Supportive Utility of Irrelevant Features in Data Preprocessing

Sam Chao, Yiping Li, and Mingchui Dong

Faculty of Science and Technology, Unversity of Macau
Av. Padre Tomás Pereira S.J., Taipa, Macao
{lidiasc,ypli,dmc}@umac.mo

Abstract. Many classification algorithms degrade their learning performance while irrelevant features are introduced. Feature selection is a process to choose an optimal subset of features and removes irrelevant ones. But many feature selection algorithms focus on filtering out the irrelevant attributes regarding the learned task only, not considering their hidden supportive information to other attributes: whether they are really irrelevant or potentially relevant? Since in medical domain, an irrelevant symptom is treated as the one providing neither explicit information nor supportive information for disease diagnosis. Therefore, the traditional feature selection methods may be unsuitable for handling such critical problem. In this paper, we propose a new method that selecting not only the relevant features, but also targeting at the latent useful irrelevant attributes by measuring their supportive importance to other attributes. The empirical results demonstrate a comparison of performance of various classification algorithms on twelve real-life datasets from UCI repository.

Keywords: supportive relevance, latent correlation, data preprocessing, feature selection, data mining.

1 Introduction

The objective of a classification problem is to accurately and efficiently map an input instance to an output class label, according to a set of labeled instances. While many aspects affect the classification performance, among all, data is a prominent one. More data no longer means more discriminative power; contrarily, they may increase the complexity and uncertainty to the learning algorithms, thus burden with heavy computational cost. On the other hand, less data may be either over-fitting or cause the learning algorithms unable to learn meaningful results. In order to learn efficiently, one of the data preprocessing algorithms – feature selection, which aims to optimize the data to be learned, can be involved to overcome such obstacles.

Various state-of-the-art feature selection algorithms are described in [1], as well as their evaluations and comparisons in [2], [3]. The existing feature selection methods are mainly divided into two categories: filter and wrapper. Filter approach evaluates the selected features independently, does not take the learning algorithm into the

Z.-H. Zhou, H. Li, and Q. Yang (Eds.): PAKDD 2007, LNAI 4426, pp. 425–432, 2007.

evaluation process. The advantage of this approach is its reasonable computational complexity and cost; while wrapper approach involves the learning algorithm as part of the evaluation function, for each subset a classifier is constructed and used for evaluating the goodness of generated subset. The advantage of this approach is its high and reliable classification accuracy. Furthermore, most feature selection methods used sequential forward/backward search [4], [5] to construct the best subset of features by starting with an empty set or a full feature set. Then, the search goes on by adding or deleting one more feature each time to or from the best feature subset, until no more performance improvement. In this paper, we adopted filter approach with sequential forward search in our method.

In the next sections, we describe our novel method LUIFS – Latent Utility of Irrelevant Feature Selection in detail, as well as the feature selection problem under medical domain. The evaluation of the proposed method on some real-life datasets is performed in the section following. Finally, we discuss the limitations of the method and present the directions for our further research.

2 Feature (Attributes) Selection

Feature selection is a process that chooses an optimal subset of features according to a certain criterion [1]. Features can be categorized into: relevant, redundant, and irrelevant. An irrelevant feature does not affect the target concept in any way; while a redundant feature does not add anything new to the target concept and a relevant feature is neither irrelevant nor redundant to the target concept [6].

2.1 Feature Selection Problem

The ordinary feature selection methods focus on selecting relevant attributes and filtering out the irrelevant ones regarding the class attribute (learned task) only. This may sometimes lose the significant supportive information hidden in the irrelevant features. For instance, a forward selection method recursively adds a feature x_i to the current optimal feature subset *OptimalA*, among those that have not been selected yet in feature set *A*, until a stop criterion is met. In each step, the feature x_i that makes evaluation measure *W* be greater is added to the subset *OptimalA*. Starting with *OptimalA* ={ }, the forward step is illustrated in equation (1).

$$\text{OptimalA} := \text{OptimalA} \cup \{A \backslash \text{OptimalA} \mid W(\text{OptimalA} \cup \{x_i\}) \text{ is maximum}\} . \quad (1)$$

The main disadvantage of the above formula is that it is impossible to have in consideration certain basic interactions among features, i.e., if x_1, x_2 are such interacted attributes, that $W(\{x_1, x_2\}) >> W(\{x_1\})$, $W(\{x_2\})$, neither x_1 and x_2 could be selected, in spite of being very useful [7]. This is because most feature selection methods assume that the attributes are independent rather than interactive, hence their hidden correlations have been ignored. However, an attribute that is completely useless by itself can provide a significant performance improvement when taken with others. Two attributes that are useless by themselves can be useful together [8], [9].