

Data mining application with case based reasoning classifier for breast cancer decision support

Abdeldjalil Khelassi

Computer Science Department

Abou Bakr Belkaied University, Temcen, Algeria

Tel: 00213550168666

Khelassi.a@gmail.com

ABSTRACT

Cytology is a complex diagnosis task which requires both expertise and experience of an oncologist for providing the cancer class and stage which is very useful in the therapy and in the surgery intervention. A case based reasoning classifier is developed with specialized agents for recognizing the malignant breast cancer. The proposed application implements a data mining method for the knowledge extraction and discovery by mining a medical database, which contains classified instances characterized by some features extracted automatically from the cytological image of the patient cancer. An original technique is implemented for enriching the retrieving process on the developed CBR system; this technique is based on the combination of global-local similarity measures and fuzzy sets for modeling the unknown response generated from the agents which increase significantly the accuracy of the system. The features selection and weighting is done by a machine learning algorithm. The efficiency of the proposed methodology has been validated through some empirical experiments applied in the cited data set which demonstrates that the developed approach achieves such average accuracies better than the current state-of-the-art approaches.

Keywords: Case Based Reasoning; data mining, Fuzzy sets; Breast cancer; Cytology.

I. INTRODUCTION

In recent years data mining has become a very popular technique for extracting information from the database in different areas due to its flexibility of working on any kind of databases and also due to the surprising results.[5] Of the data mining techniques developed recently, several major kinds of data mining methods, including generalization, characterization, classification, clustering, association, evolution, pattern matching, data visualization and meta-rule guided mining, are cited

in [2]. Data classification is an important topic in the field of data mining due to its wide applications. A number of related methods have been proposed based on the well-known learning models such as decision tree or neural network [3]. Standard commercial tools such as SPSS, SAS, Clementine or even freely available tools such as WEKA implement diverse mining techniques and algorithms, some of which are very sophisticated. Given the availability and the reliability of these tools, they are preferred by practitioners and researchers over proprietary self-development tools [4]. Also some of these techniques are implemented in the sophisticated database systems as IBM DB2, Microsoft SQL Server, MYSQL, and ORACLE.[5]

The case based reasoning approach is widely and successfully applied in many domains as games, recommendation systems, information retrieval, bioinformatics, industrial applications and others. It represents a good and easy method of knowledge extraction, discovery and modelling. It consists of using the prior similar cases for resolving the newest problems.[7]

The data mining and knowledge discovery techniques are widely used to diagnose human disease. [6] In this survey a data mining application is developed by applying a case based reasoning classifier, and machine learning algorithms to determine the class of breast cancer from a pattern extracted from the image of cancerous cellular tissue see Figure1. We have also proposed a modified similarity measures, by combining the traditional similarity functions with the fuzzy sets, for ensuring a flexible and accurate model.

II. AROUND THE PROBLEM

The problem treated in this work focus on the recognition of malignant cases of breast cancer by using the extracted knowledge from a data set constructed from some sample of breast cancer cytological image. In this section we will present the needed fundament around this problem. First of