Proceedings of the 8th IEEE Symposium on Computer-Based Medical Systems
Lubbock, Texas, USA, June 1995, pp. 111-117.

Prototyping Neural Networks Learn Lyme Borreliosis

Stefano Rovetta,* Rodolfo Zunino,* Laura Buffrini,° Guido Rovetta°

University of Genova, Italy
* Faculty of Engineering, DIBE ° Faculty of Medicine, DIMI

Abstract

In this paper, the application of neural network algorithms to the study of Lyme borreliosis is addressed. Three different methods are studied: Self Organizing Maps, Neural Gas Networks and a new approach currently under development, called Circular Back Propagation. The aim of the work is to compare the three methods in view of their use as analysis tools, to explore the inherent structure of the input data. The same procedure has been previously applied to feedforward neural models; the present work focuses on a particular form of knowledge representation, based on a set of prototypal examples rather than if-then rules. The Lyme data has been chosen as a case study and represents a common ground to allow the comparison of the different methods.

1. Introduction

In this paper, the application of advanced processing methods to the study of Lyme borreliosis is addressed. The Inst. Bruzzone Rheumatological Center of the University of Genova (Italy) has recorded a number of samples into a database, which has been preprocessed for use with numerical classification algorithms. We focused on *neural network* algorithms, a widespread technology but also a smart combination of statistical and clustering techniques, pattern recognition capabilities, and self-adaptation for autonomous programming (that we will call *learning*).

The aim of the studies here reported is not immediately applicative, but rather methodological. Our goal is not the realization of programs, but the development and the experimentation of new techniques. The general setting of the problem is familiar: we are given a set of samples from some phenomenon (e.g., Lyme disease patients), and we wish to identify the rule underlying their classification (e.g., as *ill* or *non-ill*). The samples are expressed as patterns of numerical values, and will be treated as vectors. This accounts for such concepts as *distance* among samples and *distribution* of inputs.

A previous study had been conducted on the Lyme database [1, 2], using feedforward neural networks as the main tool. The task we address is not simply to synthesize a classifier, in order to implement the diagnostic step, but rather to extract from the available data the implicit knowledge necessary to both implement the diagnosis and obtain useful rules or a characterization, in order to confirm or to complement the medical knowledge about the disease. To accomplish this task, we have previously used network models, with different layouts and training criteria, that belong to the family of Multi Layer Perceptrons. The present work reports on the application to the same problem of another class of networks, in which the knowledge representation is different. In this class, the internal representation may be interpreted as a collection of significant samples from the input space, that may or may not be present among the training examples, but are homogeneous with them. Thus we have *prototypes* as the representative elements for learning, rather than *rules* defined on the input space. The operation step can be a classification or simply the identification of the activation values (e.g. for clustering purposes), but in either case the decisions are made on the basis of the least-distance prototypes.

This representation paradigm is especially well-suited when dealing with unlabeled samples (for which the diagnosis is unknown or not taken into account). In this case, we seek for a structure in the distribution of patterns. We hope that the distribution, although independent from the actual classification, will reveal a structure correlated with the diagnosis.

2. Prototyping neural models

We studied three models: an improved version of Kohonen's Self Organizing Maps (SOM) [3], an improved version of Martinetz' Neural Gas Network (NGAS) [4] and a new approach called Circular Back Propagation (CBP) [5].

2.1. Self Organizing Maps

The network model developed by T. Kohonen is based on the paradigm of self-organization: it does not require a labeling of the input data, but acts only based on their statistical distribution. The network is composed of a set of processing units (what in the biological metaphor would be the *neurons*), structured into a lattice topology, often taking the form of a squared grid. Hence for each unit there are four immediate neighbors. This topology is a property of the network, and it cannot be modified by self-adaptation.

This description identifies the network as a map of units, hence the name. The learning process creates a correspondence between the input space (the set of all possible inputs) and the distribution of active units; this transformation preserves the topologic relations of the input space, in that two similar patterns will activate two units that are close to each other along the grid. The mapping reveals clusters in the input data by attributing more prototypes to regions in the input space where points are more frequent, and representing them typically in a two-dimensional space, which is easily visualized.

The learning procedure is based on a competition among the units. Each unit encodes a prototype, as described in the introduction. On presentation of a pattern, the distance between the pattern and each unit is computed. The unit with lowest distance wins the competition, and it is moved a bit in the direction of the received pattern. Around the winner, also a variable number of neighbors are modified in the same way, although the modification is smaller for the more peripheral neighbors. Iteration of this process yields the desired mapping by a progressive reduction of both the adaptation coefficient (modifications get smaller) and the neighborhood width (less units around the winner are modified).

2.2. Neural Gas Network

The second model, by T. M. Martinetz, is a derivation of the first one. It differs essentially in the definition of the neighborhood in the set of units, to propagate the effect of the activation. In a NGAS network the neighboring unit is not one directly connected along the lattice, but the one with the next lowest distance from the pattern. In other words, there is no fixed lattice, and the neighborhood is not defined statically, in terms of the topology, but rather dynamically, in terms of the distance ranking with respect to the incoming samples. During the adaptation, the units move in a way that is similar to the dynamics of the molecules of a gas, hence the name.

This mechanism helps avoiding some problems arising from a wrong initialization of the units. In fact, in some situations (not so unlikely when the input patterns have a high number of components) the fixed topologic structure may not allow the map to produce the required transformation, for instance because of "knots" that arise in the lattice during the evolution. The NGAS network trades the absence of topological consistence with the immunity from these problems.

2.3. Circular Back-Propagation Network

The third model is a simple but interesting modification of the standard back-propagation network, that we are currently developing. The structure is that of a layered feedforward network. The output layer is also standard, but the hidden layer is modified. Its processing elements feature an activation rule with an additive term involving the square of the input vector. Depending on the self-adapted parameters, we can interpret the obtained representation as either rule-based or prototype-based, because the presence of the quadratic term enables the activation function to implement also a distance-based activation. The result of the training is a global internal representation, as for the multi-layer perceptron family, but can be expressed either as a combination of prototypes, so that peculiar points in the input space can be identified, or as a set of surface-oriented rules.

The name is justified by the fact that a single unit can respond to vectors that are within a circular (or in general hyperspherical) region around its prototype.

2.4. Comparison of the three neural network structures

The differences and similarities among the three networks just described can be more thoroughly understood with the aid of the illustrations.

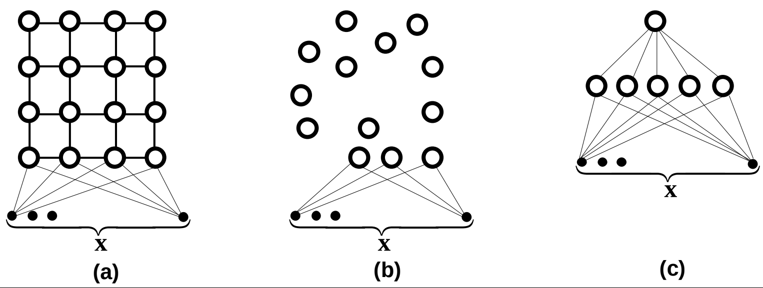

Fig. 1 – Structure of the three network models: (a) SOM - (b) NGAS - (c) CBP.

Fig. 1 shows the structure of the networks. The black dots represent the input vector components. Each one of them is fed into each unit, but for simplicity this is not completely shown. An exception to this schema is the CBP network, where there are two layers, the first of which is connected to the input and implements an *internal* representation, while the second yields the diagnosis.

Fig. 2 illustrates in more detail the different philosophy underlying the three models. The case **(a)** relates to SOM and NGAS. The units compute the distance between a stored prototype and the input vector; then a lateral inhibition process activates only one of the units; its label is used for classification. The case **(b)** is that of CBP. Here the units compute a weighted sum of the input vector components, with a *bias* term and a *sum of squares* term. The activation is a saturated function of this sum (the "membrane potential"). All units contribute to the output, as shown by the links between the representation layer (also called *hidden* layer) and the output layer, here composed only of one processing unit.

3. Data and methods

After a pre-processing phase, the data base on Lyme disease collects 684 patterns of 54 fields each (54-dimensional vectors), plus a classification target. Each pattern represents the description of a patient with suspect Lyme borreliosis. A learning methods needs such a set of samples as a training set; in accordance with it, they learn how to behave to implement some task (in our instance, diagnosing Lyme borreliosis). This is called *learning from examples*.

The learning process can be either supervised or unsupervised. *Supervised* means driven by the diagnosis. *Unsupervised* means driven only by the statistics of the data, without taking into account any knowledge about the diagnosis. The neural models applied were both of the supervised and of the unsupervised type. When an unsupervised method is applied to labeled data, it is possible to validate the significance of the obtained clusters by attaching to each prototype the labels that are more represented among the patterns it responds to (i.e., the patterns for which it wins the above described competition, those that are closest to it).

The aim of the unsupervised experiments (SOM and NGAS) is to obtain a clustering of

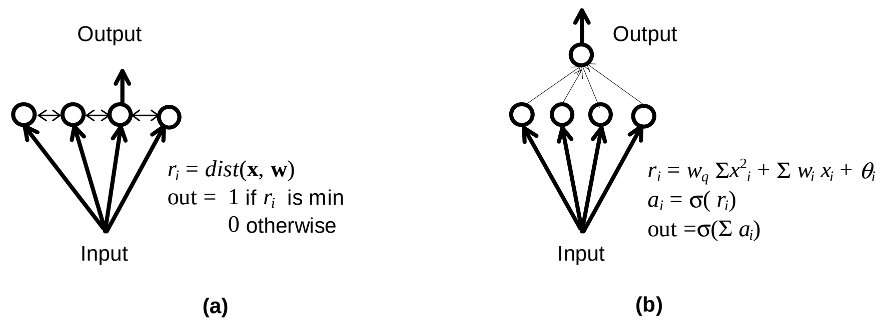

Fig. 2 – How SOM and NGAS (a) and CBP (b) produce their output when given an input pattern.

the data set based on the *structural* properties of data, without regard to their classification. Then, by associating the diagnosis labels to patterns, it is possible to inspect whether the obtained clustering has a meaning in terms of classification. This gives a hint on how the basic classes *Ill* and *Non-ill* are structured in the space of patterns.

When the supervised method (CBP) is used, the goal is similar, but the perspective is different. What we obtain is a set of clusters that do not reflect mainly the properties of data, but rather their (hopefully) optimal grouping for classification. Moreover, the adaptive representation may be used to inspect whether the optimal representation is by means of prototypes or by means of rules.

The procedure adopted for this task involves essentially the analysis of the coefficient of the sum-of-squares term. Those units that have this coefficient equal to zero represent a hyperplane, which has a first-order equation. These are clearly rules based on the surface paradigm. If this coefficient is significantly high, the stimulus equation turns into a distance-based one, like for the other models. The rule is then of the circular type, with a prototype localized in the center of the circular activation area. With a few calculations one can find the coordinates of this points from the coefficients.

Fig. 3 shows two examples of concepts. They are represented in the pattern space. The first concept is the diagnosis of osteoporosis based on subject age and bone density. The two regions can be separated by an open line (the two-dimensional version of a surface). In the second, the ECG is described by some of its component waves. The normal situation is a prototype (in the center of the closed region), and the patterns that fall too far from the prototype are classified as pathological. Of course, the situation is further complicated by the fact that the dotted lines are in practice “fuzzy” regions, and should be thought of as slowly fading from “yes” to “no,” rather than switching from one decision to the opposite one as illustrated.

4. Experimental results

All the clustering procedures discover very similar structures in the distribution of data. This refers to both the precision with which the three methods learn the training set, meas-

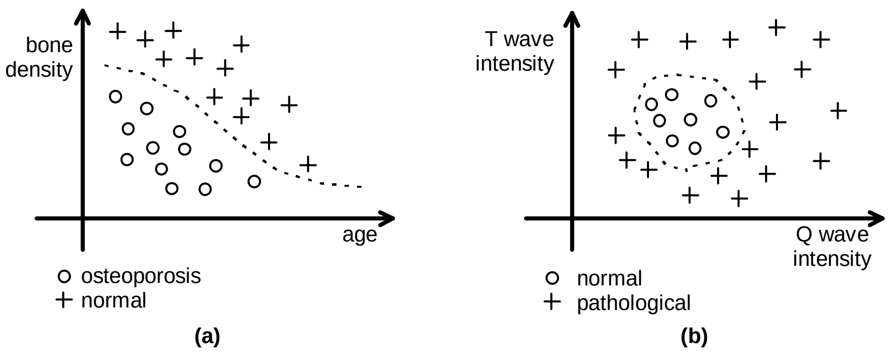

Fig. 3 – A surface-based concept (a) and a prototype-based concept (b).

ured by the error rates, and the clustering synthesized by the training. The visualization of the obtained results is not easy, because of the high dimensionality of the input vectors.

4.1. Effective dimension of the data

The first question of interest is: How many parameters are needed to describe the concept inherent in the database? In other words, how many rules are sufficient to reasonably describe the Lyme disease?

An answer is given by the comparison of the results of the three methods. To reach an average approximation error of 10%, SOM and NGAS require about 6 prototypes, and CBP requires 4 hidden units. Although the effective number of parameters is not assessed with certainty, and the experiments always include a certain degree of randomness, it is possible to observe that the 54 describing inputs are clearly not independent of each other. Some correlations among these variables are present (maybe non-linear correlations, i.e., dependency laws different from proportionality), and the data have an *effective* dimension that is much lower than the *apparent* one (approximately by an order of magnitude).

4.2. Spatial distribution of the data

Another question is: Do the data have any significant clustering in the pattern space? Is there any correlation between the position and the class of the data vectors?

By inspection of the parameter values obtained with training, we can make the hypothesis that the data set is not strongly clustered. It is possible to describe the data by a central cluster, mainly of the *Non-ill* class, and several sparse clusters, mainly *Ill*, that gather a low number of patterns each, often only an outlier pattern. The odd distribution of ill patients indirectly confirms the difficulties of the diagnostic process.

4.3. Characterization of the data

The last question we try to answer is: Are the data best characterized by means of a prototype-based or a surface-based representation?

The best method to answer this question is CBP, because of its switching capabilities between the two paradigms. Two tests can be made. We can compare the error rates of CBP with the results of standard BP, and we can examine the nature of the prototypes synthesized by the CBP training.

The first test yields as a result a small difference between the two methods. With stopped training, BP yields a training error percentage of 4.0% and a corresponding test error percentage of 9.8%; CBP yields a training error percentage of 4.2% and a corresponding test error percentage of 8.7%. The training error can be further reduced, but a worsening in the test performance (overtraining) is obtained. (A good introductory discussion of this phenomenon can be found in [6].)

The second test requires essentially to examine the weight associated to the quadratic term. If that coefficient is high, we can conclude that the rule has a prototype-based expression, otherwise it can be interpreted as rule-based. In the experiments, we found weights only of the same order of magnitude of the other weights. The radius of the circular activation region can be computed from the weights, and it is always in the range 12 - 100. The input values always lie in the range [−1, +1], so that a circle with a radius of 100 is, with good approximation, a straight line. The better performance of CBP over BP in the previous test can now be tentatively explained not with a representation based on the prototyping paradigm, but with a slight bending in the shape of the surfaces, which allows a better representation for some borderline patterns.

We can conclude that the optimal representation is by means of surface-based rules, in agreement with the fact that CBP yields better results than SOM and NGAS, and with the previous results obtained with feedforward networks of the surface-based type (see [2] and [7] for a report on the previous experiments).

5. Concluding remarks

In this paper we have presented an example of the application of neural networks techniques to medical problems. Software implementations and shells for the use of various neural network models are commonly available. However, too often they are used as a substitute for "expert systems." While an expert system can be built by means of neural networks, these two forms of knowledge-based systems should not be compared directly. This is not always clear to those neural network users who are expert in the specific applicative domain, but not in the "neural" technology.

The application we have just described is an example of how a neural methodology can be used as a tool, much like statistics in the traditional research framework. Unfortunately, most of these tools require a lot of trial-and-error procedures and empirical parameter choices, so that a help from "neural network wizards" is often required. The neurosoftware products do not help in this respect, because their ease of utilization often leads one to think that the self-adaptive properties will "self-adapt" the network to every need of the user and to every problem, which of course is not true! The current directions of research in the neural networks field aim to produce theoretical results that, either directly or indirectly, will ultimately enable researchers to use these neural tools without requiring a specific experience in that field.

References

[1] Crovato F, Nazzari G, Fumarola D, Rovetta G, Cimmino M, Bianchi G, "Lyme disease in Italy: first reported case," *Ann. Rheum. Dis.*, 44:570-571, 1985.

[2] Bianchi G, Buffrini L, Monteforte P, Rovetta G, Rovetta S, Zunino R, "Neural approaches to the diagnosis and characterization of Lyme disease," *Proceedings of the 1994 IEEE 7th Symposium on Computer-Based Medical Systems*, Winston-Salem, NC, 1994.

[3] Kohonen T, *Self organization and associative memories* (Springer Series in Information Sciences 8). Heidelberg:Springer, 1982.

[4] Martinetz T M, Berkovich S G, Schulten K J, " 'Neural-Gas' network for vector quantization and its application to time-series prediction," *IEEE Trans. Neural Networks*, vol. 4, no. 4, pp. 558-569, 1993.

[5] Ridella S, Rovetta S, Zunino R, "Adaptive Internal Representation in Circular Back-Propagation Networks," to appear in *Neural Computing and Applications*.

[6] Hecht-Nielsen R, *Neurocomputing*, Addison-Wesley, 1989.

[7] Ridella S, Speroni GL, Trebino P, Zunino R, "Class-entropy minimization networks for domain analysis and rule extraction," *Neural Computing and Applications*, 2:40–52, 1994.