

INFORMATION ON NEW ACADEMIC AND THEORETICAL CONTRIBUTIONS OF THE THESIS

Thesis's Title: Attribute reduction method and data clustering technique using rough set theory

PhD Candidate: Do Si Truong

Code: 915000393

Course: 2015

Major: Computer Science

Mã số ngành: 9480101

University: Lac Hong University

Science Instructor: Associate Professor Nguyen Thanh Tung

The contributions of thesis on science and practice, including:

Scientifically:

The research objective of this thesis focuses on two problems within the topic: (1) propose new attribute reduction methods in a decision table; (2) discovering new algorithms for clustering categorical data using rough sets theory.

By researching the algorithms proposed by researchers and finding shortcomings, the thesis has proposed a new algorithm for selecting attributes in a decision table based on clustering. ACBRC obtained the best proportion of selected attributes, and the best classification accuracy for C5.0, and Naive Bayes. The classification accuracy after attribute reduction by ACBRC even outperforms the classification accuracy using whole dataset in some cases. These experimental results show that ACBRC is promising algorithm for attribute reduction.

By researching the algorithms proposed by researchers and finding shortcomings, the thesis has proposed a new algorithm, called Minimum Mean Normalized Variation of Information (MMNVI). Experimental results on real datasets from UCI indicate that MMNVI algorithm can be used successfully in clustering categorical data. It produces better or equivalent clustering results than the baseline algorithms.

The main contributions mentioned above were published in two articles in the Journal of Computer Science and Cybernetics. In addition, I'm also the author of an international article and co-author of three reports at prestigious domestic scientific conferences related to the thesis topic.

Practically:

Knowledge Discovery in Databases (KDD) is a scientific field that researches and creates tools for discovering useful, hidden, and predictive knowledge within large databases. Research results and successful applications in recent times show that Knowledge Discovery in Databases is a potential scientific field, bringing many benefits and having numerous advantages compared to traditional data analysis tools. However, given the current rate of data growth, the research and application of data mining techniques are encountering numerous difficulties and challenges. This necessitates researchers to consistently strive in finding new tools to overcome these issues.

Rough set theory is a relatively new soft computing tool for dealing with uncertain data. Therefore, the research and contributions of the thesis in this field can be successfully applied in knowledge discovery in databases, as well as in machine learning, expert systems, and pattern recognition, and can be widely applied in practice.

Dong Nai, December 11th, 2023

SCIENCE INSTRUCTOR

PHD CANDIDATE

Assoc. Prof. NGUYEN THANH TUNG

DO SI TRUONG