Neural Networks as Pattern Recognition Systems

Lourdes Calderón

Report LSI-96-32-R

Neural Networks as Pattern Recognition Systems

Lourdes Calderon
Departament de Llenguatges i Sistemes Informatics
Universitat Politecnica de Catalunya
calderon@lsi.upc.es, lcaldero@espol.edu.ec

Abstract

The appearance of the Neural Network paradigm has brought a new era to traditional pattern recognition techniques. This paper attempts to encapsulate the basic ideas of pattern recognition, the use of neural network systems to implement pattern recognition systems and the application of these systems in biomedical sciences. On the latter point, it provides a summary of the neural applications in that field and an extended example of a connectionist model used for medical data processing.

1 Introduction

Machine intelligence involves the ability to recognize patterns. Pattern Recognition (PR) techniques are quite often used in intelligent systems for both data preprocessing and decision-making tasks. PR science has been traditionally focused on statistical methods but in recent years, due to the research on machine learning, methods such as inductive learning have brought newer models of PR systems. Most recently, the arrival of Neural Network (NN) modeling has given rise to an attractive perspective of PR systems. This type of subsymbolic-level processing seems to be well-suited for accomplishing perception tasks in which PR tasks could be included and, according to some authors, even cognitive tasks.

The main purpose of this paper is to discuss the NN perspective of PR systems, extending its scope to explore the *state of the art* of connectionist PR approaches in biomedical applications. This paper is organized as follows: in the first section an general overview of concepts and approaches to PR systems is given. The second one emphasizes the NN approach to PR problems and includes a survey of NN systems in Biomedical Applications. The next section shows an experimental example of using a NN model applied to classify Breast Cancer patients and to predict the onset of diabetes. The results indicate a superior performance of neural PR approaches in comparison with other types of system. Finally, conclusions and references are included in the remaining sections.

1

2 Pattern Recognition Systems

PR systems are frequently used in a wide range of applications, which include: image preprocessing; segmentation and signal analysis; computer vision; face, character and speech recognition; medical diagnosis; biological classification, data compression, and so on.

As a research subject, PR science is related to other research areas, such as: Artificial Intelligence (AI), Signal Processing, Neural Modeling, Optimization/Estimation Theory, Automata Theory, Fuzzy Sets, Formal Languages.

This section begins with the basic concepts in PR science, followed by a description of two existing approaches: statistical models based on the Bayes' theorem, and syntactic models based on a tree scheme.

2.1 Objectives and Basic Concepts

Pattern Recognition, Artificial Intelligence and Neural Networks have a common history from the 1950s and have a common purpose: building machines that perform "human" tasks. Thus, the concept of PR has different meanings to different people.

PR could be broadly defined as the science that concerns recognition of measurements, through an information reducing, information mapping or information labeling process (Schalkoff, 1992). For the proper performance of this process, learning is the central issue of all these concepts together with recognition, classification (or discrimination) and clustering.

A pattern is a set of measurements, often represented as a vector. Measurements are taken from the features of a particular data set. Features may be symbolic, numerical or both and they can be described by continuous, discrete or binary variables. In a real-world problem, these variables may contain noise or unknown data values. Thus, given a set of characteristics concerning a set of examples, the task of PR is to assign input data into prespecified classes, based on extraction of significant attributes. A typical PR system structure has four steps (Batchelor, 1978):

1. Transducing: translation of a pattern into a suitable form for computing;
2. Preprocessing: removal of irrelevant characteristics;
3. Pattern Description: representation of a pattern: vectors, graphs;
4. Classification Algorithm: decision-maker.

If intelligence is associated with recognition of patterns, for more complex cases, the knowledge to be represented may be acquired in different ways. Schemes to reach a successful PR system may be proposed according to the

2

nature of the feature values. In some instances, it is possible to find a quantitative statistical basis, denoting the PR model by probability theory. In other instances, structural information could be extracted from the pattern, based on relationships among features rather than their numerical values. Nevertheless, there could even be instances with no understandable structure and too extensive to be modeled using statistical functions. In such cases, a "black-box" system could be constructed so that after being trained with examples, it would be able to produce a correct response. Among others, the NN paradigm constitutes an actual implementation of a such system.

Both NN and Statistical PR schemes partition a multidimensional space of features into class labeled decision regions that indicate to which class an input belongs, the border of each region being a decision boundary. On the other hand, the syntactic approach to PR relates the pattern structure with the syntax of a formal language built on mathematical linguistic techniques, while other approaches are concerned with representations of attributed graphs or trees.

2.2 The Bayesian Approach

Bayes' theorem provides a fundamental principle for the statistical approach to Pattern Recognition. Considering a pattern described by an observation vector X_k, the Bayesian relationship states:

$$P(C_i/X_k) = P(X_k/C_i) \; P(C_i) \; / \; P(X_k),$$

where, $P(C_i)$ is the *a priori* probability that a pattern belongs to class C_i, regardless of the identity of a pattern; $P(X_k)$ is the probability that a pattern is X_k regardless of its class membership; $P(X_k/C_i)$ is the class conditional probability that a pattern is X_k, given that it belongs to class C_i; and, finally, $P(C_i/X_k)$ is the *a posteriori* conditional probability that a pattern's class membership is C_i, given that the pattern is X_k.

Taking the case of a two-class problem, a Bayes' decision rule determines that X_k belongs to class C_i or C_j if and only if $P(C_i/X_k) >< P(C_j/X_k)$, where i is different from j. This relation could be expressed in terms of decision functions. Using thes Bayes' formalism for each class:

$$P(X_k/C_1) \; P(C_1) >< P(X_k/C_2) \; P(C_2)$$

$$P(X_k/C_1) \; / \; P(X_k/C_2) >< P(C_2) \; / \; P(C_1)$$

$$l(X) = P(X_k/C_1) \; / \; P(X_k/C_2) >< P(C_2) \; / \; P(C_1)$$

3

The term $l(X)$ is called the likelihood ratio and $P(C_2)/P(C_1)$ is the threshold value of the likelihood ratio for the decision. Let $h(X) = -\ln l(X)$ be the decision rule (Fukunaga, 1972). Given that the Bayes' decision rule has been transformed into the action of discriminant functions, subsequent classification consists in minimizing the associated Bayes' error. Bayes' error measures the class separability and determines the overlap among different class densities in the measurement space. Often, it is very hard to compute unless $h(x)$ is a linear function and there are normal distributions and equal covariance matrices, characteristics present in the linear classifiers.

Linear classifiers can be implemented by many known techniques as correlations and Euclidean distances. In the latter case patterns can be classified only according to their similarities with nearest neighbors, using Euclidean distances to determine the class membership of input patterns. Linear classifiers are optimum only for normal distributions and equal covariance matrices. They are the most simple and common.

Linear classifiers work well in normal distributions, a condition that does not hold for most applications. When there are unequal covariance matrices and non-normal distributions, the design of classifiers has not been successful. Quadratic classifiers are used with covariance difference, but it is not known how to obtain the optimum. Other attempts are piecewise classifiers, which are used for a set of linear discriminant functions or multiclass problems (Fukunaga, 1993).

None of these classifiers is able to deal with the complexity of a non-linear mapping. Non-linear problems are common in many engineering areas, particulary in PR tasks, especially when the number of variables is large. There are some techniques that approximate solutions to treat non-linear mapping and to design most general classifiers (Fukunaga, 1972). However, these classifiers could be achieved only at the expense of time, excessive storage and growth in complexity.

2.3 A Syntactic Approach: ID3

The idea of a complex pattern could be described recursively in terms of simpler ones gave rise to syntactic approaches to PR problems. These approaches use a procedure similar to that used in the parsing of a text by dividing it into words.

While statistical PR handles a pattern description as an observation vector, syntactic approaches use symbolic data structures like strings, trees or graphs that represent and describe relations of pattern variables. Therefore, patterns characterized by a set of features whose interrelationships contain a structural information are better treated using syntactic PR.

4

In a syntactic PR algorithm, a new input pattern is compared by a symbolic match against an existing number of prototype models. Basic algorithms used to address syntactic PR include: string or graph matching, hidden Markov models, formal grammars, etc. (Bunke, 1993). For each one, there are numerous versions implemented. In this subsection a tree-like strategy will be cited: the ID3 algorithm introduced by (Quinlan, 1986).

Being a symbolic machine learning algorithm, ID3 treats knowledge in terms of matching rules, rather than in terms of pattern values; consequently, it is better suited for classifying patterns that have nonnumerical attributes. ID3 infers discriminants by creating a decision tree using labeled examples. At each level of the tree, the algorithm calculates an entropy value for each feature that indicates which is the feature with the most information, and depending on this value, invokes one of a set of subtrees. If it should be necessary to identify what type of information processing ID3 uses, it appears that ID3 uses sequential rather than parallel processing. ID3's steps should be stated as follows (Pao, 1989):

1. Calculate initial value of entropy.
2. Select a feature to serve as the root node of the decision tree.
3. Build the next level of the decision tree.
4. Repeat steps 1 through 3.

The principal advantage of this approach constitutes its simplicity which makes it easier to implement. It has also been proved that ID3 can discover a combination of relevant features to assign class membership. Obviously, it seems clear to infer that a decision tree is useful when only a few of the input variables are relevant to the classification. Alternatively, if all variables are relevant, a decision tree will attempt to create 2^n subtrees that might not represent important concepts, and it will become more complex. Moreover, ID3 does not provide any mechanism to allow a system to recover from errors or to improve itself by learning (Pao, 1989).

2.4 Comments and Discussion

The theoretical work elaborated in statistics gives a significant support to statistical PR. However, construction of systems of this kind is often difficult to implement in practice, especially when patterns have high dimensionality inputs, which is a general characteristic of many PR problems. Furthermore, the Bayesian approach is useful when there is no ambiguity in the patterns, which makes it impossible to include subjective measurements of belief, uncertainty or missing information.

On the other hand, the ID3 approach is easier to understand and to implement. Another advantage is the possibility of knowing the way the system has inferred discriminants, giving a medium to explain classification process, a characteristic desirable in many constructed intelligent systems. However, this approach performs a binary partition, which results in a large and more complex binary tree, efficient only for patterns with no numerical attributes.

One approach that is especially appropiate for dealing with uncertainty or missing information consists in using the fuzzy set theory in processing PR tasks. This gives a bridge between linguistic variables and quantitative variables, and uses the concept of fuzziness to represent the posibility that a pattern belongs to a determined class (Pao, 1989).

3 Neural Network Approach

General ideas of how the brain performs visual PR tasks have inspired the research of using NN for constructing PR systems. NN systems are characterized by their parallel distributed nature and high fault tolerance.

Because PR problems are often too diffuse to define in formal terms, the nature of a nonalgorithmic PR system inside a NN model offers an attractive strategy and an emerging paradigm for implementing PR systems. Many NN systems that have been developed underline the success of NN in PR applications. A list of real and potential applications of NN as PR systems is presented in (Miller & Walker, 1992).

The aim of this section is twofold. First, it provides a brief introduction to NN as pattern classifiers. Second, it gives pointers to the current literature concerning the use of NN models for biomedical tasks.

3.1 Neural Network Classifiers

As mentioned above, real problems in PR systems are characterized by a high dimensional space and complex interactions between variables. NN are presumed to be suitable for dealing with these problems. Providing there is a representative number of cases or patterns, the task of a neural classifier consists in learning a set of weight connection values so that patterns can be classified correctly.

By the way of introduction, the first approximation to a NN model known as Perceptron, was the precursor in classifying patterns using a parallel distributed processing. Despite its severe limitations, this model established the basis for

further studies. The later appearance of the Generalized Delta Rule (Rumelhart et al., 1986) opened up different approaches to neural learning discriminants, in contrast to the sequential processing present the syntactic PR.

Currently, there are various NN structures for PR research and practice. According to the functionality of network architectures and algorithms, three main areas can be identified in neural computing for traditional PR research: Unsupervised Learning, Superivised Learning and Associative Memory or Hopfield Networks.

NNs trained using an unsupervised learning algorithm construct a clustering process by self-organizing input patterns into classes or clusters based on some metric to measure the distance between two patterns. Algorithms that implement unsupervised learning are: ART algorithms, detailed in (Carpenter & Grossberg 1, 1987) and (Carpenter & Grossberg 2, 1987); and Kohonen-networks, detailed in (Kohonen 1, 1988), (Kohonen 2, 1988) and (Kohonen, 1982). In contrast, supervised learning is an inductive process to find discriminant functions, which make pattern mapping by means of connection weights in the hidden units. Algorithms that implement supervised learning are: the well-known Backpropagation algorithm and the Boltzmann Machine with Simulated Annealing. Feedforward NN with backpropagation are mostly used in classification problems. Finally, algortihms in Associative Memory model a content-addressable memory, which is excellent for distortioned patterns but it is inefficient in practical implementations, because Hopfield networks are fully connected networks.

Not all these approaches require a detailed knowledge of statistics or pattern structure, although some recent studies include the possibility of using semantic or other type of information to give meaning to the behavior of NN units. In the same way, many NN models use statistical techniques for selecting the optimal set of training examples.

A general procedure to design a NN classifier involves defining a network architecture, choosing a good training algorithm, training the network, testing its performance, and analyzing its behavior. The last point is optional and it is supposed to have the purpose of explaining NN's outcome. Training parameters are often problem-dependent and must be correctly selected: training by patterns or epochs, the use of momentum, sequential or random ordering of patterns, choosing a bias, initial range of weights, bias values, etc. Besides, since neural learning is an empirical learning system, the knowledge for supporting a classification task is acquired during the training phase, so that the success of a Neural PR is understood to be strongly influenced by the quantity and quality of the training data.

7

3.2 A Survey of Biomedical Applications

As stated in section two, PR is present in many real-world applications. An important application of PR systems is the biomedical area. This subsection attempts to present a compilation of papers that have been published recently and relate the use of NN models applied to medical data.

As a starting point for this study, it should be recalled that computer-aided diagnosis systems for medical areas have been developed using two popular methods: systems based on Bayes' theorem and Expert Systems. Systems based on Bayes' theorem require an impractical number of data and symptoms are assumed to be statistically independent, this leading to a poor performance in practice. Regarding the second methods, the first application of earlier expert systems was the medical field. These systems are based on rules attempting to emulate the reasoning of physicians. Often, these systems are criticized for this unfeasible procedure and for being designed in an *ad hoc* manner.

NN research has been working in the biomedical same field with promising results. Compiling the large number of articles published related with the use of NN systems in biomedical problems, it is easy to infer its success in three identifiable categories: prediction, classification, and imaging and analysis systems. The potentiality of NN systems for computer-aided diagnosis systems is mainly because of the subjectivity, imprecision, and incompleteness of medical data. These characteristics make this type of data very difficult to deal with using conventional methods (Jones, 1992). Informally, NN systems are seen as expert systems where there is no need to acquire either the specific knowledge in the form of production rules (DeClaris & Mu-Chun, 1993), or a specific programming (Alvager, 1994).

Searching in NN as prediction systems, prediction of outcomes in breast cancer patients is the most prominent. A large quantity of papers compare the performance obtained in connectionist models against the performance obtained with other kind of systems: on the one hand, statistical models such as Logistic Regression, Quadratic Discriminant, Cox Regression, etc.; and on the other one, the current Staging System (Burke et al., 1994; Fang, 1993; and Astion & Wilding, 1992). In these comparisons, NN models improve upon the accuracy obtained by other models, in prediction, diagnosis, or recognition of breast cancer patients. Beyond this, authors agree in that NN are potentially useful for these kind of problems, because these models can discover prognostic factors and their complex interactions (Burke et al., 1994) as well as the relevant inputs for the classification (Wilding et al., 1994).

Another interesting application under this category is the prediction of outcomes of biopsy from radiographic findings (Floy et al., 1994) in order to op-

timize patient's therapy. To close the discussion of NNs as prediction systems, there are also models that predict the postoperative outcome of implantations of artificial heart valves (Katz et al., 1994) and liver transplantations (Doyle et al., 1994), with an acceptable accuracy. Additionaly, other models predict patients who would have a prolongued Intensive Care Unit length of stay (Tu & Guerriere, 1993).

For the second category, there is a wide range of medical classification problems: classification of hepatocellular carcinoma (Thung et al., 1994), classification of autism (Cohen et al., 1993), chromosome classification (Jennings & Graham, 1993), etc.

To conclude with biomedical applications of neural PR systems, models pertaining to the third category facilate pattern recognition in medical imaging and patient classifications, such as the analysis of electroencephalogram (EEG) and electrocardiogram (ECG) patterns (Heden et al., 1994;and Jando et al., 1993). The analysis of the results from both tests is usually very complicated due to the large amount and dispersion of data, and the interpretation requires sophisticated techniques of image analysis, such as multivariate approaches, which produce systems with poor performance. In contrast, a neural PR system appears as a simpler alternative to be implemented, and has showed better results.

3.3 Comments and Discussion

Attending to the large number of publications concerning the successful application of NN in PR problems its use is more and more extended. Neural PR systems are better when pattern features are numerical and they are in a parallel processing format. Hidden units in NN can capture nonlinear relationships between input and output vectors, learning discriminants through the generalized delta rule procedure. However, the question of how many hidden layers and units are necessary to achieve a good generalization lacks of a formal answer, as well as the neural network's interpretation and explanation. Often, the former subject constitutes a common criticism due to the fact that designers must guess or imagine important parameters in a connectionist model by common sense and experimenting. In spite of that, many authors carry on with this research field and they are addressing such problems.

4 Experimental Example

Two domains of medical data were selected to test an experimental design of NN model based on multilayer perceptrons with a sinosoidal nonlinearity (Sopena & Monte, 1995). Breast Cancer and Diabetes data sets were taken from the UCI

Machine Learning Repository (Murphy & Aha, 1994). In both domains, the NN modeling seems to be suitable to achieve a better generalization in unseen patterns.

4.1 Breast Cancer Classification

The breast cancer data is characterized by a proliferation of factors, complex interactions between them and non-monoticity, field in which neural networks have been shown to be as efficient as the best traditional methods (Burke, Goodman & Rosen, 1994). The Breast Cancer database was used to measure the accuracy of two types of learning: a statistical method called the Multisurface Method of Pattern Separation (Wolberg & Mangasarian, 1990) and Instance-Based Learning method (Zhang, 1992).

The data set contains information on clinical cases of breast cancer. The number of instances is 367, and some of these have missing attribute values. Each instance has 11 numeric attributes, attributes 2 through 10 are variable inputs. Each instance has one of two possible classes: benign or malignant.

The experimental design consists of a NN model using the sine function as the activation function. The neural network architecture has nine input units, 28 hidden units and two output units. The initial range of weight connections was 0.2 for all units. The network's parameters were adjusted to obatin the best performance: the learning rate was set to 0.001 and the momentum value was set to 0.8. The original Breast Cancer data were encoded to train BP* nets. Three random resampling percentages were selected: 67%-33%, 55%-45%, and 50%-50%. Seven randomly sampled data sets were created for each case. Results showed in Table 1 were obtained from an average of the respective runs.

4.2 Diabetes Classification

The Pima Indians Diabetes Database was used by an adaptive learning routine that generates and executes digital analogs of perceptron-like devices, called the ADAP algorithm (Smith et al., 1988).

The database contains 768 instances, with no missing attribute values. Each instance has 8 numeric attributes plus a binary attribute that indicates the class. This binary-valued variable diagnoses whether the patient shows signs of diabetes according to World Health Organization criteria.

The experimental design consists of a NN model using the sine function as the activation function. The neural network architecture has eight input units, 40 hidden units and one output unit. The initial range of weight connections was 1 for units 8 to 16, 2 for units 17 to 26, 5 for units 27 to 47, and 1 for

Classification Method	Success Rate	Random Resampling
Multisurface Method	95.9% 93.7%	67% - 33% 55% - 45%
Instance-Based Learning	93.5%	50% - 50%
BP* Nets Modeling	95.9% 96.4% 95.1%	67% - 33% 55% - 45% 50% - 50%

Table 1: Classification Success Rates for Breast Cancer Data

unit 48. Network's parameters were adjusted to obtain the best performance: the learning rate was set to 0.01 and the momentum value was set to 0. The original Diabetes data were encoded to train BP* nets. One random resampling percentage was selected: 75%-25%. Seven randomly sampled data sets were created for each case. Table 2 shows the correponding results obtained from an average of the respective runs.

Random Resampling	ADAP Learning Alogithm	Modeling BP* Nets
75% - 25%	76%	81.03%

Table 2: Prediction Success Rates for Diabetes Data

5 Conclusions

The principal issue in this paper has emphasized the use of NN in PR tasks, underlying its successful implementation in biomedical applications. To this end, a general overview of the principal PR approaches was presented in order to illustrate the differences between then and the connectionist approach. Lastly, an experimental design of a connectionist model has been introduced, which has been used in two types of medical data processing: classification and prediction tasks. The current diffusion and good outcomes in the application of neural computing encourage research on this active area of Artificial Intelligence.

6 References

Alvager, T., 1994. The use of artificial neural networks in biomedical technologies: an introduction. Biomedical Instrumentation & Technology (July-Aug. 1994) Vol.28 No.4, pp.315-22.

Astion, M. and Wilding, P., 1992. Application of neural networks to the interpretation of laboratory data in cancer diagnosis. Comment in Clinical Chemistry 1992, Department of Pathology and Laboratory Medicine, Hospital of University of Pennsylvania, Philadelphia.

Batchelor, B., 1978. Pattern Recognition, Ideas in Practice. Plenum Press, New York USA.

Bunke, H., 1993. Structural and Syntactic Pattern Recognition. Handbook of Pattern Recognition and Computer Vision, edited by Chen, C.H., Pau, L.F. and Wares, P.S. World Scientific Publishing Co. Pre. Ltd., USA.

Burke, H., Goodman, P., and Rosen, D., 1994. Artifical Neural Networks for Outcome Prediction in Cancer. *Proceedings of World Congress on Neural Networks*, Volume I, pp. 53-56, San Diego, California.

Carpenter, G.A. and Grossberg,S., 1987 (1). A massively parallel architecture for a self-organizing neural pattern recognition machine. Computer, Vision, Graphics and Image Processing, vol.37, 1987, pp. 54-115.

Carpenter, G.A. and Grossberg,S., 1987 (2). ART2: Self-organization of stable category recognition codes for analog input patterns. Applied Optics, vol. 26 No. 3, 1987, pp.4919-1930.

Cohen, I., Sudhalter, V., Landon, D. and Keogh, M., 1993. A neural network approach to the classification of autism. Journal Autism-Dev-Disorders 23(3), pp. 443-466.

DeClaris, N. and Mu-Chun, S., 1993. A neural network based approach to knowledge acquisition and expert systems. *Proceedings of the International Conference on Systems, Man and Cybernetics.* Systems Engineering in the Service of Humans, Vol. 2 of 5, pp. 645-650, N.Y.

Doyle, H., Dvorchik, I., Mitchell, S., Marino, I.,Ebert, F., McMichael, J. and Fung, J., 1994. Predicting outcomes after liver transplantation, a connectionist approach. Ann-Surg. 1994 Apr; 219(4), pp. 408-15.

Fang, R., 1993. Use of neural network analysis to diagnose breast cancer patients. *Proceedings TENCON'93*, IEEE Conference on Computer Communication, Control and Power Engineering.

Floyd, C., Yun, A., Lo, J., Tourassi, G., Sullivan, D. and Kornguth, P., 1994. Prediction of breast cancer malignacy for difficult cases using an Artificial Neural Network. *Proceedings of World Congress on Neural Networks*, Volume I, pp. 127-132, San Diego, California.

Fukunaga, K., 1972. Introduction to Statistical Pattern Recognition. Academic Press, Inc., New York.

Fukunaga, K., 1993. Statistical Pattern Recognition. Handbook of Pattern Recognition and Computer Vision, edited by Chen, C.H., Pau, L.F. and Wares, P.S. World Scientific Publishing Co. Pre. Ltd., USA.

Heden, B., Edenbrandt, L., Haisty, W. and Pahlm, O., 1994. Artificial neural networks for the electrocardiographic diagnosis of healed myocardial infarction. Am-J-Cardiol. 1994 Jul 1; 74(1), pp. 5-8.

Jando, G., Siegel, R., Horvath, Z. and Buzsaki, G., 1993. Pattern recognition of the electroencephalogram by artificial neural networks. Electroencephalography Clinical Neurophysiology, 86(2), pp. 100-109.

Jennings, A. and Graham, J., 1993. A neural network approach to automatic chromosome classification. Physics in Medicine and Biology 38(7), pp. 959-970.

Jones, D., 1992. Neural Networks for Medical Diagnosis. Handbook of Neural Networks Applications, edited by Maren, A., Harston. C. and Pap, R. Academic Press, Inc., San Diego California.

Katz, S., Katz, A., Lowe, N. and Quijano, R., 1994. Neural net-bootstrap hybrid methods for prediction of complications in patients implanted with artificial heart valves. J-Heart-Valve-Dis. 1994 Jan; 3(1), pp. 49-52.

Kohonen, T., 1972. Correlation Associative Memories. IEEE Transactions on Computers, Vol. C-21, No. 4, pp. 353-357.

Miller, R. and Walker, T., 1992. Neural Networks Applications and Products. Vol. 1, pp. 21, SEAI Technical Publications.

Murphy, P. and Aha, D. (1994). UCI Repository of Machine Learning Database and Domain Theories. Irvine, CA: University of California, Department of Information and Computer Science.

Pao, Yoh-Han, 1989. Adaptive Pattern Recognition and Neural Networks. Addison-Wesley Publishing Company, Inc, New York.

Quinlan, J.R. (1986). Induction of Decision Trees. Machine Learning 1, pp. 81-106.

Rumelhart, D.E., Hinton, G.E. and Williams, R.J., 1986. Learning internal representations by error propagation. In D.E. Rumelhart and J.L. Mc.Clelland (Eds.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations, pp.318-362. The MIT Press, Cambridge, Mass.

Schalkoff, R., 1992. Pattern Recognition: Statistical, Structural and Neural Approaches. John Wuley & Sons, Inc., New York.

Sopena, J.M. and Monte, E. (1995). On the Properties of Multilayer Perceptrons with a Sinosoidal Non Linearty and Bechmarks. Submitted to International Workshop on Artificial Neural Networks, Malaga, Spain.

Smith, J. W., Everhart, J. E., Dickson, W. C., Knowler, W. C., and Johannes, R. S., 1988. Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. In *Proceedings of the Symposium on Computer Applications and Medical Care* (pp. 261–265). IEEE Computer Society Press.

Thung, S., Geller, S. and Marchevsky, A., 1994. Image analysis and diagnostic classification of hepatocellular carcinoma using neural networks and multivariate discriminant functions. Laboratory Investigation, 71(3), pp. 446-451.

14

Tu, J. and Guerriere, M., 1993. Use of a neural network as a predictive instrument for length of stay in the intensive care unit following cardiac surgery. Computing Biomedical Research 26(3), pp. 220-229.

Wilding, P., Morgan, M., Grygotis, A., Shoffner, M., and Rosato, E., 1994. Application of backpropagation neural networks to diagnosis of breast and ovarian cancer. Cancer-Letters, pp.145-153.

Wolberg, W.H. and Mangasarian, O.L. (1990). Multisurface Method of Pattern Separation for Medical Diagnosis Applied to Breast Cytology. *Proceedings of the National Academy of Science*, 9193-9196.

Zhang, J. (1992). Selecting Typical Instances in Instance-Based Learning. *Proceedings of the Ninth International Machine Learning Conference, Aberdeen, Scotland.*

Departament de Llenguatges i Sistemes Informàtics
Universitat Politècnica de Catalunya

Research Reports – 1996

LSI–96–1–R "(Pure) Logic out of Probability", Ton Sales.

LSI–96–2–R "Automatic Generation of Multiresolution Boundary Representations", C. Andújar, D. Ayala, P. Brunet, R. Joan-Arinyo, and J. Solé.

LSI–96–3–R "A Frame-Dependent Oracle for Linear Hierarchical Radiosity: A Step towards Frame-to-Frame Coherent Radiosity", Ignacio Martin, Dani Tost, and Xavier Pueyo.

LSI–96–4–R "Skip-Trees, an Alternative Data Structure to Skip-Lists in a Concurrent Approach", Xavier Messeguer.

LSI–96–5–R "Change of Belief in SKL Model Frames (Automatization Based on Analytic Tableaux)", Matías Alvarado and Gustavo Núñez.

LSI–96–6–R "Compressibility of Infinite Binary Sequences", José L. Balcázar, Ricard Gavaldà, and Montserrat Hermo.

LSI–96–7–R "A Proposal for Word Sense Disambiguation using Conceptual Distance", Eneko Agirre and German Rigau.

LSI–96–8–R "Word Sense Disambiguation Using Conceptual Density", Eneko Agirre and German Rigau.

LSI–96–9–R "Towards Learning a Constraint Grammar from Annotated Corpora Using Decision Trees", Lluis Màrquez and Horacio Rodríguez.

LSI–96–10–R "POS Tagging Using Relaxation Labelling", Lluís Padró..

LSI–96–11–R "Hybrid Techniques for Training HMM Part-of-Speech Taggers", Ted Briscoe, Greg Grefenstette, Lluís Padró, and Iskander Serail.

LSI–96–12–R "Using Bidirectional Chart Parsing for Corpus Analysis", A. Ageno and H. Rodríguez.

LSI–96–13–R "Limited Logical Belief Analysis", Antonio Moreno.

LSI–96–14–R "Logic as General Rationality: A Survey", Ton Sales.

LSI–96–15–R "A Syntactic Characterization of Bounded-Rank Decision Trees in Terms of Decision Lists", Nicola Galesi.

LSI–96–16–R "Algebraic Transformation of Unary Partial Algebras I: Double-Pushout Approach", P. Burmeister, F. Rosselló, J. Torrens, and G. Valiente.

LSI–96–17–R "Rewriting in Categories of Spans", Miquel Monserrat, Francesc Rosselló, Joan Torrens, and Gabriel Valiente.

LSI–96–18–R "Strong Law for the Depth of Circuits", Tatsuie Tsukiji and Fatos Xhafa.

LSI–96–19–R "Learning Causal Networks from Data", Ramon Sangüesa i Solé.

LSI–96–20–R "Boundary Generation from Voxel-based Volume Representations", R. Joan-Arinyo and J. Solé.

LSI–96–21–R "Exact Learning of Subclasses of CDNF Formulas with Membership Queries", Carlos Domingo.

LSI–96–22–R "Modeling the Thermal Behavior of Biosphere 2 in a Non-Controlled Environment Using Bond Graphs", Angela Nebot, François E. Cellier, and Francisco Mugica.

LSI–96–23–R "Obtaining Synchronization-Free Code with Maximum Parallelism", Ricard Gavaldá, Eduard Ayguadé, and Jordi Torres.

LSI–96–24–R "Memoisation of Categorial Proof Nets: Parallelism in Categorial Processing", Glyn Morrill.

LSI–96–25–R "Decision Trees Have Approximate Fingerprints", Víctor Lavín and Vijay Raghavan.

LSI–96–26–R "Visible Semantics: An Algebraic Semantics for Automatic Verification of Algorithms", Vicent-Ramon Palasí Lallana.

LSI–96–27–R "Massively Parallel and Distributed Dictionaries on AVL and Brother Trees", Joaquim Gabarró and Xavier Messeguer.

LSI–96–28–R "A Maple package for semidefinite programming", Fatos Xhafa and Gonzalo Navarro.

LSI–96–29–R "Bounding the expected length of longest common subsequences and forests", Ricardo A. Baeza-Yates, Ricard Gavaldà, and Gonzalo Navarro.

LSI–96–30–R "Parallel Computation: Models and Complexity Issues", Raymond Greenlaw and H. James Hoover.

LSI–96–31–R "ParaDict, a Data Parallel Library for Dictionaries (Extended Abstract)", Joaquim Gabarró and Jordi Petit i Silvestre.

LSI–96–32–R "Neural Networks as Pattern Recognition Systems", Lourdes Calderón.

Hardcopies of reports can be ordered from:

See also the Department WWW pages, http://www-lsi.upc.es/www/