Final Year Dissertation

Association Rules using Minimum Spanning Trees and Markov Clustering

April 5, 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 5, 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 Motivation

1.2 Aims and Objectives

2 Background

2.1 Apriori Algorithm

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.