

Supervised LDA モデルによるニューステキストを用いたマクロ経済不確実性指数の構築

Measuring Macro Economic Uncertainty from News Text by Supervised LDA

余野 京登^{*1}
Kyoto Yono

和泉 潔^{*1}
Kiyoshi Izumi

坂地 泰紀^{*1}
Hiroki Sakaji

島田 尚^{*1}
Takashi Shimada

松島 裕康^{*1}
Hiroyasu Matsushima

^{*1} 東京大学大学院

School of Engineering, the University of Tokyo, Tokyo, Japan

For financial market participants, uncertainty of macro economic events have crucial impact when they make decision to buy or sell financial assets because macro economic event influence the asset price. In this study, we aim to build a model to measure macro economic uncertainty from news text. We proposed extended topic model which using not only news text data but also numeric data as a supervised signal for each news articles. We constructed four macro economic uncertainty indexes by our proposed model. Each indexes matches with historical macro economic events and the correlation are higher with the volatility of market index related to the uncertainty index.

1. はじめに

市場参加者にとって、マクロ経済の不確実性は意思決定をする際に重要な役割を果たす。マクロ経済の不確実性とは、経済の今後の動きについて不透明で予測不可能なときに市場参加者がいなく懸念であり、各国の金融政策や財政政策、あるいは2国間の貿易摩擦の不確実性など多岐にわたる。あるマクロ経済の不確実性が増すと、投資家はそのマクロ経済に影響を受けやすい資産の売買の意思決定（買い増すべきか、あるいは、売すべきか）がしづらくなる。マクロ経済の不確実性により将来の資産の価格の方向性が不透明になる。例えば、米国連邦公開市場委員会（FOMC）が今後利上げをするかどうか、どのタイミングで行うか、などに関する金融政策の不透明感が高まると、米国債を買うべきか売すべきか、様々な憶測が飛び交い、米国債のボラティリティが上昇する。一方、日本銀行に関しては、当面低金利政策を維持するという考えが市場参加者間で統一されており、日本国債の価格変動はほとんどない。

1.1 マクロ経済の不確実性の測定

市場には複数のマクロ経済の不確実性が同時に存在している。もしそれを分類し、定量的に測定がかなとなれば、市場参加者にとって有意義なものとなる。ヘッジ戦略やストレステスト等に活用が可能となる。マクロ経済の不確実性を測定するには、資産の価格変動を見るのが一番簡単な方法であるが、資産価格に複数のマクロ経済の不確実性が影響する場合は、需給バランス等のマクロ経済の不確実性以外の要因により、変動する場合には測定が困難となる。近年、ビックデータの台頭、自然言語処理技術の向上により、金融経済テキストを分析し、市場に潜むマクロ経済の不確実性を測定する研究が行われている。Baker ら (2016) が自然言語処理技術を用いて、政策の不確実性指数を測定しており [Baker 16], また Also Manela ら (2017) はニュース記事に潜む市場のボラティリティの予測を行った [Manela 17].

1.2 本研究の目的

本研究では、ニューステキストを用いて、マクロ経済の不確実性指数の構築を目指す。トピックモデルを拡張し、テキストデータのみならず、数値データ (VIX インデックス) を教師信号として用いたモデルの適用を試みた。また、構築した4つのマクロ経済の不確実性指数（日本マクロ経済不確実性指数, 米

国マクロ経済不確実性指数, 中国マクロ経済不確実性指数, 欧州マクロ経済不確実性指数）を対象にした定性的、定量的な分析を行った。

2. 関連研究

数値データのみを用いて、不確実性指数の構築やボラティリティの予測を行っている研究はいくつかあり、Helena ら (2017) は株価データを用いて指数を不確実性指数を構築しており [Chuli 叩 17], また、Google トレンドの数字データを用いて、ボラティリティの予測を行った研究もある。([Castelnuovo 17],[Hamid 15]) また、Asaf ら (2017) はニューステキストを用いてボラティリティへの回帰を行っており、構築した News implied volatility 指数の構築を行った [Manela 17].

政策不確実性指数をテキストを用いて構築した事例として Baker ら (2016) の研究があり [Baker 16], 彼らは事前に定められた政策や不確実性に関する単語群を用いて、それらを含む記事の数をカウントする手法を用いた。月次で集計し、季節調整を施している。Baker ら (2016) の研究の拡張として、いくつかの研究が存在しており、日本における政策不確実性指数やベルギーにおける政策不確実性指数などターゲットとなる国を変えた研究がある [Arbatli 17],[Manela 17]. これらの政策不確実性指数を対象にした分析を行った研究も複数ある [Brogaard 15],[Bachmann 13]. また、米国連邦公開市場委員会のページブックを対象にした不確実性指数を構築した事例や本研究と同様にトピックモデルを用いた事例がある。[Saltzman 18],[Azqueta-Gavald 叩 17].

3. データセット

本研究では、ニューステキストから不確実性指数の構築を行う。トピックモデルを用いたニュースの分類を行うが、教師ありトピックモデルを用いている。使用するデータセットはニューステキストと教師信号として数値データの2種類であらう。

3.1 テキストデータ

使用するテキストデータは日本語のロイターニュースを用いた。ロイターのウェブサイトのグローバル経済のカテゴリー下にある記事を抽出し、2009年8月から2019年1月までの期間

表 1: 不確実性に関する単語群

単語群
”不確実”, ”不確定性”
”不透明”, ”不透明感”
”不確定”, ”不明確”, ”不明瞭”

のにおける約 32,000 記事を得た。記事数は一日平均で 10 記事であり、1 記事につき、約 1,200 単語ほど含まれている。グローバル経済に含まれる記事は各国の経済イベントや政府の金融政策などについてのロイター記者が書いたものや業界のエコノミストが書いた記事などが含まれており、不確実性指数の構築の観点からカテゴリーを限定した記事を用いた。

3.2 数値データ

教師ありトピックモデルに用いるシグナルとして、VIX 指数を用いた。VIX 指数とは、米国の S&P 500 指数のオプションから計算される指数であり、シカゴ・オプション取引所が公表している。VIX 指数の数値が高いほど、市場参加者が相場の先行きに不透明感を持っていると解釈されている。

4. 処理

図 1 が全体の処理を示しており、インプットデータに対する前処理、トピックモデルによる分類、不確実性指数の構築の 3 つのパートからなっている。

4.1 テキストデータに対する前処理

まず、取得したロイター記事に対して、不確実性を示し記事のみの抽出を行う。これは事前に定義した単語群を用いて、記事内にそれらの単語群を含む記事のみを抽出した。(表 1) その後、Mecab を利用した形態素解析を行い、名詞のみの抽出を行った。前処理の結果、対象コーパスは 3,115 ドキュメントとなり、全単語数 2,945,703、単語種は 1,786 である。

4.2 数値データに対する前処理

sLDA の教師データに用いる数値データとして、VIX 指数を用いているが、平均 0、標準偏差 1 となるよう正規化を施した。各記事に対して、正規化後の VIX 指数を教師信号としてモデルのインプットに使用した。

4.3 トピック分類

トピック分類として、supervised LDA を用いており、supervised LDA については、次のセクションを参照。

4.4 指数構築

supervised LDA のトピック分類の結果、各ドキュメントに対するトピック分布が計算される。トピック分布の割合を月次の平均化を行い、全期間において平均 100 となるように正規化を行った。

5. トピックモデル

5.1 Supervised Latent Dirichlet Allocation

LDA は [Blei 03] により導入された有名なトピックモデルであり、各ドキュメントが複数のトピックに含まれており、各トピックは別々の単語分布からなるというアイデアに基づいてものである。その拡張として [Mcauliffe 08] が supervised latent Dirichlet allocation (sLDA) を導入し、LDA に各ドキュメントに対する教師信号となる変数を加えることで、ドキュメント

と教師信号両方に対応する潜在変数を推論できるモデルである。図 2 が sLDA のグラフィカルモデルであり、表 2 に各変数の定義を示す。本研究では、正規化した VIX 指数を各ドキュメントの教師信号としている。

表 2: sLDA におけるノテーション

記号	定義
α, β	ハイパーパラメーター
φ	単語分布
K	トピック数
θ	T 各ドキュメントにおけるトピック分布
Z	トピック潜在変数
w	単語
N	単語数
M	ドキュメント数
Y	教師信号
μ, σ	教師信号のパラメーター

6. 結果

6.1 トピック分類

以下のパラメータセットで sLDA モデルによるトピック分類を行った。 $\alpha = 0.035, \beta = 0.01, K = 6$ 。トピック分布を表 3 に示す。トピック数 6 のうち解釈可能なものを 4 つ選択した。TOPIC 4 は米国に關係するマクロ経済不確実性指数、TOPIC 0 は日本に關係するマクロ経済不確実性指数、TOPIC 2 は中国に關係するマクロ経済不確実性指数、TOPIC 5 は欧州に關係するマクロ経済不確実性指数と解釈することができる。

6.2 マクロ経済不確実性指数の推移

図 3 に構築した 4 つのマクロ経済不確実性指数の推移を示す。日本マクロ経済不確実性指数に関しては、2013 年 1 月の物価目標時に大きく落ち込み (図 3, A)、その後 2014 年 10 月の追加緩和のアナウンス時 (Fig 3, B) 及び 2016 年 1 月のマイナス金利導入のアナウンス時に大きく上昇している (図 3, C) 中国マクロ経済不確実性指数に関しては、2013 年中ほどに年率の成長率が 10% から 7% 台に低下した時期 (図 3, D)、および 2018 年中ほどに日中貿易摩擦による今後の中国の成長率への影響が懸念された時期に大きく上昇している (図 3, E)。米国マクロ経済不確実性指数に関しては、米国の量的緩和策終了時に大きく上昇している。2011 年 6 月に QE2 が、2013 年 12 月に QE3 の終了しているが、それぞれの直前に上昇している。(図 3, F, G)。また、直近では 2018 年の中ほどに米中貿易摩擦による米国経済への影響が懸念されており、指数が上昇 (図 3, H)。

欧州マクロ経済不確実性指数に関しては 2010 年の欧州危機 (図 3, I)、2011 年、および 2015 年のギリシャ危機 (図 3, J, K) において指数が大きく上昇している。

6.3 他の指標との相関分析

表 4 において、構築した 4 つの不確実性指数と各国 10 年金利 (米国、日本、中国、ドイツ、ギリシャ)、株価指数 (S&P500、日経 225、上海総合株価指数)、ドル円レートのそれぞれのボラティリティ、および教師信号の VIX 指数との相関を示した。

これを見るに、不確実性指数と関係のあるマーケット指標のボラティリティが他に比べ高い結果となっている。中国マクロ

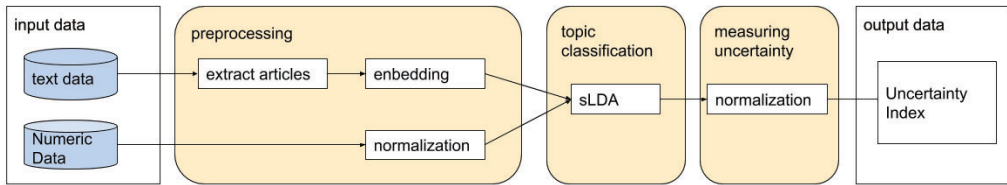


図 1: 処理の全体像

表 3: 各トピック-単語分布における上位 10 単語

TOPIC 4	TOPIC 0	TOPIC 2	TOPIC 5
FRB	日銀	中国	不透明
金融政策	政府	不透明感	問題
インフレ率	日本経済	輸出	eu
経済	日本	回復	欧州連合
インフレ	都内	企業	合意
米連邦準備理事会	経済	gdp	不透明感
利上げ	程度	増加	実施
内容	効果	伸び	欧州
総裁	期待	減少	ユーロ圏

表 4: 不確実性指数とマーケット指数のボラティリティとの相関

	VIX	10 年金利					株価指数			FX
		US	JP	CH	DE	GR	Nikkei	SH	S&P	
Japan	-0.087	-0.047	0.259	-0.158	0.059	0.136	0.335	0.194	0.073	0.152
China	0.003	0.099	0.053	0.106	0.088	0.177	-0.028	0.108	0.069	-0.054
US	-0.335	-0.343	-0.031	-0.326	-0.300	-0.203	-0.170	-0.175	-0.335	-0.141
EU	0.318	0.296	-0.158	0.317	0.369	0.338	-0.028	-0.115	0.249	0.011

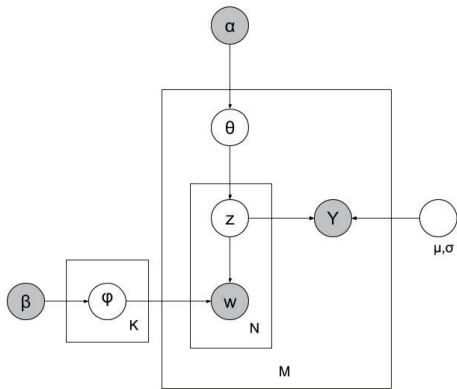


図 2: sLDA のグラフィカルモデル

経済不確実性指数は、中国の 10 年金利、上海総合株価指数との相関が高く、日本マクロ経済不確実性指数に関しては、日本の 10 年金利と日経平均との相関が高い。また、欧州マクロ経済不確実性指数に関しては、ドイツの 10 年金利とギリシャの 10 年金利との相関が高い。この結果から各国のマクロ経済不確実性

指数は各国のマクロ経済イベントを反映された指標と言える。

7. 今後の課題

本研究では、日本語のロイターニュースのみを対象にマクロ経済不確実性指数の構築を行ったが、今後はニューステキストの拡大や言語を日本語のみならず英語を加えた場合の分析等を行う予定である。

参考文献

[Arbatli 17] Arbatli, E. C.: Policy uncertainty in Japan, Technical Report 17, International Monetary Fund (2017)

[Azqueta-Gavaldn 17] Azqueta-Gavaldn, A.: Developing news-based Economic Policy Uncertainty index with unsupervised machine learning, *Economics Letters*, Vol. 158, pp. 47–50 (2017)

[Bachmann 13] Bachmann, R., Elstner, S., and Sims, E. R.: Uncertainty and economic activity, *American economic journal*, Vol. 5, No. 2, pp. 217–249 (2013)



図 3: マクロ経済不確実性指数

[Baker 16] Baker, S. R., Bloom, N., and Davis, S. J.: Measuring economic policy uncertainty (2016)

[Blei 03] Blei, D. M., Ng, A. Y., and Jordan, M. I.: Latent dirichlet allocation (2003)

[Brogaard 15] Brogaard, J. and Detzel, A.: The Asset-Pricing Implications of Government Economic Policy Uncertainty, *Management Science*, Vol. 61, No. 1, pp. 3–18 (2015)

[Castelnuovo 17] Castelnuovo, E. and Tran, T. D.: Google It Up! A Google Trends-based Uncertainty index for the United States and Australia, *Economics Letters*, Vol. 161, pp. 149–153 (2017)

[Chuli 17] Chuli, H., Guilln, M., and Uribe, J. M.: Measuring uncertainty in the stock market, *International Review of Economics and Finance*, Vol. 48, pp. 18–33 (2017)

[Hamid 15] Hamid, A.: Forecasting volatility with empirical similarity and Google Trends, *Journal of economic behavior & organization*, Vol. 117, pp. 62–81 (2015)

[Manela 17] Manela, A. and Moreira, A.: News implied volatility and disaster concerns, *Journal of Financial Economics*, Vol. 123, No. 1, pp. 137–162 (2017)

[Mcauliffe 08] Mcauliffe, J. D. and Blei, D. M.: Supervised topic models, in *Advances in neural information processing systems*, pp. 121–128 (2008)

[Saltzman 18] Saltzman, B. and Yung, J.: A machine learning approach to identifying different types of uncertainty, *Economics Letters*, Vol. 171, pp. 58–62 (2018)