Universiti Teknologi MARA

DEVELOPMENT OF FREQUENT PATTERN MINING (FPM) ALGORITHMS IN HIGH DIMENSIONAL DATASET WITH HEURISTIC SEARCHING TECHNIQUE

MUHAMMAD NORZARIMAN BIN RAZARI

Thesis submitted in fulfillment of the requirements for
Bachelor of Science (Hons) Intelligent System
Faculty of Computer and Mathematical Sciences

August 2014

Acknowledgement

Alhamdulillah, praise and thank to Allah because of His Almighty and His utmost blessings, I was able to finish this research within time duration given. Firstly, my special thank go to my supervisor, Dr. Shuzlina Binti Abdul Rahman and my co-supervisor, Pn Sofianita Mutalib.

Special appreciation also goes to my beloved parents Razari Bin Osman and

Last but not least, I would like to give my gratitude to my all my friends.

ABSTRACT

This project is about development of frequent pattern mining (FPM) with heuristic search technique. Basically, there are many algorithms that have been used in high dimensional dataset. Moreover, high dimensional dataset needs good algorithm to produce frequent patterns but it also requires lot of memory. This project employed RERII as FPM technique and best-first search as heuristic search technique. RERII with original searching technique and RERII that implement with heuristic search were compared. The result shows the RERII with original searching technique is more suitable and better than RERII that using heuristic search. There are several reason make the RERII with heuristic search is slower. One of the reason is the current RERII used DFS does not need an operation initialize weight and heuristic value. The other reason is heuristic search need a process to sort the node or item according the priority.

TABLE OF CONTENTS

CONTENT		PAGE
SUPERVISOR'S APPROVAL		ii
DECLARATION		iii
ACKNOWLEDGEMENT		iv
ABSTRACT		v
TABLE OF CONTENTS		vi
LIST OF FIGURES		ix
LIST OF ABBREVIATIONS		x
CHAPTER ONE: INTRODUCTION		
1.1	Background Study	1
1.2	Research Background	1
1.3	Problem Statement	2
1.4	Research Question	2
1.5	Objectives	3
1.6	Project Scope	3
1.7	Research Significance	3
CHAPTER TWO: LITERATURE REVIEW		
2.1	Association Rule Mining	4
2.2	Application of Association Rule Mining	6
2.3	Market Basket Analysis	6
2.4	Medical Diagnosis	7
2.5	Customer Relationship Management	9
2.6	Search Algorithm	10
2.7	Blind Search	11
2.8	Heuristic Search	11
2.9	Depth-First Search	11
2.10	Best-First Search	12

CHAPTER 1
INTRODUCTION

This chapter provides the background and rationale for the study. It also gives details on the significance of artificial intelligence in high dimensional dataset, the issues and problem that lead to this project.

1.1 Background of study

Nowadays, technology is growing very fast until it is hard for us to catch it. Every work or task that we do will involve a computer. Some of the work or task maybe will use intelligence system to complete it. This is including the frequent pattern mining (FPM) algorithms to produce frequent patterns for high dimensional dataset. According to Cong et al, high dimensional dataset will give a more complexity for existing FPM algorithms(2004). From that, there are lots of algorithms that have been created and improved in purpose to deal with this dataset. However, almost all algorithms use blind search technique. With that, there a need to seek and implement other search technique

1.2 Research background

According to Han et al and Han et al, frequent patterns are sequences or substructures of data items that frequently occur in a database no less than a user-specified threshold (2007; 2011). Data items represented as digit, characters, strings, label, records, and so forth (Shi *et al.*, 2011). This project will be focused on one of the common process in FPM algorithm and that is searching.

Searching in FPM algorithms is the process to find the frequent pattern base on minimum support in database. In artificial intelligence, search is main problem-solving mechanism. Furthermore, there has been classified three problem of searching in artificial intelligence, single-agent path finding, two-player, and constraint-satisfaction problems (Korf, 2010). Blind search (also known as uninformed algorithm) that categorize as single-agent path finding is