

Chapter 2

Feature Selection Based on Minimizing the Area Under the Detection Error Tradeoff Curve

Liau Heng Fui

University of Nottingham Malaysia, Malaysia

Dino Isa

University of Nottingham Malaysia, Malaysia

ABSTRACT

Feature selection is crucial to select an “optimized” subset of features from the original feature set based on a certain objective function. In general, feature selection removes redundant or irrelevant data while retaining classification accuracy. This paper proposes a feature selection algorithm that aims to minimize the area under the curve of detection error trade-off (DET) curve. Particle swarm optimization (PSO) is employed to search for the optimal feature subset. The proposed method is implemented in face recognition and iris recognition systems. The result shows that the proposed method is able to find an optimal subset of features that sufficiently describes iris and face images by removing unwanted and redundant features and at the same time improving the classification accuracy in terms of total error rate (TER).

1. INTRODUCTION

In real world application, the feature set is generally large in term of dimensionality. The features may be noisy and may contain irrelevant or redundant information about the target concept. This may cause performance degradation on classifiers. Besides that, large feature set also increases the storage cost and require more computation time

to process. Feature selection is crucial to select an “optimized” subset of features from the original feature set based on certain objective function. In general, feature selection removes redundant or irrelevant data while retaining classification accuracy. Although there has been a great deal of work in different areas to address this issue (Guyon & Elissee, 2003), these results have not been fully explored or exploited in emerging computer vision applications. Only recently, there has been an increased interest in implementing feature

DOI: 10.4018/978-1-4666-3628-6.ch002

selection for applications such as face detection (Sun, Bebis & Miller, 2003; Viola & Jones, 2001), face recognition (Shen & Ji, 2009; Kanan & Faez, 2008a, 2008b), auditory brainstem responses tracking (Acir, Ozdamar & Guzelisc, 2005), gene expression data (Chuang et al., 2008) etc.

Sequential forward selection (SFS) and sequential backward selection (SBS) are the two well-known feature selection methods. SFS starts with an empty feature set; the feature that benefits the performance is added iteratively. In contrast, SBS starts with the full feature set and at each step, feature whose absence causes least decreases in terms of classification performance is dropped. Combining SFS and SBS gives birth to the “plus l -take away r ” feature selection method (Stearns, 1976), which first enlarges the feature subset by adding l using SFS and then deletes r features using SBS. Sequential forward floating search (SFFS) and sequential backward floating search (SBFS) (Pudil, Novovicova & Kittler, 1994) are generalizations of the Stearn’s method since they make decisions based on single feature iteratively. Hence, they cannot be expected to find globally optimal solutions. Another famous feature selection method is the relief algorithm (Kira & Rendell, 1992) and extension of it (Wiskott et al., 1994). Features are ranked according to the hypothesis margin. Features that give large hypothesis margin are selected.

Recently, evolutionary computing methods, such as genetic algorithms (GA) (Siedlecki & Sklansky, 1989), ant colony optimization (ACO) (Dorigo & Caro, 1999) and particle swarm optimization (PSO) (Kennedy & Eberhart, 1995, 1997) have gained more popularity in the feature selection area. Genetic algorithms (GA’s) are optimization techniques based on the mechanism of natural selection. They try to mimic the operations found in natural genetics to guide itself through the paths in the search space (Siedlecki & Sklansky, 1989). Sun, Bebis, & Miller (2003) employed GA to select a subset of principal components analysis (PCA) features for vehicle and face

detection. The fitness evaluation is based on accuracy and number of principle components. The subset that consists of fewer numbers of features and gives comparable result is favored. Kanan & Faez (2008a) proposed a scale, rotation and translation invariant face recognition system using Pseudo Zernike moment invariant (PZMI). PZMI feature set is very large and an optimal subset of PZMI features subset is selected using GAs. ACO algorithm was inspired by ant’s social behavior. Ants are capable of finding the shortest route between a food source and their nest by chemical materials called pheromone that they release when moving. Kanan & Faez (2008b) employed ACO to perform feature selection on a PZMI and discrete wavelet transform (DWT) features that represent facial image. PSO was inspired by the graceful but unpredictable movements of a flock of birds and a school of fishes. Inertia weight was later introduced into the particle swarm optimizer to produce the standard PSO (Shi & Eberhart, 1998). Huang and Dun (2008) employed PSO to performed feature selection and at the same time tuning the parameter of SVM classifier. The binary PSO is incorporated to select the features sets while a common PSO is used to tune SVM’s parameter and the kernel’s parameter. PSO has been used to perform feature selection for gene expression data (Chuang et al, 2008).

Receiver operation characteristic (ROC) curve has been widely used in measuring the performance of binary classifiers. The performance of a classifier can be evaluated based on a single terms, which is the area under the ROC curve. Greater the area under the curve (AUC) implies better robustness of the classifier. Several methods have been proposed based on the idea of maximizing the area under the ROC curve. Toh, Kim and Lee (2008a) proposed a parameter selection method similar to that of the classifiers that maximized the AUC of ROC curve. Zhang et al (2006) perform feature selection based the AUC of ROC curve for each features. The AUC of ROC features below certain threshold value will be discarded. Features

14 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/feature-selection-based-minimizing-area/74920

Related Content

The Formal Design Models of a Universal Array (UA) and its Implementation

Yingxu Wang, Jason Huang and Jingsheng Lei (2013). *Advances in Abstract Intelligence and Soft Computing* (pp. 241-259).

www.irma-international.org/chapter/formal-design-models-universal-array/72785

Supporting CSCW and CSCL with Intelligent Social Grouping Services

Jeffrey J.P. Tsai, Jia Zhang, Jeff J.S. Huang and Stephen J.H. Yang (2009). *International Journal of Software Science and Computational Intelligence* (pp. 51-63).

www.irma-international.org/article/supporting-cscw-cscl-intelligent-social/2785

Scalable Edge Computing Environment Based on the Containerized Microservices and Minikube

Nitin Rathore and Anand Rajavat (2022). *International Journal of Software Science and Computational Intelligence* (pp. 1-14).

www.irma-international.org/article/scalable-edge-computing-environment-based-on-the-containerized-microservices-and-minikube/312560

Machine Learning in Personalized Anemia Treatment

Adam E. Gaweda (2010). *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques* (pp. 265-276).

www.irma-international.org/chapter/machine-learning-personalized-anemia-treatment/36989

Feature Selection Algorithms for Classification and Clustering

Arvind Kumar Tiwari (2017). *Ubiquitous Machine Learning and Its Applications* (pp. 143-167).

www.irma-international.org/chapter/feature-selection-algorithms-for-classification-and-clustering/179092