

金融市場の方向性とニュース報道の量的パターン

田平好文¹, ○水野貴之^{2,3,4}

¹ 中央大学大学院理工学研究科

² 国立情報学研究所

³ 総合研究大学院大学複合科学研究科

⁴ JST さきがけ

Amount of news before stock market fluctuations

Y. Tahira¹, ○T. Mizuno^{2,3,4}

¹ Chuo University

² National Institute of Informatics

³ The Graduate University for Advanced Studies (SOKENDAI)

⁴ JST PRESTO

概要— 近年, Wikipedia で Dow Jones Industrial Average (DJIA)の構成銘柄である 30 社が検索された回数と, 将来の DJIA の変動に相関があることが示された. 検索数の上昇は, それらの企業に関する新しい情報が報じられたことを示唆する. 本論文では, 2007年12月から2012年4月までの THOMSON REUTERS 社によって配信された全ての英語ニュース記事, 915 万記事, と主要株式市場の株価指数を用いて, ある株式市場に関連するニュース記事数の変動と将来のその市場の株価指数の変動に相関があることを明らかにする. ある週に NYSE と NASDAQ に上場する銘柄に関連するニュース記事数が増加/減少すると, その翌週に S&P500 は下降/上昇する傾向にあるが, 一方で, NYSE や NASDAQ と関連の低い他の市場の株価指数の変動はランダムに近い. 特に, 企業の業績と密接に関係する Business sector のニュース記事数の変動には, 市場は顕著に反応する. これらの傾向は各年で観測される.

キーワード: 経済物理学, 株式市場, ビジネスニュース, 外生的ショック

1 はじめに

金融市場ではしばしば正規分布に従わない激しい価格変動が起き, 市場経済を混乱させている. 計量ファイナンスや経済物理学では, このような激しい変動が生じるメカニズムについて研究している.

金融市場の価格は, トレンドフォロワーが作り出す過去の価格変動を追従するような内生的なメカニズムと, 突発的に起きたニュースに対する反応に代表される外生的なメカニズムによって動いている^(1,2). 過去の取引価格を記録したTickデータの分析によって, 主要な内生的なメカニズムが乗算過程に従っていることが明らかにされた⁽³⁾. 乗算過程は価格変動分布のファットテイルを生み出す. 一方で, 外生的なメカニズムについても, ニュース報道のアーカイブと価格のHistorical Tickデータを用いて研究されている^(4,5,6,7,8,9). ニュース本文に対する自然言語処理により, ニュースのセンチメントやトピックを抽出し, それらと価格変動に相関があることが示された^(10,11,12). さらに, ニュースの内容そのものではなく, ニュースに対

する人々の反応と価格変動の関係についても調査されている.

行動経済学におけるプロスペクト理論では, 人々は株の売買により利益を得ることよりも損失を回避することを優先的に考えると仮定する⁽¹³⁾. この仮定に従えば, 人々は株価に下落のおそれがある時, その株に関連する話題に関してWeb上で必死に情報収集を行い, 下落による損失を防ごうとする. つまり, 株価が下落する前にその株に関連する話題のWeb上での検索数・閲覧数の増加が想定される. Preisらは, Dow Jones Industrial Average (DJIA)の構成銘柄である30社についてのWeb上での検索数・閲覧数の変動とその後のDJIAの変動に関連があり, DJIAの下落の前に上記の話題に関する検索数・閲覧数が増加することを示した^(14,15,16).

本論文では, ある株式市場に上場する企業についてのニュースの記事数の変動とその後のその市場の株価指数の変動に相関があることを示す. 様々な報道機関は人々が注目するであろう話題を中心に報道するため, Web上での検索数・閲覧数と同様に株価指数の下落の前に特定の話題に関するニュースの記事数が増加すると

考えられる。この仮定をTHOMSON REUTERS社によって投資家向けに配信された全ての英語ニュース記事と株価指数の時系列を用いて証明する。

以下の章は次のように構成される。2章では、本論文で扱うTHOMSON REUTERS社のニュースデータについて述べる。3章では、ニュースの記事数の変動と株価指数の変動の関係を調べるメソッドについて述べる。4章では、NYSEとNASDAQの上場銘柄に関するニュースの記事数の変動とS&P500の変動に関する関係があることを示す。5章と6章では、この関係のニューストピックおよび、時期の依存性を調査する。7章はまとめである。

2 ビジネスニュースデータ

我々は世界最大の金融情報サービスTHOMSON REUTERS社が機関投資家向けに配信する2007年12月から2012年4月までの英語のニュースを用いる^(17,18)。ニュース記事はALERTとHEADLINEとSTORY_TAKE_OVERWRITEの3種類に分類される。ALERTは緊急性の高いニュースで、記事にはタイトルのみが含まれる、HEADLINEはその他のニュースで、同じく記事にはタイトルのみが含まれる。STORY_TAKE_OVERWRITEはALERTとHEADLINEが配信された後に配信されるそれらの記事の本文である。我々は、実質的なニュース記事数を数えるためにALERTとHEADLINEに分類される915万記事を使う。

それぞれのニュース記事にはトピックコードと呼ばれる記者によってつけられたキーワードが付随する。各トピックコードは”Arts / Culture / Entertainment”, “Asset Class / Property”, “Business sector”, “Commodity”, “Crime”, “Currency”, “Disaster / Accident”, “Event Type”, “Genre”, “Geography”, “Health / Medicine”, “Indicator Type”, “Intellectual Property”, “Language”, “Legacy News Topic”, “News Flag / Status”, “Organization”, “Physical Asset Type”, “Religion”, “Sport”, “Sport combined with Geography”, “Sporting Competition”の22分類に属する。また、それぞれのニュースには銘柄コードと呼ばれる関連銘柄を示すコードもある。

本論文では、2007年12月から2012年4月にニュース記事が100週以上存在しない銘柄・トピックは分析に利用しない。この条件に該当しないNYSEとNASDAQの銘柄数とトピック数はそれぞれ500銘柄と676個である。

3 手法

PreisらのGoogleの検索数・Wikipediaの閲覧数にもとづいた取引戦略^(14,15,16)をニュース記事数に適用することによって、ある特定の株式市場に関するニュース記事数の変動とその後の株価指数の変動との関係を明らかにする。

ある週 t の上場企業 i に関するニュースの記事数を $n_i(t)$ とする。ここで、週の終わりは日曜日とする。次に、過去 Δt 週間の記事数の平均 $N_i(t-1, \Delta t) = \sum_{\tau=1}^{\Delta t} n_i(t-\tau) / \Delta t$ を計算する。 t 週の記事数が過去より増加した $n_i(t) > N_i(t-1, \Delta t)$ とき、 $t+1$ 週のはじめにその企業が上場している株式市場における1単位の株

価指数を $p(t+1)$ で売り、 $t+2$ 週のはじめに $p(t+2)$ で買い戻す。逆に、 t 週の記事数が過去より減少した $n_i(t) \leq N_i(t-1, \Delta t)$ とき、 $t+1$ 週のはじめに1単位のその株価指数を $p(t+1)$ で買い、 $t+2$ 週のはじめに $p(t+2)$ で同じだけ売る。我々は、この売買戦略を特定の期間で行い、累積収益率を計算する。我々は累積収益率をランダムに売買を選択する戦略で得られる累積収益率の標準偏差で規格化する。

我々は、ある株式市場に上場する全ての企業について規格化された累積収益率 R_i を算出する。 R_i の分布が、ニュースの記事数の変動に依存せずにランダムに売買を選択する戦略で得られる規格化された累積収益率の分布と異なることを示すことによって、我々は、ニュース記事数の変動がその後の株価指数の変動と有意に関係していることを明らかにすることができる。

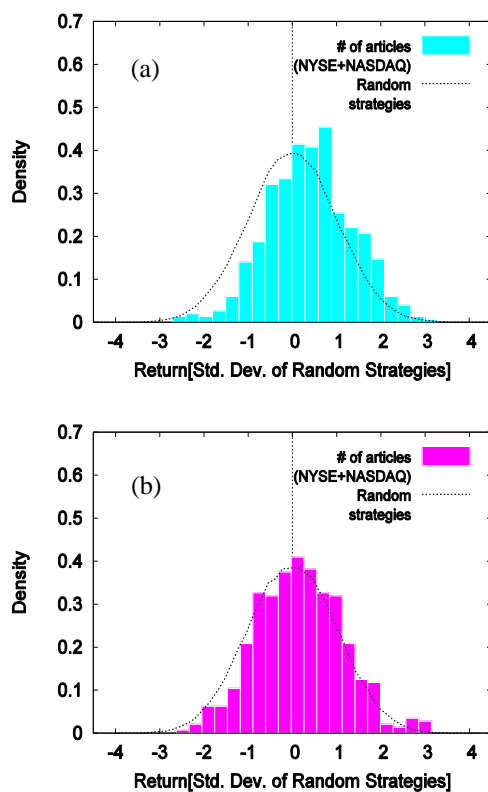


Figure 1 NYSEとNASDAQの上場企業に関するニュース記事数の変動にもとづいて(a)S&P500と(b)日経225を売買したときの累積収益率 R の確率密度分布。破線の分布は、ランダムな売買による累積収益率 R の分布を示す。売買戦略のパラメータ Δt に7週をセットした(3節)。S&P500と日経225を売買したときの平均累積収益率(R)は、それぞれ、0.40と0.17である。破線の分布を帰無仮説としたKolmogorov-Smirnov検定の p 値は、それぞれ、 1.58×10^{-11} と0.041である。

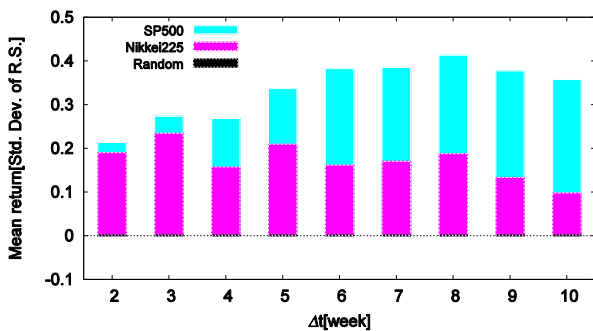


Figure 2 NYSE と NASDAQ の上場企業に関する過去 Δt 週間のニュース記事を用いた売買戦略による平均累積収益率 $\langle R \rangle$. 青と赤は、それぞれ、S&P500 と日経 225 を売買したときの平均累積収益率 $\langle R \rangle$ を表す. ほとんど $\langle R \rangle = 0$ を示す黒は、ランダムに売買したときの平均累積収益率である.

4 ニュース記事数の変動と株価指数の変動の関係

3 章で述べた売買戦略を用いることで、ある特定の株式市場に関するニュース記事数の変動とその後の株価指数の変動に関係があることを示す. 我々は、NYSE と NASDAQ と東京証券取引所と香港証券取引所と上海証券取引所と韓国証券取引所に上場する銘柄に関するニュース記事と S&P500 と日経 225 の指数を使用する.

Figure 1(a)はNYSE と NASDAQ に上場する銘柄に関する過去 $\Delta t = 7$ 週間のニュース記事数の変動にもとづいて S&P500 に対して売買戦略を実行したときの累積収益率 R_i の確率密度分布である. ランダムな売買戦略での平均累積収益率 1.3×10^{-3} に対して、これらの市場の上場企業で平均したニュースの記事数にもとづいた売買戦略の累積収益 $\langle R \rangle$ は高く 0.40 である. Kolmogorov-Smirnov 検定により、帰無仮説「ニュース数にもとづいた戦略での分布とランダムな売買戦略での分布は同じである」は棄却される. 一方で、Figure 1(b)はNYSE と NASDAQ に上場する銘柄に関するニュース記事数の変動にもとづいて、日経 225 に対して売買戦略を実行したときの累積収益率 R_i の確率密度分布である. 平均累積収益率 $\langle R \rangle$ は低く 0.17 である. そして、Kolmogorov-Smirnov 検定の p 値は 0.041 であり帰無仮説は有意水準 5% では棄却される. S&P500 での p-value= 1.58×10^{-11} よりは明らかに高く、よりランダムな売買での分布に近いことが分かる. 次に、異なる株式市場に関するニュース記事数にもとづく売買戦略を S&P500 の指数に対して実行する. 過去 $\Delta t=7$ 週間の東京証券取引所と香港証券取引所と上海証券取引所と韓国証券取引所に上場する銘柄に関するニュース記事数では平均累積収益率 $\langle R \rangle$ は 0.13 であり、NYSE と NASDAQ に関するニュース記事数の変動にもとづく売買戦略に比べて顕著に低い. また、Kolmogorov-Smirnov 検定の p 値は 0.15 で帰無仮説は有意水準 5% で棄却されない. これらの結果は、将来のある株価指数の変動は、過去の関係のより深い市場

のニュース記事の変動と相関があることを意味している.

我々は平均累積収益率 $\langle R \rangle$ の Δt への依存性を調査する. Figure 2 は過去 Δt 週間の NYSE・NASDAQ に関するニュース記事数の変動にもとづいて、S&P500 と日経 225 の指数を売買した場合の平均累積収益率 $\langle R \rangle$ である. ランダムな売買戦略に比べてニュース記事数の変動にもとづいた売買戦略の平均累積収益率 $\langle R \rangle$ は Δt に依存せず常に高い. また、すべての Δt について S&P500 の方が日経 225 よりも $\langle R \rangle$ が高い.

5 累積収益率のニューストピックへの依存性

各トピックのニュース記事数の変動にもとづいて売買戦略を実行することにより、どのようなトピックが将来の価格変動と関係があるかを明らかにする. 2 章で述べたトピック分類ごとに過去 $\Delta t = 7$ 週間のニュース記事数と S&P500 を用いて平均累積収益率 $\langle R \rangle$ を算出する*.

Table 1 は各トピック分類に含まれるトピック数とそれらの平均累積収益率 $\langle R \rangle$ を表す. 平均累積収益率 $\langle R \rangle$ は"Business sector"が一番高く 1.10 で、"Sport & Sporting Competition"が一番低く 0.30 である. "Business sector"の累積収益率の分布はランダムな売買の累積収益率の分布と有意に異なり、Kolmogorov-Smirnov 検定の p 値は 1.57×10^{-33} を示す. 一方、"Sport & Sporting Competition"の累積収益率の分布は Kolmogorov-Smirnov 検定の p 値が 0.25 であるから、ランダムな売買の累積収益率の分布に近い. 企業の業績と密接に関係する Business sector のニュース記事数の変動には、市場は顕著に反応する. Business sector でも、特に、Table 2 で示される 10 トピックは累積収益率が高い.

*ただし、トピック数が 10 未満である下記の分類は解析に用いない. ("Genre", "News Flag / Status", "Legacy News Topic", "Indicator Type", "Arts / Culture / Entertainment", "Language", "Health / Medicine", "Disaster / Accident", "Physical Asset Type", "Crime", "Religion", "Currency", "Intellectual Property", "Sport combined with Geography")

Table 1 トピックの分類におけるトピック数と平均累積収益率 . 売買戦略のパラメータ Δt に7週をセットした(3章). 2007年12月から2012年4月までの230週間の中でニュースがない週が100週以上あるトピックは除いた.

トピックの分類	トピック数	Mean of return[σ of R.S.]
Business Sector	199	1.10
Event Type	24	1.08
Commodity	53	0.93
Organization	16	0.46
Asset Class / Property	85	0.37
Geography	207	0.33
Sport & Sporting competition	25	0.30

Table 2 Table 1 の Business sector のニュースにおいて累積収益率が高いトップ10トピック.

	Topic (in Business sector)	Return [σ of R.S.]
1	Metal / Mining	3.58
2	Entertainment Production	3.46
3	Food Processing	3.17
4	Food & Drug Retailing	3.01
5	Leisure & Recreation	2.97
6	Aluminum	2.94
7	Real Estate	2.84
8	Restaurants	2.78
9	Utilities	2.74
10	Media / Publishing	2.50

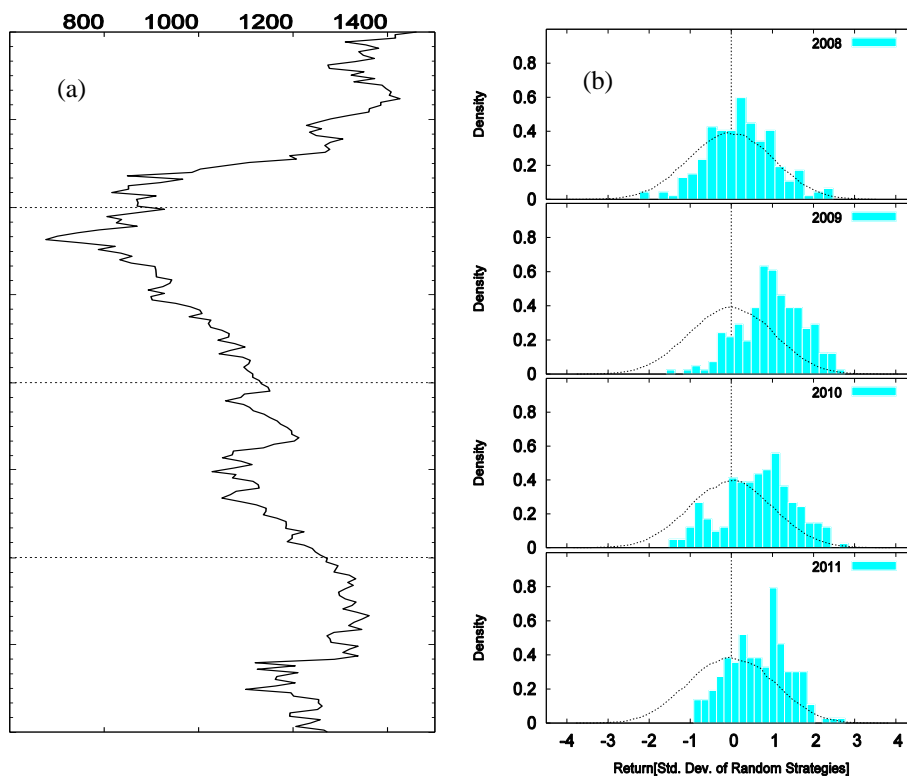


Figure 3 (a) 2008年から2011年までのS&P500の時系列, (b)Business sectorに属するトピックのニュース記事数の変動にもとづいてS&P500を売買したときの年間累積収益率 R の確率密度分布. 破線の分布は, ランダムな売買による累積収益率 R の分布を示す. 売買戦略のパラメータ Δt に7週をセットした(3章). 年間累積収益率の平均 $\langle R \rangle$ は, 2008年が0.22, 2009年が0.97, 2010年が0.66, 2011年が0.67である. Kolmogorov-Smirnov検定の p 値は, 2008年が 1.40×10^{-3} , 2009年が 6.03×10^{-35} , 2010年が 2.94×10^{-15} , 2011年が 5.33×10^{-16} であり, 全ての年で帰無仮説"ランダムに売買したときの累積収益率の分布と等しい"は有意水準1%で棄却される.

6 累積収益率の時期依存性

累積収益率の時期依存性を調べるために、2008年から2011年の各期間の累積収益率の分布を観測する。Figure 3(a)はS&P500の時系列である。2008年9月から2009年3月の時期と2011年8月から2011年10月の時期に大きな下落が起きている。Figure 3(b)はBusiness sectorに属するトピックのニュース記事数の変動にもとづいてS&P500を売買したときの年間累積収益率の確率密度分布である。平均累積収益率(R)は常に正の値を示す。大きな下落が起きている2009年と2011年は平均累積収益率が高い傾向にある。ランダムな売買による累積収益率の分布を帰無仮説とするKolmogorov-Smirnov検定では、帰無仮説は全ての年において有意水準1%で棄却される。つまり、時期ごとに収益率の差はあるが、ニュース記事数にもとづいた売買戦略は常に有効である。

7 まとめ

我々は、2007年12月から2012年4月までにTHOMSON REUTERS社が機関投資家向けに配信する英語のニュース915万記事と主要株式市場の株価指数を用いて、ある株式市場に関連するニュース記事数の変動と将来のその市場の株価指数の変動に相関があることを示した。ある週にNYSEとNASDAQに上場する銘柄に関連するニュース記事数が増加/減少すると、その翌週にS&P500は下降/上昇する傾向にあるが、一方で、NYSEやNASDAQと関連の低い他の市場の株価指数の変動はランダムに近い。特に、企業の業績と密接に関係するBusiness sectorのニュース記事数の変動には、市場は顕著に反応する。これらの傾向は各年で観測される。行動経済学のプロスペクト理論によると、投資家は損失回避を優先する。したがって、株価に下落のおそれがあるとき、投資家はその銘柄に関連する多くの情報を特に必要とし、報道機関はその要望に応じる。その結果として、ニュース量と株価指数の変動に関するこれらの統計的な特徴が現れると考えられる。

ニュースの記事数にもとづいた売買戦略はWeb上の検索数・閲覧数にもとづいた売買戦略と同水準の収益が得られる。例えば、Moat et al. (2013)によれば、2008年から2011年のWikipediaの”Outline of economics”の中の”General economic concept”に分類される全てのページの閲覧数を用いてDow Jones Industrial Averageを売買したときの平均累積収益率は、2008年が0.89、2009年が0.19、2010年が0.19、2011年が0.55である。一方、Business sectorに関するニュースを用いた売買戦略の平均累積収益率は、2008年が0.22、2009年が0.97、2010年が0.66、2011年が0.67である。若干の傾向の違いは、投資家向けニュースから情報を得る機関投資家とGoogleやWikipediaから情報を得る一般人

での行動の違いを表しているかもしれない。これを明確にすることは今後の課題である。

ニュース記事数の変動による市場価格変動の予測は株式市場以外でも可能であろう。例えば、Business sectorに関する日本語のニュース記事数を用いて日本国債10年物の金利を取引した場合、2007年12月から2012年4月までで、平均累積収益率(R) = 0.35を得ることができる。様々な金融市場とニュースの記事数の関連性を明確にすることも今後の課題である。

謝辞

水野は本研究の一部で、石井記念証券研究振興財団および、JSPS 科研費 24710156 の助成を受けた。

参考文献

- 1) Jiang, Z. Q., Guo, L., & Zhou, W. X.: Endogenous and exogenous dynamics in the fluctuations of capital fluxes. *The European Physical Journal B*, 57(3), 347/355 (2007)
- 2) Sornette, D.: Endogenous versus Exogenous Origins of Crises, *Extreme events in nature and society*. (95/119). Springer Berlin Heidelberg (2006)
- 3) M. E. J. Newman: Power laws, Pareto distributions and Zipf's law. *Contemporary Physics* (2005)
- 4) Mizuno, T., Ohnishi, T., & Watanabe, T.: Novel and topical business news and their impact on stock market activities. *arXiv preprint arXiv:1507.06477* (2015)
- 5) Hisano, R., Sornette, D., Mizuno, T., Ohnishi, T., & Watanabe, T.: High quality topic extraction from business news explains abnormal financial market volatility. *PLoS ONE* 8(6): e64846 (2013)
- 6) Alanyali, M., Moat, H. S., & Preis, T.: Quantifying the relationship between financial news and the stock market. *Scientific reports*, 3, 3578 (2013)
- 7) Mizuno, T., Takei, K., Ohnishi, T., & Watanabe, T.: Temporal and cross correlations in business news. *Progress of Theoretical Physics Supplement*, 194, 181/192 (2012)
- 8) Petersen, A. M., Wang, F., Havlin, S., & Stanley, H. E.: Market dynamics immediately before and after financial shocks: Quantifying the Omori, productivity, and Bath laws. *Physical Review E*, 82(3), 036114 (2010)
- 9) Rangel, J. G.: Macroeconomic news, announcements, and stock market jump intensity dynamics. *Journal of Banking & Finance*, 35(5), 1263/1276 (2011)
- 10) Bollen, J., Mao, H., & Zeng, X.: Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1/8 (2011)
- 11) Schumaker, R. P., & Chen, H.: Textual analysis of stock market prediction using breaking financial news: The AZFin text system. *ACM Transactions on Information Systems (TOIS)*, 27(2), 12 (2009)
- 12) Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas: Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61(12), 2544/2558 (2010)

- 13) Tversky, A., & Kahneman, D.: Loss aversion in risk-less choice: A reference-dependent model. *The quarterly journal of economics*, 106(4), 1039/1061 (1991).
- 14) Preis, T., Moat, H. S., & Stanley, H. E.: Quantifying trading behavior in financial markets using Google Trends. *Scientific reports*, 3, 1684 (2013)
- 15) Moat, H. S., Curme, C., Avakian, A., Kenett, D. Y., Stanley, H. E., & Preis, T.: Quantifying Wikipedia usage patterns before stock market moves. *Scientific reports*, 3, 1801 (2013)
- 16) Curme, C., Preis, T., Stanley, H. E., & Moat, H. S.: Quantifying the semantics of search behavior before stock market moves. *Proceedings of the National Academy of Sciences*, 111(32), 11600/11605 (2014)
- 17) Thomson Reuters.: Reuters Newsscope Sentiment Engine: Guide to sample data and system overview. *Thomson Reuters White Paper* (2008)
- 18) Thomson Reuters.: Reuters Newsscope Sentiment Engine: Output image and file format. *Thomson Reuters White Paper* (2008)