

Editorial

Bernard Chen
University of Central Arkansas, USA

Data Mining is a technique to find useful and uncommon information from large amount of data. To make mined information meaningful, the domain knowledges about the data is always the key. This issue of the journal includes two studies of adopting specific domain knowledge in classification algorithms to build explainable models: one paper focuses on medical images analysis while the other paper works on manufacturing rare quality event detection. Both paper emphases on one particular challenge in data mining: feature selection. How to find the most relevant features to explain the model and extract information for the model is well studied in this issue.

Picture Archiving and Communication Systems (PACS) were developed to provide a convenient system to analysis, exchange and store medical images. The first paper use data mining on PACS to support medical image diagnosis. Decision tree with pruning techniques and syndrome-like decision rules are applied in the task of early differential diagnosis of pulmonary nodules in lung tomograms and allowed to support image analysis and classification by the expert. The key to the success of this research lies in the meticulously collaboration with medical experts since all features selected by the classification model are carefully evaluated. The good results together with effective decision models are produced. The different models based on decision trees and syndrome-like rules get compared and evaluated. The decision tree model outperforms the syndrome-like rule model when the data are not very noisy while the syndrome-like rule model is better when the data are more uncertain. These models are able decomposing a complex medical image analysis task into several simpler subtasks. Therefore, they are useful for both diagnose and education purposes.

Data about ultrasonic metal welding of battery tabs for the Chevrolet Volt is studied in the second paper. Since the welding process is a very stable procedure, only a few defects per million operations, the data used in this work is extremely imbalanced. In order to successfully and accurately identify the defects, feature selection plays an important role in the data mining process. A Hybrid Feature Selection and Pattern Recognition framework aimed at rare quality event detection is developed in this research work. All true negative instances are identified with only a few false negative instances are reported in the result. Authors provide a good quality solution to detect rare quality event.

Both papers are closely related to our daily life, the readers will find here interesting ideas.