Design of a Fuzzy Classifier Network Based on Hierarchical Clustering

Arne-Jens Hempel, Steffen F. Bocklisch

Abstract— **In this article the issue of data based modeling is dealt with the help a network of uniform multivariate fuzzy classifiers. Within this framework the innovation consists in the specification of a hierarchical design strategy for such a network. Concretely, the two network specifying factors, namely the layout of the network structure and the classifier nodes configuration, will be addressed by a hierarchical clustering and selection strategy. The resulting network will represent the model's complexity in terms of interconnections between fuzzy classifier nodes whereas the model's vagueness and imprecision is captured within each fuzzy classifier node. Throughout the paper the network design and operation is illustrated with the help of an example.**

I. INTRODUCTION

Nowadays more and more complex and networked phenomena are subject to analysis. The goal of such an analysis is the creation of a model or the classification of the considered phenomenon. Basically there are two main philosophies to deduce such a model, theoretical and experimental modeling. When applying the latter it can be assumed that measurement data reflect the complexity of a phenomenon under consideration at least partially. Unfortunately, this data might also exhibit imprecision or depict interesting phenomena characteristics just vaguely. A possibly way to incorporate occurring imprecision is provided within the framework of fuzzy set theory [12]. Another important point is the detail of a data deduced model. On theone hand detailed modeling is costly, one the other hand one would favor the most general model. To circumvent this problem different layers of detail can be introduced. Coping with different levels of detail a network oriented representation has been proven promising [2]. A network of fuzzy pattern classifiers is such a combination of the network and the fuzzy set approach of modeling. Its basic design will presented section by section, starting with an introductive example.

II. INTRODUCTIVE EXAMPLE

In order to become acquainted with the purpose of this work, let us consider the following introductive example where the exemplary set of data depicted in Fig.1 forms the starting point. Obviously the 1200 objects of this data set contain more or less well separated, noisy data structures which could have been formed by an underlying phenomenon. The task at hand is to determine a model which captures the existing data structures respecting the data immanent uncertainties. The goal of such a model could be to assign unknown objects to their corresponding data structure. In general such a task is referred to as pattern recognition, and as [5] points out, there are a lot of sophisticated solutions for such a task. However [5] states as well that there is little research in structure describing fuzzy models. This work addresses the setup of such a fuzzy model.

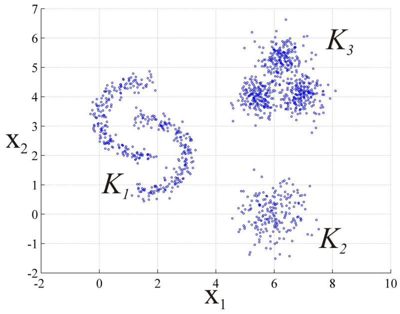

Fig. 1. Data basis of the example

Instead of applying a heavy mathematical ordnance the modeling approach featured here is guided by a more intuitive perception of patterns favoring comprehensibility. For the given data set consider a human observer. At a first glance the observer will recognize three major subsets (K_1, K_2 and K_3). After a second, more detailed analysis, the observer would state that the left subset emerges from two entangled halve circle shaped subsets, and the upper right subset consists of three minor subsets.

In terms of modeling the lines above can be paraphrased as follows: in a first step the observer sets up a coarse model encompassing the main features of the entire data set. In a second step, and only if it is necessary, the observer would incorporate further specification to the coarse model creating a new layer of more elaborated models. The result of the intuitive modeling process on the exemplary data set is depicted in Fig.2.

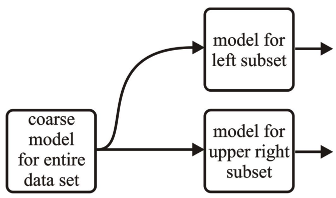

Fig. 2. Intuitive model for the example

Furthermore Fig.2 clarifies that new layers of detail lead

This work was not supported by any organization

Arne-Jens Hempel is with Faculty of Electrical Engineering, Chair of System Theory, Chemnitz University of Technology, 09107 Chemnitz, Germany hear@hrz.tu-chemnitz.de

Steffen F. Bocklisch is with the Department of Electrical Engineering, Chair of System Theory, Chemnitz University of Technology, 09107 Chemnitz, Germany steffen.bocklisch@etit.tu-chemnitz.de

to a further levels of hierarchy and consequently to a network of models. The advantage of such a network is its comprehensibility. Additionally such a network works in a human–like manner (from coarse to fine) even when classifying unknown objects. Thus granting a transparent and cost optimal classification process.

Due to its simplicity this intuitive approach will be the template for the automatic setup. Decomposing this intuitive approach in terms of an automated modeling task, two major obstacles arise: firstly to find suitable data inherent structures and secondly to appropriately model these structures.

III. Data Inherent Structures

A. Cluster Analysis

The first subtask to be dealt with is to locate data inherent structures, which is referred to as an unsupervised learning task [5]. As Fig.3 summarizes, a versatile and well researched access to such a type of learning is provided by methods of cluster analysis.

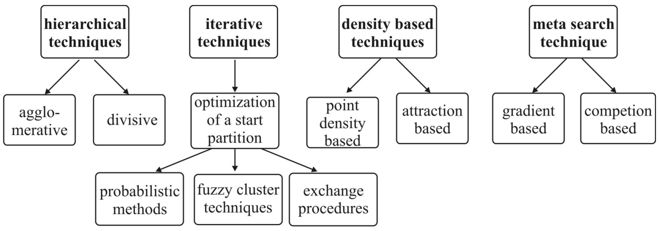

Fig. 3. Approaches to cluster analysis

As [11] points out each cluster method uses its specific strategy to discover such structures. Accordingly it is afflicted with drawbacks, for example some phenomena-typical structures remain indiscoverable by a certain method. Consequently, to overcome these drawbacks, at least an ensemble of sufficiently diverse cluster algorithms is applied when dealing with more complex data [1].

However, for the sake of clarity only the three agglomerative cluster algorithms (single,complete and wards linkage) will be considered in this paper, but, as a matter of principle, any other cluster algorithm is suitable. Background of this choice are the properties provided by the correspondent algorithms. Single linkage provides the chaining property, complete linkage provides compactness and Wards linkage is a variance criterion, providing noise resistance [7]. Moreover all three algorithms originate from the same cluster approach, facilitating their evaluation.

B. Cluster Evaluation

Since all three clustering algorithms are hierarchical agglomerative approaches applying different merging criteria they share the same kind of monotony, that is clusters grow monotone non–decreasing. This type of monotony can be expressed in terms of a distance measure (d_m), where every merging process (object to cluster, cluster to cluster) is associated with a new merging distance. The sequence of resulting merging distances can be interpreted as a notion of cluster stability, whereas cluster configuration associated with a long merging distance are referred to as stable and well separated configurations, and cluster configuration associated with a short merging distance are referred to as unstable and loose configurations. Hence studying the stability of emerging cluster configurations provides a formal cluster configuration assessment and selection criterion.

An easy understandable and yet illustrative way to gain an overview about the stability of all resulting cluster configurations is to map their corresponding merging distances in a so called dendrogram (see Fig.4). Since each cluster algorithm applies a different merging criteria the magnitude of the resulting merging distances varies significantly, rendering a quantitative comparison futile. The normalization (1) of the merging distances d_{mi} provides a circumvention to this obstacle. It results in quantitative comparable results (d_{ni}) as well as it preserves the qualitative relation, hence it will be used throughout the paper.

$$d_{ni} = \frac{d_{mi}}{\max\limits_{i} d_{mi}} \tag{1}$$

The index $i \in [1 \dots M]$ and M is equal to the number of merging processes during the clustering.
The exertion of all three cluster methods on the introductive example results in the normalized dendrograms according to Fig.4, with the single linkage dendrogram on the upper left hand side, complete linkage on the upper right hand side and the Wards linkage dendrogram on the lower left hand side. Considering the distances d_{ni} between emerging cluster configurations based on the dendrogram (Fig.4), it becomes apparent that, for all three cluster methods, the three–cluster configuration depicted on the lower right hand side is the most stable one, since it is associated with the longest relative merging distance. In concrete numbers the most stable configuration is obtained by single linkage with a maximal relative merging distance of 0.5032 (see Fig.4). Accordingly the single linkage cluster configuration can be taken as the most sensible choice for a structure proposal.

With the selection of a data inherent structure the first subtask has been completed and the second subtask can be focused on.

IV. Fuzzy Modeling

Given a data inherent structure (e.g. a cluster configuration), the second subtask is to determine an appropriate model. A desired model should be characterized by the following features: easy interpretability, explicitness/transparency, memory efficiency, and, as Fig.1 illustrates, it has to capture the imprecision and vagueness reflected in the given data set. There are plenty approaches to solve this task (e.g. Artificial Neural Networks [6], Bayesian Networks [8], probabilistic models, fuzzy set models [3], Support Vector Machines [10]), yet considering the attributes of a desired model, Artificial Neural transparency lack and Bayesian Networks lack explicitness, as well as Support Vector Machines lack interpretability. Consequently the remaining approaches are probabilistic models and the fuzzy

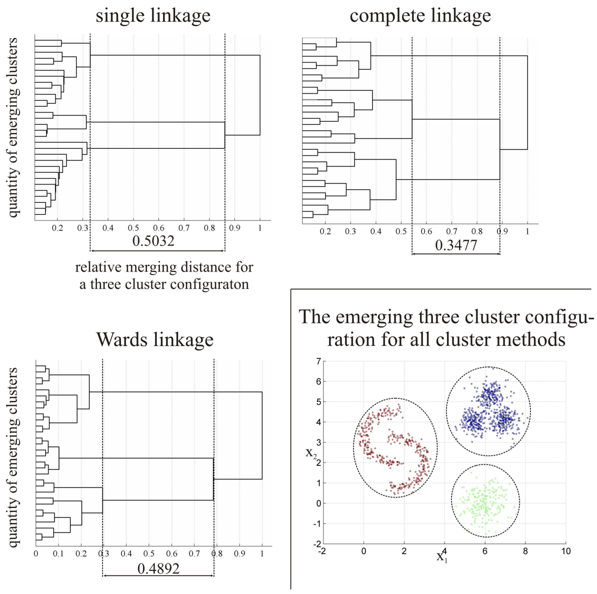

Fig. 4. Normalized dendrograms of the example

set approach. Since there is little prior information, and if additionally set-intern distributions are taken into account, a uniform probabilistic model is likely to be unavailable. On the contrary, a fuzzy set model results only from a description about the facts which are given/known, which in this case are the data at hand. It is therefore that the demands favor an approach applying fuzzy set theory.

As it is known [3] many fuzzy sets lend themselves to model data structures howsoever natured, throughout this paper the general fuzzy pattern class model introduced by [4] will be used. Besides meeting all above mentioned features it provides additional properties and methods which are crucial for the entire modeling process, as will be exposed later on. Due to its importance the fuzzy pattern class model will investigated closely in the following.

A. The Fuzzy Pattern Class Model

The multidimensional fuzzy pattern class A is expressed in every dimension of its individual class space by a parametrical function concept (as in Eq. 2) based on a set of seven parameters.

$$\mu^A(u, a, \vec{p}) = \begin{cases} \dfrac{a}{1 + \left(\dfrac{1}{b_l} - 1\right)\left(\dfrac{u}{c_l}\right)^{d_l}}, & u < 0 \\ \dfrac{a}{1 + \left(\dfrac{1}{b_r} - 1\right)\left(\dfrac{u}{c_r}\right)^{d_r}}, & u \geq 0 \end{cases} \tag{2}$$

Where the parameters denoted by a and $\vec{p} = (b_l, b_r, c_l, c_r, d_l, dr)$ possess the following meaning: The parameter a is representing the maximum value of the membership function μ^A. Regarding a whole class structure the parameter a expresses the weight of a specific class. a also embodies the topicality or authenticity of the information represented by that class. The parameter a is characterizing a whole class, whereas the parameters combined in $\vec{p}$ are related to a specific dimension of a fuzzy pattern class.

The parameters b_l, b_r of $\vec{p}$ assign left and right-sided membership values at the borders $u = c_l$ and $u = c_r$ for a normalized potential function ($a = 1$).

c_l, c_r characterize the left- and right-sided expansions of a fuzzy pattern class. Both parameters mark the range of a class in a crisp sense.

The parameters d_l, d_r specify the continuous decline of the membership function starting from the class center. d_l, d_r determine the shape of the membership function and hence the fuzziness of a class. Furthermore d_l, d_r are mapping the class internal distributions onto its shape. Fig. 5 summarizes the introduced concept of the potential type membership function considering a general one–dimensional example.

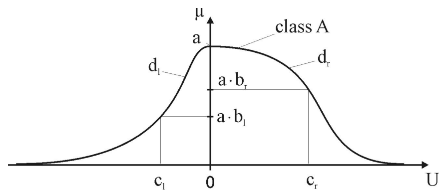

Fig. 5. Membership function and parameters.

Additionally the class describing set of parameters is supplemented by a position $\vec{u}_0$ and a class specific orientation $\vec{\varphi}$ in the original feature space. In order to obtain a multidimensional fuzzy pattern class A, the basis functions of each class dimension are connected using the n-fold compensatory Hamacher intersection operator (as in 3), with n denoting the index of the basis function.[9]

$$k_{Ham\cap}\, \mu_n^A = \frac{1}{\dfrac{1}{N}\displaystyle\sum_{n=1}^{N} \frac{1}{\mu_n^A}} \tag{3}$$

A result of this intersection is the conservation of the parametrical potential function concept as a multidimensional class description, hence it remains a same type model[9]. Since the introduced membership model is convex in nature, an adequate description of data set requires most likely a set of fuzzy pattern classes, (see Fig.8). Such a set of fuzzy pattern classes then forms a semantically and formally closed module, the so called fuzzy pattern classifier (FPC).

B. Fuzzy Pattern Classifier Setup

However a more intriguing question is, how the fuzzy pattern class model can be fitted to a proposed data structure. In the case of FPCs all class parameters can be assigned automatically by a two step aggregation procedure based upon a labeled learning data set. Exactly this prerequisite is fulfilled by data inherent structures resulting from cluster algorithms.

To generalise and facilitate the calculations, the crisp learning data set is extended to a set of fuzzy objects, using the introduced function concept (2).

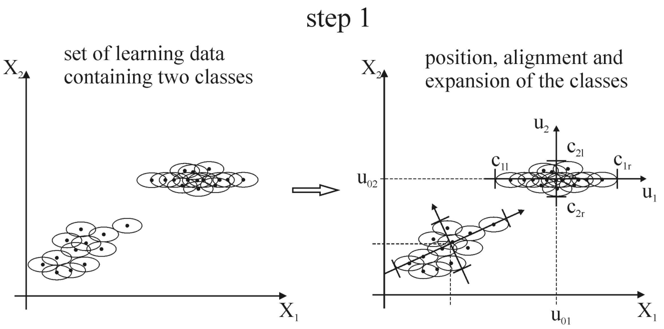

Fig. 6. Step 1: aggregation procedure

In the first step the class position $\vec{u}_0$, extension c_l, c_r and alignment φ are calculated.

As Fig. 6 exemplifies with a two–class example the position and alignment φ of each class is obtained by regression, and the extensions of the class are determined by the outmost objects.

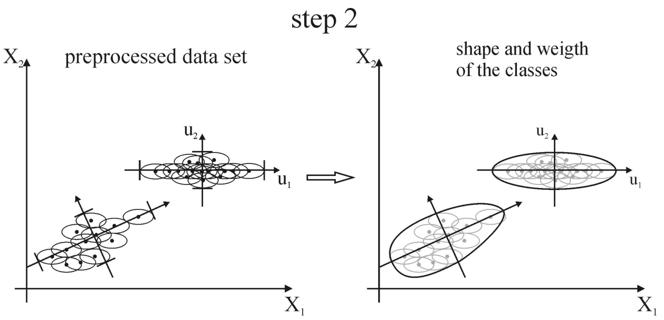

Fig. 7. Step 2: aggregation procedure

The class shape (d_l, d_r, b_l, b_r) and the "weight" (a) is specified in the second aggregation step. Based on the results of the first step the class shape is derived by the conservation of the object cardinality. The class "weight" is determined by the number of objects supporting the class [4] .

Fig. 8 illustrates the resulting FPC for the favored three class structure from the introductive example after applying the introduced aggregation procedure.

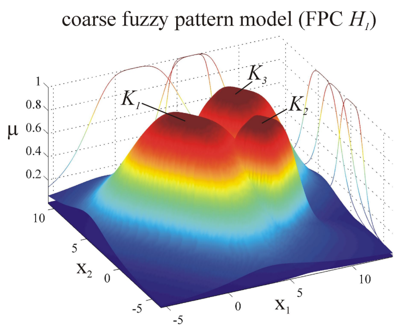

Fig. 8. FPC resulting from the automatic construction

The result of both subtasks (Fig.8) is a coarse FPC model H_1 of the introductive example, matching the required features.

V. Emergence of the Model Network

Remembering the intuitive approach, after recognizing three main subsets/groups in the first step, all subsets have been examined again for substructures in the second step, and, if it would have been necessary, further steps would have been carried on providing more and more elaborated models. Starting from the current point (that is the coarse fuzzy pattern model of the example), it is without any problem to stick with the intuitive idea. Everything has already been given to pioneer the next level of detail. The only thing to do is to treat the discovered subsets separately, but in the same manner as the entire data set. Concretely this means, based on the coarse three class fuzzy pattern model of the example, the subsets K_1,K_2 and K_3 are now subject to closer investigation.

In oder to avoid over complex and unreasonable models the modeling should be stopped when the emerging data inherent structures contain less than 5% of the entire data sets objects.

A. Detailed Analysis of Subsets

1) Analysis of Subset K_1: The reapplication of the entire treatment to the first subset K_1, reveals different data inherent structures for each cluster algorithm. Tab.I) summarizes the results with respect to the merging distances.

TABLE I

Maximum relative merging distances for the subset K_1

linkage	single	complete	Wards
max d_{ni}	0.4312	0.2558	0.4232
structure type	2 cluster	3 cluster	2 cluster

As highlighted, the usage of the maximum stability selection favors the single linkage two–cluster configuration, which captures the entangled halve circle structure due to its chaining property (see left hand side of Fig. 9). The subsequent construction of the fuzzy pattern model results in the FPC (H_{21}) pictured on the right hand side of Fig.9.

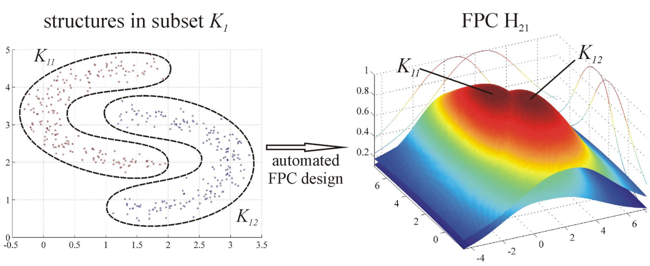

Fig. 9. Most stable data structure and resulting fuzzy pattern model

As mentioned earlier, FPC H_{21} constitutes a deeper insight of the underlying data structure in subset K_1, yet due to its convex nature the fuzzy pattern model is unable to capture the circle shape. On the one hand this insufficiency is partially compensated by the fuzziness of the model, on the other hand it reveals demands of further research.

A third iteration of analysis (for the subsets K_{11}, K_{12}) results in data structures with less than 5% of the entire data sets objects and therefore the analysis of K_1 stops at the second level of detail.

2) Detailed Analysis of Subset K_2: For the second subset K_2 the analysis stops at the first level of detail, since all the data structures induced in the second iteration contain less than 5% of the objects.

3) Detailed Analysis of Subset K_3: As Tab. II summarizes, the more detailed investigation of subset K_3 clearly quarries a three cluster configuration induced by Wards linkage. The corresponding peak out of the relative merging distance results from the variance motivation of the algorithm.

TABLE II

MAXIMUM RELATIVE MERGING DISTANCES FOR THE SUBSET K_3

linkage	single	complete	Wards
max d_{ni}	0.1589	0.2379	0.7012
structure type	4 cluster	3 cluster	3 cluster

It is this variance motivation that captures the intuitively perceived three cluster configuration, as Fig.10 illustrates. The associated fuzzy pattern model (FPC H_{22}) contains three classes. According to Fig.10 all these classes are highly overlapping around the center, leading to high degrees of class memberships for objects in this region. Exactly those high degrees of membership render the FPC H_{22} to be a more "natural" model in so far as a human observer would argue in the same manner since it uncertain to which cluster a object in the center region belongs to.

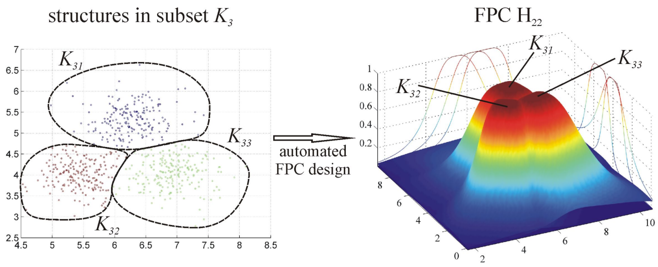

Fig. 10. Most stable data structure and resulting fuzzy pattern model

The continuative analysis (for the subsets K_{31}, K_{32} and K_{33}) results in data structures with less then 5 % of the objects and therefore the analysis of K_3 stops.

B. Fuzzy Model Network Composition

After finishing the closer examination of all subsets the iteration stops and the resulting sub-models can be assembled to the model network. Based on the level of detail, that is FPC model H_{22} specifies subset K_3, and FPC H_{21} substantiates subset K_1, the acquired fuzzy pattern models are arranged according to the setup in Fig.11. It becomes obvious that basic setup of the fuzzy model network matches the intuitively elaborated model of section II.

In summary the network design can be specified by the following course of actions:

1.) For the available hierarchical cluster algorithms derive a dendrogram with relative merging distances.

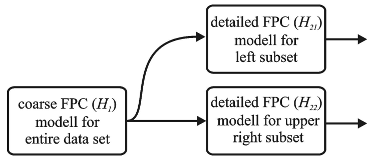

Fig. 11. Fuzzy model network based on hierarchical clustering

2.) Choose the most stable cluster configuration and create the corresponding fuzzy pattern classifier. If a structure contains less than a minimal number of objects stop the analysis and treat the next subset.

3.) Separate the training data according to its class label. Treat the each subset in the same manner (go back to 1.)).

4.) Connect each resulting classifier to its originating subset.

VI. FUZZY MODEL NETWORK OPERATION MODE

In order to provide a comprehensive overview the elaborated model network will be applied within the scope of the introductive example.

From a general point view Fig.11 clarifies that FPCs form the nodes of the model network or respectively constitute the functional core of the network. This means they determine the properties and capabilities of the network as well as the "‘signals"’ governing the network operation. Consequently, to become acquainted with the network operation, it is necessary to understand the operation of the FPC.

A. Fuzzy Pattern Classifier Operation

In operating mode the FPC classifies unknown objects using the class structure introduced before. The objects to be classified are denoted by a vector $\vec{x}$ of their features:

$$\vec{x} = (x_1, x_2, ...x_N)^T, \tag{4}$$

where N denotes the number of feature dimensions. The results of the classification process are denoted by a vector of sympathy $\vec{s}$, where the components of $\vec{s}$ express the membership of the classified object to the corresponding class:

$$\vec{s} = (s_1, s_2, ...s_K)^T, \tag{5}$$

where K is the total number of classes. The gradual membership of an object to a given class is calculated using (2).

$$s_k = \mu^k(\vec{x}) \quad \text{for } k = 1, 2, ..., K \tag{6}$$

Figure 12 illustrates the process of classification with the help of a one-dimensional three class structure. The object to be classified is situated in the right outskirt of the first class, in the center of class two but also in the left center of the third class. Alongside with the classification task the classification results are listed.

It becomes obvious that the vector of sympathy contains three membership values, describing a unique assignment of the object to the class structure with respect to its location in the feature space.

As it has been elaborated above, the signals available to govern the network operation are the features of an unknown

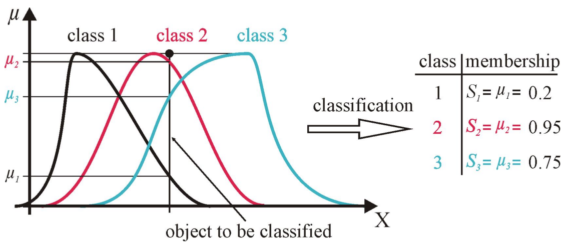

Fig. 12. Object classification

object as inputs and the memberships of the according classes as output. The most simple but yet most understandable way to govern the network operation is to route the signals exclusively based on their maximal membership. Since every sub–model or node originates from a specific subset, it will only be triggered if its corresponding sympathy is maximal.

B. Network Operation

The operation of the created model network is illustrated with the classification of the highlighted object in Fig. 13. It has been randomly selected from the learning data. According to the cluster analysis it is situated in between the classes K_{31} and K_{32} formed in the second level of the hierarchy.

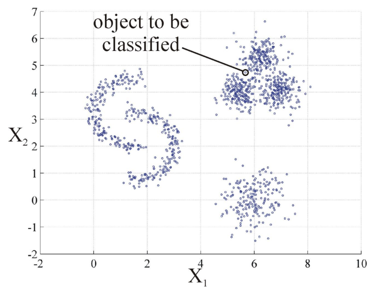

Fig. 13. Object to illustrate network operation

The test object is fed into the network via the data source block which is situated in the upper left corner. The results of the network operation are stored in separate output units forming the terminal nodes of the tree structure.

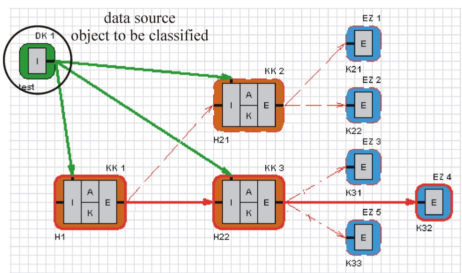

Fig. 14. Network operation for the highlighted object

The classification of the test object is highlighted with different types of arrows. Continuous lined arrows display the path actually taken, while the dashed arrows are representing all neglected options. For the test object the processing starts at the most abstract level, there it belongs with a membership of $\mu_{K_3} = 1$ to K_3, with $\mu_{K_1} = 0.1115$ to K_1 and with $\mu_{K_2} = 0.036$ to class K_2. Due to the maximum sympathy selection it proceeds to the classifier node H_{22} of the second level (see Fig. 14). This terminal node models the classes K_{31}, K_{32} and K_{33} and classifies the object with memberships of $\mu_{K_{31}} = 0.978$, $\mu_{K_{32}} = 0.968$ and $\mu_{K_{33}} = 0.853$.

VII. Summary and Conclusions

This paper presents a cluster analysis driven approach toward hierarchical data based modeling using a standardized model network. The main building block of the network is based on a multivariate and parametric classification concept (FPC). Considering the network of FPCs, the classifier concept provides the network nodes with a local and fuzzy model or knowledge base combined with the ability of fuzzy classification. In contrast to the local character of a single classifier node, the net–like interconnection of such nodes provides the following possibilities: structuring and combination of local models; scaling of the detail of a model;interpretative and intuitive decomposition and representation of complex tasks; integration of a priori or structural knowledge. It is, the instance that the discovery and selection of data inherent structures meets exactly the prerequisites for the automated FPC design which establishes the basic condition for a recursive and thus very intuitive automatic design strategy for networks of structure mapping fuzzy models.

Objects for future research are the development of a more sophisticated stopping criterion and the determination of a class description based cluster selection criterion. Another area of interest is the learning ability of such FPC networks.

References

[1] *Analysis of Clustering Algorithms for Web-Based Search*, London, UK, 2002. Springer-Verlag.

[2] Albert-Lszl Barabsi. *Linked: How Everything is Connected to Everything Else and What it Means for Business, Science, and Everyday Life*. Plume, NewYork, 2004.

[3] James C. Bezdek, James Keller, Raghu Krisnapuram, and Nikhil R. Pal. *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing (The Handbooks of Fuzzy Sets)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.

[4] Steffen F. Bocklisch. *Prozeßanalyse mit unscharfen Verfahren*. VEB Verlag Technik, Berlin, 1987.

[5] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification (2nd Edition)*. Wiley-Interscience, November 2000.

[6] Kevin Gurney. *An Introduction to Neural Networks*. Taylor & Francis, London and New York, 1997.

[7] A. K. Jain, M. N. Murty, and P. J. Flynn. Data clustering: a review. *ACM Computing Surveys*, 31(3):264–323, 1999.

[8] Judea Pearl. *Causality - Models, Reasoning, and Inference*. Cambridge Univ. Press, 2001.

[9] Ulrich Scheunert. *Neue Verfahren der Fuzzy− Mengen−Verknüpfung und Fuzz−Arithmetik und ihrer Anwendung bei der Sensor−Daten−Fusion*. PhD thesis, TU Chemnitz, 2001.

[10] B. Schölkopf and A. J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. The MIT Press, 2006.

[11] Benno Stein and Oliver Niggemann. On the nature of structure and its identification. In *WG '99: Proceedings of the 25th International Workshop on Graph-Theoretic Concepts in Computer Science*, pages 122–134, London, UK, 1999. Springer-Verlag.

[12] Lotfi Asker Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, 1965.