IDEA GROUP PUBLISHING
701 E Chocolate Avenue, Hershey PA17033, USA
Tel: 717/533-8845; Fax: 717/533-8661; URL: www.idea-group.com
ITB7251

Chapter II

A Heuristic Algorithm for Feature Selection Based on Optimization Techniques

A.M. Bagirov, A.M. Rubinov and J. Yearwood
University of Ballarat, Australia

The feature selection problem involves the selection of a subset of features that will be sufficient for the determination of structures or clusters in a given dataset and in making predictions. This chapter presents an algorithm for feature selection, which is based on the methods of optimization. To verify the effectiveness of the proposed algorithm we applied it to a number of publicly available real-world databases. The results of numerical experiments are presented and discussed. These results demonstrate that the algorithm performs well on the datasets considered.

INTRODUCTION

Feature extraction and selection is an important stage in the solution of pattern recognition problems. The feature selection problem involves the selection of a subset of features that will be sufficient in making predictions. There are various reasons for refining a feature set leading to the selection of suitable feature variables. Schaafsma (1982) gives a comprehensive description of reviews for feature selection. Among these are: the use of a minimal number of features to construct a simpler model, to give a simplest interpretation of such a model and to accelerate the decision making process. In some datasets the number of features and observations can reach several thousand and in such a situation the solution of classification problems without feature selection becomes fairly hard.

There exist various approaches to the solution of feature selection problems. We can note statistical (John, Kohavi & Pfleger, 1994; Kira & Rendell, 1992; Kittler, 1986; Koller & Sahami, 1996; McLachlan, 1992; Siedlecki & Sklansky, 1988), decision tree (Quinlan, 1993), neural network (Hecht, 1990; Hagan et al., 1996) and mathematical programming (Bradley, Mangasarian & Street, 1998; Bradley & Mangasarian, 1998; Chang, 1973; Fu, 1968; Hand, 1981) approaches among them. The book by Liu & Motoda (1998) gives a comprehensive description of statistical, decision tree and neural network approaches. In the papers by Bradley et al. (1998), and Bradley and Mangasarian (1998) the feature selection is formulated as a mathematical programming problem with a parametric objective function. Feature selection is achieved by generating a separating plane in the feature space. In the paper by Kudo and Sklansky (2000) a comparison of various feature selection algorithms is presented.

In this chapter, we suggest an algorithm for the solution of the feature selection problem based on techniques of convex programming. We consider feature selection in the context of the classification problem. In the above references, as a rule, datasets with two classes are considered. For statistical methods the selection problem with more than two classes is much more difficult than the problem with two classes where there is a close tie with multiple regression (McLachlan, 1992). The algorithm suggested in this paper allows one to consider datasets with an arbitrary number of classes.

The algorithm calculates a subset of most informative features and a smallest subset of features. The first subset provides the best description of a dataset whereas the second one provides the description which is very close to the best one. A subset of informative features is defined by using certain thresholds. The values of these thresholds depend on the objective of the task. Numerical experiments with several real-world databases have been carried out. We present their results and discuss them.

FEATURE SELECTION

The purpose of a feature selection procedure is to find as small a set as possible of informative features of the object under consideration, which describes this object from a certain point of view. The following issues are very important for understanding the problem.

1. It is convenient to consider (and define) informative features in the framework of classification. In other words it is possible to understand whether a certain feature is informative for a given example if we compare this example with another one from a different class. The following example confirms this observation.

Assume we consider a group A of people, who suffer from heart disease and recover in a hospital in a city E_1. We consider two features a_1 and a_2 of patients from this group. The feature a_1 describes a certain patient characteristic

Related Content

Principles for the Analysis of Large Complex Secondary Databases in Educational Settings

Yen Toand Hansel Burley (2012). *Cases on Institutional Research Systems (pp. 132-145).*

www.irma-international.org/chapter/principles-analysis-large-complex-secondary/60844

Introducing Data Mining and Knowledge Discovery

R. Sarker, H. Abbassand C. Newton (2002). *Heuristic and Optimization for Knowledge Discovery (pp. 1-12).*

www.irma-international.org/chapter/introducing-data-mining-knowledge-discovery/22146

Developing an Assessment Plan for the Arts and Sciences Division of a Comprehensive Community College

Annette E. Smithand Stephanie J. Jones (2012). *Cases on Institutional Research Systems (pp. 260-274).*

www.irma-international.org/chapter/developing-assessment-plan-arts-sciences/60853

Learning SKOS Relations for Terminological Ontologies from Text

Wei Wang, Payam M. Barnaghiand Andrzej Bargiela (2011). *Ontology Learning and Knowledge Discovery Using the Web: Challenges and Recent Advances (pp. 129-152).*

www.irma-international.org/chapter/learning-skos-relations-terminological-ontologies/53884

A Comparative Study of Associative Classifiers in Mesenchymal Stem Cell Differentiation Analysis

Weiqi Wang, Yanbo J. Wang, Qin Xin, René Bañares-Alcántara, Frans Coenenand Zhanfeng Cui (2011). *Knowledge Discovery Practices and Emerging Applications of Data Mining: Trends and New Domains (pp. 223-243).*

www.irma-international.org/chapter/comparative-study-associative-classifiers-mesenchymal/46898