Information Sciences 223 (2013) 102–118

Contents lists available at SciVerse ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Feature subset selection using separability index matrix

Jeong-Su Han [a], Sang Wan Lee [b,*], Zeungnam Bien [c]

[a] Samsung Electronics, Suwon, Republic of Korea
[b] Computation & Neural Systems, Behavioral & Social Neuroscience, California Institute of Technology, Pasadena, CA, USA
[c] School of Electrical and Computer Engineering, Ulsan National Institute of Science and Technology (UNIST), Ulsan, Republic of Korea

ARTICLE INFO

Article history:
Received 1 July 2009
Received in revised form 23 November 2011
Accepted 26 September 2012
Available online 10 October 2012

Keywords:
Feature subset selection
Filter method
Separability index matrix
EMG signal
Gait phase recognition

ABSTRACT

Effective Feature Subset Selection (FSS) is an important step when designing engineering systems that classify complex data in real time. The electromyographic (EMG) signal-based walking assistance system is a typical system that requires an efficient computational architecture for classification. The performance of such a system depends largely on a criterion function that assesses the quality of selected feature subsets. However, many well-known conventional criterion functions use less relevant features for classification or they have a high computational cost. Here, we propose a new criterion function that provides more effective *FSS*. The proposed criterion function, known as a *separability index matrix* (*SIM*), provides features pertinent to the classification task and a very low computational cost. This new function produces to a simple feature selection algorithm when combined with the forward search paradigm. We performed extensive experimental comparisons in terms of classification accuracy and computational costs to confirm that the proposed algorithm outperformed other filter-type feature selection methods that are based on various distance measures, including inter–intra, Euclidean, Mahalanobis, and Bhattacharyya distances. We then applied the proposed method to a gait phase recognition problem in our EMG signal-based walking assistance system. We demonstrated that the proposed method performed competitively when compared with other wrapper-type feature selection methods in terms of class-separability and recognition rate.

1. Introduction

Feature subset selection (FSS) is a process for identifying n most informative features $V = \{\upsilon_1, \ldots, \upsilon_n\}$ from N known features $S = \{x_1, x_2, \ldots, x_N\}$ ($n < N$), assuming that the C target classes $W = \{w_1, w_2, \ldots, w_C\}$ are given and an observed dataset described by M samples (instances) is available. The selection of the most informative feature set leads to an improvement in classification accuracy, faster and more cost-effective classification performance, and a better understanding of the underlying process of the observed dataset [11,19].

It would not be an overstatement to claim that FSS is the most important step when designing an engineering system because to its necessity for online classification of complex data. The electromyographic (EMG) signal-based walking assistance system is a typical system that requires an efficient computational architecture with effective *FSS*. Previous reports of EMG signal applications to lower body movements [4,8,10,16,17], suggest that the limited computational capacity of hardware and the high computational costs of the recognition task are crucial determinants of success when developing a walking assistance system [9]. Many factors affect the computational costs (e.g., input dimensionality, and the complexity of a

* Corresponding author.
E-mail address: swlee@caltech.edu (S.W. Lee).

0020-0255/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ins.2012.09.042

feature extraction and classification) but the size of a selected feature subset is of great importance mainly because the feature subset size dictates the memory allocation, which is the main bottleneck in mobile computing systems.

The division of effective FSS into a *filter* method and a *wrapper* method [19] is a key paradigm that provides a functional difference when evaluating the quality of a selected feature subset. In the filter method, the criterion function utilizes quantitative information such as the interclass distance of selected features [11]. However, the criterion function used by the wrapper method relies on performance metrics for the classifier such as accuracy, specificity, and precision.

The wrapper method often performs better at classification compared with the filter method, but it requires significantly higher computational costs because the fitness evaluation of a subset requires cross-validation or a bootstrapping procedure during the error estimation for each subset [19,23]. Furthermore, the choice of the classifier inevitably biases the characteristics of the selected feature subset, which often leads to the loss of any generalization capability [15].

Various criterion functions have been used for the filter method and they can be categorized into two groups, i.e., *distance-based* measures and *relation-based* measures. The Fisher ratio [3,26], Mahalanobis distance [5], and Bhattacharyya distance [5] are typical examples of distance-based measures that use a distance metric to measure class separability. These measures assign an average value of separability to classes for each feature. Thus, they may select features that are highly correlated if the selected features have a large average separability value. Relation-based measures are represented by correlation, mutual information [1,23], or fuzzy dependency based on a penalized Euclidean distance [7]. They extract the relation between features and classes. Using these approaches, a good feature is highly correlated within classes, but uncorrelated with other features. However, there are high computational costs when acquiring good features using these measures, especially when mutual information is involved, although some heuristics can reduce this cost [1,13,14,27].

We propose a computationally efficient criterion function based on a novel concept of a "separability index matrix (SIM)". The proposed method effectively distinguishes relevant features from irrelevant and/or redundant features.

This paper is organized as follows: In Section 2, we review various well-known criterion functions that are typically used by the filter method. In Section 3, we explain the concept of the separability index matrix (SIM) and we propose a new criterion function for effective FSS. Section 4 details the SIM-based feature selection method (SIMF) and its properties. Section 5 provides extensive experimental results, which are compared with benchmark datasets. We demonstrated the validity of the proposed method in a realistic situation and we present its application to an EMG signal-based gait phase recognition problem in Section 6. Concluding remarks are provided in Section 7.

2. Criterion functions of a filter method

Choosing a good feature subset is an important step in pattern classification tasks. A specific criterion function is required to evaluate the quality of a selected feature subset. In particular, it is always possible to find an optimal feature subset using the Branch-and-Bound technique if a monotonic criterion function is provided [5,25]. In general, the overall performance of FSS depends largely on the selected criterion function.

Various criterion functions have been reported, such as interclass distance [11], statistical dependence [19], and information-theoretic measures [1,23]. These measures can be categorized into two groups depending on the method they use to evaluate the quality of feature subsets. We refer to these groups as "*distance-based* measures" and "*relation-based* measures".

Distance-based measures characterize the quality of a feature based on its ability to discriminate instances of a class from instances of other classes. Instances from different classes (between-class) should have feature values that are more distinctive than values from the same class (within-class). Measures such as the Fisher ratio, Euclidean distance, Mahalanobis distance, and Bhattacharyya distance are representative examples in this group.

For example, it is well-known that the Fisher ratio is defined as the ratio of the between-class difference to the within-class spread [3] as follows:

$$\lambda_{i,j,l} = \frac{(m_{i,l} - m_{j,l})^2}{(\sigma_{i,l}^2 + \sigma_{j,l}^2)} \tag{1}$$

where $m_{i,l}$, $m_{j,l}$, $\sigma_{i,l}^2$, and $\sigma_{j,l}^2$ are the means and variances of instances of the ith class and jth class in the direction of the lth feature, while $\lambda_{i,j,l}$ indicates the class separation degree between the ith class and jth class in the direction of the lth feature. The Fisher ratio provides a good measure of class separability because it increases as the between-class difference increases and as the within-class spread decreases. The following generalized Fisher ratio [12] can be used in the case of a multi-class problem:

$$\lambda_l = \frac{1}{C(C-1)} \frac{\sum_{i=1}^{C}\sum_{j=1}^{C} \phi_i \phi_j \lambda_{i,j,l}}{\sum_{i=1}^{C}\sum_{j=1}^{C} \phi_i \phi_j}, \quad i \neq j. \tag{2}$$

Here, λ_l is the average class separability measure in the direction of the lth feature and C is the total number of classes, while i and j indicate the class indices with $1 \leqslant i, j \leqslant C$. Also, φ_i and φ_j are the mixing weights for the ith and jth class, respectively, and $\lambda_{i,j,l}$ is the Fisher ratio defined in (1).

The rationale for devising relation-based measures is that a good feature subset contains features that are highly correlated with classes, but uncorrelated with other features. The relevance of this approach is usually characterized in terms of