

金融レポート、およびマクロ経済指数によるリアルタイム 日銀センチメントの予測

Real time sentiment analysis of Bank of Japan using text of Financial report and
macroeconomic index

余野 京登¹ 和泉 潔¹

Kyoto Yono¹, Kiyoshi Izumi¹

¹ 東京大学大学院

¹Graduate School of Engineering, The University of Tokyo

1 はじめに

金融市場を取り巻く情報は日々増加している。物価指数や産業活動指数など数値データもあれば、企業の有価証券報告書、ロイターなどの配信している経済ニュースなどのテキストデータも含まれている。この膨大な量の情報から、投資家やトレーダーは、それぞれの投資戦略に従い、必要な情報を選び出し、投資判断を下している。

ニューラルネットワークや遺伝的アルゴリズム等を数値データに用いて市場分析を行うものがあり、一定の成果を上げている[1]。一方で、近年、データマイニング技術を用いて、市場動向を分析する研究が多く行われている。ニュースのヘッドラインを分析し、為替市場の短期予測を行った研究[2]や、中央銀行の発行するテキストを用いた研究などがある。日本銀行のテキストを対象にしたものでは、ニューラルネットワークを用いることでレポートを指数化する研究[3]、政策変更確率を予測する研究[4]、米国のFOMCのテキストを対象にしたものでは、トピックモデルによる分類と極性語による指数化に関する研究[5]がある。

本研究では、日本銀行の景気に対するセンチメントを日本銀行の発行するテキストに対して、トピックモデルおよびニューラルネットワークを用いて、トピック事に分解した指数化を行った。

今後の展望として、よりリアルタイム性を出すためにテキストの発行以前において、マクロ経済指数を外挿することによりリアルタイムな指数の予測を目指す。

2 手法

2.1 データ

日本銀行の景気に対するセンチメントを知るために分析の対象となるテキストは主に以下の通りであるが、今回の分析では金融経済月報を対象とする。

分析の期間は1998年1月から2015年12月である。なお、2016年以降、金融経済月報は「経済・物価情勢の展望」(展望レポート)に集約された。

表 1 日本銀行の発行するテキスト

テキストの種類	タイミング	頻度	分量	フォーマット
金融政策決定会合議事要旨	1~2ヶ月後	月次	20-30	定形
主な意見	1週間後	月次	5	定形
総裁定例記者会見要旨	翌営業日	月次	20	不定形
経済・物価情勢の展望	翌営業日	四半期毎	10/100	定形
金融経済月報	翌営業日	月次	20-30	定形
講演	不定	月1回程度	10~20	不定形
記者会見	不定	月3回前後	1~2	不定形

2.2 LDA によるトピック分類

金融経済月報は、金融政策決定会合における政策判断の背景となる金融経済情勢を説明する資料として公開されており、その内容は、実体経済、物価情勢や海外経済など多岐にわたる。Latent Dirichlet Allocation (LDA) [6] を用いてこれらのトピックの抽出を行う。

LDA とは、一つの文章に複数のトピックが存在すると仮定したトピックモデルである。文章 d ごとにトピック分布 θ_d があり、このトピック分布 θ_d はパラメータ α のディリクレ分布に従う。トピック z ごとに単語分布 Φ_z があり、これらによって単語生成 w がさせる。なお、単語分布 Φ_z はパラメータ β のディリクレ分布に従う。全文章 D が生成される確率は以下の式のように表される。

$$p(D|\alpha, \beta) = \prod_{d=1}^M \int p(\theta_d|\alpha) \left(\prod_{n=1}^{N_d} p(z_{dn}|\theta_d) p(w_{dn}|z_{dn}, \beta) \right) d\theta_d.$$

この生成確率が最大化されるようにパラメータ推定を行う。

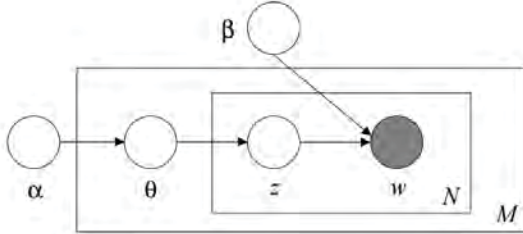


図 1 LDA のグラフィカルモデル

LDA を用いることにより、トピックの単語の生成確率が算出される。このモデルを各月の金融経済月報に適用し、トピックの比率を計算する。

2.3 文章のセンチメント付与

文章のセンチメント指数を計算する際に、極性語を用いたアプローチ[5]もあるが、今回の分析においては、先行研究[3]で用いられた景気ウォッチャー調査を学習データに LSTM RNN で構築したモデルでのセンチメント付与を行う。

内閣府の景気ウォッチャー調査は業種ごとのにおける景気判断（良い、やや良い、変わらない、やや悪い、悪いの 5 段階評価）とその理由となるテキストがセットとなっており、かつデータ量が十分にあり、センチメント付与の学習データとしては良質なものである。

このデータセットに対して、LSTM RNN により、分類器を構築する。

RNN は音声や言語、動画などの系列データを扱うのに適したニューラルネットであり、その中でも LSTM は中間層にメモリーユニットを有しているため、長い文脈をとらえてモデル構築が可能である。

LSTM ブロックは記憶セル $c(t)$ 、入力ゲート $i(t)$ 、出力ゲート $o(t)$ 、忘却ゲート $f(t)$ の 4 つの要素からなり、それぞれ以下の式で記述される。

$$a(t) = f(W_c \mathbf{x}(t) + U_c \mathbf{h}(t-1) + b_c)$$

$$i(t) = \sigma(W_i \mathbf{x}(t) + U_i \mathbf{h}(t-1) + b_i)$$

$$f(t) = \sigma(W_f \mathbf{x}(t) + U_f \mathbf{h}(t-1) + b_f)$$

$$o(t) = \sigma(W_o \mathbf{x}(t) + U_o \mathbf{h}(t-1) + b_o)$$

$$c(t) = i(t) \odot a(t) + f(t) \odot c(t-1)$$

$$h(t) = o(t) \odot g(c(t))$$

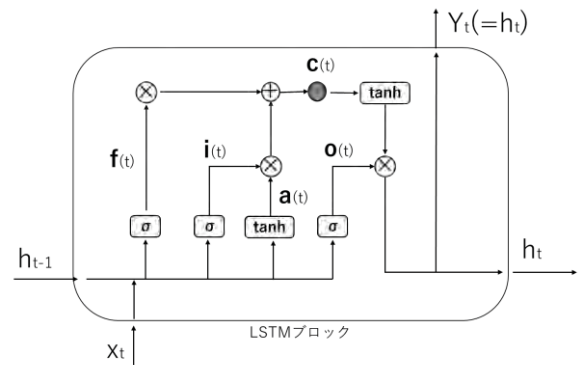


図 2 LSTM のグラフィカルモデル図

学習した LSTM RNN モデルに金融経済月報を一文ずつ判断させ、その平均値をその金融経済月報のセンチメントとする。

3 結果

3.1 トピック分類の結果

LDA の結果は以下の通りとなった。

表 2 各トピックの上位単語

個人消費	公共投資	在庫	設備投資
個人消費	公共投資	資本財	機械投資
住宅投資	公共工事請負金額	部品	実質
海外経済	実質輸出	財別	船舶
所得環境	所得環境	情報関連	輸送機械
設備投資	公共工事出来高	実質輸入	設備投資
景気	有効求人倍率	自動車関連	電力
公共投資	完全失業率	電子部品	機械受注
鉱工業生産	労働需給	中間財	全国百貨店売上高

物価	マネーサプライ	鉱工業生産	企業金融
消費者物価	マネーサプライ	在庫調整	金融機関
物価	電子部品	鉱工業生産	貸出態度
国内企業物価	C D	実質輸出	企業金融
国際商品市況	デバイス	最終需要	設備投資
ターム物金利	輸送機械	素材	資金供給面
生鮮食品	マネタリーベース	情報関連財	資金調達コスト
短期金融市場	業種別	鉄鋼	信用力
輸入物価	在庫バランス	海外経済	製造業

マーケット	貿易	資金供給	企業収益
株価	実質貿易収支	C P	製造業
国債	実質ベース	社債	企業収益
流通利回リスプレッド	サービス収支	民間銀行貸出	設備投資
為替相場	世帯	発行環境	雇用者所得
社債	海外経済	発行残高	中小企業
対米ドル相場	黒字幅	資金調達	賃金
流通利回り	二人以上	資金需要面	常用労働者数
日経平均株価	消費財	資金需要	所定外給与

金融経済月報を対象に LDA を用いて、トピック分類すると、個人消費、公共投資、在庫、設備投資、物価、マネーサプライ、鉱工業生産、企業金融、マーケット、貿易、資金供給、企業収益の 12 トピックに分類することができた。各月のトピック割合の推

移を図示すると以下になる。

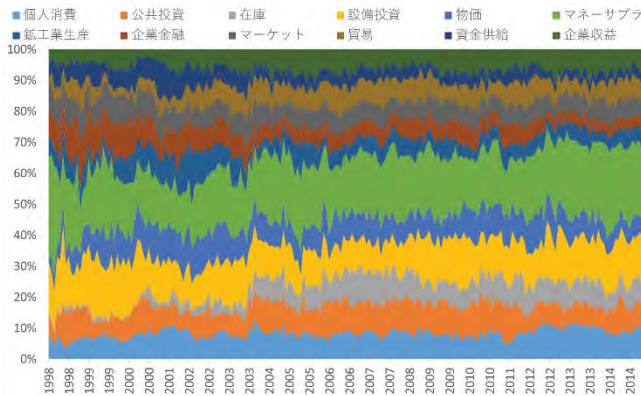


図 3 トピック推移

3.2 センチメント付与

まず、金融経済月報をトピックに分類せずに全文に対して LSTM RNN モデルでセンチメント指数を算出すると以下のようなになる。先行研究[3]とほぼ同じ結果となった。

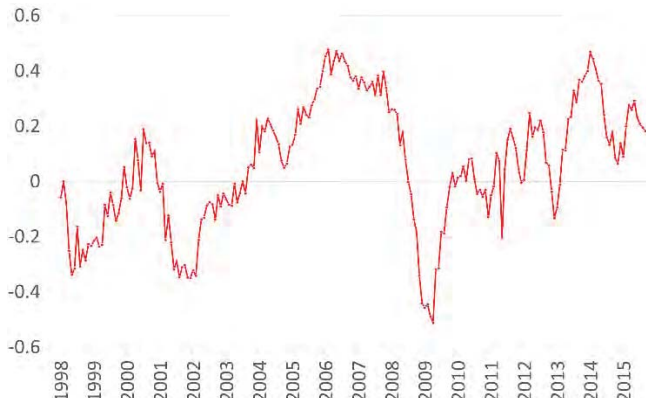


図 4 金融経済月報のセンチメント指数

一方、LDA でトピック分類を行い、トピック毎に対してセンチメントを付与し、全体に対する寄与度をトピック別に算出すると以下のようなになる。



図 5 トピック別の寄与度

トピック別のセンチメントを見ると、ほとんどのトピックは同様な正負の推移を見せている。2015 年のセンチメントの寄与度に注目すると。物価トピックは全体に対してマイナスの寄与を与えている。これは、アベノミクスにより景気は回復し、物価以外のマクロ経済は良い状態に対して、物価は政策目標である 2%に届いていないため、金融経済月報にもそれが表れているのではないかと考える。

4 まとめ

本研究では、先行研究[3]で行われた、金融経済月報を対象にした日本銀行の景気に対するセンチメントの指数化に対して、LDA によるトピック分類を加えることで、センチメントをトピック別に算出し、寄与度に分解することで、どのマクロファクターがセンチメントに影響を及ぼしているかについて、考察を行った。

今後の展望としては、よりリアルタイム性を出すためにテキストの発行以前において、マクロ経済指数を外挿することによりリアルタイムな指数の予測を目指す。

参考文献

- [1] 山口 和孝: ニューラルネットと遺伝的アルゴリズムを用いた株式売買支援システム, (2002)
- [2] D. Peramunetilleke, R. K. Wong. : Currency exchange rate forecasting from news headlines, Australian Computer Science Communications Volume 24 Issue 2, (2002)
- [3] 山本 裕樹, 松尾 豊: 景気ウォッチャー調査の深層学習を用いた金融レポートの指数化, 第 30 回人工知能学会全国大会,(2016)
- [4] 塩野 剛志: 文書の分散表現と深層学習を用いた日銀政策変更の予想, 第 16 回 SIG-FIN,(2016)
- [5] 伊藤諒, 須田真太郎, 和泉潔: トピック別極性値付与方法による FOMC 議事録の評価, 第 17 回 SIG-FIN,(2016)
- [6] Blei, D. M., Ng, A. Y., Jordan, M. I. : Latent Dirichlet Allocation, Journal of Machine Learning Research, pages 993-1022, (2003)