Contribution to the "IEEE International Conference on Neural Networks"
in San Francisco, March 28 to April 1, 1993

Classification of Parkinson Rating-Scale-Data Using a Self-Organizing Neural Net

T. Fritsch[1], P.H. Kraus[2], H. Przuntek[2], P. Tran-Gia[1]

[1]Institute of Computer Science, University of Würzburg, Am Hubland, D-8700 Würzburg, F. R. Germany
[2]Department of Neurology, University of Bochum, Gudrunstr. 56, D-4630 Bochum 1, F. R. Germany
[1] Tel.: +49-931-8885513, Fax: +49-931-8884601,
e-mail: [fritsch-trangia]@informatik.uni-wuerzburg.dbp.de
[2] Tel.: +49-234-5093925, Fax: +49-234-5092414

***Abstract*—The overall score of assessment of clinical stages with rating scales is problematic. The unknown weights of the items and the non-linearity of the interrelation between real expression of the symptoms and rated values cause only weak correlation between sum-scores and an integrative assessment by an expert.**

In this paper we present an application of a self-organizing neural net of Kohonen type to the data of 666 de-novo Parkinsonian patients of a multi-center study. The data to be learned are the 10 items of the Webster rating scale and one additional item with 4 stages, following the classification by Hoehn and Yahr. For reasons of comparison multivariate linear statistical methods have been applied to the data, yielding linear models, which are able to derive the Hoehn and Yahr staging from the staging of the Webster rating scale. The methods succeeded with a quote of correct classification of about 50 %. In contrast to these unsatisfying results, a Kohonen net with 40x40 neurons achieved a surprisingly high classification rate of approximately 90 % for the 4 stages of Hoehn and Yahr. The patients, which could not be classified correctly, can be identified as outliers. Further experiments on generalization,variation of net dimensions and single feature representation have been carried out, producing different results. For practical purposes, a similar proceeding with an integrative expert rating as 'predictor' can yield a better alternative to the assessment of sum-scores.

I. Introduction

In medical diagnosis a number of problems occur concerning the correct assignment of symptoms or combinations of symptoms to a specific form of expression of a disease. It is very difficult to diagnose the specific form, if the mechanisms of the disease are unknown. Furthermore, the expert generally has to validate data, which stem from patient self-observation. It is obvious, that this kind of classification is rather vague or in technical terms, the data are superimposed with "noise". Therefore non-linear classification methods are required, which take into account the uncertainty of this type of data. Up to now only sum-scores of the rating-scale values are used, which do not regard the correlations between the different real expressions of the symptoms and the rated values.

II. Problem statement

In our study [1], whose medical implications still have been discussed at the annual meeting of the German Society of Neurology, we investigated data of 666 de-novo Parkinsonian patients of a multicenter study at the University of Bochum [3] concerning the question how different scores reflect the same clinical state. The rating scale used was that of Webster [6], consisting of ten items, which are the following ones: bradykinesia, rigidity, posture, swinging arms, walking, tremor, gesture, seborrhea, speech, and independence. Each of these 10 items possesses 4 stages. More integrative is the rating of Hoehn and Yahr [2], which puts main emphasis on different symptoms. Due to the detected weak symptoms of the patients, we found only Hoehn and Yahr stages from 1 to 4 (129 patients stage 1, 265 patients stage 2, 231 patients stage 3 and 41 patients stage 4.) All 10 items of the Webster rating scale and the Hoehn and Yahr rating were analysed with multivariate linear statistical methods like multiple regression and discriminant analysis in an explorative way. It should be estimated, to what degree the results of both scales represent the same information. Both proceedings yielded linear models which are used to "predict" the Hoehn and Yahr stages from the staging of the Webster rating scale items. Multiple regression as well as discriminant analysis yielded a quote of correct classification of about 50 %. Discriminant analysis was superior to multiple regression for extreme ratings (Hoehn and Yahr stage 1 and 4.) For reasons of comparison, it should be noticed that random classification would deliver values of 25 % for the 4 stages of the Hoehn and Yahr scale.

III. Application of a neural net

Since these results are very unsatisfactory, a non-standard kind of non-linear classification is proposed. Generally it is the aim of diagnosis to recognize the expression of a specific type of the Parkinson syndrom in an early state of its evolution. This aim can only be reached, if the vagueness of the input data can be separated from their characteristics. Therefore the idea is, to use an artificial neural net which is capable to extract characteristic features from data though having only human-made observations of the Parkinson syndrom. These observations are blotted out by individual self-assessments which can possibly be wrong and cause uncertainties. They can be interpreted as an abstract kind of noise. The task of classification the Hoehn and Yahr stages from learned input vectors, whose components are equivalent to the items of the Webster rating scale was performed by a neural net of Kohonen type, following the application of H.Ritter and T.Kohonen on semantic relationships between abstract entitities like words or sentences [4]. In this application the semantic relationships in the data are reflected by the relative distances of corresponding neuron clusters in the topologically ordered map, which represent different semantic entities. As the self-organizing feature map algorithm of Kohonen is well-known in the neural net literature we refer for detailed information to [4] or [5]. A brief description of the algorithm is given to outline the process of non-linear classification.

The net consists of 2 layers of neurons, which are fully interconnected. Each neuron of the input layer is connected with each neuron of the mapping array. The number of input neurons is determined by the dimension of the input vectors, 10 components for the Webster items and 4 for the Hoehn and Yahr stages. The number of neurons in the mapping array has to be chosen suitably. Some experiments will soon show, which problem-dependent number of neurons will suffice, since further increasing of the neuron number would not improve the classification quality.

A. Algorithm

The algorithm can be characterized by the following steps:

1. Presentation of a new input vector $\mathbf{v(t)}$
2. Computation of the distance d_j between all input neurons and all mapping array neurons j according to
$$d_j = \sum_{i=1}^{N}(v_i(t) - w_{i,j}(t))^2$$
with $v_i(t)$ as the $i-$th component of the N-dimensional input vector and $w_{i,j}(t)$ as the connection strength between input neuron i and mapping array neuron j at time t corresponding to the Euclidean metric.
3. Choosing the mapping array neuron j^* with minimal distance d_{j^*}
4. Update of all weights, restricted to the actual topological neighbourhood $N_{j^*}(t)$
$$w_{i,j}(t+1) = w_{i,j}(t) + \eta(t)(v_i(t) - w_{i,j}(t))$$
for $j \in N_{j^*}(t)$ and $1 \leq i \leq N$. Here $\eta(t)$ represents a monotonically decreasing function of the actual environment of the winner neuron.
5. Iteration of the steps above until a predetermined error criterion is met.

In our application the 14-dimensional input vectors consist of 10 Webster items with values lying in the interval between 1 and 4 and 4 additional components for the stages of Hoehn and Yahr, each having values either 1 or 0. This coding scheme is similar to that one in Ritter [4]. In this application the input vector, say $\mathbf{v}$, is assumed to be the concatenation of two or more fields, one specifying a symbol code, denoted by $\mathbf{v_S}$, and the other one an attribute set, denoted by $\mathbf{v_a}$, respectively. The input vectors used in this paper are constructed in a similar way. The following equation illustrates in a vector notation that the encodings of the Webster items, denoted by $\mathbf{v_W}$, and the Hoehn and Yahr stages, denoted by $\mathbf{v_H}$, form a vector sum of two orthogonal components.

$$\mathbf{v} = \begin{bmatrix} v_W \\ v_H \end{bmatrix} = \begin{bmatrix} v_W \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ v_H \end{bmatrix}$$

The main idea of this symbolic mapping is that the two parts are weighted adequately. This must be done such that the weighted Hoehn and Yahr part or the weighted Webster part can predominate each other, depending on which relationship is to be examined. As usual for the class of Kohonen nets, the entire input vector is normalized to an Euclidean length of one and the Gaussian function is used as neighborhood function.

B. Net dimension and weighting of components

The experiments we carried out focus on the following questions:

1. Which net dimension is appropriate?
2. Which weighting of the components is suitable?
3. Is the net capable to generalize the features, i.e. to predict an unknown second data set from the learned data set?
4. Does the net show generalization capability when learning and prediction are restricted to an unique data set?
5. Can outliers be determined?
6. How does the net represent items of the Webster rating-scale, which are weakly correlated?

For reasons of computing time we started with nets of dimension 20x20 neurons with fixed learning rate $\eta = 0.8$ and varied the weighting of the Hoehn and Yahr (HY)- components between 0.001 and 10. Although the recognition rate was low (about 50 %), it turned out that the weighting with 1 led to the best results, but high weighting resulted in a map dominated by the 4 HY-components. The recognition rate was only about 15 % in this case. Weights lower than 1 did not reduce the recognition rate seriously.

Neural net with 30 x 30 neurons			
type	*# patients*	*# recognitions*	*percent*
1	129	98	76.0
2	265	170	64.2
3	231	141	61.0
4	41	35	85.4
total	666	444	66.7

Table 1: Results for neural net with 30 x 30 neurons

Neural net with 40 x 40 neurons			
type	*# patients*	*# recognitions*	*percent*
1	129	104	80.6
2	265	224	84.5
3	231	203	87.9
4	41	40	97.6
total	666	571	85.7

Table 2: Results for neural net with 40 x 40 neurons

The net size was then increased by steps of 10 neurons in each direction starting with 30x30 neurons. The experiments varying the net dimensions between 20x20 and 80x80 neurons by steps of 10x10 have shown that nets both with 40x40 and 60x60 neurons delivered good results. Further increasing of the number of neurons up to 80x80 neurons did not achieve effective improvements. The results are shown accordingly in Tab. 1 and Tab. 2.

C. Generalization

The generalization capability of the net was tested in two cases:

i) The first task was the prediction of an unknown data set, consisting of 520 input vectors obtained by data of the same patients half a year later. The data of those patients, having finished the medical treatment were not included in the data set used. The aim was to find out, whether the net had represented common features of the data sets. The values of the data changed according to the influence of the medication process, so we expected the recognition rate to be reduced. The results of the tests confirmed this assumption, as can be seen in Tab. 3.

ii) As a second experiment another neural net learned the new data separately. It delivered results similar to those above. A further test should show whether a better generalization can be achieved if

Neural net with 40 x 40 neurons			
type	*# patients*	*# recognitions*	*percent*
1	168	66	39.3
2	248	103	41.5
3	89	45	50.6
4	15	7	46.7
total	520	221	42.5

Table 3: Generalization results for neural net with 40 x 40 neurons

Neural net with 60 x 60 neurons			
type	*# patients*	*# recognitions*	*percent*
1	129	106	82.2
2	265	238	89.8
3	231	216	93.5
4	41	40	97.6
total	666	600	90.1

Table 4: Results for neural net with 60 x 60 neurons

the same old data set would randomly be splitted in two or three parts. One of these parts should be learned by the self-organizing feature map and the others should be used as unknown data, serving as test vectors. In this experiments the recognition rate further reduced down to 40 % in the average.
The conclusion we obtain from this case study is that the neural net shows low generalization capability and can be described as a "specialist" for a given distribution of input vectors. Nevertheless the net shows optimal results when it is applied to the set of remaining input vectors. Therefore it can serve as a tool to discover outliers i.e. those patients which cannot definitely be assigned to one of the 4 Hoehn and Yahr stages. In our case only 10 % of outliers are detectable.

D. Representation of single features

Since the items of the Webster rating-scale are highly correlated except two of them, tremor and seborrhea, we put emphasis on the question how the net represented these items, especially tremor. The net we used consisted of 60x60 neurons with a recognition rate according to Tab. 4.

The final mapping of the neural net is shown in Fig. 1, where the bestmatching neurons for the different HY-types are represented by 4 symbols, where bestmatching means that those neurons are depicted by a special symbol, whose HY-component is 1.

The assignment of the 4 different HY-types to the 10 level gray-scale picture representation of item 6 (tremor) is shown in Figs. 2- 5. The gray scale levels vary between 0 (low) and 9 (high), represented by white and black pixels.

It can be deduced that larger connected areas in the mapping array of neurons in Fig. 1 represent local subgroups, whereas in mixed areas no definite interpretation can be made.

From a clinical point of view it is comprehensible that in Figs. 2- 5 the areas with high responsibility for tremor can be assigned to HY- types 1 and 2, whereas for HY-types 3 and 4 no significant intersection is visible.

IV. Conclusions

Applying a self-organizing neural net to Parkinson data shows significant improvements of the recognition rate, comparing with linear predicting models. The application of the neural net to unknown data is less promising. The neural net appears to be a specialist for the learned data set. Patients which cannot be represented by the neural net can be interpreted as outliers. Regarding the feature "tremor" it is shown, that local subgroups of neurons exist, which are comprehensible from clinical viewpoint. The results justified a combination of an expert assessment with a proper evaluation of rating-scores.

Acknowledgment: The authors would like to thank V. Heuveline and J. Wolf for their emphasized help in the implementation of the simulators.

References

[1] **Kraus P.H., Fritsch T., Tran-Gia P., Przuntek H.,** Classification of rating-scale-data using an artificial neural net, 65. Annual meeting of the German Society of Neurology, Sep. 1992, extended abstract published in Journal of Neurology

[2] **Hoehn, M.M., Yahr M. D.,** Parkinsonism: onset, progression and mortality, Neurology 17, pp 427-442, 1967

[3] **Przuntek H., Welzel D., Schwarzmann D., Letzel H., Kraus P.H.,** Primary combination therapy of early Parkinson's disease, European Neurology 32 (suppl.1), p. 36-45, 1992

[4] **Ritter H, Kohonen, T.,** Self-Organizing Semantic Maps, Biological Cybernetics 61, p. 241-254, 1989

[5] **Ritter H., Martinetz T., Schulten, K.**, AN INTRODUCTION TO THE NEURAL INFORMATION PROCESS OF SELF-ORGANIZED NETWORKS , Addison-Wesley, 1991

[6] **Webster D. D.**, CLINICAL ANALYSIS OF THE DISABILITY IN PARKINSON'S DISEASE, Modern treatment, Vol. 5, pp. 257-282, 1968

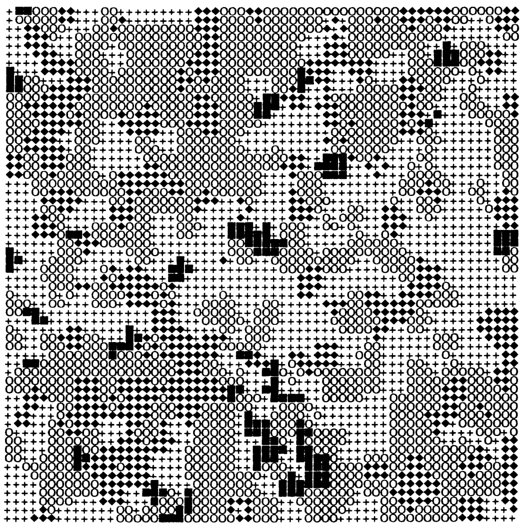

Figure 1: *Final mapping of HY-types in self organising feature map*

Type 1: Representation by solid ◇

Type 2: Representation by ∘

Type 3: Representation by +

Type 4: Representation by solid □

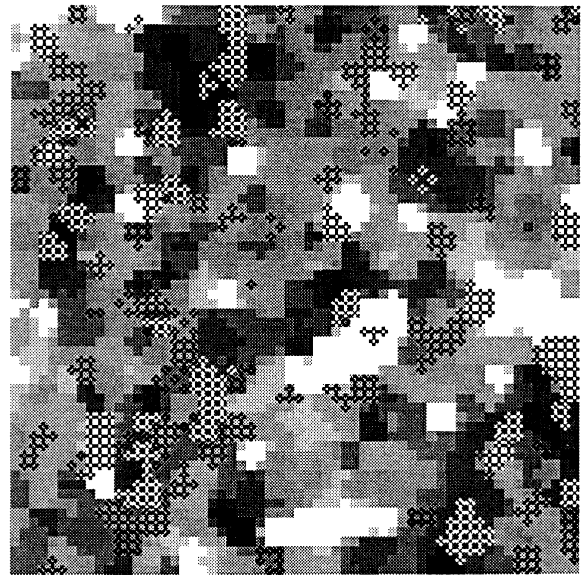

Figure 2: feature 6 map, overlapped with bestmatching neurons for type 1

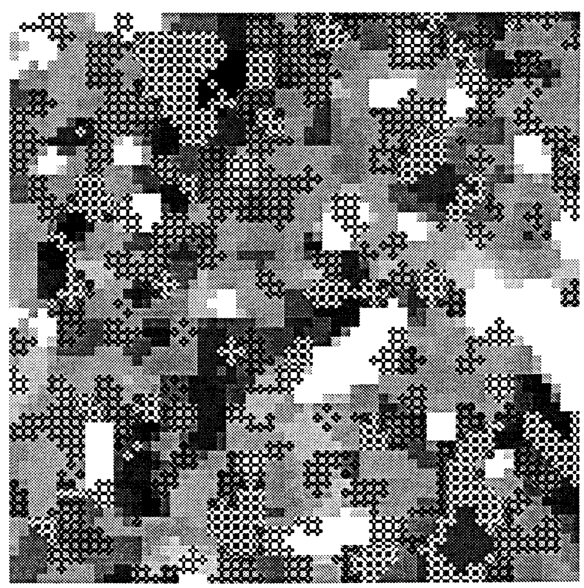

Figure 3: feature 6 map, overlapped with bestmatching neurons for type 2

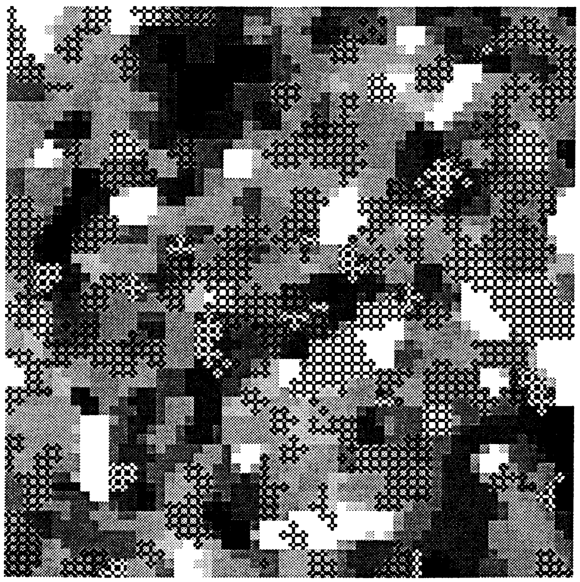

Figure 4: feature 6 map, overlapped with bestmatching neurons for type 3

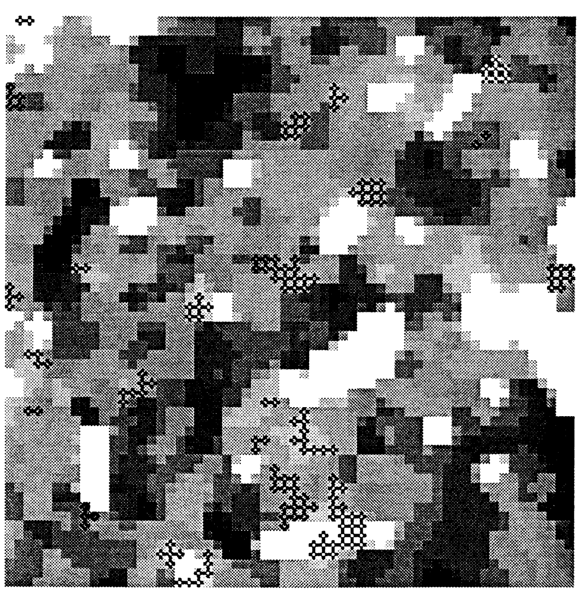

Figure 5: feature 6 map, overlapped with bestmatching neurons for type 4