Proceedings of the 9th International Conference on Neural Information Processing (ICONIP'02), Vol. 2
Lipo Wang, Jagath C. Rajapakse, Kunihiko Fukushima,
Soo-Young Lee, and Xin Yao (Editors)

STRUCTURE-ADAPTIVE SOM TO CLASSIFY 3-DIMENSIONAL POINT LIGHT ACTORS' GENDER

Sung-Bae Cho

Dept. of Computer Science, Yonsei University
134 Shinchon-dong, Sudaemoon-ku, Seoul 120-749, Korea
Fax: +82-2-365-2579
E-mail: sbcho@cs.yonsei.ac.kr

ABSTRACT

Classifying the patterns of moving point lights attached on actor's bodies with self-organizing map often fails to get successful results with its original unsupervised learning algorithm. This paper exploits a structure-adaptive self-organizing map (SASOM) which adaptively updates the weights, structure and size of the map, resulting in remarkable improvement of pattern classification performance. We have compared the results with those of conventional pattern classifiers and human subjects. SASOM turns out to be the best classifier producing 97.1% of recognition rate on the 312 test data from 26 subjects.

1. INTRODUCTION

Self-organizing map (SOM) proposed by Kohonen is well known for its topology preserving capability and clustering performance and has been applied in many fields of data mining, visualization and so on. However, in real-world problems, especially for the classification tasks, there exist difficulties how to determine the structure and size of the map.

Several researchers have dealt with the problem of structure adaptation of SOM [1, 2]. Some proposed tree-structured neural networks, but this approach was not able to represent the global order of nodes consistently because the learning takes place only in the sub trees. Fritzke proposed a "growing cell" which inserted another columns or rows of nodes on the map to adapt the structure, but this also might lead to too many extra nodes. In the previous work [3], we proposed the node-splitting scheme which let SOM adjust its structure adaptively by splitting only nodes where different class labels were mixed, and presented the usefulness by showing the experimental results of the unconstrained handwritten digit recognition.

In this paper, we make use of the structure-adaptive self-organizing map (SASOM) to classify the gender of arm movement data and compare the classification performance of SASOM with other conventional machine learning classifiers and humans themselves.

2. BACKGROUNDS

2.1 Discriminating Gender from Moving Point Lights

It is interesting for the psychologists that people can identify a friend by his walk at such a distance as face, hair style and clothes are obscure. After Wolff's work for how people recognize friends by their walk, in order to overcome the confounding role of familiarity cues such as size and shape of objects, filmed display of moving point lights called point light display (or moving light display) was introduced, which present the displacements of locomotion such as walking, running or jumping. This encouraged studies on human motion and motion perception.

In Johanssen's experiment [4], he used glass-bead retro reflective tape attached on the main joints of human body at 10 different sites. Filmed displays of the point light walkers that only the lights reflected from the tape can be seen as illuminated dots in a dark background were used to test people's perception of biological motion. People could not recognize the walkers from the static set of point lights, but once they were moving, people could do it somehow. Cutting and Kozlowski used moving point lights to discriminate the gender of the walkers, mounting point light sources on people's major joints, they could obtain 60~70% of average accuracy [5]. They also found that lights on the upper body's joints were more useful to recognize the movement.

2.2 Pattern Classifiers

Many promising machine learning techniques and algorithms have been successfully adopted in pattern classification and recognition problems. Error backpropagation neural network is a feed-forward multilayer perceptron (MLP), which learns the training examples by adjusting the synaptic weight of neurons by the errors occurred on the output layer (delta learning).

The power of the back-propagation algorithm lies in two main aspects: locality to update the synaptic weights and biases and efficiency to compute all the partial derivatives of the cost function with respect to these free parameters.

Self-organizing map (SOM) defines an I/O mapping through unsupervised learning process. SOM has an output layer consisting of N nodes, each of which represents a vector that has the same dimension as the input pattern. For a given input vector x, the winner node m_c is chosen using Euclidean distance between x and the prototypes, m_i, and the weights of the neighbor nodes surrounding the m_c are updated as follows.

$$\|x - m_c\| = \min_i \|x - m_i\| \tag{1}$$

$$m_i(t+1) = m_i(t) + \alpha(t) \times n_{ci}(t) \times \{x(t) - m_i(t)\} \tag{2}$$

where $\alpha(t)$ is the learning rate and $n_{ci}(t)$ is a neighborhood function.

k-nearest neighbor (KNN) is one of the most common methods among memory based induction system. Given an input vector, KNN chooses k closest vectors in the reference set based on similarity measures, and decides the corresponding label of input vector using the distribution of labels that k neighbors have and its similarity.

Support vector machine (SVM) optimizes a decision boundary hyperplane in such a way to maximize the margin of separation between positive and negative exemplars. SVM achieves this using the structural risk minimization principle that the error rate on the test data is bounded by the sum of the training-error rate and a term that depends on the Vapnik-Chervonenkis (VC) dimension. Given a labeled set of M training samples (X_i, Y_i), where $X_i \in R^N$ is a vector and Y_i is the associated label, $Y_i \in \{-1, 1\}$, the discriminant hyperplane is defined by:

$$f(X) = \sum_{i=1}^{M} Y_i \alpha_i k(X, X_i) + b \tag{3}$$

where $k(\ .)$ is a kernel function and the sign of $f(X)$ determines the membership of X. Constructing an optimal hyperplane is equivalent to finding all the nonzero α_i (support vectors) and a bias b.

Quinlan's C4.5 builds the decision tree according to an information-theoretical approach based on the energy entropy as follows: select an attribute, divide the training set into subsets characterized by the possible values of the attribute, and follow the same partitioning procedure recursively with each subset until no subset contains objects from more than one class. The single class subsets correspond them to the leaves. The entropy-based criterion that has been used for the selection of the attribute is called the gain ratio criterion. Let X be a possible test (attribute selection) that partitions the training set T into n subsets (T_1, T_2, ..., T_n), split_info(X), as the entropy of a message where information is given in terms of outcomes, and gain_ratio(X) can be defined as follows:

$$\mathrm{split\_info}(X,T) = -\sum \left(\frac{|T_i|}{|T|}\right) \log_2 \left(\frac{|T_i|}{|T|}\right) \tag{4}$$

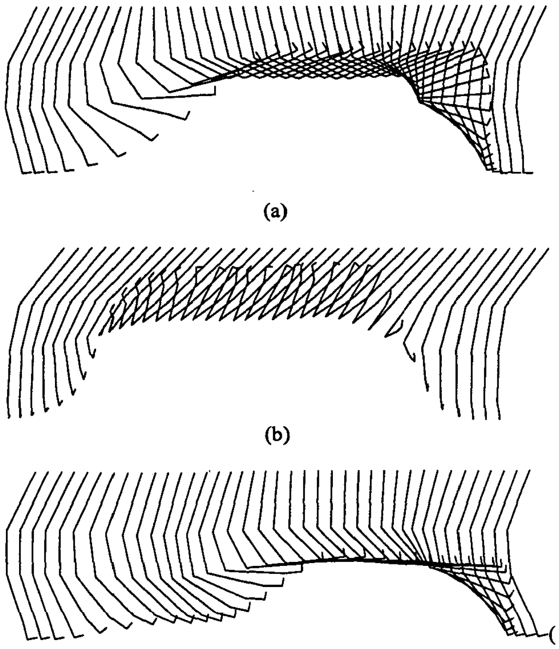
$$\mathrm{gain\_ratio}(X) = \frac{\mathrm{gain}(X)}{\mathrm{split\_info}(X)} \tag{5}$$

The gain ratio criterion selects the test X so that the gain_ratio(X) is maximized.

3. HUMAN MOVEMENT DATA

3.1 Data Acquisition

The movement data were obtained using a 3D position analysis system (Optotrak, Northern Digital). Positions of the temple, right shoulder, elbow, wrist and the first and fourth metacarpal joints were recorded at a rate of 60 Hz while an actor performed the movement. Actors were instructed to perform knocking, waving, and lifting movements in neutral and angry styles, resulting in stylistic behaviors, such as 'neutral knocking,' 'neutral waving,' 'neutral lifting,' 'angry knocking,' 'angry waving' and 'angry lifting,' as shown in Figure 1.



(a)

(b)

c)

Figure 1. : Examples of arm movements. (a) knocking, (b) waving and (c) lifting movements are shown. These are performed under specified style. We presented the side (knocking and lifting) or frontal (waving) view for the convenience, but 3D representation has been used in the experiments.

These movements have been chosen because they are all upper body's movements, easy to be played, have short time duration and take place often around us. 26 people participated to build the data set, and half of them were males and the others were females. Each movement was recorded 10 times repeatedly, and 8 of them have been used as the training data and the rest as the test data.

3.2 Preprocessing

Each movement has been processed to obtain the start and end points. The start of movement is defined as the moment the tangential velocity of the wrist rise above 5% of the peak and the end by the moment when passed below 5% of the peak. The missing data resulted from a marker going out of view of the cameras have been replaced with the right one manually.

It is important to get invariant information source from the data for any learning machine. We do not use the point light display but use 3-dimensional point light actors because there are infinite numbers of 2D directions, so that 2D features are not invariant by viewing directions. As the features, position and velocity are used. Since velocities are the derivative of positions, they are independent sources of information in the feature space. Velocity can be calculated by taking distances between the frames of motions.

All the patterns are normalized to 150. The input dimension is reduced to 150, and the amplitude is adjusted between 0 and 1.

4. STRUCTURE ADAPTIVE SOM

In this section, we present the dynamic node-splitting scheme for the structure-adaptive self-organizing map (SASOM). Figure 2 shows the flow chart of how the algorithm works.

SASOM starts with a 4×4 rectangular map using Kohonen's learning algorithm, as shown in the equations 1 and 2. We have used the hit ratio which is the threshold determining whether a node is to be split or not. Suppose n_i is the ith node on the output layer of the map and c_j is the exemplar which belongs to the jth class. The hit ratio of the n_i is defined as follows:

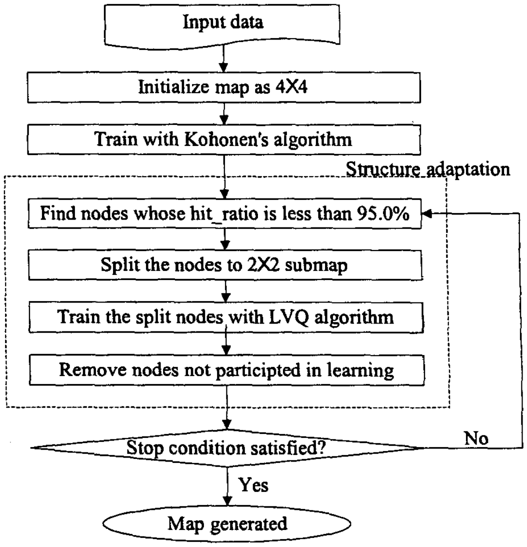

Input data
Initialize map as 4X4
Train with Kohonen's algorithm
Structure adaptation
Find nodes whose hit_ratio is less than 95.0%
Split the nodes to 2X2 submap
Train the split nodes with LVQ algorithm
Remove nodes not participted in learning
Stop condition satisfied?
No
Yes
Map generated

Figure 2. SASOM algorithm.

$$\text{hit\_ratio}_i = \max_j P(c_j \mid n_i) \tag{6}$$

where $i = 1, 2, \cdots, M$ and $j = 1, 2, \cdots, N$

$P(c_j \mid n_i)$ indicates the proportion of exemplars labeled as class c_j within node n_i. In this paper, the nodes less than 95.0% of hit ratio are split. If the hit_ratio$_i$ is too large, we need to adjust the hit_ratio$_i$ so that the map be trained optimally because SASOM may be overfitted. Selected nodes are split and replaced with 2×2 maps, and the weights of child nodes are initialized based on the weights of the parent and neighborhood nodes. When the parent node P is split to the child node C, the initial weight of C is calculated as follows:

$$C = \frac{(P \times 2) + \sum N_c}{S} \tag{7}$$

where N_c is the weights of neighbors and S is the total number of nodes that participate in weight initialization of C. The radius of N_c can vary on time, but it is advantageous to get the radius as large in the beginning and shrink it monotonically with time for a good global ordering. In the case of rectangular topology, if the radius of N_c is 1 S can be up to 5: one parent and 4 neighbors (up, down, left and right).

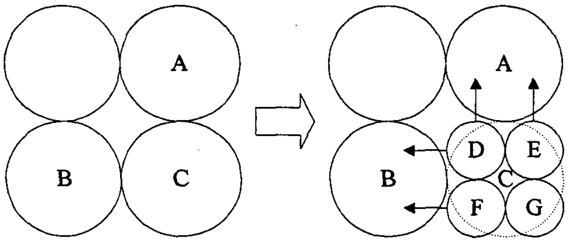


Figure 3. Weight initialization of child nodes. Node C is split into node D, E, F and G.

Figure 3 shows an example of weight initialization of child nodes. Node C is determined to be split. Child node D has neighbors A and B, E has one neighbor A, F has B, and G has no neighbors. Finally, the weights of child nodes become D = (A+B+2C)/4, E = (A+2C)/3, F = (B+2C)/3 and G = C. Split nodes are further trained as follows, until every node gets above 95.0% of hit ratio or the maximum number of iteration is reached:

$$m_i(t+1) = m_i(t) + \alpha(t) \times N_c(t) \times h_c(t) \times \{x(t) - m_i(t)\} \quad (8)$$

$$h_c(t) = \begin{cases} 1, & \text{if } x(t) \text{ and } m_i(t) \text{ belong to the same class} \\ 0, & \text{if } x(t) \text{ and } m_i(t) \text{ belong to different classes} \end{cases} \quad (9)$$

where $x(t)$ is input vector, $m_i(t)$ is a split child node in equation 7 and $\alpha(t)$ is a learning rate, $0<\alpha(t)<1$. The original LVQ algorithm uses $h_c(t) = -1$ where $x(t)$ and $m_i(t)$ belong to different classes, but there exist some cases where learning leads to a non-optimal solution. In our case, no training takes place in such situations. Also, neighborhood function $N_c(t)$ is used to preserve the topological order, which LVQ does not have. Finally, the nodes without input exemplars assigned are removed from the map.

5. EXPERIMENTS

SOM, MLP, single layer perceptron (SLP), decision tree, k-nearest neighbor and support vector machine have been used as classifiers for the performance comparison. All the classifiers are written in Unix C language. All parameters have been chosen as shown in Table 1. For KNN, we have tried k = 10~25, but k = 20 yields the best result. Both MLP and SLP have been trained until the recognition rate on the training data reaches at 98.0%. For SVM, SVM$^{\text{light}}$ module is used, which imports quadratic programming techniques. Decision tree uses 0.25 of pruning confidence level, and a rule is pruned if it contains a condition whose probability of being irrelevant is greater than the confidence level. In order to label nodes on SOM, voting method is adopted (nodes are labeled with the majority class). Male is encoded as 1 or [1 0] and female as 0 or [0 1].

Table 1: Parameters of classifiers.

Classifier	Parameters
SASOM	Rectangular topology Bubble neighborhood Maximum iteration: 1,000,000 Learning rate: 0.02
KNN	k = 20, Euclidean distance similarity function
MLP	Input nodes: 900 (velocity), 2700 (position) Hidden nodes: 30 (velocity), 50 (position) Output nodes: 1 (velocity), 2 (position) Learning rate: 0.01 Momentum: 0.7(velocity), 0.007 (position)
SLP	Input nodes: 900 (velocity), 2700 (position) Output nodes: 2
SVM	Kernel: linear Maximum size of QP-subproblems: 10 Maximum iteration: 100 XiAlpha-estimate rho: 1.00, depth: 0
DT	Gain ratio criterion used Confidence level for pruning: 0.25
SOM	Map size: 10×10 (empirical optimal size) Rectangular topology Learning rate: 0.02 Maximum iteration: 1,000,000

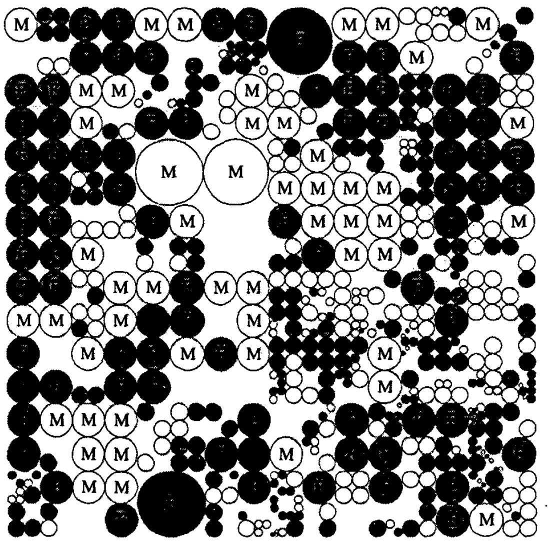


Figure 4. SASOM of the final state during training. Male is represented as "M" or white circle and female as "F" or black circle.

20 people have participated to test the human efficiency to discriminate gender from moving light actors. 1,248 patterns are used for training and 312 for test. The results are obtained from the average of the experiments in 5 times.

Figure 4 shows the map configuration of SASOM in the final state. As previously described, SASOM splits nodes continuously until the final condition of hit ratio is satisfied, so that the size and structure of map have been automatically adjusted. The blank space among the black and white nodes is the place where no exemplars have been assigned and no learning has took place.

The final result of recognition rates for the classifiers on the test data is shown in Table 2. SASOM is better than other conventional classifiers used in this work. This indicates that applying dynamic node-splitting scheme has given better efficiency of classification performance than conventional classifiers including SOM. There are 4 not-answered cases with SOM, which is the weakness of voting scheme, while there is none with SASOM. Classifiers with the position feature produce better results than those with velocity in gender classification, which implies that position is more informative feature than velocity. KNN and MLP also show relatively high recognition accuracy.

Human subjects perform very poor on the same test data, obtaining just above 50% of recognition rate. Human may not have certain principle or criterion to discriminate the gender from arm movements we used. It might be because we have excluded such familiar clues with human as size or shape of objects.

Table 2: Recognition rates of classifiers [%].

Classifier	Feature		Average
	Velocity	Position	
SASOM	85.9	97.1	91.5
KNN	81.4	92.9	87.2
MLP	81.4	84.6	83.0
SLP	75.7	76.1	75.9
SVM	71.8	73.2	72.6
DT	67.0	72.8	69.9
SOM	60.6	76.9	68.8
Human	51.3		51.3

Figure 5 shows the recognition rates of SASOM with respect to the six motions. The classification with 'neutral lifting' motion is the easiest task. For the neutral style, "lifting > waving > knocking," while "knocking ≈ waving > lifting" for the angry style.

Consequently, SASOM yields up to the 97.1% of recognition rate, and machine classifiers performed much better than human.

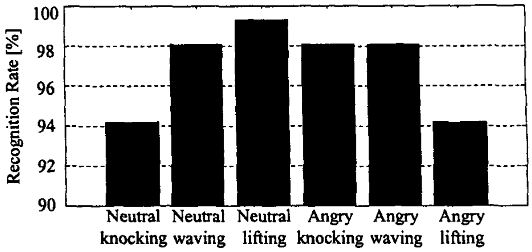


Figure 5. Recognition rates by conditions.

6. CONCLUDING REMARKS

In this paper, to classify the human's gender from biological motion, we have used SASOM and showed the superiority of classification efficiency. Measuring recognition rate, SASOM is always the best. Position is the most informative data representation for most of classifiers. The results have also shown that the machine classifiers are much better than human to recognize the actor's gender.

Acknowledgement

This paper was supported by Brain Science and Engineering Research Program sponsored by Korean Ministry of Science and Technology.

REFERENCES

[1] T.D. Sanger, "A tree-structured adaptive network for function approximation in high-dimensional spaces," *IEEE Trans. on Neural Networks*, vol. 2, no. 2, pp. 285-293, 1991.

[2] B. Fritzke, "Growing cell structures: A self-organizing network for unsupervised and supervised learning," *Neural Networks*, vol. 7, no. 9, pp. 1441-1460, 1994.

[3] S.-B. Cho, "Self-organizing map with dynamical node splitting: Application to handwritten digit recognition," *Neural Computation*, vol. 9, no. 6, pp. 1343-1353, 1997.

[4] G. Johanssen, "Visual perception of biological motion and a model for its analysis," *Perception & Psychophysics*, vol. 14, no. 2, pp. 201-211, 1973.

[5] J.E. Cutting and L.T. Kozlowski, "Recognising friends by their walk: Gait perception without familiarity cues," *Bulletin of the Psychonomic Society*, vol. 9, pp. 353-356, 1977.