Chapter 5

Feature Selection Using Neighborhood Positive Approximation Rough Set

Mohammad Atique
Sant Gadge Baba Amravati University, India

Leena Homraj Patil
Priyadarshini Institute of Engineering, India

ABSTRACT

Attribute reduction and feature selection is the main issue in rough set. Researchers have focused on several attribute reduction using rough set. However, the methods found are time consuming for large data sets. Since the key lies in reducing the attributes and selecting the relevant features, the main aim is to reduce the dimensionality of huge amount of data to get the smaller subset which can provide the useful information. Feature selection approach reduces the dimensionality of feature space and improves the overall performance. The challenge in feature selection is to deal with high dimensional. To overcome the issues and challenges, this chapter describes a feature selection based on the proposed neighborhood positive approximation approach and attributes reduction for data sets. This proposed system implements for attribute reduction and finds the relevant features. Evaluation shows that the proposed neighborhood positive approximation algorithm is effective and feasible for large data sets and also reduces the feature space.

INTRODUCTION

A huge amount of documents on internet is growing immensely. To manage such a huge volume of data, data mining yield an important skill. Rapidly growth in the document preprocessing tool, database increases rapidly in rows and in column. Feature selection, called as attribute reduction turn out to be difficult in the field of pattern recognition machine learning and data mining. Pawlak proposed a specific conceptual structure rough set model for feature selection (Liu & Yu, 2005). Hundreds of attributes are available in the database. A major issue lies executing attribute reduction in rough set theory. Rough set
theory is considered soft computing approach to examine data, and defines a structure which becomes difficult to recognize the task of relevant attributes (Liang, Wang, Dang, & Qian, 2012). It also declines the performance of machine learning problem. Features selection means sustaining the unfair influence of features. Hence, it becomes significant to decrease dimensions of the data to smaller size. It also reduces the appropriate information for declining the storage cost and time. Attribute reduction recommended a logical approach for consistency-based feature selection (Lingras & Peters, 2012). To overcome such a large issue of feature selection and attribute reduction, few attributes are eliminated.

To overcome the issue of feature selection and attribute reduction, few attributes can be omitted, which will not seriously effect on classification accuracy. Much work has been done on the subject of attribute reduction, both theoretical and experimental.

The main theme of this scheme is to propose an Neighborhood feature selection positive approximation algorithm (NFSPA) using rough set technique. This algorithm is based on the proposed neighborhood positive approximation approach. Most of the researchers have focused on several attribute reduction using Rough set. However, it has been found that none of the attribute reduction algorithm is the best. It has been observed that the several issues of attribute reduction can be overcome by suggesting the algorithm. Since the key lies in reducing the attributes and selecting the relevant features. For this purpose, we proposed and demonstrated neighborhood positive approximation approach. The idea behind using this approach is to optimize the number of feature and improves the computational time. It has been demonstrated that the proposed scheme outperforms the high dimensionality of data. Hence achieves the curse of dimensionality.

The proposed system implements for attribute reduction and finds the relevant features. An experimental result shows that the proposed neighborhood positive approximation provides the improvement in computational time and reduces the number of feature.

RELATED WORK

Feature selection, known as attribute reduction has turn out to be a significant step. In rough set theory, features selection aims to maintain the discriminatory power of original features. Hence, it is therefore important to reduce dimensionality of the data to smaller set of features and relevant information for decreasing the cost in storing and reduction in the processing time. To overcome the issue of feature selection and attribute reduction, few attributes can be omitted, which will not seriously effect on classification accuracy (Patil & Atique, 2015). Rough set theory handles data sets with imprecision and uncertain information. It utilizes a study of attribute reduction information system. It achieves concept approximation from the universe through which two defined subsets are produced. So far, many researchers have studied an attribute reduction algorithm.

Pawlak's Rough Set Model

Rough set was initially instigated by Pawlak as a efficient approach which deals with uncertainty. It works on decision analysis, knowledge discovery, and conflict analysis. The two-defined subset lower and upper are obtained through concept approximation. These two operators define an equivalence relation (Pawlak, 1991). Rough set utilizes the similarity to partition space data and create jointly equivalence class as the essential concepts. It’s applicable only to data with small attributes.