

相関ルールマイニングの概要

大阪大学 工学部 電子情報学科 3 年

情報システム工学コース

08D23091 辻 孝弥

2025 年 4 月 29 日

1 はじめに

相関ルールマイニング (Association Rule Mining) は, 大量の取引データ (トランザクション) から「あるアイテムの組み合わせが同時に出現しやすい」というパターンを発見する手法である. 小売業におけるバスケット分析に端を発し, マーケティング施策やレコメンデーションの基盤として広く用いられている.

2 基本概念

2.1 トランザクションとアイテム集合

トランザクション: 顧客の購買履歴など, アイテムの集合を表す (例: {ビール, おむつ, パン}).

アイテム集合 (Itemset): 複数のアイテムをまとめた集合. サイズ k のアイテム集合を k -アイテム集合と呼ぶ.

2.2 指標定義

支持度 (Support): 項目集合 X が出現する頻度を表す.

$$P(X) = \frac{|\{T \mid X \subseteq T\}|}{\text{全トランザクション数}}$$

信頼度 (Confidence): X を含むトランザクションのうち Y も含む割合.

$$P(Y|X) = \frac{P(X \cup Y)}{P(X)}$$

リフト (Lift): 期待値以上に同時出現する強さ. 1 以上で正の相関.

$$\frac{P(Y|X)}{P(Y)}$$

3 アルゴリズム概要

3.1 Apriori アルゴリズム

1. L1: すべての 1-アイテム集合の支持度を計算し、最小支持度 (minsup) 以上の集合を抽出.
2. 候補生成 (Ck): 頻出 (k-1)-アイテム集合同士を結合し、k-アイテム集合候補を作成.
3. プルーニング: 候補の部分集合がすべて頻出でないものは除去.
4. 支持度計算: トランザクションを再スキャンし、Ck の支持度を数えて Lk を得る.
5. k を増やし、Lk が空になるまで繰り返す.

3.2 ルール生成

各頻出アイテム集合 ℓ をその非空部分集合 s と $\ell - s$ に分割し、ルール候補 $s \Rightarrow (\ell - s)$ を作成. 信頼度を計算し、最小信頼度 (minconf) 以上のルールを採択.

4 応用例と利点・課題

4.1 応用例

- 小売業のバスケット分析による陳列最適化
- Web ログ解析でのページ遷移パターン抽出
- レコメンデーションシステムの基盤

4.2 利点

- 手軽に意思決定ルールを発見できる
- 可視化しやすくビジネス部門との連携に適す

4.3 課題

- ルール数の爆発的増加に対処が必要
- 当たり前のルール (例: パン \Rightarrow 牛乳) が大量に出力される
- 意外性の高いルールを見つけるにはリフトなどの指標でフィルタリングが必須

5 おわりに

相関ルールマイニングは、シンプルな指標と反復的探索によりデータ内の潜在的パターンを発見する有力手法である. 一方、候補数の爆発や自明なルールの多発といった課題もあるため、その他の手法と組み合わせることで、実用的な知見を抽出できる.

参考文献

- [1] DM08-04: Association Rules, Keio University (2008)
<https://web.sfc.keio.ac.jp/~maunz/DM08/DM08-04.pdf>