

# Groceries に関する R 言語を使用した頻出パターン抽出及び相関ルール分析

文理学部情報科学科

5419045 高林 秀

2021 年 7 月 23 日

## 概要

本稿では、今年度データ科学 2 で学習した「頻出パターン抽出」及び「相関ルール分析」の手法を使用して、R 言語のライブラリである `arules` に付属しているデータ `Groceries` を対象とした頻出パターン抽出、相関ルール分析を行うものである。

## 1 目的

本稿では実際に、R 言語を使用しライブラリ `arules` 付属のデータである `Groceries` の頻出パターン抽出、相関ルール分析を行うことで、本年度データ科学 2 で学習した頻出パターン、相関ルール分析の手法への理解を深め、その定着を図ることを目的とする。また、1 年次に学習した `latex` を用いた PDF 作成の復習も兼ねるものである。

## 2 理論説明

今回の実験で用いた、計算理論をそれぞれ説明する。

- 2.1 バスケット分析
- 2.2 頻出パターン・相関ルール分析の概要
- 2.3 頻出パターンの計算法
- 2.4 支持度
- 2.5 確信度
- 2.6 パターン空間について
- 2.7 （補足）深さ優先探索・幅優先探索
- 2.8 バックトラック法
- 2.9 アプリオリアルゴリズム
- 2.10 相関ルール抽出の計算法
- 2.11 頻出パターン抽出の問題点
- 2.12 相関ルール分析の評価基準

## 3 計算機実験

### 3.1 実験準備

#### 3.1.1 実験環境

今回の実験は仮想マシン上で R 言語を起動し行った。下記に実験時の環境を示す。

- ホスト OS : Window10 Home Ver.20H2
- 仮想 OS : Ubuntu 20.04.2 LTS
- CPU : Intel(R)Core(TM)i7-9700K @ 3.6GHz
- GPU : Nvidia Geforce RTX2070 OC @ 8GB
- ホスト RAM : 16GB
- 仮想 RAM : 4GB

3.1.2 実験データ

3.1.3 R 言語での頻出パターン・相関ルール分析の手法

3.2 実験結果

3.3 結果の説明

4 考察

5 まとめ

参考文献