

単語ネットワークを用いた経済と感情に関わる表現の分析

村田 真樹^a 金子 徹^b 上東 嵩^c 馬 青^d

要約

Bollen らは、ツイッターでの感情分析を利用して、株価の予測がある程度可能であることを示した。われわれは、経済に関わる概念の関係を取得できる単語ネットワークの研究を行っており、感情に加えて経済に関わる概念関係の情報も利用して株価の予測を試みたいと考えている。その初歩的研究として、本稿では、経済と感情に関わる表現を、単語ネットワークにおいて分析した。実際にその分析を、「経済」「景気」「財政」「感情」に関わる単語を対象に行った。「経済」に関わる分析では、「中国」「首相」「大統領」が経済と関係が深く見えることがわかった。「景気」に関わる分析では、景気がよいと消費が増える、金利が上昇すると株価は下落すること、さらに国債に影響することを示唆する結果が得られた。「市場」「企業」「日銀」が景気と関係が深いこともわかった。「財政」に関わる分析では、消費を増やすために財政を利用していること、財政のために増税が必要ということ、財政の観点でギリシャ支援が必要であることを示唆する結果が得られた。「感情」に関わる分析では、「事件」「映画」「ドラマ」「戦争」「テロ」「判決」「受賞」「恋愛」など、感情と関係がある事柄を把握する際に役立つ単語群が得られた。今後は、この分析結果の情報も利用して、株価の予測に関わる研究を進めたいと思っている。

JEL 分類番号： C82

キーワード：単語ネットワーク，経済，感情，株価

*本研究は、公益財団法人石井記念証券研究振興財団の助成金を受けて実施された。

a 鳥取大学大学院工学研究科情報エレクトロニクス専攻、および、鳥取大学工学部附属クロス情報科学研究センター murata@ike.tottori-u.ac.jp

b 鳥取大学工学部電気情報系学科 s152032@ike.tottori-u.ac.jp

c 鳥取大学大学院持続性社会創生科学研究科工学専攻 s132018@ike.tottori-u.ac.jp

d 龍谷大学理工学部数理情報学科 qma@math.ryukoku.ac.jp

1. はじめに

Bollen らは、ツイッターでの感情分析を利用して、株価の予測がある程度可能であることを示した(Bollen et. el., 2011)。われわれは、経済に関わる概念の関係を取得できる単語ネットワークの研究を行っており、感情に加えて経済に関わる概念関係の情報も利用して株価の予測を試みたいと考えている。その初歩的研究として、本稿では、経済と感情に関わる表現を、単語ネットワークにおいて分析した結果について述べる。

2. 単語ネットワーク

単語ネットワークでは、最初に与える単語のことを、テーマキーワードと呼ぶ。与えたテーマキーワードと関係の深い単語を芋づる式に取り出して、単語ネットワークを形成する。関係の深い単語の取り出しには同じ文書に同時に出現しやすい単語の組を利用する。単語ネットワークの作成手順を以下に示す（文献(Kamihigashi et. el., 2017)）。

手順 1

構築したいネットワークの主となる概念を、テーマキーワードとして設定する。

手順 2

データ A から、テーマキーワードを含む記事群 B を取り出す。

手順 3

記事群 B に形態素解析を用い、ネットワークのノード候補となる単語を抽出する。その際に 1 文字、数字のみ、ひらがなのみの単語を除外する。

手順 5

得られたノード候補の中から、*TF-IDF* の値を用いて、その値が最も大きい 5 単語を次のノードとする。（*TF-IDF* は、単語と文書の頻度から求めることができる値であり、重要な単語を取り出す際に役立つ文書検索技術で有名な手法である。）

手順 6

テーマキーワードと次のノードとなった単語を含む記事を、データ A から取り出す。これを記事群 B として、手順 3 に戻り、処理を繰り返す。

3. 実験と考察

2 節で述べた単語ネットワークを実際にいくつか作成した。データ A には、2007 年から 2016 年の毎日新聞（「CD-毎日新聞データ集」）を用いた。テーマキーワードとして、「経済」「景気」「財政」「感情」を利用した。それぞれのテーマキーワードで作成した単語ネットワークを以下に示すとともに、それぞれのネットワークを考察する。

3.1. 「経済」のネットワーク

「経済」のネットワークを図1に示す。

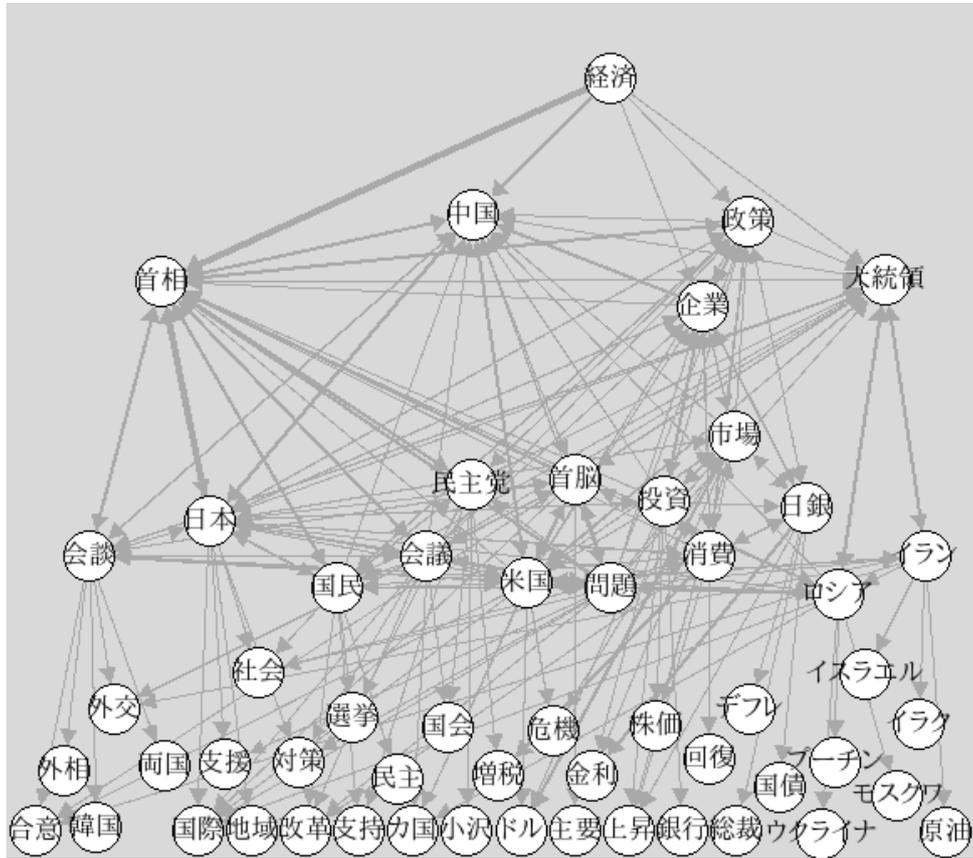


図1 「経済」のネットワーク

「経済」の単語に対して比較的早く「中国」がつながっている。中国の影響力がわかる。また、「首相」や「大統領」や「政策」も経済と関係があることがわかる。

3.2. 「景気」のネットワーク

「景気」のネットワークを図2に示す。

「景気」から「消費」につながっている。景気がよいと消費が増えるがそのことを示唆しているように見える。「金利」と「株価」もつながっており、金利が上昇すると株価は下落するがそれを示唆しているように見える。さらに、左下に「国債」があり、国債にも影響することを示唆しているように見える。「景気」から「市場」「企業」「日銀」にもつながっており、「景気」が「市場」「企業」「日銀」と関係が深いこともわかる。

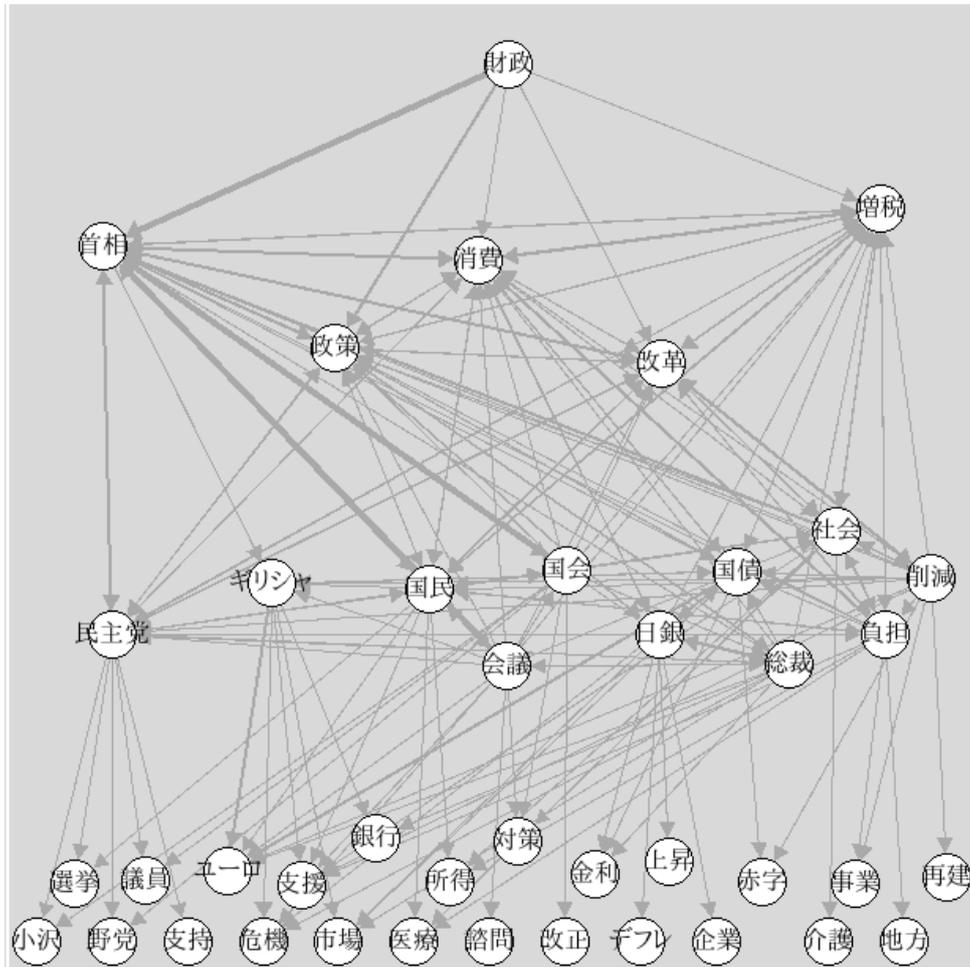


図3 「財政」のネットワーク

なものとして、「戦争」「テロ」「判決」「受賞」「恋愛」なども得られている。図4の単語群は感情と関係がある事柄を把握する際に、役立つと思われる。

4. おわりに

本稿では、単語ネットワークを利用して、経済と感情に関わる概念を分析した。その分析を、「経済」「景気」「財政」「感情」に関わる単語を対象に行った。「経済」に関わる分析では、「中国」「首相」「大統領」が経済と関係が深く見えることがわかった。「景気」に関わる分析では、景気がよいと消費が増える、金利が上昇すると株価は下落すること、さらに国債に影響することを示唆する結果が得られた。「市場」「企業」「日銀」が景気と関係が深いこともわかった。「財政」に関わる分析では、消費を増やすために財政を利用していること、財政のために増税が必要ということ、財政の観点でギリシャ支援が必要であることを示唆する結果が得られた。「感情」に関わる分析では、「事件」「映画」「ドラマ」「戦争」

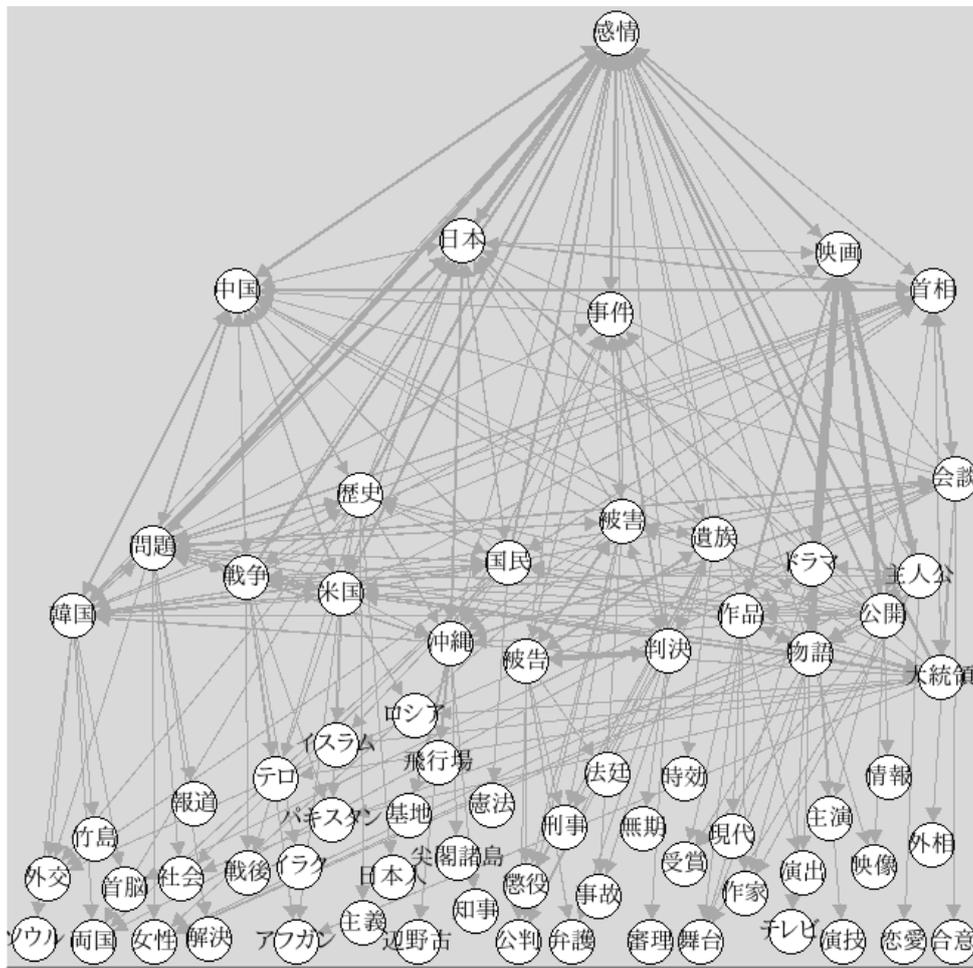


図4 「感情」の単語ネットワーク

「テロ」「判決」「受賞」「恋愛」など、感情と関係がある事柄を把握する際に役立つ単語群が得られた。今後は、この分析結果の情報も利用して、株価の予測に関わる研究を進めたいと思っている。

引用文献

- Bollen, J., H. Mao, X. Zeng, 2011. Twitter Mood Predicts the Stock Market, *Journal of Computational Science*, Volume 2, Issue 1, pp. 1-8.
- Kamihigashi T., M. Murata, Q. Ma, 2017. Use of Web Search Engines in TF-IDF based Word Network Construction for Extracting Useful Information, *Joint 17th World Congress of International Fuzzy Systems Association and 9th International Conference on Soft Computing and Intelligent Systems (IFSA-SCIS 2017)*, pp. 1-4.