International Conference on Innovative Applications in Engineering and Information Technology(ICIAEIT-2017)

International Journal of Advanced Scientific Technologies ,Engineering and Management Sciences (IJASTEMS-ISSN: 2454-356X) Volume.3,Special Issue.1,March.2017

Review On Statistical Pattern Recognition Methods

Padmavathi Vanka, Research Scholar, SPMVV (email: vanka.padmavathi@gmail.com)
Prof. T.Sudha, Department of Computer Science, SPMVV,Tirupathi, India.
(email: thatimakula_sudha@yahoo.com)

Abstract: The main goal of pattern recognition is supervised or unsupervised classification. Pattern recognition has been conventionally expressed using the following main approaches like template matching, statistical methods, syntactic methods and neural networks. Among these statistical approach has been mostly studied and also used in practice. The design of a pattern recognition system requires careful attention to the following issues: definition of pattern classes, pattern representation, feature extraction and feature selection, cluster analysis, classifier design and learning, selection of training samples, selection of test samples, and performance evaluation. The emerging applications, such as data mining, web searching, retrieval of multimedia data, biometrics, face recognition, and handwriting recognition, requires efficient pattern recognition techniques. In this paper various statistical pattern recognition methods and applications are reviewed.

Key Terms: Statistical pattern recognition, classification, clustering, feature extraction, feature selection, error estimation, classifier combination, neural network.

I.INTRODUCTION

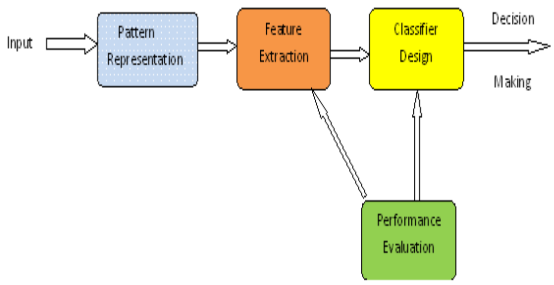
Pattern can be defined as a physical object which is represented by a set of descriptions. Pattern recognition is a subject deals with object description and classification methods. It is also a collection of statistical, mathematical, heuristic and inductive techniques of fundamental role in executing the tasks like human on computers.The practice of recognizing the patterns and classifying the data accordingly has been gaining interest from a long time and human beings have developed highly sophisticated skills for sensing from their environment and take actions according to what they observe. So a human can recognize the faces without worrying about the varying facial rotation, illuminations, facial expressions, and facial biometrical changes. It becomes a difficult task, if the point of implementing such recognition artificially came. By making the systems as intelligent as human to recognize patterns in varying environmental conditions, the fields of artificial intelligence have made this complex task possible. Such a branch of artificial intelligence is known as pattern recognition. Pattern Recognition provides the solution to a lot of problems that fall under the category of either recognition or classification, such as speech recognition, face recognition, classification of handwritten characters, medical diagnosis etc,.

The rapidly growing and available computing power, while enabling faster processing of huge data sets, has also facilitated the use of elaborate and diverse methods for data analysis and classification. At the same time, demands on automatic pattern recognition systems are rising enormously due to the availability of large databases and stringent performance requirements (speed, accuracy, and cost). In many of the emerging applications, it is clear that no single approach for classification is optimal and that multiple methods and approaches have to be used. Consequently, combining several sensing modalities and classifiers is now a commonly used practice in pattern recognition.

The classifier should perform well inspite of inherent variability of patterns and noise in feature extraction and/or in class labels as given in training set. Statistical Pattern Recognition – An approach where the variabilities are captured through probabilistic models. There are other approaches, e.g., syntactic pattern recognition, fuzzy-set based methods etc. Classification and regression (function learning) problems has to be considered in the statistical framework. Statistical Pattern Recognition X is the feature space. (We take X = $\Re$ n). We Consider a 2-class problem for simplicity of notation. A given feature vector can come from different classes with different probabilities.

We consider PR as a two step process – Feature measurement/extraction and Classification. A classifier is to map feature vectors to Class labels. Function learning is a closely related problem. The main information we have for the design is a training set of examples.

Fig1: Block structure of pattern recognition

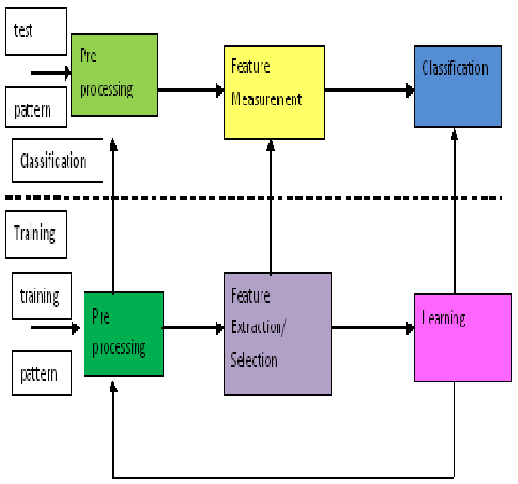


The statistical pattern recognition technique is very popular because most problems in this area deals with noisy data and data uncertainty. The statistics and probability are the best tools to deal with noisy data and data uncertainty. In case of statistical pattern recognition, the vector spaces are used to represent patterns and classes. Pattern recognition is an important and mature area of scientific activity. The two important activities under pattern recognition are pattern classification and clustering. The most important step in pattern recognition

International Journal of Advanced Scientific Technologies ,Engineering and Management Sciences (IJASTEMS-ISSN: 2454-356X) Volume.3,Special Issue.1,March.2017

is the representation of both patterns and classes. It is important to understand that there is no general theory for representing patterns in any domain. However, a good representation scheme helps in achieving better classifiers.

Fig: Model for Statistical Pattern Recognition



The Bayes classifier: In statistical pattern recognition, we model variations of feature values through probability distributions. The statistical viewpoint gives us one way of looking for 'optimal' classifier. We saw Bayes classifier – put the pattern into the class with highest posterior probability. The Bayes classifier is optimal in the sense of minimizing probability of misclassification. The notation X – feature space, Usually ℜn. Y – set of class labels. X = (x1, · · ·xn)T – feature vector. A classifier is a function h : X → Y (= {0, 1}) A classifier maps feature vectors to class labels f0, f1 – class conditional densities (over X) pi= Prob[y(X) =i], i = 0, 1 – prior probabilities. qi(X) = Prob[y(X) =i|X], i = 0, 1. – posterior probabilities. Bayes Theorem: fY |X(y|x) = fX|Y (x|y)fY (y) fX(x) . By Bayes theorem: qi(X) = fi(X)pi Z where Z = f0(X)p0 + f1(X)p1 is the normalising constant. The Bayes classifier hB(X) = 0 if q0(X) > q1(X) = 1 otherwise. q0(X) > q1(X) is same as p0f0(X) > p1f1(X). Minimizes probability of error in classification. Optimality • Each classifier h maps X to {0, 1}. For any classifier h, let Ri(h) = {X ∈ X : h(X) = i}, i = 0, 1. That is, R0(h) is the set of all feature vectors that get classified as Class-0 by the classifier h. Let F(h) denote probability of error for h as defined earlier. F(h) = P[X ∈ R1(h), X ∈ C-0] + P[X ∈ R0(h), X ∈ C-1] = p0P[X ∈ R1(h)|X ∈ C-0] + p1P[X ∈ R0(h)|X ∈ C-1] = p0 ZR1(h) f0(X)dX + p1 ZR0(h) f1(X)dX .Note that for each X we add either p0f0(X) or p1f1(X) in calculating the error integral.

Nearest Neighbour classifiers: Nearest neighbour classifiers are the most popular in literature as they are very simple for both the human and the machine (no learning) and at the same time are robust. A major difficulty associated with them is the classification time. There is no design time; however, classification (testing) time is linear in the number of training patterns. In order to improve the situation, several efficient pre-processing schemes have been proposed in literature. They employ either a reduced data set or a data structure and an associated efficient algorithm to find the nearest neighbours. Nearest neighbour classifier is a simple classifier that often performs very well. We store some feature vectors from the training set as prototypes. (Can be the whole training set!). Note that we know the (correct) class label for each prototype. Given a new pattern (feature vector) X we find the prototype X′ that is closest to X. Then classify X into the same class as X. A variation: k-NN rule. Find the k prototypes closest to X. Classify X into the majority class of these prototypes. There are two main issues in designing an NN classifier. Selection of Prototypes and Distance between feature vectors. A very simple classifier to design and operate k-NN. Time and memory needs depend on number of prototypes and complexity of distance function.

Neural Networks: Neural network is a 'good' parameterized class of nonlinear discriminant functions. Multilayer feed forward neural nets are one such class. Nonlinear functions are built up through composition of summation and sigmoids. It is useful for both classification and Regression. Neural networks can be viewed as massively parallel computing systems consisting of an extremely large number of simple processors with many interconnections. Neural network models attempt to use some organizational principles (such as learning, generalization, adaptivity, fault tolerance and distributed representation, and Pattern Recognition Models computation) in a network of weighted directed graphs in which the nodes are artificial neurons and directed edges (with weights) are connections between neuron outputs and neuron inputs. The main characteristics of neural networks are that they have the ability to learn complex nonlinear input-output relationships, use sequential training procedures, and adapt themselves to the data.

Decision Tree: Decision trees are friendly data structures to both the programmer and the manager (decision maker). They are frequently used in both pattern classification and data mining. The decision tree can deal with both numerical and categorical features. We have covered the axis parallel decision tree based classifiers in detail with appropriate examples. There are other categories of decision trees including oblique decision trees. They are computationally expensive but we have covered them briefly and the interested reader may look up the references for more details. Linear discriminant function based classifiers have played an important role in the recent development of pattern recognition. The perceptron is one of the earliest classifiers which has been studied both from analytical and practical angles. It led to the development of artificial neural networks, specifically, multilayer perceptrons. Divide feature space so that a linear classifier is enough in each region (e.g. Decision Trees). 0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1 0 0.1 0.2 0.3

0.4 0.5 0.6 0.7 0.8 0.9 1 H1 H2 H3 H1 H2 H3 C0 C1 C0. Such tree-based models are possible for regression also.

Support Vector Machine: A more recent exploit, in this area, is the support vector machine (SVM). SVM can be easily identified as the most successful and popular classifier in the past decade. Statistical pattern recognition gained a significant prominent place because of SVMs. Another prominent direction in pattern recognition is the use of more than one classifier to decide the class label of a test pattern. There are several possible schemes for combining classifiers. An important contribution here is the ADABOOST algorithm.

SVM Map X nonlinearly into a high dimensional space and try a linear discriminant function there. (e.g. Support Vector Machines). Let X = [x1 x2] and let $\varphi : \Re 2 \rightarrow \Re 5$ given by Z = φ(X) = [1 x1 x2 x21 x22 x1x2]. Now, g(X) = a0 + a1x1 + a2x2 + a3x21 + a4x22 + a5x1x2 is a quadratic discriminant function in ℜ2; but g(Z) = a0 + a1z1 + a2z2 + a3z3 + a4z4 + a5z5 is a linear dscriminant function in the ‘φ(X)’ space.

Clustering is an important tool in recognition. Even though it is studied as an end product in itself, it is not always so in practice. Typically, the results of clustering are useful for further decision making. In such a context, clustering may be viewed as an abstraction generation tool—for example, pre-processing of the training data to reduce its size.

Table: Classification Methods

Method	Property	Comment
Bayes Classifier	Assign pattern to the class, which has the maximum estimated posterior probability	This classifier yields best for Gaussian Distribution (linear) is sensitive to density based errors. It minimizes the risk under general loss function
Nearest Neighbour Classifier	Assigns pattern to the class which contains nearest mean	It finds a prototype that is closest to the given pattern. No Training is needed and it undergoes fast testing.
k- Nearest Neighbour classifier	Assigns pattern to the majority class among k nearest neighbour classes using optimized value of K.	It is asymptotically optimal, it undergoes slow testing, Scale dependent
Neural Networks	Iterative Mean Square Estimation of two or more layers of perceptrons using sigmoidal transfer function.	Sensitive to training parameters, slow training, nonlinear classification, may produce confidence values, over training sensitive, needs regularization
Support Vector Machines	Maximizes the margin between the classes by selecting a minimum number of support vectors.	Scale dependent, iterative, slow training, non linear, over training sensitive good generalization performance,

Conclusions:

In its broadest sense pattern recognition is the heart of all scientific inquiry, including understanding ourselves and the real-world around us. And the developing of pattern recognition is increasing very fast, the related fields and the application of pattern recognition became wider and wider. In this paper we amplify statistical pattern recognition in the round, include the definition of PR, statistical PR, different classifiers of statistical PR, and the application of pattern recognition. In addition, it is an important trend to use pattern recognition on engineering; we should make efforts on this. And pattern recognition scientists should pay attention to new technique of PR, and enlarge the application areas of PR.

References:

[1] A.K. Jain, J. Mao, K.M. Mohiuddin, "*Artificial Neural Networks: a Tutorial*," Computer, vol. 29, no. 3, 1996, 31-44.

[2] J. Wood, "*Invariant pattern recognition: A review*," Pattern Recognition, vol. 29, no. 1, 1996, 1-17.

[3] P.F. Baldi and K. Hornik, "*Learning in linear neural networks: A survey*," IEEE Transactions on Neural Networks, vol. 6, no. 4, 1995, 837-858.

[4] Yu.G. Smetanin, "*Neural Networks as Systems for Pattern Recognition: a Review*," Pattern Recognition and Image Analysis, vol. 5, no. 2, 1995, 254-293.

[5] A.C. Tsoi and A.D. Back, "*Locally recurrent globally feedforward networks - A critical review of architectures*," IEEE Transactions on Neural Networks, vol. 5, no. 2, pp. 229-239, 1994.

[6] S.J. Raudys and A.K. Jain, "*Small sample size effects in statistical pattern recognition: recommendations for practitioners*," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 13, no. 3, pp. 252-264, 1991.

[7] S.B. Gelfand and E.J. Delp, "*On Tree Structured Classifiers*," in: Artificial neural networks and statistical pattern recognition, ed. I.K. Sethi and A.K. Jain, pp. 51 - 70, North-Holland, Amsterdam, 1991.

[8] A.K. Jain and B. Chandrasekaran, "*Dimensionality and Sample Size Considerations in Pattern Recognition Practice*," in: Handbook of Statistics, vol. 2, ed. P.R. Krishnaiah and L.N. Kanal, pp. 835 - 855, North-Holland, Amsterdam, 1987.