

# ニュース情報が資産価格に反映されるプロセスの解明ーディープラーニングによるテキスト分析ー

研究代表者 高橋大志 慶應義塾大学 大学院経営管理研究科 教授  
共同研究者 五島圭一 東京工業大学 大学院総合理工学研究科 博士課程

## 1 はじめに

資産運用の分野において資産価格の評価は主要な関心事の一つである<sup>16 19 20 6</sup>。近年、計算機科学の急速な進展および分析可能なデータの多様化などを背景として、株式市場などを対象とした市場分析においても計算機を活用した分析手法が関心を集めている。資産価格に関する分析はこれまで数多く行われてきているが、それら研究の多くが比較的取り扱いの容易である数値データを対象としたものであった。近年の計算機科学の進展を背景とし、数値データ以外のデータを対象とした分析も可能となっている<sup>12 12 13</sup>。本研究でもこれらの背景を受け、ニュース記事を対象とした分析を試みている<sup>8 9 14 17</sup>。ニュース記事を用いた分析は、資産価格に関する情報が、どのようなプロセスで価格に反映されていくのかという資産価格理論における中心的な問いと密接に関連するものであり、その重要性は大きく、とりわけ近年、社会科学の分野においても急速に学術的関心は高まっている<sup>22 23 24</sup>。

## 2 データ

本分析では、分析対象としてトムソンロイターが提供するニュースデータ（以下ロイターニュース）を採用した。ロイターニュースは、メディアが発信する情報に分類されるものであるが、そのようなメディアが発信する情報は、各団体が発信する一次情報に比べると、各メディアの記者やアナリストによる情報の取捨選択が行われており、社会や市場に対して相対的に重要な情報が含まれていると考えられる<sup>18</sup>。本データは、金融市場における価格付けにおいて主要な役割を果たしている機関投資家の基にも配信されている膨大な量のデータであり、即時性の高いニュースデータであることから、資産価格にも高い影響力を有することが期待されるデータである。

分析においては、2003年から2014年の期間の配信されたニュースデータを対象として分析を行った。ロイターニュースには、ニュースの属性に応じて、いくつかの分類が与えられているが、本分析では、即時性の高いニュース記事に対する資産価格の反応を分析するため、再送記事と訂正記事については分析から除外した。また、ニュース記事のテキスト情報に注目したため、決算情報のみのニュース記事についても、除外して分析を行うこととした。

## 3 分析結果

### 3-1 テキスト分析システム

図1は、テキスト分析システムの概略を示したものである。当システムは大きく3つのプロセスに分けられる。図中の「ニュース記事」「ニュース記事のベクトル表現」の部分は、ニュース記事を分析するプロセスである。当プロセスにおいては、ニュース記事のベクトル表現を、Bag of WordsやRecursive Neural Networkなどの手法により獲得する<sup>21 24 15</sup>。図中の「資産価格」「ニュース記事の評価」の部分は、資産価格の変動をもとにニュース記事の評価を測定するプロセスである。当プロセスにおいては、ニュースに対する市場の評価を資産価格の変化を基に計測する。「分類アルゴリズム」以下の部分は、両者の情報を基に記事の分類を行うプロセスである。当プロセスにおいては、ニュースのベクトル表現およびニュースに対する評価を基に、ニュース記事の分類アルゴリズムを構築し、その評価を行う。評価においては、分析対象となるニュース記事のベクトル表現を獲得し、それらの情報を基に記事の分類を行い、ニュース発信前後における資産価格の

挙動を分析する。

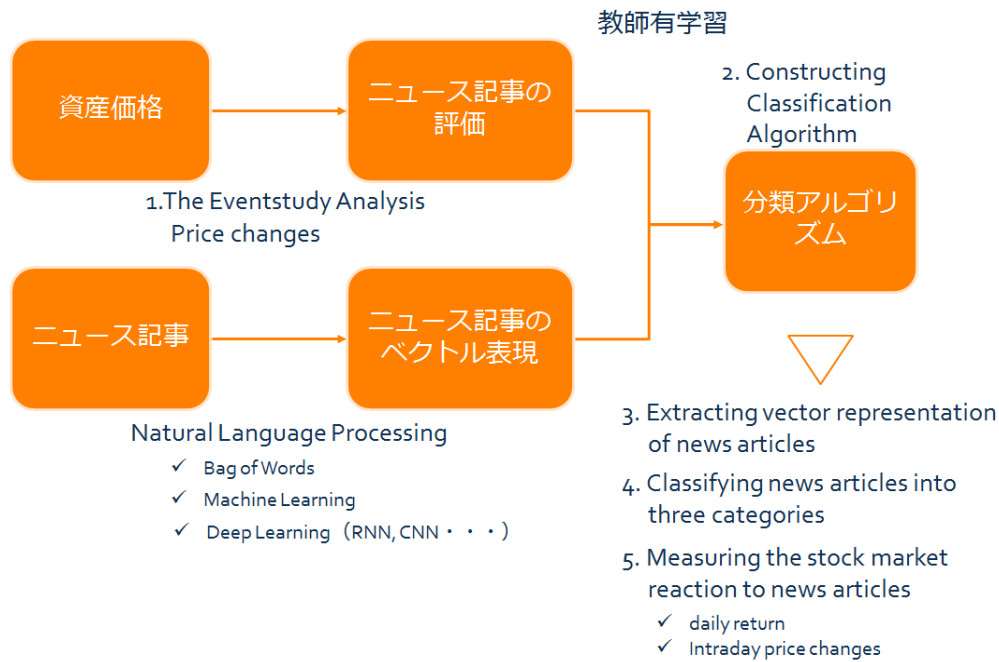


図1 テキスト分析システムの概略

### 3-2 分析結果

はじめに、ニュースを内容に基づきポジティブな記事、ネガティブな記事に分類し、それら分類結果を基に資産価格の関連性について分析を行った。具体的には、ニュースを5つに分類し (Very Positive, Positive, Neutral, Negative, Very Negative)、分析を実施した。本分析においては、ディープラーニングモデルの一つである Recursive Neural Network による分析を行った<sup>21 15</sup>。当モデルは、テキストデータの評判分析において高い性能が報告されているモデルの一つである。東京株式市場に関する英文ニュース記事を分析対象として分析を実施し、ニュースと株式価格が双方向に影響を与えていることを確認した。

更に、本分析では、より短期間の価格変動に焦点を当てた分析を実施するために日中データを採用した分析を実施した。分析においては、東京証券取引所に上場されている企業の日本語ニュースを対象とした分析を実施した。ニュースが記事内容としてとりあげる企業には、ばらつきが存在する。本分析では、相対的に数多くのニュース記事が配信されている主要な自動車メーカーを対象として分析を行った。なお、ニュースの中には、複数の企業の内容について報じているニュース記事も存在するが、本分析では、付随している企業数の分だけニュース記事を増やし、一つのニュース記事に一つの企業を対応付け、分析を進めることとした。本分析のニュースの分類においては機械学習の代表的な手法の一つであるサポートベクターマシンを採用した<sup>3</sup>。

本分析では、ファイナンスの分野において広く用いられているイベントスタディの手法を用いてニュースのアナウンスメントに対する株式価格変化の計測を行った<sup>11 7</sup>。図2は、当プロセスの概略を示したものである。推定ウィンドウにおいて、ファクターリターンと個別株式のリターンの関連性を見積もり (ベータの見積もり)、各資産のベータおよびファクターリターンから、ニュースのアナウンスメントに対する株式価格の変化 (異常リターン) の計測を行う。本分析では、ここで計測した異常リターンの値をニュース記事のアナウンスメントに対する株式市場の評価として分析に採用する。

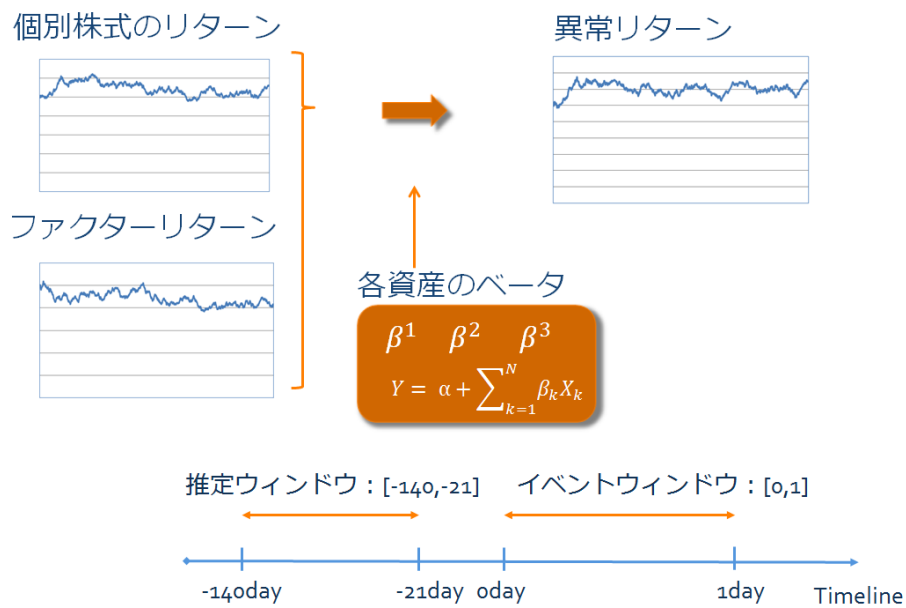


図2 ニュースに対する株式価格変化の測定

図3は、分類アルゴリズム構築の概略を示したものである。ニュース記事から獲得したベクトル表現とイベントスタディの手法を用い計測したニュースのアナウンスメントに対する市場の評価の情報を用い、ニュース記事の分類アルゴリズムを構築する。

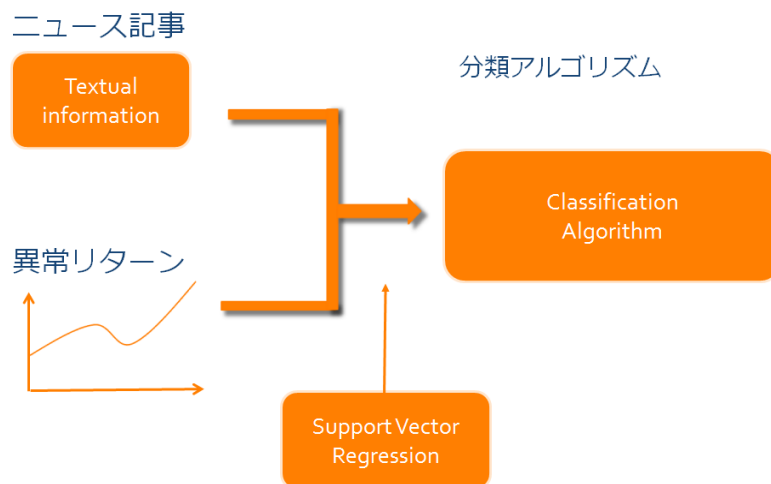


図3 分類アルゴリズムの構築

図4は、ニュース記事のベクトル表現のイメージを示したものである。ニュース記事のベクトル表現の方法に関しては、いくつかの方法が挙げられるが、同図は、オリジナルのニュース記事に含まれる単語を基にベクトル表現を行ったものである (Bag of Words)<sup>21,14,15</sup>。これらベクトル表現および構築した分類アルゴリ

ズムを基に、各記事に対する評価の算出が行われる。本分析では、Positive、Neutral、 Negative の3つのクラスに分類する分類アルゴリズムを構築し、各ニュース記事の分類を行った(図5)。

## オリジナルのニュース記事

- News.1 : Due to fall in the yen, stock prices of export companies have risen.
- News.2 : The increase in the consumption tax lead to falling in consumptions in retail.
- News.3 : A corp increased B corp's capital.
- News.4 : Today corporate social responsibility is important theme for management.

## ニュース記事のベクトル表現

	capital	compa	consu	consu	corptio	corp's	corp	corpor	ate	due	export	fall	falling	for	have	import	incleas	increa	lead	mana
	.	nies	n	ns	ns	corp's	corp	ate	due	export	fall	falling	for	have	import	incleas	increa	lead	mana	
News.1	0.00	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	0.00
News.2	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
News.3	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
News.4	0.00	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	0.00



図4 ニュース記事のベクトル表現

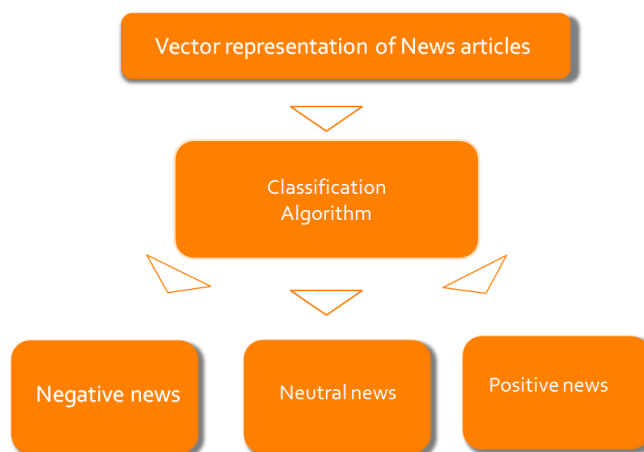


図5 ニュース記事のアルゴリズムによる分類

図6は、分類されたニュース記事のアナウンスメントに対する株式価格変化の計測の概略を示したものである。分類アルゴリズムにより3つのクラス(ポジティブ、ニュートラル、ネガティブ)にニュースを分類した後、各クラスの市場の反応を計測する。例えば、ポジティブと分類されるニュース記事のアナウンス後は、株式市場はポジティブに反応することが期待される。一方、ネガティブと判断されるニュース記事のアナウンス後は、株式市場はネガティブに反応することが期待される。

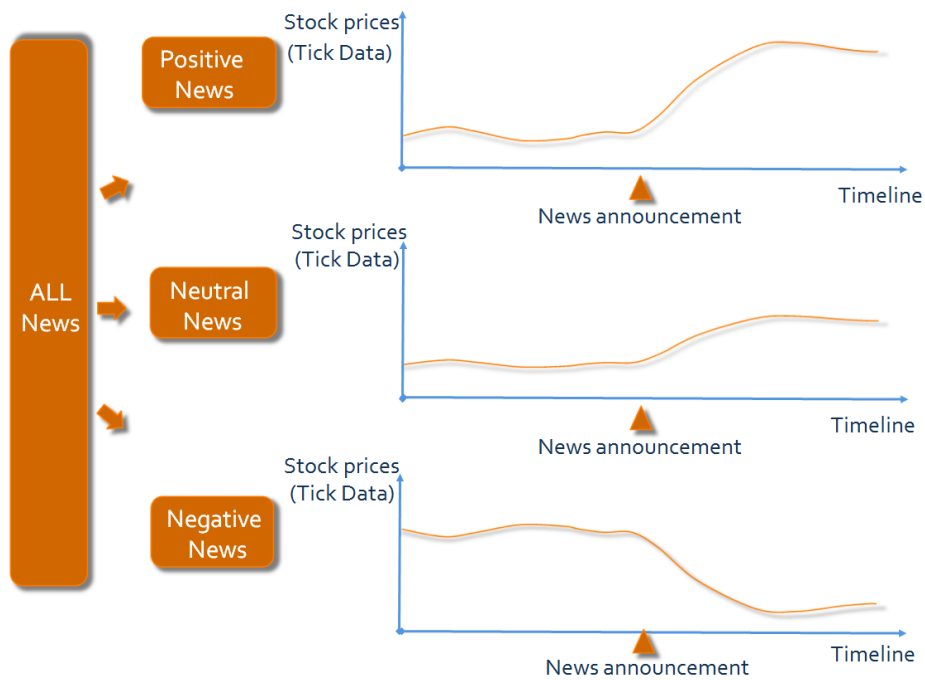


図6 ニュース記事に対する市場の反応の計測

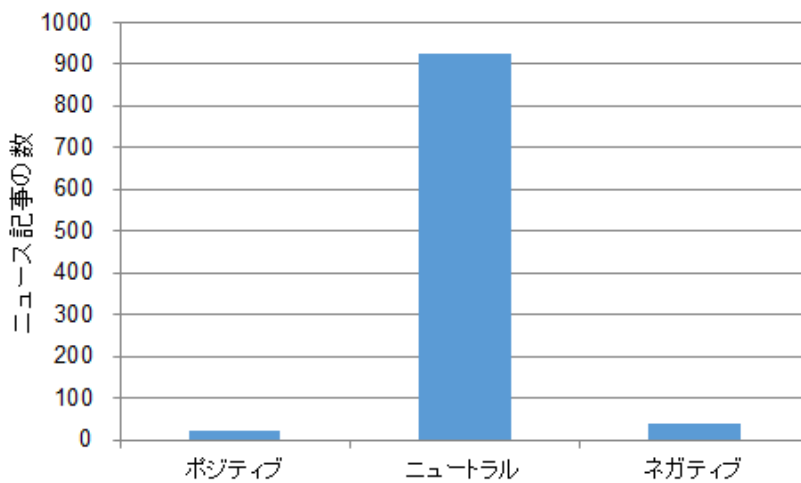


図7 各分類のニュース記事数

図7は、分類アルゴリズムにより各クラスに分類されたニュース記事数を示したものである。図よりニュートラルに分類される記事が多く、ネガティブ、ポジティブに分類されるニュース記事数は相対的に小さいことを確認できる。図8は、各クラスのニュース記事に対する株式市場の反応を示したものである（累積異常リターンの推移）。図中の横軸は時間、縦軸は株式市場の反応を示したものである。図中の赤線は、ポジティブと分類されたニュース記事に対する株式市場の反応を示したものである。ニュースのアナウンスメント

の直後に、大きくその値が上昇していることを確認できる。これらの結果は、ニュースのアナウンスメントに対し市場が速やかに反応していることを示す結果である。図中の青線は、ネガティブと分類されたニュース記事に対する反応を示したものである。こちらに関しては、ニュースのアナウンスメントの直後に、大きくその値が下落していることを確認できる。これらの結果は、市場において利用可能な情報を資産価格が速やかに反映していることを示す結果である。当結果は、伝統的な資産価格理論において中心的な仮説である市場の効率性と密接に関連した議論であり、情報と資産価格の関連性について興味深い結果を示したものとなっている<sup>10 4 5</sup>。

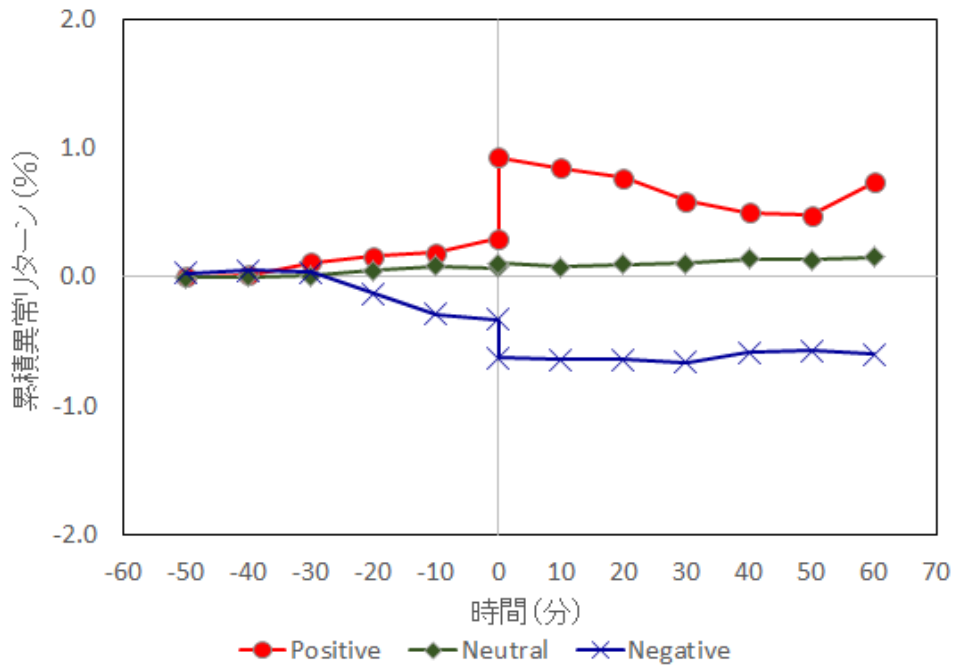


図8 ニュース記事に対する市場の反応(自動車会社)

#### 4 おわりに

本研究は、電気通信によりタイムリーに投資家に届けられるニュースデータを対象として、情報が価格に反映されるプロセスについて分析を行った。分析においては、ディープラーニング、日中価格データを用いた分析を通じ、情報と資産価格の関連性について分析を行った。情報の果たす役割は、資産価格理論において重要な役割を果たしている。本研究は、近年急速に進展してきている情報技術を活用することで、情報と価格に関し興味深い結果を見出すことができる可能性を示唆するものである。とりわけ、ニュース記事のアナウンスメントに対し、資産価格が速やかに反応している可能性を示した結果は、伝統的な資産価格理論における主要な関心事である市場の効率性と密接に関するものであり、実務的、学術的に、興味深い結果である。

今後の課題としては、分析対象の拡大(分析対象企業の拡大、対象国の拡大)、多様な手法を用いた分析(CNN, RNN 等)などが挙げられる。とりわけ、近年、分析手法に関しては、数多くの手法が活発に提案されていることから、これらの成果を取り込むことで、情報技術および資産価格理論に関し、興味深い研究が行われることが期待される。

## 【参考文献】

- [1] Allee, K. and M.D. DeAngelis: The Structure of Voluntary Disclosure Narratives: Evidence from Tone Dispersion. *Journal of Accounting Research* 53(2), 241--274 (2015)
- [2] Antweiler, W., and Frank, M.Z.: Is All That Talk Just Noise? The Information Content of Internet Stock Message Board. *Journal of Finance* 59(3), 1259--1294 (2004)
- [3] Bishop, C. M. : Pattern Recognition and Machine Learning. Springer (2006)
- [4] Black, F. and Scholes, M. : The Pricing of Options and Corporate Liabilities. *Journal of Political Economy* 81(3), 637--654 (1973)
- [5] Brealey, R.A., Myers, S.C. and Allen, F.: Principles of Corporate Finance, McGraw-Hill (2010)
- [6] Campbell, J.Y., Grossman, S.J., and Wang, J.: Trading Volume and Serial Correlation in Stock Returns, *Quarterly Journal of Economics* 108(4), 905--939 (1993)
- [7] Campbell, J. Y., Lo, A. W. and MacKinlay, A. C. : The Econometrics of Financial Markets, Princeton University Press (1997)
- [8] Dougal, C., Engelberg, J., Garcia, D., and Parsons, C. A.: Journalists and The Stock Market, *Review of Financial Studies* 25(3), 639--679 (2012)
- [9] Engelberg, J., Reed, A. V., and Ringgenber, M. C. : How are Shorts Informed? Short Sellers, News, and Information Processing, *Journal of Financial Economics* 105(2), 260--278 (2012)
- [10] Fama, E. : Efficient Capital Markets: A Review of Theory and Empirical Work, *Journal of Finance* 25(2), 383--417 (1970)
- [11] Fama, E. F. and French, K. R. : Common Risk Factors in The Returns on Stock and Bonds, *Journal of Financial Economics* 33(1), 3--56 (1993)
- [12] Garcia, D. : Sentiment during Recessions. *Journal of Finance* 68(3), 1267--1300 (2013)
- [13] Goshima, K., Takahashi, H., Terano, T.: Estimating financial words' negative-positive from stock prices. The 21st International Conference Computing in Economics and Finance (2015)
- [14] Henry, A. : Are Investors Influenced by How Earnings Press Releases Are Written? , *Journal of Business Communication* 45(4), 363--407 (2008)
- [15] Hinton, G.E., Osindero, S., and Teh, Y.: A Fast Learning Algorithm for Deep Belief Nets, *Neural Computation* 18(7), 1527--1554 (2006)
- [16] Ingersoll, J. E. : Theory of Financial Decision Making. Rowman & Littlefield (1987)
- [17] Kearney, C. and Liu, S. : Textual Sentiment in Finance : A Survey of Methods and Models, *International Review of Financial Analysis* 33, 171--185 (2014)
- [18] Healy, A.D. and Lo, A.W. : Managing Real-Time Risks and Returns: The Thomson Reuters NewsScope Event Indices. In: Mitra, G. and Mitra L. (eds.) *The Handbook of New Analytics in Finance*. John Wiley & Sons, West Sussex (2011)
- [19] Luenberger, D.G. : Investment Science. Oxford University Press, Oxford (2000)
- [20] Sharpe, W.F. : Capital Asset Prices:A Theory of Market Equilibrium under Condition of Risk. *Journal of Finance* 19(3), 425--442 (1964)
- [21] Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C., Ng, A., and Potts, C.: Recursive Deep Models for Semantic Compositionality Over Sentiment Treebank. *Conference on Empirical Methods in Natural Language Processing* (2013)
- [22] Tetlock, P.C. : Giving Content to Investor Sentiment:The Role of Media in the Stock Market. *Journal of Finance* 62(3), 1139--1168 (2007)
- [23] Tetlock, P.C., Saar-Tsechansky, M. and Macskassy, S. : More Than Words: Quantifying Language to Measure Firms's Fundamentals, *Journal of Finance* 63(3), 1437--1467 (2008)
- [24] Loughran, T. and McDonald, B. : When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *Journal of Finance* 66(1), 35--65 (2011)

〈発表資料〉

題 名	掲載誌・学会名等	発表年月
Text Analysis System for Measuring the Influence of News Articles on Intraday Price Changes in Financial Markets	10 <sup>th</sup> International KES Conference on Agents and Multi-agent Systems - Technologies and Applications	2016年6月15日
資産価格分析と計算機科学	大阪大学 数理・データ科学セミナー 金融・保険セミナーシリーズ 第75回	2016年9月29日
金融市場分析における計算機科学の活用	応用経済時系列研究会・チュートリアルセミナー	2016年10月24日