

2021年3月下旬 白田由香利

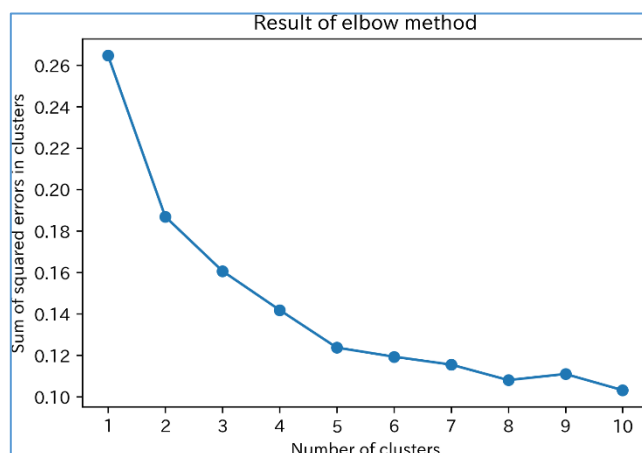
よんどころない事情のため、緊急事態宣言解除になったので、福岡に行ってきた。忙中閑ありで「そういえばコロナのため太宰府天満宮にお願いごとだけしてお礼参りをしていなかった」ことを思い出し、大宰府とその隣の九州国立博物館に行ってきた。この国立博物館の建物は超近代的でガラスのドームのような形状をしている。非常に長いエスカレーターに乗り、周囲に人は皆無。ガラスから満開の桜と陽光がそそぐが、ヒトの姿は無い。まるでSFで人類がいなくなった後、残ったりっぱな博物館の施設に調査に来た探検隊の気分だった。アシモフの Foundation and Empire の図書館での探索シーンを思い出した。朝いちばんであったからであろうが、観客は本当にいなかった。人類のロックダウンの営みとは無関係に桜の花は咲く、それが大変よく理解できてしまった九州国立博物館だった。天神様が歌に詠まなくても、梅も桜も春が来れば咲くのだ。自然界の四季の移り変わりのメカニズムは本当に神秘だと思った。

機械学習の分野のひとつにクラスタリングがある。株価の変動パターンをクラスタリングしてみる。株価変動パターンのように動きのパターンでクラスタリングする場合、一般的なクラスタリングとは違い、時系列データクラスタリングというジャンルになる。時系列データの動きが似ている(類似している)ものをまとめてクラスターに分ける。類似していることは、距離が近い、と言う。類似度の測り方(距離の定義の仕方)として、DTW (Dynamic Time Warping) 動的時間伸縮法[1]が広く使われている。こういった類似度の測り方は機械学習の世界では日進月歩で、次々と手法が開発されていく。ここでは、DTW よりも後に考案された k-Shape という距離定義の手法を用いてクラスタリングを行ってみた。K-Shape 法の特長は Invariant to scaling and shifting、つまりピークが時間的に多少ずれていても類似と認識するし、ピークの山の高さが大きくても小さくてもその時点でピークがあったので類似している、と認識してくれる点である[2]。その結果、人間だったら、これとこれは似ていると認識しますよね、という2つのパターンをAIも「似ている」と判断してくれるわけである。

電子機器製造業(コンピュータを含む)の世界 TOP129 社の株価の変動パターンをクラスタリングしてみた。データ期間は2018年4月2日から2020年3月19日の約2年間という長期間である。株価データはビューロー・ヴァン・ダイク社の ORBIS というデータベースから検索した。グローバル企業が多いので、証券取引所の休日も異なる。土曜日曜及び元日は除き、その他のローカルな休日(例えば、日本のGW)は、前後のデータを線形補間した。始め250社のデータを取ったが、データが欠落している企業は除いたので、結果として129社となった。

クラスタリングは機械学習の教師なし学習のひとつである。クラスタリングでは始めに何個のクラスターに分けるのか数を指定する。「このデータを適当に8個のクラスターに分けてください」というようにリクエストする。適当に、というところで距離の定義を指

定する。クラスター数の個数を決める際は、エラーのレベルがクラスター数に応じてどのように変化していくかをグラフに描いて、適当な数を決める(右図参照)。右図を見て、5



でもよいかと思ったが、もう少しエラーが減る8個を選択した。

次のページに、電子機器製造業(コンピュータを含む)の世界TOP129社の2年の株価をk-Shape手法により8個のクラスターに分けた結果を示す。株価のパターンを比較する際は、生の株価データではなく、リターン値を使う。リターン値とは、株価の前日比の自然対

数を取った値である。例えば、100円が120円に値上がりしたら、 $120 \div 100 = 1.2$ が前日比で、その自然対数を取った値がリターン値である。リターン値を企業ごとに標準化してから、その標準化されたリターン値を入力データとする。そして、クラスタリングのプログラムがリターン値の2年間の変動パターンが類似している会社クラスターを8個作って返してくれる。

リターン値はそのままグラフにしても人間には理解しがたいので、標準化された株価でプロットした(次ページ参照)。標準化(平均を0、標準偏差を1とすること)したのは、6000円平均で動いている株価と500円平均で動いている株価のパターンを比較できるようにするためである。また、各クラスターの平均を折れ線グラフで示した。129社の2年に渡る株価を8つのクラスターに分けるという作業は人間では処理が多くて面倒でできない。しかし、機械学習は簡単にやってくれる。クラスター数を3, 4, 5, 6, 7, 8と変化させるというようなことも厭わずやってくれる。全く便利で有難い。

このクラスタリングの結果、面白いことが分かった。

1. クラスター# 1は全27社が日本企業であった。
2. クラスター# 2は13社中11社がUS企業であった。
3. クラスター# 3は全8社が韓国企業であった。
4. クラスター# 4は全42社の殆どがUS企業であった。APPLEはここに含まれていた。
5. クラスター# 5の19社にはフランス、ドイツなど様々な国の企業が含まれていた。
6. クラスター# 6は全13社が中国企業であった。
7. クラスター# 7は全7社が台湾企業であった。
8. クラスター# 8は全10社が台湾企業であった。鴻海(ホンハイ)精密工業はこちらに含まれていた。

クラスタリングは教師無しデータ処理で、もちろん、クラスタリングする企業がどこの国籍が知らない。それでも、このように国別クラスターとなったことは興味深い。投資家は個々の企業のパフォーマンスよりも、その企業の所属する国をみて投資を決めている可能性が高いと推察できる。

終わり

引用元： 此の道や行く人なしに秋の暮 芭蕉

参考文献

- [1] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," in *KDD workshop*, 1994, vol. 10, no. 16: Seattle, WA, USA, pp. 359-370.
- [2] J. Paparrizos and L. Gravano, "k-shape: Efficient and accurate clustering of time series," in *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, 2015, pp. 1855-1870.