



ヘッドラインニュースに対する 株価の反応について

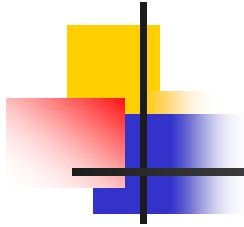
2007年2月10日

第6回 行動経済学ワークショップ

高橋 悟 (三井アセット信託銀行／筑波大学)

高橋 大志 (岡山大学)

津田 和彦 (筑波大学)



目次

➤ 研究目的	P 2
➤ 研究の背景	P 3 ~ P 4
➤ 分析データ	P 5 ~ P 7
➤ 分析方法	P 8 ~ P11
➤ 分析結果	P12 ~ P19
➤ まとめ	P20
➤ 今後の取り組み	P21



研究目的

- 金融市場においてテキストデータのもつ情報価値を分析
 - 日々更新されるニュースが株価に与える影響を分析
 - テキストマイニングの適用可能性の調査
 - ニュースの内容と株価リターンから影響のあるキーワードを抽出



研究の背景

- ▶ ファイナンスにおける大量の情報処理の問題点
 - インターネット等の情報端末, およびPCの処理能力の向上から, ファンドマネージャーの扱う情報は飛躍的に増加
 - 株価はすべての情報を反映して価格付けされていると言われるが, すべての情報を把握するのは非常に困難
 - ファイナンスにおける多くの研究では, 会計情報等の数値データが分析対象
 - 数値データ以外の情報は技術的に分析が困難

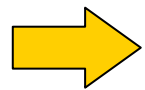
- ▶ テキスト情報の持つ効果を分析
 - テキスト情報も数値情報同様重要な情報源である
 - 決算報告書の文章(特に脚注などのコメント)
 - 新聞等のマスメディアのニュース
 - 証券アナリストの投資判断レポート(アナリストレポート)
 - インターネットにおける書き込み
 - 時事通信等・クイック等が更新するヘッドラインニュース



研究の背景

➤ テキスト情報の分析における問題点

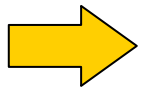
- データ数・情報量が非常に多い
- 自由記述形式であり, 単一のフォーマットがなく分析が困難
- 書き手の癖により, 同一内容であっても表現が異なる



情報処理の分野における技術である, 「テキストマイニング」を用いることにより, テキスト情報の分析の困難さを克服

➤ テキストマイニングを用いた研究例

- テキストマイニングによるアナリストレポートの分析[高橋, 津田2004]
 - アナリストレポートを対象にテキストマイニングを行い, 効果を分析
- テキストマイニング ～～ [栗田, 真壁2006]
 - 会社四季報を対象にテキストマイニングを行い, 効果を分析

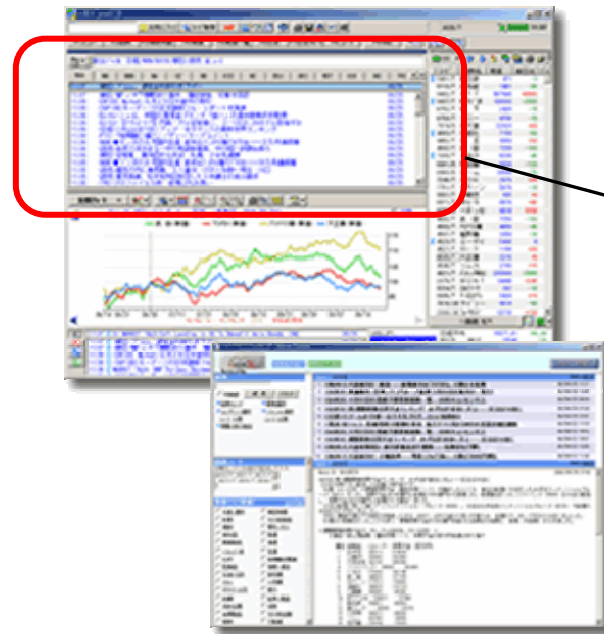


もっとも即時性のある情報元である, ヘッドラインニュースを対象としてテキストマイニングを行い, ニュースの効果について分析

分析データ

➤ ヘッドラインニュースとは

- クイック, ロイターやブルームバーグ等の情報端末によって, タイムリーに提供される情報
- 金融市場で実務を行う担当者にとって, もっとも即時性のある情報源のひとつ.
- 情報により, 株価が急激に変動する場合もある



【クイック端末の例】

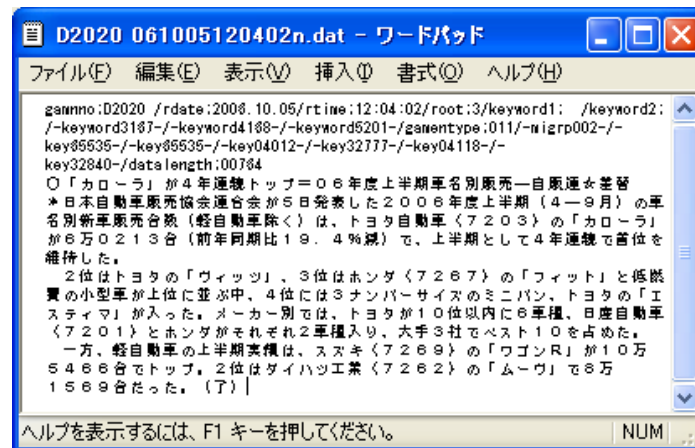
- ・この部分にリアルタイムでニュースが流れる.
- ・情報の更新タイミングが非常に早い.

(出所 株式会社QUICK ホームページ)

分析データ

➤ 時事通信のヘッドラインニュース

- 時事通信のヘッドラインニュースを分析対象として選択
- 代表例として以下の内容を含む
 - 国内マスメディアのニュース, ロイター通信などの国際時事, 天気予報といった一般のテレビなどで目にするニュース
 - 日銀や政府関係者の発言
 - 金利や為替レートなどのマーケット情報
 - 証券会社や格付け会社の企業評価に対するアナリストコメントや格付け情報
 - 企業経営者の発言や, 発言に対する評価・コメント



D2020 061005120402n.dat - ワードパッド

ファイル(F) 編集(E) 表示(V) 挿入(I) 書式(O) ヘルプ(H)

```
gammo:D2020 /rdate:2008.10.05/rtine:12:04:02/root:3/keyword1: /keyword2:
/-keyword3187-/-keyword4188-/-keyword5201-/agenttype:011/-nigrp002-/-
key89535-/-key89535-/-key04012-/-key32777-/-key04118-/-
key32840-/datalength:00784
○「カローラ」が4年連続トップ=08年度上半期車名別販売一自販運☆差替
*日本自動車販売協会連合会が5日発表した2008年度上半期(4-9月)の車
名別新車販売台数(軽自動車除く)は、トヨタ自動車<7203>の「カローラ」
が6万0213台(前年同期比19.4%減)で、上半期として4年連続で首位を
維持した。
2位はトヨタの「ヴィッツ」、3位はホンダ<7267>の「フィット」を低燃
費の小型車が上位に並び、4位には3ナンバーサイズのミニバン、トヨタの「エ
スティマ」が入った。メーカー別では、トヨタが10位以内に6車種、日産自動車
<7201>とホンダがそれぞれ2車種入り、大手3社でベスト10を占めた。
一方、軽自動車の上半期実績は、スズキ<7269>の「ワゴンR」が10万
5466台でトップ、2位はダイハツ工業<7262>の「ムーヴ」で8万
1569台だった。(ア)
```

ヘルプを表示するには、F1 キーを押してください。 NUM

【時事通信ニュースの例】

- ① ニュースタイプ
- ② 発信日付
- ③ 発信時刻
- ④ システム用コード
- ⑤ ニュースの長さ
- ⑥ 本文

(企業を対象とする場合は、本文中に「<銘柄コード>」の形で銘柄コードが記録される。

分析データ

➤ 分析データ

● 分析期間

- 2006年8月10日～2006年11月24日

● 分析対象企業

- 東証1部上場企業

分析対象期間中に東証1部に上場していた企業は、1,719社(上場廃止となった企業は22社, 新規で上場した企業は18社).

● 分析対象ニュース

- 期間中に時事通信から発信されたニュースの中から, 分析対象企業の企業情報関連ニュースを選択

	ニュース全体	企業情報 関連ニュース	企業情報関連 ニュース(東証1部)	企業情報 関連ニュース の割合	企業情報関連 ニュース(東証1部) の割合
ニュース の総数	346,975	30,052	13,402	8.661%	3.863%
1日当たり 平均ニュース数	3,336	320	144	9.583%	4.319%
1日に発信された 最大ニュース数	8,588	741	450	-	-
1日に発信された 最小ニュース数	172	0	0	-	-



分析方法

▶ ニュースの分類

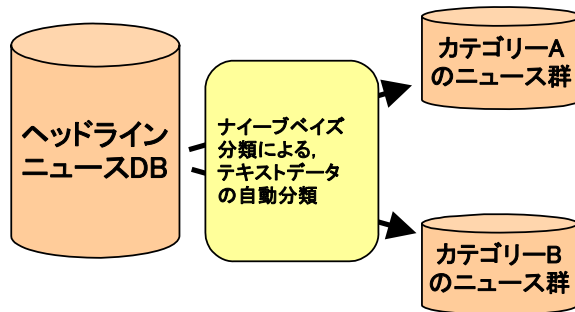
- ヘッドラインニュースを内容により、「Good News」, 「Bad News」, 「Neutral News」に自動分類
- ニュースの自動分類には, テキストマイニングで用いられる, 「ナイーブベイズ法」を使用
- 「ナイーブベイズ法」は直感的な解釈が容易なテキスト分類法であり, かつ比較的高い分類精度を有する

▶ 株価リターンの測定

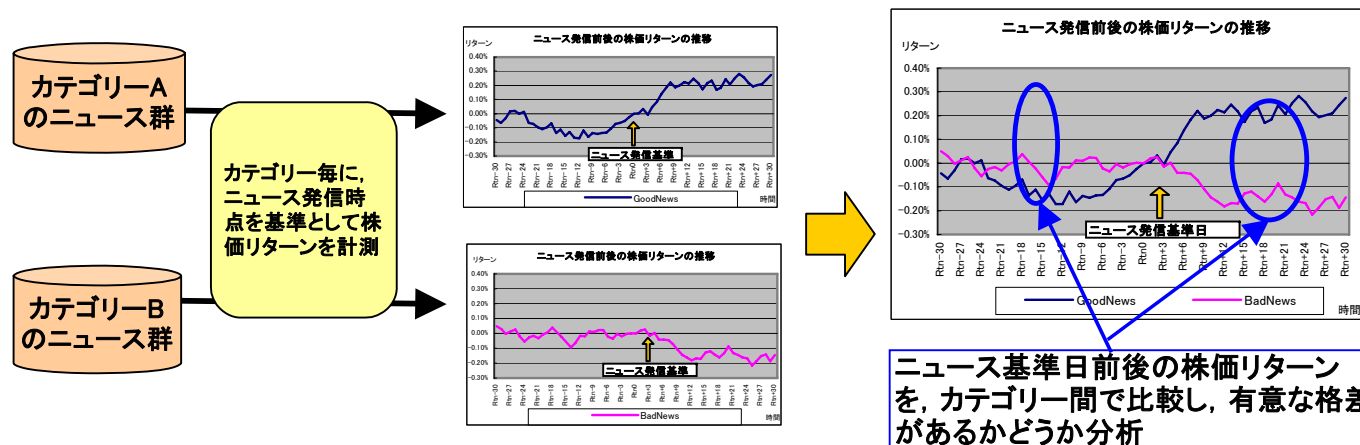
- ニュースが発信されたタイミングを基準とし, 対象企業の株価リターンを計測
 - 株価リターンは, 企業規模による影響を避けるために, 東証1部全銘柄のリターンの単純平均を計測し, その単純平均に対する超過リターンを分析対象とした
- ニュースのタイプにより, 株価リターンに有意な差が発生するかどうか, 統計的検定により分析
 - 検定には, Welch検定を使用

分析方法

➤ ステップ1 テキストの自動分類



➤ ステップ2 株価リターンを比較しテキスト情報の効果を分析





分析方法

➤ ナイーブベイズによる文章自動分類アルゴリズム

- テキストデータの特徴付けるキーワードが n 個と仮定し、テキスト内にキーワードが存在するかどうかを用いて、テキストデータをベクトルで表現

$$\mathbf{x} = (x_1, x_2, \dots, x_n)$$

x_i はテキストデータの中に i 番目のキーワードが存在するかどうかを表したもの

- 分類するクラスを c と仮定.
- テキスト x が与えられた条件のもとで、 x の属するクラスは、下式で求まる事後確率を最大化するクラスを求めることにより決定

$$\hat{c} = \arg \left\{ \max_c (P(c | \mathbf{x})) \right\} = \arg \left\{ \max_c \left(\frac{P(\mathbf{x} | c)P(c)}{P(\mathbf{x})} \right) \right\} = \arg \left\{ \max_c (P(c)P(\mathbf{x} | c)) \right\}$$

$P(\mathbf{x})$ はすべてのクラスで一定であること、およびベイズの定理を用いて算出



分析方法

▶ ナイーブベイズによる文章自動分類アルゴリズム

- 各キーワードの発生は、互いに独立であると仮定(ナイーブな仮定)すると,
 $P(\mathbf{x} | c)$ は以下のように展開

$$P(\mathbf{x} | c) = P(x_1, x_2, \dots, x_n | c) \approx \prod_{i=1}^n P(x_i | c)$$

- x の属するクラスは、以下の式で表現

$$\hat{c} = \arg \left\{ \max_c (P(c | \mathbf{x})) \right\} = \arg \left\{ \max_c (P(c) \prod_{i=1}^n P(x_i | c)) \right\}$$

- 学習データを選び、各クラスの相対頻度から $P(c)$ を決定し、また各クラスにおける x_i の相対頻度から $P(x_i | c)$ を決定
- テストデータを用いて、分類アルゴリズムの精度を検証
- 学習データとして、分析対象テキストデータの約3%にあたる、400のニュースを抽出し、またテストデータも同様に、分析対象テキストデータから400のニュースを抽出

分析結果

～テキスト分類の精度～

▶ インサンプルテストによる分類

- 学習データを用いて、分類アルゴリズムを作成
- 作成したアルゴリズムにより、学習データを再度分類し、精度を分析

		実際のタイプ					実際のタイプ		
		Good	Bad	Neutral			Good	Bad	Neutral
ベイズ学習結果	Good	179	24	24	ベイズ学習結果	Good	90.86%	24.74%	22.64%
	Bad	2	66	9		Bad	1.02%	68.04%	8.49%
	Neutral	16	7	73		Neutral	8.12%	7.22%	68.87%

▶ アウトオブサンプルテストによる分類

- 学習データを用いて作成したアルゴリズムを、テストデータに適用し、分類精度を分析

		実際のタイプ					実際のタイプ		
		Good	Bad	Neutral			Good	Bad	Neutral
ベイズ学習結果	Good	168	27	35	ベイズ学習結果	Good	90.81%	30.00%	28.00%
	Bad	3	58	5		Bad	1.62%	64.44%	4.00%
	Neutral	14	5	85		Neutral	7.57%	5.56%	68.00%

分析結果

～ニュースによるリターン格差の検証～

- Good NewsとBad Newsにおける株価リターンの比較
 - ニュースを発信した日を基準とし、前後30営業日の株価リターンを計測（株価リターンは、東証1部全銘柄のリターンの単純平均からの超過リターン）
 - ±30営業日、±20営業日、±10営業日毎に累積リターンを計測し、Good NewsとBad Newsにおいて有意な格差が存在するかどうか分析（カテゴリ間で分散が不均一な場合があるため、Welch検定を使用して分析）

		Return -30day	Return -20day	Return -10day	Return +10day	Return +20day	Return +30day
平均 リターン	Good News	1.90%	1.54%	1.11%	0.24%	0.52%	1.04%
	Bad News	-2.12%	-2.06%	-1.75%	-0.61%	-0.69%	-1.19%
	差	4.03%	3.60%	2.86%	0.84%	1.21%	2.23%
Welch 検定	t値	19.00	20.21	20.23	6.75	7.22	11.24
	検定結果	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD

- すべての計測時点において、Good NewsがBad Newsより有意にリターンが高い
- 発信日前の方がリターン格差が大きい

分析結果

～ニュースによるリターン格差の検証～

➤ 企業規模を用いた詳細な分析

- 企業規模を用いて、データを分類して分析(企業規模は、東証が公表している規模2分類を使用)
- 規模ごとに効果が変わるかどうか分析

【大型株に対する分析】

		Return -30day	Return -20day	Return -10day	Return +10day	Return +20day	Return +30day
平均 リターン	Good News	1.88%	1.17%	0.65%	0.51%	1.24%	2.22%
	Bad News	0.40%	-0.26%	-0.72%	0.39%	0.83%	0.65%
	差	1.48%	1.43%	1.37%	0.12%	0.42%	1.57%
Welch 検定	t値	5.34	6.47	7.92	0.77	2.09	6.34
	検定結果	GOOD	GOOD	GOOD	差がない	GOOD	GOOD

【中型株に対する分析】

		Return -30day	Return -20day	Return -10day	Return +10day	Return +20day	Return +30day
平均 リターン	Good News	2.12%	1.66%	1.02%	0.34%	0.47%	0.87%
	Bad News	-2.72%	-2.39%	-1.54%	-0.59%	-0.49%	-0.26%
	差	4.84%	4.04%	2.56%	0.92%	0.96%	1.13%
Welch 検定	t値	12.00	12.03	10.59	4.28	3.26	3.26
	検定結果	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD

【小型株に対する分析】

		Return -30day	Return -20day	Return -10day	Return +10day	Return +20day	Return +30day
平均 リターン	Good News	1.64%	1.86%	1.85%	-0.24%	-0.35%	-0.30%
	Bad News	-4.25%	-3.66%	-2.95%	-1.68%	-2.38%	-3.87%
	差	5.88%	5.52%	4.80%	1.44%	2.03%	3.58%
Welch 検定	t値	14.12	15.25	15.45	5.49	5.72	8.72
	検定結果	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD

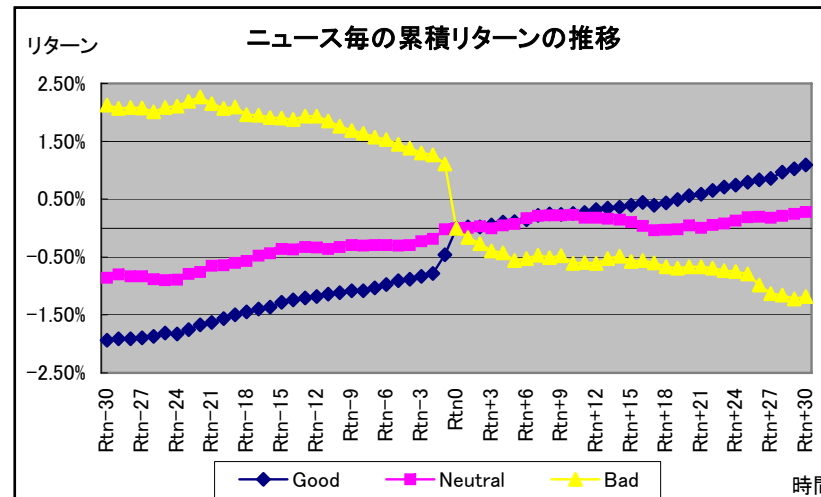
- 企業規模によらず、Good Newsの方がBad Newsよりリターン高い
- リターン格差は小型株の方が大きい(⇔アナリストレポートの結果と異なる)

分析結果

～ニュースによるリターン格差の検証～

➤ 累積リターン推移の分析

- 各グループにおける累積リターンの日次平均値を求め、グラフ化



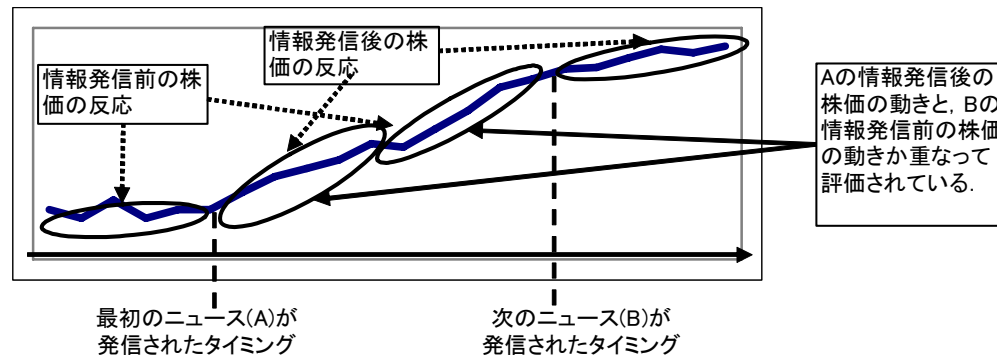
- Good Newsは右上がり、Bad Newsは右下がりとなっている
- Bad Newsの下落率の方がGood Newsの上昇率より大きい
- ニュースが発信される前の時点から、株価が反応
- ニュース発信日の当日に大きく反応している
 - ニュースの発信時間および日中の価格変動を用いた分析が必要

分析結果

～ニュースによるリターン格差の検証～

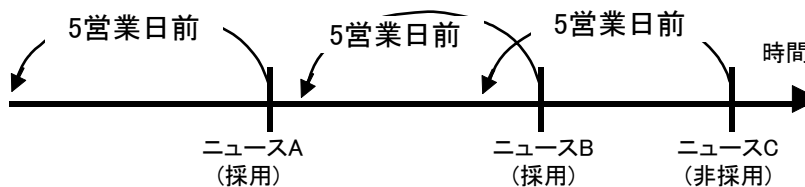
➤ ニュースの重なり処理

- ニュースの更新前から、株価が反応
 - 同一の内容に対するニュースが連続して更新されている可能性がある



- ニュースの重なりによる影響を排除するために、過去5営業日に同一タイプのニュースが更新されているかどうかをチェックし、更新されていないニュース(新しいニュース)を抽出

ある企業に対するニュースの更新履歴

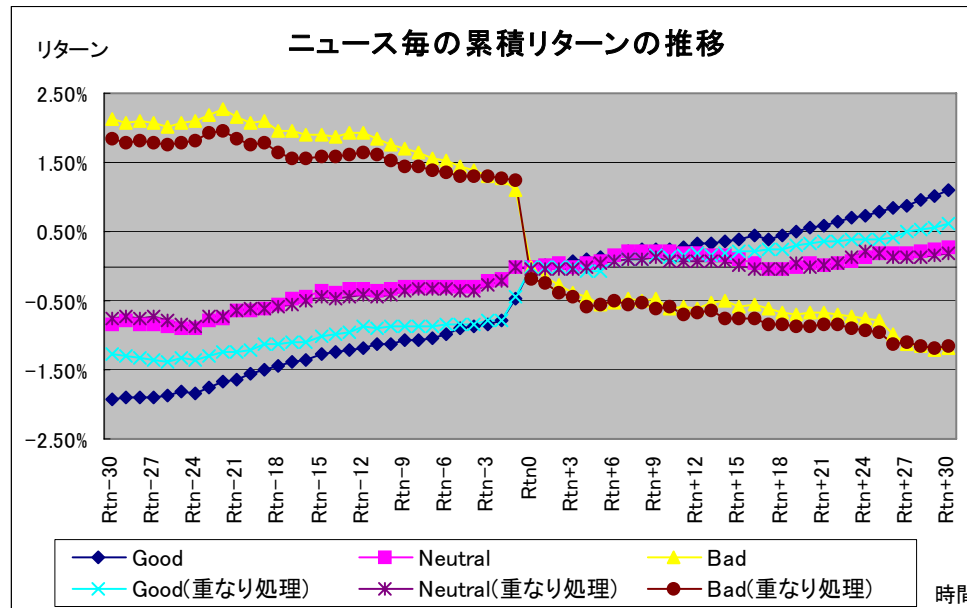


		Good News	Neutral News	Bad News
レコード数	重なり処理なし	10,098	4,463	2,418
	5営業日の重なり処理	3,276	2,344	1,173
減少率	5営業日の重なり処理	67.56%	47.48%	51.49%

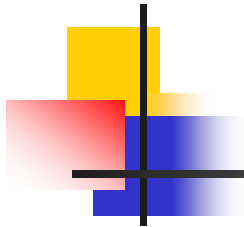
分析結果

～ニュースによるリターン格差の検証～

- 累積リターン推移の分析(情報の重なり処理後)



- ニュースに対する株価の反応の方向は、重なり処理前と同様
- Good News, Bad Newsのニュース発信日前の反応は抑制
- Neutral Newsは重なりの処理をしても反応は変わらない



分析結果

～ニュースの内容の検証～

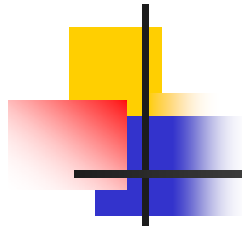
➤ ニュースの内容を詳細に検証

- Bad Newsに分類されても、株価が上昇するニュースが存在
- Bad Newsの中で、株価が下落したものと上昇したものの特徴を分析
 - Bad Newsを、ニュース更新後に株価が下落したニュースと上昇したニュースに分類
 - 分類したニュースの中に3%以上出現するキーワードについて、分類毎に出現頻度が有意に異なっているかどうか分析

【Bad内での単語のばらつき】

NO	キーワード	キーワード出現数		キーワード出現率		カイ2乗 検定統計量	NO	キーワード	キーワード出現数		キーワード出現率		カイ2乗 検定統計量
		Bad Newsで 株価下落	Bad Newsで 株価上昇	Bad Newsで 株価下落	Bad Newsで 株価上昇				Bad Newsで 株価下落	Bad Newsで 株価上昇			
1	下方修正	382	135	52.982%	31.765%	26.7	12	第三者割当	31	11	4.300%	2.588%	2.1
2	事故	57	64	7.906%	15.059%	13.0	13	低下	22	20	3.051%	4.706%	2.0
3	最終赤字	65	14	9.015%	3.294%	12.7	14	リコール	21	19	2.913%	4.471%	1.9
4	問題	60	62	8.322%	14.588%	9.9	15	引き下げ	40	16	5.548%	3.765%	1.7
5	再発防止	16	23	2.219%	5.412%	8.0	16	請求	34	25	4.716%	5.882%	0.7
6	損害	26	29	3.606%	6.824%	5.8	17	注意	25	11	3.467%	2.588%	0.7
7	対策	25	28	3.467%	6.588%	5.6	18	容疑者	36	17	4.993%	4.000%	0.6
8	恐れ	24	24	3.329%	5.647%	3.4	19	下落	25	14	3.467%	3.294%	0.0
9	けが	23	23	3.190%	5.412%	3.3	20	容疑	46	28	6.380%	6.588%	0.0
10	処分	35	32	4.854%	7.529%	3.3	21	逮捕	40	23	5.548%	5.412%	0.0
11	回収	44	37	6.103%	8.706%	2.6							

■ 有意水準5%



分析結果

～ニュースの内容の検証～

➤ ニュースの内容を詳細に検証

- Good Newsに分類されても、株価が下落するニュースが存在
- Good Newsの中で、株価が上昇したものと下落したものの特徴を分析
 - Good Newsを、ニュース更新後に株価が上昇したニュースと下落したニュースに分類
 - 分類したニュースの中に3%以上出現するキーワードについて、分類毎に出現頻度が有意に異なっているかどうか分析

【Good内での単語のばらつき】

NO	キーワード	キーワード出現数		キーワード出現率		カイ2乗 検定統計量	NO	キーワード	キーワード出現数		キーワード出現率		カイ2乗 検定統計量
		Bad Newsで 株価下落	Bad Newsで 株価上昇	Bad Newsで 株価下落	Bad Newsで 株価上昇				Bad Newsで 株価下落	Bad Newsで 株価上昇			
1	成長	73	118	3.862%	5.606%	6.3	12	トップ	68	87	3.598%	4.133%	0.7
2	開発	193	265	10.212%	12.589%	4.9	13	増収	75	74	3.968%	3.515%	0.5
3	提供	82	67	4.339%	3.183%	3.6	14	拡大	244	255	12.910%	12.114%	0.5
4	引き上げ	114	157	6.032%	7.458%	3.0	15	提携	124	150	6.561%	7.126%	0.5
5	伸び	104	139	5.503%	6.603%	2.0	16	評価	73	90	3.862%	4.276%	0.4
6	新商品	94	125	4.974%	5.938%	1.7	17	向上	87	90	4.603%	4.276%	0.2
7	回復	75	68	3.968%	3.230%	1.5	18	安定的	115	135	6.085%	6.413%	0.2
8	好調	184	230	9.735%	10.926%	1.4	19	堅調	82	97	4.339%	4.608%	0.2
9	買い付け株数	80	74	4.233%	3.515%	1.3	20	増益	81	85	4.286%	4.038%	0.1
10	買収	57	77	3.016%	3.658%	1.2	21	上方修正	491	558	25.979%	26.508%	0.1
11	改善	102	129	5.397%	6.128%	0.9	22	期待	146	167	7.725%	7.933%	0.1

有意水準5%



まとめ

- ヘッドラインニュースは、資産運用においてもっとも即時性のある情報であり、データ量は非常に大きい
- ナイーブベイズ分類法により、ヘッドラインニュースを自動分類し、約80%の分類精度を持つアルゴリズムを構築
- Good Newsに分類されたニュースのニュース対象銘柄の株価リターンは、Bad Newsに分類されたニュースのニュース対象銘柄の株価リターンより高い
- ニュースのリターン格差は、ニュース発信前から発生
- 小型企業の方がニュースの効果が大きい
- ニュース公開当日の反応が最も大きい



今後の取り組み

- ニュース分類方法の高度化
 - ニューラルネット等のより分類能力の高いアルゴリズムの利用
- ニュース分類と株価リターンによる詳細な分析
 - Good News(Bad News)の中で, 株価が上昇したニュースと下落したニュースの特徴を分析
- 業種等の属性を利用した詳細な分析
- アナリストレポートのテキスト情報と組み合わせた分析
- 会計情報等の数値データを組み合わせた分析
- イントラデイの株価を利用し, 日中のニュース効果を分析



サンプル

▶ サンプル1

● あ

- あ
- あ
- あ
- あ

▶ あ

● あ

- あ
- あ
- あ
- あ