FINDING STRUCTURE IN FITNESS DATA

Satu Lipponen, Timo Mäkikallio*, Mikko Tulppo*, Juha Röning

Machine Vision and Media Processing Group, Infotech Oulu, University of Oulu, Finland
*Merikoski Rehabilitation and Research Center, Oulu, Finland

Abstract. A method for measuring aerobic fitness is considered. The data used in the approximation contain accurate measurements of maximum oxygen uptake as reference values and physical features including R-R intervals of the heart beat measured at rest. The physical system of the human being is highly nonlinear, the number of the variables easily becomes very large and the information content of the data is quite poor.

Effective methods are therefore needed for finding information. The fitness approximation task is first visualized with self-organizing maps and maximum oxygen uptake is then approximated by means of neural networks.

Since the number of variables in the data set is impracticably large relative to the number of subjects, it is reduced by linear and nonlinear principal component analysis. The results show that this reduction in dimension can accelerate and intensify the learning process in the neural network.
Key words: PCA, NLPCA, neural networks, aerobic fitness.

1. Introduction

An individual's aerobic fitness can be approximated using certain physical features. Weight, height, age, sex and heart rate, for example, serve moderately well to characterize the maximum oxygen uptake, which can in turn be regarded as a measure of fitness.

The physical system of the human being is highly nonlinear in nature, and therefore the method used in approximation should be able to recognize nonlinear relations between variables. There is usually no a priori knowledge available about the nature of the nonlinearity, however, and the system cannot be linearized with special transformations. Neural networks will solve this problem.

Self-organizing maps (SOM) can be used to visualize the task of fitness approximation, and thereby to judge whether the task is realistic. Men and women have to be considered separately; otherwise sex becomes the only categorial variable and will dominate the mapping and confuse other features.

The number of variables is large (18) and they are rather poor in information, so that dimension reduction should be used before approximation to intensify the learning process. Nonlinear principal component analysis (NLPCA) is considered for this problem because of the highly nonlinear relations between most of the variables. Linearly correlated variables (7) are first reduced with linear PCA, and the resulting two principal components are used together with other features (11) as inputs to the NLPCA network. NLPCA is performed with an autoassociative neural network containing three hidden layers. The bottleneck layer represents the lower dimensional projection of the data.

2. Aerobic Fitness

The most important determinant of an individual's capacity to perform prolonged muscular exercise is maximum aerobic capacity, which indicates the working efficiency of the respiratory and circulatory organisms. Thus one of the methods most commonly used to approximate an individual's aerobic fitness is to measure maximum oxygen uptake (MAXL, l/min) (Niemelä 1983; Tulppo *et al.* 1996; Väinämö *et al.* 1996).

If aerobic fitness is to be measured accurately, this can be done in a special clinic where clinical methods can be used to measure maximal oxygen uptake. These measurements are expensive and time-consuming, however. Another alternative is to take part in a less accurate fitness test by running or walking under a work load. It is obvious that an inexpensive and reasonably accurate system for measuring aerobic fitness is needed.

An expert can draw conclusions about an individual's fitness from that person's appearance, using physical features such as height, weight, age and sex for approximation purposes. There is also a lot of information available on the resting heart rate, which correlates moderately well with aerobic power (Kenney 1985; Väinämö *et al.* 1996). The hypothesis can be proposed that aerobic fitness can be estimated by reference to physical features and statistical features calculated from R-R intervals (Fig 1.).

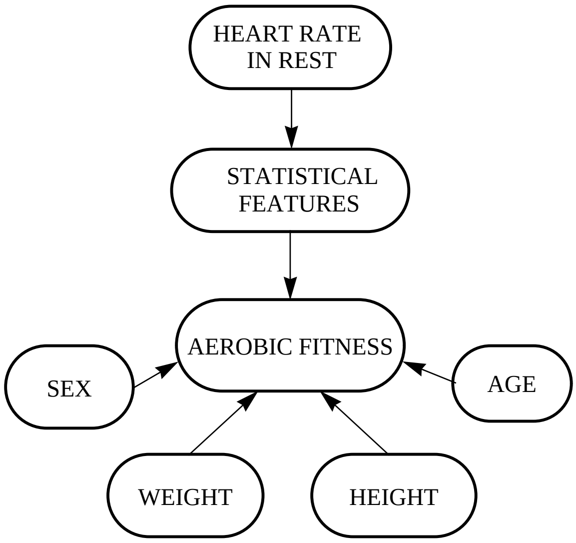

Fig 1. The hypothesis in fitness approximation.

The material for this purpose, obtained from the Merikoski Institute of Heath and Rehabilitation, Oulu, Finland, comprised 237 sets of R-R interval measurements and accurate oxygen uptake measurements (with a bicycle ergometer or treadmill) performed on adult men and women aged 15-65 years. All the subjects were healthy and none of them was receiving any medication.

The series of R-R intervals were passed through a filter to eliminate artifacts (intervals that differed from the last qualified interval by over 30%). Many statistical features were calculated from the filtered series. Features showing a high correlation with maximum oxygen uptake were selected.

The dimension of the feature space was reduced by linear and nonlinear principal component analysis. Since over 20 different statistical features can be calculated from R-R intervals, the number of variables would have been impracticably large. The best-correlated statistical and physical features (18) were chosen for fitness approximation, which was performed with and without dimension reduction. The results were compared with an approximation in which only physical features (5) were used.

One of the problems is to recognize subjects who have a low maximal oxygen uptake, as the approximations are often too optimistic for these subjects. Lean subjects who do not perform physical exercise are especially difficult to recognise, and it is therefore important to find features that will facilitate their recognition. For example, if the bmi (body mass index) is less than 20 and if the mean of the R-R interval series is less than 1 s (the pulse is higher than 60 beats/min at rest), the subject's maximal oxygen uptake will probably be less than 3 l/min.

If the R-R variance or SD-SD (the st. deviation of the differences between consecutive R-R intervals) is high, the subject will have a higher than average fitness level. The pulse is not such a good discriminative feature. Respiration is also regarded as a feature, but it is difficult to derive from R-R intervals at rest because the effects of other physical features interfere with it. The situation is different during exercise.

3. Related work

Principal components are commonly used with image analysis, where vast amounts of data are dealt with. PCA is also useful as one of a number of orthogonal transformations for data storage, e.g. if periods of ECG data are stored in digital form (Ahmed *et al.* 1975).

Data used in multidimensional regression analysis (whether performed statistically or with neural nets) should be as orthogonal as possible to avoid the multicollinearity problem. A few variables can be carefully selected, or orthogonal transformations can be used. Khoshgoftaar and Szabo (1994) report that principal component analysis can improve the predictive quality of neural network modelling, and Kurtanjec (1995) proposes PCA for pattern compression in the on-line estimation of industrial biomass production.

Nonlinear principal components are used in modelling many industrial processes. An important characteristic of many processes is that they are data rich but information poor. These processes are usually nonlinear in nature as well. Dong and McAvoy (1996) describe a NLPCA application in process monitoring.

In many applications principal components are constructed by means of neural networks in order to serve as effective tools when dealing with the nonlinear PCA, but traditional methods are sufficiently effective in the linear case.

4. Data visualization

The preprocessing of the data involves visualization before any modelling takes place. Self-organizing maps are well suited for this task, as they can be used to define a mapping from input data onto a two-dimensional lattice of nodes. This kind of SOM will enable a complex set of experimental data to be analysed even if the relations between the data elements are highly nonlinear. SOM is an unsupervised nonlinear clustering method in which the content of the clusters can be characterized by means of proper labelling.

For further information, the excellent book of Kohonen (1995) is recommended. The theory of self-organizing maps is discussed here only briefly. SOM is a network with an input layer and a competitive layer. No hidden or output layers are needed. The competitive layer of the network consists of a lattice formed by neurons. A special neighbourhood function is used to determine how the weights of the neurons should be updated during learning. For each subject in the input the winning neuron and its topographically close neighbours are rewarded in accordance with this neighbourhood function.

A hexagonal map is usually preferred for visualizing because its structure does not favour perpendicular directions. The lattice should to be rectangular (oblong) rather than square; otherwise the network may not stabilize during the learning process.

The maximum oxygen uptake or fitness class can be used as a label for each subject, and it can then be visually estimated how well the different levels of fitness can be distinguished. The map shows how the subjects forming the data units are related to each other. Valuable information concerning the feasibility of fitness approximation by means of such data is also provided.

Two maps (12x9 and 9x6) for fitness data are presented in Fig 2. The dimension of the map should not be too small, or the desired information may not be seen, nor should it be too large, or too many neurons will remain empty and it will be more difficult to interpret. The networks were constructed with SOM_PAK 3.1 because of its versatile visualization tools, the mapping being performed separately for men and women. It can be seen that the highest fitness level (5) is distinguishable moderately well from the lowest level, but the borders between the middle levels are not so clear. The intention is not to find levels of fitness, however, but to approximate the absolute value of maximum oxygen uptake, and since this absolute value is more difficult to visualize, mapping with fitness levels can be used to show that fitness can be approximated with such data.

a)

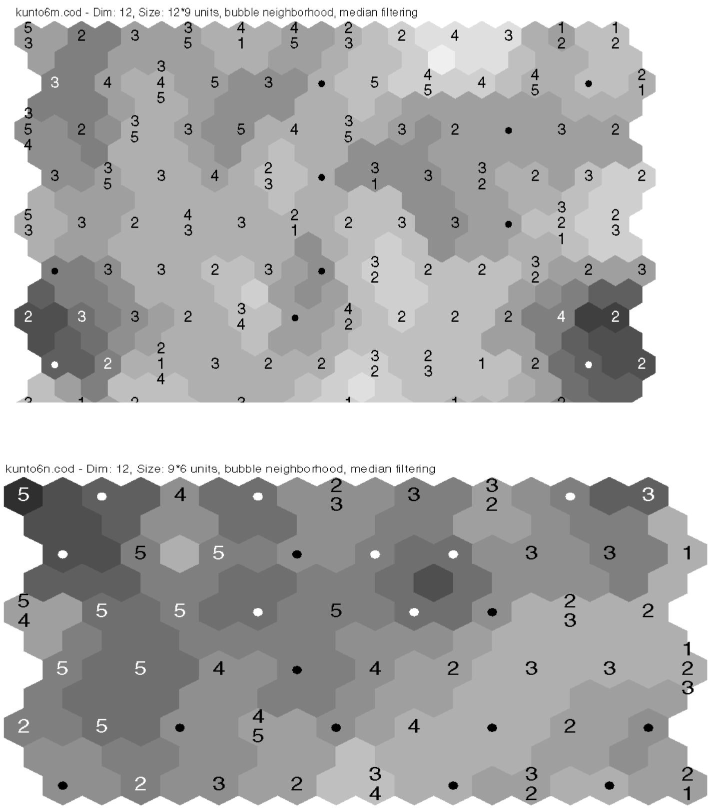

Fig 2. Self-organizing maps for men (a) and women (b). Labels 1-5 indicate the fitness level (1=lowest,..., 5=highest).

5. Principal Component Analysis

Principal component analysis (PCA) is a dimension reduction technique for use in multivariate statistical analysis which deals with data that consist of measurements applying to a number of individuals or objects. The number of variables in such cases is often impracticably large, and one way of reducing it is to take linear combinations of variables and discard those with small variances. The linear combination with maximum variance then becomes the first principal component, the second component is orthogonal to it and so on. PCA looks for a few linear combinations which can be used to summarize the data while losing as little information as possible.

Let Y represent a $n \times m$ matrix of data (n = number of observations, m = number of variables). PCA is an optimal factorization of Y into matrices T (scores $n \times f$) and P (loadings $m \times f$) plus a matrix of residuals E ($n \times m$)

$$Y = TP^T + E ,$$

where f is the number of factors ($f < m$). The Euclidean norm of the residual matrix E must be minimized for the given number of factors. This criterion is satisfied when the columns of P are eigenvectors corresponding to the f largest eigenvalues of the covariance matrix of Y.

PCA can be viewed as a linear mapping of the data from $\Re^m$ to a lower dimensional space $\Re^f$. The mapping has the form

$$t = yP ,$$

where y represents a single row of Y and t represents the corresponding row of T. The loadings P are the coefficients for the linear transformation. The projection can be reversed back to $\Re^m$ with

$$y' = tP^T ,$$

where $y' = y - e$ is the reconstructed row of data. The smaller the dimension of the feature space, the greater the resulting error E.

When using linear PCA the variables involved should be linearly correlated. If they are correlated nonlinearly, polynomial principal components can be used instead. In simple cases nonlinear transformations can be used before PCA, but polynomials represent only a very limited class of nonlinear functions. Neural networks can be used to solve the problem.

In the nonlinear case the mapping has the form

$$t = g(y) ,$$

where g contains f individual nonlinear functions $g = \{g_1, \ldots, g_f\}$, and g_i is the ith nonlinear factor of y.

The inverse transformation is implemented by a second nonlinear vector function $h = \{h_1, \ldots, h_m\}$ that has the form

$$y' = h(t) .$$

The loss of information is again measured by $e = y - y'$ and is analogous to PCA. The functions g and h are selected to minimize the Euclidean norm of E (Kramer 1991; Mardia *et al.* 1979).

Nonlinear principal component analysis can be performed with an autoassociative neural network, which has a special hidden layer called the bottleneck layer. The network has three hidden layers altogether, and its input is used as the desired output (Fig 3.). The network is therefore supervised in nature and can be taught with a back-propagation learning algorithm.

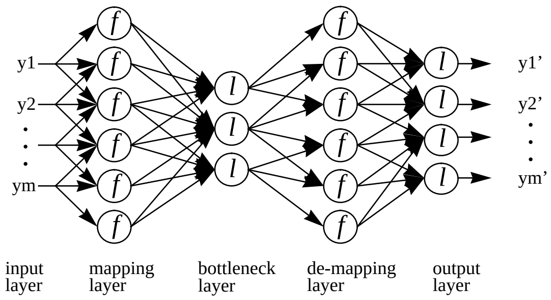

Fig 3. An autoassociative NLPCA network with three hidden layers. Transfer functions *f* are nonlinear and transfer functions *l* are linear.

As no assumptions are needed about the nature of nonlinearity between the variables, the network can be used in situations where common transformations (e.g. logarithm, square root) cannot be used. The nonlinearity is introduced into the network by sigmoidal transfer functions in the mapping and de-mapping layers.

The bottleneck layer will perform the dimension reduction, because the number of neurons in this layer is smaller than that in the input and output layers, so that the network is forced to develop a compact representation of the input data. Nonlinear activation functions should be used in the mapping and de-mapping layers, or else the network will produce linear principal components, and the iterative calculation for the linear case is not very effective compared with traditional straightforward methods.

The goal of the network is to minimize the error function

$$E = \sum_{j=1}^{n} \sum_{i=1}^{m} (X_i - X'_i)_j^2 ,$$

where X_i is an observation in the data set X and X'_i is an output of the network (Kramer 1991).

6. Approximating Maximum Oxygen Uptake

Linear regression can be used when the features have a linear correlation with the dependent variable. The regression can be formulated as

$$y_i = f(x_i, \theta) + \varepsilon_i ,$$

where i=1, ... ,n and $\theta = (\theta_1, \dots ,\theta_k)^T$ is the vector of the parameters to be estimated. The errors ε_i , i=1, ... ,n are i.i.d. with a mean of zero and an unknown variance σ^2 . The function f can be linear or nonlinear (Ratkowsky 1983).

Since the relations of many features to aerobic fitness are nonlinear, good results cannot be obtained using a plain linear regression model. Nonlinear regression does not guarantee an optimal solution, however, and it is very complicated. The functional expression has to be written, good initial values have to be found for the parameters and the derivatives of the model may have to be specified with respect to the parameters.

Neural networks are nonlinear in nature and can perform approximation better than traditional mathematical models. In the present modelling the method has to be nonlinear, but because the nature of the nonlinearity is not known, the proper nonlinear model cannot be formulated in the traditional manner.

The 3-layer MLP (multilayer perceptron with one hidden level) is a model of the form

$$y = f_2(W_2 f_1(W_1 x))$$

which is fitted to the set of couples x,y comprising the training set. The transfer function f_1 in the hidden layer is usually nonlinear, differentiable and increasing. Sigmoidal functions are often used. When performing regression with neural networks, the nonlinearity is introduced into the model by means of transfer functions. The function f_2 in the output layer may be linear or nonlinear, and the weight matrices W_1 and W_2 are estimated (Murtagh 1994).

7. Results

A Sun Ultra 2 workstation and SAS 6.12 software were used to construct the linear principal components and MATLAB 5.1. for the neural networks. Following this a MLP network with the Levenberg-Marquardt algorithm and one hidden level was employed to approximate the maximum oxygen uptake. Backpropagation with momentum and an adaptive learning rate were also tested.

Before the NLPCA stage, PCA was performed on the linearly correlated variables. Seven linearly correlated variables were reduced to two linear principal components, which accounted for almost 99% of the total variance. These principal components were used, among other variables, as inputs to the NLPCA network. In this way the resources of the NLPCA network are directed towards the nonlinearities.

If only statistical variables and two linear principal components were used as inputs to the NLPCA (8 variables), only 3 nonlinear principal components were needed together with 5 physical variables to approximate the maximum oxygen uptake.

First MAXL was approximated with 5 physical features to check whether statistical features were needed or not. The correlation between MAXL and the predicted value for the training set was 0.84 and that for the test set 0.82 (Fig 5a.). The network had 9 neurons in the hidden layer and was trained for 20 epochs.

With the NLPCA 13 variables were reduced to eight principal components which were used more successfully to approximate maximum oxygen uptake. The correlation between MAXL and the predicted value for the training set was 0.92 and that for the test set 0.87 (Fig 4a.). The network also had 9 neurons in the hidden level and was trained for 19 epochs.

When using three nonlinear principal components extracted from the statistical variables along with the physical features the correlations were 0.92 for the training set and 0.89 for the test set (Fig 4b.). The network had 8 neurons in the hidden layer and was trained for 16 epochs.

If 11 original variables and 2 linear principal components were used the correlations were 0.93 for the training set and 0.89 for the test set (Fig 5b.), but the larger number of inputs greatly increased the execution time. The network had 8 neurons in the hidden layer and was trained for 20 epochs.

As the approximating neural networks were of almost the same size, the differences in execution time between the networks were caused by the numbers of inputs. Dimension reduction cuts down the number of inputs and therefore shortens the execution time. The advantage is considerable relative to the information loss caused by the reduction. When approximating MAXL with 13 original variables the execution time was over twice that with six nonlinear principal components and the correlations were almost the same.

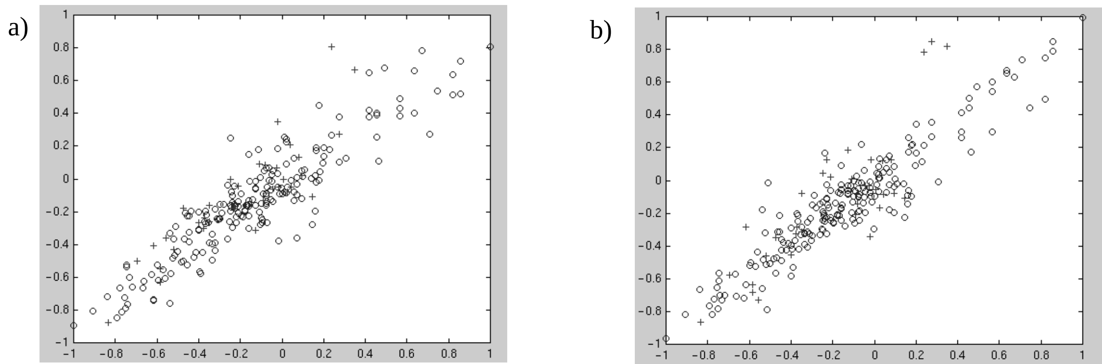

Fig 4. Approximation results (a) with eight nonlinear principal components and (b) with three nonlinear principal components and five physical features (‘o’ is the training set and ‘+’ the test set). The target values are on the x-axis and the predicted values on the y-axis.

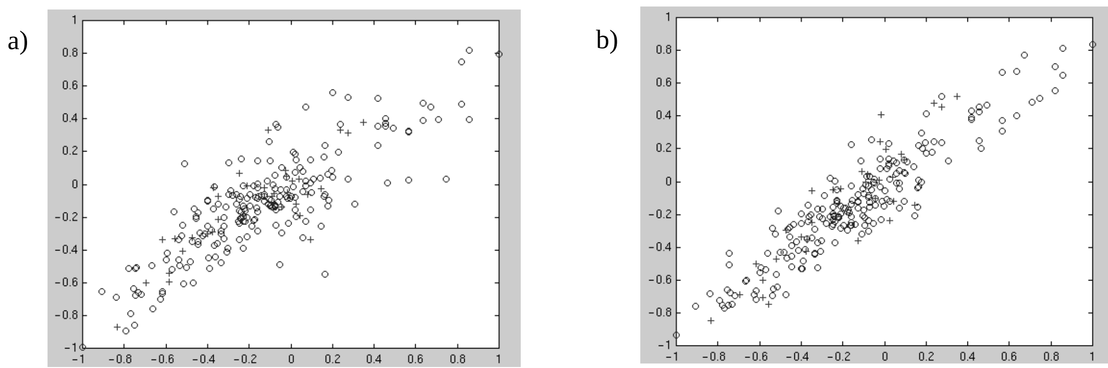

Fig 5. Approximation results (a) with five physical features and (b) with 13 original features.

8. Discussion

It is difficult to approximate aerobic fitness because the physical system of the human being is highly nonlinear and complex. Self-organizing maps were therefore used to investigate whether it was possible to approximate aerobic fitness with physical features and R-R series measured at rest. Neural networks were used in the approximation because of their nonlinearity, and these evidently give the best results for subjects. Like SOM also the neural network had some dispersion in estimating the middle class subjects. Depending on the input data set, there were also some problems in estimating the extreme values.

The correlations between the input sets did not vary much with these data, but the execution time was extended markedly when the number of inputs was increased. When dimension reduction is used, enough information remains to guarantee good approximation results, and in this way the approximation can be done in half of the time relative to the use of original features.

References

Ahmed N., Milne P., Harris S.G.: Electrocardiographic Data Compression via Orthogonal Transformations. IEEE Transactions on Biomedical Engineering **22** (1975) 484-487

Bishop C.: Neural Networks for Pattern Recognition. Oxford University Press, Inc. USA (1995)

Childers D.G.: Biomedical Signal processing, Selected Topics in Signal Processing. Editor Simon Haykin. Prentice Hall, Inc. USA (1989)

Diamantaras K.I.: Principal Component Neural networks: Theory and Applications. John Wiley & Sons, Inc. USA (1996)

Dong D., McAvoy T.J.: Nonlinear Principal Component Analysis - Based on Principal Curves and Neural Networks. Computers Chem. Engin. **10** (1996) 65-78

Gerbrands J.J.: On the Relationships between SVD, KLT And PCA. Pattern Recognition **14** (1981) 375-381

Kenney W.L.: Parasympathetic Control of Resting Heart Rate: Relationship to Aerobic Power. Medicine and Science in Sports and Exercise **17** (1985) 451-455

Khoshgoftaar T.M., Szabo R.M.: Improving Neural Network Predictions of Software Quality Using Principal Components Analysis. IEEE International Conference on Neural Networks - Conference Proceedings 5, NJ USA (1994) 3295-3300

Kohonen T.: Self-Organizing Maps. Springer-Verlag, Germany (1995)

Kramer M.A.: Nonlinear Principal Component Analysis Using autoassociative Neural Networks. AIChE Journal **37**(1991) 233-243

Kurtanjec Z.: Bioreactor Modeling by Neural Nets: Singular Value and Principal Component Decomposition. Proceedings of the 17th International Conference on Information Technology Interfaces, Pula, Croatia (1995) 443-448

Mardia K.V., Kent J.T., Ribby J.M.: Multivariate Analysis. Academic Press, Inc. London (1979)

Murtagh F.: Neural Networks and Related ‘Massively Parallel’ Methods for Statistics: A Short Overview. International Statistical Review **62**, 3 (1994) 275-288

Niemelä K.: Role of A Progressive Bicycle Exercise Test in Evaluating The Functional Capacity. Acta Universitatis Ouluensis, Finland (1983)

Ratkowsky D.A.: Nonlinear Regression modeling, a Unified Practical Approach. Marcel Dekker, Inc. USA (1983)

Tulppo M.P., Mäkikallio T.H., Takala T.E.S., Seppänen T., Huikuri H.V.: Quantitative Beat-to-beat Analysis of Heart Rate Dynamics During Exercise. American Journal of Physiology **271** (1996) H244-H252

Väinämö K., Röning J., Nissilä S., Mäkikallio T., Tulppo M.: Artificial Neural Networks for Aerobic Fitness Approximation. International Conference on Neural Networks (ICNN ‘96), June 3-6 (1996)