ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

Comparative Study of Pattern Classifiers

Jagbeer Kaur*[1], Deep Kamal Kaur Randhawa[2]

M.Tech Student, Department of ECE, GNDU RC Jalandhar, India*[1]

Assistant Professor, Department of ECE, GNDU RC Jalandhar, India*[2]

ABSTRACT: This paper presents a comparative study of different pattern classifiers with its assumptions, advantages and limitations. Patterns classifiers are commonly used for pattern recognition which has wide scope in medical, industrial and commercial applications such as face recognizer, fingerprint detection, hand writing recognizer, noise detection, video surveillance in computer vision and many more. It has overlap scope with machine learning, data mining and knowledge discovery in Databases. The motivation is to have an overview of different published algorithms and analysing its potential future trends.

KEYWORDS: Classifier, supervised, Support Vector Machine, Gaussian Mixture Model, K- Mean, Adaptive Navie Bayes, Desicion Tree

I. INTRODUCTION

Generally classifier means a thing or a person that classifies something in different classes from previously acquired knowledge on the basis of similarities. In technical sense it is a category of unified modelling Language (UML) which consists of common data, features i.e. methods or attributes.

The Pattern classification Technique consists of six steps in general [1]. First is to define a problem and determine number of classes required. The Second Step is to extract Features or known as Descriptor. It is also the step where the designer of a machine vision solution is faced with many options of types of features (e.g., color based, texture based, boundary oriented, etc.) and has the specific methods to extract them. Third is the Feature Selection method using various tools and techniques in the field of data mining and machine learning and choose one that best suits our needs, based on their complexity, computational cost, training capabilities, and other properties. The Next step is Pattern classification technique. It can have two categories, Supervised and Unsupervised Classification. Supervised Classification is first applying knowledge, then classifying. Unsupervised Classification is first classifying and then applies knowledge. Choose the appropriate Technique. The Fifth Step is to have a data set which will help to train and test the solution. The Sixth Step is to Select a subset of images and use them to train the classifier. Many pattern classification strategies require a training stage, where a small subset of images is used to “teach” the classifier about the classes it should be able to recognize, as well as adjust some of the classifier’s parameters. Tests are conducted to improve accuracy by precision and recall ratio to reduce error rate. Based on Application Specific Feedback report refinement and improvement is done.Classifier first of all train by using known features then test the data to identify the class by comparing their features with trained data.

Pattern Classification has various applications [2] including medical and industrial and commercial application. It is use in handwriting recognizers, pattern detection, figure print detection, noise classification, face recognizers, speech recognizers, and biomedical image recognizers etc.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

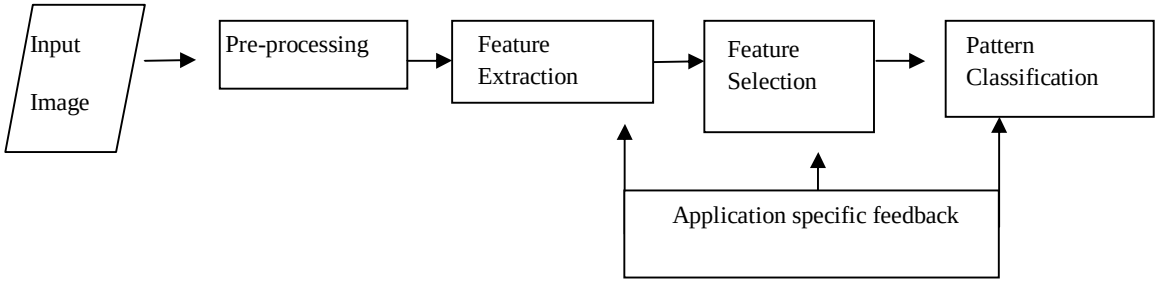

Figure 1 The Interplay between feature extraction, feature selection and pattern classification as a function of the application at hand [1].

II. DIFFERENT TYPE OF CLASSIFIER

Different types of classifier are described below in two categories according to their classification stage as Supervised and Unsupervised Classifier.

A. SUPERVISED CLASSIFIERS

Supervised classifiers are the classifier in which firstly knowledge is applied then classification is done. The Supervised classifiers are as described below.

a) Minimum Distance Classifier

In Minimum distance classifier each class is represented by its mean vector. Training is done using the objects of the known class. Mean of the feature vectors for the object with in the class is calculated. New object finding are classified by finding the closest mean vector.

Consider that $\{X_{j1}, X_{j2}, \ldots X_{jn_j}\}$ be n_j feature vectors which is used to represent the class c_j . Every feature vector contains a label of class j that has been selected for representing the class. The center of the class j is given in (1)

$$M_j = \frac{1}{n_j} \sum_{i=1}^{n_j} X_j^i \qquad \text{eq. (1)}$$

An input feature vector X of unknown class was classified in the classification phase depends on the distance to each class center and given in (2)

$$X \in \text{Class } c_j, \text{ if } d(X, M_j) = \min_i d(X, M_j) \qquad \text{eq. (2)}$$

Where, $d(X, M_j)$ is the distance function based on the distance metric selected during the training phase.

Minimum Distance Classifier includes SVM Classifier.

- SVM Classifier

The SVM stands for Support Vector Machine. It is a powerful classifier[2] and work on wide range of classification problems, including particular problems and more complex once compare to other classifier such as KNN, GMM, or ANBC algorithms. It also provides a front end to the LIBSVM library [3]. Its algorithm is a part of the GRT classification modules. It work with any of the four basic kernels that are linear, polynomial, radial basis function and sigmoid.

The Support Vector Machine classifier works on wide range of classification of problems including problem in high dimension and overcome the problems of GMM, KNN or ANBC classifier and that are not linearly separable.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

The disadvantage involve in the SVM is parameter for which is must be clearly and correctly define to achieve the best and accurate results of classification of the given problem. As the result for problem A is defines excellent, may be poor for problem B.

So the parameter for problems must have parameter setting according to the problem defined to achieve the required results.

b) Maximum Likelihood Classifier

Classify according to the largest probability taking variance and covariance into consideration. Assume that distribution within each class is Gaussian .The distribution in each class can be described by a mean vector and covariance matrix[4].

For each class from training data compute: Mean vector Covariance matrix .Form decision function for each class .New objects are classified to class with highest probability.

Variance is spread a randomness for class and Covariance is influence which is dependency between different features described by covariance matrix.

Table 1: Covariance matrix.

	Feature 1	Feature 2	Feature 3
Feature 1	1	1&2	1&3
Feature 2	1&2	2	2&3
Feature 3	1&3	1&3	3

Features : $X_1, X_2,..., X_n$

Feature vector for object i: $x_1, x_2,.....,x_n$

Mean for each class $x_{mean1}, x_{mean2},.....x_{meanm}$

Covariance: $cov(x_i, x_j) = \frac{1}{n-1}\sum_{k=1}^{n}(x_i, k - x_{meani})\,(x_j, k - x_{meanj})$

$$\Sigma = \begin{bmatrix} cov(x_1,x_1) & cov(x_1,x_2) & cov(x_1,x_3) \\ cov(x_2,x_1) & cov(x_2,x_2) & cov(x_2,x_3) \\ cov(x_3,x_1) & cov(x_3,x_2) & cov(x_3,x_3) \end{bmatrix}$$

The classifier include in Maximum Likelihood Classifier is GMM Classifier.

- GMM Classifier

GMM stands for Gaussian Mixture Model Classifier [5]. It is a basic classifier and its algorithm is part of the GRT (Gesture Recognition Tool kit) classification modules which is useful in supervised learning classifications that use in recognition various N-Dimension signals. Its algorithm is simple and best useful for classification of static posture and non-temporary pattern recognition.

GMM algorithm consist the limitation at computational reasons, it fails in work if dimensionality of the problem is too high that is greater than 6 dimensions for instance. To overcome this problem different classifier can be use either the ANBC or Support Vector machine classification algorithms instead.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

Second disadvantage is training dataset. Many time users not know how many mixture models are suitable for classification. Previously user must require to know suitable number of mixture model but it is difficult to have this prerequirement in every in case.

c) Nearest Neighbour Classifier

Nearest-neighbour classifiers are based on learning by resemblance that is by comparing a given test sample with the available training samples which are similar to it. For a data sample X to be classified, its K-nearest neighbours are searched and then X is assigned to class label to which majority of its neighbours belongs to. The choice of k also affects the performance of k-nearest neighbour algorithm [6]. If the value of k is too small, then K-NN classifier may be vulnerable to over fitting because of noise present in the training dataset. On the other hand, if k is too large, the nearest-neighbour classifier may misclassify the test sample Stores all training samples. Assigns pattern to majority class among k nearest neighbour. Nearest Neighbour Classifier include KNN Classifier. K-NN fundamentally works on the belief that the data is connected in a feature space. Hence, all the points are considered in order, to find out the distance among the data points. Euclidian distance or Hamming distance is used according to the data type of data classes used [7]. In this a single value of K is given which is used to find the total number of nearest neighbours that determine the class label for unknown sample. If the value of K=1, then it is called as nearest neighbour classification.

As it is a basic classifier and provides accurate results to simple problems. Easy to understand and implement. Training process is faster. It is robust to noisy training data.

The main disadvantage of the KNN algorithm is it is lazy learner because it simply use the training data for classification. To predict the label of a new instance the KNN algorithm will find the *K* closest neighbors to the new instance from the training data, the predicted class label will then be set as the most common label among the *K* closest neighboring points. The main disadvantage of this approach is that the algorithm must compute the distance and sort all the training data at each prediction, which can be slow if there are a large number of training examples. Another disadvantage of this approach is that the algorithm does not learn anything from the training data, which can result in the algorithm not generalizing well and also not being robust to noisy data. Further, changing *K* can change the resulting predicted class label.

d) ANB Classifier

The Adaptive Naive Bayes Classifier technique is based on Bayesian Theorem and it is used when the dimensionality of the inputs is high [8]. Bayesian classification is based on Bayes Theorem.

Let X is a data sample whose class label is not known and let H be some hypothesis, such that the data sample X may belong to a specified class C.

Bayes theorem is used for calculating the posterior probability P(C|X), from P(C), P(X), and P(X|C). Where P(C|X) is the posterior probability of target class. P(C) is called the prior probability of class. P(X|C) is the likelihood which is the probability of predictor of given class. P(X) is the prior probability of predictor of class.

$$P(C|X) = \frac{P(X|C).P(C)}{P(X)} \qquad \text{eq.(3)}$$

The Naive Bayes classifier [9] works as follows:

1. Let D be the training dataset associated with class labels. Each tuple is represented by n-dimensional element vector, $X=(X_1, X_2, X_3,....., X_n)$.

2. Consider that there are m classes $c_1, c_2, c_3...., c_m$. Suppose that we want to classify an unknown tuple X, then the classifier will predict that X belongs to the class with higher posterior probability, conditioned on X. i.e., the Naive Bayesian classifier assigns an unknown tuple X to the class Ci if and only if P(Ci|X) > P(Cj|X) For $1 \leq j \leq m$, and $i \neq j$, above posterior probabilities are computed using Bayes Theorem.

The main advantage of ANBC algorithm is that it can be use in classification of the static postures and temporal pattern recognition. It requires short computational time for training

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

The main disadvantage of the ANB classifier is when the data is not linearly separable, it not work well for classification as in Gaussian distribution algorithm classifier to represent every class and SVM classifier can be tried. This means less accurate in comparison to other classifier. Also require large amount of data.

B. UNSUPERVISED CLASSIFIERS

Unsupervised classifiers are the classifier in which firstly classification is done then knowledge is applied. The Unsupervised classifiers are as described below.

a) K Mean Classifier

A Kernel-Based Fuzzy C-Means clustering (KFCM) algorithm has been proposed by Zhang et al [10] with strong noise robustness. In this algorithm, similarity measure in FCM, i.e. Euclidean norm metric, is replaced by a new kernel-induced metric which makes the weighted sum of data points more robust. So this algorithm is a robust clustering approach if an appropriate value for sigma would be chosen. It's obtained by trial-and-error technique or experience or prior knowledge which neither too large nor too small.

It works with small data sets. It is helpful to rerun the classification using the same as well as different K values, to compare the achieved results. Suitable for noisy data set.

Different initialization can result in different final clusters. Fixed number of clusters can make it difficult to predict what K should be. Clustering with 10 intialization in 2D but with N dimension than 10 different initialization is not also enough.

b) Hierarchical Clustering

Hierarchical clustering - construct clustering tree (dendrogram). Start with each object/pixel as its own class. Merge the classes that are closest according to some distance measure. Continue until only one class is achieved. Decide the number of classes based on the distances in the tree.

Hierarchical Clustering includes decision tree classifier.

- Decision Tree Classifier

A decision tree structure is made of root, internal and leaf nodes. It is a flow chart like tree structure, where every internal node denotes a test condition on an attribute, each branch represents result of the test condition, and each leaf node (or terminal node) is assigned with a class label. The topmost node is the root node. Decision tree is constructed in a divide and conquer approach. Each path in decision tree forms a decision rule. Generally, it utilizes greedy approach from top to bottom.

Decision tree classification technique is performed in two phases: tree building and tree pruning [11]. Tree building is performed in top-down approach. During this phase, the tree is recursively partitioned till all the data items belong to the same class label. It is very computationally intensive as the training dataset is traversed repeatedly. Tree pruning is done in a bottom-up manner. It is used to improve the prediction and classification accuracy of the algorithm by minimizing over-fitting problem of tree. Over-fitting problem in decision tree results in misclassification error. This class implements a basic Decision Tree classifier. Decision Trees are conceptually simple classifiers that work well on even complex classification tasks. Decision Trees partition the feature space into a set of rectangular regions, classifying a new datum by finding which region it belongs to.

The Decision Tree algorithm is a good algorithm to use for the classification of static postures and non-temporal pattern recognition. The main advantage of a Decision Tree is that the model is particularly fast at classifying new input samples. It gives a easy representation model and accurate with results.

The main limitation of the Decision Tree algorithm is that very large models will frequently over fit the training data. It needs long training time. Not suitable for replication problem and suffers from over fitting issues. Not suitable for Noisy Data set.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

III. CONCLUSION

In this paper we have discussed about classifier. Analysing Various Types of Classifiers based on two categories supervised and unsupervised. A Comparison is made by reviewing their advantages and disadvantages respectively. As classifier are used to enhance and improve the results accordingly to their problem domain.

They are highly application dependent. For simple data problem KNN classifier can be use, for high degree of problem SVM classifier can be use with large data set requirement, for static posture and non-temporary pattern recognition GMM classifier can be use for high degree of Accuracy. In Unsupervised classification KFCM is best with small dataset and when Large data set is required than Decision tree can be used .

Table 2: Comparison of Classifier

Parameter	SVM	GMM	KNN	ANB	KFCM	Decision Tree
Deterministic /Non Deterministic	Deterministic	Deterministic	Deterministic	Deterministic	Non Deterministic	Non Deterministic
Effectiveness on	Huge data	Small Data	Small Data	Huge Data	Small Data	Large Data
Speed	Slower for large Data	faster	Slower for large Data	Faster Than KNN	Faster Than Decision Tree	Fast
Dataset	Noisy Data	Noisy Data	Non Noisy Data	Noisy Data	Noisy Data	Noisy Data
Accuracy	Accurate with Specific kernel value	Highly Accurate	Highly Accurate	Good	Accurate for small data set	Highly Accuracy

REFERENCES

[1]. Oge Marques, " Practical Image and Video Processing Using MATLAB", John Wiley & Sons, 2011.
[2]. R.P.W. Duin, "A Note on Comparing Classifiers", Pattern Recognition Letters 17 ,1996.
[3]. Chang, Chih-Chung, and Chih-Jen Lin. "LIBSVM: a library for support vector machines", ACM Transactions on Intelligent Systems and Technology (TIST) 2.3: 27, 2011.
[4]. Foody, Giles M., et al. "Derivation and applications of probabilistic measures of class membership from the maximum-likelihood classification." Photogrammetric Engineering and Remote Sensing 58.9: 1335-1341, 1992.
[5]. Y.Huang, Kevin B. Englehart and Adrian D. C. Chan, "A Gaussian Mixture Model Based Classification Scheme for Myoelectric Control of Powered Upper Limb Prostheses,"IEEE Transactions On Biomedical Engineering, Vol. 52, NO. 11,pg 1801-1811, November 2005.
[6]. M. Soundarya and R. Balakrishnan, "Survey on Classification Techniques in Data mining", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, Issue 7, July 2014.
[7]. R. Duda, and P. Hart, "Pattern Classification and Scene Analysis", John Wiley and Sons, New York, 1973.
[8]. N. Friedman, D. Geiger, and Goldazmidt, "Bayesian Network Classifiers", Machine Learning, vol. 29, pp. 131-163, 1997.
[9]. BhaveshPatankar and Dr. Vijay Chavda, "A Comparative Study of Decision Tree, Naive Bayesian and k-nn Classifiers in Data Mining", International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 4, Issue 12, December 2014.
[10]. Rastgarpour, M.; Alipour, S.; Shanbehzadeh, J. "Improved Fast Two Cycle by using KFCM Clustering for Image Segmentation." Proceedings of the International MultiConference of Engineers , Vol. 1, p640,2012.
[11]. Matthew N. Anyanwu and Sajjan G. Shiva, "Comparative Analysis of Serial Decision Tree ClassificationAlgorithms", Researchgate.