

A Neuro-Fuzzy Approach as Medical Diagnostic Interface

R. Brause, F. Friedrich

J.W.Goethe-University, Frankfurt a. M., Germany
brause@cs.uni-frankfurt.de

In contrast to the symbolic approach, neural networks seldom are designed to explain what they have learned. This is a major obstacle for its use in everyday life. With the appearance of neuro-fuzzy systems which use vague, human-like categories the situation has changed. Based on the well-known mechanisms of learning for RBF networks, a special neuro-fuzzy interface is proposed in this paper. It is especially useful in medical applications, using the notation and habits of physicians and other medically trained people. As an example, a liver disease diagnosis system is presented.

1 Introduction

The decision and diagnostic power of neural networks are needed in many tasks of everyday life. Successful example applications show that especially in medicine human diagnosis abilities are significantly worse than those of neural diagnosis systems, see [1]. Nevertheless, the set of tools of neural networks is rarely used by physicians. Why?

First, the models and language of neural networks are more mathematically formulated, a language that most doctors do not know how to use and are not interested in to learn. Second, in contrast to the symbolic approach, neural networks seldom are designed to explain what they have learned. The possibility using the experience of the physician and explaining the diagnosis by the proper medical terms is crucial for the question whether a diagnostic system is used or ignored.

Now, with the appearance of fuzzy systems which use vague, human-like categories [2] the situation has changed. Based on the well-known mechanisms of learning in RBF networks, a special neuro-fuzzy interface is proposed in this paper. The intuitive and instructive interface is useful in medical applications, using the notation and habits of physicians and other medically trained people. In Fig. 1 this concept is visualized.

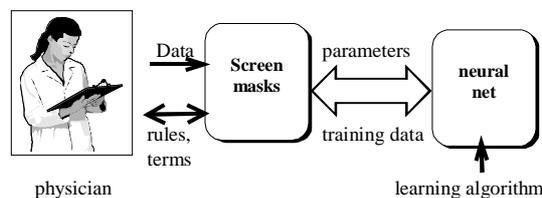


Fig. 1 Interactive transfer of vague knowledge

Here, the user interface must use the typical human properties and formulate the diagnosis by the vague, inexact language of physicians.

2 Interface Design Decisions

The neuro-fuzzy interface design has to meet several constraints.

1. It has to serve for a non-computer experienced public which has neither knowledge nor time or interest to invest in program design
2. It has to use the human attention as a major resource, i.e. it has to be understood very fast by medical people
3. The interface must be simple and consistent
4. The user necessary actions should be minimized, e.g. default values should be used whenever possible
5. The language and the mental model of the user group should be used

These are the most important design principles. When we examine some of these constraints, we remark that we might cover points 1,2 and 5 just by using appropriate medical terms, keywords and assumptions: the medical mental model. Thus will be covered by the next section.

2.1 Medical terms and notations

There are several kinds of terms which are involved in a medical description of a patient state. The chemical analysis covers

- ◆ The exact quantitative notation, e.g. *blood sugar 120 mg/dl*
- ◆ The semi-quantitative notation, e.g.

semi-quantitative result	Value range [g/dl]	Linguistic meaning
0	0 - 0,1	Not existent
+	0,1 - 0,25	Slightly elevated
++	0,25 - 0,5	elevated
+++	0,5 - 2	Strongly elevated
++++	> 2	Very strongly elevated

Table 1 Semi-quantitative results for urine sugar test

- ◆ The qualitative, categorical notation, e.g. test result positive/negative

All medical diagnosis for a variable reference an interval which is typical for the state of "healthy" people. For a Gaussian distribution of mean m and variance σ this is the interval of $[c-2\sigma, c+2\sigma]$ with 95.5 % of the population. All values distributed between $c \pm 2\sigma$ and $c \pm 3\sigma$ lie in the limit region while all values outside are pathological.

2.2 Transferring the Medical Knowledge to the Net

In contrast to the precisely defined fuzzy sets, we use only the idea of fuzzy terms, called "vague terms", and is description by membership functions. All other mathematical implications like computing rules are not used in this paper. As example, in Fig. 2 the assignment of the vague linguistic terms of the medical variable SGOT to the vague variable x (concentration IE/l) is shown. Formally, this can be defined as

membership function $\mu_i(x) : X \rightarrow [0,1]$ with the vague variable $X = [0,50]$ and the vague set $F(X) = \{ \mu_i(x) \} = \{ \text{low, normal, slightly elevated, highly elevated} \}$ with the core intervals "low"=[0,4], "normal"=[5,18], "slightly elevated"=[20,30], "highly elevated"=[35,50]

which results in a vague set of membership functions showed in Fig. 2.

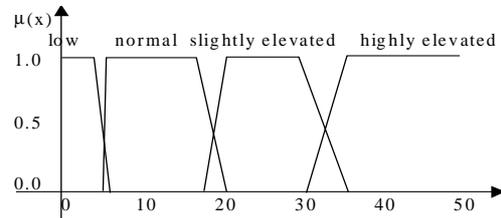


Fig. 2 The vague set of the linguistic variable SGOT

For each term of a vague set, we have to attach a name. This name or linguistic term should follow some restrictions. According to section 2.1, our system provides

- ◆ *Exact notation.* The membership function is defined exactly for each value of a vague variable.
- ◆ *Semi-quantitative notation.* The multiple values of the example in Table 1 can be mapped to standard membership functions as shown in Fig. 2.
- ◆ *Quantitative, categorical notation.* This can be easily mapped by using a rectangular membership function.

For the sake of an easy, robust interface for network initialization purposes (information stream from left to right in Fig. 1) we do not assume any knowledge about appropriate membership functions by the user which is certainly true for most of the physicians. Instead, we use the most simple membership function which is in coherence with the medical expert intuition: a simple trapezoidal bell shaped function which is directly assigned to the Radial Basis Function of a RBF neuron. A trapezoidal shaped function with its final support converges faster than Gaussian RBFs [3]. Additionally, the ramp inclination provides a simple mean of shifting from vague notation to categorical notation. In conclusion, by using a trapezoidal function as standard membership function we can easily satisfy all the demands of the medical interface. As free parameters the lower and upper core limits [min,max] are chosen as the medical range limits while the ramps are assumed standard.

So far, we considered the mapping of vague terms to single radial basis functions. Now, several terms of different variables might also be coupled to a vague rule like "IF SGOT=strongly_elevated AND GGT=pathological THEN sensitive_for_liver_disease". This can be modeled by combining the two one-dimensional membership functions to a two-dimensional radial basis function of one RBF neuron. When the two terms are active, the RBF neuron will output a high activity. So, the corresponding RBF network is a conventional two layer network; the first layer implements the rules in parallel while the second layer weights each conclusion and combines it to the final diagnosis.

2.3 Transferring the Neural Network knowledge to the User

The initialization of the network by the user is one of the interface functions. The other, more important function is to provide the user with the results of the network learning in an understandable way. Here, we encounter several problems. The first one is the representation of the network state by the medical linguistic terms, and the second problem which we have to deal with is the vast number of possible rules.

Network state approximation with linguistic terms

Since we might use a growing neural network approach (see e.g. [4],[5]), the network

might slightly or heavily differ from the initial configuration

- ◆ The location and extensions of the membership functions may have changed; they may have shifted or expanded.
- ◆ The number of neurons and therefore the number of possible membership functions may be smaller or bigger after the learning process

These two problems are hard to solve. Certainly, we might ignore all user supplied linguistic information and generate pure synthetic linguistic terms like "blood sugar \in [250,400]" which indicates that there is a vague term with the core region in the interval of [250,400] of the variable "blood sugar". But this approach is not very user friendly and ignores completely nearly all interface design demands listed in section 2.0. Instead, we developed the following mechanisms:

1. The core region of a vague term is chosen such it covers the $[c-2\sigma, c+2\sigma]$ interval of the data. The ramps are chosen such that they fit the $[c-3\sigma, c+3\sigma]$ interval.
2. If the core region has been broadened, the linguistic term " t " is changed to "roughly t " (abbreviated " $\sim t$ ") which reflects this property in everyday life.
3. For simplifying a rule we sort the terms of one variable according their order.
4. If the new membership function covers two old core regions t_1 and t_2 , we might take as new term t the combination of the old ones $t = t_1 \text{ OR } t_2$. Nevertheless, if we cover more terms than two, the linguistic approximation should simplify the term by using a "TO" binding between the first and the last term.
5. Small shifts of terms can be reflected by the preposition "very", but we have to pay attention if the term is higher or lower than the norm.
6. Negated terms are too complicated for the ordinary user. We omitted all negations and use only positive linguistic terms.

The linguistic approaches mentioned so far do facilitate the task of mapping new terms to old ones, but do not avoid this completely. In the case where the core region s of the new term is not sufficiently covered by the old core regions s_i , i.e. when the linguistic approximation quality $Q = (s_1 + \dots + s_n) / s$ is less than 50%, we have to generate a new synthetic term " $t[\text{min}, \text{max}]$ " with t the linguistic variable and min, max the values of the lower and upper bounds of the core region of the vague variable.

Since this approach might not be acceptable for all users, the interface allows to "freeze" some of the user defined terms during training.

Reducing the number of rules

The neural network might have found a very good classification configuration containing several hundreds of neurons. Following our concept of direct mapping between the rules and the network, this means also a vast number of rules. Therefore, an important part of the interface consists also of simplifying the rules in order to be readable.

We took the following approach to the problem:

1. All generated rules are weighted by the learned weights of the second network layer. Thus, the most important rules are presented first. By a user-definable threshold all rules beyond a certain importance are not presented.
2. The reference interval or "normal" term is superfluous and should not be mentioned in a rule. Instead of presenting all terms, only the extraordinary terms are used. Additionally, all rules which becomes empty by this elimination are dropped.
3. Noisy training data may cause a target region to cover a region outside the possi-

ble range of the vague variable. In this case, the rule is dropped.

4. When the target region is nearly identical to the possible range of the vague variable, the rule is always true. In this case the rule can be dropped.

3 Simulation and results

In this section we present the results of an implementation of the interface and its usage for the diagnosis of liver disease.

3.4 The implementation

The interface design principles were implemented by the means of a standard Visual C++ development system and can be downloaded from [9]. As network type we used here the Dynamic Decay Adjustment (DDA) network [6],[7],[8] based on common RBFs and the Reduced Coulomb Energy (RCE) net which implements a growing network, depending on the training data. This network uses also trapezoids and rectangles as Radial Basis Functions. It

- ◆ adapts the different diameters of the bell shaped functions differently in the dimensions by *individual learning rates*.
- ◆ implements the weights of the rules in the second layer as the number of training samples where this rule is valid, i.e. the share of the rule.
- ◆ *adds* a new RBF neuron to the net each time the diagnostic approximation error is too high. The new neuron RBF has as weight vector (center) the erroneous pattern coordinates and a width which does not include the neighbor centers.
- ◆ *shrinks* a RBF width if the neighbored data belongs to another diagnostic class.

Each trapezoidal RBF is represented by two rectangular RBFs: one for the core region and one for the ramps.

3.5 A diagnostic example

As demonstration example we took the liver disease data of BUPA Medical Research. Here, we have 345 data classified for 144 (41.74%) cases to "less sensitive for liver disease" and for 201 (58.26%) cases to "sensitive for liver disease". The membership of each class should be expressed by only a few, simple rules.

The following six input variables were used

- ◆ MCV mean corpuscular erythrocytical volume, reference [76-96] fl
- ◆ AP alkaline phosphatase [60-170] IE/l
- ◆ SGPT glutamat-pyruvat-transaminase in serum [5-22] IE/l
- ◆ SGOT glutamat-oxalacetat-transaminase in serum [5-18] IE/l
- ◆ GGT gammaglutamyltranspeptidase [5-28] IE/l
- ◆ AD Alcoholic drinks: the number of daily consumed alcohol portions. One portion is the equivalent of 250ml of a 5vol% alcoholic beverage, e.g. beer.

The system generated several sets of rules. To give an idea for the rules, we present those with a weight >6 which were extracted after the adaptation process for the single class "sensitive for liver disease". In brackets we show the weight of each rule.

[27] -- GGT is normal to pathological AND AD is few to many

[18] -- AP is pathologically too small AND SGPT is normal to elevated AND SGOT is normal to slightly elevated AND GGT is normal to pathological AND AD is few to many

[17] -- SGPT is ~ elevated to strongly elevated AND SGOT is slightly elevated to ~ strongly elevated AND GGT is pathological AND AD is many to very many

- [10] -- SGPT is ~ elevated AND SGOT is slightly elevated AND GGT is pathological AND AD is ~ moderate to many
- [9] -- MCV is [94 - 101] AND SGPT is normal to elevated AND SGOT is slightly elevated AND GGT is normal to pathological AND AD is few to many
- [9] -- SGPT is normal to ~ elevated AND SGOT is ~ slightly elevated to strongly elevated AND GGT is normal to pathological AND AD is moderate to many
- [7] -- SGPT is ~ elevated to strongly elevated AND SGOT is slightly elevated to strongly elevated AND GGT is normal to pathological AND AD is moderate to ~ many
- [7] -- SGPT is strongly elevated AND SGOT is ~ strongly elevated AND GGT is pathological AND AD is nothing to many

These are the first 8 rules of 32 covering 104 of the 184 samples. The remaining samples caused the generation of one rule per sample and were therefore discarded.

4 Discussion

The adaptive methods of Neural networks can solve many problems in everyday life. Nevertheless, it is seldom used due to the insufficient or often lacking user interface. In this paper we showed that for the important example of medical diagnosis a neuro-fuzzy approach allows to map the medical mental expert model of the diagnostic data to the input parameters of a neural network.

Furthermore, the diagnosis based on the data adapted resulting network parameters can be easily converted to user understandable phrases which use the terms and habitudes of the medical experts.

For the presentation of the rules, the most important ones are selected. Even those were too many. Additionally, the dropped rules each one representing only a small number of samples are too many but important: they represent half of the samples. Although we chose a constructive approach which generate rules only by the need to represent the data well, future work will integrate techniques for melting several regions (RBF neurons) into one.

References

- [1] Rüdiger Brause: Revolutionieren Neuronale Netze unsere Vorhersagefähigkeiten? Zentralblatt für Chirurgie 124, S. 692-698 (1999)
- [2] Lotfi A. Zadeh, Janusz Kacprzyk: Fuzzy logic for the management of uncertainty, John Wiley & Sons Inc., New York, 1992
- [3] Lei Xu, Adam Krzyzak, Alan Yuille: On Radial Basis Function Nets and Kernel Regression: Statistical Consistency, Convergence Rates, and Receptive Field Size; Neural Networks, Vol. 7, No.4, pp.609-628, (1994)
- [4] B. Fritzke: Growing cell structures - a self-organizing network for unsupervised and supervised learning. Neural Networks, 7(9): 1441-1460, 1994
- [5] U. Pietruschka, R. Brause: Using growing RBF nets in Rubber Industry Process Control, Neural computing & Applications, Springer Verlag, Vol 8, No. 2 (1999)
- [6] Michael R. Berthold, Jay Diamond: Boosting the Performance of RBF Networks with Dynamic Decay Adjustment, In G. Tesauro, D.S. Touretzky and T. K. Leen (eds): Advances in Neural Information Processing System, 7, Cambridge MA, MIT Press, 1995
- [7] Michael R. Berthold, Klaus-Peter Huber: From Radial to Rectangular Basis Functions: A new Approach for Rule Learning from Large Datasets, Internal Report, Universität Karlsruhe, 1995
- [8] Michael R. Berthold, Klaus-Peter Huber: Building Precise Classifiers with Automatic Rule Extraction, Internal Report, Universität Karlsruhe, 1995
- [9] <http://www.cs.uni-frankfurt.de/~brause/software/Analyst1.exe>