**Report: Comparation between Apriori Algorithm and Brute-force**

IN6227-2024-Assignment-2

G2401631D

Liang Hou

**Introduce**

In this report, I used the Orange Python library as my association rule mining tool to implement the Apriori algorithm. Additionally, I wrote a brute\_force\_association\_rule\_mining function to implement a brute-force algorithm and compared their runtimes. I downloaded five datasets from Kaggle: [Market Basket Analysis](https://www.kaggle.com/datasets/aslanahmedov/market-basket-analysis) (Dataset 1), [Groceries dataset](https://www.kaggle.com/datasets/heeraldedhia/groceries-dataset) (Dataset 2), [E-commerce Business Transaction](https://www.kaggle.com/datasets/gabrielramos87/an-online-shop-business) (Dataset 3), [Transaction Data](https://www.kaggle.com/datasets/vipin20/transaction-data) (Dataset 4), and the [Sales transaction dataset of Google Merchandise Store in a Binary Matrix](https://www.kaggle.com/datasets/dinosilooy/sales-transaction-of-an-online-retailer) (Dataset 5).

The results are presented in the following tables. Fig. 1 show the results for 200 items, while Fig. 2 show the results for 15 items. The minimum support and confidence values used were [0.1, 0.5] and [0.4, 0.7], respectively.

|  |  |
| --- | --- |
| Fig.1 | Fig.2 |

**Explanation of N/A:** The brute-force algorithm examines every possible combination of items, so the time required grows exponentially as the item count increases. With an item count of 200, the computation becomes too large to complete within the limits of available computing resources, making the exact runtime unknown (N/A).

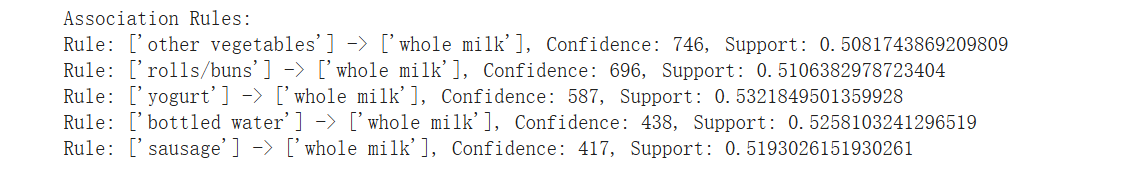
**Compared Results between Aprior and Brute-force**

A simple sample of mining results when implementing Aprior and Brute-force in Dataset 2 under min\_support = 0.1 and min\_confidence = 0.5.

|  |  |
| --- | --- |
| AR | Brute-force |

**Mining Result Sample**

And one of the mining results of Dataset 2 when min\_support = 0.1 and min\_confidence = 0.5.



**Code**

I wrote the code in a Jupyter Notebook file and ran my program in Google Colab. You can find the code on GitHub at [GitHub - RiverLiangH/Apriori-vs-BruteForce: Assignment-2 in IN6227: Comparation between Apriori & Br](https://github.com/RiverLiangH/Apriori-vs-BruteForce). The dataset is available on Google Drive at [drive.google.com](https://drive.google.com/drive/folders/1snZ6yQ349qZ2cb9asEhi5P5u7FnVdzdy?usp=drive_link).