

# テキスト情報を用いた金融市場分析の試み

## New Approach on Market Analysis using Text Mining

和泉 潔\*<sup>1</sup>      後藤 卓\*<sup>2</sup>      松井 藤五郎\*<sup>3</sup>  
Kiyoshi Izumi      Takashi Goto      Tohgoroh Matsui

\*<sup>1</sup>産業技術総合研究所  
National Institute of Advanced Industrial Science and Technology

\*<sup>2</sup>三菱東京 UFJ 銀行  
Bank of Tokyo-Mitsubishi UFJ

\*<sup>3</sup>東京理科大学  
Tokyo University of Science

In this study, we proposed a new approach on analysis of Japanese government bond (JGB) market using text-mining technology. First we extracted principal components from monthly financial reports of Bank of Japan (BOJ). Then, trends of the JGB market were inferred by regression analysis using the principal components. As a result, determination coefficients were 70-89%, and market trends were explained well by the information that was extracted from the text data.

### 1. はじめに

テキストマイニングによる市場環境分析手法の研究として、金融機関と共同でテキストマイニング技術を用いた国債市場の動向分析を試みた [三菱東京 UFJ 銀行 07]。経済情報に関するテキストデータを説明変数とし、債券価格の時系列データという被説明変数を分析するための手法を考案した。その結果、経済的に意味付け可能な各主成分が抽出され、債券市場に対して一定の説明力を持つ可能性が高いことを示した。以下に分析法および分析結果の詳細を示す。

### 2. 金融市場とマイニング

金融市場では常に様々な情報が溢れている。トレーダー達は、市場に影響を及ぼす多様な情報を取捨選択し、現在の市場の状況を分析・予測している。市場の分析に用いる情報には大きく分けて 2 種類がある。一つは、経済指標、マーケットのテクニカル指標等の数値情報である。もう一つは、市場に関わる要人の発言、中央銀行や他の市場参加者の解析記事などのテキスト情報である。これらの多様な情報が瞬時にトレーダー達のもとに、オンラインで送られてきているのである。送られてきた情報の全てを、現場のトレーダーが自分で目を通して市場分析に用いることは不可能に近い。そのため、いくつかの情報技術を市場分析に適用する研究が行われてきた。例えば、数値情報を用いて現在の市場情報を推論するようなエキスパートシステムの構築を行う研究 [日本 93] やニューラルネットや遺伝的アルゴリズムを数値情報による市場分析に用いた研究もある [電気 02]。これらの研究は一定の成果をあげてきた。しかし、数値情報には指標化されていない情報が多量に含まれていないので、分析対象の範囲がテキスト情報よりも狭くなる可能性がある。しかも、指標を集計して発表するには、どうしてもタイムラグが生じてしまうので、分析への反映も遅れがちである。近年、テキスト情報による市場分析に関して、ライターなどのオンラインの経済ニュースに対する市場の反応を推測する研究もでてきた [Seo 04, Ahmad 05, 高橋 07]。これらの研究は、1 日以内や数日の短期的な市場の反応を分析対象

としており、より長期的な市場動向の分析には用いられてこなかった。そこで、我々はオンラインのテキスト情報から、数年にわたる比較的長期の市場動向の変化を分析するための補助を目的とした解析技術を開発した。こういった観点から、市場参加者が特に注目する日本銀行の金融経済月報を題材に、テキストマイニング技術を用いて経済市場分析を試み、また金融経済月報が実際の金利動向をどの程度説明しているのかについて検証を行った。

### 3. テキストデータによる時系列データの推定

本研究ではテキスト分析の対象として、日本銀行が毎月中旬にサイト上で発表している金融経済月報 [日本] を題材として、図 1 にある下記の 3 つのステップからなる解析技術を提案する。

1. 共起関係に基づく主要単語の抽出と可視化
2. 主成分分析による単語のグループ化
3. 重回帰分析による金利データの動向分析

金融経済月報は、日本銀行が金融・経済情勢を分析した資料である。金融政策の方針を決める金融政策決定会合で内容を審議し、政策決定の基礎資料とする。この情報によって、日本銀行が、当面の経済動向をどう分析しているか対外的に明らかにしている。今回、金融経済月報を分析対象にした理由は 3 つある。第一に金融経済月報は、実際の金融市場のトレーダーが多かれ少なかれ着目している共有の重要テキスト情報であるからである。第二の理由は、会員制の有料マーケットリポート等のテキスト情報と違って、毎月の中旬にサイト上で定期的に発表されていて、誰でもアクセス可能な情報であることである。三番目の理由は、ブログ等のほとんど決まった形式のないテキスト情報と異なり、解説内容の順番や段落構成等がほぼ定式化されていて、月ごとのテキスト内容の変化が比較しやすいからである。本研究では、このように分析に好条件であると思われるテキスト情報を用いたが、今後は本手法をマーケットリポートやブログ等のより条件の厳しい情報に適用を試みる予定である。

A: 和泉 潔, 産業技術総合研究所 デジタルヒューマン研究センター, 〒 135-0064 東京都江東区青海 2-41-6, kiyoshi@ni.mints.ne.jp

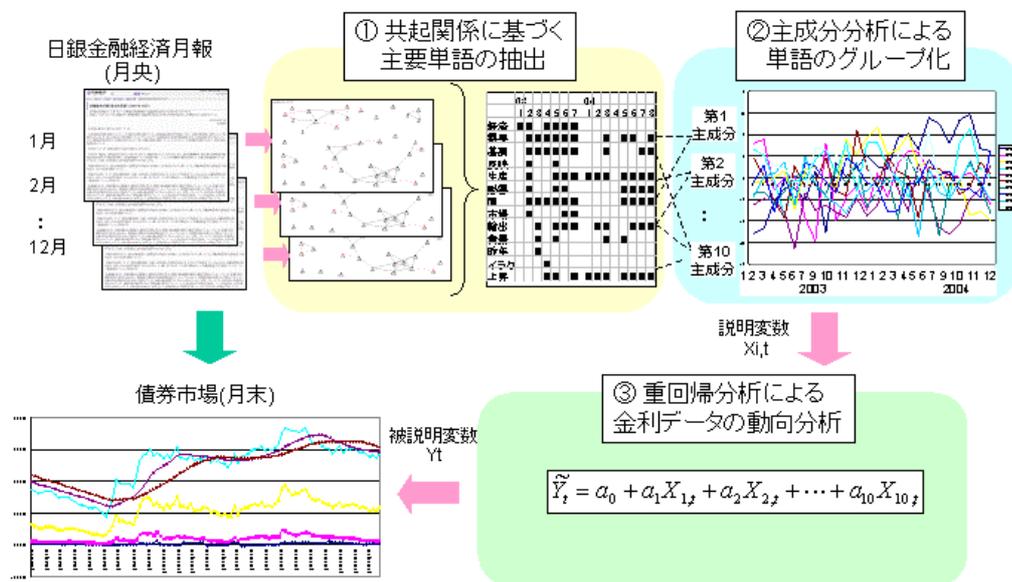


図 1: テキストデータによる時系列データの推定手法

### 3.1 共起関係に基づく主要単語の抽出と可視化

最初に、各月のテキストデータに KeyGraph [大澤 06] を適用し、共起関係を解析した。具体的にはまず、日本語形態素解析システムである Chasen [ChaS] による形態素解析を行い、出現頻度順に名詞・動詞・形容詞等を抽出した。次に、Jaccard 係数 ( $= p(A \text{ and } B) / p(A \text{ or } B)$ ; ただし A, B は抽出した単語) を段落毎に適用し、段落毎に同時に出現する単語と単語を繋ぎ、共起グラフを作成する。その後、単結合 (A, B 間のみの結合部分) を切断し、結合による「島」を作成する。またその後、各単語間の共起度に基づき、上位順に「橋」を作成する。

### 3.2 主成分分析による単語のグループ化

Keygraph に出現した単語のパターン (単語を毎月の出現状況に従いパターン分類したもの) に対し主成分分析を実施し 10 個の合成変数 (主成分) にまとめる。主成分分析の際には、単語に関して品詞を区別せずに分析を実施する。ここで注意してほしいのは、ここまで金利データは全く用いず、純粋に単語の出現パターンのみでの分析を行っていることである。つまり、ここまでの分析は、債券市場や株式市場、外国為替市場などの分析対象となる市場の種類に依存せずに、共通であるということである。

### 3.3 重回帰分析による金利データの動向分析

最後に、各主成分の毎月の動きから月次での金利の動きを解析する。具体的には、さきほどの 10 個の主成分の時系列データを説明変数として、月次の市場データを被説明変数とする重回帰分析を行う。ここでは、月中に発表される金融経済月報を用い、月末の金利の説明を行わせることとする。

## 4. 金融経済月報のテキストマイニング

上述の手法を用いて、国債価格の変動が比較的大きかった 2 つの時期 (2003 年 1 月-2004 年 12 月, 2006 年 1 月-2007 年 10 月) を分析した。

### 4.1 2003 年-2004 年の金融経済月報解析

この期間は、日本経済が底を打ち上昇に転じたとされる節目となった年である。こういった転換の前後において、金融経済月報にどのような変化が生じたのかを解析する目的で、この期間を採用した。単語について分析してみると、例えば「改善」という単語は 2003 年 8 月以降、2004 年 1 月に 1ヶ月だけ途切れたが、その後 2004 年 12 月 (観測期間の最後) まで抽出され続けた。「改善」という単語は金利が底打ちした 2003 年 6 月に浮上し、7 月には共起ネットワークが強化された様子が見えた。これは実際の金融実務者の実感とも近いものであった。ここでは詳細な分析はしていないが、「国内」「輸出」といった分野を表す単語と「改善」「厳しい」といった状態を表す単語を分類し関連性を追いかけることで、「どんな」分野が「どのような」状態にあるのかを定義できるかもしれない。回帰による推定結果を見ると、この 2 年間はゼロ金利期間だったことから、特に  $-0.01 \sim 0.07\%$  で推移した 1 年物の債券価格に対する説明力が弱かった (決定係数  $R^2: 22\%$ ) が、10 年物に関しては一定の説明力 ( $R^2: 78\%$ ) を持つことができた。

### 4.2 2006 年-2007 年の金融経済月報解析

2006 年は、「回復」「改善」といった単語が終了し、「拡大」「高い」といった単語が出現した年である。つまり景気回復局面から景気拡大局面へと変化した年であった。主成分分析の結果、各主成分における上位の単語が、それぞれある経済の一分野を表している可能性を示す状況が見られた。例えば、第 1

主成分については「全体」「海外」が上位に見られることから、「全体及び海外の状況」の分野を示し、第2主成分では「投資」「市況」が上位に見られることから「株式、為替他市場の状況」、同様に見ていくと第3主成分では「価格動向」、第4主成分「金融環境」、第5主成分「商品市況」、第6主成分「輸出動向」、第7主成分「雇用状況」…といった分野の状況を示している可能性がある。これは、共起解析を段落毎に行っているため、段落毎に各分野についてコメントされている内容が反映され、各分野において主要な位置を占める単語が主成分として上位に分けられやすいことことを示していると考えられる。この時期の債券市場は、2006年にゼロ金利が解除され、金利はゼロから離れ通常の水準に戻る過程にあった。回帰による推定結果を見ると、2003,2004年とは逆に、1年物(JGB 1Y)が決定係数89%と最も説明力が高く、2年物(JGB 2Y),5年物(JGB 5Y),10年物(JGB 10Y)と長期になるにつれ決定係数が低くなり、10年物では決定係数は70%となった(図2)。通常の金利環境下においては、金融経済月報はより短期金利に影響を与えている、という結果となった。

## 5. まとめ

2度のテキストマイニングを通じ、金融経済月報は市場金利に対して一定の説明力を持つ可能性が高いことを示した。しかし、同手法で外挿テストを行った場合には説明力がかなり低下した。その原因を分析したところ、テキストのパターンと債券価格の関係が時期によって動的に変化していることが主な要因となっていることが分かった。つまり、情報の構造を静的に抽出するのではなく、市場現象の動的な過程を考慮しながら分析する必要性が明らかになった。

## 参考文献

- [Ahmad 05] Ahmad, K., Gillam, L., and Cheng, D.: Textual and Quantitative Analysis: Towards a new, e-mediated Social Science, in *Proc. of the 1st International Conference on e-Social Science* (2005)
- [ChaS] ChaSen ホームページ: <http://chasen.naist.jp/hiki/ChaSen/>
- [Seo 04] Seo, Y.-W., Giampapa, J. A., and Sycara, K.: Financial News Analysis for Intelligent Portfolio Management, Technical Report CMU-RI-TR-04-04, Carnegie Mellon University (2004)
- [高橋 07] 高橋 悟, 高橋 大志, 津田 和彦: 株式市場におけるヘッドラインニュースの効果についての研究, ファイナンス学会第15回大会, pp. 373-383 (2007)
- [三菱東京UFJ銀行 07] 三菱東京UFJ銀行 円貨資金証券部: 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), [http://www.bk.mufg.jp/report/focus2007/FotM\\_24.pdf](http://www.bk.mufg.jp/report/focus2007/FotM_24.pdf)
- [大澤 06] 大澤 幸生: チャンス発見のデータ分析 モデル化+可視化+コミュニケーション シナリオ創発, 東京電機大学出版局 (2006)
- [電気 02] 電気学会(編): 学習とそのアルゴリズム, 第6章, 森北出版 (2002)

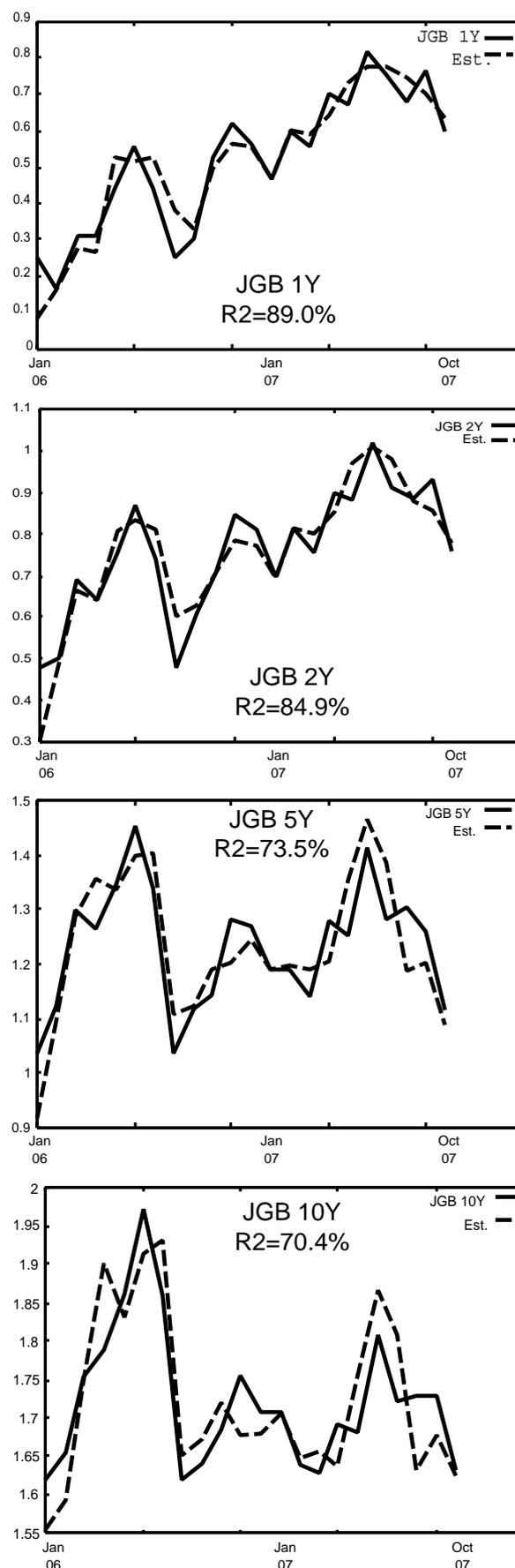


図 2: 2006年-2007年の債券市場の回帰結果

[日本] 日本銀行：金融経済月報 背景説明を含む全文：<http://www.boj.or.jp/type/release/teiki/gp/>

[日本 93] 日本ファジィ学会（編）：ファジィ・エキスパート・システム，日刊工業新聞社（1993）