8th International Conference on DEVELOPMENT AND APPLICATION SYSTEMS
S u c e a v a, R o m a n i a, M a y 25 – 27, 2 0 0 6

PATTERN RECOGNITION METHODS IN PHYSICAL TRAINING EVALUATION AND PLANNING

Ştefan-Gheorghe PENTIUC[1], Elena RAŢĂ[2]

"Stefan cel Mare" University of Suceava
str.Universitatii nr.13, RO-720229 Suceava
[1]pentiuc@eed.usv.ro, [2]elenar@eed.usv.ro

Abstract. *The paper presents some pattern recognition methods used for the analysis of the data in the unsupervised learning context. The aim of the data processing is to assess the level of the physical condition of the sportsmen and their chances to gain future performance. These methods are based on several unsupervised algorithms, In the first part of the paper a model of an unsupervised learning system is underlined based on hierarchical classification. The definitions, axioms and theorem given in the paper provide a strong formalism that assures the reliability of this technique. The automatic decision making process based on the method of unsupervised learning in pattern space was experimented in the case of sportsmen from swimming team.*
Keywords: *pattern recognition, unsupervised learning, hierarchical classification, decision making support.*

Introduction

Any training system for characteristic to a certain cyclical sport relies on the theoretical, methodological and practical training knowledge, on scientific research and on the experience of the best national and international coaches.

Scientific research in the field of sportive training, show that the process of obtaining excellent sports performances is closely related/ in close connection to the most favourable/ adaptable type of management of the multianual training relies on new principial conceptions that ensue/result from systems theory, cybernetics, information theory and, at the same time, from physiological conception and biological activities.

One of the problems of coaches is to choose the best team to represent their clubs in various contests. This decision is made before an actual competition and is based on the previous individual performances.

The paper try to describe a system for decision making process, based on unsupervised classification which helps the trainer to select its best sportsmen for a specific contest. The pattern recognition methods used to build the system assumes that a pattern is an abstract representation of a person, object, or phenomena, not necessary an visual entity. As in [1] a pattern may be viewed as a vector

$$\boldsymbol{x} = (x_1, x_2, ..., x_p) \qquad (1)$$

where x_i are called features, and their values are issued from a set of observation and/or measures applied on the real object. Let $\boldsymbol{E}$ the patterns' set containing the patterns representing all the objects from the studied real object space ($\boldsymbol{O}$). Let $\boldsymbol{F}$ the pattern space ($\boldsymbol{E} \subset \boldsymbol{F}$). It is supposed that $\boldsymbol{F}$ is a metric space. The system analyst must make a good choice for the metric because the distances evaluated by this metric between the patterns (points in $\boldsymbol{F}$) are asked to reflect the relations from corresponding objects in $\boldsymbol{O}$. The two principals context of pattern recognition learning are [2]:

i) the supervised learning - a number of prototypes, with a classification indicated by the expert (or supervisor) are grouped into a training set $\mathbf{T} = \{(\mathbf{x},\sigma) \ / \ \mathbf{x} \in C_\sigma \subset \mathbf{E}\}$ that is presented to the system; based on the information contained in **T**, the system will compute the classification model, able to classify new patterns;

ii) the unsupervised learning context – the **E** set is presented to the system that must detect discriminant structures on **E,** revealing groups of objects in **O** having similar properties.

The problem of the coach is very frequently situated in the unsupervised context [3].

Pattern Classification

The similar patterns are grouped in subsets of ***E***. It is convenient to consider these subsets an equivalence classes structure over ***E***. Let $\boldsymbol{P_k(E)}$ a partition in k equivalence classes of ***E***. A homogenous partition is one that has all its classes homogenous. A class $\boldsymbol{C_j} \in \boldsymbol{P_k(E)}$ is homogenous if $\mathbf{x,y} \in \boldsymbol{C_j}$ and $z \notin \boldsymbol{Cj}$ then $d(\boldsymbol{x,y}) < d(\boldsymbol{x,z})$ and $d(\boldsymbol{x,y}) < d(\boldsymbol{y,z})$ where d is a metric over ***F***. Let a set of labels $\boldsymbol{L} = (\lambda_1, \lambda_2, \ldots, \lambda_m\}$ that may indicate the significance of pattern classes. An identification map ϕ will associate a label to every object in ***O***

$$\phi : \boldsymbol{O} \rightarrow \boldsymbol{L} \tag{2}$$

The function ϕ represents the associations between objects indicatedby the expert. Using ϕ, the objects in ***O*** are divided into many equivalent classes of objects having the same significance λ

$$\boldsymbol{C_\lambda} = \{ \alpha \,/\, \alpha \in \boldsymbol{O}, \phi(\alpha) = \lambda \} \tag{3}$$

The data processing deployed in an information system [4] was designed for decision making support may be represented by the following map:

$$\Pi : \boldsymbol{E} \rightarrow \boldsymbol{L} \tag{4}$$

that associates to each pattern a label. This association generates an equivalent classes structure over ***E***

$$\boldsymbol{C_\sigma} = \{\boldsymbol{x} / \boldsymbol{x} \in \boldsymbol{E}, \Pi(\boldsymbol{x}) = \sigma\} \tag{5}$$

A good decision making process resides in a classification scheme that assigns the right significance to each object [3]

$$\phi(o) = \Pi(\zeta(o)) \tag{6}$$

A good performance is obtained when the function ϕ representing the expert classification by the composition of ζ is very well approximated by function Π. (determined by the system from the training data set. Thus the objective of an unsupervised learning system is to produce a partition of a pattern set E presented at input reflecting the equivalence relation induced by when the function ϕ. In the situations where this number of equivalence classes is not given by the user, a partition string $P(\boldsymbol{E}) = \{P_1(\boldsymbol{E}), P_2(\boldsymbol{E}), \ldots, P_n(\boldsymbol{E})\}$ is a more convenient result.

In this approach it is necessary to have a metric defined over E, selected by the expert, in accordance with the semantic of the problem. The choice of the metric is crucial for the entire classification process [2], [5]. It is also possible to have at input a data representation of the relations established between patterns, expressed frequently by a metric [3].

Unsupervised Learning and Hierarchies

In a great number of particular cases the most preferred structure is represented by a sequence of homogenous partitions: $P_i(\boldsymbol{E}), P_{i+1}(\boldsymbol{E}), \ldots, P_j(\boldsymbol{E})$. A hierarchy of patterns is a generous source of information for system analyst [2].

Definition. Let $\boldsymbol{P}(\mathbf{E})$ the set of all partitions in equivalence classes of **E**. It will be called hierarchy of **E** the set of sets $\boldsymbol{H} \subset \boldsymbol{P}(\mathbf{E})$ if $\mathbf{E} \in \boldsymbol{H}$, and $\forall x \in \mathbf{E}$, $\{x\} \in \boldsymbol{H}$, and if $h \cap h' \neq \emptyset$, $h, h' \in \boldsymbol{H}$, then $h \subset h'$ or $h' \subset h$.

Definition. It is called an indexed hierarchy a couple $(\boldsymbol{H}, f())$, where $f : \boldsymbol{H} \rightarrow R_+$ and

i) $(\forall) h \in \boldsymbol{H}$ with $f(h) = 0$, then $card(h) = 1$

ii) $(\forall)$ $h, h' \in \boldsymbol{H}$ if $h \subset h'$, then $f(h) < f(h')$

The function f() is called index function.

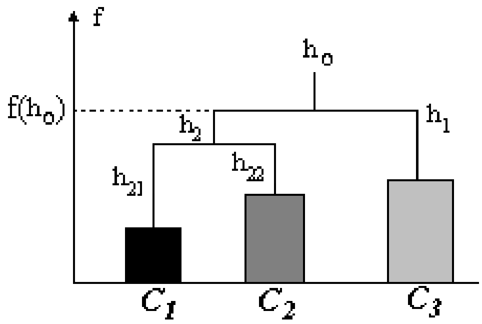

Figure 1. Schematic representation of an indexed hierarchy.

In the case of an indexed hierarchy a tree representing the inclusion relations may be generated. It is also possible to define a dissimilarity coefficient $\varphi(x,y) = \{f(h))/x,y \in h$ and $h \in \boldsymbol{H}\}$.

In figure 1 each element of the hierarchy was mapped with the values of a function f() in order to obtain an indexed hierarchy. It can be observed that the class C_1 is at an index lower than C_2 ($f(h_{21}) < f(h_{22})$). As consequence may be stated that C_1 is more homogenuous than C_2 and than C_3.

In general a metric distance d:**ExE** $\rightarrow R^+$ cannot always assure a partition with homogenous classes over a pattern set **E** [3]. From this reason an ultrametric distance is used. An ultrametric δ: **ExE** $\rightarrow R^+$, respects the first two axioms of the distances, $\delta(x,x)=0$, $\delta(x,y)= \delta(y,x)$ and additionally: $\delta(x,y) \leq \sup[\delta(x,z), \delta(z,y)]$.

Lemma 1. Let ($\boldsymbol{H}$,f) an indexed hierarchy on **E**, It can be defined a map δ : **ExE** $\rightarrow R^+$, with $\delta(x,y)=\min\{f(h) / h\in \boldsymbol{H}, x,y\in h\}, \forall(x,y)\in \mathbf{E}$. The map δ is an utrametrics.

From the practical point of view it is important to observe that a hierarchy generates an ultrametric distance over the pattern space [4].

Definition. Let δ an ultrametric over **E**. It can be defined the equivalence relation R_α over **E**: $(\forall)$ $x,y \in \mathbf{E}$, $x R_\alpha y \Leftrightarrow \delta(x,y) \leq \alpha$ with $\alpha \in R_+$.

Lemma 2. The set of all the equivalence classes generated by R_α for various values of α is an hierarchy indexed by α.

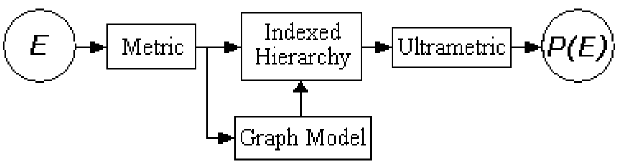

Figure 2. Data processing flow.

With these two lemmas the following theorem may be easily proven.

Theorem. Let ϕ a function that maps the indexed hierarchy onto the ultrametrics set, and let ψ the function mapping the ultrametrics onto the indexed hierarchy, then $\psi = \phi^{-1}$ and $\phi = \psi^{-1}$.

This theorem says that if an ultra metric is given then an indexed hierarchy may be obtained, and vice-versa. Based on these theoretical results a program system having a data processing flow as in figure 2 was developed. The program system accepts patterns from **E** and the metric indicated by the expert, and produces an ultrametric, various α equivalent relations and partitions' strings. Other significant research in this topic may be found in [6]. Different approaches for diagnosis using hierarchies pattern recognition are also presented in [3], or in [7] based on fuzzy methods, or using neural networks as in [8].

The Pattern Graph Model

The patterns from $\boldsymbol{E}$ belong to a metric space $\boldsymbol{F}$ with a metric d that is intended to better represent the relations established between the objects in $\boldsymbol{O}$. It can be defined as a complete graph [3] having as nodes all the patterns in $\boldsymbol{E}$

$$\boldsymbol{\Gamma} = (\boldsymbol{E}, \{(\xi\iota, \xi\varphi) / \xi\iota, , \xi\varphi \in E, \varphi > \iota\}) \qquad (7)$$

Each node is linked with all of the other and the cost of the edge (x_i, x_j) is the value of the metric d(xi, xj).

The minimum spanning tree (MST) of the graph $\boldsymbol{\Gamma}$ is the tree $\boldsymbol{\Sigma}=(\boldsymbol{E},\boldsymbol{T})$, $\boldsymbol{T} \subset \boldsymbol{E} \times \boldsymbol{E}$, where the sum of the metric values d(xi, xj). for each edge in T is minimum.

The MST is a more comprehensible representation than the graph G, having the property that if an edge is eliminated from T two trees are obtained. The nodes of each tree represent the patterns belonging to the same class.

If the distances between the nodes of the graph G are distinct, then the MST is unique. For the determination of the MST of a graph G various algorithms exists, among them the well knwon are Kruskal or Prim.

The MST offers a symbolic representation of a string partitions. The initial MST represents $\boldsymbol{P_1(E)}$, the partition with a unique class - the E set. If the edge with the great cost is eliminated the partition $\boldsymbol{P_2(E)}$ is obtained. Each class of the $\boldsymbol{P_2(E)}$ contains the patterns represented by the nodes of these two trees. And the process may continue until the partition $\boldsymbol{P_n(E)}$ is obtained, where n is the number of patterns in $\boldsymbol{E}$.

Application. Sportsmen Assesment

The evaluation of the specific physical training can be performed through the international physical endurance/capacity test known as "F.I.E.P". We have chosen the following tests:

- 50 m speed running;
- standing long jump;
- standing high jump;
- pushups

The *Speed Running Test* can be done using an up start or down start and moving timing. The distance is run in a straight line to the finish line. The time is registered or checked in seconds.

The *Standing Long Jump Test* consists in bending of the knees, swinging of the arms backwards which are went to provide forward drive. A two foot spring from the ground and the stretching of the arms forward represent the two foot take off. The fly consists in the first half, in lifting and stretching the feet ahead with the bended knees at the same time with stretching of the arms ahead and in the second half, in stretching the knees ahead. The landing is the stage when both feet reach the ground, simultaneously with the triggering of both arms backwards and the projection of the pelvis in order to prevent the body from falling backwards.

The *Standing High Jump Test* consists in reaching a certain point marked on the standardized strips of the meter/measuring apparatus. We calculate/ measure the difference between the highest point reached while jumping and the height with the arm stretched. It is counted in centimeters.

The *Pushups Test* is performed in the specific /characteristic position of this exercise. The execution pace is chosen by the subject. The result is evaluated in accordance with the number of pushups that are carried out correctly. There is no timing needed.

Because the number of values in the data series is too low, in order to obtain a good approximation the information system performs after the feature selection stage a classification of the sportsmen in classes containing the pattern representation of sportsmen with similar results of the four tests.

The schema of the pattern processing is presented in the figure 3.

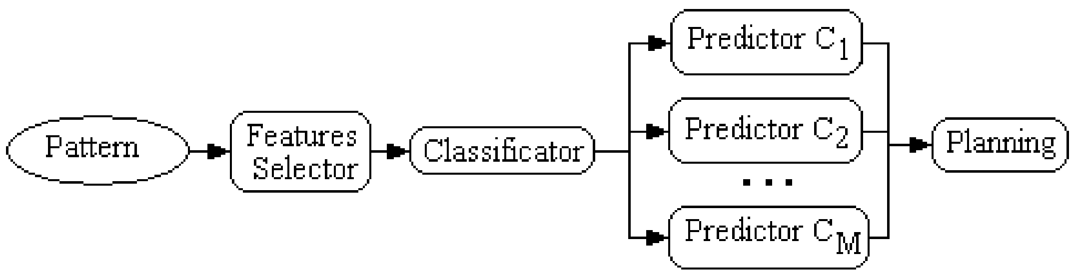

Figure 3. General Pattern Processing System for Planning the Physical Training.

The inputs of the system for planning the physical training of a swimming team are the values resulted from the physical tests and the performances obtained in competitions by each sportsmen. In the first stage the feature selector builds a pattern normalized and homogenous. The role of the *Classificator* is to establish the class where this sportsmen may be included.

After the classification is applied a prediction function with parameters specific to the class. The result are used by the *Planning* module that establish a plan of physical training for a short term.

The classes were defined after an unsupervised learning stage comprising a hierarchical classification. The resulted string of partitions was analysed as in [9] in order to determine the semantic of classes.

Once the appropriate partition was selected, the classification was compared by the classification issued from an algorithm acting on the graph model of pattern space.

Experiments

The tests that we have carried out aimed at proving the specific physical training at swimming sport teenagers aged between 14 and 16 years old. The evaluation was repeated monthly. For a sample of results see table 1.

Table 1. Results of the first assessment

Name	Age	Speed Running 50 m	Standing Long Jump	Standing High Jump	Pushups
A.A	14	8.0	1.82	30	9
A.R	15	7.9	1.85	29	11
C.N	14	8.2	1.79	28	10
B.R	14	8.4	1.77	28	8
SD	16	7.9	1.93	30	13
VE	15	8.1	1.81	29	12
RS	15	8.3	1.73	32	12
GB	16	8.0	1.84	35	12
SB	16	7.8	1.89	29	14
CV	15	8.1	1.87	24	12
SA	14	7.8	1.90	28	13
TA	14	8.4	1.91	29	9

For each sportsman / sportswoman it was obtained a data series that is was correlated with the performance in the competition organized in the same month. The correlation matrix between the values of these four tests is as below

1	0.74623	-0.19355	-0.55763
	1	0.116819	0.461761
		1	0.330048
			1

where the first row and column are dedicated to the speed test, the next, to the standing long jump test, followed by the standing high test and pushups (force).

It can be seen that only a single linear dependence may exists, between the speed test, and the standing long jump test, the others tests seem to be quasi independent.

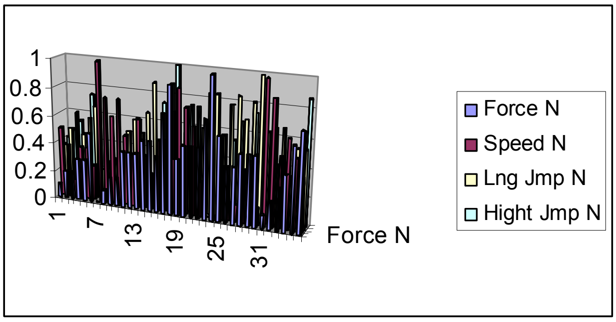

Figure 4. A graphic representation of the four tests values.

The variation of the all 4 parameters can be viewed in the figure 4.

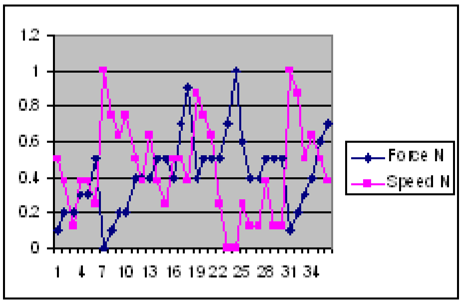

Figure 5. A comparison between the normalized values of the Speed (running on 50 m) and Force (pushups) tests.

If it is compared the results of the Force and Speed tests it can be observed the independence within these two pattern features (figure 5).

Conclusions

As is stated in [12] the physical training is everything for all training factors. On the basis of theoretical and methodical argumentation regarding the technical training of sportswomen on computer modelling grounds, it has been proposed [13-15] a new conception concerning the use of the computer in the modelling of sports training system. The system described allows the gathering and processing of huge amounts of data in a very short period of time.

Each sportsman is classified together with other colleagues with similar contest behaviour in a class based on a discriminant technique as in

[16] with a specific prediction function. As a consequence, we are thus able to choose from the sportsmen most important parameters with respect to somatic, functional, motive, psycho motive, psychological aspects no hide are characteristic to a certain contest.
As a future approach, this system based on Pattern Recognition methods will be integrated in a complex information system dedicated to the training of the sportsmen, that its workflow may be modelled by agents as in [17].

References

[1] Romul Vancea, Ştefan Holban, Dorin Ciubotar (1989) *Pattern Recognition. Applications (in Romanian)* - Editura Academiei, Bucuresti, 289 pag.

[2] Duda, R., Hart, P., Stork D.(1998) - *Pattern Classification*, the 2nd edition. John Wiley & Sons, Inc: California, USA

[3] Şt. Gh. Pentiuc (1997) - *Applications of Pattern Recognition in Automatic Diagnosis (in Romanian),* Editura Tehnică, Bucuresti, 155 pag.

[4] K.W. Lau and Q.H. Wu (2003) *Online training of support vector classifier*, Journal of Pattern Recognition Society, 36, pp. 1913-1920

[5] S. Asharaf, a and M. Narasimha Murty (2003) *An adaptive rough fuzzy single pass algorithm for clustering large data sets*, Pattern Recognition Volume 36, Issue 12 , (December), p. 3015-3018

[6] J.T.L. Wang (2001) - *Finding similar consensus between trees: an algorithm and a distance hierarchy* - The Journal of the Pattern Recognition Society. - Vol. 34, 1 (January); p. 127-139

[7] Alexandru Bârleanu - (2005) *Fault Diagnosis of Wall Hang Boliers Using a Neuro-Fuzzy Model,* Advances in Electrical and Computer Engineering, Suceava, Romania, ISSN 1582-7445 - No 1

[8] Sorin Vlad, Nicolae Morariu (2005) - *Using Neuroshell Neural Networks Simulator for Prediction Problems Solving* , Advances in Electrical and Computer Engineering, Suceava, Romania, ISSN 1582-7445 - No 1

[9] Aeberhard S., Coomans D., de Vel O. - *Comparative analysis of statistical pattern recognition methods in high dimensional settings* The Journal of the Pattern Recognition Society : Published by Pergamon Press - Vol. 27(1994) ,no.8; p.1065-1079.

[10] Hand D.J., Oliver J.J. and Lunn A.D. (1998) - *Discriminant analysis when the classes arise from a continuum* The Journal of the Pattern Recognition Society : Published by Pergamon Press - Vol. 31 ,no.5; p.641-650.

[11] Rigoutsos I. and Delis A. (1998), *Managing Statistical Behavior in Large Data Sets in Shared-Nothing Arhitectures* Parallel and Ditributed systems : IEEE COMPUTER SOCIETY - Vol. 9,no.11; p.1073-1087

[12] Bompa, T., O. (2002) *Theory and Methodology of the Training (in Romanian)*, Editura EX PONTO, Bucureşti

[13] Dragu, M.,V.(1999), *Dirijarea procesului de antrenament la înotătorii de 11-12 ani pe baza prognozării criteriilor de psihomotricitate*, Chişinău,

[14] Rață, G.(1999), *Aptitudinile motrice de bază, probleme teoretice*, Editura Plumb, Bacău

[15] Rîşneac, B., Milici, D., Rață, E. (,2004), *Utilizarea tehnicii de calcul în evaluarea performanțelor sportive*, Editura Universității Transilvania, Braşov

[16] Aristidis Likas, Nikos Vlassis and Jakob J. Verbeek (2003), *The global k-means clustering algorithm* J. of Pattern Recogn. Soc. Vol.36 pp.451-461

[17] Mihai Horia Zaharia, Florin Leon (2005), *A New Approach in Workflow Modeling Using Intelligent Agents*, Advances in Electrical and Computer Engineering, Suceava, Romania, ISSN 1582-7445 - No 2 pp. 57-60