

経済財政政策部局の動き：経済の動き 機械学習を活用した経済 分析

政策統括官(経済財政分析担当)付
参事官(総括担当)付参事官補佐
小寺 信也

技術革新が進展する中、ビッグデータ・機械学習が景気分析に用いられるようになってきている。本稿では、小売のPOSデータ及び新聞記事のテキストデータを用いて、AI技術の一つである、コンピュータにデータを解析・予測させる手法（機械学習）を活用した消費動向の分析について、最近の内閣府の取組と暫定的な分析結果を紹介する。

1. POSデータを用いた経済指標の予測

スーパー等の小売店では、購入する際に商品をバーコードで読み込み、各商品の価格・数量等の情報を取組・記録している。このようなシステムはPOS (point-of-sale) と呼ばれ、売上や在庫管理に役立てられている。ここでは、全国のスーパーマーケット約1,200店舗から収集された日用品・食料品（除く生鮮品）のPOSデータ（売上高）を分析する。

POSデータを利用することの利点としては、速報性が非常に高いことや、高頻度に観察が可能であることが挙げられるが、マクロ分析に利用するにはPOSデータのカバレッジが限定的である点等に留意する必要がある。今回利用したPOSデータも、217品目分類が日次で利用できること（高頻度）や、その値が数日後には利用できること（速報性）に強みがある一方、POSデータの大部分を占める食料品等のシェアは消費の2割程度であり（GDPベース）、消費全体をカバーできるわけではない。

そこで、POSデータの強みを活かして、弱みを補うための分析を行う。通常、マクロ経済指標は、その月が終了してから1~2か月後に公表されるものが多いが、迅速な景気動向把握のためには、ある月が終わった後すぐにその月のマクロ経済指標の値が予測できること（ナウキャスト）が望ましい。この点において、POSデータは速報性という強みをもっているが、カバレッジの低さもあり、単純にPOSデータの値を比

較しただけでは、ナウキャストを行うことは難しい。

ただし、マクロ経済指標の動きと似通った動きを、POSデータの動きの中から抽出することができれば、カバレッジ等の問題を解決し、ナウキャストを行うことが可能になると考えられる。そこで、売上高のPOSデータから、特徴的な動きを機械に抽出・学習させることで、経済産業省「商業動態統計」の小売業計（前年比）をナウキャストできるかを試みた。また、小売業計には、家電や衣類等の売上動向が気温等の天候に左右されやすい商品が含まれることや、天候データも速報性の高い情報であることを踏まえ、POSデータと天候データを併せて機械に学習させる分析も行った。

機械学習の方法は、ランダムフォレスト（Random forest）や勾配ブースティング（Gradient boosting）等の手法を用いた。これらの分析手法は、アンサンブル学習と呼ばれ、複数の学習モデルを融合させることで高いパフォーマンスを出せることで知られている。なお、以下では、他の機械学習の手法と比較して、総合的にパフォーマンスが良かった勾配ブースティングの結果を紹介する。

2007年1月~2016年12月までの10年間の月次のデータを機械に学習させ、2017年1月~2018年3月の小売業計の前年比のナウキャストを行ったところ、POSデータのみを用いて分析した場合、予測期間における相関係数は0.63であった。次に、POSデータに加え、天候データを加えて分析した結果を見たのが図1であるが、予測期間における相関係数は0.78と予測精度の改善がみられた。気温等に敏感な商品の売上高動向等を補正することができたと考えられる。図1をみると、前年比の水準には差がみられるものの、前年比前月差では概ね似たような動きをしていることから、大まかではあるが、小売業計の前年比について一定程度のナウキャストを行うことができたと考えられる。

ただし、今回の分析でも水準の差に加え、2018年以降等で動きが異なる部分もみられることから、精度向上に向けては更なる分析を進めていく必要がある。

2. ディープラーニングによるセンチメント指数

近年では、デジタル化されたテキストデータの利用が容易になったことから、テキストを用いた分析も活

発化している。定性的な情報を数値データに変換するため、分析手法によって結果が変わり得る等との課題があるものの、POSデータ同様に速報性の高さ等から、既存の分析を補完する分析が可能となることが期待されている。

ここでは、消費者マインドに影響を及ぼすことが指摘されているニュース記事のテキストデータを利用して、消費者マインドとの相関を確認することとしたい。

具体的には、景気の現状と先行きに対する評価とコメントがセットで利用できる内閣府「景気ウォッチャー調査」を利用し、ディープラーニング（深層学習）の手法として知られるLSTM（Long Short Term Memory）層を導入したRNN（Recurrent Neural Network）を用いて、機械にどのようなコメントが景気認識に対して良い・悪いコメントであるのかを学習させた。この手法を利用することの利点は文脈を時系列として評価できるため、テキスト等のデータを分析する際には、より精度の高いモデルの構築が可能となることである。学習は現状と先行きで別々に行い、学習の結果、景気ウォッチャー調査の新規コメントに対して93~95%の確率で正しく分類できるようになった。

こうして学習が完了した機械に、2013~17年における18万件以上の新聞記事（今回は日本経済新聞の朝刊を対象にした）を読み込ませ、それぞれの記事が良い（悪い）と分類される確率を算出させた。これにより新聞記事のテキストデータを数値化することが可能となり（以下、数値化した指標を「新聞センチメント指数」という）、消費者マインドとの相関をみることができる。

新聞センチメント指数は、経済面、政治面、社会面といった紙面分野別に作成し、各紙面と内閣府「消費者態度指数」との相関を確認した。図2は、このうち高い相関が確認されたものをみたものだが、景気ウォッチャー調査の先行きで学習させた「経済面」、「マーケット商品面（肉の卸売値、チャーター料の変化等、価格変化の記事が多い）」の新聞センチメント指数と消費者態度指数の相関が約0.6と高くなっている。同じ「経済面」でも景気ウォッチャー調査の現状で学習させると、相関が半分程度になることから、消費者マインドは見通しも踏まえて形成されていることが示唆される。また、対象記事全体では相関が低いことから、消費者マインドは経済面に掲載されるような

全般的な経済情勢の変化や、マーケット商品面に多く見られる価格変化等、特定の情報に影響を受ける傾向があると考えられる。

こうしてみると、新聞等のテキストデータについても経済分析に対して有用な情報をもっている可能性が考えられる。今後、こうしたテキスト情報も活用することで、より迅速かつ確かな景気把握が可能となると考えられることから、引き続き、研究を進めていく必要がある。

図1 POSデータによるナウキャスト

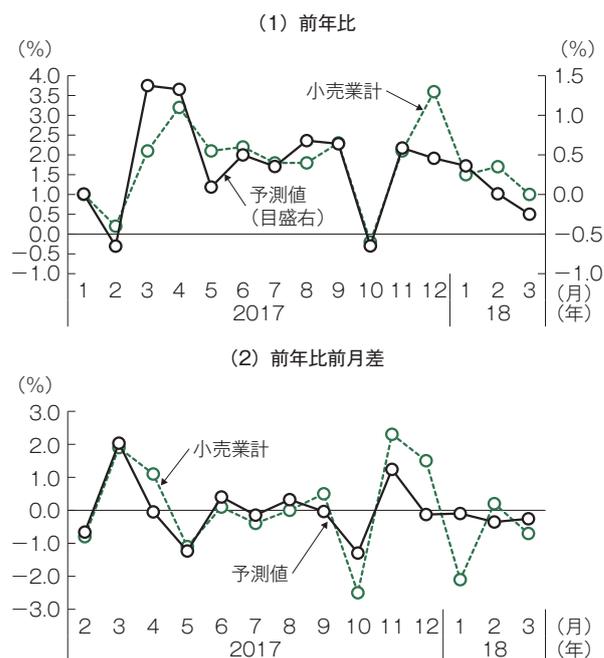
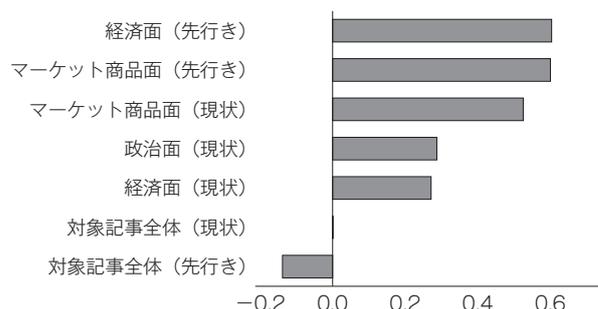


図2 新聞センチメント指数（紙面別）と消費者態度指数との相関係数（2013~17年）



主な参考文献

小寺信也・藤田隼平・井上祐介・新田堯之(2018)「POS・テキストデータを用いた消費分析—機械学習を活用して—」経済財政分析ディスカッション・ペーパー・シリーズ DP/18-1

小寺 信也（こてら しんや）