Similarity-Based Heterogeneous Neurons in the Context of General Observational Models

Valdés, Julio

National Research Council Canada
Institute for Information Technology

Conseil national de recherches Canada
Institut de technologie de l'information

NRC·CNRC

*Similarity-Based Heterogeneous Neurons in the Context of General Observational Models **

Valdés, J.
November 2002

* published in Neural Network World. Vol. 12, No. 5., pp. 499-508, November 2002. NRC 45822.

Canada

Similarity-Based Heterogeneous Neurons in the Context of General Observational Models

Julio J. Valdés

National Research Council of Canada
Institute for Information Technology
1200 Montreal Road, Ottawa ON K1A 0R6, Canada
julio.valdes@nrc.ca

Abstract. This paper presents a framework for processing heterogeneous information based on the construction of general observational domains, and similarity-based function calculi suitable for data mining in domains which can be described by the corresponding observational models. These calculi are intuitive, simple, and sufficiently general for classification and pattern recognition tasks. Functions in these calculi are represented by a particular kind of neuron models and their behavior is illustrated with examples from real-world domains showing their capabilities in processing heterogeneous, incomplete and fuzzy information.

keywords: general relational structures, heterogeneous data, similarity functions, heterogeneous neuron models.

1 Introduction

Processing heterogeneous information is a continuously growing need in many domains. Typical examples are monitoring complex systems (e.g. environmental, industrial, etc), information retrieval in multisource data basis (e.g. combined document and image search), etc. Moreover, advanced data mining operations requires processing huge masses of different types of data, all coming from a single given problem under investigation. Classification and prediction tasks are of most importance within the knowledge discovery process, but most data analysis methods work on single-type data or at most allows only very few types simultaneously. Limiting the number and/or the type of features describing the objects under study means that only a subset of the available information will be processed, leading to partial object description and important losses from the point of view of the quality of the data mining process. For example, in pattern recognition tasks, it is well known that the quantity and quality of the features used in constructing the recognition space crucial. However, discarding variables because of the inability of the used recognition method of handling features of their corresponding types might easily make the pattern recognition process fail. Clearly, similar situations arise with many other data mining techniques. All of these impose severe restrictions and hampers the opportunities to discover interesting and possible meaningful relationships, for not to mention the waste of those resources used in data acquisition, transmission, warehousing, etc. At the end, a lot of data either will not going to be used at all, or at most will be sub-used.

Constructing broader frameworks in which the nature and richness of the original information is preserved as much as possible is strongly needed for today's and tomorrow's observational problems. But as consequence, appropriate data mining techniques have to be developed for handling all heterogeneities, incompleteness, imprecision and data volumes involved. A framework for processing information of this kind is proposed here. It is based on the construction of general observational domains with similarity-based function calculi suitable for data mining in domains which can be described by the corresponding observational models. Functions are represented by specific neuron models with more elaborated constructions in the form of layered networks, in general hybrid. This approach combines the simplicity of a similarity based information processing with the adaptation, generalization and general function approximation capabilities of neural networks. It was initially proposed and tested in the context of classification problems [14], and subsequently extended and tested in different real world problems, including also theoretical studies [1], [15], [2], [16]). The aim of this paper is to put this research in the context of observational models within a soft-computing approach [22].

2 Heterogeneous Domains

According to the classical definition, a relational structure consists of a non-empty domain M and a set on relations $R_1, R_2, ..., R_n$ on M of various arities (notation: $\underline{M} =< M, R_1, R_2, ..., R_n >$). They have been used for constructing *semantic systems* like those used in the foundations of GUHA methods [7], and *information systems* (Düntsch and Orlowska in [10]). A natural generalization are V-valued structures. If V is an abstract set of values and $t =< t_1, t_2, ..., t_n >$ is a finite sequence of positive natural numbers (the *type*), a V-structure of type t is a tuple $\underline{M} =< M, f_1, f_2, ..., f_n >$, where M is the domain of $\underline{M}$ (non-empty) and each f_i is a mapping from M^{t_i} into V. In the case of both the *information systems* and the *observational model* used in the GUHA method [7], the domain is given by a set of objects, a finite non-empty set A of attributes (each $a \in A$ having a finite domain V_a), and an evaluation function f_a s.t. $f_a : M \to V_a$.

Models of interest with relevance for observational problems associated with complex processes, will be those composed by heterogeneous objects belonging to more general V-structures. In this case the attributes characterizing the objects will be given by different kinds of sets and the corresponding set of abstract values V_a for each attribute will not necessarily be finite. In observational predicate calculi, typical is the class of all $\{0, 1\}$-structures of type t whose domain is a finite set of natural numbers ($f_a(o) \in \{0, 1\}$, for all $a \in A$ and each $o \in M$). Usually all attributes are binary and therefore, models are binary matrices. For describing heterogeneous observational problems other models are required, in

this case, involving suitable mathematical description of the different *information sources* associated with the attributes (relations/ functions). These are coming from the physical nature of what is "observed" (e.g. point measurements, signals, documents, images, etc). They should be described by mathematical sets of the appropriate kind (called *source sets* and denoted by Ψ_i), constructed according to the nature of the information source to represent (e.g. point measurements of continuous variables by subsets of the reals in the appropriate ranges, structural information by directed graphs, etc). Models relevant to observational structures should account for incomplete information. Let $\mathtt{X}$ a special symbol having two basic properties: *i)* if $\mathtt{X} \in S$ (S being an arbitrary set) and f is any unary function defined on S, $f(\mathtt{X}) = \mathtt{X}$, and *ii)* $\mathtt{X}$ is an incomparable element w.r.t any ordering relation in any set to which it belongs.
A heterogeneous domain is defined as a cartesian product of a collection of source sets: $\hat{\mathcal{H}} = \Psi_1 \times \ldots \times \Psi_n$, where $n > 0$ is the number of information sources to consider. Projections and cylindric extensions are defined in the usual way.
As an example, consider the case of an heterogeneous domain where objects are characterized by attributes given by continuous crisp quantities, discrete features, fuzzy features, graphs and digital images. Let $\Re$ be the reals with the usual ordering, and $\mathcal{R} \subseteq \Re$. Now define $\hat{\mathcal{R}} = \mathcal{R} \cup \{\mathtt{X}\}$ and extend the ordering relation to a partial order accordingly. This source set may model point measurements of some variable, possibly with missing values (e.g. temperature readings). Let $\mathcal{N}$ be the set of natural numbers and consider a family of n_r sets ($n_r \in \mathcal{N}^+ = \mathcal{N} - \{0\}$) given by $\hat{\mathcal{R}}^{n_r} = \hat{\mathcal{R}}_1 \times \ldots \times \hat{\mathcal{R}}_{n_r}$ (n_r times) where each $\hat{\mathcal{R}}_j$ ($0 \le j \le n_r$) is constructed as above, and define $\hat{\mathcal{R}}^0 = \phi$ (the empty set). Now let $\mathcal{O}_j$, $1 \le j \le n_o$ be a family of finite sets with cardinalities $k_j^o \in \mathcal{N}^+$ respectively, composed by arbitrary elements, such that each set has a fully ordering relation $\le_{\mathcal{O}_j}$. Construct the sets $\hat{\mathcal{O}}_j = \mathcal{O}_j \cup \{\mathtt{X}\}$, and for each of them define a partial ordering $\hat{\le}_{\mathcal{O}_j}$ by extending $\le_{\mathcal{O}_j}$ according to the definition of $\mathtt{X}$. Analogously construct the set $\hat{\mathcal{O}}^{n_o} = \hat{\mathcal{O}}_1 \times \ldots \times \hat{\mathcal{O}}_{n_o}$ (n_o times and $\hat{\mathcal{O}}^0 = \phi$). For the special case of nominal variables, let $\mathcal{M}_j$, $1 \le j \le n_m$ ($n_m \in \mathcal{N}^+$) be a family of finite sets with cardinalities $k_j^m \in \mathcal{N}^+$ composed by arbitrary elements but such that no ordering relation is defined on any of the $\mathcal{M}_j$ sets. Now construct the sets $\hat{\mathcal{M}}_j = \mathcal{M}_j \cup \{\mathtt{X}\}$, and define $\hat{\mathcal{M}}^{n_m} = \hat{\mathcal{M}}_1 \times \ldots \times \hat{\mathcal{M}}_{n_m}$, ($n_m$ times and $\hat{\mathcal{M}}^0 = \phi$). Sets $\hat{\mathcal{O}}^{n_o}$, $\hat{\mathcal{M}}^{n_m}$ may represent the case of n_o ordinal variables, n_m nominal variables respectively (according to the statistical terminology). Similarly, a collection of n_f extended fuzzy sets $\hat{\mathcal{F}}_j$ ($1 \le j \le n_f$) , n_g extended graphs $\hat{\mathcal{G}}_j$ ($1 \le j \le n_g$) and n_i extended digital images $\hat{\mathcal{I}}_j$ ($1 \le j \le n_i$), can be used for constructing the corresponding cartesian products given by $\hat{\mathcal{F}}^{n_f}$, $\hat{\mathcal{G}}^{n_g}$ and $\hat{\mathcal{I}}^{n_i}$. The heterogeneous domain is given by $\hat{\mathcal{H}} = \hat{\mathcal{R}}^{n_r} \times \hat{\mathcal{O}}^{n_o} \times \hat{\mathcal{M}}^{n_m} \times \hat{\mathcal{F}}^{n_f} \times \hat{\mathcal{G}}^{n_g} \times \hat{\mathcal{I}}^{n_i}$. Elements of this domain will be objects $o \in \hat{\mathcal{H}}$ given by tuples of length $n = n_f + n_o + n_m + n_f + n_g + n_i$, with $n > 0$. Other kind of heterogeneous domains can be constructed in the same way, using the appropriate source sets.

3 Functions for Heterogeneous Domains Based on *H*-Neurons and Networks

Once the heterogeneous domain is defined, a natural next step would be the construction of appropriate *observational function calculi* [7]. The present discussion will focuss only on one of the steps in such process, namely the construction of functions relevant for data mining (also suitable for distributed computing). Later on they can be used within function calculii. These functions can be derived from general mappings like

$$f : \hat{\mathcal{H}} \to \mathcal{Y} \tag{1}$$

where $\mathcal{Y}$ is an abstract set.

Consider the mappings given by $h : \hat{\mathcal{H}} \times \hat{\mathcal{H}} \to \mathcal{Y}$. Accordingly, if $x, w \in \hat{\mathcal{H}}$ and $y \in \mathcal{Y}$, then $y = h(x, w)$. Now, if w is a fixed parametrizing element, then h will be compliant with (1). If $\mathcal{W} \subseteq \hat{\mathcal{H}}$ is a finite non-empty subset (with cardinality p), having elements $\{w_1, \ldots, w_p\}$, they can be used for parametrizing a corresponding collection $\{h_1, \ldots, h_p\}$ of h-mappings.

Parametrized mappings of this kind were originally introduced in [14] and called *heterogeneous neurons* (h-neurons). Let a h-neuron with parameter w be denoted by h^w. Now, take a fixed collection $\{h_1^{w_1}, \ldots, h_p^{w_p}\}$ of p h-neurons, with their corresponding parameters $\mathcal{W} = \{w_1, \ldots, w_p\}$ and having images in the abstract sets $\{\mathcal{Y}_1, \ldots, \mathcal{Y}_p\}$, such that $y_i \in \mathcal{Y}_j (1 \leq j \leq p)$. Consider the mapping $H^{\mathcal{W}} : \hat{\mathcal{H}} \to \hat{\mathcal{Y}}$ (parametrized by $\mathcal{W}$), where $\hat{\mathcal{Y}} = \mathcal{Y}_1 \times \ldots \times \mathcal{Y}_p$ and with the $H^{\mathcal{W}}$ function defined in the following way: $H^{\mathcal{W}}(x) =< h_1^{w_1}(x), \ldots, h_p^{w_p}(x) >=< y_1, \ldots, y_p >$. Call it $H^{\mathcal{W}}$*-layer*.
Now take a collection of q such layers ($q \in \mathcal{N}^+$) $\{H_i^{\mathcal{W}_i} | (1 \leq i \leq q)\}$ such that for all $(2 \leq i \leq q)$, $Dom(H_i^{\mathcal{W}_i}) = Im(H_{i-1}^{\mathcal{W}_{i-1}})$, where f is an arbitrary function and $Dom(f)$ and $Im(f)$ are functions giving the domain and image of f respectively (note that each $H_i^{\mathcal{W}_i}$ may have a different number of h^w neurons). Then, the composition $H_i^{\mathcal{W}_i} = H_1^{\mathcal{W}_1} \circ \ldots \circ H_q^{\mathcal{W}_q}$ is called a *q-layers H-neural network.*

Of particular interest are those heterogeneous neurons based on *similarity functions* [4]. These type of neurons have been studied from both theoretical and practical points of view (e.g. [14], [15], [2], [16]). Let S^0 be a similarity function with range $[0, 1]$ with the classical axioms. For the type of calculi of interest in heterogeneous domains, an extra axiom is added for accounting with incomplete information. The *extended* similarity function S is defined as $S(x, y) = \mathtt{X}$ if ($x = \mathtt{X}$ or $y = \mathtt{X}$), and $S(x, y) = S^0(x, y)$ otherwise. Thus $S : (\hat{\mathcal{H}} \times \hat{\mathcal{H}}) \to [0, 1] \cup \{\mathtt{X}\}$.

There are many ways in which the similarity function S for a heterogeneous domain $\hat{\mathcal{H}}$ can be constructed [2]. The easiest one is by having n *partial similarities* $\{s_1, \ldots s_n\}$ defined on each of n the source sets included in the heterogeneous domain ($s_j : \Psi_j \times \Psi_j \to [0, 1] \cup \{\mathtt{X}\}$), $(1 \leq j \leq n)$, and a direct product operation $\hat{s} : \hat{\mathcal{H}} \to ([0, 1] \cup \{\mathtt{X}\})^n$ constructed in the usual way from the s_j. Take an order preserving mapping $\Theta : ([0, 1] \cup [\mathtt{X}\})^n \to ([0, 1] \cup \{\mathtt{X}\})$ (called an *aggregator operator*) and now define S as the composition $S = \Theta \circ \hat{s}$. Weighted additive

measures are typical aggregator operators. In order to have a greater degree of generality, an additional unary isotone mapping $g : ([0,1] \cup \{\mathtt{X}\}) \rightarrow ([0,1] \cup \{\mathtt{X}\})$ is included in the composition. In general g will be a non-linear function (even the identity mapping) allowing control of the distribution of the similarities in the $[0,1]$ interval. Thus $h = g \circ S = g \circ (\Theta \circ \hat{s})$.
The following is an example of a g function (called *v-sigmoid*).

$$g(x,k) = \begin{cases} (\frac{-k}{(x-0.5)-a(k)} - a(k) & \text{if } x \leq 0.5 \\ (\frac{-k}{(x-0.5)+a(k)} + a(k) + 1 & \text{otherwise} \end{cases}$$

where $a(k)$ is an auxiliary function given by

$$a(k) = \frac{-0.5 + \sqrt{0.5^2 + 4 * k}}{2}$$

and k is a real-valued parameter controlling the curvature. Fig-1 shows g for several values of k.

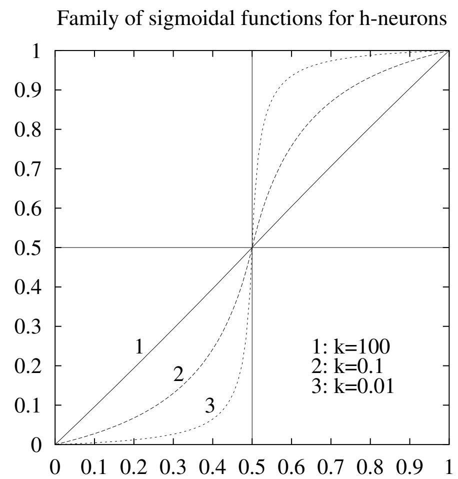

Fig. 1. Family of V-sigmoid functions for different values of k. Any of these functions can be used as a g function for the non-linear control of similarities in the definition of the h mapping.

The importance of networks constructed with similarity-based h-neurons is that they exhibit a general approximation property [2]. Thus, they can be applied to classification, pattern recognition and other data mining operations in heterogeneous domains. Their quality can be evaluated in terms of different error measures which can be constructed as generalized quantifiers over observational structures with heterogeneous domains defined as above. Within this approach, in order to construct the corresponding specific function calculi, appropriate similarity functions must be chosen for operating on the different source sets considered in the heterogeneous domain associated with the problem. Besides the classical works dealing with ratio, interval, nominal and ordinal type of variables (in the statistical sense), extensive work has been done in the study of similarity functions for other kind of information sources. Just to mention a few of these sources, see for example, [5], [11], [20] for fuzzy sets, [21], [3], [17] for graphs, [13], [19], [18] for digital images, [6], [8] for documents. Clearly, many other kind of information sources can be considered as well. Within this approach, any information source can be included in an heterogeneous observational structure, as long as there are similarity functions defined for objects from the corresponding source sets.

4 Application Examples

The resulting calculi using *H-neural networks* provides a flexible way for problem solving in complex heterogeneous domains. Its neural network nature makes it suitable for parallel and distributed implementations leading to high performance computing capabilities. This feature is crucial in many real world problems, and in order to illustrate some of its properties, several application examples are presented.

4.1 Classification with crisp vs fuzzy values

Schemes based on *H-neural networks* for classification problems have been applied successfully in different domains. The analysis of hydrochemical data the Artic (Spitzbergen) [1], showed that similarity based *H-neural networks* working with complete fuzzy neurons over fuzzy data values performed better than classical feed-forward models applied to the original data. In other words, when the observations were interpreted as imprecise information (i.e. what they really are), results were better than when they were processed as the crisp values coming from the data table (i.e. what is always done). For single layer architectures, they found that average mean squared error (MSE) for networks using classical neurons was 0.166, whereas for the heterogeneous network was 0.092. In both cases 5 neurons were used in a single layer architecture.

4.2 Influence of missing values

The behavior of H-networks in classification problems with increasing presence of missing data, is illustrated with results shown in Fig-2 [14], using Iris data (a classical data set in pattern recognition). The problem domain is classification of species of a given flower type and the data set contains 150 observations and 4 variables. They are all numeric, originally with no missing values [12]. In this case classification performance of the H-network is compared with that of the k-nearest neighbor rule. The k-nn rule is rapidly affected by the increase of missing information, whereas the H-network suffers much less. The effect of the dilution of the test set on the network is higher after the 50% barrier. Also note that, network predictions are over 80% with 50% of missing values in both the training and the test sets. Moreover, even with 90% of missing values in the training set and 70% in the test (extremely high dilution), classification accuracy is still above 50%.

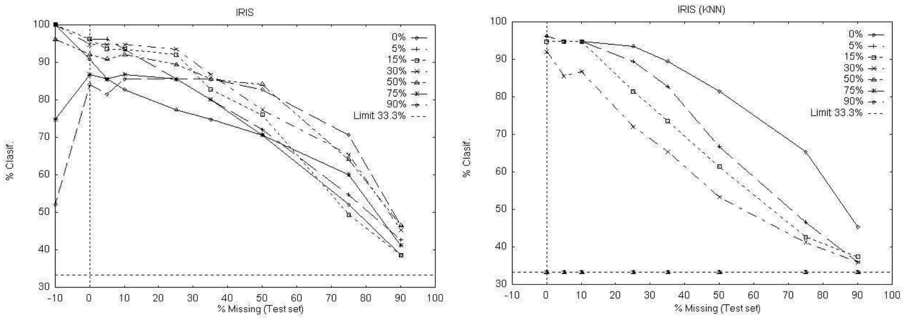

Fig. 2. Classification accuracy (test set), with increasing amount of missing data in the training set for Iris data.

4.3 Model mining in time series analysis

An application oriented to model discovery in multivariate time series allowing heterogeneous variables, fuzzy values and missing information was developed in [16]. In this system a genetic algorithm evolves entire similarity-based hybrid neural networks for discovering patterns of dependency between past values in all time series and future values in a target series. A functional representation using the same type of network is constructed for the best model(s) found, which can be used for forecasting purposes, is also obtained. Fig-3 shows the behavior of a prediction of American relative sunspot numbers (mean number of sunspots for the corresponding months in the period $1/1945 - 12/1994$), from AAVSO - Solar Division [9]. A total of 100000 models were constructed and explored by the algorithm, with their corresponding neuro-fuzzy networks. The best model found corresponds to the pattern of dependencies given by the lags $(t-1), (t-2), (t-4), (t-10), (t-12), (t-14), (t-16), (t-20), (t-28), (t-29)$, with a mean squared error of 20.45. An idea of the efficiency in the model discovery process with this approach is given by the time required to extract the model: in a Pentium III-866 Mhz PC it was 6 min 5 secs.

5 Conclusions

Relational structures can be used as a base for constructing observational domains and models suitable for describing heterogeneous, incomplete, imprecise

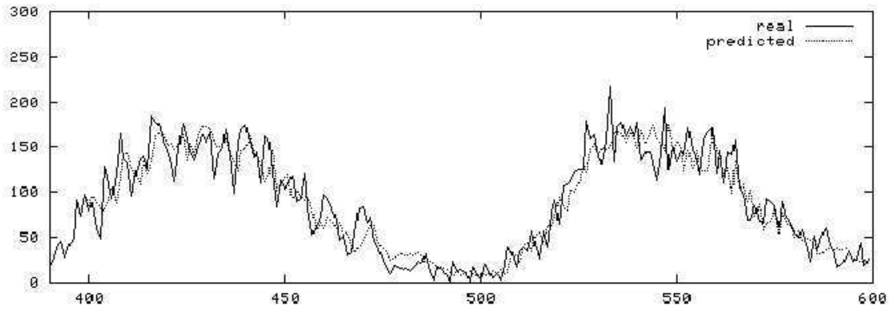

Fig. 3. Comparison of the real and predicted values for sunspot data (test set). 100000 similarity-based neuro-fuzzy networks were constructed and evaluated in 6 min 5" secs using a PIII-866 Mhz PC. Best model found relates future values at time t with past values at lags (t-1), (t-2), (t-4), (t-10), (t-12), (t-14), (t-16), (t-20), (t-28), (t-29). Mean squared error = 20.45.

and time-dependent data. Different kinds of similarity based functions are relatively simple to build and relevant for data mining on these data. Their representation by means of h-neurons and H-neural networks provides a broad and flexible way of constructing algorithms for different data mining tasks, which can be efficiently exploited in the form of parallel and distributed implementations. Real world application examples showed that these algorithms have good problem solving performance (sometimes even better than their classical counterparts), moreover tolerating imprecision and being notably robust in the presence of missing information. Results obtained with this approach looks promising, but more theoretical and applied investigation must be made in order to have a better knowledge of their properties, behavior, applicability and limitations. In the same sense, the construction of some logical calculi in the spirit of [7] using the notions described above seems promising and could be valuable data mining tools for highly heterogeneous domains.

References

1. Belanche, Ll., Valdés, J.J. : Fuzzy Inputs and Missing Data in Similarity Based Heterogeneous Neural Networks. Proc. Int. Work-Conf. on Artificial and Natural Neural Networks, IWANN'99. Engineering Applications of Bio-Inspired Artificial Neural Networks. Lecture Notes in Computer Science LNCS-1607. Springer Verlag (1999), 863-873
2. Belanche, Ll. : Heterogeneous neural networks: Theory and applications. PhD Thesis, Department of Languages and Informatic Systems, Polytechnic University of Catalonia, Barcelona, Spain, July, (2000)
3. Bunke, H., Meemer, B.: Similarity measures for structures representations. First European Workshop on Case-Based Reasoning, (1993), 106-118

4. Chandon, J.L., Pinson, S. : Analyse Typologique. Thorie et Applications. Masson, Paris, (1981)
5. Fonck, P., Fodor, J., Roubens, M. : An application of aggregation procedures to the definition of measures of similarity between fuzzy sets. Fuzzy Sets and Systems, **97** (3), (1998), 67-74
6. Gaussier, E., Goutte, C., Popat, K., Chen, F.: A hierarchical model for clustering and categorizing documents. In F. Crestani and C.J van Rijsbergen (eds) ECIR 2002, Lecture Notes in Computer Science LNCS, (2002), 229-247
7. Hajek, P., Havranek, T. : Mechanizing Hypothesis Formation. Springer Verlag, (1978)
8. Losada, D.E, Barreiro, A.: A logical model for information retrieval based on propositional logic and belief revision. The Computer Journal, **44**, (5), (2001), 410-424
9. Masters, T. : Neural, Novel & Hybrid Algorithms for Time Series Prediction. John Wiley & Sons, (1995)
10. Orlowska, E., Szalas, A. (ed.): Relational methods in computer science applications. Springer-Physica Verlag, (2001), 263-285
11. Pappis, C.P., Karacapilidis, N.I: A comparative assessment of measures of similarity of fuzzy values. Fuzzy Sets and Systems, **56**, (1993), 171-174
12. Prechelt, L.: Proben1 A Set of Neural Network Benchmark Problems and Benchmarking Rules. Fakultt fr Informatik. Universitt Karlsruhe, Germany. Technical Report 21/94. ftp://ftp.ira.uka:/pub/neuron, (1994)
13. Rubner, Y., Tomasi, C., Guibas, L.J. : Adaptive color image embeddings for database navigation. Proc. IEEE Asian Conf. on Computer Vision, (1998) 105-111
14. Valdés, J.J., García, R. : A model for heterogeneous neurons and its use in configuring neural networks for classification problems. Proc. IWANN'97, Int. Conf. On Artificial and Natural Neural Networks. Lecture Notes in Computer Science **1240**, Springer Verlag, (1997), 237–246
15. Valdés, J.J., Belanche, Ll., Alquézar, R. : Fuzzy heterogeneous neurons for imprecise classification problems. Int. Jour. Of Intelligent Systems, **15** (3), (2000), 265–276
16. Valdés, J.J. : Similarity-based Neuro-Fuzzy Networks and Genetic Algorithms in Time Series Models Discovery. NRC/ERB-1093, 9 pp. NRC 44919. (2002)
17. Valtchev, P., Euzenat, J.: Dissimilarity measure for collection of objects and values. In P. Cohen, X.Liu and M. Bertold, eds, Advances in Intelligent Data Analysis. Proc. of the IDA-97, no. Lecture Notes in Computer Science LNCS **1280**, (1997), 259-272
18. Wang, J.Z., Wiederhold, G., Firschein O., Wei X.: Content-based image indexing and searching using Daubechies' wavelets. Int. Jour. of Digit. Libr. (1), (1997), 311-328
19. Wang, J.Z., Li, J., Wiederhold, G.: SIMPLicity: Semantics-sensitive integrated matching for picture LIbraries. IEEE Trans. on Pat. Anal. and Machine Intelligence, **23**, (9), (2001), 947-963
20. Wang, W-J.: New similarity measures on fuzzy sets and on elements. Fuzzy Sets and Systems, **85**, (1997), 305-309
21. Will, T.: Switching distance between graphs with the same degrees. SIAM J. Discrete Math, **12**, 3, (1999), 298-306
22. Zadeh, L.: The role of soft computing and fuzzy logic in the conception, design and deployment of intelligent systems. Proc. Sixth Int IEEE Int. Conf. On Fuzzy Systems, Barcelona, July 1-5, (1997)