

LDA 方式による金融政策トピック抽出

白田 由香利[†] 橋本 隆子[‡] 佐倉 環^{*}

[†] 学習院大学経済学部 〒171-8588 東京都豊島区目白 1-5-1

[‡] 千葉商科大学商経学部 〒565-0456 市川市国府台 1-3

^{*} 公益社団法人 日本経済研究センター 〒100-8066 東京都千代田区大手町 1-3-7

E-mail: [†] yukari.shirota@gakushuin.ac.jp, [‡] takako@cuc.ac.jp, ^{*} sakura@jcer.or.jp

あらまし 本稿では、経済活動における不確実性の分析を、東日本大震災直後の日本経済に対して行う。手法としては、日本銀行政策委員会金融政策決定会合議事要旨のテキストデータから、トピック抽出を行ない、トピックの時系列変化を分析する。我々が抽出したトピックの一つは、震災後の株価の暴落と同じ時期にピークとなっているものがあり、我々は日本経済の先行きに関する不確実性を表すトピックと考えた。また、企業生産活動に関する不確実性と関連すると考えられるトピックも抽出した。我々はトピックの意味づけの際、オリジナル文書を読むほか、株価及び為替変動などの経済時系列データを用いた。本稿では、我々の本アプローチについて論じる。

キーワード トピック抽出, Latent Dirichlet allocation, 不確実性, 金融政策, 東日本大震災

Financial Policy Topic Extraction by using LDA

Yukari SHIROTA[†] Takako HASHIMOTO[‡] and Tamaki SAKURA^{*}

[†] *Gakushuin University, Faculty of Economics, 1-5-1 Mejiro, Toshima-ku, Tokyo, Japan 171-8588*

[‡] *Chiba University of Commerce, Chiba University of Commerce, 1-3-1 Konodai, Ichikawa-city, Chiba, Japan 565-0456*

^{*} *Japan Center for Economic Research, 1-3-7 Otemachi, Chiyoda-ku, Tokyo, Japan 100-8066*

E-mail: [†] yukari.shirota@gakushuin.ac.jp, [‡] takako@cuc.ac.jp, ^{*} sakura@jcer.or.jp

Abstract We analyze economic policy uncertainty in Japan just after the East-Japan great earthquake disaster in 2011. The target document is text data of the Policy Board of the Bank of Japan financial policy meeting proceedings summaries. From the documents, we shall extract topics by using the LDA algorithm. We extracted interesting three topics from the summaries. One topic has the peaks which overlapped the Japanese stock price fall times. The topic is considered to be related to the risk around the future of Japan. We think that another disaster-related topic expresses the uncertainty of the company production activity. In this paper, we used other economics time series data as plausible reasons to explain the extract topic. We discuss in the paper our approach by economics data to give a plausible excuse for the topic.

Keyword Topic extraction, Latent Dirichlet allocation, Uncertainty, Financial policy, the East-Japan great earthquake disaster

1. 始めに

本稿では、経済活動における不確実性の分析を、東日本大震災直後の日本経済に対して行う。手法としては、日本銀行政策委員会金融政策決定会合議事要旨の

テキストデータから、トピック抽出を行ない、トピックの時系列変化を分析する。

ビッグデータからのトピック抽出は、金融の分野でも広く用いられている[1, 2]. 株価が大きく変動した場合、その外部要因を新聞記事テキストから特定する研

究。また、トピック変化から株価データの推移を予測する研究、テキストデータから金融市場の動向を分析する研究[3]などがある。また、余野らは、英国中央銀行の金融政策委員会議事録から金利の動向を予測し、それにより通貨の取引戦略を決める研究を行っている[4]。

他方、我々の目的は、金融政策決定会議録の内容把握を支援することである。議事録を読む人は、その政策決定過程を理解するために読むのであるが、議事録には「為替変動」「震災の影響」「欧州ソブリン問題」など多種多様な単語が出現し、議論された問題を整理、把握するまでに時間がかかる。この時、トピック抽出技法により抽出されたトピックを手掛かりとすることで、議論された問題を理解することが容易になる。

分析結果として興味があることは、

(1) トピックの出現時期と持続期間、

(2) 要因が国内的要因か海外的要因かの区別、等である。(1)は、ある事件が起こり、その事件が政策に影響を与えた期間を測るものである。そして、他の事件にとって代わられるまでの時間である。また、その原因が国内的要因か、海外的要因かも重要である。一般的に言って、国内的要因の予測のほうが、海外要因に比較して容易であり、政策決定者も対処が容易である。他方、海外的要因は予測が困難であり、対処が困難である。

政策決定議事録は、経済などが安定している時期にはトピックに大きな変化は見られないと推測される。安定期とは、例えば、James Stock と Mark Watson などが指摘した *great moderation* と呼ばれる 1980 年代から 2005 年にかけての米国経済の非常に安定的な拡大の時期である[5]。この時期、米国の物価は安定しており、株価や地価も上昇を続けた。このような経済安定期には、政策決定議事録のトピックも安定しており、変化は少ないと予想できる。反対に、不確実性が高い時期には、政策決定議事録のトピック変化は激しくなることが予想される。

トピック抽出が有効であるのは、そうした、不確実性が高い時期と考えられる。不確実性が高い時期とは、例えば、東日本大震災後の日本経済、リーマンショック後の世界経済などが挙げられる。

本稿で使う不確実性という語は、経済学者ナイトが “Risk, Uncertainty & Profit” [6] の中で定義した *uncertainty* を指す。ナイトはリスクと不確実性を区別する。リスクは、それが起こる可能性についての客観的確率分布を思い描けるもので、それが起こる客観的確率分布を思い描けないものを、ナイトは不確実性と呼ぶ。不確実性の高い場合、数学的期待値は客観的には計算不可能で、主観的確率だけに依存することにな

る。ケインズ経済学は、しばしば「不確実性革命」としてとらえられる。ケインズは将来に対する知識はいつも揺れ動き不確実であるとし、不確実性の存在する中での経済行動の分析が欠かせないと主張した。奥村は、不確実性が高い場合、人間は理性的判断だけでは行動していないので、不確実性が高い場合は脳科学及び心理学的分析により、人間の直感や個人的な判断に基づく行動を解明することが重要であると論じ、サブプライムローン・バブルなどを例にあげて、説明している[7]。経済の不確実性の指標については、シカゴ大の Davis らが U.S. Economic Policy Uncertainty Index として計算している[8-10]。Davis らも、*The Wall Street Journal* などの新聞記事からのテキストマイニングを行って、この指標を計算している[10]。こうした経済や政治の分野での不確実性の指標の研究は活発に行われている。

本稿では、日本銀行政策委員会金融政策決定会合議事要旨[11](以後、議事要旨と略す)のトピック分析を行う。次節では、分析手法について説明する。第3節は抽出したトピックの時系列変化を示し、どのトピック内容を考察する。第4節はまとめてである。

2. 実験方式

分析対象は、日本銀行政策委員会金融政策決定会合議事要旨である。東日本大震災は日本経済にとっては、未曾有の大惨事であり、比類なき不確実性の高い状況であった。震災の影響だけではなく、2010年前後に起きた欧州ソブリン問題、タイの大洪水等の影響により、日本経済の不確実性は著しく高まった。よって、震災直後はトピック変化も激しいと予測されるので、トピック抽出を行う時期として、震災直後の2011年3月から12月を選んだ。

金融政策決定会合とは、日本銀行の最高意思決定機関である政策決定の金融政策の運営に関する事項を審議・決定する会合である[11]。同会合は、毎月1~2回開催され、議事事項としては、(1) 金融市場調節方針、

(2) 基準割引率、基準貸付利率および預金準備率、(3) 金融政策手段（オペレーションにかかる手形や債券の種類や条件、担保の種類等）、(4) 経済・金融情勢に関する基本的見解等がある。議事要旨は、次の決定会合で承認のうえ、その3営業日後に公表される。WEBで公開される。議事録は議事要旨に比較して、詳細な審議過程が記録されているが、各会合から10年を経過した後でないと公表されないため、震災直後の2011年の議事録はまだ公表されていない。よって我々は議事録要旨を分析対象とした。

トピック抽出は、Latent Dirichlet Allocation(潜在的ディリクレ配分法)モデルによるアルゴリズムに従った。

LDA 方式は、ベイズの推定(Bayesian inference)を用いている。LDA 方式については、オリジナル論文[12]を参照して頂きたい。この LDA 方式の探索空間は広く、計算量が大きいので、現在の多くの LDA を用いた研究では、Gibbs サンプリングを用いたランダム・ウォークによる探索が広く用いられている[13]。我々も Gibbs サンプリングを用いる計算方式を取った。

LDA 方式では、トピック数を最初に決める。普通、5 から 10 個とする。決めたトピック数が適切であるかの判定が非常に重要となるが、トピック変化が顕著に現れるトピック数、を試行錯誤で決めていく。

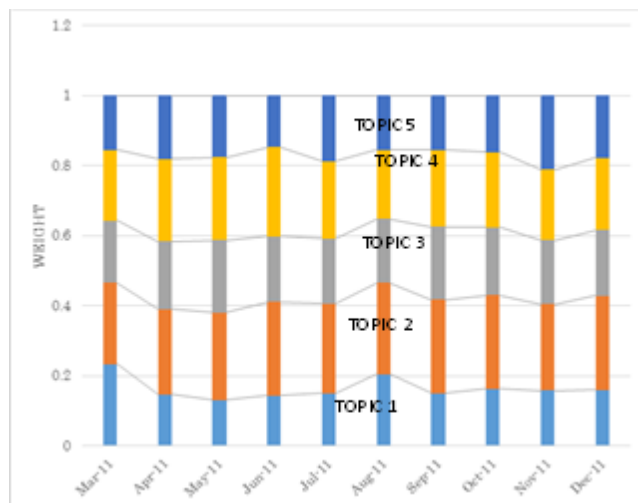


図 1. 抽出したトピックの時系列変化。比重は月ごとで合計が 1 になるように正規化している。

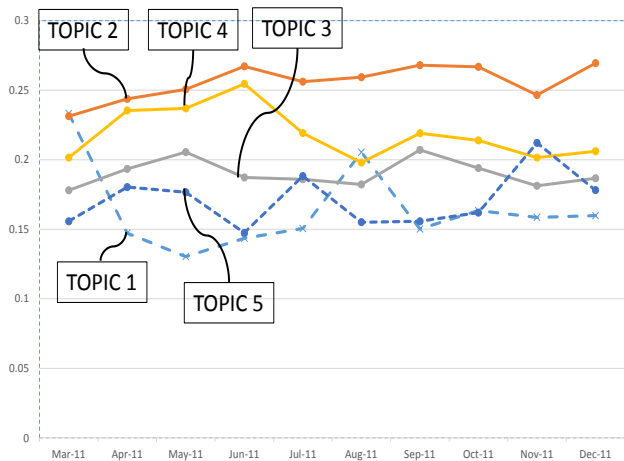


図 2. トピックごとの比重の時系列変化のようす。大きいドットのラインで示された TOPIC1 は、3 月と 8 月にピークがあること。小さいドットのラインで示された TOPIC5 は、7 月と 11 月にその前の月からの増加が大きいことが分かる。TOPIC4 は、前半である 3-6 月に増加したが、以降、比重が低い。



図 3. 対ドル為替の変化。円高が進行したのは、震災直後の 3 月 17 日(79.22 円), 8 月 3 日(77.18 円), 10 月 28 日 (75.84 円)。矢印は、7 月と 11 月を表している。

3. 抽出トピックの評価

本節では、抽出したトピックの内容について検討する。我々はトピック数は 5 が適切であると考えた。以下、トピック数 5 とした場合の評価を行う。トピックの比重の月毎の変化を図 1 に示す。

我々はこの中で TOPIC1 に注目した。このトピックは、3 月、8 月に比重が大きい、この時期は円高進行の時期と重なるが(図 3)、TOPIC5 とは、以下の点が異なる。

- (1)震災直後の 3 月 14 日の議事要旨で比重が非常に大きい。
- (2)11 月に円高進行が起こったが、その時には TOPIC1 の比重は高くない。
- (3)日経平均株価が 3 月、8 月に下落している(図 4 参照)。

このトピックの比重のピーク時は、円高進行というよりは、株価の下落時期に重なっている。そこで、我々はこのトピックは、株価下落に関係をもつトピックであると考えた。次に、このトピックに頻出する単語から、その単語を含むオリジナル文章を読んでもみると、以下のような文章等があった。

多くの委員は、やや長い目でみて、電力供給を巡る不確実性や、円高の進行などを背景に、企業の海外シフトが加速する可能性にも注意する必要があると指摘した。複数の委員は、特に、震災に伴う供給面の制約を克服しつつあるこの時期に円高が進むことは、企業マインドを大きく悪化させる可能性があるとして述べ、別の委員は、日本経済の成長期待の低下に繋がる可能性もあると述べた。(金融政策決定会合議事要旨、2011)

1年8月4日開催分より引用)

参 考 文 献

TOPIC1 が、震災直後のみに比重を上げるトピックであれば、震災による不確実性に関するトピックと説明できるが、震災から回復しつつある8月にも比重を高めていて、それが株価下落と強く相関している。8月は、電力供給問題、円高進行などの国内要因により、海外に企業シフトが起こるのではないかという不安が高まり、結果として株価下落につながったと考えるのは妥当である。よって、我々は、TOPIC1を、「日本経済の先行きに関する不安」を示すトピックと解釈した。このような、株価の下落と強い相関をもつトピックが抽出できたことは、非常に興味深い。

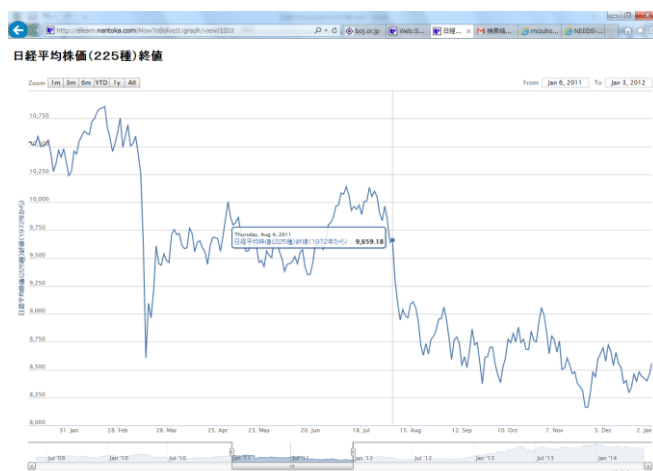


図 4. 日経平均株価の変化。震災直後と、8月に株価が大きく下落している。

4. まとめ

日本銀行政策委員会金融政策決定会合議事要旨のテキストデータから、トピック抽出を行ない、トピックの時系列変化を分析した。対象期間は、日本経済の不確実性が非常に高い、東日本大震災後の3月から12月までとした。分析はLDAモデルに基づくトピック抽出アルゴリズムを用いて行った。

トピックのタイトルを判定する手法としては、原文の文章を参照するほか、関連性の高い経済時系列データを参照した。例えば、株価、及び、為替レートの変動である。その結果、トピックのうち、一つが株価の下落と関連性が高いことを発見し、そのタイトルを「日本経済の先行きの不安」に関するトピックとした。トピック抽出分析は、トピックの正当性を科学的にいかにかに評価するかが重要であり、その証拠として我々は他の経済時系列データを用いた。

今後とも、金融政策会議の議事録の内容把握を支援するため、トピック抽出の研究を進めていきたい。

- [1] D. Peramunetilleke and R. K. Wong, "Currency exchange rate forecasting from news headlines," presented at the Proceedings of the 13th Australasian database conference - Volume 5, Melbourne, Victoria, Australia, 2002.
- [2] S. Kogan, D. Levin, B. R. Routledge, J. S. Sagi, and N. A. Smith, "Predicting Risk from Financial Reports with Regression," *Tepper School of Business, Paper 416*, <http://repository.cmu.edu/tepper/416>, 2009.
- [3] 三菱東京 UFJ 銀行, "テキストマイニング手法を用いた経済市場分析の試み," 2007.
- [4] 余野京登, 和泉潔, 後藤卓, 松井藤五郎, 陳昱, "英文経済レポートのテキストマイニング分析ツールの開発," 第6回人工知能学会ファイナンスにおける人工知能応用研究会 (SIG-FIN), 2011.
- [5] J. Stock and M. Watson, "Has the business cycle changed and why?," *National Bureau of Economic Research Macroeconomics Annual 2002*, vol. 17, 2003.
- [6] F. H. Knight, *Risk, uncertainty and profit*. Boston, New York, : Houghton Mifflin Company, 1921.
- [7] 奥村洋彦, "「場」の歴史的変容と経済分析:リアルタイム(Real-Time), イグノランス(Ignorance)を取り入れた構造分析の必要性," *学習院大学経済論集*, vol. 48, pp. 213-236, 2012.
- [8] S. R. Baker, N. Bloom, and S. J. Davis, "Measuring Economic Policy Uncertainty," *working paper*, 2013.
- [9] S. R. Baker, N. Bloom, and S. J. Davis. *Economic Policy Uncertainty*. Available: <http://www.policyuncertainty.com/>
- [10] D. Wassel, "Trying to Calculate the Cost of Uncertainty," *Wall Street Journal WORLD*, p. <http://online.wsj.com/news/articles/SB10001424127887323316804578161252496111518>, Dec. 5, 2012.
- [11] 日本銀行. (2011). *日本銀行政策委員会金融政策決定会合議事要旨*. Available: http://www.boj.or.jp/mopo/mpmsche_minu/minu_2011/index.htm/
- [12] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993-1022, 2003.
- [13] T. L. Griffiths and M. Steyvers, "'Finding scientific topics," *Proceedings of the National Academy of Sciences*, vol. 101 (Suppl. 1), pp. 5228-5235, 2004.