Scientific journal
PHYSICAL AND MATHEMATICAL EDUCATION
Has been issued since 2013.

ISSN 2413-158X (online)
ISSN 2413-1571 (print)

Науковий журнал
ФІЗИКО-МАТЕМАТИЧНА ОСВІТА
Видається з 2013.

http://fmo-journal.fizmatsspu.sumy.ua/

Воскоглу Майкл Гр. Управління нечіткими даними в освіті. Фізико-математична освіта. 2019. Випуск 1(19). С. 13-17.

Voskoglou Michael Gr. Management Of Fuzzy Data In Education. Physical and Mathematical Education. 2019. Issue 1(19). P. 13-17.

DOI 10.31110/2413-1571-2019-019-1-002

Michael Gr. Voskoglou
Graduate Technological Educational Institute of Western Greece, Patras, Greece
voskoglou@teipat.gr; mvosk@hol.gr
ORCID: 0000-0002-4727-0089

MANAGEMENT OF FUZZY DATA IN EDUCATION

ABSTRACT

Formulation of the problem. *Some years ago the unique tool in hands of the scientists for handling the situations of uncertainty that frequently appear in problems of science, technology and of the everyday life, used to be the theory of Probability. However, nowadays the theory of Fuzzy Sets initiated by Zadeh in 1965 and its extensions and generalizations followed in the recent years have given a new dynamic to this field.*

Materials and methods. *Mathematical methods of analysis are used.*

Results. *In the present work a model is developed for handling the fuzzy data appearing in the field of Education. The model is based on the calculation of the possibilities of the profiles involved in the corresponding situations, which, according to the British economist Schackle and many other researchers, are more suitable than the fuzzy probabilities for studying the human behaviour. A classroom application to learning mathematics is also presented illustrating the importance of the model in practice. The general model is extended for studying the combined results of the evaluation of fuzzy data obtained from two (or more) different sources and an example is provided to emphasize the usefulness of this extension for real situations in education.*

Conclusions. *The management and evaluation of the fuzzy data obtained by the operation mechanisms of large and complex systems is very important for real life and science applications. A developed model evaluates such kind of data in terms of the corresponding membership degrees and possibilities. The examples for the process of learning a subject-matter in the classroom and the example for a market's research illustrate the applicability and usefulness of the model to practical problems. The general character of the proposed model enables its application to a variety of other human and machine activities for a description of such kind of activities and this is one of main targets for future research.*

KEY WORDS: *Fuzzy Set (FS), Membership Degree, Possibility, Fuzzy data, Fuzzy Variable.*

INTRODUCTION

Problem formulation. Situations appear frequently in Education where many different and constantly changing factors are involved, the relationships among which are indeterminate. As a result the data obtained from the operation mechanisms of such situations cannot be easily determined precisely and in practice estimates of them are used.

While 50-60 years ago the unique tool in hands of the scientists for handling such kind of data, and situations of uncertainty in general, used to be the theory of *Probability*, nowadays the *Fuzzy Set (FS)* theory initiated by Zadeh in 1965 (Zadeh, 1965) and its extensions and generalizations that followed in the recent years (Voskoglou, 2019) have given a new dynamic to this field.

In the article at hands a model is developed for evaluating a system's fuzzy data in terms of the corresponding fuzzy possibilities. The rest of the article is organized as follows: In the second section the general model is developed and an application to learning mathematics is presented illustrating its applicability in education. In the third section the general model is extended for studying the combined results of the evaluation of fuzzy data obtained from two (or more) different sources and an example is provided to emphasize the usefulness of this extension for real situations in education.. The article closes with the final conclusions stated in the fourth section.

RESULTS AND DISCUSSION

The general model. The reader is considered to be familiar to the fundamentals of the FS theory and the book (Klir & Folger, 1988) is proposed as a general reference on the subject.

Assume that one wants to study an educational system's behavior consisting of n components, $n \geq 2$ (e.g. a class of n students), during a process involving vagueness and/or uncertainty (e.g. problem solving). Denote by S_i, i=1, 2, 3 the main steps of that process and by *a, b, c, d, e* the linguistic labels of very low, low, intermediate, high and very high success respectively of the system components in each step. Here, for reasons of simplicity, we have considered three steps only, but the model holds for any finite number of steps

Set *U = {a, b, c, d, e}*. Then a FS A_i in *U* will be associated to each step S_i, i = 1, 2, 3. For this, if n_{ia}, n_{ib}, n_{ic}, n_{id}, n_{ie} denote the numbers of the system components that faced very low, low, intermediate, high and very high success respectively at stage S_i, we define the **membership degree** $m_{A_i}(x)$ of each *x* in *U* by

$$m_{A_i}(x) = \frac{n_{ix}}{n} \quad (1)$$

Then the FS A_i in *U* associated to S_i is of the form:

$$A_i = \{(x, m_{A_i}(x)): x \in U\}, i=1, 2, 3 \quad (2)$$

In order to represent all possible **profiles (overall states)** of the system components during the corresponding process a **fuzzy relation**, say *R*, in U^3 (i.e. a FS in U^3) is considered of the form:

$$R= \{(s, m_R(s)): s=(x, y, z) \in U^3\} \quad (3)$$

Usually in practical applications the degree of success of each system's component in a certain step of the process depends upon the degree of its success in the previous step. Under this assumption and in order to define properly the membership function m_R, the following definition is given:

Definition: A profile *s=(x, y, z)*, with *x, y, z* in *U,* is said to be **well ordered** if *x* corresponds to a degree of success equal or greater than *y* and *y* corresponds to a degree of success equal or greater than *z.*

For example, *(c, c, a)* is a well ordered profile, while *(b, a, c)* is not. The membership degree of a well ordered profile *s* is defined now to be equal to the product

$$m_R(s) = m_{A_1}(x) \cdot m_{A_2}(y) \cdot m_{A_3}(z) \quad (4)$$

On the contrary, the degree of the profiles which are not well ordered is defined to be zero. In fact, if for example the profile *(b, a, c)* possessed a nonzero membership degree, then at least one of the system components demonstrating a very low performance at step S_2 would perform satisfactorily at the next step S_3, which is impossible to happen.

However, they are also real situations in Education in which the performance of each component at each step does not depend on its performance in the previous steps (e.g. see Example 2). In such cases the membership degrees of all profiles are defined by equation (4).

Next, for simplifying our notation, we shall write m_s instead of $m_R(s)$. Then the **fuzzy probability** p_s of the profile *s* is defined by

$$P_s = \frac{m_s}{\sum_{s \in U^3} m_s} \quad (5)$$

However, according to the British economist Shackle (Shackle, 1961) and many other researchers after him, the human behaviour can be better studied by using the possibilities rather of the several profiles, than their probabilities. The **possibility** r_s of the profile *s* is defined by

$$r_s = \frac{m_s}{\max\{m_s\}} \quad (6)$$

In equation (6) *max {m_s}* denotes the greatest value of m_s for all *s* in U^3. In other words the possibility of *s* expresses the "relative membership degree" of *s* with respect to *max {m_s}*.

The following application to the process of **learning** a subject matter in the classroom illustrates the applicability of the present model to real life situations:

Example 1: There is no doubt that learning is one of the fundamental components of the human cognitive action. There are very many different theories and models developed by psychologists, educators and other cognitive scientists for the description of the mechanisms of learning, Nevertheless, although the process of learning differs in details from person to person, it is in general accepted that it involves **representation** and **interpretation** of the input data in order to produce the new knowledge (step S_1), **generalization** of this knowledge to a variety of situations (step S_2) and **categorization** of the generalized knowledge by embodying it to the individual's appropriate cognitive structures, widely termed as **schemas of knowledge** (step S_3). In this way the individual becomes able to derive from memory the suitable in each case piece of knowledge for facilitating the solution of related composite and complex problems (e.g. see (Voss, 1987)).

On the other hand, the process of learning is usually connected with uncertainty and vagueness. In fact, the learner is in many cases not sure about the good understanding of a new concept or topic and also the teacher is in doubt about the degree of acquisition of a new subject matter by students. Consequently, the use of principles of the FS theory could be a valuable tool in the effort of a more effective description of the mechanisms of learning.

The following experiment took place some time ago at the Graduate Technological Educational Institute of Western Greece, in the city of Patras, during the teaching (in three teaching hours) of the definite integral to a group of 35 students of the School of Management and Economics.

In the instructor's short introduction, during the first teaching hour, the concept of the definite integral was introduced through the need of calculating the area between a curve and the x-axis, but the fundamental theorem of the integral calculus, connecting the indefinite with the definite integral of a continuous in a closed interval function, was stated without proof. Then the students were left to work alone on their papers and the instructor was inspecting their efforts and reactions giving from time to time the proper hints and instructions. His intension was to help students to understand the basic methods of calculating a definite integral in terms to the already known methods for the indefinite integral (step S_1 of the model).

It was observed that 17, 8 and 10 students respectively achieved intermediate, high and very high understanding of the new subject. In other words, in terms of the model one obtains that $n_{ia}=n_{ib}=0$, $n_{ic}=17$, $n_{id}=8$ and $n_{ie}=10$. Therefore the step of representation-interpretation of the process of learning can be represented as a FS in U in the form

$$A_1 = \{(a, 0), (b, 0), (c, \tfrac{17}{35}), (d, \tfrac{8}{35}), (e, \tfrac{10}{35})\}.$$

At the second teaching hour a series of exercises involving the calculation of improper integrals as limits of definite integrals and of the area under a curve (or among curves) was given to students for solution. The target in that case was to help students to generalize the new knowledge to a variety of situations (step S_2 of the model). Working in the same way as above it was found that the step of generalization can be represented as a FS in U in the form

$$A_2 = \{(a, \tfrac{6}{35}), (b, \tfrac{6}{35}), (c, \tfrac{16}{35}), (d, \tfrac{7}{35}), (e, 0)\}.$$

At the third teaching hour a number of composite problems was forwarded to students for solution, involving applications to economics, such as the calculation of the present value in cash flows, of the consumer's and producer's surplus resulting from the change of prices of a given good, of probability density functions, etc ((Dowling, 1980), Chapter 17). The target this time was to help students to relate the new information to their existing schemas of knowledge (step S_3 of the model). In that case it was found that the step of categorization can be represented as a FS in U in the form

$$A_3 = \{(a, \tfrac{12}{35}), (b, \tfrac{10}{35}), (c, \tfrac{13}{35}), (d, 0), (e, 0)\}.$$

Then the membership degrees of all student profiles involved in the fuzzy relation (3) were calculated. For example, for $s = (c, b, a)$ one finds that $m_s = m_{A_1}(c).\, m_{A_2}(b).\, m_{A_3}(a) = \frac{17}{35} \times \frac{6}{35} \times \frac{12}{35} \approx 0.029$.

It turns out that the profile (c, c, c) possesses the greatest membership degree, which is equal to 0.082. Therefore the possibility of each profile s is calculated by $r_s = \frac{m_s}{0.082}$. For example the possibility of (c, b, a) is equal to $\frac{0.029}{0.082} \approx 0.353$, while the possibility of (c, c, c) is equal to 1, etc.

The total number of the student profiles is obviously equal to the total number of the ordered samples with replacement of three objects taken from five, i.e. equal to 5^3. Among all those profiles the profiles possessing non zero membership degrees and their possibilities are presented in Table 1.

In Table 1 all calculations have been made with accuracy up to the third decimal point. The fuzzy data presented in that Table give not only quantitative information, but also a qualitative view of the student behaviour in the classroom during the learning process. This is obviously very useful to the instructor for organizing his/her future teaching plans.

Combined Results of Fuzzy Data. Frequently in Education it becomes necessary to study the combined results of k different groups, $k \geq 2$, during the same process (e.g. the combined performance of two or more student classes in solving the same problems).

For measuring the degree of evidence of the combined results of the k groups, it is necessary to define the **combined probability p(s)** and the **combined possibility r(s)** of each profile s with respect to the membership degrees of s in all the groups involved. The values of $p(s)$ and $r(s)$ can be defined with respect to the **pseudo-frequency**

$$f(s) = \sum_{t=1}^{k} m_s(t) \tag{7}$$

and they are equal to

$$p(s) = \frac{f(s)}{\sum_{s \in U^3} f(s)} \tag{8}$$

and

$$r(s) = \frac{f(s)}{\max\{f(s)\}} \tag{9}$$

respectively, where $max\{f(s)\}$ denotes the maximal pseudo-frequency.

Obviously the same procedure could be applied if one wanted to study the combined results of the behaviour of a single group during k different activities (e.g. the combined performance of a student class during the solution of two or more different problems).

The following example concerning a research about the degree of the student satisfaction for their school education illustrates the importance of the above procedure:

Table 1

Student profiles with non zero membership degrees

A_1	A_2	A_3	m_s	r_s
c	*c*	*c*	0.082	1
c	*c*	*a*	0.076	0.927
c	*c*	*b*	0.063	0.768
c	*a*	*a*	0.028	0.341
c	*b*	*a*	0.028	0.341
c	*b*	*b*	0.024	0.293
d	*d*	*a*	0.016	0.195
d	*d*	*b*	0.013	0.159
d	*d*	*c*	0.021	0.256
D	*a*	*a*	0.013	0.159
D	*b*	*a*	0.013	0.159
D	*b*	*b*	0.011	0.134
D	*c*	*a*	0.031	0.378
D	*c*	*b*	0.026	0.317
D	*c*	*c*	0.034	0.415
E	*a*	*a*	0.017	0.207
E	*b*	*b*	0.014	0.171
E	*c*	*a*	0.039	0.476
E	*c*	*b*	0.033	0.402
E	*c*	*c*	0.042	0.512
E	*d*	*a*	0.025	0.305
E	*d*	*b*	0.021	0.256
E	*d*	*c*	0.027	0.329

Table 2

Student profiles with non zero pseudo-frequencies

A_1	A_2	A_3	$m_s(1)$	$m_s(2)$	$f(s)$	$r(s)$
b	*b*	*b*	0	0.016	0.016	0.092
b	*a*	*b*	0	0.012	0.012	0.069
b	*c*	*b*	0	0.032	0.032	0.184
b	*b*	*a*	0	0.021	0.021	0.121
b	*b*	*c*	0	0.016	0.016	0.092
b	*a*	*a*	0	0.016	0.016	0.092
b	*a*	*c*	0	0.012	0.012	0.069
b	*c*	*a*	0	0.042	0.042	0.241
b	*c*	*c*	0	0.032	0.032	0.184
c	*c*	*c*	0.072	0.080	0.152	0.874
c	*a*	*c*	0.082	0.030	0.112	0.644
c	*b*	*c*	0.031	0.040	0.071	0.408
c	*d*	*c*	0.046	0	0.046	0.264
c	*c*	*a*	0.067	0.107	0.174	1
c	*c*	*b*	0.056	0.008	0.064	0.368
c	*a*	*a*	0.028	0.040	0.068	0.391
c	*a*	*b*	0.024	0.030	0.054	0.310
c	*b*	*a*	0.028	0.053	0.081	0.466
c	*b*	*b*	0.024	0.040	0.064	0.368
c	*d*	*a*	0.043	0	0.043	0.247
c	*d*	*b*	0.036	0	0.036	0.207
d	*d*	*a*	0.020	0	0.020	0.115
d	*d*	*b*	0.017	0	0.017	0.098
d	*d*	*c*	0.022	0	0.022	0.126
d	*a*	*a*	0.013	0.024	0.037	0.213
d	*a*	*b*	0.011	0.018	0.029	0.167
d	*a*	*c*	0.015	0.018	0.033	0.190
d	*b*	*a*	0.013	0.032	0.045	0.259
d	*b*	*b*	0.011	0.024	0.035	0.201
d	*b*	*c*	0.014	0.024	0.038	0.218
d	*c*	*a*	0.031	0.064	0.095	0.546
d	*c*	*b*	0.026	0.048	0.074	0.425
d	*c*	*c*	0.034	0.048	0.082	0.471
e	*a*	*a*	0.017	0	0.017	0.098
e	*a*	*b*	0.014	0	0.014	0.080
e	*a*	*c*	0.018	0	0.018	0.103
e	*b*	*a*	0.017	0	0.017	0.098
e	*b*	*b*	0.014	0	0.014	0.080
e	*b*	*c*	0.018	0	0.018	0.103
e	*c*	*a*	0.039	0	0.039	0.224
e	*c*	*b*	0.033	0	0.033	0.190
e	*c*	*c*	0.042	0	0.042	0.241
e	*d*	*a*	0.025	0	0.025	0.144
e	*d*	*b*	0.021	0	0.021	0.121
e	*d*	*c*	0.027	0	0.027	0.155

Example 2: An educational institution performed a research about the degree of the student satisfaction for their school education, which was characterized by the previously discussed fuzzy linguistic labels *a, b, c, d, e*. The research was performed separately for boys and girls and for three different categories of age, namely C_1: 10-12 years, C_2: 13-15 years and C_3: 16-18 years old.

Denote by A_1 (t), A_2 (t) and A_3 (t) respectively the FSs representing the students' degree of satisfaction for each of the above three categories of age, where the variable t takes the values t = 1 for boys and t = 2 for girls. Such kind of FSs, whose entries depend on the values of a variable, are usually referred as *fuzzy variables.*

According to the collected data the FSs A_i (t), for i = 1, 2, 3 and t = 1, 2 were found to be the following:

$$A_1(1) = \{(a, 0), (b, 0), (c, 0.486), (d, 0.228), (e, 0.286)\}$$

$$A_2(1) = \{(a, 0.171), (b, 0.171), (c, 0.4), (d, 0.257), (e, 0)\}$$

$$A_3(1) = \{(a, 0.343), (b, 0.0286), (c, 0.371), (d, 0), (e, 0)\}$$

$$A_1(2) = \{(a, 0), (b, 0.2), (c, 0.5), (d, 0.3), (e, 0)\}$$

$$A_2(2) = \{(a, 0.2), (b, 0.267), (c, 0.533), (d, 0), (e, 0)\}$$

$$A_3(2) = \{(a, 0.4), (b, 0.3), (c, 0.3), (d, 0), (e, 0)\}.$$

In this example the degree of the student satisfaction in each age category does not depend on the previous categories. Therefore the calculation of the membership degrees of all the student profiles is made by the product law defined by equation (4). For example, for the profile *s* = (*c, c, a*) one finds that

$$m_s(1) = 0.486 \times 0.4 \times 0.343 \approx 0.67 \text{ and } m_s(2) = 0.5 \times 0.5 \times 0.33 \approx 0.107.$$

It turns out that the above profile has the greater pseudo-frequency *f(s)* = 0.67 + 0.107 = 0.174 and therefore its combined possibility is equal to 1, while the combined possibilities of all the other profiles are calculated by $r(s) = \frac{f(s)}{0.174}$.

The membership degrees, the pseudo-frequencies and the combined possibilities of all the student profiles with nonzero pseudo-frequencies are presented in Table 2.

The above calculations have been made again with accuracy up to the third decimal point. The fuzzy data of Table 2 give a detailed idea of the student satisfaction for their school education.

CONCLUSION

The management and evaluation of the fuzzy data obtained by the operation mechanisms of large and complex systems is very important for real life and science applications. A model has been developed in the present work for evaluating such kind of data in terms of the corresponding membership degrees and possibilities. Examples were also presented, for the process of learning a subject-matter in the classroom and for a market's research, illustrating the applicability and usefulness of the model to practical problems.

The general character of the proposed model enables its application to a variety of other human and machine activities (e.g. see the book (Voskoglou, 2017) for a description of such kind of activities) and this is one of our main targets for future research.

References

1. Dowling, E.T. (1980). *Mathematics for Economists,* Schaum's Outline Series, New York, Mc Graw – Hill.
2. Klir, G. J. & Folger, T.A. (1988). *Fuzzy Sets, Uncertainty and Information*, London, Prentice Hall Int.
3. Shackle, G.L.S. (1961). *Decision, Order and Time in Human Affairs,* Cambridge, Cambridge Univ. Press.
4. Voskoglou, M. Gr. (2017). *Finite Markov Chain and Fuzzy Logic Assessment Models: Emerging Research and Opportunities*, Createspace.com. – Amazon, Columbia, SC, USA.
5. Voskoglou, M. Gr. (2019). Multi-Valued Logics: A Review, *International Journal of Applications of Fuzzy Sets and Artificial Intelligence*, 9, 5-12.
6. Voss, J.F. (1987). Learning and transfer in subject matter learning: A problem solving model, *Int. J. Educ. Research*,11, 607-622.
7. Zadeh, L.A. (1965). Fuzzy Sets, *Information and Control*, 8, 338-353.

УПРАВЛІННЯ НЕЧІТКИМИ ДАНИМИ В ОСВІТІ

Майкл Гр. Воскоглу

Вищий технологічний освітній інститут Західної Греції,
Школа технологічних застосувань, Греція

Анотація.

Формулювання проблеми. *Кілька років тому унікальним інструментом в руках вчених для обробки ситуацій невизначеності, які часто з'являються в проблемах науки, техніки і повсякденного життя, була теорія ймовірності. Однак тепер теорія нечітких множин, ініційована Заде в 1965 році, а також її розширення та узагальнення дали нову динаміку цій галузі.*

Матеріали і методи. *Використано математичні маетоди аналізу.*

Результати. *У даній роботі розроблена модель для обробки нечітких даних, що з'являються в галузі освіти. Модель базується на розрахунках можливостей профілів, що беруть участь у відповідних ситуаціях, які, на думку британського економіста Шеккла і багатьох інших дослідників, є більш придатними, ніж нечіткі ймовірності для вивчення поведінки людини. Також в роботі представлено застосування навчального класу для вивчення математики, що ілюструє важливість моделі на практиці. В подальших дослідженнях загальна модель розширена для вивчення об'єднаних результатів оцінки нечітких даних, отриманих з двох (або більше) різних джерел, і наведено приклад, який підкреслює корисність цього розширення для реальних ситуацій в освіті.*

Висновки. *Управління та оцінка нечітких даних, отриманих механізмами експлуатації великих і складних систем, дуже важлива для реального життя та наукових застосувань. Розроблена модель дозволяє оцінити такого роду дані з точки зору відповідних ступенів та можливостей участі. Приклад процесу вивчення предмета в навчальному класі та приклад дослідження ринку ілюструють застосовність і корисність моделі в практичній площині. Загальний характер запропонованої моделі дає змогу застосовувати його до інших людських та машинних дій, що і є однією з головних цілей для подальших досліджень.*

Ключові слова: *нечітка множина, ступінь участі, можливість, нечіткі дані, нечітка змінна.*