VIETNAM NATIONAL UNIVERSITY, HANOI
UNIVERSITY OF ENGINEERING AND TECHNOLOGY

Le Tuan Anh

OPTIMIZE CNF ENCODING FOR ITEMSET MINING TASKS

Major: Computer Science

HA NOI - 2024

VIETNAM NATIONAL UNIVERSITY, HANOI
UNIVERSITY OF ENGINEERING AND TECHNOLOGY

Le Tuan Anh

OPTIMIZE CNF ENCODING FOR ITEMSET MINING TASKS

Major: Computer Science

Supervisor: Dr. To Van Khanh

HA NOI - 2024

ABSTRACT

Summary: In this thesis, we apply the ”Sequential Encounter Encoding” method to optimize itemset mining tasks. This method has been proven effective in optimizing the search process for itemsets in data. By utilizing ”Sequential Encounter Encoding” we aim to enhance the performance of itemset mining algorithms while minimizing processing time.

We conduct a series of experiments on real-world datasets to evaluate the performance of the applied method. Experimental results demonstrate a significant improvement in accuracy and efficiency compared to traditional methods. The application of this method not only enhances the performance of data mining processes but also opens up potential applications for similar problems in the field of data science and information retrieval.

Keywords: *SAT, SAT Encoding, Sequential Encounter Encoding, Itemset Mining Tasks*

ACKNOWLEDGEMENT

First of all, I would like to express my deepest gratitude to Lecturer, Dr. To Van Khanh, who wholeheartedly guided, guided, encouraged, and helped me throughout the course of this thesis.

I would like to thank the teachers in the Faculty of Information Technology as well as the University of Engineering and Technology - Vietnam National University, Hanoi, for creating conditions for me to study in a good environment and for teaching me the best things to do. Knowledge is especially important for me to continue to study and work in the future.

I especially express my gratitude to my family who have always been a solid support to help and support me in every journey. Finally, I would also like to thank the members of the K65CA-CLC2 class, my friends in the course, and the siblings inside and outside the school who have been with me to study and practice and help each other throughout four years university.

Thank you sincerely!

Le Tuan Anh

AUTHORSHIP

I hereby declare that the thesis "SEQUENTIAL ENCOUNTER ENCODING FOR ITEMSET MINING TASKS" is done by me and has never been submitted as a report for Graduation Thesis at University of Engineering and Technology - Vietnam National University, Hanoi, or any other university. All content in this thesis is written by me and has not been copied from any source, nor is the work of others used without specific citation. I also warrant that the source code is my development and does not copy the source code of any other person. If wrong, I would like to take full responsibility according to the regulations of University of Engineering and Technology - Vietnam National University, Hanoi.

Ha Noi, May 26 2024

Student

Le Tuan Anh

SUPERVISOR'S APPROVAL

I hereby approve that the thesis in its current form is ready for committee examination as a requirement for the Bachelor of Computer Science degree at the University of Engineering and Technology.

Ha Noi, May 26 2024

Supervisor

Dr. To Van Khanh

Contents

List of Figures

CHAPTER 1

Introduction

This chapter will focuses on introducing the itemset mining tasks and SAT encoding, encompassing their concepts, related terms, and applications.

1.1 Itemset Mining Tasks

1.1.1 Overview

1.1.2 Technical background

1.2 SAT Encoding

1.2.1 Concept

1.2.2 SAT Solvers

1.2.3 Applications

CHAPTER 2

SAT-based Encoding of Itemset Mining

In this chapter, we present how to encode the itemset mining problem into a SAT problem. And we will discuss the limitation of the standard method.

2.1 Constraint Encoding

2.2 Standard Method in Itemset Mining

2.3 Limitation of Standard Method

CHAPTER 3

Sequential Encounter Encoding for Optimization

3.1 Sequential Encounter Encoding

3.2 Optimization CNF Encoding using Sequential Encounter Encoding

CHAPTER 4

Experiments

4.1 Experimental Setup and Datasets

4.2 Results and Analysis

Conclusions