

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 24 年 3 月 26 日現在

機関番号：13904

研究種目：若手研究（B）

研究期間：2009～2011

課題番号：21700158

研究課題名（和文）：投資判断支援のための業績要因自動抽出・分析手法の開発

研究課題名（英文）：Development of the method of extracting causal expressions from Japanese financial articles concerning business performance and analyzing them for supporting decision on investment

研究代表者

酒井 浩之（SAKAI HIROYUKI）

豊橋技術科学大学・大学院工学研究科・助教

研究者番号：70402659

研究成果の概要（和文）：

本研究では、経済新聞記事に含まれる企業業績に関する記事から、その業績要因を自動的に抽出し、さらに、抽出した業績要因に対して当該企業業績に対する重要度を自動的に付与することで、抽出した業績要因を投資判断の支援だけでなく、景気予測、技術動向予測などの用途に使用することのできる情報に変換することを目的とした。そして、業績要因表現を抽出する既提案手法を改善するために抽出した企業の Web サイトから当該企業にとって重要なキーワードを利用して業績要因に対して重要度を付与し、その重要度に基づいて最も重要な業績要因の自動判定を行った。評価の結果、本手法は 77.3% の正解率を達成し、業績発表記事において 2 番目に出現する文を最重要の業績要因として判定する Lead(s2) 手法や、業績発表記事において最初に抽出される業績要因を最重要の業績要因として判定する Lead(ce) 手法より高い正解率を達成した。

研究成果の概要（英文）：

We have developed a method of extracting causal expressions from Japanese financial articles concerning business performance of companies and assigning importance to them automatically. Causal expressions assigned importance are useful for investors in selecting companies to invest, business prediction, and technology-trends prediction.

Our method automatically extracted causal information as a form of causal expression by using statistical information and initial clue expressions. We extracted keywords from Web sites of the companies and improved the previous method by using them. Moreover, our method automatically determined the most important causal expression by using these keywords extracted from Web sites of the companies.

We evaluated our method and it attained 77.3% accuracy and the experimental results showed that our method outperforms Lead(s2) method which determines the sentence which appears in the 2nd in an article concerning business performance of companies as the most important causal expression and Lead(ce) method which determines a causal expression first extracted from an article concerning business performance of companies as the most important causal expression.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2009 年度	1,200,000	360,000	1,560,000
2010 年度	1,100,000	330,000	1,430,000
2011 年度	900,000	270,000	1,170,000

年度			
年度			
総計	3,200,000	960,000	4,160,000

研究分野：情報工学

科研費の分科・細目：情報学・知能情報学

キーワード：テキストマイニング，情報抽出，業績要因抽出，投資判断支援，技術動向予測，重要度付与

東芝10-12月、最終赤字が縮小、68億円-半導体リストラ奏功。

東芝が二十九日発表した二〇〇二年十一月十二月期連結決算は、最終損益が六十八億円の赤字だった。.....十一月十二月期の売上高は前年同期比八%増の一兆三千四百四十三億円。デジタルカメラの販売増やカメラ付き携帯電話の普及で、NAND型フラッシュメモリーなどが順調だった。.....

業績要因
(重要度:★★★★★)

1. 研究開始当初の背景

研究開始当初，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応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の支援をする技術が注目されていた。さらに，当時，個人投資家からの証券口座の開設の申し込みが急増しており，証券市場における個人投資家の比重が高まっていた。そのため，個人投資家に対して投資判断の支援をおこなう技術の必要性が高まっていた。そのような背景のもと，第22回人工知能学会全国大会において「ファイナンスにおける人工知能応用」のセッションが設けられ，また，4th IFIP International Conference on Artificial Intelligence Applications and Innovations (AIAI 2007) においては AI in Business and Finance のセッションが設けられるなど，注目が集まっていた。

適切な投資判断のために，企業の業績に関する情報を収集することは重要なことである。また，企業にとっても，同業他社の業績を分析することは企業経営において重要である。ここで，企業業績の分析において，業績情報と同等に，その業績要因（例えば「電子デバイスの売り上げが好調」）が重要である。なぜなら，業績拡大の要因が，その企業の事業が好調であるならば株価への影響は大きい，株式売却益の計上などの一時的な利益の計上が要因であるならば株価への影響は軽微であるからである。しかし，証券市場の上場企業数は東証だけでも2008年9月現在で2,400社と多いうえに，近年では年に4回の決算発表がある。さらに，大幅な業績の修正を行う場合にも業績修正発表を行う必要があるため，多くの上場企業の業績発表を分析し，その業績要因を取得するには多

大な労力を要する。さらに取得した業績要因を分析して重要性を判定するには，その企業に関する詳細な知識が必要になり，高度な専門知識がない個人投資家にとっては非常に困難な作業である。そこで，投資判断支援を行うための技術の一環として，業績要因自動抽出・重要度付与技術についての研究を開始した。

2. 研究の目的

本研究では，経済新聞記事に含まれる企業業績に関する記事から，その業績要因を自動的に抽出し，さらに，抽出した業績要因に対して当該企業業績に対する重要度を自動的に付与し，抽出した業績要因を投資判断の支援だけでなく，景気予測，技術動向予測などの用途に使用することのできる情報に変換することを目的に推進した。例えば，図1の例では「NAND型フラッシュメモリーなどが順調だった」という部分を業績要因として抽出し，さらに，それを東芝の事業と関連する業績要因として認識することで高い重要度を付与し，一方，「償却負担が増加した」といった一時的な要因であったならば低い重要度を付与することを目的とする。具体的には，まず，抽出すべき業績要因を「共通頻出表現」と「手がかり表現」に汎化し，この2つの表現を自動的に抽出することで業績要因抽出を行った。次に，業績要因に含まれる単語や企業のWebページの統計情報を使用することで重要度付与を行った。現状では業績要因の分析は，業界に精通した専門のアナリストが行っており，人手で行う場合でも難しいタスクであるが，本研究により自動的に業績要因を抽出し，重要度を付与することができれば，高度な専門知識がない個人投資家に対する投資判断支援を行うための有用な

情報源になることが期待できる。

3. 研究の方法

本研究では3年間(H21~H23)で以下の3つの技術について研究を推進した。

- (1) 経済新聞記事集合から企業の業績発表記事を認定し、その中から業績要因を識別して抽出する技術：特に再現率の向上について取り組む。
- (2) 抽出した業績要因が、その企業の事業と関連があるかどうかを自動的に判定する技術：例えば、「リストラ費用が増加」「償却負担が増加」といった一時的な要因と、事業に関連がある要因を分類する。
- (3) 業績要因に対して重要度を付与する技術：上記で事業に関連があると判定された業績要因に対して、その企業の事業との関連に応じて重要度を付与する。

4. 研究成果

本研究を推進した結果、以下に示す成果が得られた。

- (1) 経済新聞記事集合から企業の業績発表記事を認定し、その中から業績要因を識別して抽出する技術：企業の業績発表記事から業績要因表現を抽出する既提案手法の改良を行った。既提案手法の業績要因表現の抽出手法では、抽出すべき業績要因を「共通頻出表現」と「手がかり表現」の2つの表現で構成される形態素列と定義した。しかしながら、共通頻出表現は、異なった業績要因に対して共通して頻出する形態素列と定義された表現（「売り上げ」等）であり、「靴向け人工皮革も伸び悩んだ。」のように、その企業にとって重要なキーワード（この例では「人工皮革」）を含みつつ、共通頻出表現を含まない業績要因を抽出することができなかった。そこで、ある企業の業績発表記事から業績要因を抽出する際に、その企業の Web サイトから当該企業にとって重要なキーワードを抽出し、それを共通頻出表現として使用することで、既提案手法では抽出できなかった業績要因を抽出できるような手法を追加した。この改良によって、既提案手法の再現率は 67.1%であるが、改良手法の再現率は 73.7%であり、既提案手法では抽出できなかった業績要因が抽出できるようになった。
- (2) 抽出した業績要因が、その企業の事業と関連があるかどうかを自動的に判定する技術：業績要因に重要度を付与する前処理として、業績要因が当該企業の本業と関連があるか否かで分類する手法を開発した。分類には SVM を使用し、素性として形態素のユニグラム、バイグラムを使

用した。評価実験の結果、正解率 95.3%を達成し、良好な結果を得た。また、イベントスタディ法に基づく分析によって、業績要因を含む業績発表記事が株式市場に対し影響を与えている可能性があることが示された。

- (3) 業績要因に対して重要度を付与する技術：企業の業績発表記事から抽出した業績要因表現に対して重要度を付与する手法の開発を行った。例えば、「三菱電機」は多くの事業を行っているが、会社四季報によれば、三菱電機の特徴欄に「F A が収益柱」という記述がある。そのため、三菱電機の業績要因として F A（ファクトリーオートメーション）が好調（もしくは不振）であれば、投資判断を行ううえで重要な情報となる。しかし、個人投資家が必ずしも多くの企業の主力事業を熟知しているわけではない。そこで、既提案手法を改善するために抽出した企業の Web サイトから当該企業にとって重要なキーワードを利用して業績要因に対して重要度を付与し、その重要度に基づいて最も重要な業績要因の自動判定を行った。評価の結果、本手法は 77.3%の正解率を達成し、業績発表記事において2番目に出現する文を最重要の業績要因として判定する Lead(s2)手法や、業績発表記事において最初に抽出される業績要因を最重要の業績要因として判定する Lead(ce)手法より高い正解率を達成した。
- (4) 業績要因への極性付与手法の有効性の検証：我々は既に予備的検討として、業績発表記事から抽出した業績要因に対して業績に関する極性（ポジティブ、ネガティブ）を付与する技術を開発していたが、開発した手法と様々な既存手法との比較実験を実施した。その結果、既存手法に比べて大幅に性能が向上しており、開発した極性付与手法の効果を確認することができた。業績要因表現は複数の文節で構成されるため、同一の表現が業績発表記事集合中に出現しない。そのため、特定の語との共起頻度を使用する既存手法では精度の高い極性付与は困難であり、比較実験でもそのことが示された。開発した手法では、業績要因表現を手がかり表現（が好調等）と共通頻出表現（売り上げ等）の対に簡略化することで、上記の問題点を解決した。
- (5) 製品特徴情報の製品発表プレスリリースからの抽出手法の開発：本研究では、製品発表プレスリリースから「高画質を実現しました。」といった、製品特徴の情報を含む文を抽出する手法を開発した。製

品発表に関するプレスリリースに含まれる、その製品の特徴に関する情報は、製品のトレンド分析や企業の開発方針の決定等において重要な情報となり、また、技術動向予測や投資判断支援のための情報としても有用である。抽出には「実現しました。」といった手がかり表現を使用するが、本研究では数多くの手がかり表現を製品発表プレスリリースから自動的に獲得した。また、手がかり表現を自動的に獲得する際に同時に獲得される共通頻出表現（「操作性」など）を使用することで、製品発表プレスリリースでは頻出する体言止めで表される製品特徴文の抽出も可能とした。手がかり表現を自動的に獲得する際に、「発売します。」といった不適切な手がかり表現が獲得されることを防ぐために、不適切な手がかり表現をも自動的に獲得することで、不適切な手がかり表現のリストを作成し、それらが手がかり表現として獲得されることを防ぐことができた。評価の結果、精度78.4%、再現率62.9%であり、不適切な手がかり表現の除去を行わない手法に比べて精度が向上した。

5. 主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計7件）

- (1) 谷口将太, 坂地泰紀, 酒井浩之, 増山繁, 経済新聞記事から抽出した景気動向を示す根拠表現への極性付与手法の提案, 電子情報通信学会論文誌 D, J94-D, No.6, 6, pp.1039-1043, 2011. (査読有)
- (2) 坂地泰紀, 野中尋史, 酒井浩之, 増山繁, Cross-Bootstrapping: 特許文書からの課題・効果表現対の自動抽出手法, 電子情報通信学会論文誌 D, J93-D, no.6, pp.742-755, 2010. (査読有)
- (3) 鈴木佑輔, 横田隼, 酒井浩之, 増山繁, Web 掲示板からの質問・回答対応の自動抽出, 人工知能学会論文誌, vol. 25, no1, pp.168-173, 2010. (査読有)
- (4) Hiroyuki Sakai, Shigeru Masuyama, Assigning Polarity to Causal Information in Financial Articles on Business Performance of Companies, IEICE Trans. Information and Systems, vol.E92-D, no.12, pp. 2341-2350, 2009. (査読有)
- (5) 酒井浩之, 野中尋史, 増山繁, 特許明細書からの技術課題情報の抽出, 人工知能学会論文誌, vol.24, no.6, pp.531-540, 2009. (査読有)
- (6) Hirofumi Nonaka, Hiroyuki Sakai, Shigeru Masuyama, Masazumi Ao, Yutaka Hada,

Development of an automatic ecotechnology patent search method using text mining, Journal of ecotechnology research, vol.15 pp.23-26, 2009. (査読有)

- (7) 野中尋史, 酒井浩之, 増山繁, テキストマイニングを用いた判例文書の分類および情報抽出, 情報ネットワークレビュー, vol.8, pp74-85, 2009. (査読有)

〔学会発表〕（計24件）

- (1) 太田貴久, 鶴田雅信, 野中尋史, 山本悠二, 酒井浩之, 増山繁, 技術-効果型パテントマップに基づくシーズ知的探索システムの開発, 言語処理学会第18回年次大会, 広島市立大学, 2012年3月15日.
- (2) 谷口将太, 野中尋史, 酒井浩之, 増山繁, パテントマップ生成のための特許明細書中の技術構成要素候補の修飾必要性判定手法, WiNF2011, 豊橋技術科学大学, 2011年11月26日.
- (3) 酒井浩之, 増山繁, 企業業績発表記事を対象としたテキストマイニング, 第54回自動制御連合講演会, 豊橋技術科学大学, 2011年11月19日.
- (4) 酒井浩之, 増山繁, 手がかり表現自動獲得による製品発表プレスリリースからの製品特徴の抽出, 言語処理学会第17回年次大会, 豊橋技術科学大学, 2011年3月9日.
- (5) 藤村真太郎, 野中尋史, 酒井浩之, 増山繁, 製品情報のプレスリリースとその製品特徴に関連した特許文書との対応付け, 言語処理学会第17回年次大会, 豊橋技術科学大学, 2011年3月8日.
- (6) 坂地泰紀, 増山繁, 酒井浩之, 新聞記事中の文が因果関係を含むか否かの判定, 電子情報通信学会技術研究報告, 東北大学, 2010年7月23日.
- (7) Akio Kobayashi, Hirofumi Nonaka, Shigeru Masuyama, Hiroyuki Sakai, An Automatic Thesaurus Construction Method for Technological Terms in Patent Maps, The 40th International Conference on Computers & Industrial Engineering, Awaji Island, Japan, 26 July, 2010.
- (8) Hirofumi Nonaka, Akio Kobayashi, Hiroki Sakai, Yusuke Suzuki, Hiroyuki Sakai, Shigeru Masuyama, Extraction of the effect and the technology terms from a patent document, The 40th International Conference on Computers & Industrial Engineering, Awaji Island, Japan, 26 July, 2010.
- (9) Hirofumi Nonaka, Yusuke Suzuki, Hiroki Sakai, Hiroyuki Sakai, Shigeru Masuyama, Extracting EFFECT Expressions Form Patent Documents, The 25th International Technical Conference of Circuits/Systems, Computers and Communications (ITC-CSCC 2010), Pattaya,

- Thailand, 7 July, 2010.
- (10) Yuji Yamamoto, Shigeru Masuyama, Hiroyuki Sakai, Clustering via Online Passive-Aggressive Algorithm, The 25th International Technical Conference of Circuits/Systems, Computers and Communications (ITC-CSCC 2010), Pattaya, Thailand, 7 July, 2010.
 - (11) Yusuke Suzuki, Hirofumi Nonaka, Akio Kobayashi, Hiroyuki Sakai, Shigeru Masuyama, Extraction of Technology Terms from Patent Specifications for Technology-Effect Type Patent Map Generation, The 25th International Technical Conference of Circuits/Systems, Computers and Communications (ITC-CSCC 2010), Pattaya, Thailand, 6 July, 2010.
 - (12) 野中 尋史, 小林 暁雄, 坂地 泰紀, 鈴木 佑輔, 酒井 浩之, 増山 繁, 効果-技術型パテントマップ作成を目的とした特許明細書からの効果語及び技術語の自動抽出と意味的統合, 第 24 回人工知能学会全国大会, 長崎ブリックホール, 2010 年 6 月 11 日.
 - (13) 西崎 海人, 酒井 浩之, 増山 繁, 製品・部門情報の企業業績要因表現からの抽出, 第 24 回人工知能学会全国大会, 長崎ブリックホール, 2010 年 6 月 11 日.
 - (14) 酒井浩之, 松原茂樹, 増山繁, 稲垣康善, 文中の接続助詞「が」に着目した翻訳単位間の意味関係の推定, 言語処理学会第 16 回年次大会, 東京大学(本郷キャンパス), 2010 年 3 月 11 日.
 - (15) 西崎海人, 酒井浩之, 増山繁, 新聞記事の企業業績発表記事からの製品・部門情報の抽出, 言語処理学会第 16 回年次大会, 東京大学(本郷キャンパス), 2010 年 3 月 11 日.
 - (16) 梅村真史, 酒井浩之, 増山繁, レビュー記事からの評判情報抽出の一手法の提案, 言語処理学会第 16 回年次大会, 東京大学(本郷キャンパス), 2010 年 3 月 9 日.
 - (17) 谷口将太, 坂地泰紀, 酒井浩之, 増山繁, 経済新聞記事から抽出した景気動向を示す根拠表現への極性付与手法の提案, 言語処理学会第 16 回年次大会, 東京大学(本郷キャンパス), 2010 年 3 月 9 日.
 - (18) 梅村真史, 酒井浩之, 増山繁, 統計情報を使用した Web 掲示板からの製品評判情報の抽出, WinF2009 (第 7 回情報学ワークショップ), 名古屋工業大学, 2009 年 11 月 27 日
 - (19) 坂地泰紀, 野中尋史, 酒井 浩之, 増山 繁, 特許文書からのブートストラップ手法を用いた課

題・効果表現対の抽出, 情報処理学会研究報告(2009-NL-192), 北見工業大学, 2009 年 7 月 23 日.

- (20) 藤村真太郎, 酒井 浩之, 増山 繁, 企業業績要因文の経常的か否かに基づく分類とイベントスタディ法に基づく分析, 第 23 回人工知能学会全国大会, サンポートホール高松, 2009 年 6 月 19 日.
- (21) 酒井浩之, 野中尋史, 増山繁, 特許明細書からの出願目的・技術課題情報の抽出, 第 23 回人工知能学会全国大会, サンポートホール高松, 2009 年 6 月 17 日.

[その他]
ホームページ等
<http://www.la.cs.tut.ac.jp/~sakai/>

6. 研究組織

- (1) 研究代表者
酒井 浩之 (Sakai Hiroyuki)
豊橋技術科学大学・大学院工学研究科・助教
研究者番号：70402659
- (2) 研究分担者
()
研究者番号：
- (3) 連携研究者
()
研究者番号：