



FINAL YEAR DISSERTATION

Market Basket Analysis with Graph Theory

April 6, 2021

Sahil M. Pattni
Bachelor of Science with Honours in Computer Science
Supervised by Dr. Neamat El Gayar

Declaration

I, Sahil Manojkumar Pattni, confirm that this work submitted for assessment is my own and is expressed in my own words. Any uses made within it of the works of other authors in any form (e.g., ideas, equations, figures, text, tables, programs) are properly acknowledged at any point of their use. A list of the references employed is included.

Date: April 6, 2021

Signed: Sahil Manojkumar Pattni

Abstract

In this digital age, data is being generated and collected at an unprecedented rate, with data analytics employed by corporations and small businesses alike to produce actionable insights, reduce costs, optimize operations and increase revenue. Association rules allow us to identify relationships between products that can provide insights into customer spending habits and product perception.

In this study, a minimum spanning tree (MST) will be generated from a transactional database such that only the strongest relationships between products remain. A clustering algorithm will be applied to this MST to identify high co-purchase segments, and association rules will then be extracted from these segments. **TODO: ADD MORE**

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Aims	1
1.3	Objectives	1
2	Background	1
2.1	Related Work	1
2.1.1	Fast Algorithms for Mining Association Rules	1
2.1.2	Market basket analysis: Complementing association rules with minimum spanning trees	1
3	Data [Creation/Organization/Blank]	1
4	MST Generation	1
5	Testing and Evaluation	1
5.1	Metrics	1
5.2	Data Filtering	1
	References	2

1 Introduction

1.1 Motivation

For businesses such as groceries, hypermarkets, and retail outlets that deal with the trade of heterogeneous physical assets, operations such as inventory management and product placement play an instrumental role in determining the business' financial success. These involve asking questions such as:

- Which products should be placed at the entrance of the store? Which should be placed closer to the exit?
- Which products will benefit the most by being placed at eye-level?
- Which products should be placed next to each other to maximize the purchase volume?

One way to find optimal solutions to such queries is to employ the use of Association Rule Mining (also known as Market Basket Analysis). This set of techniques assess frequent itemsets (e.g. from sales data) and generate association rules between products. Several algorithms exist for association rule mining, most prominently the Apriori Algorithm [1] and FP-Growth [2], however these algorithms tend to generate an overwhelming amount of rules, rendering it inconvenient for the end-user to extract actionable information from the ruleset. A method that could produce a smaller, albeit more meaningful ruleset could prove to be more useful to such businesses and those who manage them.

1.2 Aims

The aim of this paper is to improve upon an existing method (see: Section 2.1.2) that derives association rules from minimum spanning trees, and to determine whether it can be considered a viable alternative or complement to established method: the Apriori Algorithm.

1.3 Objectives

2 Background

2.1 Related Work

2.1.1 Fast Algorithms for Mining Association Rules

2.1.2 Market basket analysis: Complementing association rules with minimum spanning trees

3 Data [Creation/Organization/Blank]

4 MST Generation

5 Testing and Evaluation

5.1 Metrics

5.2 Data Filtering

Talk about how MST different when comparing different cities.

References

- [1] Rakesh Agrawal and Ramakrishnan Srikant. “Fast Algorithms for Mining Association Rules”. In: *Proceedings of the 20th International Conference on Very Large Databases* (1994), pp. 487–489.
- [2] Jiawei Han, Jian Pei, and Yiwen Yin. “Mining Frequent Patterns without Candidate Generation”. In: vol. 29. 2. New York, NY, USA: Association for Computing Machinery, May 2000, pp. 1–12. DOI: 10.1145/335191.335372. URL: <https://doi.org/10.1145/335191.335372>.