ISSN 2047-3338

Pattern Recognition Techniques: A Review

[1]Seema Asht and [2]Rajeshwar Dass
[1,2]Department of Electronics and Communication Engineering, DCRUST, Murthal, Sonepat, India
[1]seema.asht@gmail.com, [2]rajehswardas10@gmail.com

Abstract– **Pattern Recognition has attracted the attention of researchers in last few decades as a machine learning approach due to its wide spread application areas. The application area includes medicine, communications, automations, military intelligence, data mining, bioinformatics, document classification, speech recognition, business and many others. In this review paper various approaches of Pattern Recognition have been presented and their pros-cons, application specific paradigm has been shown. On the basis of survey, pattern recognition techniques can be categorized into six parts. These include Statistical Techniques, Structural Techniques, Template Matching, Neural Network Approach, Fuzzy Model and Hybrid Models.**

Index Terms– **Pattern Recognition, Statistical Pattern Recognition, Structural Pattern Recognition, Neural Networks and Fuzzy Sets**

I. INTRODUCTION

RECOGNIZING the objects and the surrounding environment is a trivial task for human beings. But if the point of implementing it artificially came, then it becomes a very complex task. Pattern Recognition provides the solution to various problems from speech recognition, face recognition to classification of handwritten characters and medical diagnosis.

The various application areas of pattern recognition are like bioinformatics, document classification, image analysis, data mining, industrial automation, biometric recognition, remote sensing, handwritten text analysis, medical diagnosis, speech recognition, GIS and many more. Similarity between all these applications is that for a solution-finding approach features have to be extracted and then analyzed for recognition and classification purpose. Three processes take place in pattern recognition task. First step is data acquisition. Data acquisition is the process of converting data from one form (speech, character, pictures etc.) into another form which should be acceptable to the computing device for further processing. Data acquisition is generally performed by sensors, digitizing machine and scanners. Second step is data analysis. After data acquisition the task of analysis begins. During data analysis step the learning about the data takes place and information is collected about the different events and pattern classes available in the data. This information or knowledge about the data is used for further processing. Third step used for pattern recognition is classification. Its purpose is to decide the category of new data on the basis of knowledge received from data analysis process. Data set presented to a Pattern Recognition system is divided into two sets: training set and testing set. System learns from training set and efficiency of system is checked by presenting testing set to it. The performance of the pattern recognition techniques is influenced by mainly three elements (i) amount of data (ii) technology used(method) (iii) designer and the user. The challenging job in pattern recognition is to develop systems with capability of handling massive amounts of data. The various models opted for pattern recognition are:

Statistical Techniques, Structural Techniques, Template Matching, Neural Network based techniques, Fuzzy models and Hybrid Models.

II. PATTERN RECOGNITION MODELS

Models opted for pattern recognition can be categorized in to different categories depending upon the method used for data analysis and classification. Models can be independently or dependently used to perform a pattern recognition task [1]. The different models used for pattern recognition task are as follow:

A. Statistical Model

In Statistical method of Pattern Recognition each pattern is described in terms of features. Features are chosen in such a way that different patterns occupy non-overlapping feature space. It recognizes the probabilistic nature both of the information we seek to process, and of the form in which we should express it [2]. It works well when the selected features lead to feature spaces which cluster in a recognizable manner, i.e. there is proper interclass distance. After analyzing the probability distribution of a pattern belonging to a class, decision boundary is determined [3], [4]. Here patterns are projected to pre-processing operations to make them suitable for training purposes. Features are selected upon analyzing training patterns. System learns and adapts itself for unknown patterns as shown in Fig. 1.Test patterns are applied to check suitability of system to recognize patterns. Feature measurement is done while testing, then these feature values is presented to learned system and in this way classification is performed.

When conditional probability density distribution is known, parametric classification schemes are used otherwise non parametric classification scheme need to be used. Various decision rules are there to determine decision boundary like, Bayes Decision Rule, Optimal Bayes Decision Rule, The Maximum Likelihood Rule, Neyman-Pearson rule and MAP rule. As feature spaces are partitioned, system becomes noise insensitive, therefore in case of noisy patterns. The choice of statistical model is a good solution. Depending upon whether the method opted is supervised or unsupervised statistical technique can be categorized as: Discriminant Analysis and Principal Component Analysis [1].

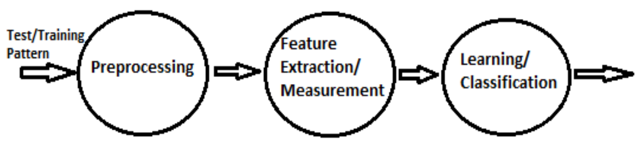

Fig. 1: Statistical Pattern Recognition Model

Discriminant Analysis is a supervised technique in which we approach for dimensionality reduction. Here linear combination of features is utilized to perform the classification operation. For each pattern class, a discriminant function is defined which performs the classification function [5] - [8]. There is not a well defined rule regarding the form of discriminant function like minimum distance classifier uses one reference point for each class, and discriminant function computes minimum distance from unknown vectors to these points, on the other hand nearest neighbor classifier uses set of points for each class. There are various kinds of Discriminant Analysis methods that are used based upon the application and system requirement such as: Linear Discriminant Analysis (LDA), Null-LDA (N-LDA), Fisher Discriminant Analysis (FDA), Two Dimensional Linear Discriminant Analysis (2D-LDA), Two Dimensional Fisher Discriminant Analysis (2D-FDA). In LDA feature set is obtained by linear combination of original features. Intra-class distance is minimized as well as inter-class distance is maximized to obtain the optimum results.LDA suffers from small sample size (SSS) problem.

In FDA ratio of variance in inter-classes to variances in intra-classes defines the separation between classes. In FDA inter-class scatter is maximized and intra-class scatter is minimized to get the optimum results [8]. FDA approach is a combination of PCA and LDA. 2D-LDA avoids small sample size (SSS) problem associated with 1D-LDA. Here matrices of input data are computed to form the feature vector. Trace of interclass scatter matrix is maximized while trace of intra-class scatter matrix is minimized to get the optimum results in 2 D-LDA. As compared to 1-D LDA; 2D-FDA provides non-singular interclass and intra-class matrices. Chen et al. [9] suggested that the null space spanned by eigenvectors of intra-class scatter matrices having zero Eigen values contains highly discriminating information. An LDA method in null space of intra-class scatter matrix is N-LDA which involves solving the Eigen value problem for a very large matrix. Principal Component Analysis (PCA) or Karhunen-Loeave expansion is a multi-element unsupervised technique in which we approach for dimensionality reduction [10]. Using PCA, patterns are detected in the data and these patterns determine the similarity measure [11]. In PCA Eigen vectors with largest Eigen values are computed to form the feature space. PCA is closely related to Factor Analysis [11]. Kernel PCA is a solution for nonlinear feature extraction [12], [13].Other non-linear feature extraction techniques are Multidimensional scaling (MDS) and Kohonen feature Map [14]. Application areas of PCA include graphically unreliable patterns. Discriminant Analysis is more efficient as compared to PCA, in terms of accuracy and time elapsed [15].

B. Structural Model

When we came across patterns with strong inherent structures, statistical methods give ambiguous results, because feature extraction destroys vital information concerning the basic structure of pattern. Therefore, in complex pattern recognition problems, like recognition of multidimensional objects it is preferred to adopt a hierarchical system, where a pattern is considered to be made up of simple sub-patterns, which are further composed of simpler sub patterns [16], [17]. In structural approach of pattern recognition a collection of complex patterns are described by a number of sub-patterns and the grammatical rules with which these sub patterns are associated with each other. This model is concerned with structure and attempts to recognize a pattern from its general form. The language which provides structural description of patterns in terms of pattern primitives and their composition is termed as pattern description language. Increased descriptive power of a language leads to increased complexity of syntax analysis system.

To recognize finite-state languages finite- state automata is used. Descriptive power of finite-state languages is weaker than that of context-sensitive languages. Context sensitive languages are described by non-deterministic procedures. Selection of type of grammar for pattern description depends upon the primitives and on the grammar’s descriptive power and analysis efficiency [18]. For description of patterns such as chromosome images, 2D-mathematics, chemical structures, spoken words, English characters and fingerprint patterns, a number of languages have been suggested [19], [20]. High dimensional patterns need high dimensional grammars such as web grammars, tree grammars, graph grammars and shape grammars for efficient description [19], [21]-[23].

Stochastic languages, approximation and transformational grammars are used to describe noisy and distorted patterns [19], [24]-[26]. This approach demands large training sets and very large computational efforts [27]. When dealing with noisy patterns, grammar defining the basic structure of complex patterns becomes too difficult to define, there in such cases statistical approach is a good option. Acceptance Error is the criterion to measure performance. This model is used in the application areas like in textured images, shape analysis of contours and image interpretation where patterns have a definite structure [28].

C. Template Matching Model

Template matching is simplest and most primitive among all pattern recognition models. It is used to determine the similarity between two samples, pixels or curves. The pattern to be recognized is matched with the stored templates while assuming that template can be gone through rotational or scalar changes. The efficiency of this model depends upon the stored templates. Correlation function is taken as recognition function and is optimized depending on the available training set. The shortcoming of this approach is that, it does not work efficiently in the presence of distorted patterns [29].

D. Neural Network Based Model

Neural networks are the massively parallel structures composed of "neuron" like subunits [29]. Neural networks provide efficient result in the field of classification. Its property of changing its weight iteratively and learning [10], [30], give it an edge over other techniques for recognition process. Perceptron is a primitive neuron model. It is a two layer structure. If output function of perceptron is step, then it performs classification problems, if it is linear than it perform regression problems. The most commonly used family of neural networks for pattern classification is the feed forward networks like MLP and RBF networks [31]. Different types of neural networks are used depending upon the requirement of the application.

Feed Forward Back-propagation Neural Network (FFBP-NN) is used to implement non-linear differentiable functions. Increase in the learning rate in back-propagation neural network leads to decrease in convergence time [32]. General Regression neural network (GRNN) is a highly parallel structure in which learning is from input side to output side [33]. General Regression Neural Network (GRNN) performs efficiently on noisy data than Back-propagation. FFBP Neural Network does not work accurately if available data is large enough. On the other hand in GRNN, as the size of data increases, the error approaches towards zero [33]. Kohnen-Networks are mainly used for data clustering and feature mapping [14]. Ripley [34] and Anderson et al. [35] stated the relationship between neural networks and statistical model of pattern recognition.

The performance of the neural networks enhances upon increasing the number of hidden layers up to a certain extent. Increased number of neurons in hidden layer also improves the performance of the system. No. of neurons must be large enough to adequately represent the problem domain and small enough to permit the generalization from training data. A trade-off must be maintained between size of network and complexity resulted because of network size. Percentage recognition accuracy of a neural network can be further enhanced if we use 'tansig'-'tansig' combination of activation functions for neurons of hidden layer and output layer opted rather than opting for other combinations [36].

E. Fuzzy Based Model

The importance of fuzzy sets in Pattern Recognition lies in modeling forms of uncertainty that cannot be fully understood by the use of probability theory [37], [38]. Kandel [39] states, "In a very fundamental way, the intimate relation between theory of fuzzy sets and theory of Pattern Recognition and classification rests on the fact that most real world classes are fuzzy in nature", Kandel defined various techniques of fuzzy pattern recognition. Syntactic techniques are utilized when the pattern sought is related to the formal structure of language. Semantic techniques are used when fuzzy partitions of data sets are to be produced. Then a similarity measure based on weighted distance is used to obtain similarity degree between the fuzzy description of unknown shape and reference shape.

F. Hybrid Model

In most of the emerging applications, it is clear that a single model used for classification doesn't behave efficiently, so multiple methods have to be combined together giving result to hybrid models. Primitive approaches to design a Pattern Recognition system which aims at utilizing a best individual classifier have some drawbacks [40]. It is very difficult to identify a best classifier unless deep prior knowledge is available at hand [40], [41]. Statistical and Structural models can be combined together to solve hybrid problems. In such cases statistical approach is utilized to recognize pattern primitives and syntactic approach is then used for the recognition of sub-patterns and pattern itself. Fu [28] gave the concept of attributed grammars which unifies statistical and structural pattern recognition approach. To enhance system performance one can use a set of individual classifiers and combiner to make the final decision. Tumer and Ghosh [42] experimentally proved that using a linear combiner or order statistics combiner minimize the variance of actual decision boundaries around the optimal boundary. Multiple classifiers can be used in several ways to enhance the system performance. Each classifier can be trained in a different region of feature space or in other way, each classifier can provide probability estimate and decision can be made upon analyzing individual results. Methods utilizing classifier ensemble design [43], [44] generate a set of mutually complementary classifiers that achieve optimal accuracy using a fixed decision function. Those methods which utilize combination function design tend to find an optimal combination of decisions from a set of classifiers. To achieve optimum results, a large set of combination functions of increasing complexity, ranging from simple voting rules through trainable combination functions is available to designer [45] - [47].

III. CONCLUSION

A comparative view of all the models of pattern recognition has been shown which depicts that for various domains in this areas different models or combination of models can be used. In case of noisy patterns, choice of statistical model is a good solution. Practical importance of structural model depends upon recognition of simple pattern primitives and their relationships represented by description language. As compared to statistical pattern recognition, structural pattern recognition is a newer area of research. For complex patterns and applications utilizing large number of pattern classes, it is beneficial to describe each pattern in terms of its components. A wise decision regarding the selection of Pattern grammar influences computations efficiency of recognition system. Pattern primitives and pattern grammar to be utilized depends upon the application requirements. Low dependence of neural networks on prior knowledge and availability of efficient learning algorithms have made the neural networks famous in the field of Patten Recognition. Although neural networks and statistical pattern recognition models have different principles most of the neural networks are similar to statistical pattern recognition models. To recognize unknown shapes fuzzy methods are good options. As each model has its own pros and cons, therefore to enhance system performance for complex applications it is beneficial to append two or more recognition models at various stages of recognition process.

REFERENCES

[1]. K.S. Fu, “A Step towards Unification of Syntactic and Statistical Pattern Recognition,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 5, no. 2, pp. 200-205, Mar. 1983.

[2]. Amin Fazel and Shantnu Chakrabartty, “ An Overview of Statistical Pattern Recognition Techniques for Speaker Verification’’, IEEE circuits and systems, pp. 61-81, magazine 2nd quarter 2011.

[3]. L. Devroye, L. Gyorfi, and G. Lugosi, "A Probabilistic Theory of Pattern Recognition." Berlin: Springer-Verlag, 1996.

[4]. R.O.Duda and P.E.Hart, Pattern Classification and Scene Analysis, New York: John Wiley & sons, 1973.

[5]. W.W.Cooley and P.R. Lohnes," Multivariate Data Analysis."New York: Wiley, 1971.

[6]. R.Fisher, "The Use of Multiple Measurements in Taxonomic Problems," Ann. Eugenics, vol. 7, no. 2, pp. 179-188,1936.

[7]. M.M. Tatsouka, "Multivariate Analysis."New York: Wiley, 1971.

[8]. Zhen Lei, Rufeng Chu, Ran He, Shengcai Liao and Stan Z. Li., “Face Recognition by Discriminant Analysis with Gabor Tensor Representation,” Vol.4642/2007, pp.87-95, Springer Berlin/Heidelberg Publishing, 2007.

[9]. L.Chen, H.Liao, M.Ko, J.Lin and G.Yu," A New LDA-Based Face Recognition System which can Solve the Small Sample Size Problem", Pattern Recognition,2000.

[10]. Ethem Alpaydin: Introduction to Machine Learning, Prentice Hall of India 2005.

[11]. Yuehui Sun, Minghui: “DT-CWT Feature Based Classification Using Orthogonal Neighborhood Preserving Projections for Face Recognition,” Volume: 1, pp.719-724.Nov.2006.

[12]. S. Haykin, Neural Networks, A Comprehensive Foundation. Second ed., Englewood Cliffs, N.J.: Prentice Hall, 1999.

[13]. B. SchoÈlkopf, A. Smola, and K.R. Muller, "Nonlinear Component Analysis as a Kernel Eigenvalue Problem," Neural Computation, vol. 10, no. 5, pp. 1,299-1,319, 1998.

[14]. T. Kohonen, Self-Organizing Maps, Springer Series in Information Sciences, vol. 30, Berlin, 1995.

[15]. Tasweer Ahmad.Ahlam Jameel,Dr. Balal Ahmad, "Pattern Recognition using Statistical and Neural Techniques", IEEE 2011.

[16]. K.S.fu, Syntactic Pattern recognition and Applications. Englewood Cliffs, N.J.: Prentice Hall,1982.

[17]. T.Pavlidis, “Structural Pattern Recognition,” New York: Springer-Verlag, 1977.

[18]. King-Sun Fu and Azriel Rosenfeld, "Pattern Recognition and Image Processing, "IEEE Trans. on computers, vol. c-25.No. 12, Dec. 1976.

[19]. Syntactic Methods in Pattern Recognition, New York;Academic,1974

[20]. K.S. Fu, "Applications of Syntactic Pattern Recognition."New York: Springer, 1976.

[21]. J.Gips, “Shape Grammers and their Uses.”Birkhauser: Verlag, Basel and Stuttgart, 1975.

[22]. P.A. Ota, "Mosaic Grammars," Pattern Recognition, vol. 7, June 1975.

[23]. K.L.Williams, “A Multidimensional Approach to Syntactic Pattern Recognition,” Pattern Recognition, vol. 7, Sept. 1975.

[24]. U.Grenander, "Foundations of pattern analysis, "Quart. Appl. Math., vol. 27, pp. 1-55, 1969.

[25]. K.S.Fu, Ed., Digital Pattern Recognition, Communications and Cybernatics, vol. 10, New York: Springe, 1976.

[26]. T.Pavlidis and F.Ali, "Computer Recognition of Handwritten Numerals by Polygonal Approximations,"IEEE Trans. Syst.,Man,cybern.vol. SMC-5, Nov. 1975.

[27]. L.I. Perlovsky, “Conundrum of Combinatorial Complexity,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20, no. 6, pp. 666-670, 1998.

[28]. K.S. Fu, Syntactic Pattern Recognition and Applications. Englewood Cliffs, N.J.: Prentice-Hall, 1982.

[29]. R.Bajcsy and S.Kovacic, “Multiresolution Elastic Matching,” Computer vision graphics image processing, vol46, pp. 1-21, 1989.

[30]. Nils J. Nilsson: Artificial Intelligence A New Synthesis, ELSEVIERA.K. Jain, J. Mao, and K.M. Mohiuddin, “Artificial Neural Networks: A Tutorial,” Computer, pp. 31-44, Mar. 1996.

[31]. Anupam Joshi, Narendram Ramakrishman, Elias N.Houstis and John R.Rice, "On Neurobiological, Neuro-Fuzzy, Machine Learning And Statistical Pattern Recognition Techniques", IEEE Trans. on Neural Networks, vol. 8, no. 1, 1997.

[32]. Burcu Erkmell, Tu.lay YlIdmm, “Improving classification performance of sonar targets by applying general regression neural network with PCA,” Science Direct.

[33]. B. Ripley, “Statistical Aspects of Neural Networks, Networks on Chaos: Statistical and Probabilistic Aspects.” U. Bornndorff-Nielsen, J.Jensen, and W. Kendal, eds., Chapman and Hall, 1993.

[34]. J. Anderson, A. Pellionisz, and E. Rosenfeld, Neurocomputing 2: Directions for Research. Cambridge Mass.: MIT Press, 1990.

[35]. Amit Choudhary,Rahul Rishi, Savita Ahlawat and Vijaypal Singh Dhaka, “Performance Analysis Of Feed Forward MLP with Various Activation Functions For Handwritten Numerals Recognition”, IEEE Trans. 2010,volume 5,pp. 852-856

[36]. J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithm, New York:Plenum Press,1981.

[37]. J.C. Bezdek And S. K., Fuzzy Models For Pattern Recognition: Methods that Search for Structures In Data.” Pal,Eds.,IEEE CS Press,1992.

[38]. Kandel,A., Fuzzy Techniques in Pattern Recognition”, John Wiley and Sons, New York,1982

[39]. J. Kittler, F. J. Ferri, J. M. I~nesta, A. Amin and P. Pudil, Eds., “A framework for classifier fusion: is it still needed in Advances in Pattern Recognition,” LNCS 1876, 45-56 (Springer-Verlag, 2000).

[40]. R. O. Duda, P. E. Hart and D. G. Stork, Pattern Classification (John Wiley & Sons 2000).

[41]. K. Tumer and J. Ghosh, "Analysis of Decision Boundaries in Linearly Combined Neural Classifiers", Pattern Recognition, vol. 29, pp. 341-348, 1996.
[42]. J. Kittler and F. Roli, (Eds.), Multiple Classifier Systems," LNCS 1857, 404 (Springer-Verlag, 2000).
[43]. A. J. C. Sharkey, "Multi-Net Systems", in Combining Artificial Neural Nets,Ensemble and Modular Multi-Net Systems, 1-27, (Springer-Verlag, 1999).
[44]. G. Giacinto and F. Roli, "Dynamic Classifier Selection based on Multiple Classifier Behaviour", Pattern Recognition, 34, 179 -181 (2001).
[45]. G. Giacinto, F. Roli and G. Fumera, "Selection of Image Classifiers", Electronics Letters 36, 420{422 (2000).
[46]. L. I. Kuncheva, "Combinations of multiple classifiers using fuzzy sets," in Fuzzy Classifier Design, 233-267 (Springer-Verlag, 2000).

Seema Asht is pursuing her M. Tech. from DCRUST Murthal in Electronics and Communication Engg. She received her B.E. (Honors) Degree in Electronics & Communication Engineering from CRSCE, Murthal (Sonepat) (Affiliated to M.D.U Rohtak) in 2010. Her interest includes Neural Network and Wireless Communication.

Rajeshwar Dass received M.E Degree in Electronics & Communication Engineering from National Institute of Technical Teachers Training and Research Center (NITTTR) Chandigarh in the year 2007 and B.E. (Honors) degree in Electronics & Communication Engineering from Apeejay College of Engineering Sohna Gurgaon (Affiliated to M.D.U Rohtak) in 2004. He is pursuing his Ph.D from DCRUST Murthal. In August 2004 he joined Department of Electronics & Communication Engineering of Apeejay College of Engineering Sohna Gurgaon. He joined Department of Electronics & Communication Engineering of Deenbandhu Chhotu Ram University of Science and Technology (D.C.R.U.S.T) Murthal Sonepat(India) as Assistant professor in October 2008. His interest includes Medical Image Processing, Neural Network, Wireless Communication and Soft Computing Techniques. He has contributed 15 technical papers in International Journals and Conferences. He has written a book on wireless communication. He is the member of IEEE (92061144).