

# 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

Classification of Causal Sentences Concerning Business Performance of Companies based on whether Ordinary or Not and Analysis Using the Event Study Method

藤村 真太郎      酒井 浩之      増山 繁  
Shintaro Fujimura      Hiroyuki Sakai      Shigeru Masuyama

豊橋技術科学大学 知識情報工学系

Department of Knowledge-based Information Engineering, Toyohashi University of Technology

We propose a method to classify causal sentences concerning business performance extracted from news paper articles. Our method classifies an extracted sentence into “ordinary” or “not ordinary.” Furthermore, we analyze articles including sentences classified based on our method, by using event study method to show how articles impact stock market. As a result, we observed that articles classified as “ordinary” impact stock market more than “not ordinary” articles.

## 1. はじめに

近年では、ネット証券の普及により、個人による投資に対する敷居も下がり、新たに投資活動を始める人も少なからず存在している。適切な投資判断を行うためには、企業の業績情報を収集することは重要である。また、投資判断だけでなく、企業経営においても同業他社の業績を収集し、分析することは重要である。しかし、東京証券取引所に上場する企業だけで2009年4月で2,372社存在し、年4回の決算発表を行っている。多くの上場企業の業績発表を分析し、適切な投資判断を行うには、高度な専門知識を要する。そのため、投資判断を支援するための技術が必要になってきている。

Sakaiら[Sakai 08]は、新聞記事から業績要因を取得する手法を提案している。ここで、取得された業績要因が経常的なもの(例えば、「自動車の売り上げが好調だった」)であれば、株価に与える影響が大きいことが期待される。一方、一時的な利益や損失の計上(例えば、「社宅跡地の売却益を特別利益に計上」)が要因であるならば、株価への影響は経常的なものと比較して軽微であることが予想される。そのため、業績要因をそのいずれであるかに分類することは、投資判断を支援するための情報として意義がある。

本研究では、Sakaiら[Sakai 08]の手法を用いて新聞記事から業績要因を含む文を取得し、それらに対して経常的な業績要因を含む文であるか否かを分類を行う手法を提案する。それに加えて、業績要因文の分類の有効性を調べるために、イベントスタディ法[MacKinlay 97]に基づく分析によって、分類された業績要因文が株式市場に対して影響を与えているか否かを明らかにする。

## 2. 関連研究

本研究のように新聞記事と株価を用いた研究として、小川ら[小川 01]は分類により新聞記事に付与されたテーマ情報を用いて、いくつかの銘柄に関して株価変動と関連のあるテーマの抽出を行っている。Koppelら[Koppel 04]はある企業に関連する記事が株価に対して、よいニュースであるか悪いニュースであるか、すなわち、それにより株価が上がるニュースであるか株価が下がるニュースであるかを語彙と株価変動を利用し、

機械学習によって分類している。張ら[張 08]は過去の新聞記事と株価変動を利用し、新聞記事中の単語対に重みを付与して新聞記事の評価を行っており、記事評価値と株価変動の相関を示した。このように、新聞記事と株価の関連に関する研究は近年、活発に行われている。しかしながら、これらの研究は新聞記事全体を対象とした研究であり、新聞記事中のどの情報が重要であるかに着目した研究ではない。本研究では、新聞記事からSakaiら[Sakai 08]の手法によって業績要因の抽出を行い、それを対象としている点が異なる。

新聞記事から業績要因を抽出する研究としては、Sakaiら[Sakai 08]は手がかり表現と共通頻出表現の2つを定義し、これらの表現をブートストラップ的に獲得し、取得された表現を用いて業績発表記事中から業績要因の抽出を行っている。そして、それらに対して極性の付与を行っている[Sakai 08-2]。しかし、これらの研究では、極性付与された業績要因が株価に対して影響を与えるかどうかは分析されていない。それに対して本研究では、業績要因文の株価への影響についてもイベントスタディ法で分析している。

## 3. 業績要因文

本研究における業績要因文とは、業績要因を含む文のことである。業績要因文はSakaiら[Sakai 08]の手法によって新聞中の企業業績発表記事から取得する。業績要因文には経常的な業績要因を含む文と、経常的な業績要因を含まない文が存在するため、本研究では、業績要因文をこの2つのクラスに分類する。

経常的な業績要因を含む文とは図1のような文である。

肉まんなど食材の外販部門が予想以上に好調だった。

図1: 経常的な業績要因を含む文の例

それに対して、経常的な業績要因を含まない文とは図2に示すような文である。

## 4. 分類手法

本研究では、業績要因文を経常的な業績要因を含む文と経常的な業績要因を含まない文に分類を行う。分類対象である業

連絡先: 豊橋技術科学大学, 豊橋市天伯町雲雀ヶ丘1-1, 0532-44-6867, 0532-44-6873, fujimura@smlab.tut.ac.jp

社宅跡地の売却益八十三億円などを特別利益に計上。

図 2: 経常的な業績要因を含まない文の例

績要因文に対する調査の結果, Sakai ら [Sakai 08] の手法で取得された業績要因文中には「販売」、「営業」、「特別」など、分類に有効であると期待できる特徴的な語が多く含まれているため、単語または単語の列で分類できる可能性が高い。そのため、2 値分類器である Support Vector Machine (SVM) を使用することによって業績要因文の分類を行う。

#### 4.1 使用する素性

先に述べたように、本研究での業績要因文の分類においては、単語または単語の列が有効であると考えたため、業績要因文を形態素解析器<sup>\*1</sup>を用いて形態素単位に分割し、形態素列を取得する。形態素列から形態素のユニグラム、バイグラムを生成し、これらの業績要因文中での頻度を求め、それらを素性とする。

例えば、「自動車の売り上げが好調。」という文からは素性として、「自動車」、「の」、「売り上げ」、「が」、「好調」、「。」、「自動車の」、「の売り上げ」、「売り上げが」、「が好調」、「好調。」が生成される。

## 5. 結果と考察

今までに述べた業績要因文分類手法の評価実験と考察を行う。本章では先に評価実験結果を示し、その後考察を行う。

### 5.1 評価実験

本手法の評価を、2001 年 1 月 5 日から 2001 年 3 月 20 日までの企業業績発表記事から抽出された業績要因文を集めた学習用データセット (データ数 434 件, 正例 217 件, 負例 217 件) と 2005 年 11 月 12 日から 2005 年 12 月 29 日までの企業業績発表記事から抽出された業績要因文を集めた評価用データセット (データ数 1,000 件, 正例 798 件, 負例 202 件) を用いて行う。正例は経常的な業績要因を含む文, 負例は経常的な業績要因を含まない文とする。また、これらの業績要因文は日本経済新聞から取得した。

これらのデータセットを利用し、評価実験を行った。評価実験では、素性にユニグラムを用いたケース、バイグラムを用いたケース、ユニグラムとバイグラムの両方を用いたケースの 3 種類について、 $SVM^{light*2}$  を使用して実験を行った。

### 5.2 考察

評価実験の結果を表 1 に示す。分類結果は正解率 95.3% となり、良好であると言える。素性の種類を変えながら実験を行った結果、正解率は形態素のユニグラム、バイグラムの全てを使用したものが最も高い値を示した。なお、形態素のユニグラムを使用しただけでも、十分な正解率を得ることができた。また、SVM のモデルファイルを解析し、正例、負例それぞれのスコアが高いものを表 2, 3 に示す。この結果から、素性として有効であったのは正例、負例ともにユニグラムであることがわかる。負例に関しては「特別」や「売却」などの内容語が上位に現れていたが、正例に関しては「が」や「ほか」などの語が上位に現れている。業績要因文では、「自動車の販売は好

調だが」や「コストの削減が奏効したほか、」といったような記述が見られ、「が」や「ほか」を用いて複数の業績要因の記述を行う場面も見られる。業績要因には、経常的な業績要因の方が多いため、複数の要因を記述しようと試みている文は経常的な業績要因を含む文であると期待できるからであるといえる。

表 1: 評価実験結果

実験名	ユニグラム		バイグラム		両方	
	正例	負例	正例	負例	正例	負例
適合率	0.969	0.885	0.936	0.925	0.962	0.914
再現率	0.971	0.876	0.985	0.733	0.980	0.847
F 値	0.970	0.881	0.960	0.818	0.971	0.879
正解率	0.952		0.934		0.953	

表 2: モデルファイル解析結果 (正例, 上位 5 件)

素性の種類	素性	スコア
ユニグラム	が	0.253409
バイグラム	ほか、	0.106376
ユニグラム	ほか	0.106376
ユニグラム	伸びる	0.094023
ユニグラム	や	0.093236

表 3: モデルファイル解析結果 (負例, 上位 5 件)

素性の種類	素性	スコア
ユニグラム	特別	-0.219319
ユニグラム	売却	-0.202714
ユニグラム	損失	-0.197737
ユニグラム	最終	-0.150065
ユニグラム	損	-0.141194

評価実験の結果は良好であったが、分類に失敗した業績要因文も少なからず存在する。図 3 で示す文は、分類器によって経常的な業績要因を含むとされた文である。この文中の業績要因の箇所は「事務所など固定資産の売却益が寄与」という箇所であるが、それに先行する「前年同期は売上高の開示で比較できないが、」の箇所があるために経常的な業績要因を含むとされてしまっている可能性がある。モデルファイルの解析結果に基づいて調べると、「売上」、「売上高」、「が」などの素性は上位 10 件内に含まれている素性である。そのため分類を誤ったといえる。

## 6. イベントスタディ法に基づく分析

業績要因文の株式市場での影響や分類されたクラス間 (経常的な業績要因を含む, 含まない) での影響の違いを分析するために、業績要因文を含む新聞記事に対してイベントスタディ法 [MacKinlay 97] に基づく分析を行った。業績要因文を本研究で述べた手法 (4 章) で分類し、その結果に基づき経常的な業績要因を含んでいる業績発表記事と経常的な業績要因を含まない業績発表記事を取り出した。イベントスタディ法は、イベント期間中の 1 日あたりの株式収益率を用いて、イベントが

\*1 ChaSen, <http://chasen-legacy.sourceforge.jp/>

\*2  $SVM^{light}$ , <http://svmlight.joachims.org/>

前年同期は売上高のみの開示で比較できないが、事務所など固定資産の売却益が寄与し、前期末の最終赤字十六億円から黒字転換した。

図 3: 分類器によって経常的な業績要因を含むとされた例

株式市場に影響を与えているか否かを調べる手法であるため、ポジティブな業績要因文（例えば、「民間設備投資の回復で試験・計測機の受注が伸びた。」）とネガティブな業績要因文（例えば、「半導体市況悪化と米国のパソコン事業の低迷が響く。」）を合わせて 1 つの集合として捉えてしまうと、各クラス内の平均を求める際に株式収益率が相殺されてしまうことが考えられる。そのため、それらを分けて扱うことにした。すなわち、以下に示す 4 つの集合について検証を行う。対象とした記事は東証一部上場企業のものとした。各記事集合の記事は 2001 年 1 月 1 日以降の記事を日付順に各 60 件、合計 240 件を取得した。この分析では本研究の分類によるクラス間の違いを把握することを目的としているため、業績要因文のポジティブとネガティブの判断は人手で行っている。また、業績要因文の分類クラスが記事内で全て同一であるもののみを対象とした限定された状況のみを想定している。

表 4: 検証対象一覧

記事集合名	経常的な業績要因を含むか	極性
記事集合 1	経常的な業績要因を含む	ポジティブ
記事集合 2	経常的な業績要因を含む	ネガティブ
記事集合 3	経常的な業績要因を含まない	ポジティブ
記事集合 4	経常的な業績要因を含まない	ネガティブ

### 6.1 イベントスタディ法

イベントスタディ法とは、あるイベントが起こった日をイベント日とし、イベント日の前後数日に渡って株式収益率を調べ、イベントが有意に株式市場に影響を与えているか否かを調べるための手法である。イベントスタディ法を適用するにあたっては、MacKinlay の論文 [MacKinlay 97] を参考にした。

イベントスタディ法を適用するにあたっての設定としてはイベント日を業績発表記事が掲載された日以降の最初の株式市場の営業日に割り当て、イベント期間については 11 日間（イベント日の 5 営業日前から 5 営業日後まで）とする。また、各銘柄の収益率の推定には、マーケットモデル（式 1）を用い、マーケットインデックスとして TOPIX を使用する。

$$R_{it} = \alpha_i + \beta_i R_t^m + \epsilon_{it} \quad (1)$$

$R_t^m$  は  $t$  日の市場収益率、 $\epsilon_{it}$  は誤差項、 $E[\epsilon] = 0$  である。

イベントスタディ法では、各銘柄の異常収益率（式 4）を求めることによって分析を行う。異常収益率は各銘柄の実際の収益率と市場収益率によって推測された収益率との差で求められる。 $P_{it}$  を銘柄  $i$  の  $t$  日の終値とすると、銘柄  $i$  の  $t$  日の収益率は次式で求められる。

$$R_{it} = \frac{P_{it} - P_{it-1}}{P_{it-1}} \quad (2)$$

マーケットモデルの  $\alpha_i$  と  $\beta_i$  を今回は 60 日間の推定期間（イベント日の 30 営業日前から 89 営業日前まで）を利用して

最小二乗法により求める。銘柄  $i$  の誤差項の分散は推定期間が十分にあれば、次式で示される。

$$\hat{\sigma}_{\epsilon_i}^2 = \frac{1}{60 - 2} \sum_{t=-89}^{-30} (R_{it} - \hat{\alpha}_i - \hat{\beta}_i R_{it}^m)^2 \quad (3)$$

銘柄  $i$  の  $t$  日の異常収益率は実際の収益率とマーケットモデルによって推定された収益率の差によって求められる。

$$AR_{it} = R_{it} - (\hat{\alpha}_i + \hat{\beta}_i R_{it}^m) \quad (4)$$

この異常収益率は帰無仮説の下で推定期間により推定された分散  $\hat{\sigma}_{\epsilon_i}^2$ 、平均 0 の正規分布に従う。

各クラスごとの影響を分析するためにクラス内の  $N$  個の異常収益率の平均を取ったものは、平均異常収益率と呼ばれ、式 5 で求められ、この平均異常収益率の分散は式 6 で示される。式 5 と式 6 から求めた統計量は標準正規分布に従うため、 $\overline{AR}_t$  が有意に 0 と異なるか否かを検定することができる。

$$\overline{AR}_t = \frac{1}{N} \sum_{i=1}^N AR_{it} \quad (5)$$

$$\text{var}(\overline{AR}_t) = \frac{1}{N^2} \sum_{i=1}^N \sigma_{\epsilon_i}^2 \quad (6)$$

$$\theta_1 = \frac{\overline{AR}_t}{\sqrt{\text{var}(\overline{AR}_t)}} \quad (7)$$

また、イベント期間を通してのイベントの影響を調べるための累積平均異常収益率とその分散は、式 8、式 9 で示される。同様に標準正規分布に従う統計量を用いて検定を行うことができる。

$$\overline{CAR}(t_1, t_2) = \sum_{t=t_1}^{t_2} \overline{AR}_t \quad (8)$$

$$\text{var}(\overline{CAR}(t_1, t_2)) = \sum_{t=t_1}^{t_2} \text{var}(\overline{AR}_t) \quad (9)$$

$$\theta_2 = \frac{\overline{CAR}(t_1, t_2)}{\sqrt{\text{var}(\overline{CAR}(t_1, t_2))}} \quad (10)$$

### 6.2 考察

6.1 節で説明した方法で求めた各記事集合のイベント期間の平均異常収益率を示したものを表 5 に、イベント期間での累積異常収益率を求めたものを表 6 に示す。表中の\*\*\*は有意水準 1%、\*\*は有意水準 5%、\*は有意水準 10%を表す。平均異常収益率とその統計量（表 5）を観察すると、経常的な業績要因を含む記事集合（記事集合 1、記事集合 2）に関しては、イベント日に有意水準 1%の平均異常収益率が観測された。また、前日にも有意水準 5%以上の反応が見られる。同様に経常的な業績要因を含まない記事集合（記事集合 3、記事集合 4）でも反応は見られるが、記事集合 3 に関しては前日に有意水準 1%ではあるが負の反応が観測され、記事集合 4 では有意水準 10%の反応が観測されるに留まった。期間中の累積平均異常収益率の推移を示したもの（図 4）を見ると、イベント日（ $t = 0$ ）に向け、経常的な業績要因を含む記事集合では絶対値に大きい変化が見られる。それに対して、経常的な業績要因を含まない記事集合では、変化は見られるものの、経常的な業績要因を含む

ものと比較して軽微であった。累積平均異常収益率とその統計量(表6)を観察すると、イベント期間での累積結果は経常的な業績要因を含む記事集合では有意水準1%であったが、経常的な業績要因を含まない記事集合の統計量に関しては、統計的な有意差が観測されなかった。

このことから、業績要因を含む記事は株式市場に対して影響を与えているといえ、経常的な業績要因を含む文から構成される記事は、経常的な業績要因を含まない文から構成される記事よりも株式市場に対する影響が大きいことが期待できるといえる。記事集合内部で異常収益率が相殺されている可能性や抽出されていない業績要因文による影響は否定できないが、2つのクラス間での差が現れていることは事実であり、本研究での分類が有用であるといえる。

表 5: イベント期間の平均異常収益率とその統計量

t	記事集合 1		記事集合 2	
	$\overline{AR}_t$	$\theta_1$	$\overline{AR}_t$	$\theta_1$
-5	-0.00028	-0.08154	0.00509	1.43205
-4	0.00468	1.35372	-0.00136	-0.38247
-3	0.00087	0.25264	-0.00098	-0.27405
-2	0.00816	2.35860**	0.00037	0.10272
-1	0.01529	4.41883***	-0.00910	-2.55982**
0	0.02085	6.02652***	-0.01793	-5.04132***
1	-0.00285	-0.82421	-0.00180	-0.50718
2	-0.00510	-1.47411	-0.00338	-0.95128
3	-0.00494	-1.42637	-0.00395	-1.10911
4	0.00229	0.66208	0.00182	0.51096
5	0.00128	0.36892	-0.00099	-0.27897

t	記事集合 3		記事集合 4	
	$\overline{AR}_t$	$\theta_1$	$\overline{AR}_t$	$\theta_1$
-5	-0.00323	-1.11944	0.00196	0.55932
-4	-0.00113	-0.39054	0.00264	0.75197
-3	-0.00023	-0.07893	0.00352	1.00399
-2	-0.00487	-1.69028*	-0.00031	-0.08785
-1	0.01213	4.20688***	-0.00277	-0.78990
0	0.00350	1.21244	-0.00628	-1.79108*
1	-0.00473	-1.64161	-0.00335	-0.95570
2	0.00069	0.24058	-0.00092	-0.26169
3	-0.00335	-1.16094	-0.00125	-0.35557
4	-0.00226	-0.78364	0.00191	0.54376
5	0.00388	1.34543	0.00284	0.81006

表 6: 累積平均異常収益率とその統計量

	$\overline{CAR}$	$\theta_2$
記事集合 1	0.04025	3.50781***
記事集合 2	-0.03222	-2.73098***
記事集合 3	0.00041	0.04236
記事集合 4	-0.00201	-0.17262

## 7. まとめ

本研究では、業績要因文を単純な分類器を用いて、経常的な業績要因を含む文と経常的な業績要因を含まない文に分類し

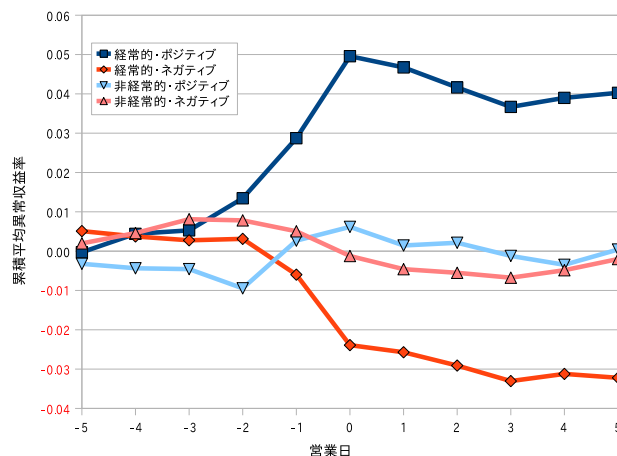


図 4: 累積平均異常収益率の推移

た。また、イベントスタディ法に基づく分析によって、業績要因文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。業績要因文のポジティブ、ネガティブの判断を人手で行っている点など、注意すべき点もあるが、本研究で分類した2つのクラス間で異常収益率に差が見られ、クラス間で異なる性質を持っている可能性があることが示された。今後の課題として、投資家にとって更に有用な情報を提供するために、業績要因文に対して重要度の付与を行うことを考えている。

## 参考文献

- [Sakai 08] H. Sakai, S. Masuyama, "Cause Information Extraction from Financial Articles Concerning Business Performance", IEICE Transactions on Information and Systems Vol.E91-D No.4 April 2008, pp.959-968, 2008.
- [Sakai 08-2] H. Sakai, S. Masuyama, "Polarity Assignment to Causal Information Extracted from Financial Articles Concerning Business Performance of Companies", Proc. of the AI-2008 Twenty-eighth SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence, pp.307-320, CAMBRIDGE, ENGLAND, DECEMBER, 2008.
- [MacKinlay 97] MacKinlay, A. C., "Event Studies in Economics and Finance", Journal of Economic Literature Vol.35, Issue 1 (March 1997) pp.13-39, 1997.
- [小川 01] 小川 和也, 渡辺 勇, "株価データと新聞記事からのマイニング", 情報処理学会研究報告, 情報学基礎研究報告会報告 IPSJ SIG Notes Vol.2001, No.20(20010305), pp.137-144, 2001
- [Koppel 04] Koppel, M. and Shtrimberg I., "Good News or Bad News? Let me Market Decide", Proceedings of AAAI Spring Symposium on Exploring Attitude and Affect in Text, pp.86-88, 2004.
- [張 08] 張 へい, 松原 茂樹, "株価データに基づく新聞記事の評価", 第22回人工知能学会全国大会論文集, 2008