Expert Systems with Applications 42 (2015) 573–583

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Maximum relevancy maximum complementary feature selection for multi-sensor activity recognition

Saisakul Chernbumroong [a], Shuang Cang [b], Hongnian Yu [a,*]

[a] Faculty of Science and Technology, Bournemouth University, Poole, Dorset BH12 5BB, UK
[b] School of Tourism, Bournemouth University, Poole, Dorset BH12 5BB, UK

ARTICLE INFO

Article history:
Available online 23 August 2014

Keywords:
Feature selection
Neural networks
Mutual information
Activity recognition

ABSTRACT

In the multi-sensor activity recognition domain, the input space is often large and contains irrelevant and overlapped features. It is important to perform feature selection in order to select the smallest number of features which can describe the outputs. This paper proposes a new feature selection algorithms using the maximal relevance and maximal complementary (MRMC) based on neural networks. Unlike other feature selection algorithms that are based on relevance and redundancy measurements, the idea of how a feature complements to the already selected features is utilized. The proposed algorithm is evaluated on two well-defined problems and five real world data sets. The data sets cover different types of data i.e. real, integer and category and sizes i.e. small to large set of features. The experimental results show that the MRMC can select a smaller number of features while achieving good results. The proposed algorithm can be applied to any type of data, and demonstrate great potential for the data set with a large number of features.

1. Introduction

The aim of feature selection is to identify the smallest subset of input features which explains the output classes. This process is important especially to the classification problems with a large number of input features. For example, a multi-sensor activity classification system normally contains a large number of input features generated from different sensors. Feature selection can help reduce the size of feature space which leads to reduction in computational cost and complexity in the classification system. In real world problems where input features contain irrelevant and redundant features, feature selection can help identify a relevance feature set which leads to improvement in classification performances.

There are three main approaches in feature selection found in wearable sensor-based activity recognition applications: intuition, filter, and wrapper. Intuition based feature selection requires a domain knowledge or understanding which is required in the classification of the interested activities. This approach is often used in conjunction with visual inspection, statistical analysis of the features e.g. histogram, distribution graph, or observation made during activity occurrence (Parkka et al., 2006 & Ward et al., Ward, Lukowicz, Troster, & Starner, 2006). Filter based-feature selection measures the relevance between features and the outputs by using techniques such as information theory, distance, correlation, receive operating curve (ROC), etc. Each feature is evaluated for its relevance then given a ranking score. For example, features which have the best performance in discriminating the interested activities were selected using ROC (Banos, Damas, Pomares, Prieto, & Rojas, 2012; Ermes, Parkka, Mantyjarvi, & Korhonen, 2008). Many of the statistical tests are used with this approach e.g. chi-square, T-test, etc. The study in Banos et al. (2012) found that features selected from the ranking quality group technique based on discrimination and robustness, ROC, T-test or the Wilcoxon with support vector machine produced remarkable results. Mutual information (MI) is another popular measurement used for measuring the relationship between two variables. Feature selection techniques which use MI are such as maximum relevance minimum redundancy (Peng, Long, & Ding, 2005), normalized mutual information feature selection-feature space (Cang & Yu, 2012), feature selection based on cumulate conditional mutual information (Zhang & Zhang, 2012), etc. Some techniques are based on neural networks to rank the features e.g. neural network feature selection (Setiono & Liu, 1997), Clamping technique (Wang, Jones, & Partridge, 2000), constructive approach for feature selection (Kabir, Islam, & Murase, 2010), etc. The main advantages of the filter approach are due to its simplicity, speed and independence of the classification algorithm (Saeys, Inza, & Larraaga, 2007). However, most of the techniques in this approach usually consider

* Corresponding author.

http://dx.doi.org/10.1016/j.eswa.2014.07.052
0957-4174/

two variables i.e. a feature and class output, thus ignoring dependencies among a set of features. This may lead to a selection of redundant features resulting in low classification accuracy. In some techniques such as MRMR (Peng et al., 2005) and NMIFS (Estevez, Tesmer, Perez, & Zurada, 2009), another criteria i.e. redundancy is used to reduce the chance of selecting redundant features.

Wrapper based-feature selection is the most popular technique in wearable sensor-based activity recognition. In this technique, various set of feature subsets are generated and evaluated using a classification algorithm. The most optimum feature subset is selected using search techniques. Examples of this approach are forward selection (Dalton & Olaighin, 2013; Peng, Ferguson, Rafferty, & Kelly, 2011; Zhang & Sawchuk, 2013), backward selection, forward–backward selection (Khan, Lee, Lee, & Kim, 2010), exhaustive search (Varkey, Pompili, & Walls, 2012), etc. In forward selection, one feature is added into a feature subset each time and the subset is evaluated for its performance. On the other hand, backward selection removes one feature from the feature subset each time and evaluates the subset performance. Forward–backward selection employs both directions where forward selection is carried out first then the subset is refined using backward selection. This approach is computationally more extensive than the filter method, however it can provide a better result as it takes into account the features dependency and interaction with the classification algorithm. Some studies combine both filter and wrapper methods. For example, the study in Hsu, Hsieh, and Lu (2011) combined the features filtered by information gain and F-score, then used the wrapper method to improve classification accuracy.

The most feature selection methods in the current literature are based on two criteria i.e. relevancy – how the feature is relevant to outputs, and redundancy – to reduce the chance of selecting redundant features. However, feature selection using these two criteria does not consider how a feature will complement the already selected features. This may result in selecting a larger number of feature than actually required. Also, in some feature selection techniques which only consider the relevancy criteria, redundant features may be selected. For techniques which use the wrapper approach, considering all possible feature subsets suffers a high computational cost.

Considering the above limitations, we propose a new feature selection algorithm with a new criterion i.e. complementary – how a feature complements the already selected features. In addition, based on our knowledge, this criterion has not yet been considered in any other feature selection algorithm. The Clamping technique is employed to measure the feature relevance. We introduce a new measurement to calculate the complementary value of the feature to the already selected feature set. The feature is selected based on the criteria of maximum relevance and maximum complementary. The main difference between the proposed technique and the other algorithms are that the complementary measurement is used instead of the redundancy measurement. Feature redundancy can be detected through the complementary measurement such that the redundant feature should give a low complementary score.

The paper is organized as follows: Section 2 presents some popular feature selection algorithms which are used for comparison in this study. Section 3 presents the proposed feature selection technique in detail. We evaluate our algorithm using two well-defined problems and four benchmark data sets and one multi-sensor activity recognition data set collected from a real home. The experimental results are presented in Section 4. Finally, the discussion and conclusion are presented in Section 5.

2. Related works

Many techniques have been proposed for feature selections as discussed in the previous section. In this paper, we look at two different approaches used for feature ranking i.e. mutual information (MI) and neural networks (NN).

2.1. Mutual information based feature selection

MI, which is based on information theory (Shannon, 2001), measures the dependency between two variables. The MI value is zero if and only if the variables are independent. Given continuous variables f_i and f_j, the MI is:

$$MI(f_i;f_j) = \iint p(f_i,f_j)\log\frac{p(f_i,f_j)}{p(f_i)p(f_j)}\mathrm{d}f_i\mathrm{d}f_j$$

In practice, it is difficult to calculate MI of the continuous values and often the variables are discretized using bins. The MI of discrete variables is:

$$MI(f_i;f_j) = \sum_c\sum_j p(f_i,f_j)\log\frac{p(f_i,f_j)}{p(f_i)p(f_j)}$$

There are many feature ranking algorithms based on the MI (Cang & Yu, 2012; Estevez et al., 2009; Peng et al., 2005; Zhang & Zhang, 2012). The maximal relevant minimal redundant (MRMR) is one of the most popular feature selection algorithms. Many algorithms have been based on MRMR. For example, the normalized mutual information feature selection (NMIFS) which enhance MRMR by using entropy of the variables to normalize the MI values when calculating the redundancy between variables. MRMR is enhanced by using the kernel canonical correlation analysis as inputs rather than the actual features (Sakar, Kursun, & Gurgen, 2012).

In this study we investigate the commonly used feature selection algorithms based on MI which are MRMR and NMIFS algorithms.

2.1.1. MRMR

The MRMR algorithm (Peng et al., 2005) ranks the features based on the minimal redundancy and maximal relevance criterion. It calculates the MI between two features to measure the redundancy and the MI between a feature and the outputs to measure the relevance. Using the MRMR concept and greedy selection, a set of feature rankings S can be obtained as follow:

(A) Given $S = \{\}$ where S is a set of selected features and $F = \{f_1, f_2, f_i, f_j \ldots, f_N\}$ where F is a set of N features, select the feature f_s in F which has the maximum mutual information between itself and output C where $C = \{c_1, c_3, \ldots, c_K\}$ and $f_s = \max_{f_i \in F} MI(f_i; C)$, and update S and F.

$$S = S \cup \{f_s\} \tag{1}$$

$$F = F \setminus \{f_s\} \tag{2}$$

(B) Select feature f_s in F which satisfies the following condition:

$$f_s = \max_{f_i\in F}\left\{MI(f_i;C) - \frac{1}{|S|}\sum_{f_j\in S}MI(f_i;f_j)\right\}$$

Update S and F using (1) and (2).
Repeat Step (B) until the desired number of features is obtained.

2.1.2. NMIFS

The NMIFS algorithm (Estevez et al., 2009) is an enhancement of the MRMR algorithm. A normalized MI (NMI) between two features are used instead:

$$NMI(i;j) = \frac{MI(i;j)}{\min\{H(i), H(j)\}}$$