Theory and Applications of Mathematics & Computer Science

(ISSN 2067-2764)
http://www.uav.ro/applications/se/journal/index.php/tamcs

Theory and Applications of Mathematics & Computer Science 1 (2011) 11–16

Classification Accuracy of Neural Networks with PCA in Emotion Recognition

Jasmina Novakovic[a,*], Milomir Minic[a], Alempije Veljovic[b]

[a] *Faculty of Public Administration, Megatrend University, Belgrade, Serbia.*
[b] *Technical Faculty Cacak, University in Kragujevac, Cacak, Serbia.*

Abstract

This paper presents classification accuracy of neural network with principal component analysis (PCA) for feature selections in emotion recognition using facial expressions. Dimensionality reduction of a feature set is a common preprocessing step used for pattern recognition and classification applications. PCA is one of the popular methods used, and can be shown to be optimal using different optimality criteria. Experiment results, in which we achieved a recognition rate of approximately 85% when testing six emotions on benchmark image data set, show that neural networks with PCA is effective in emotion recognition using facial expressions.

Keywords: emotion recognition, feature selection, neural network, PCA.
2000 MSC: 68T45, 97P20, 97R40, 68T45.

1. Introduction

Feature selection is an active field in computer science. It has been a fertile field of research and development since 1970's in statistical pattern recognition (Wyse *et al.*, 1980), (Ben-Bassat, 1982), (Siedlecki & Sklansky, 1988), machine learning and data mining (Blum & Langley, 1997), (Dash & Liu, 1997), (Dy & Brodley, 2000), (Kim *et al.*, 2000), (Das, 2001), (Mitra *et al.*, 2002). Feature selection is a fundamental problem in many different areas, especially in forecasting, document classification, bioinformatics, and object recognition or in modeling of complex technological processes. Data sets with thousands of features are common in such applications. For some problems, all features may be important, but for some target concept, only a small subset of features is usually relevant.

Feature selection reduces the dimensionality of feature space, removes redundant, irrelevant, or noisy data. It brings the immediate effects for application: speeding up a data mining algorithm, improving the data quality and thereof the performance of data mining, and increasing the comprehensibility of the mining results.

Feature selection can be defined as a process that chooses a minimum subset of M features from the original set of N features, so that the feature space is optimally reduced according to a certain evaluation criterion. As the dimensionality of a domain expands, the number of feature N increases. Finding the best feature subset is usually intractable (Kohavi & John, 1997) and many problem related to feature selection have been shown to be NP-hard (Blum & Rivest, 1992).

In pattern recognition and general classification problems, methods such as PCA, independent component analysis (ICA) and Fisher linear discriminate analysis have been extensively used. These methods find a mapping between the original feature space to a lower dimensional feature space.

[*] Corresponding author
Email addresses: `jnovakovic@megatrend.edu.rs` (Jasmina Novakovic), `mminic@megatrend.edu.rs` (Milomir Minic), `alempije@beotel.rs` (Alempije Veljovic)

The main aim of this paper is to experimentally verify, on benchmark image data set, the classification accuracy of neural network with PCA in emotion recognition. This paper is organized as follows. Section 2 contains the general approach to automatic facial expression analysis. Neural network as classifier in emotion recognition is presented in section 3. Section 4 contains general issues concerning PCA as feature selection and reduction method. Section 5 presents experimental evaluation. Final section contains discussion of the obtained results and some closing remarks.

2. Facial Expression Analysis

Facial expression analysis includes both measurement of facial motion and recognition of expression. The general approach to automatic facial expression analysis consists of three steps: face acquisition, facial data extraction and representation, and facial expression recognition.

The first step, face acquisition is a processing stage to automatically find the face region for the input images or sequences. It can be a detector to detect face for each frame or just detect face in the first frame and then track the face in the remainder of the video sequence. To handle large head motion, the head finder, head tracking, and pose estimation can be applied to a facial expression analysis system.

The next step is to extract and represent the facial changes caused by facial expressions. There are mainly two types of approaches in facial feature extraction for expression analysis: geometric feature-based methods and appearance-based methods. The geometric facial features present the shape and locations of facial components (including mouth, eyes, brows, and nose). The facial components or facial feature points are extracted to form a feature vector that represents the face geometry. With appearance-based methods, image filters, such as Gabor wavelets, are applied to either the whole-face or specific regions in a face image to extract a feature vector. Depending on the different facial feature extraction methods, the effects of in-plane head rotation and different scales of the faces can be eliminated by face normalization before the feature extraction or by feature representation before the step of expression recognition.

In the last step, the facial changes can be identified as facial action units or prototypic emotional expressions. The detection and processing of facial expression is achieved through various methods such as optical flow, hidden Markov model, neural network processing or active appearance model. More than one modalities can be combined or fused (multimodal recognition, e.g. facial expressions and speech prosody (Caridakis *et al.*, 2006) or facial expressions and hand gestures (Balomenos *et al.*, 2005)) to provide a more robust estimation of the subject's emotional state.

Table 1. Properties of an ideal facial expression analysis system (Tian *et al.*, 2001).

Robustness
Deal with subjects of different age, gender, ethnicity
Handle lighting changes
Handle large head motion
Handle occlusion
Handle different image resolution
Recognize all possible expressions
Recognize expressions with different intensity
Recognize asymmetrical expressions
Recognize spontaneous expressions
Automatic process
Automatic face acquisition
Automatic facial feature extraction
Automatic expression recognition
Real-time process
Real-time face acquisition
Real-time facial feature extraction
Real-time expression recognition
Autonomic Process
Output recognition with confidence
Adaptive to different level outputs based on input images

The properties of an ideal facial expression analysis system are summarized in Table 1.

3. Neural Networks

Neural networks have been studied for more than four decades since Rosenblatt first applied the *single-layer perceptrons* to pattern-classification learning in the late 1950's. A network structure consisting of a number of nodes connected through directional links is a neural network. In this network each node represents a processing unit, and the links between nodes specify the causal relationship between connected nodes. The outputs of these nodes depend on modifiable parameters pertaining to these nodes.

A neural network is a massive parallel distributed processor made up of simple processing units. This network has the ability to learn from experiential knowledge expressed through interunit connection strengths, and can make such knowledge available for use.

A neural network derives its computing power through its massive parallel distributed structure and its ability to learn and therefore to generalize. Generalization refers to the neural network producing reasonable outputs for new inputs not encountered during a learning process.

Fundamental to the operation of a neural network is a neuron as an information-processing unit. Figure 1 shows that a neuron consists of three basic elements: a set of connecting links, an adder and an activation function.

The first basic elements, a set of connecting links from different inputs x_i (or synapses), is characterized by a weight or strength w_{ki}. The weights of a neuron may lie in a range that includes negative as well as positive values. The first index refers to the neuron in question and the second index refers to the input of the synapse to which the weight refers. The second basic elements, an adder, summing the input signals x_i weighted by the respective synaptic strengths w_{ki}. An activation function f, as the third basic elements in neuron, limiting the amplitude of the output y_k of a neuron.

An externally applied bias, denoted by b_k, is also includes in the model of the neuron shown in Figure 1. The effect of bias is increasing or lowering the net input of the activation function, depending on whether it is positive or negative.

The architecture of a neural network is defined by the characteristics of a node and the characteristics of the node's connectivity in the network. Network architecture is specified by the number of inputs to the network, the number of outputs, the total number of elementary nodes that are usually equal processing elements for the entire network, and their organization and interconnections. Generally, neural networks are classified, on the basis of the type of interconnections, into two categories: feedforward and recurrent.

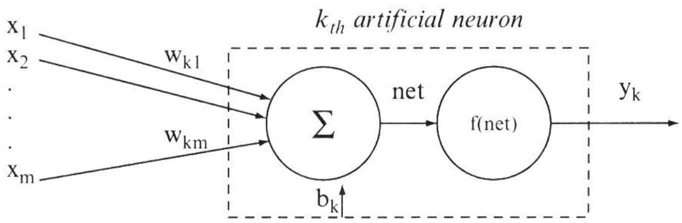

Figure 1. Model of a neuron.

If the processing propagates from the input side to the output side unanimously, without any loops or feedbacks, the network is feedforward. In a layered representation of the feedforward neural network, there are no links between nodes in the same layer; outputs of nodes in a specific layer are always connected as inputs to nodes in succeeding layers. This representation is preferred because of its modularity, i.e., nodes in the same layer have the same functionality or generate the same level of abstraction about input vectors. The network is recurrent if there is a feedback link that forms a circular path in a network (usually with a delay element as a synchronization component). In Figure 2 are shown examples of neural network belonging to both classes.

In both classes many neural-network models have been proposed, but the multilayer feedforward network with a backpropagation-learning mechanism, is the most widely used model in terms of practical applications.

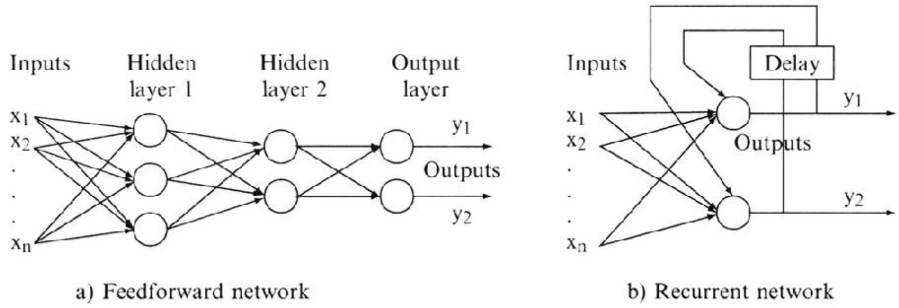

Figure 2. Typical architectures of neural networks.

4. PCA for Dimensionality Reduction

PCA, also known as Karhunen-Loeve transform, is one of the most popular techniques for dimensionality reduction. It is a standard statistical technique that can be used to reduce the dimensionality of a data set. PCA is useful tool for dimensionality reduction of multivariate data in image analysis, pattern recognition and appearance-based visual recognition, data compression, time series prediction, and analysis of biological data.

The strength of PCA for data analysis comes from its efficient computational mechanism, the fact that it is well understood, and from its general applicability. For example, a sample of applications in computer vision includes the representation and recognition of faces, recognition of 3D objects under varying pose, tracking of deformable objects, and for representations of 3D range data of heads.

PCA is a method of transforming the initial data set represented by vector samples into a new set of vector samples with derived dimensions. The basic idea can be described as follows: a set of n-dimensional vector samples $X = \{x_1, x_2, x_3, \ldots, x_m\}$ should be transformed into another set $Y = \{y_1, y_2, \ldots, y_m\}$ of the same dimensionality, but y-s has the property that most of their information content is stored in the first few dimensions. So, we can reduce the data set to a smaller number of dimensions with low information loss.

In implementation, the transformation from the original attributes to principal components is carried out through a process by first computing the covariance matrix of the original attributes and then, by extracting its eigenvectors to act as the principal components. The eigenvectors specify a linear mapping from the original attribute space of dimensionality N to a new space of size M in which attributes are uncorrelated. The resulting eigenvectors can be ranked according to the amount of variation in the original data that they account for. Typically, the first few transformed attributes account for most of the variation in the data set and are retained, while the remainder are discarded.

PCA is an unsupervised method which makes no use of information embodied within the class variable. Because, the PCA returns linear combinations of the original features, the meaning of the original features is not preserved.

However, PCA models have several shortcomings. One is that naive methods for finding the principal component directions have trouble with high dimensional data or large numbers of data points. Difficulties can arise in the form of computational complexity and also data scarcity. Another shortcoming of standard approaches to PCA is that it is not obvious how to deal properly with incomplete data set, in which some of the points are missing. To solve these drawbacks of standard PCA, a lot of methods were proposed in the field of statistics, computer engineering, neural networks etc.

Over the years there have been many extensions to conventional PCA. For example, Independent Component Analysis (ICA) is the attempt to extend PCA to go beyond decorrelation and to perform a dimension reduction onto a feature space with statistically independent variables. Other extensions address the situation where the sample data live in a low-dimensional (non-linear) manifold in an effort to retain a greater proportion of the variance using fewer components and yet other (related) extensions derive PCA from the perspective of density estimation (which facilitate modeling non-linearity in the sample data) and the use of Bayesian formulation for modeling the complexity of the sample data manifold.

5. Experimental Results

JAFFE database of facial expression images was used (available at http://www.kasrl.org/jaffe.html). Ten expressors posed 3 or 4 examples of each of the six basic facial expressions (happiness, sadness, surprise, anger, disgust, fear) and a neutral face for a total of 219 images of facial expressions. For simplicity of experimental design only Japanese female expressors were used. Sample images are shown in Figure 3.

Many classifiers have been applied to expression recognition such as neural network, support vector machines, linear discriminant analysis, k-nearest neighbor, multinomial logistic ridge regression, hidden Markov models, tree augmented naive Bayes, and others.

Some systems use only a rule-based classification based on the definition of the facial actions. Also, there are the expression recognition methods to frame-based and sequence-based expression recognition methods. The frame-based recognition method uses only the current frame with or without a reference image (mainly it is a neutral face image) to recognize the expressions of the frame. The sequence-based recognition method uses the temporal information of the sequences to recognize the expressions for one or more frames.

Figure 3. Sample images of JAFFE database.

For emotion recognition we used neural networks and PCA-based dimensionality reduction. Algorithm for facial expression recognition classify the given image into one of the seven basic facial expression categories (happiness, sadness, fear, surprise, anger, disgust and neutral). PCA is used for dimensionality reduction in input data. It retaining those characteristics of the data set that contribute most to its variance, by keeping lower-order principal components and ignoring higher-order ones. Because, low-order components contain the "most important" aspects of the data. The extracted feature vectors in the reduced space are used to train the supervised neural network classifier. This approach does not require the detection of any reference point or node grid. The proposed method is fast and can be used for real-time applications.

Table 2. Confusion matrix for classification of facial expressions.
HAP - Happiness, SAD - Sadness, SUR - Surprise, ANG - Anger, DIS - Disgust, FEA - Fear.

ANG	DIS	FEA	HAP	SAD	SUR	I/O
9	1	0	0	0	0	ANG
0	11	1	0	1	0	DIS
0	0	9	1	0	0	FEA
0	0	0	9	2	0	HAP
0	0	1	2	8	0	SAD
0	0	1	0	0	10	SUR
0	0	0	0	0	0	NEU

The expression classifier was first tested using a set of images of expressions posed by ten Japanese females expressor initials: KA, KL, KM, KR, MK, NA, NM, TM, UY and YM. Each expresser posed three or four examples of each of the six fundamental facial expressions and a neutral face. The image set was partitioned into ten segments, each corresponding to one expressor. Two facial expression images of each expression of each subject were randomly selected as training samples, while the remaining samples were used as test data. Not all expressions were equally well recognized by the system. Table 2 shows a confusion matrix showing misclassification rates for expressors. Our simulation experiment results show that neural networks is effective in emotion recognition using facial expressions, and we achieved a recognition rate of approximately 85% when testing six emotions.

It is not so convenient to compare categorization performance, because the problem that posed expressions is not always pure examples of a single expression category. It is important to realize that expression is never pure expressions of one emotion, but always admixtures of different emotions. The expression labels on the images in JAFFE database just represent the predominant expression in that image - the expression that the subject was asked to pose.

6. Conclusions and Ongoing Research

Experiment results with recognition rate of approximately 85% when testing six emotions on benchmark image data set, show that neural networks with PCA is effective in emotion recognition using facial expressions.

This research could help in future works, like capturing non-static images in real time and simultaneously analyzing these images according to affective computing techniques. By making these analyses some of the user's emotional states could be seen like joy, fear, angry, and with these probable results, assistants and computer optimizers could help users in the most different applications.

There are many questions and issues that remain to be addressed and that we intend to investigate in future work. Some improvements of the selecting methods presented here are possible. The algorithms and data sets will be selected according to precise criteria: classify algorithms and several data sets. These conclusions and recommendations will be tested on larger data sets using various classification algorithms in the near future.

References

Balomenos, T., A. Raouzaiou, S. Ioannou, A. Drosopoulos, K. Karpouzis and S. Kollias (2005). Emotion analysis in man-machine interaction systems. In: *Machine Learning for Multimodal Interaction* (Samy Bengio and Hervé Bourlard, Eds.). Vol. 3361 of *Lecture Notes in Computer Science*. pp. 318–328. Springer Berlin, Heidelberg.

Ben-Bassat, M. (1982). Pattern recognition and reduction of dimensionality. In: *Handbook of statistics-II* (P. R. Krishnaiah and L. N. Kanal, Eds.). pp. 773–791. North Holland.

Blum, A. I. and P. Langley (1997). Selection of relevant features and examples in machine learning. *Artificial Intelligence* **97**(1-2), 245–271.

Blum, A. L. and R. L. Rivest (1992). Training a 3-node neural network is np-complete. *Neural Networks* (5), 117–127.

Caridakis, G., L. Malatesta, L. Kessous, N. Amir, A. Raouzaiou and K. Karpouzis (2006). Modeling naturalistic affective states via facial and vocal expressions recognition. In: *Proceedings of International Conference on Multimodal Interfaces (ICMI06), Banff, Alberta, Canada, November 2-4, 2006*. pp. 146–154.

Das, S. (2001). Filters, wrappers and a boosting-based hybrid for feature selection. In: *Proceedings of the Eighteenth International Conference on Machine Learning (ICML 2001), Williams College, Williamstown, MA, USA, June 28 - July 1, 2001*. Morgan Kaufmann. pp. 74–81.

Dash, M. and H. Liu (1997). Feature selection for classification. *Intelligent Data Analysis* **1**, 131–156.

Dy, J. G. and C. E. Brodley (2000). Feature subset selection and order identification for unsupervised learning. In: *Proceedings of the Seventeenth International Conference on Machine Learning, June 29-July 2, 2000, Stanford University, CA*. Morgan Kaufmann. pp. 247–254.

Kim, Yongseog, W. Nick Street and Filippo Menczer (2000). Feature selection in unsupervised learning via evolutionary search. In: *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM Press. pp. 365–369.

Kohavi, Ron and George H. John (1997). Wrappers for feature subset selection. *Artificial Intelligence* **97**(1), 273–324.

Mitra, P., C. A. Murthy and S. K. Pal (2002). Unsupervised feature selection using feature similarity. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(3), 301–312.

Siedlecki, W. and J. Sklansky (1988). On automatic feature selection. *International Journal of Pattern Recognition and Artificial Intelligence* **2**(2), 197–220.

Tian, Ying-Li, Takeo Kanade and Jeffrey Cohn (2001). Recognizing action units for facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **23**(1), 97 – 115.

Wyse, N., R. Dubes and A.K. Jain (1980). A critical evaluation of intrinsic dimensionality algorithms. In: *Pattern Recognition in Practice* (E.S. Gelsema and L.N. Kanal, Eds.). pp. 415–425. Morgan Kaufmann Publishers, Inc.