

ロックダウン下では、家の中で整理整頓をする時間が増える。ピアノの調律師のかたは、この時期忙しくなっているという。粗大ごみの排出量も増加している。私もゆっくりとお着物の衣替えに時間をかけた。実母と京都旅行を計画していたのだが、デルタ株も注意したいので結局中止し、2 人分の夏帯と単衣の着物を用意していたが、筆筒に戻した。今は、罹患しないことのほうが重要だ。その時、お能「猩々」から取った猩々の柄の帯が目についた。猩々とは、霊的な力をもつ酒好きの怪異で、お能では猩々が気持ちよく酔って舞を舞う、というストーリーも殆どない、祝賀のための演目だ。私は高校の仕舞部のキャプテンのとき、文化祭や学芸会などで猩々をよく舞った。あれこれ思い出されて「ああ、この帯をしてお能を見に行きたい」と心の底から思った。コロナが怖いので映画館も能楽堂も人の集まる場所には一切いかない。何かネットで動画はないものかと検索したところ、ありました。「能のお稽古【後編】《猩々》の仕舞【能楽 シテ方金春流（こんぼるりゅう）金春流八十一世宗 金春憲和】」
<https://www.youtube.com/watch?v=57OUrB8bOy8>

これには驚きました。学習院女子中高等科の仕舞部は金春流なので、八十世宗 金春安明先生(学習院大の OB です)もまだ学生でいらしたとき、仕舞部での我々のお稽古を時折みてくださいました。憲和先生はそのご長男でいらっしゃいます。この動画の秀逸な点は、ご指導の内容も素晴らしいですが、シャツとズボンの軽装で、扇もお稽古用の金春のおだんごのような丸模様(正式には星)の稽古扇を使っている点です。着物と袴のなりではない、その点が驚きました。(女子部のお稽古のときは、先生方はすべてお着物ですが)。私が今、見たいのは、NHK の教育番組で流すようなお能の録画ではないのです。こういう日常のお稽古、が見たかったのです。コロナの時期に、文化継承のために果敢に攻めていらっしゃるのだろうと、宗家の意気込みと御意思の強さを感じました。そして、「このような時節ですが、家でしっかりお稽古いたしましょう」という宗家の声が聞こえるような気がして、「はい！」と返事をしてしまいました。私も「白田先生も頑張って教材作っているのね」と学生さんに思ってもらえるような数学の反転授業の教材(まあ、ビデオ教材です)作りに入魂していこうと、元気づけられました。学生の皆様、動画教材のリクエストお待ちしております。

さて SHAP 値を使った回帰分析の事例を紹介する。SHAP 値を使う分析は、回帰分析である。回帰分析よりも、回帰分析プラス SHAP 値による分析のほうが、深い分析が可能となる。復習ですが、回帰分析では、被説明変数(ターゲット変数とも呼ぶ)がひとつあり、その被説明変数に影響を与える変数(これを説明変数と呼ぶ)が複数与えられている。例えば、気温とビールの売上数である。気温だけでなく、入場者数も説明変数に加えると、重回帰分析となる。回帰分析では、ターゲット変数に強く影響を与えるのはどの説明変数であるのか、を求める。これから扱う回帰分析の事例は以下のようなものである。詳細はデータ工学研究会で発表した論文を読んで頂きたいが、本稿ではその事例を用いて、平易な言葉で SHAP 値の有用性を説明していく[1-3]。

- データ：日本の製造業 3 種の株価。
 - ◇ 自動車製造業(79 社)、精密機械製造業(26 社)、電気機器製造業(112 社)
- 目的：2012 年 11 月 13 日から 2015 年 8 月 18 日の期間の株価上昇率に影響を与えた説明変数は何かを調べる。

- ターゲット変数：株価上昇率＝2015年8月18日株価÷2012年11月13日株価
- 説明変数：以下の6個の経営指標を使う。

1. 売上高成長率[%]
2. 売上高営業利益率[%]
3. 自己資本当期純利益率 (ROE) [%]
4. 使用総資本事業利益率 (ROA) [%]
5. 棚卸資産 (在庫) 回転率[回]
6. 有形固定資産回転率[回]

説明変数の値は、10年間(2006年度から2016年度)の年次データの平均値を使っている。企業は年に依っては、何か大きく値を動かすことがあるので、少なくとも5年分のデータは必要である。



また、どうしてこの時期の株価の上昇率を選んだかという点、この時期は上図の日経 225 インデックスの変化を見ても分かるように、非常に安定して株価が上昇している。大きく落ち込んでいる期間はない。このような全体として大きく成長している時期に、大きく上昇した企業と少ししか成長できなかった企業と、その違いはどこから生まれたのかを調べたい。企業のレジリエンス(回復力)を調べるために、あえてコロナで大きく株価が暴落した後の株価上昇率を調べることもある。安定時期を選んだ理由は、企業の基本的な体力や企業構造を知るためには、レジリエンスではなく、基本の基礎体力を調べるためには、この安定した時期の方が適しているからである。そして、その株価上昇率には6つの説明変数のうちの説明変数が最も効いているのかを調べる。そういう回帰分析を行う。

SHAP 値による分析は以下のステップを取る。

- (1)機械学習を使って回帰モデル $f(X)$ を作る。
- (2) $f(X)$ を使って特性関数 characteristic function ν を作る。
- (3)特性関数を使って各企業の各説明変数の SHAP 値を計算する。

ここまでは前号の<Shapley 値>の復習である。忘れた人は前に戻って復習してください。それでは、きちんと理解しているか確認クイズを行いましょ。

(Q1) 以下の文が正しいか間違いか○か×をつけなさい。

- 機械学習による回帰分析のアルゴリズムとは例えば XGBOOST がある。
- 機械学習による回帰で得られた回帰モデルは一般に複雑形状をしている。
- Shapley 値の理論におけるプレーヤーとは、回帰の場合、説明変数のことである。
- SHAP 値は、S 社の ROE の SHAP 値は 0.12 というように、各社の各説明変数ごとに計算される。

れる。

(Q2) 上述した6個の説明変数がある場合、その特性関数における部分集合の数はトータルで何個になるか？ (空集合は1個と数えます)

さあ答えです。Q1 の文章はどれも正しい。SHAP 値は以下に示すように、各社の各説明変数ごとに計

算される。SHAP 値は、その説明変数がターゲット値へどれだけ貢献しているかを示しているの、ターゲット値が高い企業のほうが、概して、各 SHAP 値も高くなる。

Q2 の答えである。2 の 6 乗で 64 個である。特性関数とは、例えば、売上高成長率君と ROE 君と ROA 君だけで作業すると 1.5 しか稼げなかったが、そこに棚卸資産回転率君が加わると、稼ぎが 1.7 に増えました、というように、取りうる全ての部分集合に対して、その稼ぎ額を与える関数である。S 社の特性関数はこれ、P 社の特性関数はこれ、というように別々の特性関数が定義される。しかし、そのような特性関数は実際のところ、分かりません。「じゃ、SHAP が計算できなくなってしまうではないですか？ いったいどうやって特性関数を求めるのですか？」

いい質問です。これこそが、Shapley 博士の Shapley 値の公式を機械学習の世界でも使えるように工夫した Lundberg 氏の秀逸なところである。いない人（説明変数）のところには、全体の平均値を入れて $f(X)$ を計算して、それをその部分集合による稼ぎ、とする。これにより疑似的な特性関数が定義できる。それを使って Shapley 値を計算したものを SHAP 値と呼んでいる。

以下の X 社と Y 社の SHAP 値を比較してみる。ターゲット値は X 社が 2.58 で Y 社の 1.52 よりも高い。SHAP 値の合計を計算すると、X 社のほうが -0.0008 なので、Y 社よりも高かった。両方マイナスだったので分かりづらいかもしれないがマイナスのなので、-0.0008 のほうが大きい。概して SHAP 合計値のランキングは、ターゲット値のランキングに近くなる。個々の説明変数の SHAP 値を 2 社で比較してみよう。この 2 社の比較では、大きく影響しているのは「売上高成長率」の SHAP 値の違いである。X 社のほうが株価上昇率が Y 社の 2 倍近くあるのだから、全部で値が高いだろうという予想に反して以下の説明変数では Y 社の方が SHAP 値が高かった： 売上高営業利益率、棚卸資産回転率。

	株価上昇率	1. 売上高成長率[%]	2. 売上高営業利益率[%]	3. 自己資本当期純利益率 (ROE) [%]	4. 使用総資本事業利益率 (ROA) [%]	5. 棚卸資産 (在庫) 回転率 [回]	6. 有形固定資産回転率 [回]	SHAP 合計
X社	2.58	0.1363	-0.1525	0.0033	0.0246	-0.0487	0.0362	-0.0008
Y社	1.52	-0.8057	-0.0794	-0.1383	-0.0472	-0.0030	0.0133	-1.0604

不思議な感じがするかもしれないが、これこそが、SHAP 値が特性関数を使っているからこそであり、正にその企業の構造の特性を表している。次回から 2 回、詳細に SHAP の散布図を示していく。

終わり

参考文献

1. 辻浦衣美、白田由香利：「Shapley値による株価上昇における重要要素の分析 ～ 自動車製造企業のケースについての考察 ～」、電子情報通信学会データ工学研究会(DE)、DE2021-1、pp.1-5、2021.
2. 藤巻美舞、白田由香利：「Shapley値による株価上昇における重要要素の分析 ～ 電気機器製造企業のケースについての考察 ～」、電子情報通信学会データ工学研究会(DE)、DE2021-2、pp.6-11、2021.
3. 保科慧、白田由香利：「Shapley値による株価上昇における重要要素の分析 ～ 精密機械製造企業のケースについての考察 ～」、電子情報通信学会データ工学研究会(DE)、DE2021-3、pp.12-17、2021.

引用元：さまざまの事思ひ出す櫻かな 芭蕉

