

銘柄固有の金融極性辞書の構築

Constructing Financial Sentiment Lexicons for Individual Stocks

関 和広^{1, 2*} 柴本昌彦²
Kazuhiro Seki^{1,2} Masahiko Shibamoto²

¹ 甲南大学 知能情報学部

Faculty of Intelligence and Informatics, Konan University

² 神戸大学 経済経営研究所

Research Institute for Economics & Business Administration, Kobe University

Abstract: This paper reports on our ongoing work to construct sentiment lexicons in the financial domain. Our approach takes advantages of news headlines and a given financial variable, such as stock prices, so as to generate initial sentiment lexicons. The initial lexicons are then filtered based on their co-occurrences with financial seed words and are subsequently expanded by analogical reasoning by using distributed representation of words. Evaluative experiments on around 12 years' worth of news data show that the resulting lexicons are mostly reasonable. As a possible application of the lexicons, trading simulation is also carried out, showing promising results.

1 まえがき

近年、マイクロブログやユーザレビュー、質問応答型の様々なフォーラムなど、誰もが即座に利用可能な Consumer Generated Media (CGM) のプラットフォームが急速に普及し、世界中で利用されている。そして、これらのプラットフォームに投稿される情報は、商品の購入や訪れる店舗の選択といったユーザの多様な意思決定に日常的に影響を及ぼしている。このような背景から、WWW 上に溢れるテキスト情報を有効に活用するためのテキストマイニングの研究が盛んに行われている [Imran 15].

テキストマイニングの技術の一つとして、分析対象のテキストに表現される書き手の感情を抽出、あるいはスコア化する「感情分析」がある [Liu 15]. そして、感情分析を行う際に有用な言語資源として、極性辞書や感情表現辞書がある。極性辞書は肯定・否定という二極分化的な観点から構築され、感情表現辞書は、喜びや悲しみや怒りといった複数の感情の種類を区別して構築される。これらの辞書では、肯定や否定、あるいは喜びや悲しみといった感情の強弱を考慮して、辞書に収録される各単語に感情を点数化して付与していることが多い [Takamura 05].

一方、金融・経済の分野でもテキストマイニングの研

究が長年行われてきている。たとえば、ニュース記事を分析して株価などの金融指標を予測する研究がある [Schumaker 09]. これらの研究の手法は2つのアプローチに大別することができる。1つ目のアプローチでは、ニュース記事等のテキストと目的変数（金融指標）の関係を過去のデータから Support Vector Regression などの何らかの回帰モデルで学習し、予測を行う。もう一方のアプローチでは、予測対象とする金融指標に関してテキストが肯定的な内容を含んでいるか、否定的な内容を含んでいるかを分析し、前者であれば金融指標が上昇、後者であれば下降するものと予測する。このアプローチは前述の極性辞書を用いた感情分析と類似している。

本研究では、金融テキストマイニングにおける第一歩として、金融指標の短期的な変動に関わる語を収録した金融極性辞書の構築を目指す。なお、極性辞書に含むべき語は、個別銘柄の株価など、対象とする金融指標によって異なる可能性があるため、対象とする個々の指標に応じた個別の辞書を構築する。辞書の構築は、過去のニュース記事と金融指標の変動との関係を利用して、自動的に行う。さらに、構築した辞書を用いて金融指標の短期的な変動を予測し、その結果に基づいて行った投資シミュレーションの結果についても報告する。

本論文の構成は以下の通りである。まず、2節で感情分析や株価予測の関連研究を紹介する。続いて、3節

*連絡先：甲南大学知能情報学部知能情報学科
〒 658-8501 兵庫県神戸市東灘区岡本 8-9-1
E-mail: seki@konan-u.ac.jp

で提案手法の詳細を述べ、4節で実データを用いて行った評価実験について報告する。最後に、5節でまとめと今後の課題について述べる。

2 関連研究

2.1 極性辞書の自動構築

自動的に極性辞書を構築する方法は、辞書を用いる方法とコーパスを用いた方法に大別できる。辞書を用いた方法では、WordNet [Fellbaum 98] のようなシソーラスを用いて、極性が既知の語の関連語に極性を伝播させることによって極性辞書を構築する。なお、以降では極性が既知の語を「種表現 (seed word)」と呼ぶ。例えば、「important」が肯定的な種表現であるとき、シソーラスを参照することで、その同義語である「essential」「momentous」等は肯定的、反意語である「immaterial」「fiddling」「lilliputian」等は否定的な表現であるとみなす。実際には、シソーラスを語のネットワークとして見たときの語間の距離や、語の定義文 (gloss) の類似度を考慮して極性を伝播させることが多い [Baccianella 10, Kamps 04]。辞書を用いた方法の欠点としては、辞書あるいはシソーラスに含まれる語しか扱うことができないこと、ドメインを区別しない一般的な極性辞書しか構築できないことが挙げられる。

コーパスを用いた手法でも、辞書による方法と同様に、種表現を用いて極性が未知の語の極性を推定する。推定の手がかりとしては、大規模なコーパスに現れる語の言語的特徴や統計的特徴を利用する。前者の例としては、接続詞を用いる方法が提案されている [Hatzivassiloglou 97]。この方法は、「but」のような逆接の接続詞の前後の語は逆の極性を持つことが多く、「and」のような順接の接続詞の場合は同じ極性を持つことが多いことに着目する。たとえば、effective が肯定的な種表現であったとき、「effective but costly」という句からは costly が否定的な表現であること、「efficient and effective」からは efficient が肯定的な表現であることが推測できる。

コーパスを用いた統計的な手法としては、共起関係を用いた方法 [Turney 02] がよく利用される。この方法では、少数の種表現を用いて、極性が不明な語 w とそれら種表現との共起頻度を計数する。そして、共起頻度を基に肯定的な種表現と w の自己相互情報量、および否定的な種表現と w の自己相互情報量を算出し、その差を極性スコアと定義する。そして、このスコアの符号と絶対値の大きさをそれぞれ極性の方向 (肯定対否定) と強さとする。

また最近では、語の分散表現 [Mikolov 13a] を用いた手法がいくつか提案されている。分散表現とは、ニュー

ラルネットワーク言語モデルを構築する際に副次的に獲得される語のベクトル表現であり、単語類推問題¹などに応用できる。このような分散表現を利用し、たとえば佐藤ら [佐藤 16] は、分散表現を素性として語の極性を出力する分類器を学習することを提案している。ただしこの手法では、極性が既知な語が学習データとして相当数必要となる。

2.2 金融ドメイン極性辞書

極性辞書に収録されるべき単語やその極性は、対象ドメインや記述の対象によって異なる [Bross 13]。たとえば、ホテルのレビューを対象に感情分析を行うのであれば、(ベッドや風呂などが)「小さい」は否定表現であるのに対して、ノートパソコンなどの電化製品に関するレビューでは同じ語が肯定表現になりうる。本研究では、金融ドメインを対象とするため、一般ドメイン向けの極性辞書では適切な感情分析は難しい [Loughran 11]。

金融ドメインについては、Loughran and McDonald [Loughran 11] によって人手で構築された英語の極性辞書がいくつかの研究で利用されている [Jegadeesh 13, Lee 14, Wang 13]。極性辞書の自動構築については、日本語の文章を対象とし、Hatzivassiloglou ら [Hatzivassiloglou 97] と類似の言語的特徴を利用した手法が試みられている [前川 13]。また、前述のような分散表現を用いた手法もいくつか試みられている [片倉 15, 伊藤 16]。

以上で述べた関連研究と比較し、本研究で提案する手法は、種表現がほとんど必要なく人手で収集する必要もない点、テキストデータに加えて金融指標の時系列データも用いている点、個別銘柄の株価など金融指標固有の極性辞書を構築できる点に特長がある。

3 金融極性辞書の構築

3.1 概要

本研究では、以下の手順で、対象とする金融指標固有の極性辞書を構築する。

1. 金融指標の変動を利用した極性表現候補の抽出
2. 極性表現候補のフィルタリング
3. 反意語の獲得

以下、個々の処理について詳述する。

¹ 「フランスに対するパリ」の関係から、「日本に対する x 」の x を推定するような問題 (正解は東京)。

3.2 金融指標の変動を利用した極性表現候補の抽出

ある時間 t にニュース記事 d が配信され、その後 $t+\Delta t$ に金融変数 y が $\theta_y\%$ 以上変動したとき、ニュース記事 d はこの変動に影響を与えたと仮定する。もちろん、ニュースでは報道されない他の様々な要因も考えられるため、この仮定は必ずしも成り立たない。しかし、金融指標の変動に影響を与えた記事を正確に同定することは困難であるため、本研究ではこの単純な仮定を用いて議論を進める。

上述の仮定において、時間窓の幅 Δt と金融指標の変動の幅 θ_y を適当な値に設定すれば、目的変数 y の正方向への変動の要因となった記事集合 D^+ 、および負方向への変動の要因となった記事集合 D^- を自動的に同定することができる。これらの記事集合 $D = D^+ \cup D^-$ を用いて、 y の変動の要因となる記事に特徴的な語を見つける。

具体的には、記事に現れる各語 w について式 (1) のカイ二乗値を算出し、この値が高いほど特徴的な語であると判断する。

$$\chi^2(w) = \frac{|D| \times (n_w^+ n_w^- - n_w^+ n_w^-)^2}{(n_w^+ + n_w^-)(n_w^+ + n_w^-)(n_w^+ + n_w^-)(n_w^+ + n_w^-)} \quad (1)$$

ここで、 $|D|$ は総記事数を表し、 n_w^+ は n^+ のうち w を含む記事数、 n_w^- は n^- のうち w を含む記事数、 n_w^+ は n^+ のうち w を含まない記事数、 n_w^- は n^- で w を含まない記事数とする。

なお、この指標は D^+ と D^- のいずれかに特徴的な語であれば大きい値を取る。よって、カイ二乗値だけではどちらに特徴的な語か判断できないため、 D^+ と D^- における語 w の相対頻度を比較して、どちらに特徴的な語であるか判定する。そして、 D^- に特徴的な語である場合は、便宜上、カイ二乗値にマイナスの符号を付して区別する。このようにして算出されたカイ二乗値の絶対値が閾値 θ_χ より大きい語を「極性表現候補」と呼ぶ。

3.3 極性表現候補のフィルタリング

前節で得られた極性表現候補には、目的変数 y 、あるいは金融指標一般の変動とは関わりがないと思われるような語も含まれてしまう。このような不適切な語を除去するため、コーパスを利用した極性辞書の構築でよく用いられる指標 [Turney 02] を用いる。具体的には、式 (2) に示す自己相互情報量 I の差 δ を算出する。

$$\begin{aligned} \delta(w) &= I(w, seed^+) - I(w, seed^-) \\ &= \log_2 \frac{h(w, seed^+)h(seed^-)}{h(w, seed^-)h(seed^+)} \quad (2) \end{aligned}$$

ここで $h(\cdot)$ は、引数 (語) の出現文書数を表す。 $\delta(w)$ の絶対値が大きいほど、種表現 $seed^+$ あるいは $seed^-$ と語 w の共起がいずれかに偏っていることを意味し、従来研究ではこの値が極性値として使われている。本研究では、 $\delta(w)$ の絶対値がある閾値 θ_δ 以上の場合のみ、語 w は金融指標の変動と関連した文脈で使われることが多いと判断し、極性表現と見なす。

3.4 反意語の獲得

前節で得られた極性表現は、抽出・フィルタリングの過程を経て獲得されるため、あまり多くの表現は得られない。そこで、得られた表現を基にこれを拡張する。表現の拡張には、語の統語的・意味的特徴を表現できる分散表現 [Mikolov 13b] を用いる。具体的には、単語類推問題と同様の方法で、前節で獲得した極性表現 w の反意語 w' の候補を類推する。すなわち、「 $seed^+$ に対する $seed^-$ 」あるいは「 $seed^-$ に対する $seed^+$ 」は「 w に対する w' 」であると考え、それぞれの語に対応する分散表現の加減算によって得られたベクトルと類似した分散表現を持つ語 (本研究では上位 20 語) を w' の候補と考える。ベクトル間の類似度としては、コサイン類似度を用いる。

さらに、この過程で同意語が反意語として獲得されてしまうことを避けるため、3.3 節で利用した自己相互情報量の差 (δ) の符号の変化に着目する。具体的には、語 w についての $\delta(w)$ と、類推に基づいて得られた語 w' についての $\delta(w')$ が異符号のとき反意語、同符号のときには同意語と判断する。なお、3.3 節と同様に $|\delta(w')| \geq \theta_\delta$ が成り立つ w' のみを獲得の対象とする。

4 評価実験

3 節で述べた提案手法の有効性を検証するため、実データを用いて評価実験を行った。評価実験は、得られた極性辞書の定性的評価、および極性辞書を用いた株価動向予測による投資シミュレーションからなる。以下、4.1 節で実験設定について述べ、4.2 節で極性辞書の評価、4.3 節で投資シミュレーションの結果について報告する。

4.1 実験設定

4.1.1 データ

評価実験に用いたデータは、Thomson Reuters News (TRN) アーカイブの 2003 年 1 月 1 日～2014 年 6 月 30 日分 (11 年 6 ヶ月分) の日本語ニュース記事と、その期間に対応する株の売買を記録した歩み値 (ティッ

配信日時	2013-01-04 00:28:22
見出し	ホットストック：自動車株が軒並み高、円安加速で収益改善期待強まる
本文	[東京 4日 ロイター] 自動車株が軒並み高となっている。ドル／円が87円台後半まで上昇するなど...
関連銘柄	7203.T 7261.T 7267.T 7270.T
...	...

図 1: Thomson Reuters News アーカイブ収録の記事の一部 (記事 ID: nTK0584247)。

ク) データである。TRN のニュース記事には、記事配信の日時、記事見出し、記事本文、記事に関連する銘柄のコードなどが記載されている (図 1)。評価実験では、銘柄コードの情報に基づいて特定の企業に関するニュース記事を抽出し、その企業の株価の動きと対応を取ることで株価の変動に影響を与えるニュース記事 ($D^+ \cup D^-$) を同定した。

本節の実験では、特に他の記載がない場合、TRN の 2003~2012 年 (10 年分) の記事の見出しを極性表現候補の抽出に利用し、残り 1.5 年分のデータを投資シミュレーションに用いた。TRN の記事見出しは形態素解析器 MeCab²で解析し、(複合) 名詞と動詞だけを極性表現候補として抽出した。なお、株価増加に影響を与えた記事集合 D^+ 、減少に影響を与えた記事集合 D^- を決定するための時間窓の幅 Δt は 30 分、株価変動の幅 θ_y は $\pm 1\%$ とした。また、カイ二乗値による極性表現候補の抽出時の閾値 θ_x は 2.706 (有意水準 10%)、自己相互情報量の差に関する閾値 θ_δ は 3.0 とした。seed⁺ には「売り」、seed⁻ には「買い」をそれぞれ用いた。なお、ニューラルネットワーク言語モデル CBOW [Mikolov 13a] の構築には gensim³ を利用し、TRN の 2003~2012 年の 10 年分の記事の見出しと記事本文を入力とした。パラメタ値については、語の最低出現頻度を 10 とした他はデフォルトの設定を用いた。

4.1.2 投資戦略

自動的に構築された金融極性辞書の応用例の一つとして、単純な投資シミュレーションを行った。ニュース記事 d が配信された際、その記事の関連銘柄 $code$ についての投資の判断は、次式の極性平均値に基づいて行った。

$$s_{avg}(code, d) = \frac{1}{|d|} \sum_{w \in d} s(code, w) \quad (3)$$

ここで $s(code, w)$ は $code$ 固有の辞書を基に w の極性値を返す関数であり、肯定表現であれば +1、否定表現で

²<http://taku910.github.io/mecab/>

³<https://radimrehurek.com/gensim/>

あれば -1 とした。code の辞書に w が含まれていない場合は、全銘柄データから作成した辞書を基に極性値を返すものとした。| d | は記事 d の長さ (語数) である。

そして、 $s_{avg}(code, d)$ の絶対値が閾値 θ_s (本実験では 0.1) を超えたとき、この値が正であれば買い、負であれば売りと判断した。売買単位はいずれも 1 株とした。続いて、株価があらかじめ設定した比率 (本実験では 1%) だけ変動した際に反対売買を行い、利益確定あるいは損切りした。

4.2 極性辞書の構築

3 節で述べた方法で、全銘柄のニュース記事を対象に構築した銘柄共通の極性辞書を表 1 に示す。なお、時間の都合上、本節の実験では、TRN の 2013 年の記事だけを用いた。結果を見ると、「下方修正」「売り」「公募増資」「赤字」など、株価の低下につながるような語が負のカイ二乗値を持ち、否定的な語として獲得されている。一方、「上方修正」「好感」「自社株買い」など、株価の上昇につながるような語は正の値を示しており、肯定的な語として獲得されている。この結果から、ニュース記事の変動とその後の実際の株価との間には、直感に合致するような関連があることが見て取れる。

さらに、分散表現を用いた関連語の同定と相互情報量の差の符号変化に基づく極性判定により、「下方修正」に対する「上方修正」、「売り」に対する「買い」、好感に対する「嫌気」などの語が獲得されていることが確認できる。また、同意語としても「売り」に対する「売り越し」、「買い」に対する「買われる」などが獲得されている。また、同意語ではないものの、「株式分割」から「自社株買い」や「消却」などのいずれも肯定的な語が獲得できていることが分かる。一方で、「売り」の同意語として「買い越し」など、本来は極性が反対である語を獲得してしまっている場合もあり、さらなる改善が必要である。

表 1 の辞書は、東証一部上場のすべての銘柄に関するニュース記事と株価を利用し、銘柄を区別せずに構築している。しかしながら、同じ表現であっても、業種や個別の企業によってそれぞれの株価への影響は異なると考えられる。たとえば、為替相場が業績にどのように影響するかは、輸出産業か輸入産業かなどで異なる。前者の場合、円高が進めば海外での売り上げが日本円では目減りするため、減益につながる。すなわち、同じ「円高」という語でも、株価の上昇につながる肯定的な語と捉えられるか、否定的な語として捉えられるかは注目する銘柄によって異なる。

本研究の手法では、入力に用いる金融指標のデータを変更すれば、その金融変数に応じた極性辞書が自動的に構築される。そこで、提案手法によって銘柄ごと

表 1: 共通金融極性辞書構築の結果. 否定的な表現のカイ二乗値には便宜的に負号を付している (3.3 節参照).

極性表現 w	χ^2	δ	極性表現 w'	δ
下方修正	-89.55	-3.76	上方修正	3.81
売り	-39.16	-4.38	買い	4.21
			買い越し	-3.22
			換金	-3.04
			売り越し	-5.07
			売りで	-3.09
			押され	-4.20
公募増資	-12.20	-3.48	自社株買い	3.33
			株式分割	4.81
			消却	3.14
赤字	-9.20	-4.08	黒字	3.14
自社株	16.18	4.29	公募増資	-3.48
			自社株買い	3.33
			株式分割	4.81
			消却	3.14
上方修正	30.33	3.81	下方修正	-3.76
好感	75.64	5.16	嫌気	-3.56
自社株買い	148.77	3.33	公募増資	-3.48
			株式分割	4.81
株式分割	148.86	4.81	公募増資	-3.48
			消却	3.14
			自社株買い	3.33
材料視	223.20	3.56	嫌気	-3.56
買い	800.59	4.21	売り	-4.38
			売りで	-3.09
			売られ	-3.35
			買われる	3.86
			買われ	3.86

に極性辞書を構築し、実際に「円安」あるいは「円高」が逆の極性で獲得された銘柄を抽出して比較した。表 2 に結果を示す。

表 2: 銘柄による円安・円高の極性の違い.

円安が肯定的	円高が肯定的
キャノン ジェイエフイーホールディングス 安川電機 新日鐵住金 川崎重工業 シャープ 富士通	本田技研工業 (ホンダ) パナソニック

いずれの企業も、一般的に円安メリットを享受すると思われる企業である。しかしながら、ホンダは他の自動車メーカーと比較して海外生産の比率が非常に高く、円安効果がほとんどないと言われている。また、パナソニックは国内での売上比率が高く、大手電機メーカーの中では円高の影響を受けにくい。これらの点から、ホンダ、パナソニックでは円高のニュースの配信後に株価が相対的に上昇した可能性がある。

なお、銘柄固有の極性表現として、トヨタでは「米自動車販売」、ソフトバンクでは「アジア株」、シャープでは「新興国」、東京電力では「見送り」「不透明感」「処分」などの語が獲得されていた。

4.3 投資シミュレーション

4.1.2 節で述べた戦略に従い、投資シミュレーションを行った。売買が行われた銘柄の異なり数は 1,066 であり、取引総数は $2,367 \times 2$ 回であった。利益の増減の結果を図 2 の「ind. & common」のグラフに示す。投資期間の前半で利益が大きく伸び、最終的には約 1.7 百万円の利益が得られた。なお、「individual」で示したもう一方のグラフは、(全銘柄共通の辞書を用いず) 銘柄固有の辞書のみを利用した場合の結果である。取引回数が減少することもあり、前者ほどの利益の伸びは見られないものの、最終的には約 10 万円程度の利益が得られた。投資シミュレーションの期間が 18 ヶ月と比較的短期であり、この実験の結果を一般化することはできないものの、本研究で自動構築された金融極性辞書が有用である可能性が示された。

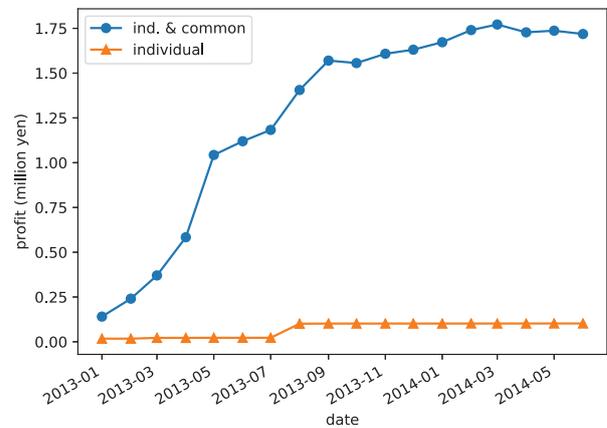


図 2: 投資シミュレーションの結果.

参考まで、投資シミュレーションで特に利益幅が大きかった 5 銘柄と小さかった 5 銘柄の個々の結果を表 3 にまとめる。

表 3: 利益幅上位・下位 5 銘柄の投資シミュレーションの結果.

銘柄	取引回数	最終利益 (千円)
プロロジ	8	142
SBI ライフ	2	141
産業ファンド	6	132
ワコム	4	105
日本高純度化	4	84
星野	2	-68
アコモ F	4	-68
ナノキャリア	6	-71
NVC	2	-74
ヴィレッジ V	2	-80

5 むすび

本研究では、ニュース速報コーパスと任意の金融指標を用いて金融極性辞書を構築する枠組みについて報告した。この枠組は、既存のシソーラスや辞書を必要とせず、与えた金融指標に応じた辞書を自動的に構築できる点に特長がある。実際に構築された辞書を確認したところ、「下方修正」「赤字」「嫌気」「公募増資」などが否定的、「上方修正」「黒字」「好感」「自社株買い」などの語が肯定的な語として獲得されており、おおむね直感に合う結果が得られていた。また、企業ごとの辞書を比較したところ、たとえば「円安」という語は企業によって極性が異なっていることが確認できた。また、極性辞書の応用例の一つとして、2013年から1.5年のデータを用いて投資シミュレーションを行ったところ、単純な方法ながらある程度の収益が得られることが示された。

一方、実際には単語（あるいは複合語）単位で極性を判断することはできない場合も多いため、節や述語項関係などより大きい単位での極性辞書を構築する必要がある。また、同一の表現であってもその極性は時間とともに変化すると考えられるため、固定的な極性ではなく、今後、時系列的な極性の変化についても検討していく。

謝辞

本研究の一部は、私立大学等経常費補助金特別補助「大学間連携等による共同研究」、およびJSPS科研費15H05729の助成によるものである。

参考文献

- [Baccianella 10] Baccianella, S., Esuli, A., and Sebastiani, F.: SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining, in *Proceedings of the International Conference on Language Resources and Evaluation 2010*, pp. 2200–2204 (2010)
- [Bross 13] Bross, J. and Ehrig, H.: Automatic Construction of Domain and Aspect Specific Sentiment Lexicons for Customer Review Mining, in *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management*, pp. 1077–1086 (2013)
- [Fellbaum 98] Fellbaum, C. D.: *WordNet: an electronic lexical database*, MIT Press (1998)
- [Hatzivassiloglou 97] Hatzivassiloglou, V. and McKeown, K. R.: Predicting the Semantic Orientation of Adjectives, in *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*, pp. 174–181 (1997)
- [Imran 15] Imran, M., Castillo, C., Diaz, F., and Vieweg, S.: Processing Social Media Messages in Mass Emergency: A Survey, *ACM Computing Surveys*, Vol. 47, No. 4, pp. 67:1–67:38 (2015)
- [Jegadeesh 13] Jegadeesh, N. and Wu, D.: Word power: A new approach for content analysis, *Journal of Financial Economics*, Vol. 110, No. 3, pp. 712–729 (2013)
- [Kamps 04] Kamps, J., Marx, M., Mokken, R. J., and Rijke, de M.: Using WordNet to Measure Semantic Orientations of Adjectives, in *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)*, pp. 1115–1118 (2004)
- [Lee 14] Lee, H., Surdeanu, M., MacCartney, B., and Jurafsky, D.: On the Importance of Text Analysis for Stock Price Prediction, in *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*, pp. 1170–1175 (2014)
- [Liu 15] Liu, B.: *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*, Cambridge University Press, Cambridge, United Kingdom (2015)
- [Loughran 11] Loughran, T. and McDonald, B.: When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks, *Journal of Finance*, Vol. 66, No. 1, pp. 35–65 (2011)
- [Mikolov 13a] Mikolov, T., Chen, K., Corrado, G., and Dean, J.: Efficient estimation of word representations in vector space, *arXiv preprint arXiv:1301.3781* (2013)
- [Mikolov 13b] Mikolov, T., Yih, W., and Zweig, G.: Linguistic Regularities in Continuous Space Word Representations, in *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT-2013)*, pp. 746–751 (2013)

- [Schumaker 09] Schumaker, R. P. and Chen, H.: Textual Analysis of Stock Market Prediction Using Breaking Financial News: The AZFin Text System, *ACM Transactions on Information Systems*, Vol. 27, No. 2, pp. 12:1–12:19 (2009)
- [Takamura 05] Takamura, H., Inui, T., and Okumura, M.: Extracting semantic orientations of words using spin model, in *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL2005)*, pp. 133–140 (2005)
- [Turney 02] Turney, P. D.: Thumbs Up or Thumbs Down?: Semantic Orientation Applied to Unsupervised Classification of Reviews, in *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pp. 417–424 (2002)
- [Wang 13] Wang, C.-J., Tsai, M.-F., Liu, T., and Chang, C.-T.: Financial Sentiment Analysis for Risk Prediction, in *Proceedings of the 6th International Joint Conference on Natural Language Processing*, pp. 802–808 (2013)
- [伊藤 16] 伊藤 友貴, 坪内 孝太, 山下 達雄, 和泉 潔 : ニュース記事を用いた経済専門用語のクラスタリングと極性付与, 第 30 回人工知能学会全国大会論文集 (2016)
- [佐藤 16] 佐藤 貴俊, 高村 大也, 奥村 学 : 分散表現を用いた単語の感情極性抽出, 情報処理学会研究報告自然言語処理 (NL) , Vol. 2016-NL-228, No. 12, pp. 1–6 (2016)
- [前川 13] 前川 浩基, 中原 孝信, 岡田 克彦, 羽室 行信 : 大規模ニュース記事からの極性付き評価表現の抽出と株価収益率の予測, オペレーションズ・リサーチ: 経営の科学, Vol. 58, No. 5, pp. 281–288 (2013)
- [片倉 15] 片倉 賢治, 高橋 大志 : 金融市場ニュースの分散表現学習による辞書作成と金融市場分析, 第 29 回人工知能学会全国大会論文集 (2015)