

フィルタリングと機械学習に基づく Web ニュースからの日経平均株価予測

一瀬 航[†] 嶋田 和孝[†]

[†]九州工業大学大学院 情報工府 先端情報工学専攻

〒 820-8502 福岡県飯塚市川津 680-4

E-mail: †{k_ichinose,shimada}@pluto.ai.kyutech.ac.jp

あらまし 近年、機械学習を用いたテキストマイニング手法によって、テキスト情報と市場変動の関係性を発見し、市場分析に応用する研究が増えている。また、Web ニュースは企業の株価に少なからず影響を与えており、世に存在する個人投資家がこのニュース記事を参考にしていていると考え、Web ニュースから未来の株価が予測できる可能性がある。そこで本論文では、Web ニュースを対象とし、より多くの投資家が市場の分析に用いていると考えられる指標である日経平均株価の予測を目的とする。テキストを用いた金融予測では膨大なテキスト情報を用いて機械学習を行うことが一般的である。しかし、投資家は市場に影響を与える多様な情報を自ら取捨選択し、市場の分析を行っている。本研究では、この取捨選択という作業をフィルタを導入することで再現し、これにより機械学習の精度が向上するかの検証を行う。

キーワード Web ニュース, 株価予測, フィルタリング, SVM

Stock market prediction from Web news using filtering and machine learning

Ko ICHINOSE[†] and Kazutaka SHIMADA[†]

[†] Kyushu Institute of Technology, Graduate School of Computer Science and Systems Engineering Kawadu 680-4, Iidukasi, Fukuoka, 820-8502 Japan

E-mail: †{k_ichinose,shimada}@pluto.ai.kyutech.ac.jp

Abstract The market analysis is one of the important tasks for text mining. Many researchers have proposed methods using text information for analyzing the market. In this situation, Web news have the important role to predict stock prices. In this paper, we propose a method to predict the Nikkei Stock Average, which is one of the most important stock market indexes. We apply a filtering process and a machine learning into the method. First, we select important news from many Web news by using the filtering. Then, we classify the selected news into “UP” and “Down” by using a machine learning method. Finally, we predict the change of the next day’s Nikkei stock price. The experimental result shows the effectiveness of the filtering process for the Web news based classification task.

Key words Web-news, Stock price prediction, filtering, SVM

1. はじめに

近年、ネット証券の誕生によって初心者でも株の取引が簡単にできるようになり、株式投資が活発になっている。活発な投資活動は経済成長に必要な成長資金の供給を助け、経済の効率化も望める。しかし、1年以内で新規の個人投資家の90%以上が退場するといわれるほど株式投資は初心者にとっては厳しい世界である。経済成長のために長期安定的な投資の定着を図るためにも、株式投資初心者が大きな損失を出すことを防ぎ、投

資の継続を支援することは社会的にも大きな意義がある。

投資家が市場の分析に用いる情報には、経済指標やテクニカル指標と呼ばれる数値で表される情報と、市場に対して影響を持つ人物の発言や企業の動向、事件や事故についての記事といったテキスト情報の2種類がある。前者の情報の分析に関しては書籍やWebサイトなどでも数多く取り扱われているが、後者の情報の分析に関しては個人の経験や感性に拠るところが大きいため分析方法を解説したものはほとんどなく、投資初心者が独学で勉強するのは容易ではない。そのため、機械でテキス

ト情報を用いた市場の分析ができれば初心者の支援になると考えられる。

テキストマイニング技術を応用して、テキスト情報と市場変動の関係性を発見し、市場分析に応用する研究がある [1] [2]。そのような金融テキストマイニングの対象となるテキスト情報源には Twitter, 新聞, インターネット掲示板, Web ニュースなど様々なものがある。この中でも Web ニュースは配信頻度が高く、即時性に優れている。辻ら [3] の研究では Web ニュース中に出てくる複数の企業に対応した株価予測をそのニュースのテキスト情報を用いて行っている。この研究のようにテキスト情報を用いて株価を予測する研究は個別企業に焦点を当てたものが多い。個別企業の株価の予測は、当然ながらその企業の株に注目している投資家にしか恩恵はない。また、限られた Web ニュースだけで全ての企業の株価を個別に予測することは現実的ではない。

そこで本論文では、より多くの企業の株価に影響を与えらると思われる日経平均株価の予測を目的とする。日経平均株価は、東証一部に上場している約 1700 社の中から選ばれた、日本を代表する 225 社の株価の平均である。この指標は、日本の景気を判断する上での重要な指標と考えられており、投資家がよく市場の傾向の分析に用いる指標でもある。この指標を正確に予測することは投資判断に役立つと考えられる。

テキストを用いた金融予測では膨大なテキスト情報を用いて機械学習を行うことが一般的であるが、Web 上に溢れる多様なテキスト情報を人が全て処理するのは不可能であり、投資家は市場に影響を与える多様な情報を自ら取捨選択し、市場の分析を行っている。この取捨選択とはつまり、重要なのは市場に影響がありそうな情報だけであり、膨大なテキスト情報の大部分が不要な情報であると捉えることができる。本研究では、この取捨選択という作業をフィルタを導入することで再現し、これにより機械学習の精度が向上するかの検証を行う。また、本手法による日経平均株価の予測は Web ニュースを扱うという特性のため、大きく 2 つの処理が必要となる。

- 1 記事単位の分類
 - Web から獲得されたある記事から翌日の株価を予測するための処理
- 1 日単位の分類
 - 1 日に得られた複数の分類結果からある日の株価を予測するための統合処理

本論文では、1 記事単位の分類と 1 日単位の分類について、それぞれ 4. 節と 5. 節で説明する。

2. 関連研究

テキスト情報を用いた市場分析に関する研究として前川ら [4] の研究がある。前川らは、大量の記事データから投資家が反応している「言葉」を探索することで、そうした言葉から構成される投資モデルを構築し、言葉の中に将来の株価の予測可能性を持つかどうかの検証を行った。Lee ら [5] は、企業が提出する金融報告書に注目し、報告された金融イベントに応じて企業株価の変化を予測するシステムを開発した。これにより、株価

予測にテキストを利用することでテキストを用いない場合よりも相対的に 10% も精度が上がることを示した。他にも、インターネット掲示板に投稿されたメッセージの分析に基づき、そこから得られる掲示板指標と株式指標の関係性を明らかにしようという丸山ら [2] の研究や、経済の専門家や金融機関が Web 上に発行するマーケットリポートから日本国債市場の月次の動向を推定する手法を開発した和泉ら [6] の研究がある。さらに、Twitter からツイート内容に含まれる単語をベースとしたテキストの特徴量とグラフ表現した際のグラフ特徴量の 2 つを抽出することで、Twitter 上の膨大な情報の中から、経済動向の分析に有益な情報を得ることで短期的な経済動向の分析を行った迫村ら [7] の研究がある。Twitter を用いたものは他にも Bollen ら [8] の研究がある。Bollen らはツイートにおける気分に注目し、6 つの心的状態を表す指数を抽出することでダウ平均株価の予測を行っている。

このように様々なテキスト情報を対象とした、金融テキストマイニングの研究が行われているが、Web ニュースを情報源とした日経平均株価の予測をメインに行っている研究はなかった。そこで、本研究では Web ニュースを情報源とした日経平均株価の予測を行う。

3. 対象データ

本研究では訓練事例となる実験データに Web ニュースと日経平均株価の終値を用いる。記事の獲得には MoneyTimes^(注1) という Web サイトを利用する。この Web サイトは複数のニュース配信サイトで配信される記事のリンクを「社会」、「経済」、「国際」、「株式」のように記事のカテゴリ毎に毎日まとめている。このサイトで 2014 年 7 月 14 日～11 月 19 日の間に配信されたものを対象に、以下の三つの条件を満たす記事を収集する。

- (1) 「株式カテゴリ」に分類されているもの
- (2) 株式市場が閉じた後に配信されたもの
- (3) 平日に配信されたもの

(1) の条件を付けた理由は「株式カテゴリ」の記事が市場に一番影響を与えようと考えたからであり、(2) の条件も市場が開いている間に配信された記事よりもその後に配信された記事のほうが次の日の日経平均に影響するよう考えたからである。また、(3) の条件は平日にしか市場は開かないために付けた。日経平均の終値のデータはインターネット上で公開されているもの^(注2)を用いる。

4. 1 記事単位の分類

4.1 提案手法

本研究では、フィルタリングした学習データを用いた機械学習による教師あり学習を行うことによって日経平均株価を予測する。教師値は、日経平均株価の終値が前日の終値と比べて上昇するか下落するかの 2 値とする。素性には、前日の各 Web ニュースに条件を満たす単語が出現したかどうかの 2 値をとる

(注1) : <http://mt.monegle.com/>

(注2) : <http://indexes.nikkei.co.jp/nkave/archives/data>

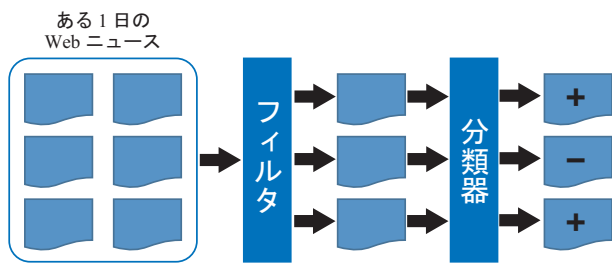


図1 1記事単位の分類手法

Bag-of-Wordsを用いる。データは3.節で説明したものをを用いる。1記事単位の分類手法を図1に示す。

4.1.1 フィルタの作成

まずは、学習データに用いる記事を選別するためのフィルタの作成方法について述べる。フィルタの作成は発見的手法によって行う。1.節で説明したように、このフィルタは投資家の取捨選択という作業を再現するものである。よって、フィルタは投資家が重要であると考えらるであろう記事を中心にして作成する。本研究では以下の3つのフィルタを作成する。

- (A) {日経平均}
- (B) {円相場, 為替, ドル}
- (C) {アメリカ株, 米国株, ダウ}

これらのフィルタは、フィルタを構成する単語がニュース記事のタイトルに1つでも含まれていればその記事を抽出する、というように使用する。

4.1.2 分類器の作成

本研究では日経平均株価の予測を機械学習によって行う。ここでは機械学習を用いた分類器の作成について説明する。分類器にはSVM [9]を、機械学習を行うツールにはWeka^(注3)を使用する。サポートベクターマシン (Support Vector Machine) は、1995年に、AT&TのVladimir Vapnikによって統計的学習理論の枠組みで提案された2クラスのパターン認識手法のことである。SVMでは、2種類のクラスのデータと、分離超平面との間の距離 (マージンと呼ぶ) が最大になるような分離超平面が、最も汎化能力の高い超平面になるということを利用して

3.節の実験データに4.1.1節で説明した3つのフィルタを適用し、学習データを作成するための記事の集合を3種類準備する。そして、得られた学習データ用の記事を用いて3つの学習データを作成する。この学習データを構成する特徴ベクトルは、同じ月日でも異なる記事間には依存関係はないとみなし1日に含まれる複数記事からその記事と同じ数だけ作成される。そして、作成された3種類の学習データを用いて機械学習を行い、分類器を作成する。

素性には2値のBag-of-Wordsを用いる。全てのWebニュースに対して形態素解析を行い、品詞が「名詞」、「動詞」、「形容詞」である単語から索引語の集合を作る。同じように、次は学習データ用の各記事に対しても形態素解析を行い、品詞が「名

表1 フィルタを利用した1記事単位の分類に関する評価結果

分類器		適合率	再現率	F 値
ベースライン	Minus	52.8	52.8	52.8
	Plus	61.8	61.7	61.7
	Ave	57.7	57.7	57.7
(A) を利用した分類器	Minus	64.1	64.1	64.1
	Plus	70.3	70.3	70.3
	Ave	67.5	67.5	67.5
(B) を利用した分類器	Minus	57.5	53.5	55.4
	Plus	60.0	63.8	61.9
	Ave	58.8	58.9	58.8
(C) を利用した分類器	Minus	61.1	47.8	53.7
	Plus	67.6	78.1	72.5
	Ave	64.9	65.5	64.6

詞」、「動詞」、「形容詞」である単語から検索語の集合を作る。そして、索引語の集合に登録されている全ての単語を一つずつ見ていき、検索語の集合にも登録されていれば1、されていないならば0としてベクトルを作成する。このように、索引語の集合の大きさが作成するベクトルの大きさとなり、検索語の集合を使ってそのベクトルの各値を1か0かに決定する。なお、形態素解析器にはMeCab^(注4)を用いた。

教師値にはその記事の次の日の日経平均株価の終値が上昇した (Plus) か下落した (Minus) かの2値を用いる。具体的には以下のようにする。

$$\text{教師値} = \begin{cases} \text{Plus} & (\text{次の日の終値} > \text{前日の終値}) \\ \text{Minus} & (\text{次の日の終値} \leq \text{前日の終値}) \end{cases}$$

この式を用いて教師値を決定する。

4.2 実験

本節では、フィルタを用いずに作成した学習データで機械学習を行った分類器をベースラインとして、提案手法で作成したフィルタに関する評価実験を行う。

4.1.2節で作成した3つの分類器とベースラインとして用いた分類器で3.節のデータを用いて評価実験を行った。評価には10分割交差検定を用いた。ベースラインの結果と3つのフィルタに関しての結果を表1に示す。評価項目はPlusクラスとMinusクラスの適合率と再現率とF値である。F値とは、適合率と再現率の調和平均である。また、表中のAveはPlusとMinusの加重平均である。

実験結果を表1に示す。表1のベースラインの評価項目と3つのフィルタを利用した分類器の評価項目を比べてみると、ベースラインのF値と比較して、(A)のフィルタを適用した分類器は9.8ポイント、(B)のフィルタを適用した分類器は1.1ポイント、(C)のフィルタを適用した分類器は6.9ポイント上回った。さらに、3つのフィルタは全ての項目でベースラインを上回っている。このようにフィルタを利用した分類器がフィルタを用いていない分類器の精度を上回ったことから、フィルタリングが有効であると考えられる。

(注3) : <http://www.cs.waikato.ac.nz/ml/weka/>

(注4) : <http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>

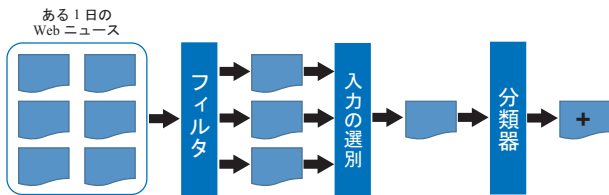


図2 入力を選別による1日単位の分類手法

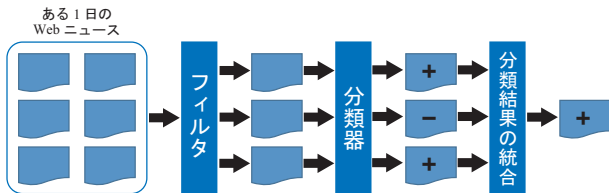


図3 分類結果の統合による1日単位の分類手法

5. 1日単位の分類

4.1.2節の分類器は、1記事単位の分類器である。つまり、1日で入力データとして利用できる記事が1個でない限り、この分類器は1日単位の分類結果ではなく複数の1記事単位の分類結果を出力する。そのため、最終的に1日単位の分類結果を得るための処理が必要となる。そこで、本節では1日単位の分類結果を得るための処理を解説する。

5.1 提案手法

1日単位の分類結果を得るための処理として2種類が考えられる。

- 入力を選別による1日単位の分類手法

- 以下の3種類の条件で選別した記事をそれぞれ入力とし、得られた分類結果を1日単位の分類結果とする

入力記事選別1: タイトルが一番長い記事

入力記事選別2: 入力記事全てを一つにして作成した記事

入力記事選別3: スコアリングによって抽出された1つの記事

- 分類結果の統合による1日単位の分類手法

- 単純な多数決を用いて、1日の入力データから得られた複数の分類結果のクラス数の多いほうを1日単位の分類結果とする

入力記事選別3に適用するスコアリング手法として、本論文では展望台システム[10]を利用する。展望台システムとは、キーワード抽出に基づく重要文抽出の1つであり、単語の出現頻度によるキーワードを手がかりとして、文章の主題となるキーワードを単語間の関連に基づいて抽出し、低頻度のキーワード中からも文章の主題に合致しかつ文章を特徴付けるキーワードを取り出す。そして、このキーワードの評価値をもとに文のスコアリングを行い、重要文を選定する。

5.2 実験

ここでは、実際に1日単位の分類に関する実験を前述の2種類の方法を用いて行う。まずは、評価方法を説明する。ここで、 n 日分のデータが存在すると仮定する。この実験の目的は1日単位での分類結果を得ることであるため、各日ごとにその日に属している記事に対して入力を選別もしくは分類結果の統合を

表2 入力を選別による1日単位の分類に関する評価結果

選別条件		適合率	再現率	F 値
入力記事選別1	Minus	42.5	47.2	49.1
	Plus	54.8	50.0	52.3
	Ave	49.4	48.8	49.1
入力記事選別2	Minus	40.0	33.3	36.4
	Plus	53.8	60.9	57.1
	Ave	47.8	48.8	48.3
入力記事選別3	Minus	47.2	47.2	47.2
	Plus	58.7	58.7	58.7
	Ave	53.7	53.7	53.7

表3 分類結果の統合による1日単位の分類に関する評価結果

分類器		適合率	再現率	F 値
ベースライン	Minus	30.0	16.7	21.4
	Plus	51.6	69.6	59.3
	Ave	42.1	46.3	44.1
(A) を利用した分類器	Minus	48.1	36.1	41.3
	Plus	58.2	69.6	63.4
	Ave	53.8	54.9	54.3
(B) を利用した分類器	Minus	40.9	34.6	37.5
	Plus	51.4	58.1	54.5
	Ave	46.6	47.4	47.0
(C) を利用した分類器	Minus	61.5	47.1	53.3
	Plus	69.0	80.0	74.0
	Ave	66.0	66.7	66.3

行う。すなわち、 n 日交差検定を行う。

5.2.1 入力を選別による1日単位の分類に関する評価実験

実際に1記事単位の精度が最も高かった(A)を利用した分類器を用いて、入力記事選別1、入力記事選別2、入力記事選別3の3種類の条件で選別された記事を入力として n 日交差検定により評価実験を行った。その結果を表2に示す。1記事単位の精度である表1の(A)の評価結果と表2を見比べてみると、入力記事を選別する前のF値と比べて、入力記事選別1は-18.4ポイント、入力記事選別2は-19.2ポイント、入力記事選別3は-13.8ポイントとなっている。このように、入力記事を選別して1つに絞った結果、1記事単位の精度よりも悪い精度となった。

5.2.2 複数の分類結果の統合に関する評価実験

実際に n 日交差検定を用いて多数決による分類結果の統合の結果を集計し、精度を評価した。その結果を表3に示す。1記事単位の精度である表1と表3を見比べてみると、分類結果を統合する前のF値と比べて、ベースラインは-10.7ポイント、(A)を適用した分類器は-13.2ポイント、(B)を適用した分類器は-11.8ポイント、(C)を適用した分類器は+1.7ポイントとなっている。このようにベースライン、(A)を利用した分類器、(B)を利用した分類器では、多数決による分類結果の統合を行ったために、1記事単位の精度よりも悪い精度が出ている。一方、(C)を利用した分類器を用いた分類結果の統合のみ精度が向上している。そこで(C)の結果について分析を行った。

まず最初に、(C)についてはフィルタを適用したことで1日

表 4 フィルタ別の欠損日数

フィルタ	欠損日
(A)	0 日
(B)	25 日
(C)	40 日

表 5 (C) を利用した分類器による分類結果の統合に関する評価結果

	適合率	再現率	F 値
1 記事のみ	67.0	67.7	67.4
2 記事以上	64.4	63.6	64.0

のうちに 1 個も記事が存在しないため予測ができない欠損日が全データ 82 日中で 40 日存在していたことがわかった。全データ 82 日中の各分類器の欠損日を表 4 に示す。フィルタによって除去された欠損日のデータは分類が失敗しやすい日であることを考えると、欠損日が半分近くもある (C) の精度が他の分類器の精度よりも上がるのは当然であり、表 3 の (C) の精度は必ずしも信頼できない。

さらに、多数決という統合法の観点から分析したところ、1 日に利用できる記事が 1 個だけの日が多く存在していた。(C) のフィルタを適用した後の 1 日に利用できる記事数が 1 個だけだった日は (C) の全データ 42 日中で 31 日であった。約 7 割が 1 日に 1 記事だったことから、(C) を利用した分類器の分類結果の統合の精度は統合前の 1 記事単位の精度の影響を大きく受けていると考えられる。そこで、1 日に 1 記事のみの日と、1 日に 2 記事以上ある日で分けて再評価を行った結果を表 5 に示す。1 記事のみの日はその記事の分類結果を 1 日単位の分類結果とし、2 記事以上ある日は多数決を用いた分類結果の統合により評価した。表 5 の結果では、1 日に 1 記事のみの日が 1 日に 2 記事以上ある日の F 値と比べて 3.4 ポイント高くなっている。また、1 日に 2 記事以上ある日の F 値 64.0 は表 1 の F 値 64.6 よりも低い値であるため、表 3 で (C) を利用した分類器のみ分類結果の統合の精度が上がったのは 1 日に 1 記事のみの日が精度を引き上げたからであると考えられる。

5.3 1 日単位の分類に関するエラー分析

まずは、入力の変別による 1 日単位の分類 (5.2.1 節) で精度が低かった原因についてエラー分析を行った。この入力記事の変別は、正しい予測値が得られるような特徴を持った記事を正しく選別できれば、分類精度もあがるのではないかとこの考えに基づいている。そこで、選別されるべき記事を人手で確認したところ、互いの記事が非常に似通っており人間でもどの記事を選択すべきか分からないような場合がほとんどであるということがわかった。具体的には、同じ事柄について異なる著者、出版元が記事を作成した程度の差異しか見られなかった。このことから、5.2.1 節で用いた入力記事選別 1、入力記事選別 2、入力記事選別 3 のような条件では正しい記事を選別することは難しいと考えられる。

次に、分類結果の統合による 1 日単位の分類 (5.2.2 節) で失敗した多数決についてのエラー分析を行った。まずは、多数決に用いた分類結果を可視化するために、1 日分の分類結果の

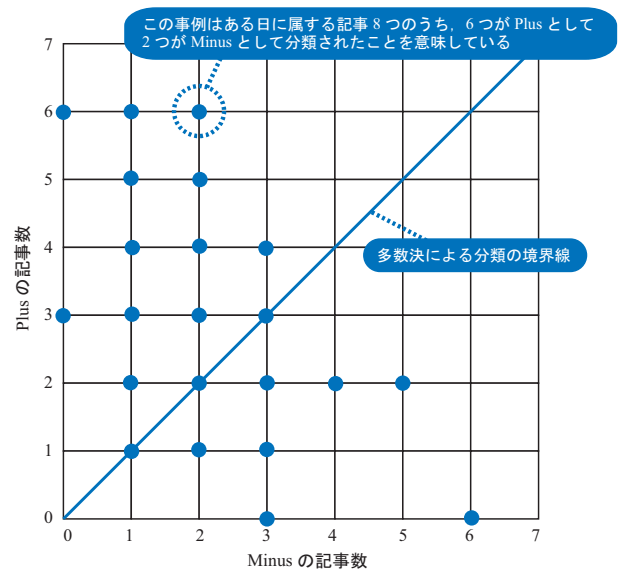


図 4 分類結果を可視化した散布図 (Plus クラス)

集合を 1 事例とし、その事例に含まれる記事の分類クラスの数を変数として散布図を作成した。散布図は表 1 で一番精度が高い (C) のフィルタを適用したデータで作成すべきだが、5.2.2 節で説明したように (C) のフィルタ適用後は約 7 割が 1 日 1 記事となっているため、多数決を分析するには適切でない。よって、表 1 で次に精度が高かった (A) のフィルタを適用したデータ (すなわち表 3 の (A) に関する結果) で正解クラス別に作成した散布図を図 4、図 5 に示す。図中の直線が多数決のための境界線を表している。すなわち、直線の上側に存在すれば Plus、下側に存在すれば Minus と分類されることを意味する。したがって、Plus の事例をマッピングした図 4 を考えれば、直線より左上に事例が多く存在することが望ましい。同様に図 5 では、右下に存在することが望ましい。しかし、両方の図からわかるように、マッピングされた事例は散布図上にまんべんなく分布しており、これは、多数決では正しく分類できないことを意味している。

6. おわりに

本研究では、Web ニュースを用いて日経平均株価の予測を行った。また、フィルタによって機械学習に用いるデータを選別することで予測精度が向上するかどうかの実験を行い、実際に精度が向上したことを確かめた。そして、1 記事単位の分類器から最終的に 1 日単位の分類結果を得るための処理として複数の入力記事を選別して入力を 1 つに絞る方法と複数の分類結果を統合する方法を提案し、実際に実験と評価を行った。その結果、この 2 種類の方法はともに、1 日単位の分類結果が 1 記事単位の分類精度を維持できず、精度が大きく下がった。入力の選別による 1 日単位の分類の精度が下がった原因を対象となる記事を人間が読んで分析したところ、人間でも正しい記事を選択できないほど記事が似ており、正しい記事を選別するのは難しいということが分かった。また、分類結果の統合による 1 日単位の分類の精度が下がった原因を散布図を用いて分析した

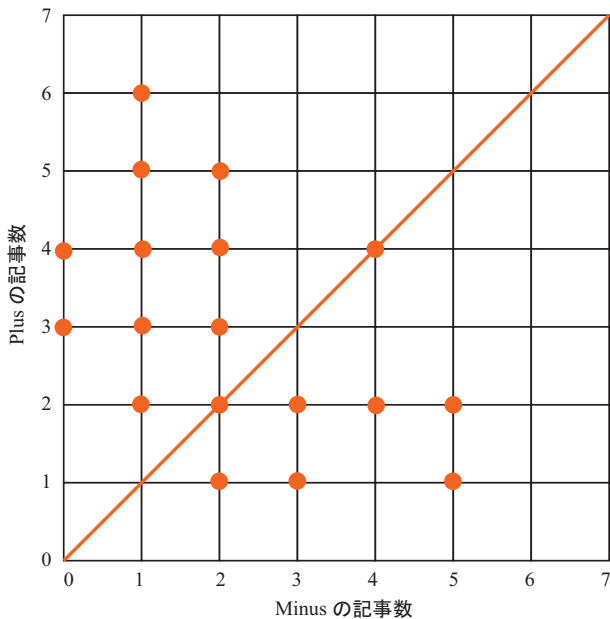


図 5 分類結果を可視化した散布図 (Minus クラス)

ところ、多数決による分類結果の統合自体が適切な方法ではないということがわかった

今後の方針としては、予測の精度をさらに上げるために、素性に現在の BoW ではなくもっと高次の言語情報を用いたり、分析に用いる媒体を Web ニュースだけでなく別の媒体も利用するなどといったことが考えられる。

文 献

- [1] 高橋宏圭, 関和広, 上原邦昭, “株価回帰と Web ニュース記事分析を組み合わせた株価動向推定”, 電子情報通信学会, 信学技報, pp.103-108, 2012.
- [2] 丸山健, 梅原英一, 諏訪博彦, 太田敏澄, “インターネット株式掲示板の投稿内容と株式市場の関連性”, 金融情報学研究会, 第 2 回研究会, 2013.
- [3] 辻洋平, 古宮嘉那子, 小谷善行, “Web ニュース中の複数企業に対応した株価予測”, 電子情報通信学会, 信学技報, pp.109-113, 2011.
- [4] 前川浩基, 中原孝信, 岡田克彦, 羽室行信, “大規模ニュースと株価収益率の予測可能性について”, 日本オペレーションズ・リサーチ学会, 2013 年春季研究発表会, pp.108-109, 2013.
- [5] Heeyoung Lee, Mihai Surdeanu, Bill MacCartney, Dan Jurafsky, “On the Importance of Text Analysis for Stock Price Prediction”, Proceedings of the Ninth International Conference on Language Resources and Evaluation, 2014
- [6] 和泉潔, 後藤卓, 松井藤五郎, “経済リポートのテキスト分析による金融市場動向推定”, 電子情報通信学会, 信学技報, pp.107-111, 2011.
- [7] 迫村光秋, 和泉潔, “twitter テキストマイニングによる経済動向分析”, 金融情報学研究会, 第 9 回研究会, 2013.
- [8] Bollen Johan, Huina Mao, Xiao-Jun Zeng, “Twitter mood predicts the stock market”, Journal of Computational Science, 2 (1), pp.1-8, 2011
- [9] Vladimir Vapnik, “The Nature of Statistical Learning Theory”, Springer-Verlag, 1995.
- [10] 砂山渡, 谷内田正彦, “文書要約のための特徴キーワードの発見による重要文抽出法-展望台システム-”, 情報処理学会研究報告, 自然言語処理研究会報告, pp.103-110, 2000.