again project onto S_1, etc. For each of these projections, the angle is equal to 2α, so after each of them, the distance from the desired point $\widetilde{s}$ is multiplied the same factor $\cos(2\alpha)$.

As a result, after k projection steps, we get a point s_k at a distance $d_k = d(s_k, \widetilde{s}) = d_0 \cdot \cos(\alpha - \beta) \cdot \cos^{k-1}(2\alpha)$ from the desired state $\widetilde{s}$.

In the second option, we start with teaching S_2, i.e., if we first project the state s_0 into the set S_2. In this option, we get the following configuration:



Here, we have $d_k = d(s_k, \widetilde{s}) = d_0 \cdot \cos(\alpha+\beta) \cdot \cos^{k-1}(2\alpha)$. Since, in general, $\cos(\alpha - \beta) \neq \cos(\alpha + \beta)$, we can see that a change in the presentation order can indeed drastically change the success of the learning procedure. Thus, our simple geometric model explains why the effectiveness of learning depends on the order in which the material is presented.

In general, starting with S_1 leads to a more effective learning than starting with S_2 if and only if $d_0 \cdot \cos(\alpha - \beta) \cdot \cos^{k-1}(2\alpha) < d_0 \cdot \cos(\alpha+\beta) \cdot \cos^{k-1}(2\alpha)$. As shown in [10], [11], this means that to make learning more efficient, we should start with studying the material which is further away from the current state of knowledge. In other words, we should start with a material that we know the least.

This ties in nicely with a natural commonsense recommendation that to perfect oneself, one should concentrate on one's deficiencies. This recommendation is also in a very good accordance with the seemingly counter-intuitive conclusion from [5], [6], [17], [27], that studying more difficult (abstract) ideas first enhances learning.

In general case, as shown in [10], [11], the problem of selecting an optimal path is computationally complex (NP-hard). So, to solve such problem, we need to use expert knowledge – and fuzzy techniques are known to be among the most successful techniques for handling such knowledge.

Comment. Other applications of fuzzy to selecting the order to the material are described in [1], [8], [14].

B. Finding the Most Efficient Way of Cross-Disciplinary Teaching

For the collaboration between researchers from different disciplines (even different sub-disciplines) to be successful, we need to *educate* collaborating researchers in the basics of each others' disciplines.

Let us give an example of a typical problem that we encountered. Suppose that a computer science has an interesting idea on how to better organize the geosciences' data and/or metadata. Alternatively, a computer scientist may simply want to teach, to a geosciences colleague, a few existing computer science ideas on how to organize data and/or metadata. How to convey a computer science idea to a geoscientist?

One possibility is simply to describe this idea in Computer Science terms. Alas, many of these terms are usually very specific. Even many computer scientists – those whose research is unrelated to cyberinfrastructure – are not very familiar with these terms and with the ideas behind them.

The only serious way for a geoscientist to understand and learn these terms, notions, ideas is to learn the material of several relevant computer science courses – i.e., in effect, to get a second degree in computer science. A few heroes may end up doing this, but it is unrealistic to expect such deep immersion from a usual student.

Alternatively, to make it clearer, a computer scientist can try to explain his or her ideas on the example of a toy geosciences problem.

The limitation of this approach is that the computer scientist is usually not a specialist in the domain science (in our case, in geosciences). As a result, his or her description of the toy problem is, inevitably, flawed: e.g., oversimplified. Hence, the problem that the new idea is trying to solve in this example is often not meaningful to a geoscientist – and since the motivation is missing, it is difficult to understand the idea.

The above problems may sound unsolvable if we restrict ourselves to a specific domain science. However, the very fact that there are already examples of successful inter-disciplinary education, shows that these communication problems are solvable.

It turns out that these outside examples themselves helped us to solve our communication problem.

Instead of trying to describe his ideas in purely computer science terms or on a toy geosciences example, a computer scientist described these ideas on the example of his applying similar ideas to a complete different area: solar astronomy.

This description was inevitably less technical – since none of us is a specialist in solar astronomy – and therefore, much more understandable. As a result, we got a much better understanding of the original computer science idea.

After we found this solution, we started thinking why it worked – and discovered an explanation – via the matter-of-degree ideology.

Every person has different degrees of knowledge in different areas. There are many potential ways to measure these degrees. A natural way is to gauge the degree of expertise the way we gauge the student's knowledge: by counting the proportion of correct answers on some test describing the knowledge. In this case, the degree is a number between 0 and 1, with 0 representing no knowledge at all and 1 meaning perfect knowledge.

Let us consider a general case, when the describer translates his or her idea (or problem) into a domain in which he or she has a degree of expertise d_1 and the respondent has a degree of expertise d_2.

In general, there are two problems that prevent us from perfect understanding:

- first, the describer's level may be too low, so his or her presentation has a lot of inaccuracies that prevent understanding;
- second, the describer's level may be too high, so his or her presentation may be too sophisticated for the responder to understand – which also prevents understanding.

Because of these two possible problems, let us consider two subcases corresponding to the above two situations:

- when the new domain is closer to the describer's area of expertise, i.e., when $d_2 \le d_1$, and
- when the new domain is closer to the respondent's area of expertise, i.e., when $d_1 \le d_2$.

In the case when $d_2 \le d_1$, the describer's degree of sophistication in the new domain is higher than the respondent's, so the respondent will not be able to detect inaccuracies in the describer's presentation. The only problem here is that since the describer's level of sophistication d_1 may be higher than the respondent's level d_2, the corresponding part of the presentation will not be clear to the respondent.

From all the knowledge corresponding to the levels of sophistication from 0 to d_1, only the parts corresponding to levels from 0 to $d_2 \le d_1$ will be properly understood. We have argued above that the knowledge is more or less uniformly distributed across different levels of sophistication. Out of d_1 different levels, only d_2 levels lead to understanding.

As a result, the proportion d of properly understood message is approximately equal to the ratio d_2/d_1.

In the case when $d_1 \le d_2$, the describer's degree of sophistication in the new domain is smaller than the respondent's, so the respondent will be able to understand all the terms that the describer is using. However, because of the possible difference of the levels of expertise, the respondent will be able to detect inaccuracies in all the levels of sophistication beyond d_1.

Thus, from all the knowledge corresponding to the levels of sophistication from 0 to d_2, only the parts corresponding to levels from 0 to $d_1 \le d_2$ will be properly understood. We have argued above that the knowledge is more or less uniformly distributed across different levels of sophistication. Out of d_2 different levels on which the recipient receives information, only d_1 levels lead to understanding.

As a result, the proportion d of property understood message is approximately equal to the ratio d_2/d_1.

In both cases, the degree of understanding d can be obtained by dividing the smallest of the degrees d_1 and d_2 by the largest of these two degrees:

$$d = \frac{\min(d_1, d_2)}{\max(d_1, d_2)}.$$

This explain the success of our empirical strategy. Indeed, when the describer formulates his or message either in his or her own domain terms or in terms of the respondent's domain, we have $\min(d_1, d_2) \approx 0$ and $\max(d_1, d_2) \approx 1$, so $d \approx 0$.

When the describer instead formulates his or her own message in the language of the third domain, in which $d_1 \approx d_2$, we have $\min(d_1, d_2) \approx \max(d_1, d_2)$ and therefore, $d \approx 1$. For details, see [24], [25].

C. Stimulating Students by Explaining Motivations Behind Concepts and Ideas

In many cases, when the new material starts, the students do not receive a convincing explanation of why this material is important. This observation helps to explain why the existing pedagogical techniques are often not as successful as predicted: no matter how interested and active the students are, they do not reach their full potential simply because they do not fully understand the need to study this particular material.

In view of this observation, a key to success is for a teacher to make sure that the students understand this need – before explaining the material itself.

In many school disciplines like English (or any other native language), the need to communicate clearly may not be always well understood by the students, but it is usually well understand by their parents and the population as a whole.

In mathematics, the situation with understandability is much more problematic. Not only the students do not understand the need for mathematics education; in general, most people do not understand the significance and importance of mathematics.

Most people see mathematics as a collection of useless rules and formulas that they need to memorize to get the correct answer to the textbook problems, rules and formulas that, in their opinion, have no use in real life.

Since students do not understand the need for mathematics education, they feel forced to study it and, as a result, many students develop strong negative attitude towards mathematics ("I hate math").

Of course, mathematics is not useless, mathematics is a foundation of science and engineering and, as a result, it lies in the foundation of technological progress. Thus, a natural way to enhance the understanding of why mathematics is important is to make sure that from the very beginning, students do not simply learn the abstract rules and formulas, they are also given numerous examples of *practical* applications of these rules and formulas. As a result, student may (somewhat grudgingly) accept that formulas are useful, but not what is the essence of mathematics for mathematicians: proofs.

From this viewpoint, what we need to enhance the students' understanding of mathematics is help them understand that not only engineering formulas are useful, rigorous mathematical proofs have their value. How can we do that? Good news is that we do not have to invent new reasons: after all, there are reasons why rigorous mathematics was designed in the first place, and why it remains an important (and well-supported) part of science and, more generally, of our quest for knowledge. What we need to do is convincingly convey these reasons to students of mathematics.

History of mathematics (see, e.g., [3]) shows that the main reason why mathematicians started making their methods more rigorous is that heuristic methods sometimes lead to seeming contradictions (*paradoxes*).

Formal approach to mathematics was first developed by the ancient Greeks, and their paradoxes explain why they felt this need for rigor. Probably the first known paradox is the *heap* paradox:

- one little rock does not form a heap;
- if we have a pile of rocks which is not a heap, and add one more rock, then we still do not get a heap;
- by induction, we can thus conclude that no matter how many rocks we add, we will never get a heap;
- on the other hand, everyone knows that heaps are possible.

This paradox can be reformulated in terms of a crowd (one person does not form a crowd, etc.). This paradox, well known to the Greeks, explained the need to restrict ourselves to well-defined notions – in contrast to vaguely defined ("fuzzy") notion such as *heap* or *crowd*.

It is interesting to mention that this same paradox was revived in the 20th century by computer scientists who were interested in making computers understand our commonsense words and terms. To describe such "fuzzy" words in precise terms – understandable to a computer – L. A. Zadeh invented a methodology called *fuzzy logic*; see, e.g., [7], [21], [29]. In fuzzy logic, in contrast to the traditional two-valued logic, notions like "is a heap" or "is a crowd" are not necessarily either true or false. In many practical situations, expert may conclude that to some degree, we have a crowd. When one more person enters, this degree of "crowdedness" increases – until this degree attains its largest possible value 1, meaning that all reasonable persons will conclude that this particular collection of people is a crowd.

This solution of a paradox enabled researchers to successfully use informal (seemingly inconsistent) expert knowledge in the design of efficient automated systems for control, data processing, decision making, etc.; see, e.g. [7], [21].

Fuzzy logic was the first example of *soft computing*, attempts to formalize parts of commonsense reasoning that were "left behind" when mathematics became more rigorous. Fuzzy logic and other soft computing techniques are used in many areas of science and engineering; see, e.g., [7], [21]. In particular, fuzzy logic can help explain many other paradoxes; see, e.g., [7], [19].

Calculus led to the discovery of several new paradoxes. An example of such a paradox was related to summation of infinite series. Before calculus, there were very few attempts to build a converging sequence or to add up an infinite series. Calculus provided a general framework within which such a summation became routine. Most functions like $\exp(x)$, $\sin(x)$, $\cos(x)$, where shown to be representable as Taylor series – and this provided a very efficient way of computing these functions, so efficient, that even nowadays, Taylor series are the main techniques used by computers to compute the values of these functions.

For many series, calculus-based ideas led to successful formulas for their sum. However, for some series, the same ideas led to paradoxes. Probably the most well known of these paradoxes was discovered by Euler who tried to compute the sum s of the infinite series

$$1 + (-1) + 1 + (-1) + 1 + \ldots$$

On the one hand, we can combine elements into pairs, and end up with

$$s = (1 + (-1)) + (1 + (-1)) + \ldots = 0 + 0 + \ldots = 0.$$

On the other hand, we can keep the first element in this infinite sum intact and combine elements starting with the second one. Then, we get

$$s = 1+((-1)+1)+((-1)+1)+\ldots = 1+0+0+\ldots = 1 \neq 0.$$

This (and similar) paradoxes has led to the need to revisit the foundations of calculus, to introduce rigor. This "revolution of rigor", largely associated with the names of Cauchy and Weierstrass, introduced the modern "epsilon–delta" definitions into calculus. For example, according to Cauchy, an infinite series $a_1 + a_2 + \ldots$ has a sum s if and only if for every $\varepsilon > 0$, there exists an N such that for all $n \geq N$, we have $|(a_1 + \ldots + a_n) - s| \leq \varepsilon$. Euler's paradox is then resolved because according to this definition, the series $1 + (-1) + 1 + (-1) + \ldots$ does not have a sum at all.

The "epsilon–delta" definitions are not easy – but they are needed because without rigorous definitions, we can get paradoxes. Students taking calculus, however, often learn these definitions without understanding why this complexity is needed – and thus, sometimes have trouble understanding the related notions.

Similar paradoxes occur not only in pure mathematics, they also occur in applications of mathematics to physics and other disciplines. The most well known example is quantum mechanics – and its mathematical counterpart, Hilbert spaces and operators in Hilbert spaces. Many counter-intuitive properties of quantum mechanics come, crudely speaking, from the fact that the same elementary particle (e.g., photons, electrons) can exhibit seemingly inconsistent properties. For example, the same particle can sometimes behave as a point particle, with no spatial distribution; on the other hand, sometimes, it can behave as a wave, a spatially distributed process.

It is interesting that in the last two decades, it turned out that these seemingly paradoxical properties of quantum objects can be used to drastically speed up computations and provide additional communication security; see, e.g., [22].

In short, what we propose is to teach student motivations – and that includes paradoxes, for which fuzzy logic is often a reasonable way to formalize; for details, see [9], [13].

III. Use of Fuzzy Techniques in Assessment

A. Towards Designing a Better Grading Scheme for Tests and Assignments

The overall grade for a class is formed by adding the grades for different assignments and tests. Similarly, the grade for a test is formed by adding the grades earned on each of the problems. (In the text of the test, it is usually described how many points each problem is worth.)

An appropriate allocation of points is very important. For example, if we allocate almost all the points to the final exam, some students will see no reason to study hard during the semester. They will try to cram the material during the finals, and as a result, even if they pass the finals, their knowledge of the material will not be as good as the knowledge of the students who studied diligently during the semester.

On the other hand, if we allocate too few points to the final exam, then some students will have no incentive to review the class material for the final exam. As a result, their knowledge of the material that was studied in the beginning of the semester will be not as good as the knowledge of those students who did review this material at the end of the class.

At present, the points for different assignments, tests, and problems are allocated based on the teacher's subjective experience. As a result, there is a high variety of point allocation. Since it is very important to properly allocate points for different assignments, tests, and problems, it is desirable to find a more objective ways for such allocation.

Our main idea is that most classes are prerequisites for other classes (or for some qualification exams). For example, Pre-calculus is a prerequisite for Calculus I, Calculus I is a prerequisite for Calculus II, etc. For such classes, it is desirable to allocate the points in such a way that a success in this class will be a good indication of the success in the next class.

If we grade too easily, students may be happier with their good grades, but some of them will pass the class without acquiring the knowledge needed for the next class – and so they may fail this next class. On the other hand, if we grade too harsh, we unnecessarily fail many students who may have not learned all the details but whole knowledge is actually good enough to successfully pass the next class.

From this viewpoint, the best way to allocate points is to select the allocations for which the resulting grade is the best predictor for the grade in the next class.

To find the best allocation of points, we must use the grades that the students got for different assignments and tests, and the grades they got in the next class.

Let us denote the total number of assignments and tests by T, and let us denote the total number of students who took this class in the past by S. Let us denote the grade of student s ($1 \le s \le S$) on assignment or test t ($1 \le t \le T$) by g_{st}. The grade of student s at the next class (for which this class is a prerequisite) will be denoted by n_s.

We want to predict the value n_s based on the grades $g_{s1}, \ldots, g_{sT}$. For such a prediction, it is natural to start with a linear regression $n_s \approx a_0 + a_1 \cdot g_{s1} + \ldots + a_T \cdot g_{sT}$. The coefficients a_t can be found from the Least Squares method, by minimizing the sum $\sum\limits_{s=1}^{S} (a_0 + a_1 \cdot g_{s1} + \ldots + a_T \cdot g_{sT} - n_s)^2$. Details are described in [16], [23].

B. Towards a More Adequate Individual Grading of Contributions to Group Projects

How can we estimate individual contributions to a group project? This problem is important in education when several students work together on a project, it is important in the business environment when several people work together on a joint project.

In all such situations, we need to know the relative contributions $E_1, \ldots, E_n$ of all n participants, relative in the sense that they represent the fraction of the overall credit – and thus, the sum of these contributions should be equal to 1: $\sum\limits_{i=1}^{n} E_i = 1$.

In many practical situation, the only available information for estimating contributions consists of the estimates that different participants give to each other's contribution. In this case, we have n^2 values e_{ij} ($1 \le i, j \le n$) – estimates made by the i-th participate of the contribution of the j-th participant.

In the ideal case when all estimates are unbiased, for each participant j, we have n estimates $e_{1j}, \ldots, e_{nj}$ for the desired value E_j. In this case, we have n approximate equalities to find E_j: $E_j \approx e_{1j}$, …, $E_j \approx e_{nj}$. To find a reasonable estimate E_j from these approximate equalities, a natural idea is to use the Least Squares technique and find the value E_j for which the sum $\sum\limits_{i=1}^{n} (E_j - e_{ij})^2$ is the smallest possible. This is a textbook use of the Least Squares method, to combine several estimates of the same quantity, and the solution to this optimization problem is well known – it is the arithmetic average of these estimates: $E_j = \dfrac{e_{1j} + \ldots + e_{nj}}{n}$.

In practice, estimates of others' contributions are often unbiased, but it is very difficult to get an unbiased estimate of one's own contribution in comparison with the contributions of others. In other words, while the estimates e_{ij} for $i \ne j$ are really unbiased, the estimates e_{ii} are too subjective and biased to be useful. Because of the bias, we can say that the only available information about each value E_j consists of estimates e_{ij} with $i \ne j$. These estimates do not necessarily add up to 1, so in the ideal case, we have $E_i \approx s_i \cdot e_{ij}$ for some normalization constant s_i.

The Least Squares approach now means that we find both E_j and s_i by minimizing the sum of the squares of the discrepancies $\sum\limits_{i \ne j} (E_j - e_{ij} \cdot s_i)^2$ under the constraint $\sum\limits_{i=1}^{n} E_i = 1$.

As shown in [26], the following algorithm solves this problem: First, we take $e_{ii} \stackrel{\text{def}}{=} 0$, then compute $c_i \stackrel{\text{def}}{=} \sum\limits_{m=1}^{n} e_{im}^2$ and $c_{j,\ell} \stackrel{\text{def}}{=} \sum\limits_{i=1}^{n} \frac{e_{ij} \cdot e_{i\ell}}{c_i}$, and solve the following system of linear equations $(n-1) \cdot e_j - \sum\limits_{\ell=1}^{n} c_{j,\ell} \cdot e_\ell = 1$. After this, we compute the sum $S = \sum\limits_{i=1}^{n} e_i$ and then, the desired estimates as $E_j = \frac{e_j}{S}$.

C. Towards a More Adequate Description of Student Knowledge and of Teaching Effectiveness

Sometimes, the efficiency of a class is assessed by assessing the amount of knowledge that the students have after taking this class. However, this amount depends not only on the quality of the class, but also on how prepared were the students when they started taking this class. A more adequate assessment should therefore be *value-added*, estimating the added value that the class brought to the students.

In pedagogical practice, there are many value-added assessment models. However, most existing models have two limitations. First, they model the effect of the class as an additive factor independent on the initial knowledge. In reality, the amount of knowledge learned depends on the amount of the initial knowledge. Second, the existing models are statistical, they implicitly assume that the assessment values are objective – and are subject to random measurement errors and noises. In reality, many assessment values are subjective. Thus, fuzzy techniques provide, in our opinion, a more adequate way of processing these values.

In this section, we describe how the use of fuzzy techniques can help us overcome both limitations of the existing value-added assessments.

Value-added assessment describes how the post-test result y depends on the pre-test result x. In the traditional approach we, in effect, assume that the post-test result y is obtained from the pre-test result x by adding a certain amount a of new knowledge (and new skills): $y \approx x+a$. Here, we say that y is only approximately equal to $x+a$, to take into account measurement errors, random fluctuations, and the effect of factors that we do not take into account in this simple model.

As we have mentioned, the actual dependence of the post-test value y on the pre-test value x is more complex, because the difference $y-x$ changes with x. To describe this dependence, we therefore need to use more general formulas than $y = x+a$. The natural next approximation is to use the general linear dependence of the post-test value y on the pre-test value x: $y \approx m \cdot x + a$. We can thus use, e.g., the Least Squares method to determine m and a based on the pr-test and post-test grades.

In the previous section, we assumed that we know the numerical grade on the exam represents an exact measure of the student knowledge. In practice, however, the number grades are reasonably subjective.

Usually, instructors allocate certain number of points to different questions and problems on the test and to different aspects of the same question or problem. As a result, when the answer to each of the problems or questions is either absolutely correct or absolutely wrong (or missing), the resulting grade is uniquely determined. The subjectivity comes when the answer is partly correct, and we need to decide how much partial credit this answer deserves. Some such situations can be described from the very beginning, but often, it is not practically possible to foresee all possible mistakes and thus, to decide how much partial credit the student deserves.

Often, when two instructors co-teach a class or teach two different sections of the same class, their grades for similar mistakes can slightly differ – because of the slightly different allocation of partial credit. Even the same instructor, when grading two different student papers with similar mistakes, can sometimes assign two slightly different numerical grades.

As a result of this subjectivity, the numerical grade given to the test is not an exact measure of the student knowledge – because other instructors may assign a slightly different number grade to the same test results.

This subjectivity is well understood by instructors. This is one of the reasons why student transcripts usually list not the exact overall number grades, but rather the letter grades.

For example, usually, a letter grade A is assigned to all the numerical grades from 90 to 100, and a letter grade B is assigned to all numerical grades between 80 and 89. This assignment is in good accordance with the fact that while the difference between, say, 85 and 95 is meaningful and most probably not subjective, and a student with a grade of 95 has a higher knowledge level than a student with a grade of 85, the difference between, say 92 and 93 can be caused by the subjective reasons – and thus, a student with a grade of 93 does not necessarily know the material better than a student with a grade of 92.

The traditional letter grades may provide too crude a picture. In many cases, the distinction between, say, low 90s and high 90s also makes sense. To emphasize such a difference, some schools, in addition to usual letter grades, also use signed letter-type grades like A$-$ or B$+$. Letter grades from the resulting set correspond to intervals which are narrower than the width-10 intervals describing the usual letter grades.

Because the distinction within each interval may be caused by the subjectivity of an individual instructor grading, it makes sense, when describing how well the students learned, to use not the original numerical grades x_i, but rather the corresponding letter grades – i.e., in other words, the intervals $\mathbf{x}_i = [\underline{x}_i, \overline{x}_i]$ that describe possible values of the student knowledge. Based on these values, we want to predict the endpoints of the interval $\mathbf{y}_i = [\underline{y}_i, \overline{y}_i]$. We consider the case of a linear dependence, so we assume that both dependencies are linear: $\underline{y}_i \approx \underline{m} \cdot \underline{x}_i + \underline{a}$ and $\overline{y}_i \approx \overline{m} \cdot \overline{x}_i + \overline{a}$ (in [28], we show that the values $\underline{y}_i$ depend only on $\underline{x}_i$ and not on $\overline{x}_i$, and similarly for $\overline{y}_i$). To find the values $\underline{m}$, $\overline{m}$, $\underline{a}$, and $\overline{a}$, we can now use the Least Squares method.

In reality, the bounds that we know are "fuzzy", i.e., they contain x only with some degree of confidence α. Usually, we have different intervals $[\underline{x}(\alpha), \overline{x}(\alpha)]$ corresponding to different degrees α. The narrower the interval, the less confident we are that x belongs to this interval. Thus, we have a *nested family* of intervals corresponding to different values α.

Alternatively, for each possible value of x, we describe the degree $\mu_(x)$ to which this value is possible. For each degree of certainty α, we can determine the set of values of x_i that are possible with at least this degree of certainty – the *α-cut* $\mathbf{x}(\alpha) = \{x \mid \mu(x) \geq \alpha\}$ of the original fuzzy set.

Vice versa, if we know α-cuts for every α, then, for each object x, we can determine the degree of possibility that x belongs to the original fuzzy set [4], [7], [18], [20], [21].

A fuzzy set can be thus viewed as a nested family of its (interval) α-cuts.

Thus, we have fuzzy numbers X_i describing the pre-test grades and fuzzy numbers Y_i describing the post-test grades. We would like to describe the corresponding dependence of the post-test grade Y on a pre-test grade X. For each α, we can take the α-cuts $\mathbf{x}_i(\alpha)$ and $\mathbf{y}_i(\alpha)$ of the corresponding fuzzy numbers X_i and Y_i.

For each α, based on these intervals, we can now use the above least Squares method to find the interval-values linear function $[\underline{m}(\alpha) \cdot x + \underline{a}(\alpha), \overline{m}(\alpha) \cdot x + \overline{a}(\alpha)]$ corresponding to this α.

The list of all these interval-values linear functions corresponding to different values α forms the desired description of the dependence of the post-test grade y on the pre-test grade x – i.e., the desired value-added teacher assessment.

Detailed algorithms are given in [28].

CONCLUSIONS

In this paper, we show that fuzzy techniques can be helpful on all the stages of the education process, from planning the order of the material, to designing a proper way of teaching this material to different audiences, to evaluating the results of the teaching process.

REFERENCES

[1] R. Aló and O.M. Kosheleva, "Optimization techniques under uncertain criteria, and their possible use in computerized education", *Proc. NAFIPS'2006*, Montreal, Canada, June 3–6, 2006.

[2] R. Biswas, "An application of fuzzy sets in students' evaluation", *Fuzzy Sets and Systems*, vol. 74, no. 2, pp. 187–194, 1995.

[3] C.B. Boyer and U.C. Merzbach, *A History of Mathematics*, Dover, New York, 1991.

[4] D. Dubois and H. Prade, "Operations on fuzzy numbers", *International Journal of Systems Science*, vol. 9, pp. 613–626, 1978.

[5] J.A. Kaminski, V.M. Sloutsky, and A.F. Heckler, "Do children need concrete instantiations to learn an abstract concept?", In: R. Sun and N. Miyake (Eds.), *Proc. 27th Annual Conference of the Cognitive Science Society, Vancouver, BC, July 26–29, 2006*, pp. 411–416, Lawrence Erlbaum, Mahwah, NJ, 2006.

[6] J.A. Kaminski, V.M. Sloutsky, and A.F. Heckler, "The advantage of abstract examples in learning math", *Science*, vol. 320, pp. 454–455, April 25, 2008.

[7] G.J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice-Hall, Upper Saddle River, New Jersey, 1995.

[8] O. Kosheleva, "Spiral curriculum: towards mathematical foundations", *Proceedings of the International Conference on Information Technology InTech'07*, Sydney, Australia, December 12–14, 2007, pp. 152–157, 2007.

[9] O. Kosheleva, "Potential application of fuzzy techniques to math education: emphasizing paradoxes as a (seemingly paradoxical) way to enhance the learning of (strict) mathematics", *Proc. NAFIPS'2008*, New York, New York, May 19–22, 2008.

[10] O. Kosheleva, "Early start can inhibit learning: a geometric explanation", *Proc. IFSA'09*, Lisbon, Portugal, July 20–24, 2009, pp. 438–442, 2009.

[11] O. Kosheleva, "Early start can inhibit learning: a geometric explanation", *Geombinatorics*, vol. 19, no. 3, pp. 108–118, 2010.

[12] O. Kosheleva, "What is wrong with teaching to the test: uncertainty techniques help in understanding (and hopefully resolving) the controversy", *Proc. NAFIPS'2011*, El Paso, Texas, March 18–20, 2011.

[13] O. Kosheleva, "Towards combining Freirean ideas and Russian experience in mathematics education", In: C.A. Rossatto, A. Luykx, and H.S. Garcia (eds.), *Teaching for Global Community*, Information Age Publishing, Charlotte, North Carolina (to appear).

[14] O.M. Kosheleva and R. Aló, "Towards economics of education: optimization under uncertainty", *Proc. 2nd Int'l Conf. on Fuzzy Sets and Soft Computing in Economics and Finance FSSCEF'2006*, St. Petersburg, Russia, June 28 – July 1, 2006, pp. 63–70, Nauka Publ., St. Petersburg, Russia, 2006.

[15] O.M. Kosheleva and M. Ceberio, "Processing educational data: from traditional statistical techniques to an appropriate combination of probabilistic, interval, and fuzzy approaches", *Proc. Int'l Conf. on Fuzzy Systems, Neural Networks, and Genetic Algorithms FNG'05, Tijuana, Mexico, October 13–14, 2005*, pp. 39–48, Tijuana Tech, Tijuana, Mexico, 2005.

[16] O. Kosheleva and V. Kreinovich, "How to assigns weights to different tests?", *Proc. Sun Conference on Teaching and Learning*, El Paso, Texas, February 27, 2009.

[17] L.M. Lesser and M.A. Tchoshanov, "Selecting representations", *Texas Mathematics Teacher*, vol. 53, no. 2, pp. 20–26, Fall 2006.

[18] R.E. Moore and W. Lodwick, "Interval analysis and fuzzy set theory", *Fuzzy Sets and Systems*, vol. 135, no. 1, pp. 5–9, 2003.

[19] H.T. Nguyen, O. Kosheleva, and V. Kreinovich, "Is the success of fuzzy logic really paradoxical? Or: Towards the actual logic behind expert systems", *International Journal of Intelligent Systems*, vol. 11, pp. 295–326, 1996.

[20] H.T. Nguyen and V. Kreinovich, "Nested intervals and sets: concepts, relations to fuzzy sets, and applications", In: R.B. Kearfott and V. Kreinovich, eds., *Applications of Interval Computations*, Kluwer, Dordrecht, pp. 245–290, 1996.

[21] H.T. Nguyen and E.A. Walker, *A First Course in Fuzzy Logic*, Boca Raton, Florida: Chapman & Hall/CRC, 2006.

[22] N.A. Nielsen and I.L. Chuang, *Quantum Information and Quantum Computation*, Cambridge University Press, Cambridge, Massachusetts, 2000.

[23] S. Niwitpong, M. Chiangpradit, and O. Kosheleva, "Towards optimal allocation of points to different assignments and tests", *Journal of Uncertain Systems*, vol. 4, no. 4, pp. 291–295, 2010.

[24] P. Pinheiro da Silva, A. Velasco, and O. Kosheleva, "Degree-based ideas and technique can facilitate inter-disciplinary collaboration and education", *Proc. NAFIPS'2010*, Toronto, Canada, July 12–14, 2010, pp. 388–393, 2010.

[25] P. Pinheiro da Silva, A. Velasco, O. Kosheleva, and V. Kreinovich, "How AI-type uncertainty ideas can improve inter-disciplinary collaboration and education: lessons from a case study", *Journal of Advanced Computational Intelligence and Intelligent Informatics JACIII*, vol. 14, no. 6, pp. 700–707, 2010.

[26] J. Sappakitkamjorn, W. Suriyakat, W. Surachardkiati, and O. Kosheleva, "How to estimate individual contributions to a group project", *Journal of Uncertain Systems*, vol. 4, no. 4, pp. 301–305, 2010.

[27] M.A. Tchoshanov, *Visual Mathematics*, ABAK Publ., Kazan, 1997 (in Russian).

[28] K. Villaverde and O. Kosheleva, "Towards more detailed value-added teacher assessments", *Proc. IEEE World Congress on Computational Intelligence WCCI'2010*, Barcelona, Spain, July 18–23, 2010, pp. 3126–3133.

[29] L.A. Zadeh, "Fuzzy sets", *Information and control*, vol. 8, pp. 338–353, 1965.