Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Feature selection using rough set-based direct dependency calculation by avoiding the positive region

Muhammad Summair Raza, Usman Qamar

Department of Computer Engineering, College of Electrical and Mechanical Engineering (E&ME), National University of Sciences and Technology (NUST), Pakistan

ARTICLE INFO

Article history:
Received 7 January 2017
Received in revised form 8 August 2017
Accepted 13 October 2017
Available online xxxx

Keywords:
Positive region
Rough set theory
Dependency rules
Feature selection
Reducts

ABSTRACT

Feature selection is the process of selecting a subset of features from the entire dataset such that the selected subset can be used on behalf of the entire dataset to reduce further processing. There are many approaches proposed for feature selection, and recently, rough set-based feature selection approaches have become dominant. The majority of such approaches use attribute dependency as criteria to determine the feature subsets. However, this measure uses the positive region to calculate dependency, which is a computationally expensive job, consequently effecting the performance of feature selection algorithms using this measure. In this paper, we have proposed a new heuristic-based dependency calculation method. The proposed method comprises a set of two rules called Direct Dependency Calculation (DDC) to calculate attribute dependency. Direct dependency calculates the number of unique/non-unique classes directly by using attribute values. Unique classes define accurate predictors of class, while non-unique classes are not accurate predictors. Calculating unique/non-unique classes in this manner lets us avoid the time-consuming calculation of the positive region, which helps increase the performance of subsequent algorithms. A two-dimensional grid was used as an intermediate data structure to calculate dependency. We have used the proposed method with a number of feature selection algorithms using various publically available datasets to justify the proposed method. A comparison framework was used for analysis purposes. Experimental results have shown the efficiency and effectiveness of the proposed method. It was determined that execution time was reduced by 63% for calculation of the dependency using DDCs, and a 65% decrease was observed in the case of feature selection algorithms based on DDCs. The required runtime memory was decreased by 95%.

1. Introduction

Feature selection is the process of selecting a subset of features from the entire dataset such that the selected subset can be presented on behalf of the entire dataset. Selecting feature subsets thus lets us reduce datasets to a manageable size by eliminating unnecessary and redundant information. The feature selection process thus provides a subset of features from the dataset that contains most of the useful information. The reduced set consequently reduces execution time for further tasks, thus enhancing the performance. This approach helps to reduce the number of noisy and irrelevant features. In the past two decades, the dimensionality of the datasets involved in machine learning and data mining applications has

E-mail addresses: summair.raza@ceme.nust.edu.pk (M.S. Raza), usmanq@ceme.nust.edu.pk (U. Qamar).

https://doi.org/10.1016/j.ijar.2017.10.012
0888-613X/© 2017 Elsevier Inc. All rights reserved.

increased explosively. There are two main approaches for such an increase in data dimensionality: transform-based reduction, also called attribute reduction, and selection-based reduction, also called feature selection. Transform-based reduction, as the name implies, transforms the underlying semantics of data. Selection-based reduction, i.e., feature selection, selects the features to represent the data instead of transforming the underlying semantics. Thus, the underlying semantics are preserved. Feature subset selection in these domains has helped to reduce the dimensionality of the feature space, improve the predictive accuracy of classification algorithms, and improve the visualization and comprehensibility of the induced concepts. Feature selection not only implies dimensionality reduction, i.e., reduction of the number of attributes that should be considered when building a model but also the choice of attributes, i.e., attributes can be selected or discarded based on criteria specifying their usefulness. Data in the real world may contain far more information than necessary, e.g., a database table may contain many attributes of a customer, out of which only a few are necessary to perform a certain type of analysis. Therefore, feature selection has become a necessary step to make the analysis more manageable and extract useful knowledge regarding a given domain [1]. It is very important in the analysis of high-dimensional data [2], where it serves several purposes, such as reducing the dimensionality of a dataset, decreasing the computational time required for classification and enhancing the classification accuracy of a classifier by removing redundant and misleading or erroneous features [3].

Various feature selection techniques have been proposed in the literature. These include correlation-based feature selection [4,5], mutual information-based feature selection [6,7], heterogeneous feature selection [8], consistency-based feature selection [9], the graph theoretic approach [10], ACO-based feature selection [11], feature selection in possibilistic modelling [12], and SVM-based feature selection [13].

Rough Set Theory (RST), proposed by Pawlak [14,15], is a mathematical tool for data analysis. RST-based approaches for attribute reduction [16–19] and feature selection [20–24] have become dominant. For feature selection, RST provides a positive region-based dependency measure called "attribute dependency" to perform feature selection. Attribute dependency determines how uniquely the value of an attribute determines the value of a dependent attribute. The value of attribute dependency ranges from zero (0) to one (1), where zero (0) means that an attribute does not depend on the other and one (1) means that an attribute fully depends on the other. However, this approach uses the positive region to calculate dependency, which is a time-consuming and complex step, adversely affecting the performance of feature selection algorithms using this measure and thus making them almost impossible to use for feature selection when datasets grow beyond smaller sizes. Rough set-based dependency requires three steps, i.e., calculation of equivalence classes using a decision attribute (decision class), calculation of equivalence classes using conditional attributes, and finally calculation of the positive region. Performing these tasks is a computationally expensive job and becomes inappropriate for datasets having large numbers of attributes or large numbers of examples. To overcome this issue, we require an alternate method to calculate dependency for which these computationally expensive steps are unnecessary. However, the accuracy of such an alternate approach should be exactly the same as that of the original positive region-based dependency measure so that it could be effectively applied to any of the feature selection algorithms without affecting their accuracy. This research proposes a new method called Direct Dependency Calculation (DDC), which directly calculates the dependency measure without performing the time-consuming positive region calculation. It directly scans the number of unique/non-unique classes in a dataset using attribute values and calculates dependency. Calculating dependency in this manner lets us avoid the positive region, which makes DDC-based feature selection algorithms suitable for average and larger datasets. The proposed approach is an alternative to the conventional positive region-based dependency measure and can be safely used in any of the feature selection algorithms using a rough set-based dependency measure.

The rest of the paper is organized as follows. Section 1 discusses preliminaries of rough set theory, and section 2 describes various related works. In section 3, DDC is discussed in detail. In section 4, various feature selection algorithms using DDC are presented. Finally, sections 5, 6 and 7 present experimental analysis.

2. Rough set theory preliminaries

Rough Set Theory (RST) was proposed by Pawlak [14,15]. Since its inception, it has been used in various domains for data analysis, including economics and finance [25], medical diagnosis [26], medical imaging [27], banking [28], and data mining [29].

2.1. Information system

An information system is the basic structure for representing the underlying information in RST. It comprises objects and their attributes. Formally, an information system $I = \{U, A\}$, where U is a non-empty finite set of objects representing the universe and A is the number of attributes, which are also called features. Every attribute has a value: $U \rightarrow V_a$, where V_a is called the value set of attribute "a".

Table 1 is an Information System (IS) where $A = \{Age, Income\}$ and $U = \{X_1, X_2, X_3, X_4, X_5, X_6, X_7\}$.

2.2. Decision system

A decision system is not only an information system; it also has decision attribute(s). A decision attribute, also called a "class", is a feature whose value depends on other attributes called conditional attributes. Formally, a decision system