DOI: 10.5281/zenodo.3949686
UDC 004.891

FUZZY FUNCTIONS OF EXPERT KNOWLEDGE ENCAPSULATED WITHIN STATISTICAL WORKFLOW DATA

Viorel Carbune*, ORCID: 0000-0002-1556-4453

Technical University of Moldova, 168 Stefan cel Mare Bvd., MD-2004, Chisinau, Republic of Moldova
**viorel.carbune@calc.utm.md*

Received: 05. 28. 2020
Accepted: 07. 22. 2020

Abstract. Taking over the skills of the human expert will make it possible to develop decision-making algorithms in conditions of uncertainty for industrial applications. Fuzzy sets are used in different fields and estimating membership functions is one of the most important issues in the design of fuzzy systems that depends directly on the identification of used method. This article presents an approach to this problem that can provide solutions in specific cases. In this context, a method of extracting the knowledge of the human expert is developed and it allows to retrieve specific expertise and the construction of algorithms for decision support systems. The conditions to apply the method and identify membership functions as well as the automation process of this stage are analyzed. There is proposed a method to determine trapezoidal or custom membership functions. The approach presented in this paper can be applied to the analysis and research of decision making in conditions of uncertainty. A case study is presented that reflects the applicability of the proposed method and algorithms.

Keywords: *human expert, membership functions, fuzzy variables, identification, fuzzy decision making, knowledge, skills.*

1. Introduction

One can notice the growing role and share of intelligent systems in the field of industrial applications, the issue of research, development and implementation of advanced solutions to take over the experience of the human expert which are becoming more and more current. Specific industrial processes can be characterized by a certain degree of ambiguity in the decision-making process [1].

The basic problem in the knowledge extraction process is data structuring in order to continue their use for the construction of decision-making systems. Particularly, the problem of determining the membership functions during the design stage of fuzzy systems is a very current one in the conditions of increasing the requirements of industrial process automation [2]. Under these conditions, the development of fuzzy systems depends directly on the qualification of the human expert and his ability to formulate the characteristics of the system. Thus, the accuracy of the designed system depends on the possibility of the human expert to formulate the rules of inference and specify the number of fuzzy variables and their membership functions [3, 4]. Some processes are characterized by both uncertainty and a certain degree of ambiguity. Respectively, a negative influence on the system properties can also have the procedure of identifying the number of fuzzy variables and their membership functions. This is because the theory underlying this methodology provides a concrete answer to the questions regarding the number of variables and the form of the membership function for each specific variable. In many cases solving these important issues depends on the human expert, and a practical solution can be often received only by an experimental way. So, the efficiency of the obtained solutions depends mainly on the expert knowledge. Under these conditions, the formalization and standardization of this procedure becomes the main problem for engineers dealing with the design of fuzzy systems, and solving that problem can streamline their work by the possibility of implementing algorithms to automate the process.

The paper is organized as follows. Firstly, a brief overview of the problem of identifying membership functions is given. This includes a short analysis of some basic methodologies related to that problem (section 2). Then, we put into discussion an approach for the identification of fuzzy membership functions from expert knowledge encapsulated within statistical data. In this context, the identification method and the algorithm for calculation of membership functions are described (section 3). After that, a case study is analyzed (section 4). This study explains how to apply the proposed approach when a real production system is considered. To this goal, there is described the procedure of taking over the human operator skills encapsulated within statistical workflow data. Finally, the conclusions are drawn in section 5.

2. Related works

As it is well known, the correct choice of the form and values of the membership function is not a trivial task at all. A unique membership function for each vague concept cannot be defined due to uncertainty, diversity and individual data differences. For these reasons, it is considered that membership functions can be selected arbitrarily, or intuitively, without following any pre-established procedures. However, both the nature of the data used to identify the membership functions and the nature of the process they characterize must be taken into account [2 - 5].

On the other hand, the organization of a production system, which would ensure the efficiency of the technological process and the quality of the end product, is the key challenge in a market economy and free competition. In this context, it is necessary to optimize the production system in terms of reaction's time in the decision-making process as well as their accuracy to allow to control costs and improve the quality of the end product [2].

The task of the decision-making system in the production process can often be characterized by a degree of uncertainty, generated by incompleteness, inaccuracy, fragmentation, validity, ambiguity or contradiction of information [1, 3]. Specifically, the human operator may not be able to accurately describe the nature of the process in terms of its state or dynamics. In other cases, he may not be able to establish exactly the cause-and-effect connection between things or events. The human may perceive and interpret information differently, or focus on different aspects of it: the accuracy of data acquisition, the negative effects of noise, or the individual particularities of the process. [2, 3]. In the case of specific industrial processes, that take place under uncertainty, there is a need to identify as accurately as possible the membership functions. Their accuracy can decisively influence the evolution of such processes, because both the real time acquisition of valid data and the application of the strategic plan are the necessary conditions for maintaining the industrial process on the desired trajectory.

In order to improve the performance of the human operator or to ensure the automation of the production process, what would improve the quality of the end product, there is a need for highly-integrated decision support systems [1, 2]. In general, many of the decision-making systems described in the literature are based on classical mathematical models for solving classification problems. However, the uncertainty and ambiguity of the input data are not taken into account in that models and also their do not reproduce the expert's decision-making process. The application of fuzzy logic for solving decision-making problems once again argues the property of this methodology to solve the problem of inaccurate and uncertain knowledge management, which ensures the support of the decision-making process [1, 5].

Fuzzy systems are based on membership functions used to represent unclear input values. These functions can be generated in different ways, one of which suppose the involvement of the expert to define them.

However, the latter method is not always applicable. Thus, the automatic determination of membership functions is a matter of current interest. There are several methods proposed for identifying such functions [6 - 10]. Some of that methods are based on the distribution of probability or possibilities when the data on which the identification method is applied are ambiguous or uncertain [3]. Often, statistical data describing the training process can be quite ambiguous and may have some uncertainty. Thus, data belonging to certain categories can be computed on the basis of clusters. A method of automatic generation of the membership functions that divides the initial set of data into classes which can then be used to obtain desired functions is presented in [6, 7]. So, this method supposes data grouping into classes, then the membership functions are generated for the obtained classes.

The proposed in [6] method for automatic generation of membership functions can lead to a fairly large number of classes and rules of inference. This fact may cause increased complexity of the inference engine. Besides, the analyzed method allows only the generation of classical membership functions and does not provide for the possibility of developing custom functions that could directly inherit the behavioral character of the expert.

Next section presents a method for the identification of membership functions, that can be used to develop decision-making systems designated for technological systems in which the presence of human operator is indispensable.

3. Approach for the identification of fuzzy membership functions

The method proposed in this article has some similarities with the method described in [6], but it represents a new approach that aims to eliminate the disadvantages mentioned above. In order to achieve this objective, it is proposed, on the one hand, to limit the maximum number of linguistic variables.

On the other hand, it is proposed to specify and create custom membership functions. For these reasons, an algorithm was developed to determine membership functions based on the processing and interpretation of data streams, which is shown in Figure 1.

Being given that a technological system is often supervised by a human expert who ensures the quality of the end product, it can be assumed that the operator is experienced one, and the data acquired over the production process are valid. Assuming that the collected data are valid and the operator who controls the process is qualified, we can conclude that the evolution of that process is acceptable.

Then the process itself can be considered as being into a good operating condition, Figure 2 (a). In order to determine the discrete membership function in that case, some manipulation of the data stream is necessary. At first, the data should be preprocessed in order to exclude the samples that do not meet the quality requirements. It should be noted that at this stage the data is filtered based on the output streams that represent the parameters of interest obtained as a result of checking the quality of the end product. This step of the procedure will affect both the input and output streams according to Eq.(1) and Eq.(2).

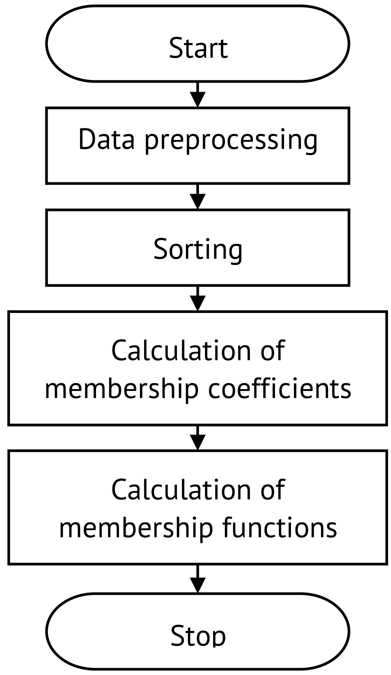

Figure 1. Algorithm for the identification of membership functions.

$$\varphi_{out}^{f} = \{x | x \in \varphi_{out}^{nf}, Min \le x \le Max\} \tag{1}$$

$$\varphi_{in}^{f} = \{x | x \to y, x \in \varphi_{in}^{nf}, y \in \varphi_{out}^{f}\} \tag{2}$$

where φ_{out}^{f} represents the filtered output stream, φ_{out}^{nf} – the unfiltered output stream, φ_{in}^{f} represents the filtered input stream, φ_{in}^{nf} – the unfiltered input stream.

As a result of the preprocessing, new input/output data streams will be obtained, which will be built on the basis of the initial data flows. So, the values of the output data that do not meet the quality requirements and the corresponding values of the input data will be excluded, Figure 2 (b).

These forward-looking operations aim to minimize deviations of the fuzzy system which is based on the behavior of the human operator involved to control the process. Obtaining clean data streams in this way, one can perform the procedure of mathematical calculation of the membership function for the state of good functioning.

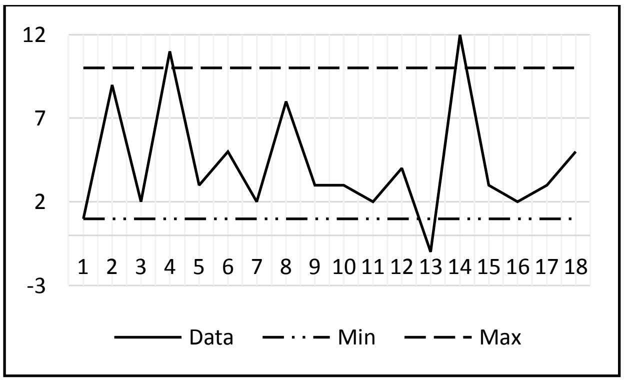

Figure 2 (a). Data stream.

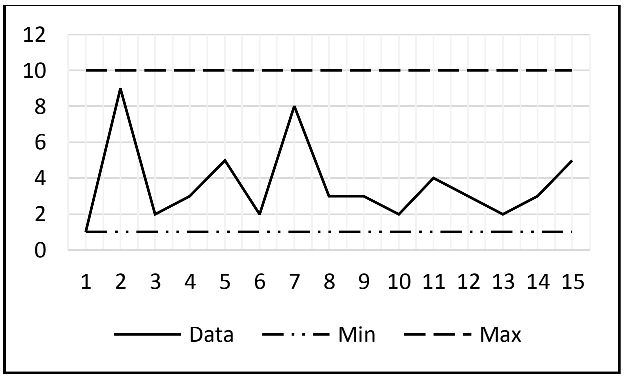

Figure 2 (b). Filtered data stream.

This procedure consists of analyzing the input and output data streams in order to calculate the value of belonging to the mentioned state for each discrete value within the range of values.

To optimize the declared procedure, the incremental sorting of data streams could be applied, Eq.(3). However, this will affect the relationship between input/output streams.

$$\varphi^{s}_{in/out} = \varphi^{f}_{in/out} \uparrow \tag{3}$$

That relationship can be preserved by saving copies of data streams and processing those copies. The determination of a set of membership coefficients in this case comes down to computing the number of repetitions of discrete values relative to the number of occurrences of the element with the maximum frequency in the data stream, Eq.(4):

$$MC(x)_{in/out} = \frac{N_x}{N_{Max}} \tag{4}$$

where $MC(x)_{in/out}$ is the membership coefficient of discrete value x, N_x – the number of occurrences of the discrete value x, iar N_{Max} – the number of occurrences of the most common (the most frequent) discrete value seen in the stream.

As a result of this process a normalized membership function of the state of good functioning is obtained, Figure 3 (a). This method of representing membership functions is easier to perceive and more convenient to use.

By analyzing the diagram of the membership function, its shape can be visually estimated. Visual assessment can serve as an indicator of the correctness of the choice of form of membership function only for a human expert and does not matter for a computer system.

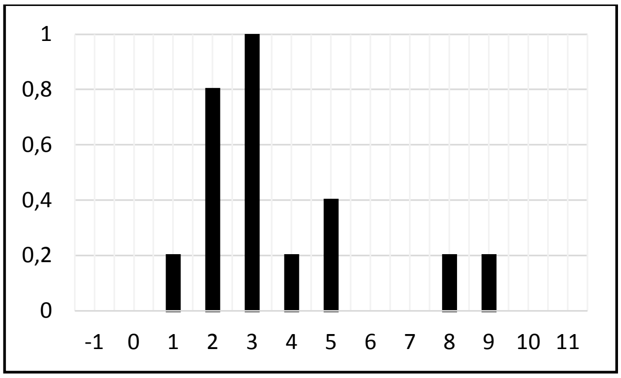

Figure 3 (a). Distribution chart of membership coefficients.

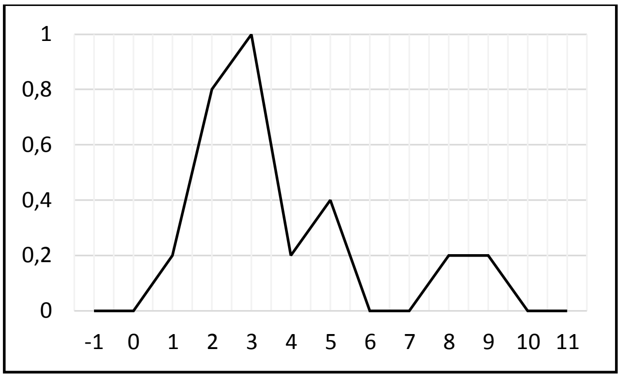

Figure 3 (b). Linear interpolated membership function.

Thus, there is a need to determine the shape of the continuous membership function based on a discrete membership function, Figure 3 (b).

The above problem can be solved quite easily in the case of simple shapes, such as triangular, rectangular or trapezoidal, but it is difficult to solve in the case of a sigmoid shape. Since the methodology itself does not give a clear answer to this question, a partial solution can be obtained by grouping all forms of membership functions into 2 categories:

- trapezoidal functions, Figure 4 (a); with triangle as particular case, Figure 4 (b); and with more maximums, Eq.(5):
- custom functions - a membership function can follow exactly the values of the membership coefficients, or can be an approximation of them, Figure 5.

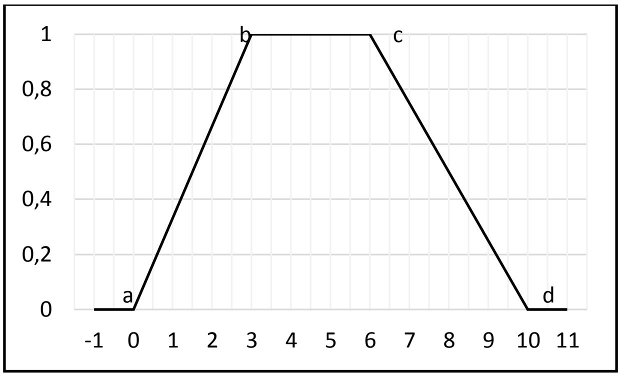

Figure 4 (a). Trapezoidal function.

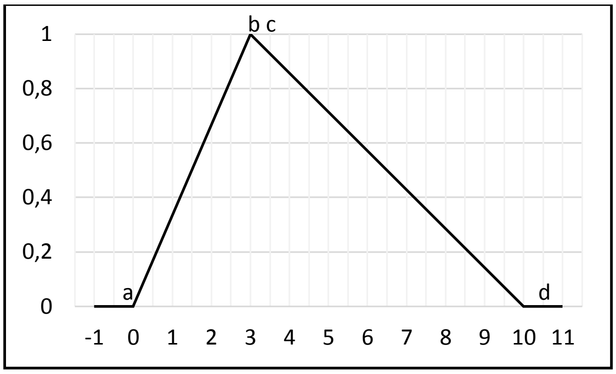

Figure 4 (b). The calculated function.

It should be noted that after the calculation of membership coefficients for each value within the range of values there are may be some values whose membership coefficient is very small or even zero, Figure 3 (a), (b). This can be caused either by the nature and dynamics of the process or equipment, or by the behavior and reaction of the human operator.

$$\mu_F(x,a,b,c,d)=\begin{cases}0, & x<a;\\ \dfrac{x-a}{b-a}, & a\le x\le b;\\ 1, & b<x<c;\\ \dfrac{d-x}{d-c}, & c\le x\le d;\\ 0, & x>d\end{cases} \tag{5}$$

As a result, a phenomenon can be observed when one or several values of the membership coefficient represent the values of local minima, which may indicate the absence of this value in the analyzed stream. Therefore, there is a problem with the correctness of this value. In the case of the trapezoidal shape of membership function, this problem can be solved by determining the points of interest of the trapezoid**,** Figure 4 (a), followed by linear interpolation of the values between them according to Eq.(6), Eq.(7), Eq.(8), Eq.(9).

$$a = MAX(x)|Mf(x) = 0, x < y, Mf(y) = 1 \tag{6}$$

$$b = MIN(x)|Mf(x) = 1 \tag{7}$$

$$\mathrm{c} = \mathrm{MAX(x)|Mf(x)} = 1 \tag{8}$$

$$d = MIN(x)|Mf(x) = 0, x > y, Mf(y) = 1 \tag{9}$$

In particular case when b = c, a triangular membership function is obtained, shown in Figure 4 (b).

In the case of a custom form of the membership function, a solution to the above problem can be achieved either using an interpolation method, which ignores that values, or by providing the data stream containing all the range's values, Figure 5.

Once the form of the membership function is defined, and respectively the interpolation method in the case of the custom form, the membership functions of the other fuzzy variables can be determined. Obviously, the membership function that belongs to the area of good functioning divides the range of values into two areas of poor functioning, the area on the left and the right.

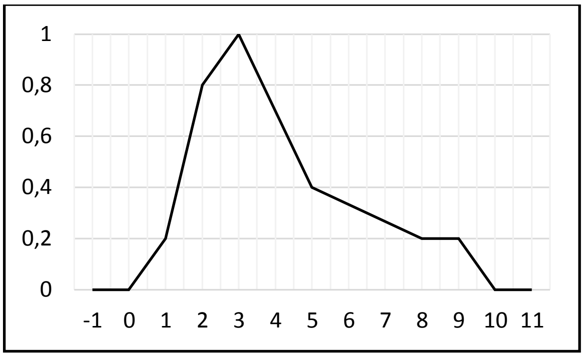

Figure 5. Custom membership function.

These areas can be associated with fuzzy variables, whose membership functions can be easily computed, given the function of the area of good functioning.

To do this, it is enough to subtract 1 from the value of the function that belongs to the area of good functioning.

The form of the function thus obtained will be the inverse of good functioning, so will represent the area of poor functioning. The resulting function can be interpreted as two variables relative to its minimum values on the left, Eq.(10) and on the right, Eq.(11), respectively, which will make it possible to operate with a large number of fuzzy variables for more precise process control.

$$Mf_L^b(x) = \begin{cases} 1 - Mf_C^g(x), & x < y,\ Mf_C^g(y) = 1; \\ 0, & x \geq y,\ Mf_C^g(y) = 1. \end{cases} \tag{10}$$

$$Mf_R^b(x) = \begin{cases} 0, & x \geq y,\ Mf_C^g(y) = 1; \\ 1 - Mf_C^g(x), & x < y,\ Mf_C^g(y) = 1. \end{cases} \tag{11}$$

where $Mf_L^b(x)$ is the computed membership function placed on the left side near the area of good functioning, $Mf_C^g(x)$ și $Mf_C^g(y)$ – the value of the membership function which refers to the area of good functioning (center) for x and y, respectively, and $Mf_R^b(x)$ – the computed membership function placed on the right side near the area of good functioning.

This method ensures the partial overlap of the adjacent membership functions. The latter feature is useful in case where it is desirable to generalize the decision-making process.

On the other hand, this is not a good choice if you want to follow the human operator's behavior as accurately as possible.

4. Case study

To show how to apply the method and algorithm proposed in the previous section, the following study was performed.

Let us consider the technological process of production of glass coated microwires [11]. The method of identifying membership functions was applied on the data verified at the stage of quality control of the end product. For this reason, statistical data were provided by the most experienced operators involved in casting of the resistive microwires of 5÷10Ω. Based on these data, a graph was plotted as shown in Figure 6 (a).

Due to the specific issues of technological process some deviations from quality standards may occur. To tackle the negative impact of some data points on the decision-making model, the acquired data were subjected to a filtering procedure. After that, the graph shown in Figure 6 (b) was obtained.

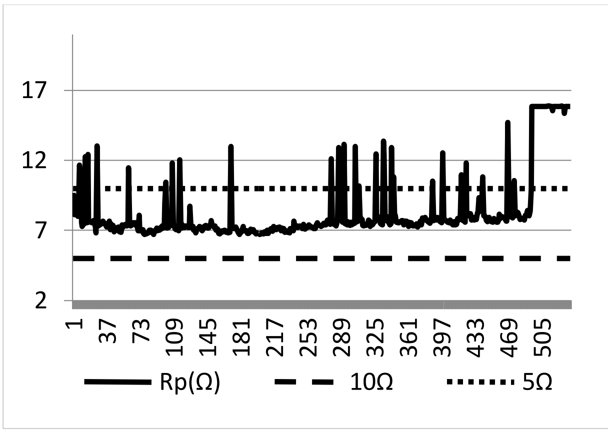

Figure 6 (a). Microwire resistance graph.

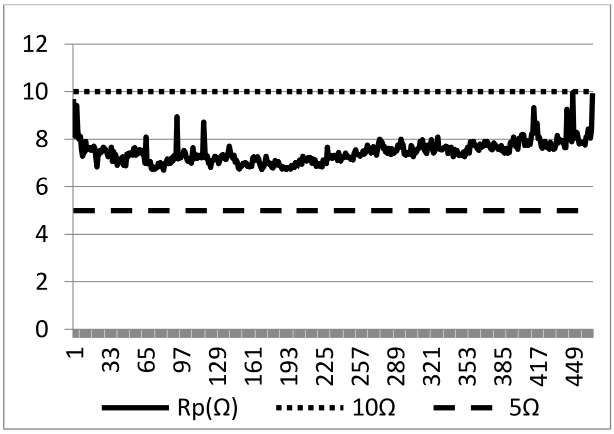

Figure 6 (b). Graph after data filtering.

During the casting of microwire the decision-making process involves the determining of a specific parameter such as R^{ref} which will be used as reference value. So, the operator is guided by that reference value while taking decisions throughout the technological process.

The need to define the above-mentioned parameter is explained by the specific behavioral properties exhibited by each human operator in the production process.

Depending on the quality requirements, the casting plant, type of microwire, external factors such as noise, each operator defines the reference value of this parameter by following the explicit or intuitive way of judgement.

Thus, the control of the microwire casting process consists in supervising and maintaining certain parameters such as the linear resistance of the microwire in the vicinity of the reference value. In our case it is about R^{ref} that must be within the range of values specified in the technical requirements. In order to determine the reference value R^{ref} associated with the behavior of human operator, the graph of the distribution of the frequency of occurrence of the linear resistance values over the entire duration of the microwire casting process was constructed, Figure 7 (a).

From the graph shown in Figure 7 (a), it can be observed that the frequency or count of the occurrences of value R = 7.22Ω is maximum and equals 11. This fact shows that, following an explicit or intuitive strategy, the operator maintains the linear resistance of the microwire close to 7.22Ω. The latter is the reference value R^{ref} and this behavior is typical for given operator.

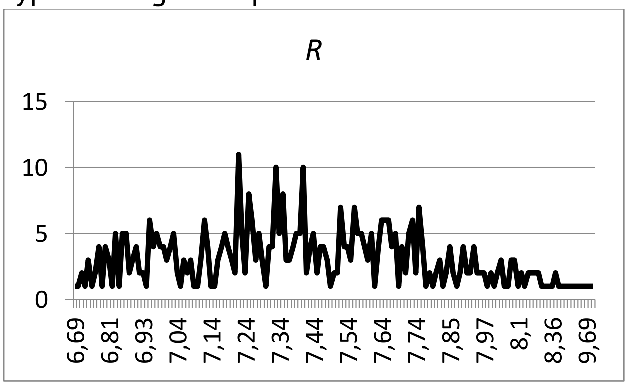

Figure 7 (a). Distribution graph.

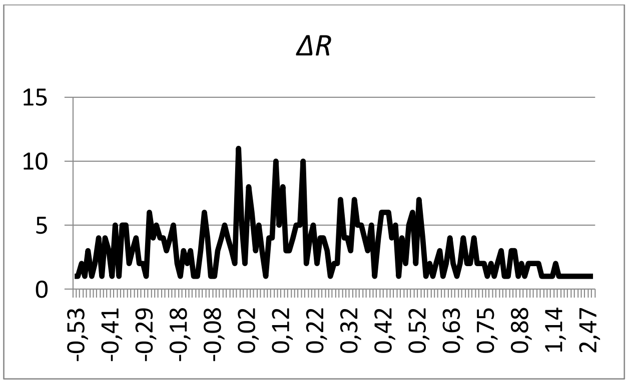

Figure 7 (b). Deviation graph.

Figure 7 (b) shows the distribution graph of the frequency or count of occurrences of deviations of the linear resistance relative to the R^{ref} and their real magnitude during the casting process. Using this graph, we can determine what magnitude appears as most frequent deviation of the resistance relative to the reference value. In our case this magnitude is 0Ω and its maximum frequency of occurrences equals 11.

The above-mentioned deviation of 0Ω reflects a good algorithm of decision-making by human and that the microwire casting process is stable one. That maximum number related to the most frequent value of deviation of linear resistance will be used at the stage of defining the membership functions for linguistic variables.

After analyzing the data collected from the human operator, a set of characteristic parameters was calculated that serve to construct membership functions in accordance with the behavior of each operator. Then, using the developed method, specific membership functions were obtained that can summarize the experience and knowledge of experts, but at the same time can allow to retain their individual skills.

Figure 8 (a) shows the triangular membership functions for the $-\Delta R$ and $+\Delta R$ fuzzy variables, calculated for the human operator whose behavior was investigated.

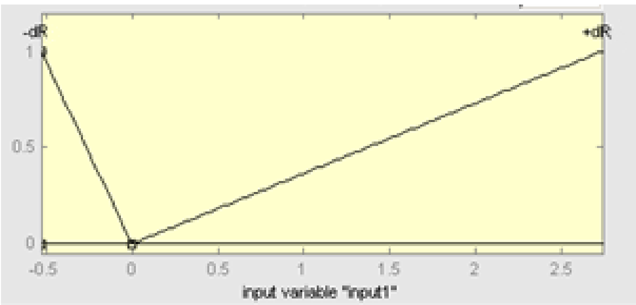

Figure 8 (a). Membership functions.

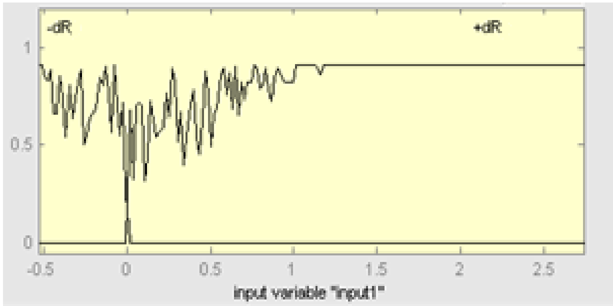

Figure 8 (b). Custom functions.

The above membership functions, Figure 8 (a), were obtained by applying the expert's behavioral characteristics, as well as using classical forms of functions. It is important to mention that the membership functions provided only partially adopt the skills of the human operator. However, this should not be considered solely as a disadvantage, since the decision-making by operator is characterized by ambiguity.

For these reasons, in order to investigate behavioral characteristics of the human operator and to take over his skills as much as possible, it is necessary to apply the proposed method for determining custom membership functions. As it was previously mentioned, this approach provides for the possibility of developing membership functions that could directly inherit the behavioral characteristics of the operator.

Thus, using the method described in section 3, and taking into account the results obtained in this section, the custom membership functions were specified. Figure 8 (b) illustrates the custom membership functions for the $-\Delta R$ (on the left side of the figure) and $+\Delta R$ (on the right side of the figure) fuzzy variables.

5. Conclusions

Taking over the skills of the human expert makes it is possible to develop decision-making systems under conditions of uncertainty for some real-life applications. For these reasons, the proposed approach in the article for the identification of fuzzy membership functions from expert knowledge encapsulated within statistical data may be of real interest. In this context, the technological process of production of glass coated microwires was considered by authors. The proposed approach was applied on the data provided at the stage of quality control of the microwire, which represents the end product.

As a result of our study, a model was developed that allows one to adopt the skills of a human operator by identifying the membership functions for fuzzy variables. Generally speaking, a successful implementation procedure depends on several factors such as the number of fuzzy variables, the type and form of membership functions, the number of inference rules, but also on the defuzzification method used. The proposed method makes it possible to choose appropriate membership functions shapes and can be automated to simplify the process of extracting knowledge from an expert person. One can note that during this research the MATLAB engineering software was used. Finally, in our opinion it is necessary to pay more attention to build custom membership functions that can offer a new perspective in the field.

References

1. Valášková K., Kliestik T., Mišanková M. The Role of Fuzzy Logic in Decision Making Process. In: 2nd International Conference on Management Innovation and Business Innovation (ICMIBI 2014), Bangkok, Thailand, 2014.
2. Caldas L., Belfiore P., Leonardi F. Application of Fuzzy Cluster Means for Decision Making in Industrial Process. In: The 20th International Congress of Mechanical Engineering (COBEM 2009), Gramado, RS - Brazl, 2009.
3. D'acierno A., Esposito M., De Pietro G. An extensible six-step methodology to automatically generate fuzzy DSSs for diagnostic applications [online]. BMC Bioinformatics, 14 January, 2013, [accessed 01.03.2020]. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3548688/
4. Sadollah A. Introductory Chapter: Which Membership Function is Appropriate in Fuzzy System? In: Fuzzy Logic Based in Optimization Methods and Control Systems and Its Applications, SADOLLAH, A., [online], ed *IntechOpen*, 31 October, 2018, [accessed 01.03.2020]. Available at: https://www.intechopen.com/books/fuzzy-logic-based-in-optimization-methods-and-control-systems-and-its-applications/
5. Gan G., LI B., LI X. Wang S. (eds). Advanced Data Mining and Applications. The Proceedings of 14th International Conference, ADMA 2018, Nanjing, China, November 16−18, 2018. Lecture Notes in Artificial Intelligence, vol. 11323. Springer, 2018, 532 p. doi:10.1007/978-3-030-05090-0
6. Cano J.C., Nava P. A fuzzy method for automatic generation of membership function using fuzzy relations from training examples. In: Proceedings of Fuzzy Information Processing Society, 2002. NAFIPS. 2002 Annual Meeting of the North American, 2002, pp.158−162.
7. Tzung-Pei H., Lee C. Induction of fuzzy rules and membership functions from training examples. In: Fuzzy Sets and Systems, Nov. 1996, Vol. 84, pp. 33-47.
8. Yoshikawa A. Improvement of Membership Function Identification Method in Usability and Precision [online]. In: Advances in Soft Computing. Roy R., Furuhashi T., Chawdhry P.K., (eds). Springer, London, 1999, [accessed 04.01.2020]. Available at: https://link.springer.com/chapter/10.1007/978-1-4471-0819-1_18
9. Nieradka G., Butkiewicz B. A Method for Automatic Membership Function Estimation Based on Fuzzy Measures [online]. In: Foundations of Fuzzy Logic and Soft Computing. Melin P., Castillo O., Aguilar L.T., Kacprzyk J., Pedrycz W., (eds) IFSA 2007. Lecture Notes in Computer Science, vol. 4529. Springer, Berlin, Heidelberg, 2007, [accessed 04.01.2020]. Available at: https://link.springer.com/chapter/10.1007/978-3-540-72950-1_45
10. Pazhoumand-Dar H., Lam C., Masek M. Automatic Generation of Fuzzy Membership Functions using Adaptive Mean-shift and Robust Statistics. ICAART, 2016, pp.160-171.
11. Zaporojan S., Plotnic C., Calmicov I., Larin V. A knowledge-based approach for microwire casting plant control. In: JOZEFCZYK, J. and ORSKI, D., (eds). Knowledge-Based Intelligent System Advancements: Systemic and Cybernetic Approaches. Hershey PA: IGI Global, 2011, pp. 419-437. doi: 10.4018/978-1-61692-811-7.ch19