

平成 27 年度

学士学位論文

潜在意味解析を用いた  
トピック抽出による株価予測

Prediction of Stock Market Price by  
Topic Extraction using  
Latent Semantic Analysis

1160318 塩見 侑司

指導教員 吉田 真一

2016 年 2 月 26 日

高知工科大学 情報学群

# 要 旨

## 潜在意味解析を用いた トピック抽出による株価予測

塩見 侑司

テキストマイニングを用いた株価予測が研究されている。それらの研究は人間が新聞記事やニュース記事などのテキストデータを投資材料としていることから、コンピュータにおいてその行動を実現することを目的としている。しかし、テキストデータは単語ベクトルと呼ばれる形で表現されるが、一般にその次元は数万を超えそのままでは機械学習に適さない。自然言語処理分野ではテキストデータから特徴的な情報の抽出を行う潜在意味解析が提案されている。潜在意味解析は本質的な特徴量を残しつつ必要のない情報量を削減することで次元削減を行う。そこで本研究ではこの潜在意味解析を用いて経済新聞記事からトピックを抽出し、そのトピックにより SVM(サポートベクターマシン) を用いて株価の予測を行う。株価予測精度の向上と株価予測に影響されるトピックの検証を目的とする。テキストデータとして 2014 年のニュース記事 (全 64564 記事) を使用する。各記事に形態素解析を行い、名詞を抽出し、各単語の出現回数を日付毎に求めた単語ベクトルを用いる。教師データとして、日経平均終値の上昇が前日比 +0.5%より増加すれば上昇、前日比 -0.5%より減少すれば下降、その間を値動きなしとした 3 クラスのデータを用いる。結果は、4, 5 個のトピックを選んだ時の 66.7%が最大の識別率であり、先行研究の奥村 [1] の相関を用いた識別結果よりも 11.6 ポイント向上したことを示す。また選出したトピックも直感的に株価の影響のあるトピックが選ばれていることから株価に影響するトピックが存在することを示す。

キーワード 株価予測, 潜在意味解析, テキストマイニング, トピック, TF-IDF, サポートベクターマシン

# Abstract

## Prediction of Stock Market Price by Topic Extraction using Latent Semantic Analysis

Stock price prediction using text mining has been studied. For human stock dealers, news articles are used to make decision for dealing. Therefore stock market price prediction using text mining has been studied in the area of artificial intelligence. Text data is generally expressed by word vector. The dimension is over ten thousands and it is not suitable for machine learning. Latent semantic analysis(LSA) has been proposed in natural language processing and it can extract the essential feature from text data. LSA reduces the dimension by keeping essential feature while unnecessary information reduced. This study propose the prediction method of stock prices by topic extraction from news articles using LSA. The purpose of this study is to improve the prediction accuracy and to clarify whether stock market price is influenced by topic. As a result, choosing 4 or 5 topics achieves the highest accuracy. The prediction accuracy is 66.7%. Also this thesis shows that the results of selected topics affect the stock market prices.

**key words** stock market prices, latent semantic analysis, text mainig, TF-IDF, support vector machine

# 目次

第 1 章	序論	1
第 2 章	テキスト分析の手法	3
2.1	形態素解析	3
2.2	単語頻出度ベクトルの作成	4
2.3	単語-文書行列	4
2.4	TF-IDF	4
2.4.1	TF(Term Frequency)	5
2.4.2	IDF(Inverse Document Frequency)	5
2.5	潜在意味解析	6
2.5.1	特異値分解	6
第 3 章	学習手法	8
3.1	サポートベクターマシン	8
3.2	グリッドサーチ法	9
3.3	交差確認法	10
第 4 章	新聞記事による LSA の適応と予測方法	12
4.1	学習データの作成	12
4.1.1	単語-文書行列	12
4.1.2	潜在意味解析によるトピック抽出	13
4.1.3	学習データの標準化	13
4.2	教師データの作成	14
4.3	学習方法	14
4.4	評価方法	15

## 目次

4.5	実験 . . . . .	16
第 5 章	結果・考察	17
5.1	結果 . . . . .	17
5.2	考察 . . . . .	19
第 6 章	まとめ	23
	謝辞	24
	参考文献	26
付録 A	名詞のみを用いた時のトピック	27
付録 B	名詞，動詞を用いた時のトピック	31
付録 C	名詞，形容詞，形容動詞を用いた時のトピック	35

# 目次

2.1	形態素解析 . . . . .	3
3.1	マージン最大とソフトマージンの概要図 . . . . .	9
3.2	交差確認法 . . . . .	11
4.1	潜在意味解析 . . . . .	13
4.2	スケーリング . . . . .	14
4.3	5-交差確認法 . . . . .	15
5.1	抽出されたトピック上位 10 トピック . . . . .	18
5.2	特異値分解にかかった時間 . . . . .	19

# 表目次

2.1	単語ベクトル	4
2.2	単語-文書行列	5
4.1	教師データ	14
4.2	SVM パラメータ	15
5.1	識別率の高いトピックの組み合わせの識別率	17
5.2	識別率上位 3 つの選出されたトピック	18
5.3	4 トピック全ての組み合わせを学習するのにかかる時間	19
5.4	名詞と動詞を用いたトピック	21
5.5	名詞, 形容詞, 形容動詞を用いたトピック	22
A.1	名詞のみのトピックとトピックに含まれる単語群	27

# 第 1 章

## 序論

コンピュータやスマートフォンなどのインターネットに接続できる機器の発展と普及により、必要な情報を手に入れることが容易になったことで、膨大な情報を得ることが可能になった。しかしそれらの情報はさらに膨大な情報量の中の一部であり、全ての情報を適切に分析し、目的の情報を得るには時間や知識が必要となる。このことは投資家の情報取得手段においても例外ではない。インターネット上で世界の経済情報を得られるようになった現在ではネット上の新聞記事やニュースサイトの情報が投資の判断材料となっている。一般に、投資家が株価の予測を行うための分析方法としてテクニカル分析とファンダメンタル分析がある。テクニカル分析は株価の取引の指標となる経済指標や株価の推移チャートなどの数値データを参考に予測を行う。ファンダメンタル分析は会社の業績や財務状況、さらにそれらに深く影響される世界情勢などを参考に株価を予測する。このように株価予測では株価などの定量的データの他にも世界情勢を公開している新聞記事などの定性的データも投資家の判断材料として用いられる。またビックデータ時代の到来として、顧客のレビューや twitter などの書き込みなどのデータをコンピュータによって解析しようとする試みもある [2][3]。このような背景から、インターネット上の膨大な定性的データから特徴的なデータを抽出し分析に活かすことを目標とした研究が行われている。

インターネット上に公開されているテキストデータに対して特徴的なデータを抽出し株価予測に活かそうとした試みが行われている [4][5]。先行研究では、2 単語の共起度に着目した 2 単語ベクトルの提案 [4] や時系列情報を加味した時系列単語ベクトルの提案 [5] がされている。自然言語処理の分野では対象のテキストデータをどのように特徴を設計するかが精度向上の鍵となっている。またデータ数が増えるにしたがって本質的な特徴量の抽出と必要



のない情報の削減を効率良く行わなければ情報量の膨大に対応しきれないという課題もある。こうした特徴設計の問題に対応するアイデアとして注目を集めているのが潜在意味解析によるテキストマイニングである。潜在意味解析では、本質的な特徴を残しつつ重要でない情報量を削減することができる手法であり、特徴量の抽出と次元削減の両方を同時に行うことができる。自然言語処理の分野で主にテキスト検索に利用されており文書と単語の索引付としての潜在意味インデキシングとして既に利用されている。本研究ではこの潜在意味解析を経済記事に適用することで、膨大なテキストデータから株価に影響のある特徴的なトピックが抽出できる可能性がある。

先行研究では、株価と単語の相関関数を用いた株価予測が行われており、サポートベクターマシンによる機械学習を用いた予測では 94%の精度で識別が行われている [1]。しかしこの予測方法では本来知ることのできない未来の単語と株価の相関を取っているために精度が正確でないと考えられる。また、先行研究の実験を適切な方法でやり直したところ精度が 57.1%にまで下がることが確認されている [6]。そこで本研究では株価と単語の相関をとる手法ではなく、経済記事から特徴的な情報の削減を最小限に抑えて株価予測に適した特徴的トピックを抽出することを目的とする。さらに先行研究から株価の予測にはサポートベクターマシンによる予想が高精度なことから、本研究でもサポートベクターマシンを機械学習として利用し、先行研究よりも低次元での株価予測の精度向上を目的とする。

本論文の構成として第 2 章ではテキスト分析の手法を説明し、本論文で取り扱う潜在意味解析の解析手法について説明する。第 3 章では株価予測に利用した学習方法について説明する。第 4 章では本研究の実験手法について説明を行い、潜在意味解析が株価予測に対して有効的であるかの検証を行う。第 5 章では本研究の実験結果と考察を述べる。

## 第 2 章

# テキスト分析の手法

テキストデータを機械学習にかける際に，特徴量の抽出を適切に行うためにテキストデータの分析を行い，特徴量が得やすい情報に変換しなければならない．本章では様々なテキスト分析の手法を解説し，本研究でどのように取り扱われるのかを説明する．

### 2.1 形態素解析

英文の文章の構成は主に英単語とスペースとカンマとピリオドで構成されている．従って単語のみを抽出する際はスペースやカンマ，ピリオドを分析するだけで抽出が可能になる．しかし日本語の文章であると上記のような方法では上手くいかない．日本語は単語と単語の区切りが英文のようにスペースで明確に示されているわけではないので，自然言語処理の一つである形態素解析によって文章を単語に分割する．図 2.1 に形態素の概要図を示す．文章を辞書によって意味を持つ最小の単位である形態素に分割して，前後の単語の関係から単語の品詞を判別することができる．

内閣の支持率が急劇に上昇した

内閣 / の / 支持 / 率 / が / 急激 / に / 上昇 / した

名詞   助詞   名詞   接尾辞   助詞   形容動詞   助詞   名詞   助動詞

図 2.1 形態素解析

## 2.2 単語頻出度ベクトルの作成

表 2.1 単語ベクトル

単語 (回数)	文書 1	文書 2	文書 3	...	文書 4
株価	12	10	5	...	32

## 2.2 単語頻出度ベクトルの作成

テキストデータから単語-文書行列を生成するために単語頻出度ベクトルを生成する。テキストデータに含まれる文書を意味のある最小の単語である形態素に分割する。形態素は文書に出てくる活用のまま抽出するのではなく原型を抽出することによって、同じ意味で利用されている単語が別々に抽出されることのないようにする。これらの形態素の出現頻度を要素とした単語頻出度ベクトル (表 2.1) のような単語頻出度ベクトルを作成する。表 2.1 の例では、文書 1 に含まれる「株価」という単語の出現回数が 12 回であると読み取れる。

## 2.3 単語-文書行列

上述で得られた単語出現頻度ベクトルは日付ごとにどのような単語が何回出現するのかという文書における単語の出現頻度をあらわす。潜在意味解析で用いるのはこの単語出現頻度ベクトルを全記事中に存在する単語分だけ並べた単語-文書行列を利用する (表 2.2)。単語-文書行列の要素には全ての記事で共通に出現する高頻出度単語に重みを低くし、ある記事に頻繁に出現する単語に対して重みを高くする TF-IDF 値を用いる。これにより定型文など、その日の出来事をあらわすことのできない特徴のない単語を除去することができる。

## 2.4 TF-IDF

TF-IDF とはいくつかの文書があった時に、文書の特徴付ける単語を決定する際に付ける値である。いくつもの文書で横断的に出現する単語はその文書の特徴付ける単語にはならないために重みを低くし、ある文書に高頻度で出現する単語には重みを高くする式 2.1 を利用する。

## 2.4 TF-IDF

表 2.2 単語-文書行列

単語 (回数)	1月4日	1月5日	1月6日	...	12月31日
上昇	0	12	0	...	15
内閣	11	24	23	...	32
支出	1	45	56	...	2
消費	3	10	34	...	57
⋮	⋮	⋮	⋮	...	⋮

$$\text{tf-idf} = \text{tf} \cdot \text{idf} \quad (2.1)$$

### 2.4.1 TF(Term Frequency)

TF(Term Frequency) はある文書中に含まれる単語の出現頻度を表す式である。式 2.2 ではある単語  $i$  の文書  $j$  における出現回数を  $n_{i,j}$  と表し分母には文書に出現する単語の総数を表している。この式により頻繁に出現する単語に重みを高くつけることができる。

$$\text{tf}_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (2.2)$$

### 2.4.2 IDF(Inverse Document Frequency)

IDF(Inverse Document Frequency) はすべての文書の数に対しある単語が出現する文書数の割合の逆数を表す式である。式 2.3 では  $D$  が総文書数を表し、 $\text{df}_i$  は単語  $i$  が出現する文書数を表している。この式により、すべての文書に共通している単語は低い値の重みの積をとることで文書全体で見た際の単語の相対的な重要度を下げることができる。したがって記事に出現する定型文などを除去することが可能となる。

$$\text{idf}_i = \log \frac{|D|}{\text{df}_i} \quad (2.3)$$

## 2.5 潜在意味解析

潜在意味解析 (Latent Semantic Analysis) は、1988 年 Susan T. Dumais らが情報検索技術として提案した技術であり自然言語処理に用いるテキスト解析の手法である [7]。文書中に含まれる複数の単語で構成された文章に注目した手法であり、単語の出現は独立ではなく、多くの単語は類似した文脈で共起することが多いことを利用して次元削減をする。解析の手法として特異値分解を用いて単語を基底へと写像する。この基底は文脈と単語の関係から潜在的な特徴が導き出されたものと考えることができる。潜在意味解析を用いることで次元の削減と潜在意味の抽出を行うことができる。数学的には主成分分析や因子分析と同等であり、単語間に共通する潜在的な因子を抽出することができる。

### 2.5.1 特異値分解

特異値分解は潜在意味解析に用いられる行列分解の手法である。主成分分析とほぼ同等の行列分解であり任意の行列を 3 つの直行行列の積に分解する [8]。式 2.4 にその式を表す。任意の  $n \times p$  行列  $X$  は

$$X = U\Lambda V^T$$

$$= \begin{pmatrix} u_1 & u_2 & \cdots & u_p \end{pmatrix} \begin{pmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\lambda_p} \end{pmatrix} \begin{pmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_3^T \end{pmatrix} \quad (2.4)$$

のように、3 つの直行行列の積に分解することができる。 $U$  は  $X$  の左特異ベクトル、 $V$  は右特異ベクトルであり、それぞれ正規直行行列である。 $\Lambda$  は対角要素が特異値である対角行列で、特異値の大きさに  $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$  となっている [9]。

文書と単語を用いた単語-文書行列を特異値分解した場合  $U$  は単語を基底とした行列、 $V$

## 2.5 潜在意味解析

は文書を基底とした行列となる。

特異値分解では次元数を減らすことで本質的な情報量を保存したまま重要でない情報量を減らす次元削減が行える。

次元削減を行う手順は  $U$  の基底のうち対応する特異値の大きいものから  $k$  個の数を取り出した部分行列  $U_{(k)}$  を用いて  $n$  次元から  $k$  次元に単語ベクトルを変換する。従って低次元空間の基底が潜在意味になるように変換を行うことができる。特異値が高い順に特徴を利用するので変換される基底は潜在意味の中でも最も重要な  $k$  個の基底だけを取る変換になっている。変換式を 2.5 に示す。

$$U_{(k)}^T X \tag{2.5}$$

単語と文書の関係では、単語を潜在意味の基底で表された行列  $U_{(k)}^T$  と元の単語-文書行列の積を取ることでトピック (潜在意味の基底)-文書行列を得ることができる。また得られた行列は  $k$  次元に削減される。

## 第 3 章

# 学習手法

本章では株価の予測に用いる機械学習の手法について述べる。

### 3.1 サポートベクターマシン

SVM(Support Vector Machine) は、1992 年に Vladimir N.Vapnik らによって提案された教師データを用いた学習手法の一つである [10]。SVM は現在知られている多くの機械学習手法の中でも高い性能を得ることができ広く用いられている。高い認識性能を発揮できる要因として未学習データに対して高い識別性能（汎化性能）を得ることができる点にある。未来の株を予測するために本研究では汎化性能が重要な点となる。従って本研究では SVM を識別器として用いる。

SVM は最適化問題を解いて、マージンを最大にする分離超平面を求める。一般に機械学習では汎化性能を高くすることが求められるが、汎化性能を維持するためには各クラスの点と分離超平面との距離（マージン）が最大になる超平面を求めることと、ある一定の割合でサンプルが識別誤りを許すように制約を緩めるソフトマージンとのバランスが重要となる。マージン最大とソフトマージンの概要図を図 3.1 に示す。図を見ると分離超平面を境におおよそクラス A とクラス B に分類されていることがわかる。しかし実際のパターン認識での実問題では線形分離可能な場合は殆どない。従って、実際のデータに対して SVM を利用するには図中  $a_1$  と  $b_1$  のように多少の識別誤りを許すように分離超平面を求める方法をとる。このように多少の誤りを認めた上で図中色付き四角と丸のような分離超平面から最も近いデータ点（サポートベクトル）を求める、この方法をソフトマージン法と呼ぶ。

### 3.2 グリッドサーチ法

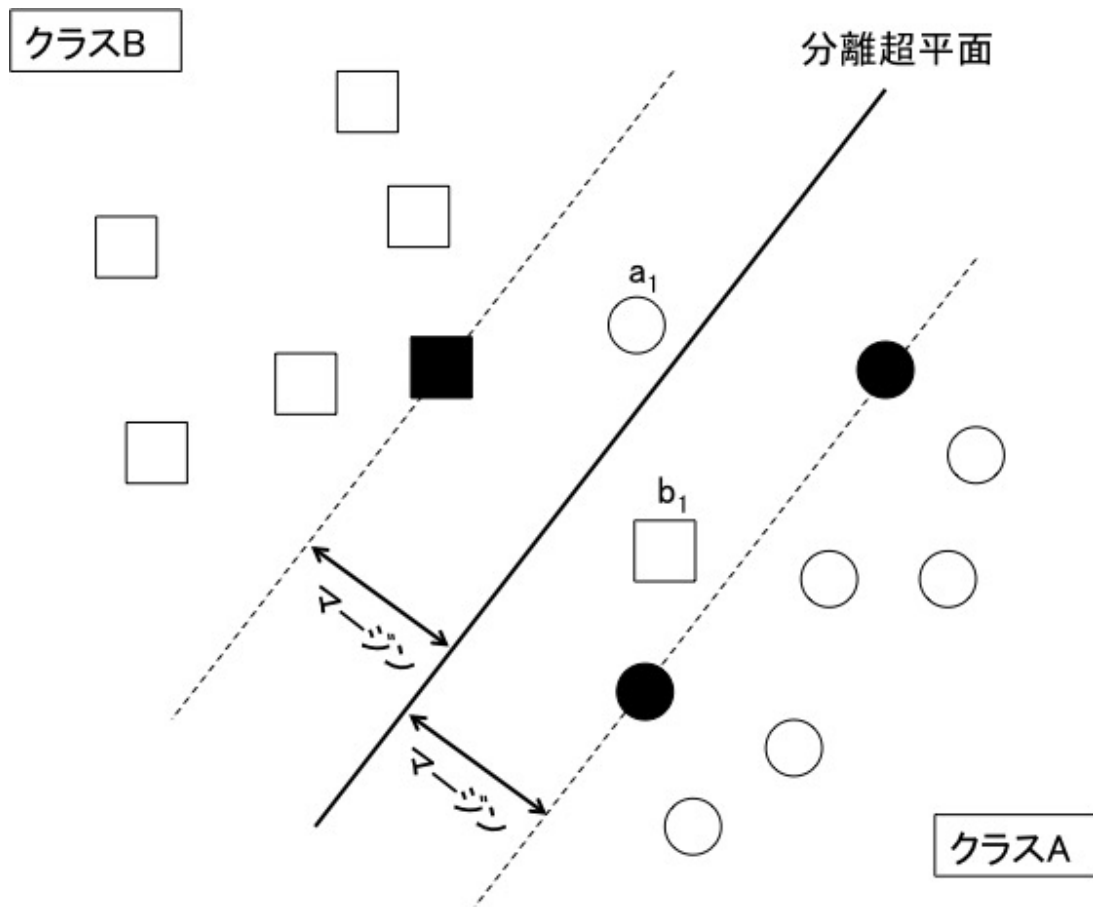


図 3.1 マージン最大とソフトマージンの概要図

### 3.2 グリッドサーチ法

グリッドサーチ法とは機械学習を利用する際にパラメータを自動で調節しながら識別する手法である。サポートベクターマシンではソフトマージンの誤り許容率を調節する  $C$  と RBF カーネル (ガウス) を用いた時の分離超平面について単純な分離面か複雑な分離面かを定める  $\gamma$  の二つのパラメータを調節する必要がある。SVM では識別率を向上させるにはこのパラメータの調節が不可欠であり、最適値を選ぶことで性能が大きく向上することもあるのでグリッドサーチ法を用いてパラメータを求める [9]。識別率が高くなるようなパラメータを求めるためにはパラメータを評価する必要がある。そこでグリッドサーチ法と組み合わせて利用されるのが交差確認法や 1 つ抜き法などの検定方法である。次の章で SVM のパラメータを評価するのに最もよく使われる交差確認法について説明する。



## 3.3 交差確認法

交差確認法 (cross validation) とは汎化能力を評価するための手法である。SVM では未知のデータに対しても識別できる汎化性能が高い。しかし、学習によって得られた学習モデルが学習データに含まれない未知のデータに対しても有効であるかどうかはわからない。ランダムに選ばれた訓練データが片方のクラスに分類されるデータばかり集まってしまうと過剰に適合してしまう学習モデルになる可能性もある。そのために、データを訓練用のデータと訓練用から得られた学習器を検証するテストデータに分けてテストデータに対する識別率で性能を評価する必要がある。そこで最も簡単な検証方法にホールドアウト法がある。ホールドアウト法ではデータを 2 つに分割して一方を訓練データにもう一方を検証用のテストデータに使用する方法である。データを分割する割合は自由に決定することができ、訓練データを多くするとテストデータが少なくなり学習モデルの性能が評価が不十分となる。テストデータを多くすると訓練データが少なくなり学習モデルの性能が悪くなる可能性が高くなる。このホールドアウト法ではデータが大量にあることが前提の評価方法であり、データが少ない場合は適切な評価を行うことができないといった欠点がある。

データが少ない場合に利用する検定方法として  $k$ -交差確認法がある。 $k$ -交差確認法の概要を図 3.2 に示す。図のように手元のデータをそれぞれ  $k$  個のグループに分割して、 $m - 1$  個のグループデータを使って識別器を学習し残りの 1 つのグループデータでテストをし、識別率精度の検証を行う。これを  $k$  回繰り返し、それらの平均値を識別率として求める。ホールドアウト法と異なって交差確認法では全てのデータを学習とテストに利用するのでより良い性能評価を行うことができる。しかし、この方法でもデータを分割するので分割によっては精度の偏りが生じる可能性が有る。従って分割を変えて交差確認法を繰り返し行い、その平均でどれだけ誤差があるのかを予測することが必要である。

### 3.3 交差確認法

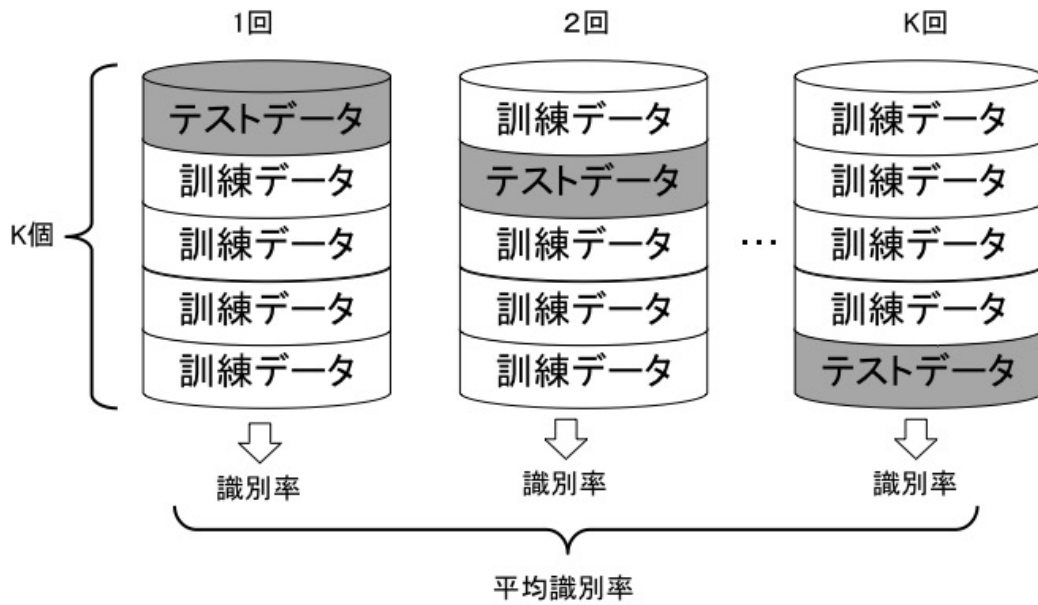


図 3.2 交差確認法

## 第 4 章

# 新聞記事による LSA の適応と予測 方法

### 4.1 学習データの作成

学習データの作成は以下の手順で行う。

- 単語-文書行列の作成
- TF-IDF 法の適応
- 潜在意味解析によるトピックの抽出
- 学習データの標準化

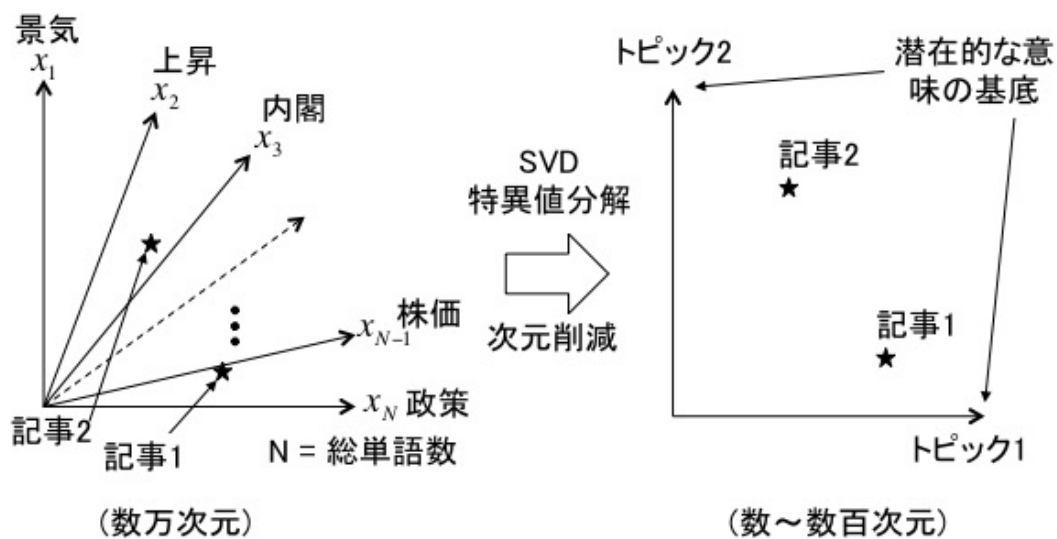
本研究では特徴の抽出に潜在意味解析を利用し，潜在意味解析を適応するための前処理として形態素解析，単語-文書行列の作成，TF-IDF 法による単語の重み付けを行う。

テキストデータとして日本経済新聞社が運営している日経テレコンから 2014 年の 1 年分の日本経新聞記事朝刊を利用する [11]。2014 年 1 月 1 日から 12 月 31 日までの期間の内株式取引が行われている期間かつ，朝刊が休刊でない期間 234 日分を対象とする 64564 記事を利用する。

#### 4.1.1 単語-文書行列

本実験で 234 日分の記事と全記事中に含まれる単語の出現回数を表した単語頻出度ベクトルを用いた単語-文書行列を利用する。単語-文書行列の要素には全ての記事で共通に出現

## 4.1 学習データの作成



$$\begin{aligned} \text{トピック1} &= 2.4 \times \text{上昇} + 1.4 \times \text{株価} + 0.5 \times \text{景気} \rightarrow \text{「経済」のトピック} \\ \text{トピック2} &= 3.2 \times \text{内閣} + 2.1 \times \text{政治} + 0.3 \times \text{政策} \rightarrow \text{「政治」のトピック} \end{aligned}$$

図 4.1 潜在意味解析

する高頻出度単語に重みを低くし、ある記事に頻繁に出現する単語に対して重みを高くする TF-IDF 値を用いる。これにより定型文など、その日の出来事をあらわすことのできない特徴のない単語を除去することができる。

### 4.1.2 潜在意味解析によるトピック抽出

得られた単語-文書行列に対して潜在意味解析を適用することでトピックが抽出される。トピックを抽出する際には図 4.1 のように単語-文書行列に対して特異値分解を行うことで次元削減され、低次元空間にデータを写像することができる。また、トピックが記事に対してどの程度の割合で表現されるかを要素とした文書-トピック行列を得ることができる。

### 4.1.3 学習データの標準化

機械学習の前処理として、統計的にデータを取り扱うために標準化処理（スケーリング）を行う（図 4.2）。学習するデータに大きな値をとる特徴量がある場合その特徴量に影響され

## 4.2 教師データの作成

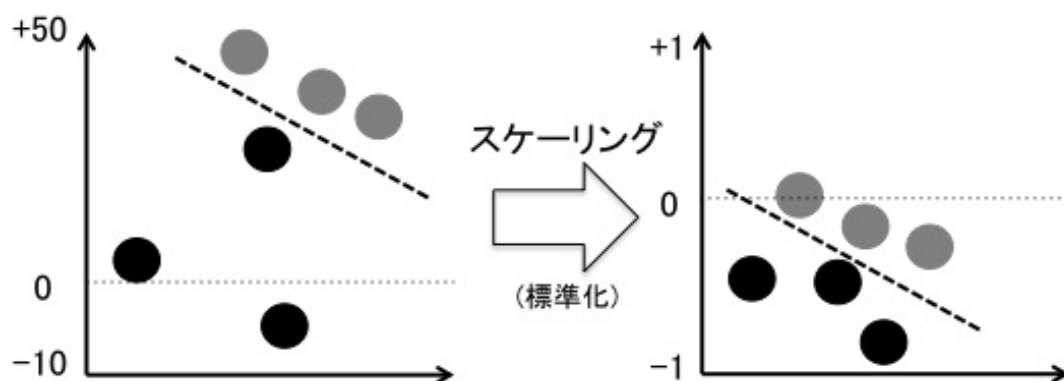


図 4.2 スケーリング

表 4.1 教師データ

下降	値動きなし	上昇
-1	0	+1

てしまうのを防ぐために行う。本実験では  $-1$  から  $1$  の値を取りうるように標準化処理を行う。

## 4.2 教師データの作成

教師データの作成には予測日前日の日経平均終値から予測日の日経平均終値が  $+0.5\%$  より増加すれば上昇  $-0.5\%$  より減少すれば下降その間を値動きなしとした 3 クラス教師データを用いる (表 4.1)。

## 4.3 学習方法

本実験では未来の株価予測を行うためにより汎化性能の高い学習器である SVM を用いて学習を行った。識別はグリッドサーチ法を用いて最も識別率の良いパラメータを用いて識別を行う。検証で用いたパラメータは表 4.2 に示す。また使用したカーネルは RBF カーネルである。

## 4.4 評価方法

表 4.2 SVM パラメータ

$\gamma$	1.0e-1	1.0e-2	1.0e-3	1.0e-4	1.0e-5
$C$	1	10	100	1000	10000

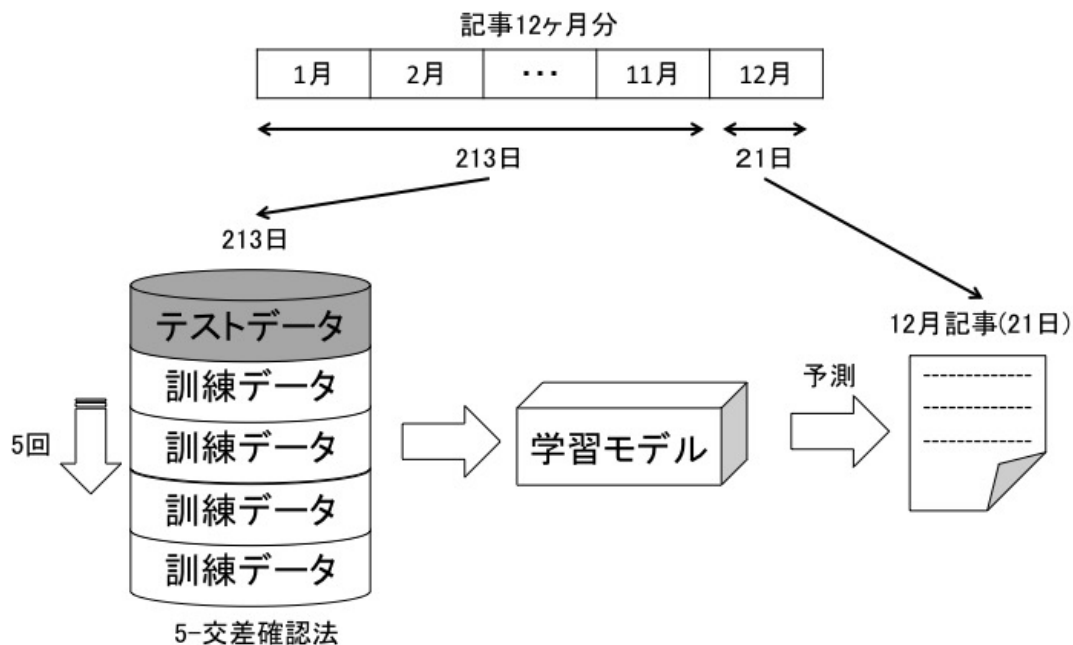


図 4.3 5-交差確認法

## 4.4 評価方法

本実験では SVM の性能評価に 5 分割交差確認法を用いた。今回は未知のデータに対して正確な汎化性能を評価するために図 4.3 に示すように 12ヶ月分 (全 234 日分) の記事のうち 11ヶ月分 (213 日分) を SVM の学習データにし、12 月分 (21 日分) を未来の記事と想定する。5 分割交差確認法では 11ヶ月分を学習、評価し最も識別率がよくなるようなパラメータを探索する。さらに、そのパラメータを利用し再度 11ヶ月分の訓練データから学習モデルを作成しその学習モデルで 12 月分の予測を行う。

## 4.5 実験

実験では SVM を用いて学習を行う．カーネル関数は RBF を用いて SVM のパラメータ  $\gamma$  と  $C$  を変化させて学習を行った．パラメータの取りうる範囲は下記の通りである (表 4.2)．本実験ではテキストデータから「名詞」のみを利用し，トピック抽出した．トピックは寄与率の高い値から上位 10 トピックを用いて学習に利用した．

追加実験として「名詞，動詞」と「名詞，形容動詞，形容詞」を利用したトピック抽出も行った．実験結果ではこれらの結果を示し，考察を行う．

# 第 5 章

## 結果・考察

### 5.1 結果

トピック数 1 から 10 個を利用して全ての組み合わせで最も識別率の高かったトピックの数とその識別率を表 5.1 に示す。トピック数 4, または 5 個の時の識別率 66.7% が最大識別率であった。

抽出されたトピックを図 5.1 に示す。トピックを見ると政党やソチオリンピック, 都知事などに意味付けられるトピックが抽出されているのがわかる。しかし中にはトピックの意味とは無関係な単語が含まれているトピックも存在していることがわかる。またトピック番号 7 は人間には意味付けできないような単語の集まりが存在している。

トピック数 4 個と 5 個の時の識別率上位 3 つの選出されたトピック番号を表 5.2 に示す。図 5.1 と合わせてみると政党や都知事のトピックが頻繁に選ばれていることがわかる。またトピック番号 7 も識別率が上位のトピックに存在していることがわかる。

表 5.1 識別率の高いトピックの組み合わせの識別率

トピック	識別率の高い上位 (%)				
2	52.4	47.6	47.6	47.6	42.9
3	52.4	52.4	47.6	47.6	47.6
4	66.7	61.9	57.1	52.4	52.4
5	66.7	61.9	57.1	57.1	52.4
6	61.9	61.9	61.9	57.1	57.1



## 5.1 結果

トピック	特異値	選出された単語 (10個)
STAP	0 4.320	細胞, 小保, 相続, 捕鯨, ペア, コイン, 農協, 噴火, 理研, 水素
政党	1 1.717	共, 民, 公, 新, 当, 麻, 市部, 無党派, 議席, 開票
ソチ五輪	2 1.575	秒, ジャンプ, メダル, ソチ, スケート, 決勝, バンクーバー, 男子, 金メダル, 京劇
時事	3 1.470	お知らせ, NEWS, 噴火, 秒, 列島, 共, 大岡山, 火山, 御嶽山, マガジン
	4 1.369	噴火, 火山, 御嶽山, クリミア, 小保, 赤崎, 小淵, 共, 青色, キューバ
外国	5 1.287	ハマス, 赤崎, クリミア, ペア, ガザ, コイン, 編入, 秒, 青色, キューバ
	6 1.263	捕鯨, キューバ, 噴火, タカタ, 火山, 赤崎, 共, 御嶽山, ルーブル, ハマス
潜在意味	7 1.245	捕鯨, キューバ, 噴火, 都知事, 蚊, 舩添, 秒, デング熱, 火山, ビーム
都知事	8 1.238	捕鯨, 細川, 噴火, 都知事, 蚊, 舩添, 秒, デング熱, 火山, ビーム
紛争	9 1.206	捕鯨, コイン, クリミア, 蚊, 佐治, デング熱, ハマス, 武力, PKO, 赤崎

図 5.1 抽出されたトピック上位 10 トピック

表 5.2 識別率上位 3 つの選出されたトピック

トピック	トピック番号		
4	0,1,5,7	1,3,7,8	1,5,7,8
5	1,3,7,8,9	0,1,5,7,8	1,3,4,7,8

特異値分解にかかった時間を図 5.2 に示す．本研究で用いた特異値分解のプログラムは記事数を  $n$  としてオーダ  $n$  で特異値分解できることがわかった．これは，記事数がある程度増えると単語数がほぼ変わらなくなるからと考える．

すべての組み合わせを SVM で学習し探索するのにかかった時間を表 5.3 に示す．用いるトピック数を  $n$  とすると，その組み合わせは  $2^n - 1$  であるので最も良い組み合わせを探索するのにかかる時間は特異値分解を行うより時間がかかることがわかる．本実験で利用した環境では次元数 12 以上であると数時間学習にかかるのでその日の終値を予測するには適していないと考えられる．したがってデータが増えた場合の組み合わせの計算量を工夫して減らさなければならない．

## 5.2 考察

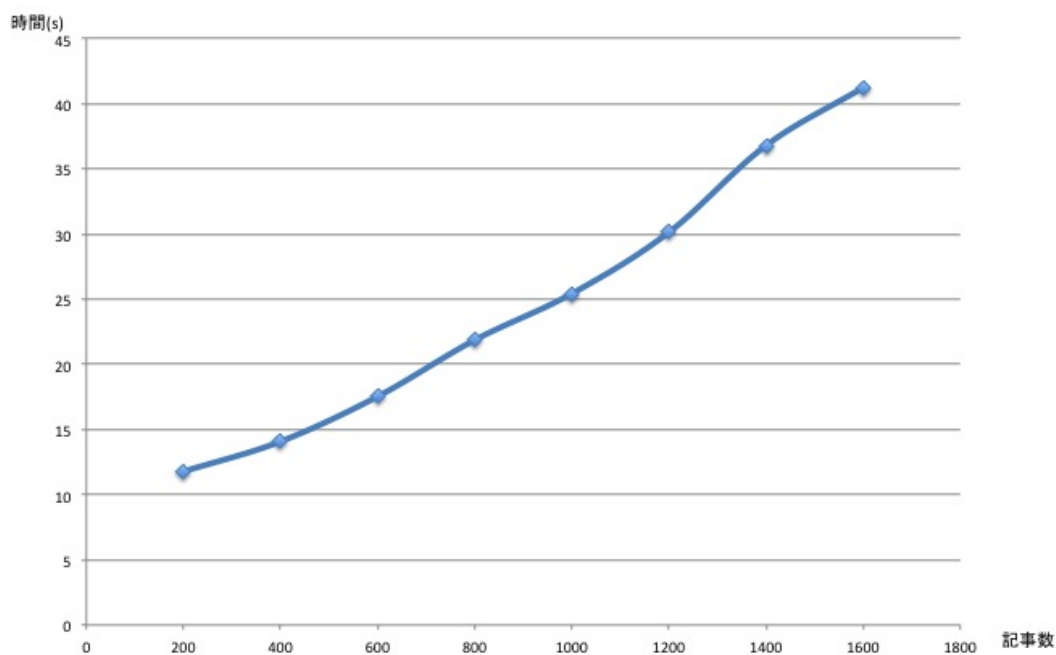


図 5.2 特異値分解にかかった時間

表 5.3 4トピック全ての組み合わせを学習するのにかかる時間

トピック数	組み合わせ	平均識別率	時間 (s)
4	${}_{10}C_4$	38.2%	582.21
5	${}_{10}C_5$	39.0%	560.11

## 5.2 考察

本実験では、日経経済新聞記事からトピックを抽出することで、オリンピックや、時事ニュース、政治などのトピックを抽出することができた。さらに、平均識別率はチャンスレベル (38.0%) 相当なものの予測精度が向上する時のトピックを見ると株価に影響があると考えられる政治トピックが選択されており、株価に影響がないと考えられるオリンピックなどのスポーツ記事が選ばれていないことから株価に影響するトピックが存在することが実験からわかった。また、人間では意味づけられることができない、単語群が株価に大きく影響するトピックとして選出されていることから、株価に影響のある潜在的な意味のかたまりの

## 5.2 考察

単語も存在していることがわかった．これにより人間では発見することができないトピックが解析によって発見されることにより，株価予測の精度向上に貢献しているのではと考えられる．

潜在意味解析の精度の検証も課題である．トピックを構成する単語を見るとスポーツや政治などに分類された単語もあったが中には共通の意味を導き出せない単語の集まりも存在した．今回は一つの記事に含まれる文書が 3000 文で構成されているために共起される単語が同文書中に複数存在する可能性が高いのがうまくいかなかった原因であると考えられる．解決策として，一つは記事に含まれる文書量を少なくして，共起する単語の数を減らすことがあげられる．しかし減らしすぎると今度はどのような単語にも共起されてしまうので適切な文書量を考える必要がある．

選出されたトピックが汎化性能をどのくらい要しているかは今回の実験では明らかにできなかった．一年分の経済記事から抽出されたトピックが数年後にも株価予測に影響のあるトピックであるかどうかはわからない．今後の課題として，記事を増やし様々な情勢の経済記事を学習し株予測に対応するために，株価の変動パターンとトピックの関係を明らかにし，より汎化性能を持ったトピックの抽出を行う必要があると考えられる．

また，トピック抽出を行う際に利用した特異値分解は高速であると考えられる．1 年分の記事約 50000 単語を 10 トピックに次元削減を行うのにかった時間 15.96 秒はその日の終値を予測するには十分有効である処理速度であると考えられる．しかし，株価に有効なトピックを抽出するためにすべての組み合わせを学習するのにかかる計算量は  $2^n$  と膨大になってしまう．したがって計算量を削減して株価に影響のあるトピックを抽出する必要がある．そこで株価とトピックの相関を取る手法により計算量の削減を行い現実的な処理時間で株価予測を可能にすることが必要となる．

用いる品詞を追加した場合のトピック 先の実験では名詞のみを用いたトピックの抽出と株価の予測を行った．しかし名詞のみのトピックであると，そのトピックが株価を上昇させるトピックであるか，下降させるトピックであるかが見えづらい．動詞や形容動詞などはそ

## 5.2 考察

表 5.4 名詞と動詞を用いたトピック

topic	最もピックをよく表す単語群 10 個
0.	. 細胞 小保 捕鯨 相続 ペア コイン 農協 噴火 理研
1.	共 ○ 民 当 公 麻 開票 , 新 議席
2.	秒 ジャンプ メダル ソチ 決勝 スケート バンクーバー 男子 金メダル 京劇
3.	. 共 秒 お知らせ NEWS ○ 噴火 列島 メダル ジャンプ
4.	噴火 火山 御嶽山 小保 クリミア 赤崎 秒 共 小淵 捕鯨
5.	, キューバ 赤崎 ペア クリミア タカタ ハマス 共 捕鯨 ルーブル
6.	噴火 キューバ , 捕鯨 ハマス 火山 御嶽山 タカタ ガザ クリミア
7.	捕鯨 , 蚊 袴田 デング熱 南極 秒 調書 クジラ 捕獲
8.	細川 都知事 秒 舩添 蚊 接種 ビーム クリミア 小保 赤崎
9.	捕鯨 クリミア コイン ハマス 武力 P K O 編入 蚊 ガザ 佐治

の単語のみで、株価に対して上昇をイメージさせる単語か下降をイメージさせる単語であるかが判断できる単語が存在する。例えば、「上がる」や「良い」、「豊かである」などは株価の上昇に繋がる単語、「下がる」や「悪い」、「悪化する」などは株価の下降に繋がるような単語であることがわかる。そのような単語が含まれるトピックが抽出された場合、株価予測に有用なトピックなのではないかと考えた。

そこで、追加実験として下記の品詞の組み合わせでのトピック抽出を行った。

- 名詞，動詞
- 名詞，形容詞，形容動詞

実験結果として名詞と動詞を用いて抽出したトピックを表 5.4 に、名詞，形容詞，形容動詞を用いて抽出したトピックを表 5.5 に示す。抽出した単語群を見るとどちらも動詞や形容詞，形容動詞が含まれず、名詞のみを用いたトピック抽出とほぼ同じ単語が抽出されていることがわかる。

## 5.2 考察

表 5.5 名詞，形容詞，形容動詞を用いたトピック

topic	最もピックをよく表す単語群 10 個
0.	細胞 小保 相続 捕鯨 ベア コイン 農協 噴火 理研 水素
1.	共 民 公 新 当 麻 市部 無党派 議席 開票
2.	秒 ジャンプ メダル ソチ 決勝 スケート バンクーバー 男子 金メダル 京劇
3.	お知らせ NEWS 秒 噴火 列島 共 大岡山 火山 御嶽山 マガジン
4.	噴火 火山 御嶽山 小保 クリミア 赤崎 小淵 秒 共 理研
5.	ハマス 赤崎 クリミア ベア ガザ コイン 編入 青色 イスラエル 秒
6.	捕鯨 キューバ 噴火 タカタ 共 赤崎 火山 ハマス 御嶽山 ルーブル
7.	捕鯨 キューバ 噴火 タカタ 火山 御嶽山 秒 袴田 赤崎 ルーブル
8.	細川 都知事 秒 捕鯨 舩添 ビーム 小保 クリミア 蚊 噴火
9.	コイン 捕鯨 クリミア 蚊 ハマス 武力 佐治 P K O 小保 デング熱

今回の実験では株価の変動に繋がるような動詞，形容動詞が含まれたトピックが抽出されることはなかった．考えられるのは，動詞や形容詞，形容動詞などは名詞と異なって，どのような文脈にも用いられる一般的なものが多いため，特定の名詞と頻繁に出現することがないのではないかと考える．つまり，様々な名詞と共起する場合は，多くのトピックに少しずつ寄与することによりある特定のトピックのみに含まれないのではと考える．したがってこの追加実験により，潜在意味解析では，多くの単語と共起するようなものはトピックから排除され共通の意味を持っている単語群からトピックを構成されることが確認できる．

## 第 6 章

### まとめ

本研究では、潜在意味解析によるトピック抽出により株価に影響のあるトピックの抽出を行うことができた。膨大なテキスト量から抽出したトピックは人間が判断しても株価に関連するトピックと考えられることから、潜在意味解析で抽出されたトピックは人間が株を予測するための判断材料の手助けになると考える。また人間では発見することができない株価とトピックの関係性も確認した。これは人間では分析のできないコンピュータによる解析結果であり、新たな株予測の可能性を示しているのではないかと考える。本研究では 2014 年度の新聞記事を対象にトピックの抽出を行ったが、短期間での株価予測ではこの潜在意味解析によるトピック抽出を用いた株価予測は有効であると結論付けられる。今後はより汎用的に用いることのできる予測を行うために、過去のデータから株価の変動傾向を分析し、長期予測にも対応するトピックの選出を行う必要がある。また潜在意味解析は解析対象を SNS などの他のテキストデータへの適用も可能であり、予測精度の向上につながるようなテキストデータが存在する可能性もあることが考えられる。

# 謝辞

本研究を進めるにあたり、ご指導して頂きました高知工科大学 情報学群 吉田真一准教授に心より感謝申し上げます。吉田先生には、3年生の頃から卒業研究に向けた取り組みの一環として、輪講を通して発表や質疑応答の指導をして頂きました。この輪講は私の卒業研究のアイデアの一つとなりました。また発表が苦手であった私が卒業研究発表の際にうまく発表ができたことにつながりました。また研究以外のことで人と触れ合う大切さとして様々なイベントを他の研究室と合同で楽しませてもらいました。一つ一つの出会いが最後の研究の成果と私の成長に繋がったと感じています。心より深く感謝するとともに、お体に気をつけてこれからの研究室のご発展と進化をお祈り申し上げます。

また本研究の副査を引き受けて頂きました高知工科大学 情報学群 篠森敬三教授と高田善朗准教授に深く感謝申し上げます。篠森先生には、技術的な部分をご指摘いただき大変勉強になりました。高田先生は、梗概の至らない部分をご指摘いただきました。卒業研究発表時までに研究の理解が深まり良い発表を行うことができました。両先生に深く感謝いたします。

研究室の方々にも感謝いたします。先輩である修士の松尾氏には、研究の技術的課題から、発表、梗概まで様々なところでご迷惑とお世話になりました。心が広い先輩であり、楽しく研究室での生活を送ることができました。ありがとうございました。

同学年の川上氏、西本氏、前原氏、田中氏は4年間の間、共に勉強し、研究室での時間を過ごしました。川上氏は同じ研究分野で、プログラムの技術的な問題や課題で行き詰まった時の相談にのっていただきました。西本氏は、就職活動の面接の練習や友人として普段の様々な場面でお世話になりました。前原氏は、研究室の盛り上げ役として明るく元気な性格が研究室での楽しい雰囲気となっており、楽しく卒業研究を終えることができました。田中氏は3年の間でしたが、趣味の話で盛り上がりました。研究室ではとても接しやすい性格であり田中氏を中心に研究室が和みました。短い間でしたがとても充実して楽しく過ごすこと

## 謝辞

ができました。

留学生の卓氏は私の不慣れな英語を一生懸命聞き取っていただきたくさんお話しすることができました。共通の趣味でお話しすることができてとても楽しい研究室生活を送ることができました。

後輩の方々は研究室のイベントでたくさんお世話になりました。後輩の思いつきの節分の行事は卒業研究の良い一息になりました。残り1年間卒業研究と就職活動を頑張って大学生活を楽しんでください。

そして、大学4年間を経済面や心の面から支えてくださいました両親には心から感謝いたします。次は自分が社会人として、両親に恩返しができるように頑張っていきたいと思えます。

最後に大学入学からの4年間でお世話になりました全ての方々に、改めて御礼申し上げます。



## 参考文献

- [1] 奥村順哉, “ディープラーニングによる経済記事テキストデータを用いた株価予測,” 高知工科大学学士学位論文, 2014 年度 .
- [2] Johan Bollen, “Twitter mood predicts the stock market,” *Journal of Computational Science*, 2(1), March 2011, Pages 1-8.
- [3] Qingyu zhang “International Journal of Information Technology & Decision Making,” Vol.7, No.4 (2008) 683720.
- [4] 山口裕輝, “テキストマイニングによる株価予測に適した機械学習”, 高知工科大学, 2012 年度学位論文 .
- [5] 中井淳人, “株価の時系列変化の予測のための特徴選択”, 高知工科大学, 2013 年度学位論文 .
- [6] 川上雄仁 “ニューラルネットを用いたテキストデータからの低次元特徴の抽出”, 高知工科大学, 2015 年度学位論文 .
- [7] Scott Deerwester, Susan T Dumais, George W Furnas, Thomas K Landauer “Indexing by Latent Semantic Analysis,” *Journal of the American Society for Information Science (1986-1998)* Sep 1990 41 6 ABI/INFORM Globalpg. 391.
- [8] Greenacre, M.J.(1984), “特異値分解 (SVD) と多次元解析”, “[http://elsur.jpn.org/reading\\_notes/Greenacre1984.pdf](http://elsur.jpn.org/reading_notes/Greenacre1984.pdf).”
- [9] 平井有三, “初めてのパターン認識”, 森北出版株式会社, 2012 .
- [10] Vladimir Naumovich, and Vladimir Vapnik, “Statistical learning theory,” Vol. 1. New York: Wiley, 1998.
- [11] 日経テレコン, “日本経済新聞朝刊”, “<https://t21.nikkei.co.jp/g3/CMNDF11.do>,” 2 月 4 日閲覧 .

# 付録 A

## 名詞のみを用いた時のトピック

表 A.1 名詞のみのトピックとトピックに含まれる単語群

topic	特異値	最もピックをよく表す単語群 10 個
0.	4.320	-0.047*細胞 + -0.045*小保 + -0.044*相続 + -0.043*捕鯨 + -0.040*ベア + -0.040*コイン + -0.040*農協 + -0.039*噴火 + -0.038*理研 + -0.036*水素
1.	1.717	0.669*共 + 0.253*民 + 0.162*公 + 0.142*新 + 0.136*当 + 0.130*麻 + 0.120*市部 + 0.113*無党派 + 0.107*議席 + 0.101*開票
2.	1.575	0.244*秒 + 0.169*ジャンプ + 0.159*メダル + 0.130*ソチ + 0.127*スケート + 0.127*決勝 + 0.118*バンクーバー + 0.113*男子 + 0.105*金メダル + 0.093*京劇
3.	1.470	0.139*お知らせ + 0.138*NEWS + 0.127*噴火 + -0.125*秒 + 0.114*列島 + -0.085*共 + 0.083*大岡山 + 0.080*火山 + 0.080*御嶽山 + 0.078*マガジン
4.	1.369	0.197*噴火 + 0.129*火山 + 0.123*御嶽山 + -0.111*クリミア + -0.108*小保 + 0.102*小淵 + 0.100*赤崎 + -0.099*共 + 0.093*秒 + -0.092*捕鯨
5.	1.287	-0.168*ハマス + 0.136*赤崎 + 0.132*クリミア + -0.127*ガザ + 0.124*ベア + 0.101*コイン + 0.092*捕鯨 + 0.091*編入 + -0.086*秒 + 0.085*キューバ
6.	1.263	0.273*捕鯨 + -0.247*キューバ + 0.175*噴火 + -0.159*タカタ + -0.115*赤崎 + 0.111*火山 + 0.109*共 + 0.101*御嶽山 + -0.097*ループル + -0.084*リコール
7.	1.245	0.348*捕鯨 + 0.192*キューバ + -0.122*細川 + -0.114*噴火 + -0.101*都知事 + 0.100*秒 + 0.085*タカタ + 0.084*袴田 + 0.083*赤崎 + 0.077*ループル
8.	1.238	-0.217*捕鯨 + -0.141*細川 + 0.133*噴火 + -0.123*蚊 + -0.117*都知事 + 0.107*小保 + -0.093*舩添 + -0.092*デング熱 + 0.086*火山 + 0.084*秒
9.	1.206	-0.139*捕鯨 + 0.125*コイン + 0.115*クリミア + 0.093*蚊 + -0.092*赤崎 + -0.082*佐治 + 0.077*編入 + -0.077*小保 + -0.074*武力 + 0.073*デング熱

10.	1.179	-0.191*捕鯨 + -0.181*ハマス + -0.145*ガザ + 0.142*小保 + 0.117*蚊 + -0.104*イスラエル + 0.102*理研 + 0.096*デング熱 + -0.090*空き家 + 0.089*実録
11.	1.158	0.267*赤崎 + 0.172*青色 + -0.117*キューバ + 0.113*天野 + 0.103*小保 + 0.099*ハマス + 0.092*ガリウム + -0.089*農協 + 0.086*学賞 + 0.083*図案
12.	1.153	0.236*キューバ + -0.195*赤崎 + 0.156*噴火 + 0.148*小保 + -0.126*青色 + 0.118*理研 + 0.095*火山 + -0.090*天野 + 0.088*御嶽山 + -0.084*ガリウム
13.	1.128	0.111*小保 + -0.096*水素 + 0.096*マスターズ + -0.094*可視 + 0.092*キューバ + -0.077*調書 + -0.076*石仏 + -0.075*接種 + 0.074*理研 + 0.068*赤崎
14.	1.117	-0.218*キューバ + 0.110*細胞 + -0.086*ルーブル + 0.072*解散 + 0.071*竿 + 0.071*小保 + -0.070*ハマス + 0.063*O P E C + 0.062*新 + 0.060*臓器
15.	1.109	-0.100*マスターズ + 0.079*空き家 + 0.073*P K O + -0.069*相続 + -0.068*贈与 + 0.068*良斎 + 0.062*土砂 + -0.062*可視 + -0.057*ウッズ + 0.057*マハティール
16.	1.104	-0.174*キューバ + -0.097*実録 + 0.076*相続 + 0.073*小保 + 0.071*小淵 + -0.070*健保 + -0.063*章司 + -0.062*天皇 + -0.062*良斎 + -0.060*マスターズ
17.	1.091	-0.179*小淵 + -0.106*松島 + 0.093*黒澤 + -0.083*パティック + -0.082*観劇 + -0.076*うちわ + 0.072*農協 + -0.069*マスターズ + -0.058*キティ + 0.058*土砂
18.	1.085	-0.106*キューバ + 0.103*水素 + 0.080*コイン + 0.078*解散 + 0.068*きち + 0.065*とうしろ + 0.064*マスターズ + 0.060*蚊 + 0.060*クロマグロ + 0.059*たい
19.	1.076	-0.138*コイン + 0.085*編入 + -0.082*ビット + 0.082*クリミア + 0.079*めぐみ + -0.067*回文 + -0.065*生薬 + 0.065*浪 + -0.065*柏 + 0.064*衆
20.	1.075	-0.097*マスターズ + -0.094*キューバ + 0.081*農協 + -0.080*可視 + 0.072*石仏 + 0.071*農地 + -0.069*古紙 + -0.066*良斎 + -0.064*ハンモック + -0.062*箸置き
21.	1.067	-0.133*キューバ + -0.084*佐治 + -0.078*空き家 + 0.076*良斎 + 0.076*古紙 + -0.071*接種 + -0.070*カジノ + 0.067*解散 + -0.064*勸請 + -0.062*ピーム
22.	1.066	-0.136*接種 + -0.076*生薬 + -0.076*良斎 + -0.075*ワクチン + 0.070*受精卵 + -0.068*東大寺 + -0.066*細川 + -0.063*相続 + -0.062*小保 + -0.057*贈与
23.	1.060	-0.093*空き家 + 0.086*調書 + 0.074*マスターズ + 0.067*ハマス + -0.066*相続 + -0.065*後見人 + -0.060*受精卵 + -0.056*釣り + 0.056*開門 + -0.054*ドアン
24.	1.055	-0.107*実録 + -0.086*パティック + -0.079*空き家 + -0.074*土砂 + 0.069*プリスペン + 0.069*新 + -0.065*タカタ + -0.063*献金 + -0.063*違憲 + -0.062*小淵

25.	1.052	0.130*空き家 + -0.124*保護司 + -0.090*キューバ + 0.082*マスターズ + -0.073*カジノ + 0.063*君野 + -0.063*コイン + 0.063*タカタ + 0.060*ダイエー + 0.057*コシヨウ
26.	1.050	0.089*調書 + 0.088*キューバ + 0.084*実録 + 0.068*献金 + -0.067*開門 + 0.067*古紙 + 0.067*違憲 + 0.064*天皇 + -0.061*良斎 + 0.061*彗星
27.	1.045	-0.084*コイン + 0.076*受精卵 + -0.071*新 + -0.070*実録 + 0.068*製糸 + 0.064*噴火 + -0.063*黒澤 + 0.057*流産 + -0.056*年功 + 0.056*精子
28.	1.043	-0.107*秒 + 0.081*コイン + -0.074*問 + 0.069*実録 + -0.061*開門 + 0.059*錦織 + 0.058*口バ + 0.057*法科 + -0.056*父子 + -0.055*大飯原発
29.	1.040	0.087*地価 + 0.080*空き家 + 0.067*小淵 + -0.063*ビーム + -0.061*秒 + -0.060*東大寺 + -0.058*囚 + 0.056*適 + -0.056*銚 + -0.055*実録
30.	1.037	-0.098*キューバ + -0.086*受精卵 + -0.076*黒澤 + 0.071*実録 + -0.071*アホウドリ + 0.070*運河 + -0.068*移設 + -0.068*仲井 + -0.065*コイン + -0.063*章司
31.	1.034	0.115*実録 + 0.079*天皇 + 0.066*遠泳 + -0.064*衆 + -0.063*法科 + 0.058*ゲラ + 0.057* 浮体 + 0.056*秒 + 0.055*献金 + -0.054*診療
32.	1.033	-0.109*小淵 + 0.085*生薬 + 0.076*スリランカ + 0.065*農協 + 0.062*蚊 + -0.058*松島 + 0.058*石綿 + -0.057*解散 + -0.057*観劇 + 0.056*錦織
33.	1.029	0.099*実録 + -0.086*遠泳 + -0.072*調書 + 0.071*つらら + -0.069*ゲラ + -0.061*朝顔 + -0.061*暑熱 + 0.060*天皇 + -0.055*黒人 + 0.052*瓦
34.	1.026	-0.105*調書 + 0.091*対馬 + 0.088*回文 + 0.078*バレエ + 0.076*古紙 + 0.073*タカタ + 0.066*総統 + -0.065*空き家 + 0.063*竿 + 0.060*中台
35.	1.021	0.079*実録 + 0.079*新 + -0.071*ゲラ + -0.069*農協 + -0.059*ジャッキー + 0.059*法科 + 0.057*重吉 + -0.056*ロツテ + 0.055*茶会 + 0.055*家なき子
36.	1.020	0.095*黒澤 + 0.076*保護司 + -0.069*イチョウ + -0.065*開門 + -0.063*石仏 + 0.063*対 馬 + -0.061*八雲 + -0.058*瓦 + -0.057*少林寺 + 0.056*口バ
37.	1.018	-0.109*開門 + 0.082*農協 + -0.069*重吉 + -0.066*瓦 + 0.061*ゲラ + 0.061*古墳 + -0.060*秒 + 0.058*竿 + -0.056*贈与 + 0.054*袴田
38.	1.016	-0.129*秒 + 0.091*実録 + -0.086*重吉 + -0.062*総統 + 0.060*移設 + -0.060*開門 + -0.057*中台 + -0.056*エチオピア + 0.056*縮緬 + 0.052*贈与
39.	1.014	-0.145*袴田 + 0.083*保護司 + -0.073*再審 + 0.068*マハティール + 0.067*開門 + 0.065* 義太夫 + -0.059*シシ + 0.056*差し止め + -0.056*千家 + -0.054*復

40.	1.013	-0.155*保護司 + -0.076*竿 + -0.067*実録 + -0.063*適 + -0.060*接種 + 0.060*コイン + 0.060*弥兵衛 + 0.057*重吉 + -0.056*天皇 + 0.053*空き家
41.	1.011	-0.091*小淵 + 0.088*パティック + -0.066*石仏 + 0.066*竿 + 0.059*保育 + 0.059*対馬 + -0.058*観劇 + -0.057*黒澤 + -0.057*砂時計 + -0.056*農協
42.	1.010	0.090*運河 + 0.079*贈与 + -0.074*キューバ + 0.070*ヤナギ + 0.067*つらら + -0.066* パティック + -0.064*空き家 + 0.063*古墳 + 0.061*調書 + -0.059*良斎
43.	1.006	0.133*保護司 + -0.084*実録 + 0.066*法科 + 0.066*家財 + -0.064*開門 + -0.061*義太夫 + 0.059*石仏 + 0.056*ダイエー + -0.051*天皇 + 0.050*黒澤
44.	1.003	0.069*保護司 + 0.068*新 + -0.066*適 + -0.064*ウナギ + 0.063*袴田 + -0.055*タカタ + 0.054*開門 + 0.054*子宮 + 0.053*後見人 + 0.053*解散
45.	1.001	-0.105*調書 + 0.074*竿 + -0.070*ベータ + -0.068*水素 + -0.063*献金 + 0.059*父子 + 0.056*衆 + -0.055*つらら + 0.055*ウイグル + -0.054*角川
46.	1.000	0.171*保護司 + 0.136*生薬 + -0.092*黒澤 + 0.091*新 + 0.070*石仏 + -0.060*勧請 + -0.054*農協 + 0.054*ベータ + -0.052*コイン + 0.052*滞納
47.	0.999	0.072*キューバ + -0.071*瓦 + -0.059*ゲラ + 0.059*生薬 + -0.058*遺品 + -0.058*実録 + 0.054*義太夫 + -0.050*クマ + -0.050*釣り船 + -0.047*開門
48.	0.997	-0.073*法科 + -0.066*風鈴 + -0.065*病床 + -0.064*イチョウ + -0.059*遠泳 + -0.058*自 販機 + -0.058*違憲 + 0.053*茂田井 + -0.051*特区 + 0.051*コイン
49.	0.997	-0.085*銚 + 0.074*生薬 + -0.068*大船 + -0.068*つらら + 0.066*違憲 + 0.064*保護司 + 0.063*APEC + 0.060*秒 + -0.059*踏切 + -0.056*超電導

## 付録 B

# 名詞，動詞を用いた時のトピック

topic	特異値	最もピックをよく表す単語群 10 個
0.	4.377	0.127*, + 0.045*細胞 + 0.043*小保 + 0.042*捕鯨 + 0.042*相続 + 0.038*ベア + 0.038*コイン + 0.038*農協 + 0.037*噴火 + 0.037*理研
1.	1.637	-0.548*共 + -0.479*○ + -0.210*民 + -0.147*当 + -0.128*公 + -0.115*麻 + -0.110*開票 + 0.100*, + -0.093*新 + -0.091*議席
2.	1.557	-0.243*秒 + -0.158*ジャンプ + -0.151*メダル + -0.123*ソチ + -0.122*決勝 + -0.122*スケート + -0.111*バンクーバー + -0.110*男子 + -0.100*金メダル + -0.094*京劇
3.	1.472	0.193*, + 0.133*共 + 0.129*秒 + -0.128*お知らせ + -0.128*NEWS + 0.119*○ + -0.105*噴火 + -0.105*列島 + 0.078*メダル + 0.078*ジャンプ
4.	1.355	-0.183*噴火 + -0.123*火山 + -0.116*御嶽山 + 0.111*小保 + 0.111*クリミア + -0.107*赤崎 + -0.102*小淵 + 0.095*共 + -0.089*秒 + 0.087*○
5.	1.29	0.287*, + -0.159*キューバ + -0.126*赤崎 + -0.115*クリミア + 0.115*ハマス + -0.114*ベア + -0.103*タカタ + 0.101*共 + -0.099*捕鯨 + -0.083*ルーブル
6.	1.261	-0.212*噴火 + 0.183*キューバ + -0.162*, + -0.146*捕鯨 + 0.145*ハマス + -0.124*火山 + -0.123*御嶽山 + 0.115*タカタ + 0.112*ガザ + -0.091*クリミア
7.	1.243	-0.449*捕鯨 + 0.137*, + -0.124*蚊 + -0.106*袴田 + -0.101*デング熱 + -0.099*秒 + -0.095*調書 + -0.093*南極 + -0.086*スコットランド + -0.083*捕獲
8.	1.228	-0.193*細川 + -0.153*都知事 + 0.122*秒 + -0.113*舩添 + -0.102*接種 + -0.094*ピーム + 0.093*小保 + 0.090*赤崎 + 0.087*クリミア + -0.082*蚊
9.	1.193	0.125*捕鯨 + -0.111*クリミア + 0.102*赤崎 + -0.093*コイン + -0.090*ハマス + 0.085*PKO + -0.081*蚊 + 0.080*武力 + 0.076*佐治 + -0.076*編入

10.	1.186	0.317*, + -0.162*噴火 + 0.155*キューバ + 0.153*赤崎 + -0.118*小保 + -0.109*火山 + 0.094*青色 + -0.090*理研 + -0.087*御嶽山 + 0.084*捕鯨
11.	1.168	0.209*捕鯨 + 0.186*ハマス + 0.146*ガザ + -0.118*キューバ + -0.113*蚊 + 0.110*イスラエル + -0.098*小保 + -0.093*調書 + 0.092*空き家 + -0.091*デング熱
12.	1.152	-0.286*赤崎 + -0.181*青色 + -0.123*天野 + -0.118*小保 + 0.101*解散 + -0.094*ガリウム + -0.092*○ + -0.090*学賞 + -0.088*図案 + -0.082*ノーベル
13.	1.14	0.415*キューバ + 0.128*噴火 + 0.118*ループル + -0.108*赤崎 + 0.100*小保 + 0.088*, + 0.079*理研 + 0.074*御嶽山 + 0.073*マスターズ + -0.070*プリズベン
14.	1.121	0.152*小保 + 0.095*理研 + -0.090*キューバ + 0.086*細胞 + 0.073*農協 + -0.072*浪 + -0.071*クリミア + -0.068*後見人 + -0.068*○ + -0.067*地価
15.	1.107	-0.137*マスターズ + -0.128*○ + -0.118*キューバ + -0.084*パティック + 0.084*, + -0.073*ウッズ + -0.072*赤崎 + -0.069*箸置き + -0.065*優勝 + 0.065*水素
16.	1.105	-0.125*相続 + 0.124*○ + -0.099*贈与 + -0.084*とうしろ + -0.078*ハマス + 0.075*章司 + 0.068*細胞 + 0.068*キューバ + -0.066*小保 + -0.066*きち
17.	1.085	0.134*○ + -0.109*, + 0.100*小淵 + -0.086*捕鯨 + 0.079*生薬 + -0.077*受精卵 + 0.063*マハティール + 0.061*PKO + 0.059*キティ + -0.057*とうしろ
18.	1.084	0.279*○ + -0.076*共 + 0.074*農協 + -0.070*当 + 0.066*調書 + -0.063*政務 + -0.062*開票 + -0.059*マスターズ + -0.059*衆 + -0.055*水素
19.	1.078	0.119*小淵 + 0.102*可視 + -0.089*農協 + -0.085*実録 + 0.078*ハンモック + 0.077*松島 + 0.076*傍受 + -0.074*健保 + 0.068*法制審 + 0.065*,
20.	1.069	-0.126*佐治 + -0.116*キューバ + -0.097*浪 + -0.083*スリランカ + 0.078*石仏 + -0.078*章司 + -0.077*わが国 + -0.076*空き家 + -0.076*後見人 + -0.074*ビーム
21.	1.066	-0.115*コイン + 0.085*良斎 + 0.080*相続 + 0.080*贈与 + -0.069*○ + -0.067*農地 + -0.064*ピット + -0.060*石仏 + -0.060*蚊 + 0.058*黒澤
22.	1.061	0.092*コイン + 0.091*接種 + 0.082*良斎 + 0.081*適 + 0.076*カジノ + 0.073*黒人 + 0.070*カタルーニャ + -0.067*めぐみ + -0.063*空き家 + 0.060*東大寺
23.	1.059	-0.136*実録 + 0.097*, + 0.091*接種 + -0.073*マスターズ + -0.071*天皇 + -0.070*縮緬 + -0.069*土砂 + 0.064*問 + -0.062*空き家 + -0.060*年功
24.	1.053	0.083*適 + 0.077*マスターズ + -0.070*受精卵 + 0.067*錦織 + -0.064*コイン + 0.063*ハマス + 0.063*クリミア + -0.058*生薬 + 0.058*実録 + -0.058*流産

25.	1.051	-0.271*○ + 0.151*共 + 0.130*キューバ + -0.103*タカタ + 0.096*当 + 0.075*開票 + -0.065*無党派 + -0.064*リコール + 0.061*瓦 + 0.057*秒
26.	1.048	0.097*実録 + -0.083*黒澤 + -0.081*ダイエー + -0.076*君野 + -0.071*竿 + -0.066*角川 + 0.062*天皇 + 0.062*保護司 + -0.059*空爆 + 0.059*賢人
27.	1.041	-0.110*コイン + 0.105*小淵 + -0.079*新 + 0.076*黒人 + 0.071*章司 + 0.068*タカタ + -0.066*実録 + 0.064*空き家 + -0.063*スコットランド + 0.060*松島
28.	1.037	-0.093*空き家 + 0.079*実録 + -0.061*瓦 + -0.058*熊 + -0.056*釣り + 0.055*東大寺 + -0.055*贈与 + 0.053*農協 + -0.052*開門 + 0.052*トレンチ
29.	1.036	-0.089*実録 + 0.081*秒 + 0.080*キューバ + -0.071*生薬 + 0.070*イチョウ + 0.066*移 設 + -0.062*石仏 + 0.061*長氏 + -0.059*天皇 + 0.058*つらら
30.	1.033	0.109*○ + -0.107*共 + 0.095*保護司 + 0.079*竿 + 0.076*小淵 + -0.075*生薬 + -0.073* 空き家 + -0.071*ゲラ + 0.071*合板 + 0.071*病床
31.	1.031	0.094*実録 + 0.090*秒 + -0.074*農協 + -0.070*衆 + 0.061*献金 + 0.060*キューバ + -0.057*適 + -0.056*谷垣 + 0.055*砂時計 + 0.055*パティック
32.	1.03	-0.095*○ + -0.079*受精卵 + -0.077*めぐみ + -0.071*贈与 + -0.064*重吉 + -0.064*製糸 + 0.061*調書 + -0.057*タカタ + 0.054*良斎 + -0.053*梅酒
33.	1.026	-0.090*キューバ + 0.080*保護司 + 0.074*○ + -0.068*コイン + 0.054*パティック + -0.053*空き家 + -0.052*同窓会 + 0.048*対馬 + -0.048*囚 + 0.045*プロレス
34.	1.025	0.078*調書 + -0.074*製糸 + 0.066*父子 + -0.062*黒澤 + -0.061*対馬 + 0.061*嫡出 + 0.059*ゲラ + 0.058*秒 + 0.054*インターン + -0.054*受精卵
35.	1.023	0.100*コイン + 0.080*遠泳 + 0.073*小淵 + 0.071*重吉 + -0.066*つらら + 0.065*違憲 + -0.062*ベア + -0.057*袴田 + 0.057*暑熱 + 0.056*キューバ
36.	1.021	0.130*○ + 0.080*古紙 + 0.077*水素 + -0.071*地価 + -0.069*傍受 + -0.069*共 + 0.066* 秒 + -0.062*ゲラ + 0.059*つらら + -0.057*調書
37.	1.019	-0.154*保護司 + 0.116*○ + 0.103*ゲラ + -0.083*開門 + 0.080*クマ + -0.077*秒 + -0.064*瓦 + -0.058*共 + 0.055*解散 + -0.055*マハティール
38.	1.015	0.104*保護司 + -0.073*農協 + -0.067*小淵 + 0.063*特区 + -0.057*ゲラ + 0.056*贈与 + -0.054*献金 + 0.054*原電 + -0.053*調書 + 0.051*生薬
39.	1.012	-0.069*袴田 + 0.065*賢人 + 0.065*重吉 + 0.064*空き家 + 0.060*家財 + 0.056*談合 + -0.055*調書 + -0.049*生薬 + 0.047*柏 + -0.046*接種



40.	1.009	-0.122*保護司 + -0.104*生薬 + -0.089*石仏 + -0.062*健保 + 0.056*開門 + -0.053*柏 + 0.052*鑑定 + -0.050*国保 + 0.049*遠泳 + 0.049*嫡出
41.	1.007	-0.111*黒澤 + -0.077*ダイエー + -0.076*調書 + -0.071*嫡出 + -0.065*違憲 + -0.064*重吉 + 0.062*東大寺 + -0.059*父子 + 0.057*朝顔 + -0.054*つらら
42.	1.007	-0.113*保護司 + 0.084*銚 + 0.081*踏切 + -0.074*嫡出 + -0.071*石仏 + 0.069*大船 + -0.068*生薬 + -0.065*黒澤 + -0.064*父子 + -0.054*血縁
43.	1.004	-0.127*黒澤 + 0.084*実録 + -0.082*調書 + 0.073*空き家 + 0.062*回文 + 0.062*千家 + -0.057*小張 + -0.056*垣 + 0.055*嫡出 + -0.054*美玲
44.	1.003	0.104*接種 + 0.072*マハティール + 0.070*ワクチン + 0.066*回文 + -0.065*つらら + 0.054*○ + -0.053*地価 + 0.053*G 2 + 0.053*子宮 + -0.051*角川
45.	1.001	-0.105*瓦 + -0.089*開門 + -0.083*保護司 + 0.078*実録 + 0.076*対馬 + 0.072*竿 + 0.067*回文 + -0.064*石仏 + 0.061*法科 + -0.061*衆
46.	0.998	0.115*○ + 0.084*ゲラ + 0.062*クマ + -0.060*コイン + 0.058*風鈴 + 0.057*瓦 + -0.056*石仏 + -0.053*当 + -0.052*共 + 0.051*スコットランド
47.	0.997	-0.070*八雲 + 0.065*キューバ + 0.064*ヤナギ + -0.056*タウト + 0.056*古墳 + 0.055*返礼 + -0.052*実録 + -0.052*旅券 + -0.051*若田 + 0.047*対馬
48.	0.995	-0.098*保護司 + -0.083*遠泳 + 0.077*解散 + 0.074*○ + -0.071*黒澤 + -0.057*ロツテ + -0.056*イチョウ + -0.054*暑熱 + 0.053*ドアン + -0.052*教団
49.	0.994	-0.101*保護司 + -0.082*秒 + -0.068*遠泳 + -0.058*風鈴 + -0.055*竿 + 0.052*クロマゲ口 + 0.051*義太夫 + 0.051*運河 + 0.046*農協 + 0.044*傍受

## 付録 C

# 名詞，形容詞，形容動詞を用いた時のトピック

topic	特異値	最もピックをよく表す単語群 10 個
0.	4.324	-0.047*細胞 + -0.045*小保 + -0.043*相続 + -0.043*捕鯨 + -0.040*ベア + -0.040*コイン + -0.039*農協 + -0.039*噴火 + -0.038*理研 + -0.036*水素
1.	1.716	-0.670*共 + -0.253*民 + -0.162*公 + -0.141*新 + -0.134*当 + -0.130*麻 + -0.120*市部 + -0.112*無党派 + -0.105*議席 + -0.100*開票
2.	1.574	0.243*秒 + 0.168*ジャンプ + 0.159*メダル + 0.131*ソチ + 0.128*決勝 + 0.127*スケート + 0.117*バンクーバー + 0.113*男子 + 0.105*金メダル + 0.094*京劇
3.	1.469	0.140*お知らせ + 0.138*NEWS + -0.125*秒 + 0.121*噴火 + 0.115*列島 + -0.085*共 + 0.082*大岡山 + 0.078*御嶽山 + 0.077*火山 + 0.076*マガジン
4.	1.368	0.201*噴火 + 0.135*火山 + 0.123*御嶽山 + -0.111*クリミア + -0.108*小保 + 0.105*赤崎 + 0.102*小淵 + -0.099*共 + 0.094*秒 + -0.091*捕鯨
5.	1.287	0.158*ハマス + -0.144*赤崎 + -0.135*クリミア + -0.126*ベア + 0.121*ガザ + -0.105*コイン + -0.095*編入 + 0.095*秒 + -0.088*青色 + -0.086*キューバ
6.	1.263	-0.317*捕鯨 + 0.247*キューバ + -0.161*噴火 + 0.153*タカタ + -0.101*火山 + 0.097*赤崎 + -0.095*共 + -0.093*御嶽山 + 0.092*ループル + 0.089*ハマス
7.	1.240	0.346*捕鯨 + 0.213*キューバ + -0.135*噴火 + -0.113*細川 + 0.094*タカタ + -0.091*都知事 + 0.084*赤崎 + -0.080*火山 + 0.079*ループル + -0.078*御嶽山
8.	1.238	-0.168*捕鯨 + -0.163*細川 + 0.134*噴火 + -0.131*都知事 + -0.120*蚊 + -0.104*舩添 + 0.092*秒 + -0.088*デング熱 + 0.087*火山 + -0.085*ピーム

9.	1.202	0.131*捕鯨 + -0.117*コイン + -0.110*クリミア + -0.095*蚊 + 0.083*佐治 + -0.077*デング熱 + -0.076*ハマス + 0.076*武力 + 0.074*PKO + 0.073*赤崎
10.	1.176	-0.202*捕鯨 + -0.178*ハマス + 0.148*小保 + -0.144*ガザ + 0.121*蚊 + 0.110*理研 + -0.106*イスラエル + 0.100*キューバ + -0.099*空き家 + 0.096*デング熱
11.	1.160	-0.257*赤崎 + -0.163*青色 + -0.150*小保 + -0.109*天野 + -0.103*理研 + -0.087*ハマス + -0.085*ガリウム + 0.084*解散 + -0.083*学賞 + 0.080*農協
12.	1.155	0.295*キューバ + -0.200*赤崎 + 0.162*噴火 + -0.129*青色 + 0.123*小保 + 0.103*火山 + 0.097*理研 + -0.096*天野 + 0.088*御嶽山 + -0.083*ガリウム
13.	1.130	-0.124*小保 + 0.092*地価 + 0.089*水素 + 0.085*浪 + -0.083*農協 + 0.078*可視 + -0.078*理研 + 0.076*クリミア + 0.071*イラク + 0.070*調書
14.	1.120	-0.224*キューバ + 0.081*細胞 + -0.081*ループル + -0.080*PKO + 0.076*解散 + -0.074*武力 + 0.071*調書 + 0.069*可視 + -0.067*わが国 + 0.067*新
15.	1.110	-0.129*マスターズ + -0.105*キューバ + -0.086*健保 + -0.075*ウッズ + -0.075*嫡出 + -0.071*全英オープン + 0.070*空き家 + -0.064*箸置き + -0.063*優勝 + -0.063*父子
16.	1.106	-0.122*相続 + -0.102*贈与 + -0.090*とうしろ + 0.079*実録 + -0.070*小淵 + -0.068*きち + -0.062*後見人 + -0.058*たい + 0.058*章司 + -0.057*小保
17.	1.087	-0.144*小淵 + 0.087*農協 + -0.086*松島 + -0.080*マスターズ + 0.076*とうしろ + 0.072*農地 + -0.070*可視 + -0.069*観劇 + -0.069*傍受 + -0.068*パティック
18.	1.084	0.189*キューバ + -0.103*解散 + -0.091*水素 + -0.080*小保 + -0.078*石仏 + -0.072*相続 + 0.069*共 + -0.066*きち + -0.063*理研 + 0.061*年功
19.	1.079	-0.095*佐治 + 0.080*石仏 + -0.079*マスターズ + -0.077*空き家 + -0.077*キューバ + 0.076*小淵 + 0.075*生薬 + -0.075*浪 + -0.075*受精卵 + -0.071*章司
20.	1.073	-0.099*キューバ + 0.088*古紙 + 0.088*贈与 + 0.086*良斎 + 0.080*土砂 + -0.076*蚊 + -0.066*錦織 + 0.066*笹井 + -0.065*ビーム + 0.062*東大寺
21.	1.068	0.118*コイン + 0.098*回文 + 0.088*水素 + -0.087*実録 + 0.080*接種 + 0.075*古紙 + 0.073*ビット + -0.069*クリミア + 0.067*良斎 + -0.061*編入
22.	1.067	0.113*接種 + 0.075*ワクチン + -0.062*コイン + 0.060*キューバ + -0.058*年功 + -0.056*ウナギ + 0.055*自販機 + -0.055*理財 + 0.055*生薬 + -0.054*袴田
23.	1.063	-0.101*小淵 + 0.092*調書 + 0.089*マスターズ + -0.075*空き家 + -0.074*章司 + -0.070*釣り + -0.068*受精卵 + 0.062*ハンモック + 0.061*黒澤 + -0.060*生薬
24.	1.056	-0.127*実録 + -0.074*小淵 + -0.074*土砂 + -0.073*献金 + -0.072*縮緬 + -0.067*天皇 + 0.067*空爆 + -0.065*違憲 + -0.064*黒人 + -0.063*パティック

25.	1.050	0.144*空き家 + -0.108*キューバ + -0.082*コイン + 0.079*ベア + 0.070*タカタ + -0.065*捕鯨 + -0.064*秒 + 0.059*君野 + 0.057*竿 + -0.056*当
26.	1.049	-0.127*実録 + 0.105*小淵 + -0.092*天皇 + -0.081*コイン + -0.074*スコットランド + 0.070*丸太 + -0.070*解散 + -0.069*古紙 + 0.067*松島 + -0.065*錦織
27.	1.043	-0.091*キューバ + 0.085*農協 + -0.078*黒澤 + -0.075*開門 + -0.074*秒 + 0.074*製糸 + -0.073*生薬 + -0.072*ダイエー + -0.071*つらら + -0.060*瓦
28.	1.040	-0.090*ロバ + -0.068*銚 + -0.068*失踪 + -0.066*贈与 + -0.062*シシ + 0.062*地価 + -0.061*法科 + -0.061*瓦 + -0.060*拉致 + -0.059*ビーム
29.	1.039	-0.116*錦織 + 0.089*黒澤 + -0.074*献金 + 0.071*良斎 + -0.071* Deng 熱 + -0.068*スリランカ + 0.065*東大寺 + -0.063*空き家 + -0.059*蚊 + -0.057*法科
30.	1.037	-0.098*秒 + 0.093*コイン + -0.088*ベア + 0.082*黒澤 + -0.065*保護司 + 0.063*タカタ + -0.060*問 + 0.059*キティ + -0.056*イチョウ + 0.054*ビット
31.	1.033	0.078*石仏 + -0.069*実録 + -0.066*クローム + 0.060*キューバ + -0.058*嫡出 + 0.057*受精卵 + -0.056*開門 + -0.054*成城 + 0.053*ジャッキー + 0.052*落語
32.	1.031	-0.091*秒 + -0.078*石仏 + -0.069*生薬 + -0.063*精子 + 0.059*踏切 + 0.059*衆 + -0.057*水素 + 0.057*運河 + -0.056*移設 + -0.054*受精卵
33.	1.027	0.091*調書 + 0.088*ゲラ + -0.083*重吉 + -0.075*開門 + 0.070*クマ + -0.065*銚 + -0.061*秒 + -0.059*竿 + 0.055*遠泳 + 0.054*細川
34.	1.026	-0.101*実録 + -0.082*古墳 + 0.074*父子 + -0.074*農協 + 0.073*キューバ + -0.073*竿 + 0.072*空き家 + -0.066*天皇 + -0.066*ビーム + 0.065*黒澤
35.	1.022	0.082*黒澤 + 0.072*ダイエー + -0.062*細川 + -0.062*朝顔 + -0.055*クリミア + 0.053*調書 + -0.051*キューバ + -0.050*小淵 + -0.047*黒人 + 0.046*スコットランド
36.	1.019	-0.087*対馬 + -0.077*キューバ + -0.076*保護司 + -0.070*違憲 + 0.067*袴田 + -0.067*黒澤 + -0.066*パティック + 0.063*瓦 + -0.062*ダイエー + 0.060*熊
37.	1.018	0.064*キューバ + -0.060*浮体 + -0.058*洋上 + 0.056*実録 + -0.056*開門 + -0.055*瓦 + -0.053*プルトリウム + 0.052*HDD + -0.051*空き家 + -0.050*遠泳
38.	1.017	0.097*ゲラ + -0.095*タカタ + -0.088*重吉 + 0.081*保護司 + -0.079*梅酒 + 0.073*クマ + -0.067*プルトリウム + -0.065*賢人 + 0.063*解散 + -0.062*バナナ
39.	1.014	0.103*保護司 + 0.080*梅酒 + -0.078*生薬 + 0.066*献金 + 0.065*ドアン + 0.064*実録 + 0.062*京劇 + -0.056*長氏 + 0.055*藩札 + -0.053*パティック

40	1.013	-0.109*瓦 + -0.108*開門 + 0.075*つらら + -0.063*黒澤 + 0.059*浮体 + 0.055*原電 + -0.055*タイル + -0.055*金貨 + 0.054*法科 + -0.054*君野
41	1.011	0.106*実録 + 0.074*遠泳 + 0.068*天皇 + 0.062*王座 + -0.058*調書 + -0.053*くじ + -0.051*クローム + -0.051*受精卵 + -0.051*ダイエー + 0.051*舩添
42	1.010	0.088*ガラ + 0.072*義太夫 + 0.069*調書 + 0.066*空き家 + 0.066*保護司 + -0.061*黒澤 + 0.060*クマ + -0.059*京劇 + -0.059*嫡出 + -0.058*石仏
43	1.007	0.176*保護司 + 0.103*調書 + 0.084*生薬 + -0.070*袴田 + -0.062*ドアン + -0.057*古墳 + 0.054*健保 + -0.051*ジャッキー + -0.051*父子 + -0.051*返礼
44	1.006	0.079*調書 + -0.071*男山 + 0.070*法科 + -0.062*囚 + 0.061*プロレス + -0.055*ウナギ + -0.055*同窓会 + -0.054*重吉 + -0.053*関電 + 0.052*適
45	1.003	0.079*保護司 + 0.067*空き家 + -0.066*法科 + -0.065*回文 + -0.062*竿 + -0.058*実録 + -0.056*小淵 + 0.056*インボイス + 0.055*生薬 + -0.055*ウイグル
46	1.002	0.134*保護司 + 0.090*生薬 + -0.083*空き家 + -0.079*つらら + -0.076*章司 + -0.075*角 川 + -0.071*秒 + -0.057*調書 + -0.056*銚 + 0.052*石仏
47	0.999	-0.078*空き家 + -0.077*パティック + -0.071*重吉 + 0.064*遠泳 + 0.060*黒澤 + 0.059* 接種 + 0.052*つらら + -0.052*ウナギ + -0.051*調書 + -0.048*嫡出
48	0.997	-0.076*袴田 + 0.063*開門 + 0.062*生薬 + 0.059*つらら + -0.055*贈与 + 0.054*編入 + 0.052*特区 + 0.052*法科 + -0.051*コイン + -0.051*踏切
49	0.996	-0.266*保護司 + -0.071*踏切 + -0.066*団長 + 0.065*ウナギ + 0.061*古紙 + -0.061*違憲 + 0.059*水素 + -0.057*つらら + -0.057*生薬 + -0.057*藍