

非階層クラスタリングを用いた市場シェア率とリピート率による マーケティング分析

Marketing Analysis by Market Share Rate and Repeat Rate using Non-Hierarchical Clustering

2041082 千野根 琉成

Ryusei CHINONE

指導教員 秋葉 知昭

In recent years, the value of segmentation in marketing has been increasing due to the diversification of customer needs. In addition, with the spread of big data, the use of machine learning in the field of marketing, such as forecasting and clustering, has been attracting attention.

In this study, I conduct segmentation of products using non-hierarchical clustering based on the market share and repeat rate. First, I collect data on market share and repeaters, and conduct six non-hierarchical clustering methods: k -means, x -means, GMM, Nonparametric Bayesian, DBSCAN, and Mean-Shift. Next, the silhouette score is computed using silhouette analysis. The clustering model with the highest silhouette score is selected as the optimal model. Finally, the results are interpreted. The optimal model is characterized for each cluster and each market.

The clustering results are analyzed for each cluster and each market, and the characteristics are clarified.

1. 緒言

マーケティングという言葉は AMA やコトラーなどによって様々な定義がされているが、共通していえることは「モノやサービスが売れる仕組みづくり」ということである。

コトラー[1]によるとマーケティングは時代とともに進化しており、マーケティングの誕生から現在までをマーケティング 1.0 からマーケティング 4.0 までの 4 つの段階に分けられると述べている。近代になるにつれ、マーケティングの中心は製品から人へと変化した。そして顧客一人一人の多様化するニーズに対応するため、現在ではワントゥワンマーケティングに焦点が置かれている。

このような時代背景もあり、近年では顧客の共通性を抜き出し、グループ化するセグメンテーションの重要性が高まっている。またビックデータの取り扱いが容易になったことで機械学習の活用にも注目が集まっている。

本研究では非階層クラスタリングを用いて市場シェア率とリピート率によるセグメンテーションを行い、クラスターおよび市場の特徴を明らかにするマーケティング分析を行う。

2. データ収集と非階層クラスタリングの実装

2.1 データ収集

本研究では株式会社 True Date が提供するウレコン[2]から 2023 年 7 月から 9 月における各商品の「市場シェア率」、「リピート率」を利用する。また各カテゴリは「加工食品」、「菓子類」、「飲

料・酒類」、「その他食品」、「日用雑貨」、「OTC 医薬品」、「化粧品」、「家庭用品」、「ペット用品」の 9 つの中分類に分類される。

2.2 非階層クラスタリングの実装

本研究では 6 つの非階層クラスタリングの手法を実装する。実行環境は Jupyter Notebook, プログラム言語は Python を使用している。各手法の説明は以下のとおりである。

① k -means 法

重心を基準としたクラスタリングであり、最も代表的な手法である。

② x -means 法

k -means 法の拡張版。 k -means 法とベイズ情報量規準を用いることで、クラスター数を自動決定し、クラスタリングする手法である。

③ 混合ガウスモデル

確率分布を基準としたクラスタリング手法である。データが複数の正規分布から生成されたと仮定し、クラスタリングする手法である。

④ ノンパラメトリックベイズ

無限次元の確率分布を基準としたクラスタリング。クラスター構造をベイズ推定によってモデル化し、クラスタリングする手法である。

⑤ DBSCAN

密度を基準としたクラスタリング。データポイントからの半径内にある他のデータポイントの数によってクラスタリングする手法である。

⑥ Mean-shift

密度を基準としたクラスタリングであり k -means 法の有限版といえる手法である。

⑦エルボー法

エルボー法はクラスタの数を決定する基準を示す手法である。クラスタの重心からデータポイントまでの距離の総和であるクラスタ内誤差平方和を求めることでクラスタ数を決定する。

3. 最適モデルの選択

3.1 シルエット分析

2.2 節で実装した 6 つの非階層クラスタリングについて、シルエット分析を適用することで最も適切にクラスタリングできているモデルを判断し、選択する。シルエット分析とはクラスタリングしたモデルの性能を定量的に評価する手法である。シルエット分析では凝集度と乖離度からシルエットスコアを算出し、その値が 1 に近づくほどモデルの性能が良いと判断する[3]。なおシルエットスコアは $[-1,1]$ の区間に収まる。

3.2 シルエットスコアの算出

本研究で求めたシルエットスコアを表 1 に示す。この結果から「加工食品」，「飲料・酒類」，「その他食品」，「日用雑貨」，「OTC 医薬品」については k -means 法，「菓子類」，「ペット用品」については x -means 法，「化粧品」，「家庭用品」については GMM が最適モデルとなる。

表 1 シルエットスコアの算出結果

	中分類/手法	k-means法	x-means法	混合ガウスモデル (GMM)	DBSCAN	Mean-shift	ノンパラメトリック ベイズ
食品	加工食品	0.9502 (n=4)	0.9359	0.9390 (n=4)	0.9117 (eps=0.01-0.50) (min_points=1-10)	0.9117 (bandwidth=0.01-0.76)	0.9117 (gamma=1.5)
	菓子類	0.8856 (n=4)	0.9074	0.8864 (n=4)	0.8845 (eps=0.01-0.75) (min_points=1-7)	0.8845 (bandwidth=0.01-0.75)	0.8845 (gamma=0.4)
	飲料・酒類	0.9664 (n=4)	0.9534	0.9473 (n=4)	0.5222 (eps=0.01-0.50) (min_points=1-9)	0.4689 (bandwidth=0.01-0.07)	0.2086 (gamma=1.0)
	その他食品	0.8189 (n=3)	0.6576	0.8183 (n=4)	0.6996 (eps=0.01-0.27) (min_points=2)	0.6996 (bandwidth=0.01-0.27)	0.6354 (gamma=1.0)
日用品	日用雑貨	0.9464 (n=5)	0.9152	0.9423 (n=5)	0.9035 (eps=0.01-0.50) (min_points=1-18)	0.9035 (bandwidth=0.01-0.59)	0.4235 (gamma=1.0)
	OTC医薬品	0.9400 (n=5)	0.9018	0.9306 (n=5)	0.8254 (eps=0.01-0.49) (min_points=1-16)	0.8254 (bandwidth=0.01-0.50)	0.4528 (gamma=1.0)
	化粧品	0.9608 (n=5)	0.9389	0.9732 (n=5)	0.8053 (eps=0.01-0.13) (min_points=1-16)	0.7887 (bandwidth=0.01-0.14)	0.4792 (gamma=1.5)
	家庭用品	0.8792 (n=4)	0.7369	0.9883 (n=4)	0.1435 (eps=0.01-0.02) (min_points=2)	0.5350 (bandwidth=0.01-0.02)	0.2864 (gamma=1.2)
	ペット用品	0.6519 (n=3)	0.7369	0.6311 (n=3)	—	—	0.6713 (gamma=1.0)

4. 結果と考察

市場シェア率とリピート率について、中分類に含まれる全データの平均値を閾値に設定することで市場を 4 つに分類する。各クラスタの所属する市場は表 2 から読み取ることができる。

結果の解釈について「化粧品」を例に挙げると、クラスタ 0 は女性用化粧品，クラスタ 1 はヘアケア用品，クラスタ 2 は外れ値，クラスタ 3 は基礎化粧品，クラスタ 4 は男性用化粧品という特徴を読み取ることができた。また市場の特徴については市場シェア率とリピート率ともに高い市場は食に関する商品，市場シェア率が低くリピート率が高い市場は購入頻度が高い商品，市場シェア率とリピート率ともに低い市場は需要が限定的な商品であった。市場シェア率が高くリピート率が低い市場については外れ値が属していたため特徴の把握が難しかった。これらのことから、各クラスタおよび各市場の共通性は比較的多く発見することができ、特徴の理解がしやすかった。ただし、外れ値を含むクラスタについては共通性を発見することは難しかった。

表 2 各クラスタの傾向まとめ

		クラスタ0	クラスタ1	クラスタ2	クラスタ3	クラスタ4
食品	加工食品	低/高	高/高	低/低	高/高	—
	菓子類	低/高	高/低	—	—	—
	飲料・酒類	高/低	低/高	低/高	高/低	—
	その他食品	低/低	高/低	低/高	—	—
日用品	日用雑貨	低/高	高/低	高/低	低/高	低/低
	OTC医薬品	低/高	高/低	高/低	高/低	低/低
	化粧品	低/低	低/高	高/低	低/高	高/低
	家庭用品	低/低	低/高	高/高	高/低	—
	ペット用品	低/高	高/低	—	—	—

5. 結 言

本研究は、市場シェア率とリピート率を基に非階層クラスタリングを実施することで、クラスタおよび市場についての共通性を発見し、セグメンテーションにおいて新たな視点から知見を提供することができた。

今後の展望として、今回用いた市場シェア率とリピート率以外の変数の利用や商品の分類方法を変更することによって、また違った観点から興味深い共通性の発見をすることが期待できる。

文 献

- [1] フィリップ・コトラー: コトラーのマーケティング 4.0 スマートフォン時代の究極法則, 朝日新聞出版 (2017)
- [2] 株式会社 True Data : ウレコン, <https://urecon.jp/> (2024 時点)
- [3] Technical Note : シルエット分析, <https://hkawabata.github.io/technicalnote/note/ML/Evaluation/silhouette-analysis.html> (2024 時点)