

接触履歴を用いた地方景況感と業種間構造の分析

Analysis of Local Business Confidence indices and Inter-Industry Relations using Contact Histories

坂地泰紀^{1*} 蔵本涼太¹ 和泉潔¹
松島裕康¹ 島田尚¹ 砂川恵太²

Hiroki Sakaji¹ Ryota Kuramoto¹ Izumi Kiyoshi¹
Hiroyasu Matsushima¹ Takashi Shimada¹ Keita Sunakawa²

¹ 東京大学

¹ The University of Tokyo

² 沖縄銀行

² The Bank of Okinawa

Abstract: 自然言語処理の発展に伴い、膨大なテキストデータから有益な情報を抽出することが可能となりつつある。テキストデータを活用した分析には、その更新頻度の高さやリアルタイム性から金融経済情勢を迅速かつ的確に把握できるという強みがある。そのため、既存指標を補完する速報性の高い指標生成はテキストマイニングとの親和性が高い。本論文では、日次で蓄積されるテキストから景況指標を生成する手法を提案した。そして、実際に沖縄県を対象に景況指標を生成し、既存指標を高い精度で再現していることを確認した。加えて、得られた景況指標から業種間における景況感の連動性を分析した。

1 はじめに

経済・金融において、種々のデータは将来のマーケット動向を予測し、その分析を行うために用いられている。一般的に、マーケット動向の予測手法は扱うデータの差異からテクニカル分析とファンダメンタル分析に大別される。前者は、将来の市場は過去の取引実績の時系列データから予測可能であるという立場の下、過去の取引価格や出来高を入力データとして予測を行う。一方で後者は、テクニカル分析で活用されるデータに加え、該当企業の業績予想など幅広い情報を利用して将来予測を図る手法である。投資家やアナリストは両手法を組み合わせて活用してきたが、その扱い易さからいずれも数値データのみにも焦点を当てることが主であった。しかしながら、数値情報以外にもマーケット分析に有益なデータは存在する。その代表例がテキスト情報である。我々は日々様々なテキストから情報を得ている。ファイナンスの分野においても、専門家によって書かれたレポートには経済分析に関する重要な知見が豊富に含まれているであろうし、新聞記事は過去の出来事やその影響を報じている。また、SNSにおけるコメントにも人々の経済に対する印象が反映されている

と考えることができる。これらテキスト情報を将来動向予測に役立てようとする動きが進んでおり、機械学習を活用した様々な研究が提案されている [Bollen 11]。

地方銀行もまた貴重な金融テキストデータを持っている。例えば、業務日報や決算概況、接触履歴などがある。業務日報や接触履歴の中には、日々の業務の中で行員が感じた些細な変化や気づきが含まれている場合がある。さらに、接触履歴には業種に関する情報も付与されている場合もある。そこで、本研究では、地方景況感と業種間構造の分析を接触履歴を用いて行う。

2 関連研究

機械学習によるインデックス生成に関する研究として、山本らの研究 [山本 16] と余野らの研究 [余野 17] がある。山本らは LSTM を用いて景気ウォッチャー調査を学習し、一例として学習したモデルを用いて月例経済報告 (<http://www5.cao.go.jp/keizai3/getsurei/getsurei-index.html>) を指標化している。ここで、月例経済報告は日本政府が毎月発行する経済レポートである。また、余野らは山本らと同様に景気ウォッチャー調査で学習したモデルを用いて、月例経済報告から指数を生成している。彼らは LDA を用いることにより、個人消費や公共投資、物価などのトピックごとに指標

*連絡先：東京大学大学院工学系研究科
〒113-8656 東京都文京区本郷 7-3-1
E-mail: sakaji@sys.t.u-tokyo.ac.jp

を生成することで、どのファクターが影響しているかを調査できるような手法を提案している。これらの研究では、景気ウォッチャー調査と月例経済報告という類似した文書を利用しているのに対して、我々は接触履歴というテキストを利用している点が異なる。

極性付与に関する研究として、Sakaji et al. の研究 [Sakaji 08] がある。Sakaji et al. は新聞記事から景気動向の根拠となる表現を抽出し、抽出した根拠となる表現に極性を付与する手法を提案している。Turney [Turney 02] は、“excellent” や “poor” などの語を使って、Positive・Negative フレーズを抽出し、それを用いて文書分類を行っている。Pang et al. [Pang 02] は機械学習を用いて極性分類を行う手法を提案している。これらの研究とは異なり、本研究では、地域景況インデックスを生成することを目的にしている。

他にも関連研究として、迫村らの研究 [迫村 13] や山澤の研究 [山澤 18] がある。迫村らは Twitter データを用いて経済動向を分析する研究を行っている。山澤は景気ウォッチャー調査を対象に、単語と景気の現状判断 DI の相関を調べ、さらに主成分分析をすることで、どの単語が景気の現状判断 DI に有効かを調べている。それに対して、我々は主成分分析などの統計的手法ではなく、機械学習を利用することで地域景況インデックスを生成している。

また、文や文書に極性を付与するための辞書を生成する研究として、Ito et al. [Ito 18] の研究がある。Ito et al. は文の極性を判定しながら、辞書も生成する特殊なニューラルネットワークを構築する手法を提案している。加えて、Kaji et al. [Kaji 07] は、人手で作成した手ごかり表現リストやパタン、規則から評価文を抽出して極性 (好評, 不評) を付与し、各極性中に評価表現 (名詞+格助詞+形容詞) が出現する頻度から評価表現の極性を決定する手法を提案している。Kanayama et al. [Kanayama 06] は、自動的に辞書を拡張する手法を提案している。

3 提案手法

本節では、我々が提案する解析手法の概要について説明する。本手法はまず、Bidirectional long short-term memory (BiLSTM) [Graves 05] を用いて地方景況感を算出する。その後、地方景況感と以下に示すいくつかの手法を用いて、業種間構造の分析を行う。

Step 1: インプットデータとして、景気ウォッチャー調査を用いて BiLSTM モデルを学習する。その後、学習済みモデルを用いて接触履歴に極性を付与し、地方景況感を生成する。

Step 2: 生成した地方景況感に対して、グレンジャー因果分析、インパルス応答分析を用いて業種間構造を分析する。

3.1 地方景況感生成モデル

本節では、地方景況感の生成に用いた BiLSTM モデルについて述べる。我々は、Recurrent Neural Network (RNN) の一種である LSTM [Hochreiter 97] を機械学習手法として採用した。BiLSTM への入力には、分散表現を用いる。分散表現には 2003 年から 2008 年と 2013 年から 2014 年までのロイターニュース記事と、2014 年から 2017 年までの琉球新報記事を用いる。事前調査でロイターニュース記事だけで分散表現を作ったところ、地域景況インデックス生成において不十分な性能であったため、琉球新報記事を追加した。分散表現生成には gensim¹ を、形態素解析には MeCab² を用いた。分散表現は、次元数は 200 次元、5-window で生成した、分散表現に変換した入力に対しては、LSTM (\overleftarrow{LSTM} , \overrightarrow{LSTM}) を通して $\{\overleftarrow{h}_i\}_i^n$ と $\{\overrightarrow{h}_i\}_i^n$ を得る。その後、 \overleftarrow{h}_1 と \overrightarrow{h}_n と、0 と 1 で構成されたベクトルに変換した入力 Z を結合する。

$$s = [\overleftarrow{h}_1; \overrightarrow{h}_n; Z] \quad (1)$$

$$t = \tanh(W_s \cdot s + b_s) \quad (2)$$

$$u = \tanh(W_t \cdot t + b_t) \quad (3)$$

$$v = \tanh(W_u \cdot u + b_u) \quad (4)$$

$$Y = W_v \cdot v + b_v \quad (5)$$

ただし、 $Z \in \mathbb{R}^z$, $s \in \mathbb{R}^{2m+z}$, $t \in \mathbb{R}^{400}$, $u \in \mathbb{R}^{200}$, $v \in \mathbb{R}^{100}$. z はベクトルの次元数、 m は隠れ層のユニット数である。ここで、 W_s, W_t, W_u, W_v は重み行列、 b_s, b_t, b_u, b_v はバイアスベクトルである。 Y は $Y = (y_1, y_2, y_3)$ で構成される出力層である。最後に、出力層の中から最大の値を選び、それに対応する極性を入力に対する出力とする。BiLSTM モデルを図 1 に示す。詳細は、次の論文 [坂地 19] を参照されたい。

3.2 業種間構造分析手法

本節では、業種間構造を分析する手法について述べる。

3.2.1 グレンジャー因果分析

業種間の因果関係をグレンジャーの因果性を検定することで確認する。その具体的な過程は、帰無仮説として業種 A から業種 B へのグレンジャー因果性がないと

¹<https://radimrehurek.com/gensim/>

²<https://taku910.github.io/mecab/>

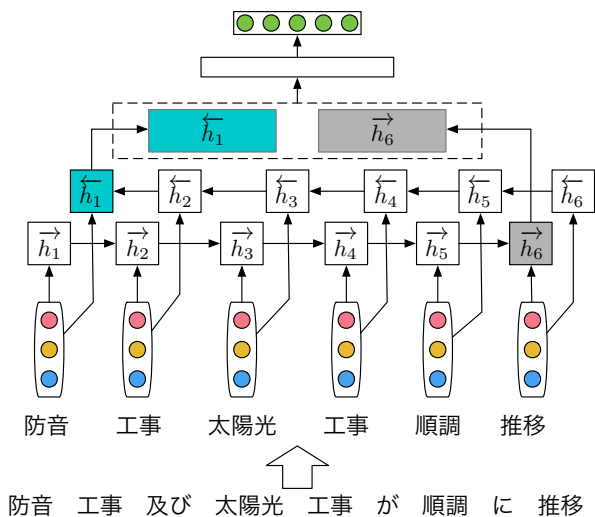


図 1: Bidirectional LSTM を用いての極性判定 [坂地 19]

仮定し、同仮説を棄却するか否かをその p 値と有意水準との比較により判断するというものである。本研究では、グレンジャー因果を多変量へ拡張した偏グレンジャー因果分析 [Guo 08] を採用した。偏グレンジャー因果分析は外生要因や潜在変数の影響を減じることが可能であり、多変量データ解析において有用である。

3.2.2 インパルス応答分析

グレンジャー因果検定は時系列データ間の因果性を判定する手法であるが、その関係の強さを測定することは出来ない。インパルス応答関数は、その関係性を定量的に捉える手段として用いられる。ある変数における変動が他の変数の変動に寄与する度合いを分析することで、定量的な解析が可能となるのである。本論文では、複数業種の景況指数が与えられたときに、ある業種での変動が他業種にどのような影響を及ぼすのか定量的評価を行うため、直交化インパルス応答関数 [Sims 80] を活用し分析する。

4 評価実験

本研究では、沖縄県に着目し、沖縄銀行の接触履歴を用いて地方景況感と業種間構造を分析する。

4.1 地方景況感生成実験に用いるデータ

我々は地方景況感生成実験に以下のデータを用いる。

表 1: 接触履歴の例

他行住宅ローン借換推進で訪問するが不在。名刺、チラシ投函する。
修学旅行も引続き順調であり来期も期待できる。新ホテル棟建設についてであるが投資金額が当初予定額より大幅に上回り、再度計画予定である。一応は9月までを目処としているが、建築確認等を考えると厳しいかもしれない。今後の展開を考え、＜組織＞内等にホテル等の売却案件があったら運営及び買取を行いたいので是非紹介をお願いしたい。
＜組織＞社長面談 現況確認 台風の影響で前期は売上減少 健康食＜組織＞による売上増加を見込むが動きなし 同氏の土地収用に伴う入金で借入金返済を見込む
キャッシュカード暗証番号変更でご来店。定期商品や保険を少しご案内しました。老後のことが不安との事なので時間があるときに検討する。

琉球新報 沖縄県の地方新聞紙として、琉球新報がある。本研究では、2014年から2017年の琉球新報を用いる。

接触履歴 接触履歴は、行員が顧客と何か知らのやり取りを行った際に記録されるデータであり、様々なことが記述されている。表1にその例を示す。

4.2 地方景況感生成実験設定

我々は、接触履歴、もしくは、琉球新報を用いて地方景況感を生成する実験を行う。琉球新報においては、文、段落、記事全体のそれぞれを用いて生成実験を行う。これは、新聞記事中のどの箇所が地方景況感を表しているのかを調べるためである。正解データとして、おきぎん経済研究所が発行している企業動向調査³を用いる。

5 結果と考察

表2に地方景況感生成実験の結果を示す。表2より、対象テキストの範囲変化による効果検証実験において、景況に関連のある記事・段落・文に大きな差異はないことが分かった。それに対して、接触履歴を用いた地方景況感生成は非常に高い精度で再現できている。その結果を図2を示す。

³http://www.okigin-ei.co.jp/report_DI.html

表 2: 地方景況感生成実験の結果

	相関係数
琉球新報：文単位	0.301
琉球新報：段落単位	0.275
琉球新報：記事単位（景気）	0.300
琉球新報：記事単位（全体）	0.175
接触履歴	0.856

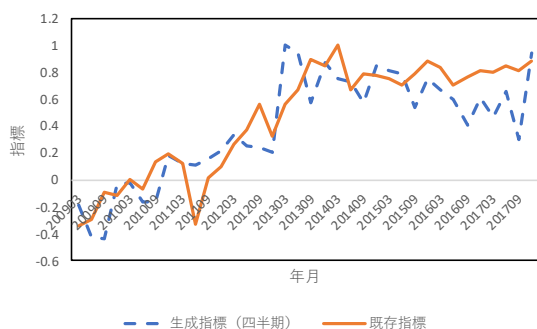


図 2: 接触履歴により生成した景況指標

次に、業種間分析の結果を図 3 と図 4 に示す。それぞれ、接触履歴から生成した地方景況感を用いて分析した結果である。沖縄県において、業種の影響力係数を大きい順に列挙すると、商業（小売業と卸売業）、建設業、サービス、製造業、医療・保健・社会保障・介護、不動産業の並びになる。そして、この並びは景況感変動の分析結果と結びつけて考えることが可能である。上位 3 つの業種、すなわち、商業（小売業と卸売業）と建設業、サービスは、図 3 の上流に位置する業種と対応している。また、図 4 に示すように、インパルス応答分析においても他業種に対する大きな変動要因になっている。そのような業種の景況感が向上することで、他業種に生産を喚起し、それに伴い他業種の景況感も向上する。この流れは理にかなったものであり、他業種への景況変動要因となる業種は生産波及効果の高い業種に対応することが確認された。

6 まとめ

本論文では、日次で蓄積されるテキストから地方景況感を生成する手法を提示した。そして、実際に沖縄県を対象に地方景況感を作成し、既存指標を高い精度で再現していることを確認した。加えて、得られた地方景況感から業種間における景況感の連動性を明らかにした。

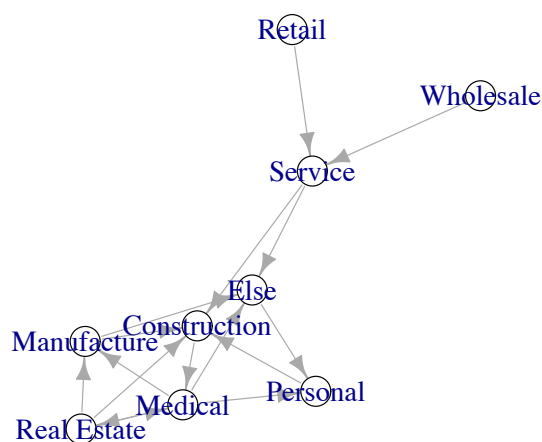


図 3: 偏グレジャー因果分析による業種間因果の可視化 [Sakaji 19]

謝辞

本研究は、三井情報株式会社と株式会社沖縄銀行、琉球新報社の協力により行うことができた。三社に感謝いたします。

参考文献

- [Bollen 11] Bollen, J., Mao, H., and Zeng, X.: Twitter mood predicts the stock market, *Journal of computational science*, Vol. 2, No. 1, pp. 1–8 (2011)
- [Graves 05] Graves, A. and Schmidhuber, J.: Frame-wise phoneme classification with bidirectional LSTM and other neural network architectures, *Neural Networks*, Vol. 18, No. 5, pp. 602–610 (2005)
- [Guo 08] Guo, S., Seth, A., Kendrick, K., Zhou, C., and Feng, J.: Partial Granger causality? Eliminating exogenous inputs and latent variables, *Journal of Neuroscience Methods*, Vol. 172, No. 1, pp. 79–93 (2008)
- [Hochreiter 97] Hochreiter, S. and Schmidhuber, J.: Long short-term memory, *Neural computation*, Vol. 9, No. 8, pp. 1735–1780 (1997)
- [Ito 18] Ito, T., Sakaji, H., Izumi, K., Tsubouchi, K., and Yamashita, T.: GINN: gradient interpretable neural networks for visualizing financial texts, *International Journal of Data Science and Analytics* (2018)

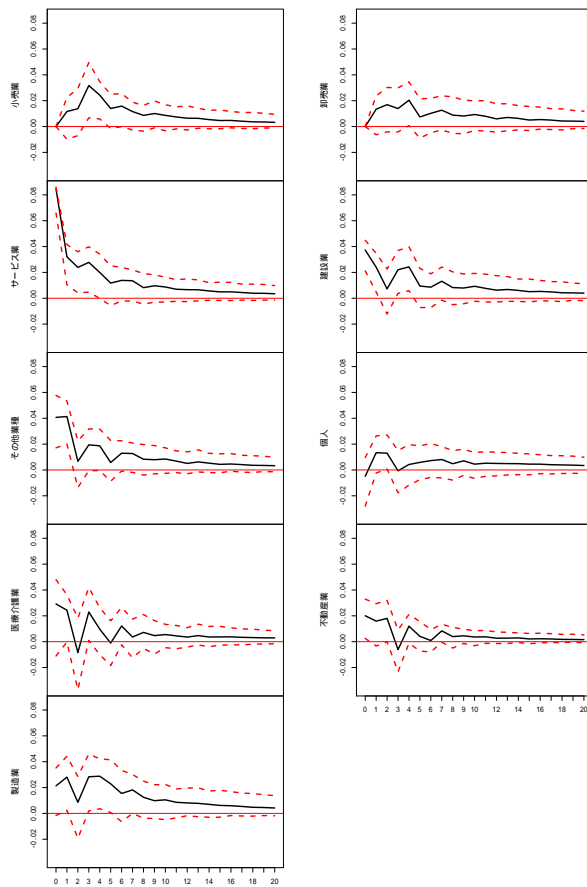


図 4: インパルス応答関数 (サービス業)

- [Kaji 07] Kaji, N. and Kitsuregawa, M.: Building Lexicon for Sentiment Analysis from Massive Collection of HTML Documents, in *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pp. 1075–1083 (2007)
- [Kanayama 06] Kanayama, H. and Nasukawa, T.: Fully Automatic Lexicon Expansion for Domain-oriented Sentiment Analysis, in *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, EMNLP '06*, pp. 355–363, Stroudsburg, PA, USA (2006), Association for Computational Linguistics
- [Pang 02] Pang, B., Lee, L., and Vaithyanathan, S.: Thumbs Up?: Sentiment Classification Using Machine Learning Techniques, in *Proceedings of the ACL-02 Conference on Empirical Methods in Natu-*

ral Language Processing - Volume 10, EMNLP '02, pp. 79–86, Stroudsburg, PA, USA (2002), Association for Computational Linguistics

- [Sakaji 08] Sakaji, H., Sakai, H., and Masuyama, S.: Automatic Extraction of Basis Expressions That Indicate Economic Trends, in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp. 977–984 (2008)
- [Sakaji 19] Sakaji, H., Kuramoto, R., Matsushima, H., Izumi, K., Shimada, T., and Sunakawa, K.: Financial Text Data Analytics Framework for Business Confidence Indices and Inter-Industry Relations, in *Proceedings of the First Workshop on Financial Technology and Natural Language Processing*, pp. 40–46, Macao, China (2019)
- [Sims 80] Sims, C. A.: Macroeconomics and reality, *Econometrica: journal of the Econometric Society*, pp. 1–48 (1980)
- [Turney 02] Turney, P. D.: Thumbs Up or Thumbs Down?: Semantic Orientation Applied to Unsupervised Classification of Reviews, in *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, ACL '02*, pp. 417–424, Stroudsburg, PA, USA (2002), Association for Computational Linguistics
- [坂地 19] 坂地 泰紀, 和泉 潔, 松島 裕康, 川瀬 和哉, 林 寛: 接触履歴を用いた地域景況インデックスの自動生成, *知能と情報*, Vol. 31, No. 2, pp. 626–635 (2019)
- [山本 16] 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, 2016 年度人工知能学会全国大会 (2016)
- [山澤 18] 山澤 成康: 計量テキスト分析による景気判断—コーディングルールや主成分を使った時系列分析—, *ESRI Discussion Paper Series* (2018)
- [迫村 13] 迫村 光秋, 和泉 潔: twitter テキストマイニングによる経済動向分析, 第 9 回人工知能学会フェイナンスにおける人工知能応用研究会 (2013)
- [余野 17] 余野 京登, 和泉 潔: 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, 2017 年度人工知能学会全国大会 (2017)