

金融テキストマイニングの基づいた投資家支援プラットフォームの開発

Development of Platforms for Investors using Financial Text Mining

坂地 泰紀^{1*} 和泉 潔¹ 松島 裕康¹
Hiroki Sakaji¹ Kiyoshi Izumi¹ Hiroyasu Matsushima¹

¹ 東京大学大学院工学系研究科

¹ Graduate School of Engineering, The University of Tokyo

Abstract: 近年、人工知能分野の手法や技術を、金融市場における様々な場面に应用することが期待されており、例えば、膨大な金融情報を分析して投資判断を支援する技術が注目されている。特に、個人投資家・機関投資家を問わず、大量の情報の中から投資に有用な情報を抽出し、提示するシステムが求められている。そこで、我々は個人、もしくは、専門家が容易に新聞記事や決算短信などの金融テキストを分析できるテキスト解析プラットフォームを構築することを目指す。

1 はじめに

近年、人工知能分野の手法や技術を、金融市場における様々な場面に应用することが期待されており、例えば、膨大な金融情報を分析して投資判断を支援する技術が注目されている。さらに、最近では証券市場における個人投資家の比重が増大しており、個人投資家に対して投資判断の支援を行う技術の必要性が高まっている。そのような中、酒井ら [酒井 15] は個人投資家支援のため、企業の決算短信 PDF から業績要因を抽出し、それを検索できるシステム¹を公開している。また、日本経済新聞社は人工知能技術を用いて決算データを自動的に要約して出力する AI 決算サマリー²を開発・公開している。このように、人工知能技術を投資家支援に用いた研究・開発は盛んである。

また、特許分野においては、株式会社 NTT データ数理システムが Text Mining Studio³を特許文書などを解析するテキストマイニングプラットフォームとして開発している。さらに、株式会社プラスアルファ・コンサルティングは、特許文書を専門とした解析プラットフォームである見える化エンジン Paterio⁴を開発している。このように、金融分野だけではなく、特許分野などの他の分野においても、特許分析支援のために文書を解析するシステムが開発されている。

上記で紹介したシステムやプラットフォームのい

れも、単語や表現の頻度、もしくは、係り関係を元に分析するものとなっている。それに対して、我々が作成した原因・結果表現検索システム⁵は、論理関係の一つである因果関係に基づき、原因と結果を表す表現を決算短信 PDF から抽出し、それを検索できるシステムである [坂地 15]。システムを図 1 に示す。単語や表現の頻度だけではなく、因果関係を利用することで、単語頻度だけでは得られない情報を得ることができる。例えば、「猛暑」という語で検索すること、原因表現として「猛暑の影響」、結果表現として「除草関連用品、散水用品、日除け用品が好調に推移しました」を見つけることができる。この結果から猛暑になると、除草関連用品が売れることを知ることができる。この結果の背後には、「猛暑により雑草が大きく成長する」という隠れた関係がある。我々は、このような思いがけないような因果関係を抽出する手法も提案している [Sakaji 17]。

このような中、我々は因果関係の原因と結果をつなげることで因果チェーン構築の試みを行っている [西村 18]。本論文では、原因・結果表現の抽出や因果関係の判定、因果チェーン検索機能などを持ったテキスト解析プラットフォームの構想について述べる。

2 テキスト解析プラットフォーム

本節では、我々が想定しているテキスト解析プラットフォームについて述べる。本テキスト解析プラットフォームでは、因果関係の抽出と判定、さらには因果をリンクさせて因果チェーンを構築し、それを利用した

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

¹<http://hawk.ci.seikei.ac.jp/cees/>

²<http://pr.nikkei.com/qreports-ai/>

³<https://www.msi.co.jp/tmstudio/aboutTMS.html>

⁴<http://www.pa-consul.co.jp/mieruka/scene/paterio.html>

⁵<http://socsim.t.u-tokyo.ac.jp/CS/>

検索などの機能を備える予定である。現在は、決算短信と日経新聞記事を用いた因果チェーン検索システムの開発を試みている。加えて、景気ウォッチャー調査を学習したモデル [余野 17] を用いた極性判定 API や、プラットフォームとは異なるがロイターニュース記事を学習して生成した極性辞書 [Ito 18] も公開予定である。極性辞書に関しては、金融特化の辞書となっているため、金融テキストの分析においては有効的な結果を示すことができている。

さらに我々は、過去の研究で用いた手法を Python パッケージにして公開する計画を立てている。これにより、お手軽に金融テキストマイニングを試すことができるようになり、実務に応用しやすくなると期待している。



図 1: 決算短信における因果関係検索システム

3 まとめ

本論文では、テキスト解析プラットフォームの構想について述べた。特に因果関係抽出を用いた因果チェーンの構築について紹介した。いくつかのシステムは既に実装・公開しており、今後は完成次第、順次公開していく予定である。

4 今後の展望

今後は、金融分野のテキストだけではなく、特許や Web ブログや掲示板などのテキストデータからも因果関係を抽出し、因果チェーンを構築したい。さらに、因果関係だけではなく、仮定や目的などの他の論理関係も扱えるように拡張する予定である。そのように拡張することで、分野横断論理チェーン (図 2) が構築でき、新しく作られる製品の予想などができるようになると期待できる。

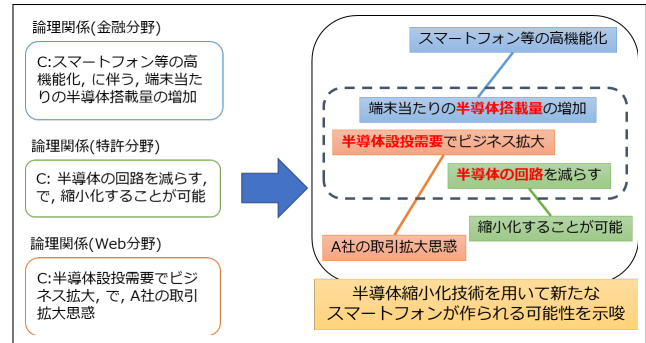


図 2: 分野横断論理チェーンの例

参考文献

- [Ito 18] Ito, T., Sakaji, H., Tsubouchi, K., Izumi, K., and Yamashita, T.: Text-Visualizing Neural Network Model: Understanding Online Financial Textual Data, in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp. 247–259 Springer (2018)
- [Sakaji 17] Sakaji, H., Muroto, R., Sakai, H., Bennett, J., and Izumi, K.: Discovery of Rare Causal Knowledge from Financial Statement Summaries, in *Proceedings of the 2017 IEEE Symposium on Computational Intelligence for Financial Engineering and Economics (CIFER)*, pp. 602–608 (2017)
- [坂地 15] 坂地 泰紀, 酒井 浩之, 増山 繁: 決算短信 PDF からの原因・結果表現の抽出, 電子情報通信学会論文誌 D, Vol. J98-D, No. 5, pp. 811–822 (2015)
- [酒井 15] 酒井 浩之, 西沢 裕子, 松並 祥吾, 坂地 泰紀: 企業の決算短信 PDF からの業績要因の抽出, 人工知能学会論文誌, Vol. 39, No. 1, pp. 172–183 (2015)
- [西村 18] 西村 弘平, 坂地 泰紀, 和泉 潔: 表現類似度を用いた因果ネットワークの構築, 2018 年度人工知能学会全国大会 (2018)
- [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)