URN etd-0407109-183717-1411
Author Yu-Hsien Wang
Author's Email Address No Public.
Statistics This thesis had been viewed 44 times. Download 0 times.
Department
Year 2006
Semester 1
Degree Master
Type of Document Master's Thesis
Language zh-TW.Big5 Chinese
Title Medical Diagnosis Knowledge Extraction using Feature Selection:An Example of Coronary Heart Disease Diagnosis Knowledge Extraction
Date of Defense 0000-00-00
Page Count 42

Keyword
knowledge extraction
feature selection
Induction decision tree

Abstract
Because of the innovation and development of computer technology, researchers attempt to build an expert system to help human to solve the problems like an expert. However, it is difficult to unify and to organize the experts' knowledge to store the knowledge in an information system. In the current medical environment, physicians are asked to increase the quality of medical care while reducing the medical cost. In order to help physicians to make good diagnosis efficiently based on patients' history and lab examination, this study evaluates feature selection methods, including feature selection by expertise and automatic feature selection, for the diagnosis of coronary artery heard disease. The evaluation results may help the medical knowledge extraction from historical examples.
In this study, we employ a coronary artery heard disease dataset in the UCI repository as for our evaluation. This dataset includes 920 patient records. We invited three cardiac physicians to participate in this study for identifying important diagnosis factors about the coronary artery heard disease. We first used Delphi-like method to identify the important diagnosis factors with helps of the three physicians. Then, we also used an automatic feature selection mechanism, CFS, to select important factors. Finally, we built the C4.5 decision-tree models based the feature sets.
Base on performance evaluation results of C4.5 models, we find that we can first employ the automatic feature selection method to quickly select important features. Then, we revise the feature set according to the experts' opinions. The C4.5 models based on the revised feature set will have good disease classification performance. At the same time, the readability and explanation of the prognosis models will also increase. This approach also simplifies the diagnosis knowledge extraction work.
Advisory Committee Tsang-Hsiang Cheng - advisor

Files No Any Full Text File.
Date of Submission 2007-08-21