

機械学習における教師なしクラスタリングの 地方債銘柄への適用

丹 波 靖 博*
原 口 健 太 郎**

1. はじめに

本稿では、機械学習モデルの一つである教師なしクラスタリングの手法を用いて、日本の地方債銘柄をその特徴を表すデータによりクラスターに分類する方法について検討を行う。

わが国においては、2016年に日本銀行によるマイナス金利政策やイールドカーブコントロール政策の開始により、国債のイールド水準が非常に低い水準で推移している。これにより、収益性の面で国債に比べ地方債の優位性が相対的に高まり、国債から地方債への投資の移行を行っている金融機関も見られる。日本においては、地方債は政府非保証債であるもののそのリスクは事実上国債と同じであるとする暗黙の政府保証論が存在し、現段階においては一般的にデフォルトする可能性は非常に低いと見られている。しかし、過去には北海道の夕張市が財政再建団体に指定され、財政状況が危ぶまれる事態に陥った事例があるほか、現在でも、一部の格付機関は自治体ごとに異なる格付けを付与しており、発行体固有のクレジットリスクは無視できない。また、直近では新型コロナウイルス感染症の影響により各地方公共団体の財政支出が増加しており、一部の地方公共団体においては長期的視点での財政状況の深刻化が懸念されている。

* 経済学部教授 y-tanba@seinan-gu.ac.jp

** 商学部准教授 haraguch@seinan-gu.ac.jp

このような環境下で地方債投資において重要になるのが、適切なリスク評価により投資先を選別するという視点である。具体的には、社債投資で行われているような投資対象の信用リスクの定量評価が重要になるとともに、地方債市場には取引額が少ない銘柄も存在することから、流動性リスクの定量評価も必要となる。地方債以外の金融商品投資において、投資家は一般に、リスク対比における収益性を評価し投資先を決定するが、地方債に関しては、特に新規における地方債の発行に係る市場であるプライマリー市場において、このようなリスク対比に基づく収益性評価が十分に行われていない恐れがある。つまり、地方債発行時、各金融機関は地元の地方公共団体や、つながりの深い地方公共団体のみに投資するケースも多いと考えられるのである。暗黙の政府保証が維持され、イールド水準にもそれほど差がない場合には、そのような投資行動も合理性がないとは言えないが、今後発行体固有のクレジットリスクがより重要視され、地方公共団体間のイールドの差がより拡大し、投資対象の選別が必要となる場合には、セカンダリマーケットのみならずプライマリーマーケットにおいても、上述のようなリスク量の定量化と投資先の合理的な選別の手法が重要となるだろう。このようなリスクの定量評価が行われた上で、多くの地方公共団体や多くの属性の地方債銘柄が存在する中、効率的な地方債銘柄の選別の手法が必要となってくると考える。

このような観点から、本稿では、クレジットリスクを定量化したうえで、効率的な地方債銘柄の選別手法を開発することを目的とする。具体的には、投資対象である地方債銘柄の選別について、機械学習のクラスタリングモデルを用い、地方債銘柄の特徴を表す収益性、属性、信用リスクの特徴量により類似銘柄のクラスタリングを行うための手法を検討する¹。

1 本章で議論した通り、流動性リスクの定量評価も重要なテーマであるが、紙幅の制約から本稿では取り扱わず、別稿での議論としたい。

2. 地方債における先行研究レビュー

地方債市場に関連する研究は会計学、財政学、ファイナンスの3つの領域に大別され、特にファイナンス分野では債券市場の信用リスク評価・流動性リスク評価に係る研究などが行われている。地方債に関する研究は主に米国で進展しており、Goldstein, and Woglom (1991), Wang, Wu, and Zhang (2008), Schwert (2017), Downing and Zhang (2004), Capeci (1991), Liu and Thakor (1984), Harris and Piwowar (2006), Novy-Marx and Rauh (2011a, 2011b) などの研究が見られる。研究の視点としては、財務健全性、流動性リスク、格付けなどに関連する内容が挙げられる。

わが国においては、地方債の信用リスク評価に関連した地方債のスプレッド推定モデルなどの研究が行われてきた。地方債スプレッドに関する研究は主に財政学の分野で行われており、石田・中里 (2019), 田中 (2012), 中里 (2008), Hattori and Miyake (2015), 三宅 (2017), 石川 (2007), 井潟・三宅 (2007) などの研究が見られる。著者らの近年の研究では、原口 (2017), 原口 (2018), 原口 (2019), Haraguchi and Oishi (2019), 原口 (2021), 原口・丹波 (2021), 丹波・原口 (2021), 丹波・原口 (2022) などの研究が挙げられ、格付けや地方債スプレッドに関連する研究を進めている。しかし、米国に比べるとわが国における地方債の先行研究事例は限定的であり、特にファンナンスの観点からの研究はほとんど見られない。

3. 機械学習と教師なしクラスタリング

3-1. 機械学習の概要

現在、急速な発展を見せている人工知能 (AI: artificial intelligence) は、「人間と同様な知的処理を実行可能な技術や機械」という若干広い概念で使用されることが多い。人工知能という概念を更に分類すると、強い人工知能と弱い人工機能が存在する。強い人工知能は「人間知能の模倣により人間と同様な認知能力を持った機械」の意味で使われる。一方、弱い人工知能とは「人間の知能

的な行動を模倣することにより人間の能力の一部を代替できる機械」のことを指す。機械学習は人工知能の分野のひとつであり、人工知能より狭い概念として「効率的・効果的にコンピューターが学習を行うための理論体系」を指す。つまり、機械学習は分析の際にモデル式を用いず、コンピューターの計算機能を利用してデータから学習し、潜むパターンや規則性を見つけ出す手法である。機械学習を用いることにより、これまで人間が判断してきた作業を効率化できる可能性や、複雑な構造を持ったデータの中に決まったパターンを見つけることで人間の認知以上の規則性を見つけ出すことができる可能性がある。秋庭他（2019）によると、機械学習には、入力データと分析の目的により、教師あり学習、教師なし学習、教科学習の3つの主要な種類に分類される。

教師あり学習は、正解となる答えを他の変数で説明するためのモデルを学習させる方法のことである。分析対象である問題の答えである目的変数（教師データ）と特徴を表すデータ（特徴量）をモデルに与え、目的変数を対象として特徴量による推定を行い、機械学習のモデルの推定精度を向上するために学習を行う。学習においては、目的変数である正解のデータと学習により得られた推定値の誤差をゼロに近づけることで学習し、徐々に正解できるデータの数を増やしていく。教師あり学習には、回帰モデル、決定木モデルなど様々なモデルが存在するが、画像認識等によく利用されるディープラーニングもその1つである。

2つ目の教師なし学習は、答えである目的変数を与えない機械学習モデルである。教師なし学習では、特徴量を使用し目的変数を推定するのではなく、データの中に特徴の近い部分集合を見つけたりすることで入力データの構造を理解することなどを目的とする。つまり、教師なし学習では与えられたデータの本質的な構造や法則をモデルのアルゴリズムが抽出していくことになる。教師なし学習において、複数の性質の似ているデータの集まりに分類する手法はクラスタリングと呼ばれる。機械学習を利用しなくても、2変数の場合には2次元平面にデータをプロットすることでデータの全体の散らばりや特徴を把握することができる場合も多い。一方、3つ以上の変数からなる場合は、データの散らばり具合やデータの構造を直感的に人間が理解するのは難しい場合が多

いが、クラスタリングによって似た属性により分類することでデータの構造を解釈することが可能となり、様々な応用へ繋がる可能性がある。

3つ目の手法である強化学習は、設定された環境の中で行動するエージェントが得られる報酬を定義し、報酬を最大化するように行動し学習していく手法である。この際、設定された環境において自らの成果を最大化するように試行錯誤を繰り返すことで最適な行動を行うように学習していく。強化学習では、多数の繰り返しが必要になる作業に関して、コンピューター上で実際に行動シミュレーションを行いながら最適な戦略を学習することができる。強化学習で有名な例としては、囲碁対局を行うアルゴリズムである AlphaGO が挙げられる。AlphaGO 同士がコンピューター上で囲碁の対局を繰り返しながら強くなっていき、現在では人間を超える強さに達している。近年では、将棋の世界でも人間の能力と同等かそれを超えるような将棋対局アルゴリズムが開発されている。

3-2. 教師なしクラスタリング手法

教師なしクラスタリング手法には、その分類手法により大きく分けて2つの方法が挙げられる。特徴量により単に似た属性のグループに分類するクラスタリングを非階層的クラスタリングと呼ぶ。非階層クラスタリングでは、始めに分析者がクラスター数を設定し、選定した特徴によって最もよく分類できるようにデータのクラスタリングを行っていく。一方、観測データの特徴によってデータがクラスターにまとまっていく過程を分析する場合は、階層的クラスタリングを用いることもある。階層的クラスタリングでは特徴の似ているクラスターを段階的に結合させていき、一つの大きなクラスターになるまで、操作を繰り返すことでクラスタリングを行っていく。その場合、データ間の近さを距離として定義し、計量化することによりクラスターにまとめていくことになる。教師なし学習は、教師あり学習と比べると正解となる答えがないため、結果の解釈が必要かつ重要となる場合もある。

クラスタリング以外の教師なし学習の手法としては、特徴量の数であるデータの次元を削減する手法（次元削減）が挙げられる。多次元のデータ情報量か

ら次元削減を行うことにより、より本質的なデータを抽出することができる。次元削減により、元々の特徴量数より少ない次元の情報でデータの特徴を表現できるとともに、データの持つノイズが除去されることで、分析精度を上げることが出来る可能性もある。また、データサイズが小さくなることにより、機械学習における計算時間の短縮も期待できる。次元削減の代表的な手法として、主成分分析が挙げられる。ここでの注意点として、次元削減で得られた主成分の意味については分析者が解釈する必要があり、解釈が難しい場合もある。

4. 分析データと特徴量

4-1. 特徴量について

本稿では、投資対象となる地方債をその特徴により選別するために、教師なし学習を用いてクラスタリングを行う。本稿のクラスタリング分析では、地方債銘柄の特徴量として、「対東京債スプレッド (TokyoSpread)」「地方債銘柄の年限 (YRSTOMAT)」「実質公債費比率 (jissitsu)」を用いる。対東京債スプレッドは同年限における各地方自治体の地方債イールドと東京債イールドとの差である。対東京債スプレッドは地方債銘柄の収益性を表す指標として採用している。地方債銘柄の年限は各地方債銘柄の残存期間であり、通常は年限が長くなるほどイールドが高くなるため、地方債銘柄の属性を表す特徴量として採用する。実質公債費比率は従前から総務省が「地方公共団体の財政の健全化に関する法律 (健全化法)」に基づき公表してきた地方公共団体の主要財政指標のうちの1つで、地方公共団体の財政状況を判断する重要な特徴量であるため、各地方公共団体の信用リスクを表す指標として採用している。

社債を対象とした信用リスク研究においては、無リスクと考えられる国債を基準とした対国債スプレッドを被説明変数に用いることが多い。一方、日本における地方債の場合、地方債のイールド水準が低いこと、年限10年以下の国債のイールドのマイナス幅が小さくないことなどから、対国債スプレッドは地方債イールドの変動よりも国債イールドの変動の影響を大きく受けてしまう。そのため、本稿においては、リスク水準が低く流通額の多い東京都債を基準とし

たスプレッド（対東京都債スプレッド）を特徴量として採用する。スプレッドの算出には、各地方債銘柄と同一残存年限の東京都債のイールド情報が必要となるため、東京都債をスプライン補間してイールドカーブを導出し、任意の残存年限のスプレッドを計算している。

4-2. 分析データユニバース

本分析では金利情報の取得は債券分析データベース・ソフトウェア「イールドブック」を使用し、データ期間は2019年4月15日から2019年5月31日までとする。本研究では、セカンダリマーケットにおける日次の地方債データを分析対象としている。データユニバースは、公会計情報等を全て入手可能であった49地方公共団体に係る57,354銘柄の地方債データである。

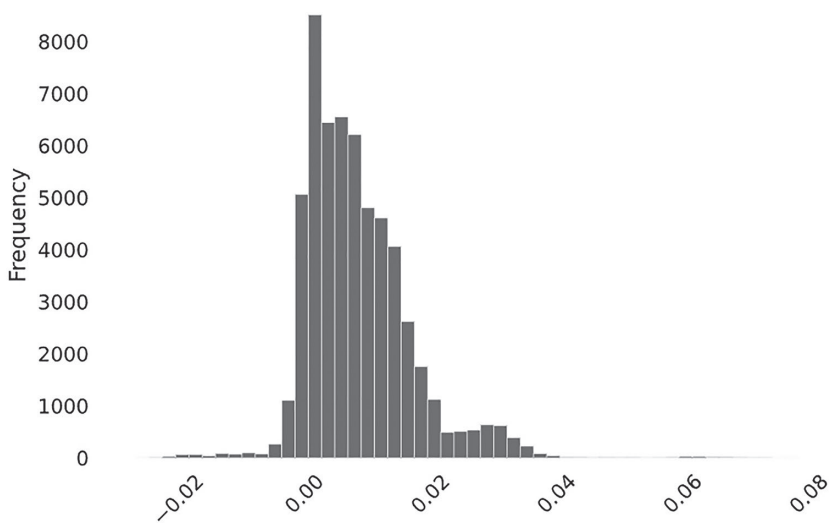
4-3. 特徴量の基本統計量と特性

本節では、クラスタリングに用いる特徴量について基本統計量などの調査を行う。各変数の特徴を知ることは、得られるクラスターの特徴を解釈する上で必要となる。表1は各特徴量の基本統計量、図1～3は各特徴量の頻度分布である。対東京債スプレッドは、最小値と最大値の差である値の範囲が約0.1%で、Kurtosis（尖度）が5を超えており正規分布（尖度3）に比べ尖った分布であり、Skewness（歪度）が約1.4と正であるため、左に偏った分布となっている。年限は平均が約8.6年で10年以下の銘柄が多く、20年超の銘柄は少ないことがわかる。実質公債費比率は平均が約11.7%、最大値が21%程度で、分布

表1 特徴量の基本統計量

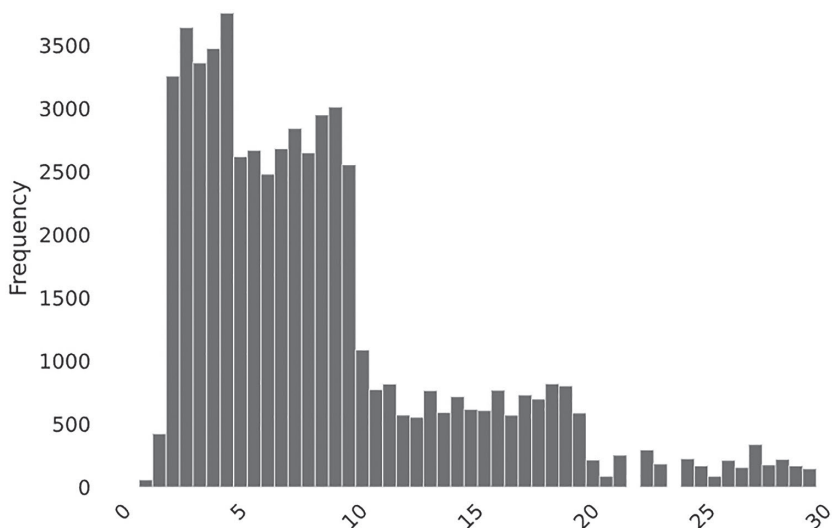
変数名	Mean	Minimum	Maximum	Standard deviation	Kurtosis	Skewness
対東京債スプレッド (%)	0.0084	-0.0268	0.0782	0.0085	5.3404	1.4362
年限 (年)	8.5755	0.666	29.866	5.9150	1.5485	1.3622
実質公債費比率 (%)	11.74291	1.6	21.1	4.4819	0.3902	-0.5077

図1 対東京債スプレッド分布



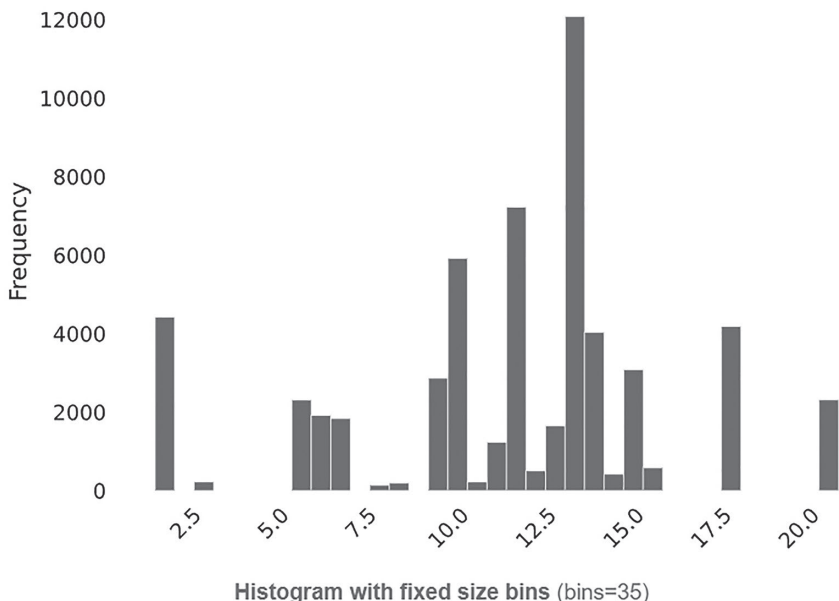
Histogram with fixed size bins (bins=50)

図2 年限分布



Histogram with fixed size bins (bins=50)

図3 実質公債費比率分布



にはばらつきがあることがわかる。これらの統計量や分布の状況から、これら3つの特徴量はそれぞれ異なった特徴を有しており、クラスタリングを行う上では、大きな障害とはならないと推察する。

図4は対東京債スプレッド（縦軸）と年限（横軸）の関係を散布図形式でわかりやすく表したものである。これによると、概ねこれら2変数には正の関係があることが確認できる。年限が長くなるほどスプレッドが大きくなるという一般的な関係性が確認できる。

図5は対東京債スプレッド（縦軸）と実質公債費比率（横軸）の関係を散布図形式で表したものである。これによると、概ねこれら2変数には正の関係があることが確認できる。実質公債費比率が大きくなると信用リスクが大きくなると考えられるため、このようにスプレッドが大きくなることが一般的である。

図6は3変数の相関値（スピアマンの相関）水準を表したものである。これによると、対東京債スプレッドと年限の相関は高い正の水準にあり、対東京債

図4 対東京債スプレッドと年限の関係

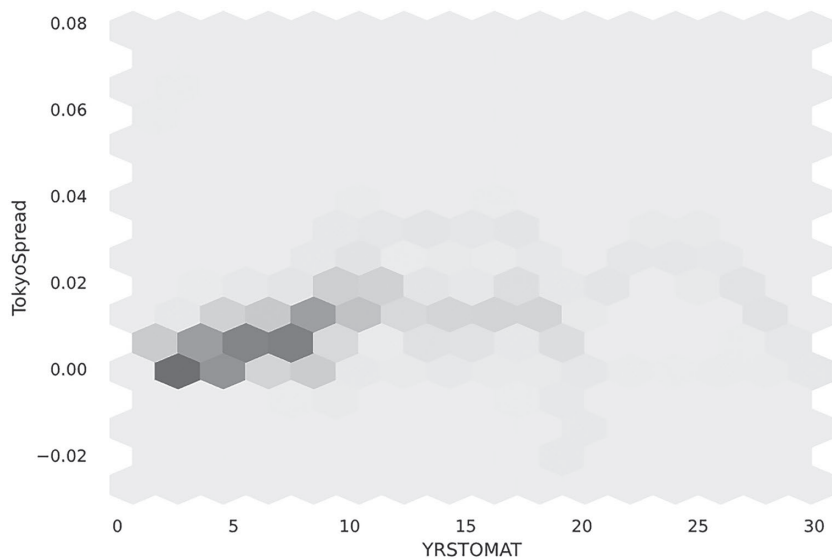


図5 対東京債スプレッドと実質公債費比率の関係

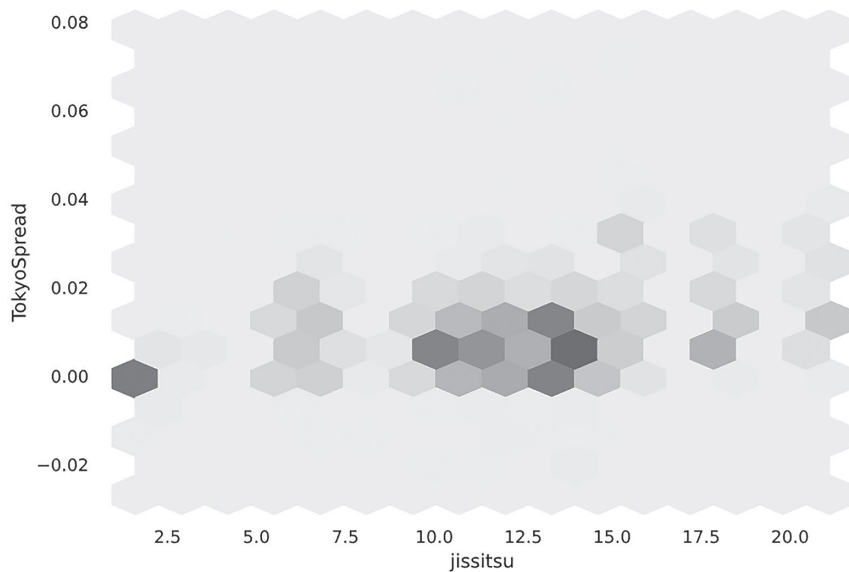
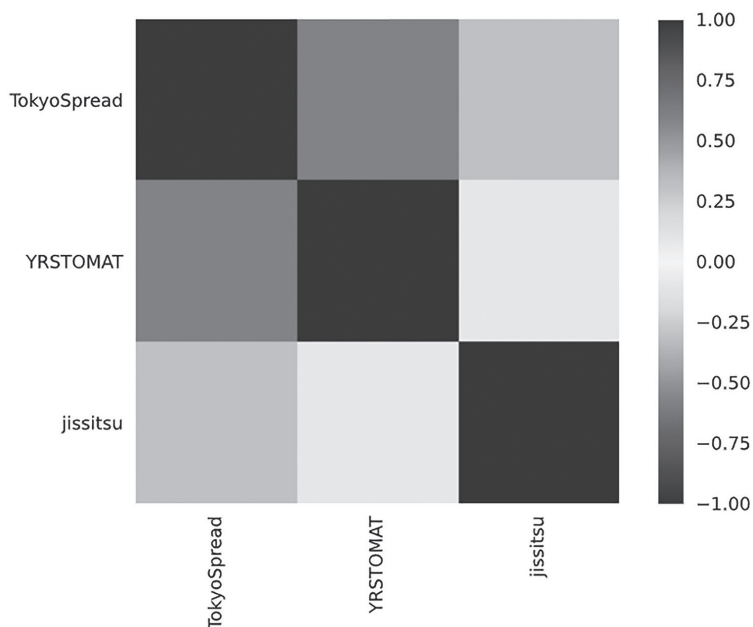


図6 3変数の相関関係



スプレッドと実質公債費比率は中程度の正の相関で、年限と実質公債費比率の相関はほぼ無いことがわかる。本分析のように用いる特徴量が3つと少ない場合で、高い相関の特徴量がある場合は、クラスタリングの精度に影響が出る可能性も考慮しておいたほうが良い。

5. クラスタリングモデルの構築

5-1. クラスタリングモデルの概要

本稿では、非階層クラスタリングモデルである k-means モデルと k-modes モデルを用いてクラスタリングを行う。まず、「対東京債スプレッド」「地方債銘柄の年限」「実質公債費比率」の3変数により、k-means モデルを適用し、3クラスタ（0～2）に分類を行った。加藤（2018）によると、k-means は、各点の重心までの距離の和を最小化する手法である。つまり、クラスターの重心

からの距離によりクラスターが分かれるように調整していくアルゴリズムで、近い属性を持つデータをクラスター化する手法である。一方、k-modes モデルは、データ間の非類似度を計算し、非類似度が小さいほどデータは類似していると認識する。k-modes アルゴリズムには、カテゴリデータにも適用可能であるというメリットがある。クラスター数は初期設定する必要のあるパラメータであるが、今回は両モデルとも3で設定を行った。

5-2. クラスタリングモデルの適用結果

両モデルを適用した結果を図7、図8に示す。これらの図は、各モデルにおける各クラスターのサンプル数である。これらを比較すると、k-means モデルのほうがクラスター間のサンプル数の偏りが小さいことが確認できる。通常はクラスター間のサンプル数の偏りが小さいほうが良くクラスタリングできていると解釈されることが多い。図8のk-modes モデルのように1つのクラスターにサンプルが偏っている結果が良いかどうかは、入力するデータ内容や分析の目的と合わせて吟味する必要があるだろう。

表2はk-modes と k-modes モデルによるクラスタリングの評価指標を表す。Silhouette と Davies-Bouldin は k-modes, Calinski-Harabasz は k-modes のほうが良い結果であることがわかる。クラスタリング評価指標はクラスタリングの良し悪しの判断に関して一定の尺度とはなるが、分析の目的と合わせてどの指標を重視するか判断が必要と思われる。

本稿では、地方債銘柄のデータを用いてクラスタリングを行う際のシンプルな実践方法をまとめることを目的としているため、特徴量数や適用モデルもシンプルな設定としている。更に複雑で精緻な分析を行うためには、採用する特徴量の吟味、適用モデルの比較、クラスター数の設定、分析目的を考慮した最適なモデルの選択が必要になる。例えば、本稿では両モデルとも、クラスター数は3で初期設定を行っているが、最適なクラスター数の設定の方法には、クラスタリングの改善度を統計的に判断できる Elbow 法の結果も確認する必要がある。また、クラスター内のサンプルがどの程度密になっているか（凝集度）の目安となるシルエット分析もクラスター数の設定に利用することができる。最終的には、得られたクラスタリング結果を見ながら、意図したクラスタ

図7 k-means モデルによる各クラスタリングのサンプル数

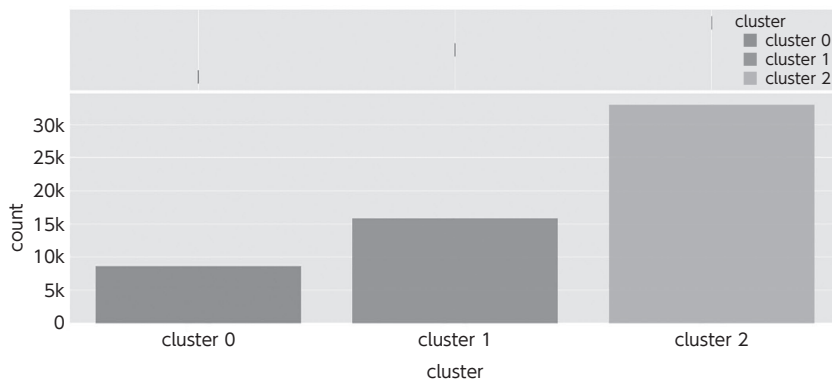


図8 k-modes モデルによる各クラスタリングのサンプル数

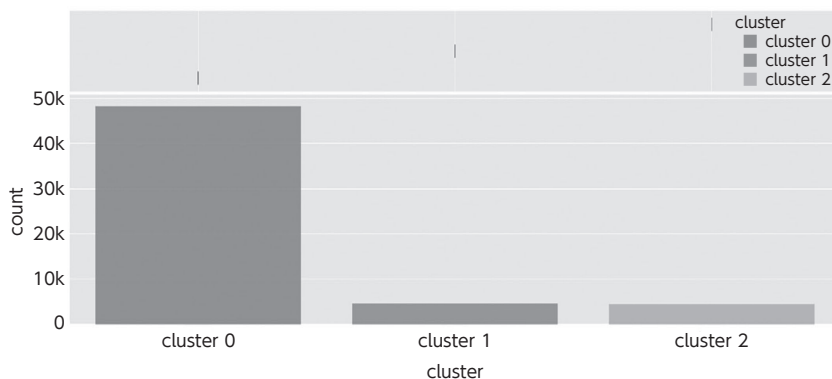


表2 k-means と k-modes モデルによるクラスタリング評価指標

Model	Silhouette	Calinski-Harabasz	Davies-Bouldin
k-means	0.3946	30350.6597	0.9675
k-modes	0.0174	5938.3011	4.8724

Silhouette (シルエット値)：クラスター内の平均距離 a と最も近い別のクラスターとの平均距離 b に対して $(b - a) / \max(a, b)$ で計算される。他のクラスターと離れているとき数値が1に近くなる。

Calinski-Harabasz：クラスター間の分散とクラスター内の分散の比を合計したもので計算される。数値が大きいほどクラスターが分離している。

Davies-Bouldin：クラスター内と最も類似しているクラスター間の類似度の比の平均で計算される。数値が小さいほどクラスターが分離している。

リングが実施できているかを確認する必要がある。例えば、各クラスターにおける変数ごとの分布を確認することで、そのクラスターがどのような特徴を持っているかを確認することができ、前述の「4-3.特徴量の基本統計量と特性」の結果と比較しながら、分析の目的に沿ったクラスタリングの有効性を確認する必要がある。

6. まとめと今後の展望

本稿では、地方債銘柄のデータを用いて、「対東京債スプレッド」「地方債銘柄の年限」「実質公債費比率」の3つの特徴量を用いたクラスタリング分析を行った。具体的には、k-modes と k-modes モデルによるクラスタリングを行い、それぞれのモデルにおける各クラスターのサンプル数と評価指標の確認を行った。その結果、適用するモデルによってクラスタリングの結果は異なることが確認された。どのモデルを採用するかは判断は評価指標のみで必ずしも可能なものではなく、分析の目的と照らし合わせて評価を行う必要があることが示唆される。具体的なクラスター数の設定や得られたクラスターの特徴などの詳細な分析は、今後の課題としたい。

さらに、機械学習のクラスタリングモデルを用いた分析結果を利用し、地方公共団体間のスプレッド水準の差について分析を行うことで、投資銘柄により地方公共団体間で恒常的な収益性の差が生じているか、どのような投資対象の選別方法が効率的であるのかについて分析を進める予定である。

参考文献

- 秋庭伸也, 杉山阿聖, 寺田 学, 加藤公一 (2019) 『見て試してわかる機械学習アルゴリズムの仕組み機械学習図鑑』翔泳社。
- 石田三成, 中里 透 (2019) 「地方債の信用リスクとスピルオーバー」『地方債の経済分析』有斐閣。
- 加藤公一 (2018) 『機械学習のエッセンス — 実装しながら学ぶ Python, 数学, アルゴリズム』SBクリエイティブ。
- 田中宏樹 (2012) 「地方債をめぐる地方公共団体間信用連関 — 市場公募債パネルデータを用いた実証分析」『証券経済研究』第78号, 69~79頁。

- 丹波靖博, 原口健太郎 (2021) 「機械学習を用いた我が国における地方債の信用リスク分析」『JAFEE2021夏季大会予稿集』65～76頁。
- 丹波靖博, 原口健太郎 (2022) 「わが国における地方債スプレッド推定モデル構築に対する機械学習の適用可能性」『経済学論集 (西南学院大学)』第56巻第1・2合併号, 近日刊行。
- 中里 透 (2008) 「財政収支と債券市場 — 市場公募地方債を対象とした分析」『日本経済研究』第58号, 1～16頁。
- 原口健太郎 (2017) 「公会計財務諸表分析による地方公共団体の財務健全性の測定」『会計検査研究』第55号, 117～132頁。
- 原口健太郎 (2018) 「地方公共団体における財務健全性の国際比較 — 日本と米国の公会計財務諸表に係る統計分析 —」『会計検査研究』第57号, 13～35頁。
- 原口健太郎 (2019) 「地方公共団体における公会計財務諸表と地方債市場との関連性の発現過程 — 米国各州のデータを用いた時系列分析 —」『会計プロGRESS』第20号, 16～31頁。
- 原口健太郎 (2021) 「米国地方債市場における会計学研究の発展とわが国での展開可能性」『産業経理』近刊。
- 原口健太郎, 丹波靖博 (2021) 「わが国における公会計と地方債市場に関する分野横断的研究の展開可能性」『商学論集 (西南学院大学)』第67巻第3・4合併号, 169～199頁。
- Capeci, J. (1991) Credit Risk, Credit Ratings, and Municipal Bond Yields: A Panel Study, *National Tax Journal*, 44(4), pp.41-56.
- Downing, C. and F. Zhang (2004). Trading Activity and Price Volatility in the Municipal Bond Market, *The Journal of Finance*, 59(2), pp.899-931.
- Duffie, D. and K. J. Singleton (2003) Credit Risk, *Princeton University Press*.
- Goldsterin, M. and G. Woglom. (1991) Market-Based Fiscal Discipline in Monetary Unions: Evidence from the U.S. Municipal Bond Market, *IMF Working Paper*, WP/91/89.
- Hattori, T. and H. Miyake (2015) Empirical Analysis of Yield Determinants in Japan's Municipal Bond Market: Does Credit Risk Premium Exist? *MPRA Paper 67127*, University Library of Munich, Germany.
- Haraguchi, K. and K. Oishi (2019) The International Differences in the Relationship between Public Financial Reports and Bond Markets: A Comparison between Japanese Local Governments and U.S. States, in Proceedings of the 17th Biennial CIGAR Conference in Amsterdam.
- Harris, L. E. and M. S. Piwowar. (2006) Secondary Trading Cost in the Municipal Bond Market, *The Journal of Finance*, 61(3), pp.1361-1397.
- Liu, P. and A. V. Thakor (1984) Interest Yields, Credit Ratings, and Economic Characteristics of State Bonds: An Empirical Analysis, *Journal of Money, Credit and Banking*, 16(3), pp.344-351.
- Novy-Marx, R. and D. R. Joshua (2011a) The crisis in local government pensions in the United States, in Robert Litan and Richard Herring, eds.: *Growing Old: Paying for Retirement and Institutional Money Management after the Financial Crisis* (Brookings Institution, Washington, DC).
- Novy-Marx, R. and D. R. Joshua (2011b) Public pension promises: How big are they and what are they worth? *The Journal of Finance*, 66(4), pp.1211-1249.
- Schwert, M. (2017) Municipal Bond Liquidity and Default Risk, *The Journal of Finance*, 72(4), pp.1683-1722.
- Wang, J. C. Wu, and F. X. Zhang. (2008) Liquidity, Default, Taxes, and Yields on Municipal Bonds, *Journal of Banking & Finance*, 32(6), pp.1133-1149.

参考 URL

総務省「地方債 Q&A」

https://www.soumu.go.jp/main_sosiki/c-zaisei/chihosai/chihosai_qanda.html

総務省「地方公共団体の主要財政指標一覧」

https://www.soumu.go.jp/iken/shihyo_ichiran.html

日本証券業協会

<https://www.jsda.or.jp/>

ロンドン証券取引所 FTSE Russell グループ「The Yield Book Inc」

<https://www.yieldbook.com/m/products/yield-book/index-jpn.shtml>

謝 辞

本研究は科学研究費補助金（JSPS KAKENHI Grant Number JP19K23214, 20K02058, JP21K13412）の交付を受けて行ったものである。また、ロンドン証券取引所・FTSE グループ The Yield Book Inc. には、債券分析ソフトウェアドブックを通じて効率的な分析ツールと貴重なデータを提供いただいた。この場を借りて御礼申し上げたい。