

Classification Techniques in Pattern Recognition

Lihong Zheng and Xiangjian He
Faculty of IT, University of Technology, Sydney
PO Box 123, Broadway NSW 2007, Sydney, Australia
{Lzheng, sean}@it.uts.edu.au

ABSTRACT

In this paper, we review some pattern recognition schemes published in recent years. After giving the general processing steps of pattern recognition, we discuss several methods used for steps of pattern recognition such as Principal Component Analysis (PCA) in feature extraction, Support Vector Machines (SVM) in classification, and so forth. Different kinds of merits are presented and their applications on pattern recognition are given. The objective of this paper is to summarize and compare some of the methods for pattern recognition, and future research issues which need to be resolved and investigated further are given along with the new trends and ideas.

Keywords

Pattern recognition, feature extraction, feature selection, mapping, kernels, support vector machines

1. INTRODUCTION

Pattern recognition can also be seen as a classification process. Its ultimate goal is to optimally extract patterns based on certain conditions and is to separate one class from the others. Pattern recognition was often achieved using linear and quadratic discriminants [1], the k-nearest neighbor classifier [2] or the Parzen density estimator [3], template matching [4] and Neural Networks [5]. These methods are basically statistic. The problem of using these recognition methods is having to construct a classification rule without having any idea of the distribution of the measurements in different groups. Support Vector Machine (SVM) [6] SVMs have gained prominence in the field of pattern classification. They are forcefully competing with other techniques such as template matching and Neural Networks for pattern recognition.

This paper is organized as follows. We first introduce some general process of pattern recognition and basic techniques in section 2. Conclusions are made in section 3.

2. GENERAL PROCESS OF PR

A pattern is a pair comprising an observation and a meaning. Pattern recognition is inferring meaning from observation. Designing a pattern recognition system is establishing a mapping from measurement

space into the space of potential meanings, whereby the different meanings are represented in this space as discrete target points. The basic components in pattern recognition are preprocessing, feature extraction and selection, classifier design and optimization.

2.1 Preprocessing

The role of preprocessing is to segment the interesting pattern from the background. Generally, noise filtering, smoothing and normalization should be done in this step. The preprocessing also defines a compact representation of the pattern.

2.2 Feature Selection and extraction

Features should be easily computed, robust, insensitive to various distortions and variations in the images, and rotationally invariant. Two kinds of features are used in pattern recognition problems. One kind of features has clear physical meaning, such as geometric or structural and statistical features. Another kind of features has no physical meaning. We call these features mapping features. The advantage of physical features is that they need not deal with irrelevant features. The advantage of the mapping features is that they make classification easier because clear boundaries will be obtained between classes but increasing the computational complexity.

Feature selection is to select the best subset from the input space. Its ultimate goal is to select the optimal features subset that can achieve the highest accuracy results. While feature extraction is applied in the situation when no physical features can be obtained. Most of feature selection algorithms involve a combinatorial search through the whole space. Usually, heuristic methods, such as hill climbing, have to be adopted, because the size of input space is exponential in the number of features. Other methods

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or to publish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

*Conference proceedings ISBN 80-903100-8-7
WSCG'2005, January 31-February 4, 2005
Plzen, Czech Republic.
Copyright UNION Agency – Science Press*

divide the feature space into several subspaces which can be searched easily.

There are basically two types of feature selection methods: filter and wrapper [7]. Filters methods select the best features according to some prior knowledge without thinking about the bias of further induction algorithm. So these methods performed independently of the classification algorithm or its error criteria.

In feature extraction, most methods are supervised. These approaches need some prior knowledge and labeled training samples. There are two kinds of supervised methods used: Linear feature extraction and nonlinear feature extraction. Linear feature extraction techniques include Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), projection pursuit, and Independent Component Analysis (ICA). Nonlinear feature extraction methods include kernel PCA, PCA network, nonlinear PCA, nonlinear auto-associative network, Multi-Dimensional Scaling (MDS) and Self-Organizing Map (SOM), and so forth.

2.3 Classifiers design

After optimal feature subset is selected a classifier can be designed using various approaches. Roughly speaking, there are three different approaches [14]. The first approach is the simplest and the most intuitive approach which is based on the concept of similarity. Template matching is an example. The second one is a probabilistic approach. It includes methods based on Bayes decision rule, the maximum likelihood or density estimator. Three well-known methods are K-nearest neighbour (KNN), Parzen window classifier and branch-and bound methods (BnB). The third approach is to construct decision boundaries directly by optimizing certain error criterion. Examples are fisher's linear discriminant, multilayer perceptrons, decision tree and support vector machine. The important advantage of SVM is that it offers a possibility to train generalizable, nonlinear classifiers in high dimensional spaces using a small training set. SVMs generalization error is not related to the input dimensionality of the problem but to the margin with which it separates the data. That is why SVMs can have good performance even with a large number of inputs. There are many kinds of methods aiming at reducing the computational burden for pattern recognition. Examples are K-nearest neighbor method, Parzen Window, Clustering, PNN and Branch-and-bound. KNN's significant disadvantage is that the distance must be calculated between an unknown and every prototype each time a sample is recognized. Parzen Window depends on the kernel function and on the value of the window-width h . It allows us to obtain complex nonlinear decision boundaries. Clustering method

aims at partitioning a given set of N data into M groups so that similar vectors are grouped together. PNN's main idea of the PNN can be generalized so that we can optimize multi-merging steps. BnB technique uses a search tree for finding the optimal clustering and generates clustering through a sequence of merging operations.

2.4 Optimization

The optimization is not a separate step, it is combined with several parts of the pattern recognition process. In preprocessing, optimization guarantee that the input pattern have the best quality. Then in the feature selection and extraction part, optimal feature subsets are obtained under some optimization techniques. Furthermore, the final classification error rate is lowered in the classification part.

3. CONCLUSION

The basic idea we get is: the more relevant patterns at your process, the better features subsets you obtain, the more simple your classifier will be applied, finally the better your decisions will be. Based on our analysis of various methods, a combination of various techniques may be a better way for our final goal that will utilize available domain knowledge to make decisions automatically and accurately. In summary, we should attempt to design a hybrid system combining with multiple models.

4. REFERENCES

- [1] R.A. Fisher, "The Use of Multiple Measurements in Taxonomic Problems," *Annals of Eugenics*, vol. 7, part II, pp. 179-188, 1936.
- [2] Dasarathy, B.V.; "Minimal consistent set (MCS) identification for optimal nearest neighbor decision systems design", *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 24, Issue: 3, pp:511 – 517, March 1994.
- [3] Girolami, M.; Chao He; "Probability density estimation from optimally condensed data samples" *Pattern Analysis and Machine Intelligence*, *IEEE Transactions on*, Volume: 25, Issue: 10, pp:1253 – 1264, Oct. 2003.
- [4] Meijer, B.R.; "Rules and algorithms for the design of templates for template matching", *Pattern Recognition*, 1992. Vol.1. Conference A: Computer Vision and Applications, 11th IAPR International Conference on, pp: 760 – 763, Aug. 1992.
- [5] Hush, D.R.; Horne, B.G.; "Progress in supervised neural networks", *Signal Processing Magazine*, *IEEE*, Vol. 10, Issue: 1, pp:8 – 39, Jan. 1993
- [6] Vapnik, V., *The Nature of Statistical Learning Theory*, Springer, 1995.
- [7] Julia Neumann, Christoph Schnorr, "SVM-based feature selection by direct objective minimization", 2004.