



Modeling a fuzzy system for assisting the customer targeting decisions in retail companies

Angela-Eliza Micu

Abstract

This paper reveals a proposal of a decision support system, enabling retail companies' marketing managers with powerful tools for the customers' portfolios targeting. The innovative methodological approach highlights a fuzzy algorithm based system that captures and computes relevant information about customers, providing a managerial support for making the most effective decisions, according to different degrees of affiliation assigned to customers. The contributions of fuzzy algorithms on targeting decisions reflects the affiliation of customers to a specific cluster, based on computation of the input vectors, income and seniority, leading to performing marketing decisions.

1 Introduction

In 1965 the American professor **Lotfi Zadeh** published the article "Fuzzy Sets" (Zade65) laying the mathematical foundations of fuzzy logic. The author has given up the bivalent logic (only yes or no) and laid the mathematical foundations for multivalent logic (yes and no) by defining the fuzzy set as an extension of the classical set, a precise one (crisp). The fuzzy set, unlike normal classic sets does not have well defined boundaries, its elements belonging to the set only to a certain extent. The term "fuzzy" would be translated into Romanian by vague, imprecise, inaccurately diffused. As an already well-known term it is further used as it is. Lotfi Zadeh (1997) considered necessary

Key Words: fuzzy system
Received: 30.05.2015
Revised: 17.10.2015
Accepted: 10.12.2015

to introduce fuzzy algorithms due to the fact that he noticed that: the more complex a problem is, the harder it is to be solved algorithmically, and even if we arrive at a mathematical result, this one in terms of volume of calculation is not convenient (SuYa93). In one of his first articles, L.Zadeh enunciated the principle of incompatibility between prediction and complexity, which manifests itself strongly in the case of humanoid systems (Zade73). This principle was already known by Albert Einstein: "As long as the laws of mathematics refer to reality, they are not accurate. And until they are accurate, they do not refer to reality anymore". (Erfahrung wave Geometry). When a system or a process is very complex and/or is affected by deep uncertainties, the exact classical mathematical methods, do not lead to convenient solutions. Although there are no known exact methods to optimally solve complex problems, the human operator very often successfully solves these problems using heuristics imprecise and intuitive algorithms. Even very complex systems and processes can be solved by the human operator using approximate reasoning. Let us consider the well-known example of a driver that using different information like speed, road condition and traffic can always adopt the adequate behavior required a safe travel. However, the fully automatic control of a car moving on a highway is still an open question. A system consisting of a car, a certain road, traffic conditions, etc. is way too complicated to be modeled and controlled properly using known exact methods. So it is observed that the precision - used in classical automatic control - may be unnecessary and even disturbing, and an approximate approach to the problem - used by the human operator - unquestionably solves the problem (Mirkin, 2003). If we want to model the "algorithm" used by the driver, probably the best estimation would be the description of knowledge, information contained by the rules used, the basic rules are formed during the training interval. The accumulated collection of rules widens continuously with the increase of the experience. These rules are, for example: "Reduce live for, if the road is wet", "Step on the brake if the traffic light is red ", "Press more on the acceleration, if the car goes up a hill", etc. The fuzzy logic represents an approximate method by which this vague knowledge, stored in a rule base, can be modeled formally. The implementation of fuzzy systems is due to their advantages they have in the following specific situations: - allow the modeling of nonlinear, complex or imprecise known systems, - allow the implementation of human experience for building inference rules, by using linguistic variables. The fuzzy logic is basically a calculation technique with which in the case of solving specific problems, you can get superior performances compared to the accurate conventional methods. The fuzzy systems also have very good behavior in the presence of uncertainty, imprecision and noise. How well these fuzzy systems work it is demonstrated by their worldwide widespread in recent years (Nauck

et al., 1997). We already know a number of fuzzy logic applications in various fields of science: automatic control (temperature adjustments, subway speed control, video cameras autofocus) recognition of a pattern (fuzzy classification algorithms), measurements (processing the information provided by the sensors), medical (pacemakers control), economy (fuzzy decision methods), cognitive psychology (fuzzy modeling of the sight system). Fuzzy Sets In everyday language in order to describe a person we use adjectives like "tall", "young", "energetic", on the basis of which we approximately know the height, the age and respectively the behavior of that specific person. If we want more information, we can usually indicate that the person is "very tall" or "less young". These adjectives, so-called linguistic variables are used in qualitative descriptions bear a certain degree of ambiguity, because it is not known precisely what is the value of that size ("tall" = ?cm) and the adjectives can be subjectively interpreted in several ways. The linguistic variables like "tall" or "young" do not specify exactly the height or the age of the described person, but are still connected in some way with the values of heights or ages. Further on the linguistic variables will be expressed using the concept of set. In a research paper (Li, 2000), the fuzzyfication of marketing strategy factors and the fuzzy reasoning for setting marketing strategy are addressed.

2 Setting the problem

We want to create a software system for decision support for the segmentation of the portfolios of customers for retail companies. The approach consists from providing a system that processes a series of information considered relevant about the customer, providing the loans granting personnel a support tool for taking the adequate decision.

Two types of algorithms will be used:

- 1. The fuzzy algorithm** - in this case each relevant information about the customer is fuzzyfied, resulting in a degree of affiliation to that field. By the interference mechanism we obtain a classification of the customers, the fuzzyfication not being clear, but the customer will belong to different segments with different degrees of affiliation. The approach consists of providing a system which processes some information considered relevant about the customer, providing to the marketing manager a decision support tool. We start from the manner in which the customer segmentation is made in the retail companies. The method involves assigning a score for each customer. When determining a customer's score for a particular category, points are awarded or not, without the possibility to refine this choice. When using fuzzy techniques, relevant information about each client is fuzzyfied, resulting in a degree of affiliation to that field. By the interference mechanism we obtain a classi-

fication of the customers, the fuzzyfication not being clear, but the customer will belong to different segments with different degrees of affiliation. First we considered the following sizes that can be fuzzyfied: income and "seniority" in the retail company (criterion used to determine the loyalty points). The other sizes were considered as eligibility sizes. For the income were considered four linguistic terms (Figure 1a): Small, Medium, Large and Very Large, and for the seniority three linguistic terms: Small, Medium and Large (Figure 1b). The rule base is shown in Figure 2; we present the manner in which the inference is applied. The information that results is actually a score between 25 and 75 (random chosen limits). The advantage of this method is that it allows the continuous variation of the score of a customer.

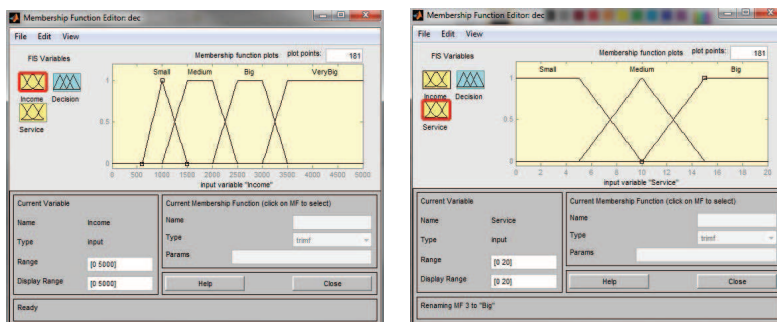


Figure 1: a. Linguistic terms for income; b. Linguistic terms for "seniority" in the retail company

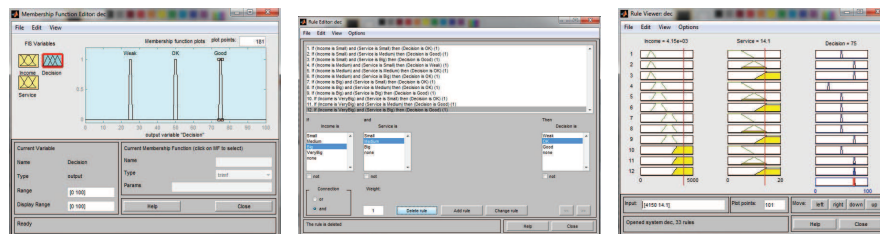


Figure 2: a. Score according to the income and the "seniority"; b. Rule base; c. The manner in which the inference is applied

The previous application was further developed by introducing a new variable used in the customer segmentation field: the number of days of delay for the payment of installments corresponding to previous purchases. It was

considered that the maximum acceptable delay is of 60 days, which is in accordance with the lending strategies of financial institutions. Also there were redefined the linguistic terms corresponding to the other two input variables ("seniority" in the retail company and income) and we introduced a new linguistic term corresponding to the output variable. The introduction of this new term was needed in order to describe using a wider spectrum the possibility to include the customers in the scoring chart. In Figure 3 there are shown the considered input and output variables and after these figures it is presented the implemented rule base. We can notice that the rule base has now a total of 36 rules, and the introduction of a new entry with 3 linguistic terms would lead to a rule base of 108 rules. Thus, the approach of this empowerment thesis is justified: some sizes are considered only for eligibility (the conditions are met or not), and a limited number in order to allow a reasonably sized must be considered as *fuzzyfiabile*.

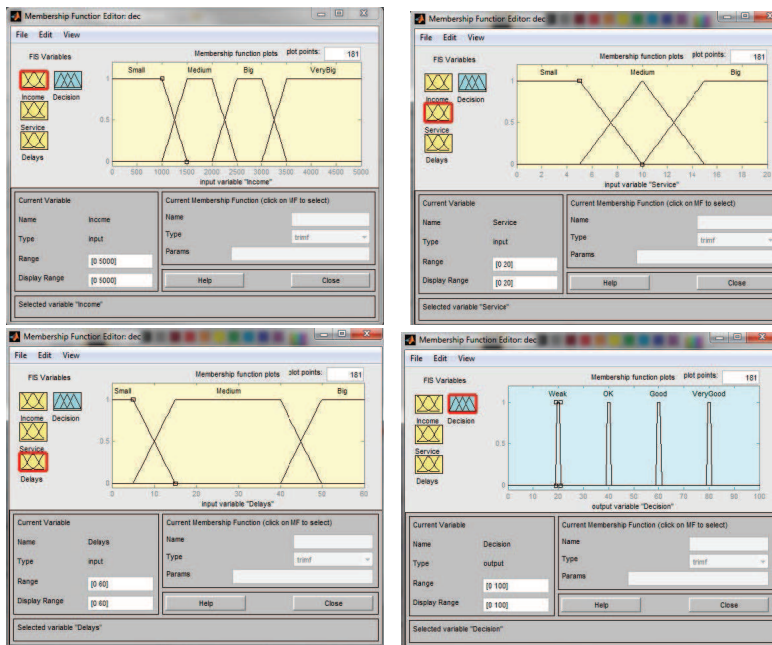


Figure 3: Linguistic terms for the input and output variables

The rule base of the new fuzzy block:

1. (income==small) & (seniority==small) & (delay==small) => (output=ok) (1)
2. (income ==small) & (seniority==medium) & (delay==small) => (output=good) (1)
3. (income==small) & (seniority==big) & (delay==small) => (output=good) (1)
4. (income==medium) & (seniority==small) & (delay==small) => (output=ok) (1)
5. (income==medium) & (seniority==medium) & (delay==small) => (output=good) (1)
6. (income==medium) & (seniority==big) & (delay==small) => (output=good) (1)
7. (income==big) & (seniority==small) & (delay==small) => (output=ok) (1)
8. (income==big) & (seniority==medium) & (delay==small) => (output=good) (1)
9. (income==big) & (seniority==big) & (delay==small) => (output=verygood) (1)
10. (income==verybig) & (seniority==small) & (delay==small) => (output=good) (1)
11. (income==verybig) & (seniority==medium) & (delay==small) => (output=verygood) (1)
12. (income==verybig) & (seniority==big) & (delay==small) => (output=verygood) (1)
13. (income==small) & (seniority==small) & (delay==medium) => (output=poor) (1)
14. (income==small) & (seniority==medium) & (delay==medium) => (output=ok) (1)
15. (income==small) & (seniority==big) & (delay==medium) => (output=ok) (1)
16. (income==medium) & (seniority==small) & (delay==medium) => (output=poor) (1)
17. (income==medium) & (seniority==medium) & (delay==medium) => (output=ok) (1)
18. (income==medium) & (seniority==big) & (delay==medium) => (output=ok) (1)
19. (income==big) & (seniority==small) & (delay==medium) => (output=ok) (1)
20. (income==big) & (seniority==medium) & (delay==medium) => (output=ok) (1)
21. (income==big) & (seniority==big) & (delay==medium) => (output=good) (1)
22. (income==verybig) & (seniority==small) & (delay==medium) => (output=ok) (1)
23. (income==verybig) & (seniority==medium) & (delay==medium) => (output=good) (1)
24. (income==verybig) & (seniority==big) & (delay==medium) => (output=good) (1)
25. (income==small) & (seniority==small) & (delay==big) => (output=poor) (1)
26. (income==small) & (seniority==medium) & (delay==big) => (output=poor) (1)
27. (income==small) & (seniority==big) & (delay==big) => (output=poor) (1)
28. (income==medium) & (seniority==small) & (delay==big) => (output=poor) (1)
29. (income==medium) & (seniority==medium) & (delay==big) => (output=ok) (1)
30. (income==medium) & (seniority==big) & (delay==big) => (output=ok) (1)
31. (income==big) & (seniority==small) & (delay==big) => (output=poor) (1)
32. (income==big) & (seniority==medium) & (delay==big) => (output=ok) (1)
33. (income==big) & (seniority==big) & (delay==big) => (output=ok) (1)
34. (income==verybig) & (seniority==small) & (delay==big) => (output=poor) (1)
35. (income==verybig) & (seniority==medium) & (delay==big) => (output=ok) (1)
36. (income==verybig) & (seniority==big) & (delay==big) => (output=good) (1)

Figure 4 graphically illustrates how to apply the inference for a considered input.

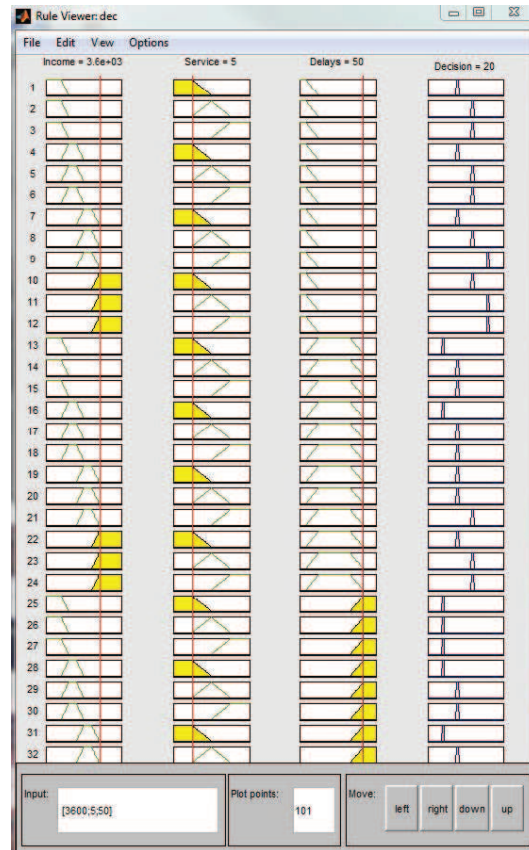


Figure 4:

2. The crisp algorithm - in this case study we use the classical approach of the situation awareness systems. In this case, customers will be divided into different segments, segments determined by relevant information gathered about them. By the data processing in relation to the minimum distance to the segments, the customer will be included in one of the segments. Below we present the situation awareness algorithm and an example for two-dimensional case, when we have two input information and have considered four segments (for examples with multiple inputs, the graphical representation is not possible).

3 The data processing algorithm

This algorithm falls into the classic situation of situations awareness systems. The stages covered for the implementation of the algorithm are:

- The normalization of the measured variables, i.e. bringing them into the $[0, 1]$ field;
- The weighting of the features based on a priori information regarding their importance: $\tilde{x} = w^T x$, in which the components of the vector are included in the $[0, 1]$ field;
- The definition of poles, $P_{ik}, k = \overline{1, m_i}$, which characterize each class $S_i, i = \overline{1, p}$, as well as some radial functions (potential functions) afferent to the poles:

$$K(\tilde{x}, p_k) \equiv \varphi_k(\tilde{x}) = \frac{1}{1 + \alpha d^2(\tilde{x}, p_k)} \quad (1)$$

where $d(\tilde{x}, p_k)$ represents the Euclidian distance between \tilde{x} and p_k ;

- The definition of the discriminant functions of the S_i segments as:

$$g_i(\tilde{x}) = \gamma_i^T \varphi(\tilde{x}) \quad (2)$$

In which

$$\varphi(\tilde{x}) = [\varphi_1(\tilde{x})\varphi_2(\tilde{x})\dots\varphi_M(\tilde{x})]^2 \quad (3)$$

and

$$\gamma_i = [\gamma_{i1}\gamma_{i2}\dots\gamma_{iM}]^T \quad (4)$$

is the representative vector of the S_i segment. The size of these vectors is $M = \sum_{i=1}^p m_i$;

- The decision rule regarding the affiliation of the vector \tilde{x} to the segments S_i is

$$Max_i = \{g_i(\tilde{x})\} = g_s(\tilde{x}) \Rightarrow \tilde{x} \in S_s \quad (5)$$

The representative (prototype) vector of a segment S_i can be established in two ways:

- *By imposing the components of the vector φ_i the values +1 or -1, depending on the affiliation, or the non-affiliation of the poles $p_{ik}, k = \overline{1, M}$ to the class S_i . For example, the S_1 class, the prototype vector has the form*

$$\gamma_i = [\underbrace{11 \dots 1}_{de\ m_1\ ori} \ \underbrace{-1 - 1 \dots - 1}_{de\ M-m_1\ ori}]^T \quad (6)$$

- *By training adjusting the vector φ_i . In this case, the prototype vectors are initialized according to the procedure shown in the previous case, initially noted as γ_i^0 . Training is required if, after fixing the prototype vectors, errors exist in the allocation of some vectors \tilde{x} to the segments adjacent to those they belong to in reality. To correct the boundary surfaces of the segments, the prototype vectors are adjusted by a training algorithm. In this case you need to have, along with the M vectors used initially as poles of segments, also a set of training data, consisting of vectors with known affiliation. Let N be the number of vectors in the training group. The training algorithm is usually*

$$\gamma_i^k = \gamma_i^{k-1} + \beta^k (\gamma_i(\tilde{x}^k) - \varphi^T(\tilde{x}^k) \gamma_i^{k-1}) \varphi(\tilde{x}^k), k = \overline{1, N} \quad (7)$$

in which k represents the current step of the training iteration, β^k establishes the training rate, and using $\varphi(\tilde{x}^k)$ we define the affiliation of the vector \tilde{x}^k , so:

$$y_i(\tilde{x}^k) = \begin{cases} +1, & \tilde{x}^k \in S_i; \\ -1, & \tilde{x}^k \notin S_i. \end{cases} \quad (8)$$

In the simplest version of the algorithm, the training rate is chosen as constant. In this case, the adoption of parameter β takes into account the conflicting requirements of convergence and stability.

A more efficient solution is to use β^k as:

$$\beta^k = \frac{1}{\|\varphi(\tilde{x}^k)\|^2} \quad (9)$$

which leads to the error correction algorithm.

The structure of the training system for the awareness of the situation corresponds to a RBF neural network type, emulated by the computer.

In order to use the situation awareness algorithm we consider the same example used for the fuzzy algorithm. They will be considered essential information for the customer classification, income and "seniority" in the retail

business, and will be considered 4 segments (classes), $C_j, j = \overline{1,4}$, out of which we can find the specific customer. For the training we will consider the data set shown in Figure 5. From the figure it can be noticed the obvious separation of the 4 segments considered. In order to cover the training procedure we

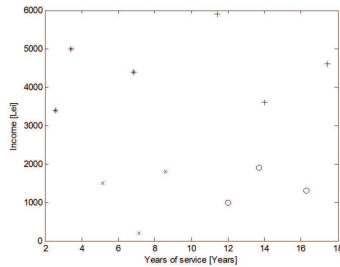


Figure 5: Data set used for training

scale the input vectors, income and "seniority" in the retail business, so that there will result two vectors with values present in the interval $[0,1]$. Following the training we obtain the affiliation of the input data set from the segments, as shown in Figure 6, the delimitation between the segments is presented in Figure 7.

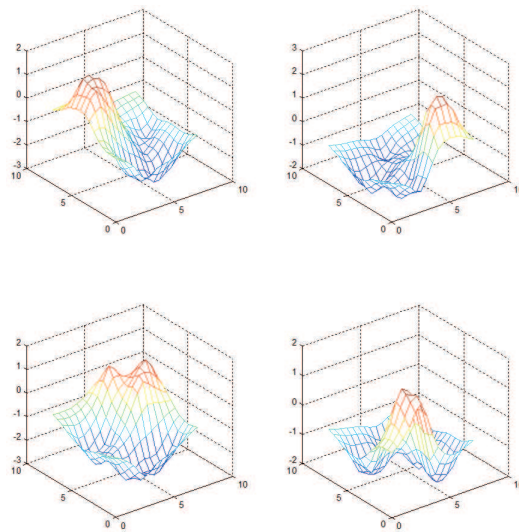


Figure 6: The affiliation of the data set to the 4 segments considered

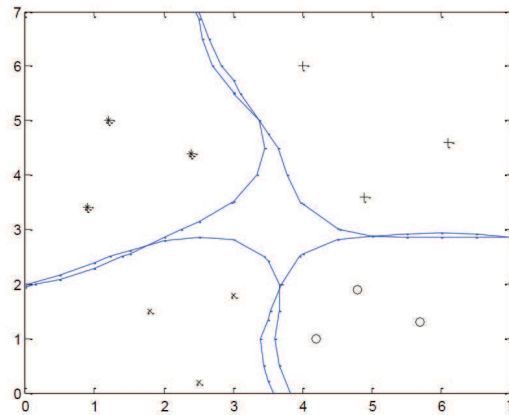


Figure 7: The partitioning of the space after the training related to the 4 segments considered

In order to test the recognition algorithms we considered the following input data:

Case 1: a customer with 6 years seniority and an income of 1000 RON. After scaling the following input vector results [2 1].

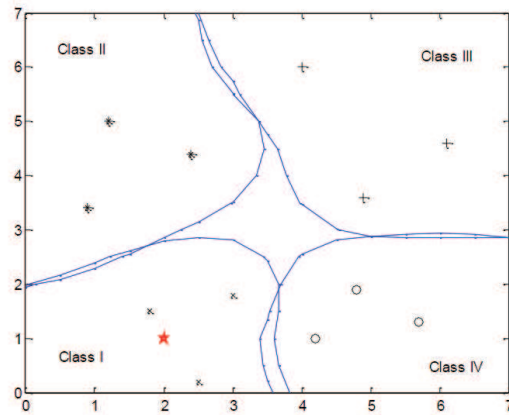


Figure 8: Situations recognition example

By applying the recognition algorithm the following discriminant functions of the segments results:

$$FP1 = 1.7236$$

$$FP2 = -1.6300$$

$$FP3 = -1.9141$$

$$FP4 = -1.4527$$

These values of the discriminant function show that the considered vector belongs to the Segment 1, corresponding to the maximum value of the discriminant function obtained. The situation was presented graphically in Figure 8, the test vector is represented by the red dots.

Case 2: a customer with 6 years seniority and an income of 5000 RON. After scaling the following input vector results [2 5]. By applying the recognition algorithm the following discriminant functions of the segments results:

$$FP1 = -1.4894$$

$$FP2 = 1.4989$$

$$FP3 = -1.2469$$

$$FP4 = -1.6231$$

These values of the discriminant function show that the considered vector belongs to the Segment 2, corresponding to the maximum value of the discriminant function obtained.

The situation was presented graphically in Figure 9, the test vector is represented by the red dots.

Case 3: a customer with 15 years seniority and an income of 55000 RON. After scaling the following input vector results [5 5]. By applying the recognition algorithm the following discriminant functions of the segments results:

$$FP1 = -1.1748$$

$$FP2 = -0.9700$$

$$FP3 = 0.9524$$

$$FP4 = -1.0148$$

These values of the discriminant function show that the considered vector belongs to the Segment 3, corresponding to the maximum value of the discriminant function obtained. The situation was presented graphically in Figure 10, the test vector is represented by the red dots.

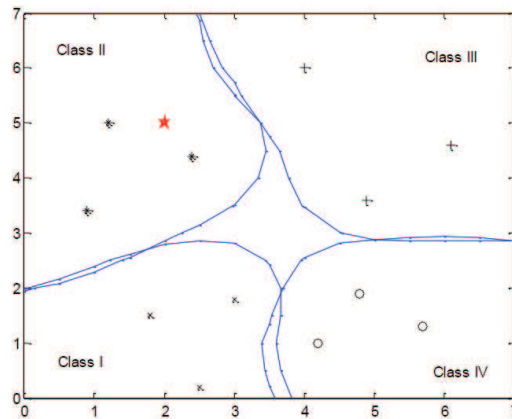


Figure 9: Situations recognition example

Case 4: a customer with 12 years seniority and an income of 2000 RON. After scaling the following input vector results [4 2]. By applying the recognition algorithm the following discriminant functions of the segments results:

$$FP1 = -0.6510$$

$$FP2 = -1.9309$$

$$FP3 = -1.6513$$

$$FP4 = 0.5668$$

These values of the discriminant function show that the considered vector belongs to the Segment 4, corresponding to the maximum value of the discriminant function obtained. The situation was presented graphically in Figure 11, the test vector is represented by the red dots.

4 Conclusion

The decision support systems represent a powerful software tool that can be used successfully in applications of marketing research, as this approach benefits from fuzzy logic to construct the models in order to choose the leading model.

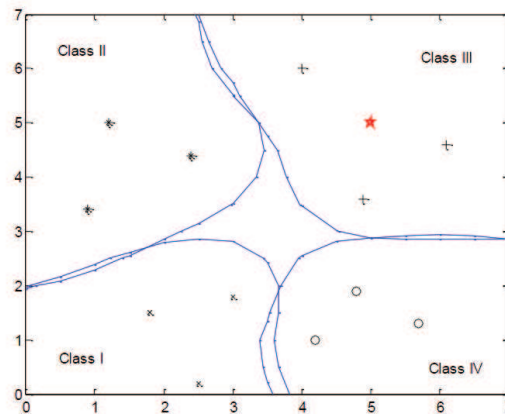


Figure 10: Situations recognition example

References

- [1] Agrawal, D. (1996) *Effect of Brand Loyalty on Advertising and Trade Promotions: A Game Theoretic Analysis with Empirical Evidence*, Marketing Science, 15, 86-108.
- [2] Araya, S., Silva, M., Weber, R. (2004) *A methodology for web usage mining and its application to target group identification*, Fuzzy sets and systems, 148(1), 139-152.
- [3] D'Urso, P., Gastaldi, T. (2002) *An "orderwise" polynomial regression procedure for fuzzy data*, Fuzzy Sets and Systems, 130(1), 1-19.
- [4] Hruschka, H. (1986) *Market definition and segmentation using fuzzy clustering methods*, International Journal of research in Marketing, 3(2), 117-134.
- [5] Kuo, R. J., Chi, S. C., Kao, S. S. (1999) *A decision support system for locating convenience store through fuzzy AHP*, Computers and Industrial Engineering, 37(1), 323-326.
- [6] Levy, J. B., Yoon, E. (1995) *Modeling global market entry decision by fuzzy logic with an application to country risk assessment*, European Journal of Operational Research, 82(1), 53-78.
- [7] Li, S. (2000) *The development of a hybrid intelligent system for developing marketing strategy*, Decision Support Systems, 27(4), 395-409.

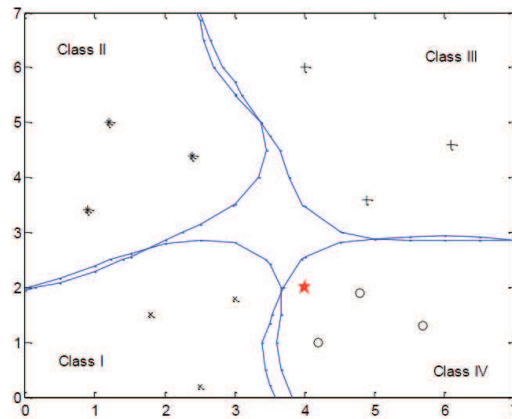


Figure 11: Situations recognition example

- [8] Mirkin, L. (2003) *On the extraction of dead-time controllers and estimators from delay-free parametrizations*, IEEE Transactions on Automatic Control, 48(4), 543-553.
- [9] Nauck, D., Klawonn, F., Kruse, R. (1997) *Foundations of neuro-fuzzy systems*, John Wiley and Sons, Inc.
- [10] Turksen, I. B., Willson, I. A. (1995) *A fuzzy set model for market share and preference prediction*. European Journal of Operational Research, 82(1), 39-52.
- [11] Zadeh, L. A. (1997) *Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic*, Fuzzy sets and systems, 90(2), 111-127.

“Ovidius” University of Constanta,
 Faculty of Economic Sciences,
 124 Mamaia Blvd., 900527 Constanta, Romania,
 e-mail: angelaelizamicu@yahoo.com

