# CLASSIFICATION OF CONSTRUCTION INFORMATION WITH FUZZY ATTRIBUTES

Özer Ciftcioglu
*Delft University of Technology, Faculty of Architecture*
*Building Technology, 2628 CR Delft, The Netherlands*
*o.ciftcioglu@bk.tudelft.nl*

## SUMMARY

Construction industry deals with various kinds of diverse information, which have to be synergistically handled. Information can be engineering data from exact sciences as well as linguistic or qualitative data from soft sciences. In such a wide spectrum to deal with the context dependent information optimally in perspective is a formidable task. The loss of benefit from available information is reflected on the cost effectiveness and efficiency of the construction. Referring to the complexity of this information, intelligent technologies can be of important help to assess information in a particular context in perspective for enhanced assessment of the construction process while it is in progress. In this respect, role of intelligent technologies, in particular, fuzzy logic, neural network and evolutionary search algorithms, in dealing with construction information is discussed and exemplified. In particular, in fuzzy logic terms, classification of construction information with fuzzy attributes/semantic labels is described.

## INTRODUCTION

The rich information environment of construction industry makes it a complex activity where many different individuals take part in it. The advances in the information and communication technology provide possibility of distributing information in a wide range of construction community in a cost-effective way with rapid dissemination. The advent of such desirable possibilities however urges the improvement of the existing means of information processing to cope with the information in appropriate proportions. Due to the size of complexity of information, the essential aid to deal with it may come from intelligent technologies since they are able to deal with complexity while they impose heavy demand on the computational power of the computers. Parallel to the advancements in the computer technology, there are considerable advancements in information acquisition and database management technologies. Due to this, data mining and knowledge discovery in databases is another emerging technology, which aims to maximize the gain of information utilization and knowledge acquisition. For construction industry presumably two important information-handling bottlenecks are *information ordering* and *information processing in perspective*. Information ordering engages in optimal distribution of information to individuals involved in the construction. That is, information deliveries to individuals in a right proportion as much as it is appropriate without information redundancy. Such a policy in information distribution provides efficiency and clarity in communication while the redundant information-load imposed on the construction community is kept to a minimum. Information processing in perspective deals with information duly so that any piece of information is handled with due attention in the context of concern without any inappropriate gravity attachment to it. The essential gain from information processing in perspective is the ability of impartial planning and decision-makings for cost effective and efficient construction. The work deals with this issue from the viewpoint of classifying construction information with fuzzy attributes in the terminology of fuzzy logic. Here, some more explanation of the construction information classification is in order. As the construction data is available in the form of attribute values of well-defined construction items' attributes, such a structure already conveys certain information to certain actors. In the present context classification refers to the categorization of the data with some semantic labels attached to the constructional attributes. Each of such combination forms a piece of information. In this respect, the categorization of the data refers to the classification of the associated piece of information. Such classification process is important, because precise description of each piece of information item requires a formidable computer storage on one hand and costly search process for a certain information, on the other. The classification can be accomplished in two different forms. It may be context dependent, that is, the fuzzy sets are allowed to change their position as well as context independent where fuzzy sets are fixed. Both may have different merits and in both cases, if the classification is done as fuzzy attribute values which is the case in this work, then, the classification is further simplified and the information becomes easier to access due to the efficiency of the information search algorithms. If necessary, there may be another database for more accurate and/or exact piece of information for situations requiring more matching precision.

The rest of the paper is organized as follows. In the following section, various soft-computing technologies are briefly described from the viewpoint of construction industry. This is followed by a section explaining information classification in perspective with example. Thereafter, a case study related to the method of information classification is presented. Finally, the conclusions follow.

## SOFT COMPUTING TECHNOLOGIES IN CONSTRUCTION

Soft computing is a keyword in information technologies. It is a concept of computation similar to that involved in fundamental human activities. From soft computing viewpoint three basic features in a human activity become conspicuous. These are

- parallel processing of information
- ability of processing vague, imprecise even incomplete information, and
- optimizing activities continuously based on some dynamic criteria which form the base for reasoning.

These three features can be simulated with soft computing methodologies to some extend with a remarkable success. Based on this success, soft computing is a potential information processing means to deal with exponentially increasing information-processing demand due to the advances in information and communication (IC) technologies. Three basic methodologies analogous to the features mentioned above are neural networks, fuzzy logic and evolutionary algorithms respectively. These methodologies have already established their associated technologies and they are briefly mentioned below.

### Artificial Neural Networks

Artificial neural networks (ANN) are systems [1] that can perform information processing in a manner similar to human brain. This is accomplished by means of a number of basic processors known as artificial neurons, which are inherently nonlinear, highly parallel, robust and fault tolerant. ANN can learn from examples and generalize the knowledge acquired to unknown tasks. In this way, they can make remarkably accurate predictions in that domain. Such identification and modeling through the use of available data is a considerable concern in all areas of construction. For instance, this spans structural engineering in one side, knowledge management on the other; dynamic system modeling on one side, expert system for decision support or fault diagnosis on the other, and so forth.

### Fuzzy Logic

Fuzzy logic explicitly aims to model the imprecise form of human reasoning and decision making. It is based on the concepts of fuzzy sets [2,3]. A fuzzy set is a generalization of a conventional set in that the memberships are assigned between 0 and 1 in contrast with being purely boolean. The fundamental concept of fuzzy logic is known as *linguistic variable*. A linguistic variable is a variable that takes values from spoken language. It can be described by

- qualitatively using an expression involving linguistic terms, and
- quantitatively using a corresponding membership function

A linguistic term is useful for communicating concepts and knowledge with human. A membership function is useful for processing numeric input data. Consider $x$ is some variable over some domain of discourse U called *universe of discourse* and X is a fuzzy set over U, the $\mu_X(x)$ is defined as the degree of membership of $x$ in $X$. The function $\mu_X = f(x)$ is referred to as membership function and it represents the degree of association of the variable $x$ with the fuzzy set $X$. Given a domain of discourse U, fuzzy sets over U can be identified with a set of names of linguistic variables. Typical fuzzy sets of speed of a car are shown in figure 1(a).

Considering the example of driving a car, such a variable can be assigned as *slow*, *medium* or *fast*. Although these values do not have precise meaning, a certain distribution between zero and one can be defined and associated with the values. This distribution is represented by a membership function. The membership functions are the fuzzy attributes or semantic labels of the quantity of concern. They are the basic elements of fuzzy computation. Fuzzy logic can be used in various ways. Some examples are *rule base* in an expert system, *control* in engineering systems, *information modeling* in a knowledge base, *semantic labeling* in a database and so forth. The essential machinery of fuzzy logic is the production of an output (consequent), based on given premises (antecedants) where in this process reasoning plays the major role. This is accomplished by means of statements, which are referred to as *rules*. The general expression for a set of fuzzy rules is

$$R^i : IF\ x_1\ is\ A_1^i\ and\ x_2\ is\ A_2^i\ \ldots.\ and\ x_n\ is\ A_m^i\ THEN\ y^j\ is\ B^i$$

where $R^i$ (i=1,2,…,l) denotes the i-th fuzzy rule. A basic example is illustrated in figure 1(b) where two properties related to two items in a construction process are considered. The items are door and window and the properties are the associated width and height. The fuzzy attribute values are represented by the associated membership functions. Figure 1(b) shows a fuzzy partition of the height (of the door) × width (of the window) space with fuzzy sets. Each division is represented by a similar fuzzy partition scheme where each triangle represents a fuzzy set. The shaded regions represent the overlapping areas of the fuzzy sets where each fuzzy variable, i.e. height and width is represented by two fuzzy sets. The universe of discourse is two-dimensional and it can be represented by four rules.

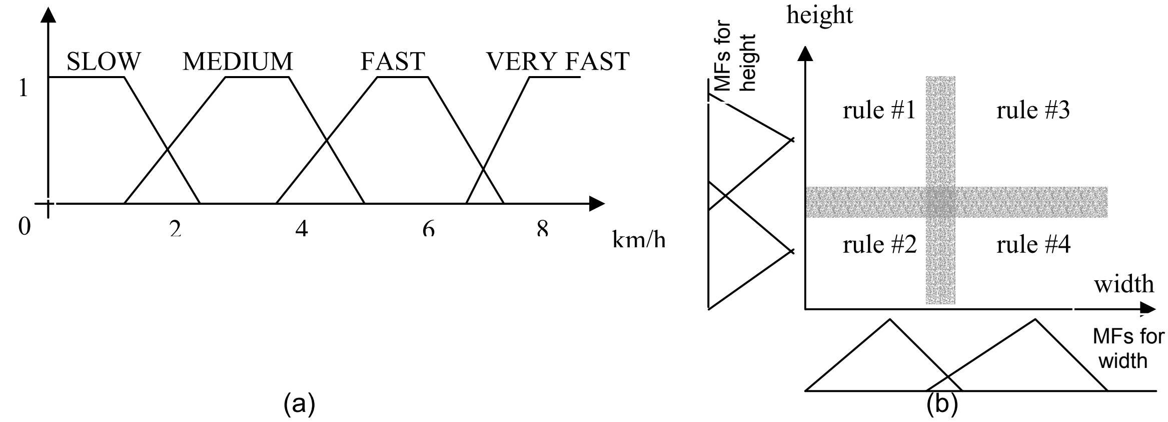


**Figure 1** Typical fuzzy sets of speed (a) and fuzzy memberhip functions for height and width (b)

## Genetic Algorithms

Genetic algorithms (GAs) are combinatorial optimization techniques inspired by the mechanism of evolution and natural genetics [4,5]. Their feature is the ability of parallel search of the state space for optimization in contrast with the point-by-point search by the conventional optimization. In GAs, the set of possible solutions of the optimization problem is called *population*. In the population any individual is a string of symbols or bits which are readily represented by a digital computer. The symbols are called *genes* and each string of genes is called *chromosome* so that each chromosome represents a solution. During the parallel search process, the validity of each solution individual of the population is graded by some criterion called *fitness*. The fitness evaluation marks the end of the iteration and according to the evaluation results a new population is formed. In a natural way, in the consecutive population there are more individuals inclined to acceptable solutions than the ones in the preceding population. The initial preparation of the population is done randomly if there is no information on the location of the solution. Otherwise, based on some available information, initial approximate representation of solutions to the problem as string of symbols is performed. By doing so, the search process is greatly facilitated. The ensuing population is prepared according to the genetic rules and implemented by means of genetic operators. These operators are basically *reproduction*, *crossover* and *mutation*. The genetic operators are applied on the selected parents to generate new possible solutions called *offspring*. Obviously, selected parents are the best representative candidates for a solution in the population. Afterwards, the individuals of the current population and its offspring form the new population. These basic concepts are implemented as follows.

We assume that, the initial population is created randomly at the time t=0. Let P(t) be the population at the time t and consisting of chromosomes the size of which is P as a constant scalar quantity. Crossover with an associated probability $P_C$ recombines two chromosomes by cutting them at a random position and partially exchanging the genes of the chromosomes. Mutation, with an associated probability $P_M$ changes the values of the some randomly selected genes. The "*fitness*" of each newly formed chromosome is evaluated and chromosomes with a low fitness score are replaced with those having high scores. This process is called *reproduction*. By the end of the reproduction a new population is formed that it becomes population P(t+1) at the time t+1. Note that the reproduction ensures that the population size is maintained as constant. The application of GA requires genetic parameters, namely, population size together with both crossover and mutation probabilities. Each of these greatly influences the performance of the GA. From the implementation viewpoint there is one important parameter, which is length of a string of bits representing each parameter to be optimized in binary form. For example in binary representation, if the length is equal to twelve, and the parameter range is between zero and unity then it means the resolution of this particular parameter identification is $2^{-12}$=1/4096. Below, is a pseudo code that represents the algorithm.

```
Start Program GA
        Read Genetic Parameters (popsize, Pm, Pc, length)
        Initialize P(t=0)
        Evaluate fitness of P(t=0)
           While ~(stop_conditions) do
                t=t+1
                Reproduction of P(t)
                Crossover P(t)
                Mutation P(t)
                Evaluate fitness of P(t)
           End
End
```

In the above code, each iteration defined by the *while loop* is called generation. Alternatively, in place of while loop, number of iterations can be given as a stopping criterion. Originally the computer implementation of GA is conveniently based on bits of a computer word so that a chromosome consists of a series of binary bits. The bits of each string are called *genes* and their varying values as *alleles*. This brief description conveys the message that evolutionary computation, in particular GAs can effectively handle many optimization problems of interest in construction industry. The areas can be rather diverse such as design of truss structures in structural engineering or optimal decision-making in construction management, etc.

## INFORMATION HANDLING WITH FUZZY CLASSIFICATION

### General Considerations

Based on the brief description of essential intelligent technologies for construction industry, in this section, classification of information in perspective will be described. Handling information in perspective is a concept where relevant construction information is considered as an entity of its own. The size of this entity depends on the information intended for use and it may encompass the total information available or a part of it. Depending on the size of information, a related information model is considered. In this model, fuzzy attributes are used. In this respect one can distinguish between *approximative approach* and *descriptive approach*. In the approximative approach fuzzy attributes can vary to some extent during the fuzzy model. In the descriptive approach the prescribed fuzzy sets are either not allowed to be modified or are permitted to have minor modifications. In the classification of information in perspective, the bounded amount of information implies that it is application dependent and therefore approximative approach is appropriate. In the literature, transformation methods from approximative to descriptive fuzzy classifiers are discussed [6]. The number of fuzzy attribute values may be in high numbers beyond the human comprehension due to human's short time memory at a time. The attribute values are determined during the formation of the model. To keep the fuzzy attributes rather moderate and consistent as to transparency, the sets can be aggregated [7-9]. Therefore attachment of the fuzzy attributes/linguistic labels and hence *a posteriori* interpretability may be necessary while this ought to be minimal. For construction industry, the information is essentially context dependent. In this respect the *knowledge in context* is an important concept [10]. This gives the possibility to consider the information in perspective while forming the fuzzy model [11]. In this consideration the fuzzy attributes of each aspect is determined relative to others in perspective. The range or nominal interval of each aspect is subdivided into appropriate sub-intervals by means clustering and point-wise projecting [12] to identify the center of full membership segment of this information. By the fuzzy modeling of construction information following gains are achieved.

- Semantic labels for the attribute are obtained for easy description, classification or use for information modeling, and
- Gravity points of the attributes are obtained as to the context dependent information. These are the areas where the membership functions take the maximum values. They are important local places within the nominal range of the attributes in the presence of total construction information being handled. This implies that the determination of the fuzzy attribute values is accomplished in perspective rather than independent assignments.

To illustrate this, consider figure 1(b) where total information as an entity has two attributes, namely height and width. The full membership locations in this illustration are the points corresponding to the maxima of the membership functions in the respective attributes. Note that information can be on any attribute of any construction component, like height of doors and width of windows as is the case in the present illustration. It can be qualitative as well as quantitative with qualities to be used as construction information and a semantic label can be attached presumably for convenience, in this example. However, the semantic labeling may be imperative in some other construction information employment.

The information about construction components is vastly available today thanks to construction sites on the internet. Basically, the construction items are defined as objects in the form of internet-based construction lexicon. Generically these objects are instantiated by taxonomizers via a taxonomy server [13]. A number of pieces of information on any construction item can be obtained from these sites and combination of these items in the context of concern form the information of interest as an information entity subject to analysis or process in a right perspective. From the instantiated information of the objects, any related joint information can be described by fuzzy attribute values. For instance, in very simple terms by figure 2, one can think of the total cost of the doors and windows in the whole construction project based on the given attributes of these items. In the meanwhile, in this bounded information it is important to realize the followings

- What is the functional relationship between the cost and the attributes, i.e. a fundamental model so that in a dynamic construction environment cost effectiveness is kept under control?
- What are the important sub-intervals of the attributes? Which fuzzy attribute values play the essential role on the basic model; that is, what are the center of full membership segments that play the essential role on the cost evaluation in above example? This question identifies the attribute values of natural demand.
- In a variety of choices, what attribute values should be selected for a cost-effective construction?

Referring to the basic example above, the fully satisfactory answers of above questions are necessary for a cost-effective construction. In a generic sense, these questions should be answered in a vast construction information environment and the answers can be added to the database as an easy generic reference in construction. To make this information generic enough, some categorical standardization of the construction information might be necessary. In a complex construction information environment, the answers for the questions above come from the soft computing methodologies. Broadly, neural network is used as to model realization, fuzzy logic is used as to important sub-intervals of the attributes and finally genetic algorithms are used for the identification of optimal attributes as to cost-effective construction. These are further explained below by means of the basic example considered in figure 1(b) where two attributes namely width and height are considered. We assume that we are given a data set of 25 points in a three dimensional space; namely the space of height, width and cost where all the ranges of the attributes are normalized to unity, that is, the range is between 0 and 1. For convenience, we can think of that the first two attributes are collected from a construction database and cost is computed from the construction project description at hand. The data set and the information model thereby established are shown in figure 2 (lower plot). Note that, by means of the model, predictions for the different composition of the attributes is now possible. The contour plot is also included in this figure (upper plot) pointing out that the cost has a minimum at the location where apparently both attributes have values about 0.7. However, this tentative observation is subject to blind search later by a combinatorial optimization method since the model is discrete. Here, such a basic model may serve as a metaphor for a complex multidimensional search space subject to a genetic search.

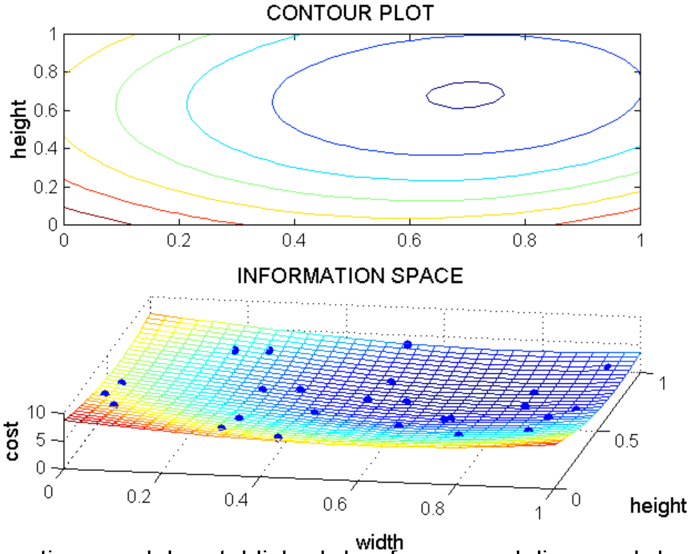


**Figure 2** Information model established by fuzzy modeling and based on given construction information as shown by dots (lower plot). The contour plot of the model is also shown (upper plot) where small ellipse indicates the solution region with cost optimality.

The fuzzy attribute values as membership functions attached to width and height underlying the model in figure 2 are obtained by clustering [14] and pointwise projection [12] where 3 clusters are assumed.

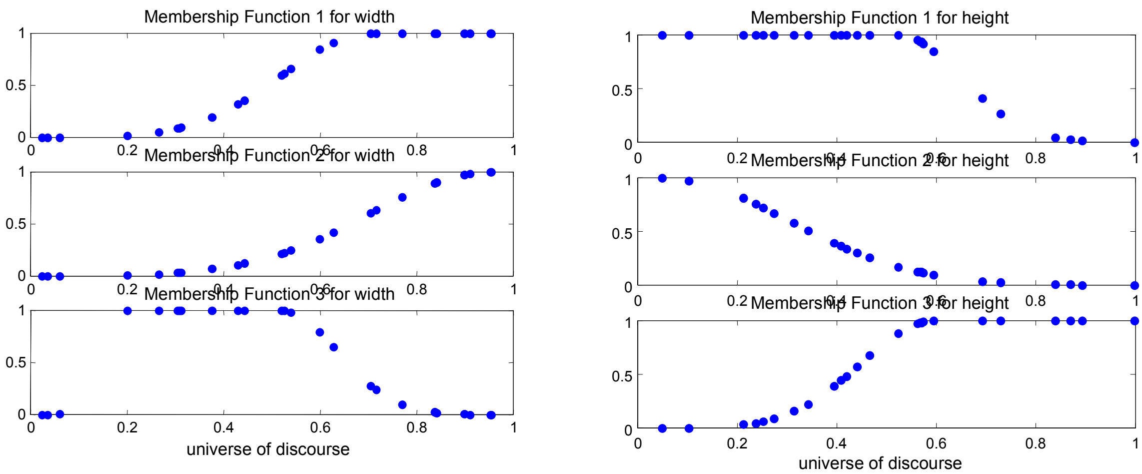


**Figure 3** Modeling with *width* and *height* attributes relating the attributes to cost

These are shown in figure 3. Based on the model established and shown in figure 2, one can search for the appropriate attribute values for cost effectiveness as pointed out before. The genetic search results are summarized in figure 4. In the contour plot, the trajectory of the genetic search is seen inside the smallest ellipsoidal area. The detailed progress of the genetic search is seen at the middle plot. The variation of the associated cost during the genetic search is shown in lowest plot. After 100 generations of the genetic search, the crisp attribute values for the minimized cost are obtained as 0.70 for the width and 0.68 for the height attributes. The cost for these attribute values is found to be 1.95 (monetary units).

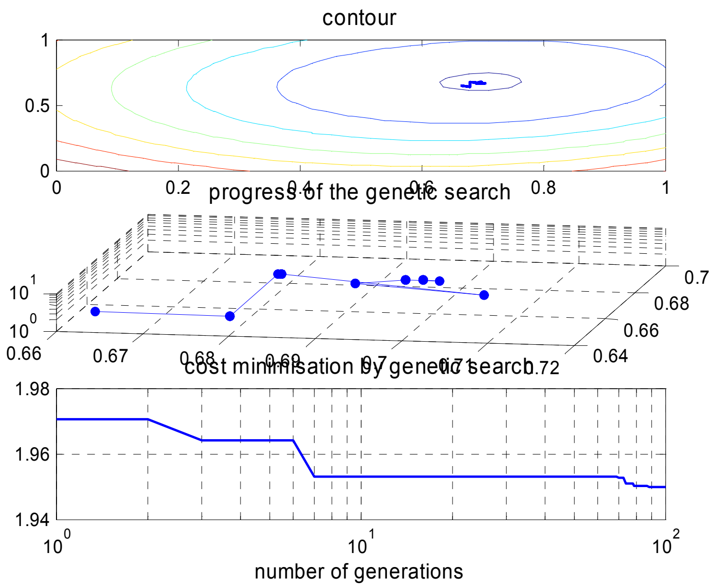


**Figure 4** A component width and a component height data connected to cost in construction (contour plot) and a genetic search for a minimal cost (middle and lower plot)

**A Case Study**

Next to the illustrative basic example given above, a study was made using a real-life data set from the database. The data set is connected to a metro station (Blaak) in Rotterdam, in the Netherlands. The details of the set are described in another research work [15]. In the present research four perceptional attributes from the data set are considered. These are *platform height*, *pleasantness of entrance*, *pleasantness of train platform* and *safety*. The fuzzy membership functions for the classification of the design information as fuzzy attributes in relation to safety are presented in figure 5. For clustering, initially 2 and later 4 centers are considered yielding two and four membership functions for each attribute, respectively. Each membership function comprises 60 points as result of the projection. Note that, each piece of information is formed with an attribute and categorized as attribute values. The categories are defined by fuzzy sets. The database is formed compiling the information from a questionnaire distributed to a number of targeted people. By means of appropriate transformation, all the answers are brought within 0 and 1. In the terminology of fuzzy logic, this interval is the universe of discourse. The fuzzy sets can be labeled as ‘small’, ‘medium’, ‘large’ and so on. In figure 5, it is important to note that, the shapes of the fuzzy sets are obtained from the data and therefore they are in a way optimal. The linguistic description of the attributes provides a natural way of describing the attributes with most conventional convenience in a design process. The fuzzy sets determined can be used in variety of different ways and contexts in construction, depending on the area of application in mind. In any case they serve as classified construction information and can be connected to specifications section in a web-based construction lexicon, for instance. In the present study, the fuzzy sets are used to form certain design rules. The rules form a model where the model relates the design variables, such as *platform height*, *pleasantness of entrance* and *pleasantness of train platform* to *safety*. This is a knowledge model since it possesses not only the information on the

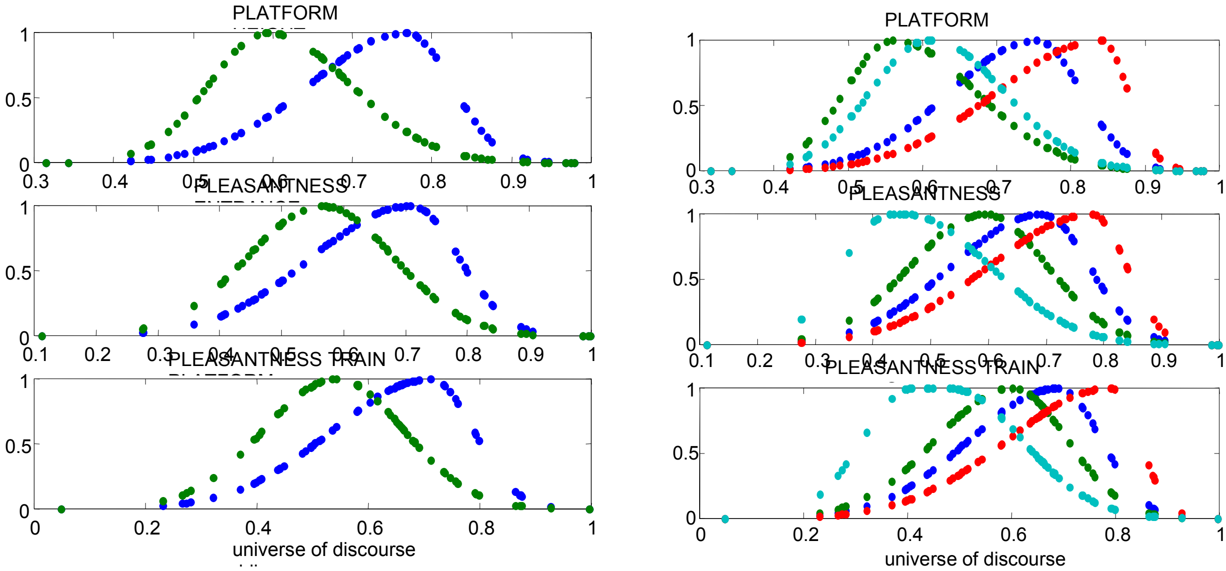


**Figure 5** Classified fuzzy attributes as fuzzy sets with two (left) and four (right) fuzzy membership functions for a data set related to perceptive safety of a metro station.

specific design variables, but also it provides probable complex interrelations expressed as rules among them. These rules may serve as basic design information or guidelines for prospective designs in the future. Due to the complexity, in some situations a machine learning process rather than expert driven rules should determine interrelations. Due to the same reason, in the original study [15] the machine learning process determines the rules. Namely, the number of input variables is 43 and the number of model output variables is 2. For such rich design environment, to accomplish a complete expert rules for a knowledge model practically is not feasible without machine learning. Note that, in figure 5, as the fuzzy sets are stipulated by the data, they are application dependent and therefore the fuzzy model is *approximative*. Therefore, the classification of construction information as a function of the related available information implies a relative classification of the fuzzy attributes. In contrast with this, for an absolute classification and correspondingly for a descriptive model, fuzzy set should be established by means of other means, elaborated experts' opinions, for instance. However, the consistency and accuracy of the model is hard to maintain, as the complexity of the model goes beyond certain proportions.

The knowledge model can provide a number of specific design related items by appropriate investigation of the model. For instance, it can provide strength of relations between each model input variable and each model output variable. That is any of *platform height*, *pleasantness of entrance* and *pleasantness of train platform* variables and the model output which is *safety* in this case. The design guidelines extracted from the model can be used in various ways as explained below.

- It may be used to make a general assessment of the existing metro station from the user viewpoint since the certain user perception aspects are known via the model. From this assessment, various improvements or modifications can be planned in order to enhance the effectiveness of the station for public use.
- The existing case study can serve as a basis for other similar designs pointing out the desirable and undesirable or unforeseen architectural and structural effects present in such designs.
- It may serve as a comprehensive case study material for design using the method of *case-based reasoning*
- The design guidelines may be used to meet certain optimality conditions in design. As this was exemplified in figure 2 for cost effectiveness.

## CONCLUSIONS

Classification of construction information with fuzzy attributes/semantic labels is described. Such classification is important as significant complementary information in the construction databases due to fuzzy logic implementations in constructions. Among these implementations mention may be made to

- development of rule based systems as decision support in construction,
- context dependent fuzzy attributes for construction databases by *approximative* fuzzy modeling or context independent fuzzy attributes by *descriptive* fuzzy modeling
- fuzzy modeling for knowledge acquisition from the construction information in the databases

and so on. Fuzzy logic is a part of soft computing which underlies the intelligent technologies. Therefore it ought to be incorporated with associated intelligent technologies in the context of soft computing. In this respect, the incorporation of the classification of construction information by fuzzy attributes into genetic algorithms is exemplified as to cost optimization by means of a demonstrative example. Referring to fuzzy modeling of construction information, the requirement for incorporating fuzzy logic with neural network technology is pointed out. Also, the classification with fuzzy attributes is presented for a real-life data set related to the perceptional design attributes of a metro station in the Netherlands. The utilization of the design guidelines extracted from the model is explained for enhanced construction design process.

## REFERENCES


[1] Hush D. R. and Horne B. G. (1993) Progress in Supervised Neural Networks, *IEEE Signal Processing Magazine*, pp.8-39, January

[2] Zadeh L. (1965), Fuzzy Sets. *Information and Control* **8**, pp.338-353

[3] Zadeh L. (1973) An Outline of a New Approach to the Analysis of Complex Systems and Decision Processes, *IEEE Trans. Systems Man and Cybernetics,* SMC-3, pp. 29-44

[4] Holland J.H. (1975) *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, MI

[5] Goldberg D. E. (1989) *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley Publishing Co. Inc., Reading, MA

[6] J. G. Marin-Blazquez and Q. Shen and A.F. Gomez-Skarmeta (2000) From Approximative to Descriptive Models, *9th IEEE Int. Conf. Fuzzy Systems*, pp.829-834, May

[7] Roubos J.A., Setnes M., and Abonyi J. (2000) Learning Fuzzy Classification Rules from Data in *Recent Advances in Soft Computing* (RASC2000), pp. 6-12, Leicester, UK,

[8] Setnes M., Babuska R. and Verbruggen H.B. (1998) Rule-Based Modeling: Precision and Transparency, *IEEE Trans. Systems, Man and Cybernetics-Part C: Applications and Reviews*, 28(1) pp.165-169, Feb.

[9] Setnes M., Babuska R. and Verbruggen H.B. (1998) Transparent Fuzzy Modeling, *International Journal of Human-Computer Studies*, 49(2), pp.159-179

[10] Fruchter R. (2002) Metaphors for Knowledge Capture, Sharing and Reuse, in *Proc. ECPPM 2002, eWork and eBusiness in Architecture, Engineering and Construction*, pp. 17-26, 9-11 Sep '02, Slovenia, Z. Turk and R. Scherer (eds)

[11] O. Ciftcioglu (2002) Ordering Rules and Complexity Reduction for Fuzzy Models, *Proc. NAFIPS 2002 (North Atlantic Fuzzy Information Processing Society),* pp. 535-540, June 27-29, New Orleans, LA,

[12] Kruse R., Gebhardt J. and Klawonn F. (1994) *Foundations of Fuzzy Systems*, Wiley, New York

[13] Woestenenk K. (2002) From LexiCon to XTD, in *Proc. ECPPM 2002, eWork and eBusiness in Architecture, Engineering and Construction*, pp. 387-391, 9-11 Sep '02, Slovenia, Z. Turk and R. Scherer (eds)

[14] Gustafson D. E. and Kessel W. C. (1979) Fuzzy clustering with a fuzzy covariance matrix, in *Proc. IEEE CDC*, San Diego, CA, pp.761-766

[15]Durmisevic S. (2002) Perception Aspects in Underground Space by Means of Intelligent Knowledge Modeling, PhD Thesis, Delft University of Technology, May