

平成 26 年度

学士学位論文

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

Prediction of Stock Market Price from Economical
News Text Data Using Deep Learning

1150294 奥村 順哉

指導教員 吉田 真一

高知工科大学 情報学群

要旨

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

奥村 順哉

現在，機械学習を用いた株価予測が注目を集めており，数値データだけでなくテキストデータを用いた株価予測も行われている．しかし，テキストデータの処理では，対象となるテキストの特徴の表現が重要である．このような特徴量の設計の問題に対して有効な手法として，多数の階層を有する構造のディープラーニングが提案されており，自然言語処理にも応用されている．本研究では，ディープラーニングの代表的な手法である DBN(Deep Belief Network) を用いて経済記事の単語ベクトルから株価の予測を行い，多くの実問題で用いられている SVM(サポートベクトルマシン) との比較を行い，テキストマイニングにおけるディープラーニングの有効性を検証することが目的である．学習データとして 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，動詞，形容詞を抽出し，各単語の出現回数を日付毎に求め，単語ベクトルを生成する．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．単語ベクトルの各成分と教師データとの相関を求め，相関係数の絶対値の高い順に単語を選出し，SVM (線形およびガウシアン (RBF) カーネル) と DBN について k-分割交差確認法を用いて比較を行った．結果として，DBN が学習時間については SVM より劣るが，識別率については DBN の方が 3 ポイント高くなることを示す．

キーワード 株価予測，Deep Belief Network，テキストマイニング，単語ベクトル

Abstract

Prediction of Stock Market Price from Economical News Text Data Using Deep Learning

Junya Okumura

Recently, stock prices prediction using machine learning has been studied. Stock traders use not only numerical data, but also the qualitative data. Qualitative data, however have a problem that it is difficult to extract features. Deep learning has been proposed and it can extract the feature automatically using unsupervised learning as a pretraining. This study propose the prediction method of stock prices using DBN(Deep Belief Network), which is a major deep learning. Word vectors are used as input vector. Word vectors are retrieved from newspaper articles in the Internet. The training data are given as three categories , price up, price down, and same price. Effective words for prediction are selected by the correlation to training data. Proposed method is compared with SVM(Support Vector Machine). As a result, learning of DBN is much slower than that of SVM, and precision of DBN is higher than that of SVM. Therefore DBN is effective for stock prices prediction.

key words stock prices prediction , Deep Belief Network , text mining , word vector

目次

第 1 章	序論	1
1.1	研究背景	1
1.2	関連研究	2
1.3	研究目的	2
1.4	本論文の構成	3
第 2 章	テキスト分析手法および特徴ベクトル	4
2.1	形態素解析	4
2.2	単語ベクトル生成	5
2.3	単語選出	6
第 3 章	学習手法およびモデル評価法	7
3.1	Support Vector Machine	7
3.2	ニューラルネットワーク	8
3.3	Deep Belief Network	9
3.3.1	RBM	11
3.3.2	誤差逆伝播法	12
3.3.3	Dropout 法	13
3.4	交差確認法	13
第 4 章	ニュース記事による予測実験	16
4.1	学習用データセットの作成	16
4.1.1	教師データの生成	16
4.1.2	単語ベクトルの生成	16
4.1.3	相関に基づく単語の選出	17

目次

4.1.4	学習データの標準化	17
4.2	学習条件	17
4.2.1	DBN	18
4.2.2	SVM	18
4.3	性能評価	18
4.4	実験	19
第 5 章	結果・考察	20
5.1	結果	20
5.2	考察	22
5.2.1	追加実験	23
第 6 章	まとめ	24
	謝辞	25
	参考文献	27
付録 A	グリッドサーチ	29
付録 B	選出単語	40

目次

2.1	形態素解析処理の概要	5
3.1	マージン最大化	8
3.2	フィードフォワード型の構造	9
3.3	リカレント型の構造	9
3.4	DBN の概略図	10
3.5	RBM の概略図	11
3.6	Dropout 法の概略図	14
3.7	交差確認法	15
5.1	DBN の識別結果	21
5.2	層の数と識別精度の推移	22

表目次

2.1	単語ベクトルの例	5
4.1	教師データ	17
4.2	検証用パラメータ	18
5.1	各手法の識別率 (\pm S.D)	20
5.2	各手法の単語数と学習時間 [s]	21
5.3	各手法の識別率	23
A.1	グリッドサーチによるパラメータと精度の変化	29
A.1	グリッドサーチによるパラメータと精度の変化	30
A.1	グリッドサーチによるパラメータと精度の変化	31
A.1	グリッドサーチによるパラメータと精度の変化	32
A.1	グリッドサーチによるパラメータと精度の変化	33
A.1	グリッドサーチによるパラメータと精度の変化	34
A.1	グリッドサーチによるパラメータと精度の変化	35
A.1	グリッドサーチによるパラメータと精度の変化	36
A.1	グリッドサーチによるパラメータと精度の変化	37
A.1	グリッドサーチによるパラメータと精度の変化	38
A.1	グリッドサーチによるパラメータと精度の変化	39
B.1	選出単語一覧	41

第 1 章

序論

1.1 研究背景

近年，ネット証券会社の増加の影響を受けて個人投資家による株取引が増加している．このため，株式投資のための支援システムの需要も増加している．このような背景を受けて，重回帰分析による予測や，サポートベクトルマシンを使った予測など，コンピュータを使った株価の予測に関する研究が積極的に行われている [1]．株価の上昇か下降かを予測する単純予測では 90%以上の精度を出した研究もある [2]．しかし，多くの場合，予測には過去の株価の推移や経済指標などの定量的な数値データのみを扱うことが多かった．

投資家が株価の予測を行うための分析手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．テクニカル分析では過去の株価や出来高などの時系列パターンから予測を行う．ファンダメンタル分析は企業の業績や財務状況および，それらの背景となる世相や経済情勢などに基づいて予測を行う手法である．このように，人間が株価の予測をする際には，出来高などの市場指数を代表とする定量的データのみでなく，ニュース記事などの定性的な情報も判断に用いられる．しかし，ビッグデータ時代の到来に伴い，電子化されたデータも増大し，一日に数万件もあるニュース記事を人が分析することは困難である．このため，テキストデータがもたらす情報量も数値データ同様に膨大なものとなり，定性的データも解析対象として重要なファクターになっている．このような背景を受けて，テキストマイニングによる株価予測に関する研究が行われている．

1.2 関連研究

1.2 関連研究

インターネット上で取得した新聞記事のテキストデータから、機械学習を用いて株価の値動きを予測する試みが行われている [3][4]. 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 [3] や時系列情報を加味した時系列単語ベクトルの提案 [4] がされており、識別率を上げる特徴量の設計が一つの課題となっていた。このようにテキストデータを扱う自然言語処理の分野では、対象となるテキストをどのような特徴量で表現するかが重要な問題となる。単語ベクトルなどのテキストデータにおける特徴はデータの増大とともに、特徴の数も増え、有効な特徴を見いだすことが困難となる。こうした特徴設計の問題に対するアプローチとして近年、注目を集めている技術がディープラーニングである。

ディープラーニングは既に音声認識や画像認識、自然言語処理の分野でも優れた認識精度を発揮することで注目を集めている [5][6]。経済指標予測の分野でもディープラーニングの研究が行われており、為替レートの変動の予測で従来のニューラルネットワークと比較して、良い結果となることも報告されている [7]。ディープラーニングを用いることによって、従来人の手によって設計されていた特徴量を、教師なし学習器のネットワークによって自動的に学習させることができる。また、ディープラーニングではネットワークの層を重ねることで特徴量のより複雑な組み合わせが考慮できる [8]。このような性質からテキストデータをディープラーニングで学習させることによって、複数の単語の組み合わせを考慮した複雑な文章構造を考慮した識別をすることができる可能性がある。

1.3 研究目的

先行研究ではテキストマイニングからの株価予測に適した機械学習手法の比較が行われており、サポートベクトルマシンによる予測が高精度なことが報告されている [3]。本研究では、ディープラーニングを株価の予測に適用する。ディープラーニングとサポートベクトルマシンについて識別精度と学習時間についての比較を行い、更なる精度の向上とテキストマイニングにおけるディープラーニングの有効性を示すことが目的である。また、識別に用

1.4 本論文の構成

いる単語数 (次元数) やディープラーニングの層の数などのパラメータの変化と精度の変化についての検証も行う。

1.4 本論文の構成

本論文の構成として第 2 章ではテキストマイニングに関する技術について説明し、株価の予測にテキストデータを使用するための手順について説明する。第 3 章では予測に使用する機械学習の手法について説明を行う。第 4 章で予測実験についての説明を行い、第 5 章で結果と考察を述べる。

第 2 章

テキスト分析手法および特徴ベクトル

本章ではテキストマイニング技術を用いて株価の予測を行うための手順について説明を行う。テキストマイニングとは大量のテキストデータから価値のある情報を見つけ出すことである。

テキストマイニング技術を用いて、定性的データであるテキストデータを使用して機械学習を行うために、以下の手順で定量的データへの変換を行った。

- 形態素解析
- 単語ベクトル生成
- 単語選出

2.1 形態素解析

まず、取得したニュース記事を統計的に分析するために形態素解析を行う。形態素解析とは、自然言語処理技術の一つで文を単語に分割し、各単語の品詞を求めることである。図 2.1 に形態素解析処理の概要を示す図のように形態素解析によって文章を単語ごとに区切ることができる。日本語文章の場合は英語のように単語と単語の区切りが明確に示されている訳ではないので、文章を辞書によって意味を持つ最小単位である形態素に分割し、前後の単語の出現関係からそれぞれの単語の品詞を判別する。

2.2 単語ベクトル生成

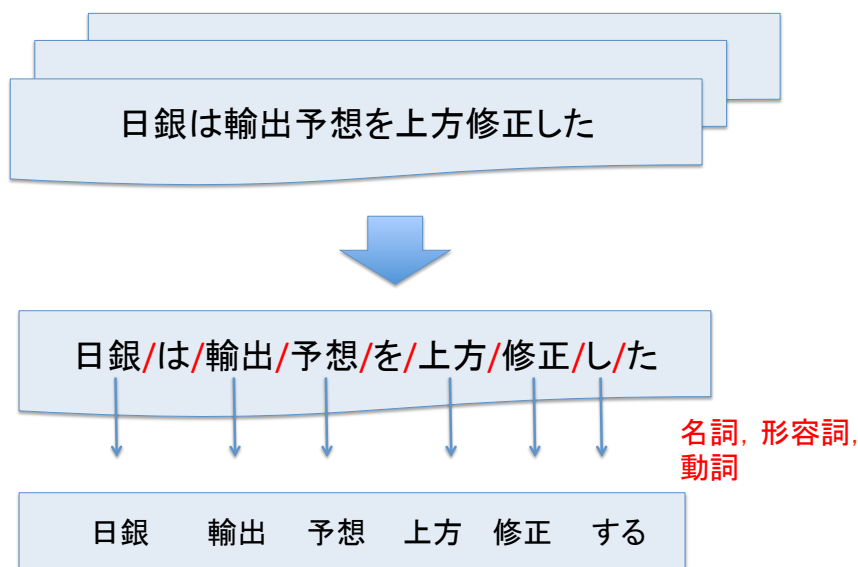


図 2.1 形態素解析処理の概要

2.2 単語ベクトル生成

単語ベクトルとは、日付毎の各単語の出現回数をリスト化したものである。形態素解析によって分割した単語から名詞、動詞、形容詞のみを取り出し、単語ベクトルを作成する。表 2.1 に単語ベクトルの例を示す。表 2.1 の例では、1月1日のニュース記事に「回復」という単語が出現した回数が3回であると読み取れる。

表 2.1 単語ベクトルの例

	回復	消費	内閣	...
1月1日	3	0	1	...
⋮	⋮	⋮	⋮	...
12月31日	0	3	0	...

2.3 単語選出

作成された単語ベクトル中には、株価の値動きと関連の低い単語が多く含まれている。また、1年分の記事を使用して単語ベクトルを生成するとおよそ5万次元の高次元データとなり、次元の呪いの影響を受けてしまう。そのため、株価の値動きと関連の高い単語を選出し、単語ベクトルの次元を削減する必要がある。そこで株価の値動きと関連の高い単語を選出するために相関係数を用いる。2組のデータ列を $(x_i, y_i) (i = 1, 2, \dots, n)$ とすると、相関係数は式 (2.1) で表される。 \bar{x}, \bar{y} はそれぞれのデータの平均である。

$$\frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}} \quad (2.1)$$

相関係数が正の値を取るときを正の相関関係と呼び、一方の値の増加に伴ってもう一方のデータの値が増加する。逆の場合を負の相関関係と呼ぶ。本研究では単語の出現回数と株価の値動きとの相関係数を求め、絶対値の高い順に単語の選出を行う。

第 3 章

学習手法およびモデル評価法

本章では，株価の予測に用いる機械学習アルゴリズムについての説明を行う．具体的には実世界で広く利用されている SVM とディープラーニングについて述べる．ディープラーニングは従来の機械学習より深い層構造をしている機械学習手法全般のことを指すため，基礎技術であるニューラルネットワーク，研究で用いたディープラーニングの代表的な手法である DBN の概要，および DBN に用いられる学習手法について述べる．また，機械学習の性能評価に用いる交差確認法についての説明も述べる．

3.1 Support Vector Machine

SVM(Support Vector Machine) は，1992 年に Vladimir N.Vapnik らによって考案された教師あり学習を用いる識別手法の一つである [10]．SVM は高次元特徴空間において線形識別関数を用いる学習システムである．最適化問題を解き，各クラスのデータ点との距離(マージン)が最大になる分離超平面を求める．図 3.1 にマージン最大化の概略図を示す．分離直線に最も近い各クラスのデータ点をサポートベクトルと呼び，サポートベクトル間の距離を最大にすることで高い汎化性能を実現することが出来る．SVM は本来，線形分離可能な問題に対する分類器である．しかし実問題の識別において線形分離可能となる場合は限られる．そのため，ベクトルを高次元特徴空間に写像することで非線形な識別を行う．このような手法をカーネルトリックと呼び，最適なカーネル関数を用いることによって非線形分類問題に対しても優れた性能を発揮することができる．

3.2 ニューラルネットワーク

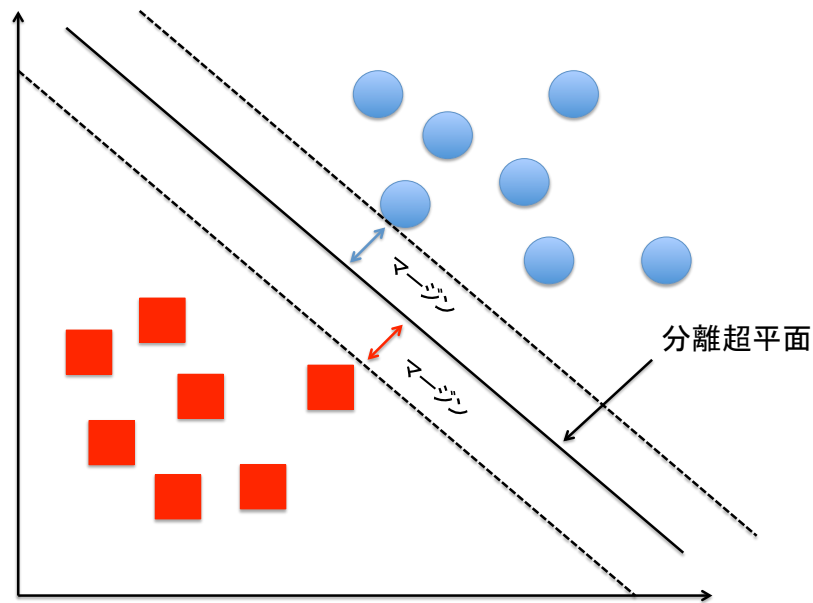


図 3.1 マージン最大化

3.2 ニューラルネットワーク

ニューラルネットワークは人間の神経回路網を模倣した学習手法である。ニューラルネットワークはニューロンの振る舞いをモデル化したユニットの結合から成り立っている。ユニットの結合には強度と方向性があり、ニューラルネットワークは結合の強度を変えることによって学習を行う。結合の方向性によってニューラルネットワークは大別することができ、フィードフォワード型とリカレント型の2種類に分けられる [11]。フィードフォワード型を図 3.2, リカレント型を図 3.3 に示す。フィードフォワード型はネットワークが階層構造になっていて、ニューロンに流れる信号が入力層から出力層への1方向のみに制限される構造となっている。これに対しリカレント型はネットワークの各ニューロンが相互に結合しており、双方向に信号が流れる構造となっている。後述する DBN はフィードフォワード型ニューラルネットワークである。

3.3 Deep Belief Network

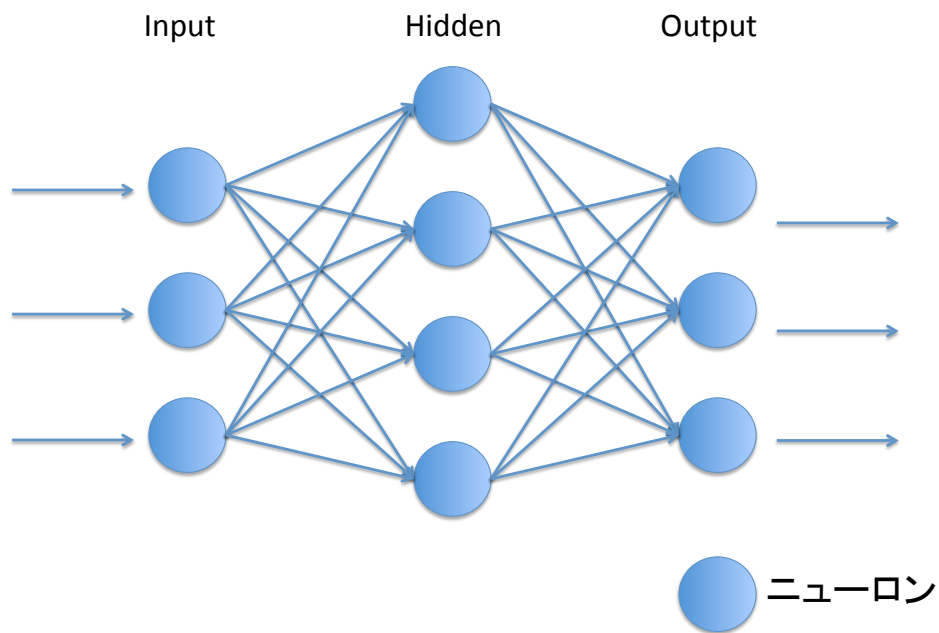


図 3.2 フィードフォワード型の構造

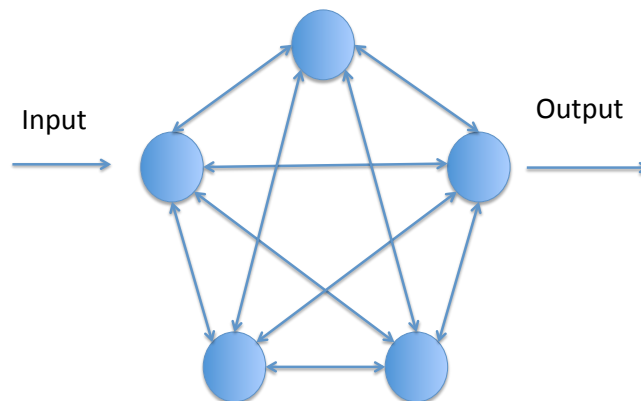


図 3.3 リカレント型の構造

3.3 Deep Belief Network

DBN(Deep Belief Network) は 2006 年に Hinton によって考案されたディープラーニングの手法の一つである [12] . DBN は RBM(Restricted Boltzmann Machine) を積み上げた

3.3 Deep Belief Network

フィードフォワードニューラルネットワークと見なすことができる。図 3.4 に DBN の概略図を示す。DBN による学習は Pre-training と Fine-tuning という 2 段階で行う。図に DBN の学習方法を示す。Pre-training では RBM の層ごとに教師なし学習を行う。第 1 層の RBM は入力されたベクトルを基に教師なし学習を行い隠れ層のベクトルを算出する。第 2 層の RBM は第 1 層が算出したベクトルを入力ベクトルとして学習する。この行程を繰り返すことによって特徴抽出を行う。Fine-tuning では Pre-training で学習した全ての RBM を一つのフィードフォワードニューラルネットワークとみなして誤差逆伝播学習 (Back Propagation, BP) を行う。多層ニューラルネットワークは表現力が高いため、過学習に陥りやすいという問題がある。また、誤差逆伝播法による学習を行うと、勾配が拡散し、重みの更新が十分にできなくなってしまうという問題もある。これらの問題を DBN では、RBM による Pre-training と Dropout 法で解決している。

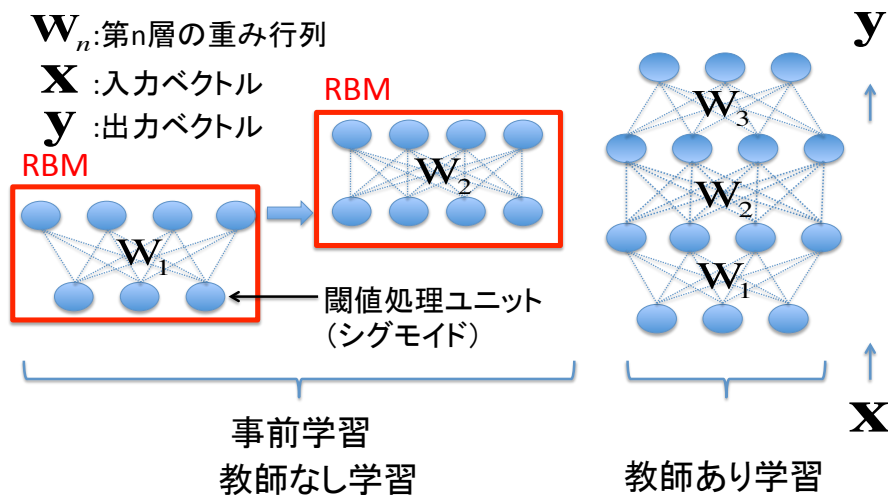


図 3.4 DBN の概略図

3.3 Deep Belief Network

3.3.1 RBM

RBM(Restricted Boltzmann Machine) とは, 本来の Boltzmann Machine の入力層, 隠れ層のユニット間の結合を制限することによって効率的な計算を可能とした確率的なニューラルネットワークである [9]. 図 3.5 に RBM の概略図を示す. RBM では入力ベクトルの層を可視層, 特徴ベクトルが抽出される層を隠れ層と呼ぶ. v_i は入力ユニット, h_j は隠れユニット, W_{ij} はユニット間の結合の重みである. 可視層のユニットのバイアスを b_i , 隠れ層のユニットのバイアスを c_j とする. 各ユニットは式 (3.1) のシグモイド関数によって出力値を算出する.

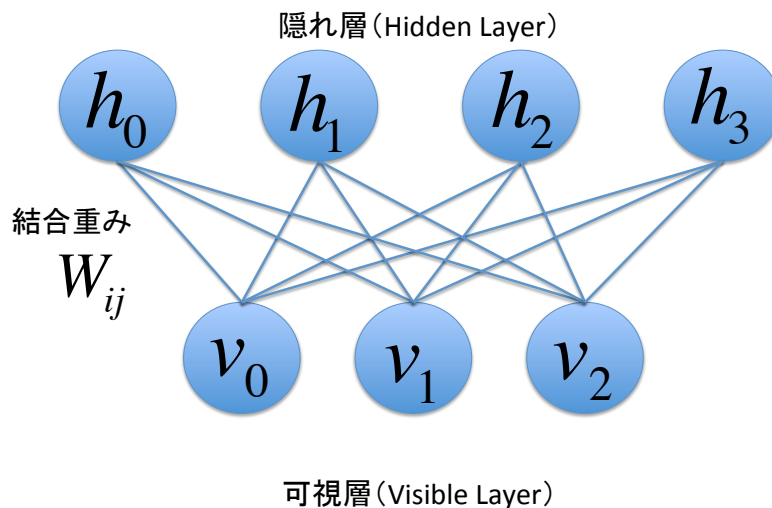


図 3.5 RBM の概略図

$$\text{sigmoid}(x) = \frac{1}{1 + \exp(x)} \quad (3.1)$$

RBM では式 (3.2), (3.3) の条件付き確率を用いる.

$$P(h_i = 1 | \mathbf{v}) = \text{sigmoid} \left(\sum_m^{j=1} \mathbf{W}_{ij} \mathbf{v}_j + \mathbf{c}_i \right) \quad (3.2)$$

3.3 Deep Belief Network

$$P(v_j = 1|\mathbf{h}) = \text{sigmoid}\left(\sum_n \mathbf{W}_{ij} \mathbf{h}_j + \mathbf{b}_j\right) \quad (3.3)$$

RBM の学習は対数尤度の最大化を行い，結合重みとバイアスを決定する．重みとバイアスは以下の更新式で行われる．

$$\mathbf{W} \leftarrow \mathbf{W} + \epsilon(\mathbf{h}^{(1)T} \mathbf{v} - P(\mathbf{h}^{k+1} = 1|\mathbf{v}^k)^T \mathbf{v}^k) \quad (3.4)$$

$$\mathbf{b} \leftarrow \mathbf{b} + \epsilon(\mathbf{v} - \mathbf{v}^{(k)}) \quad (3.5)$$

$$\mathbf{c} \leftarrow \mathbf{c} + \epsilon(\mathbf{h}^{(1)} - P(\mathbf{h}^{k+1} = 1|\mathbf{v}^{(k)})) \quad (3.6)$$

この時， ϵ は学習率， k はサンプリングの回数である．このように k 回のサンプリングで打ち切ることによって計算量を削減する手法を k-step Contrastive Divergence と呼び， k の値が 1 の場合であっても良い結果が得られることが確認されている [13].

3.3.2 誤差逆伝播法

誤差逆伝播法 (Back Propagation, BP) とは Rumelhart によって考案されたニューラルネットワークの学習手法である．BP では教師データと出力層の誤差を基に出力層側から入力層側に誤差を順に伝えていき，ネットワークの重みを更新する．BP では最急降下法に従って 2 乗誤差を最小化することによって学習を行う．教師データを t ，ユニットの出力を o とする．出力ユニットの個数を K 個とした時， n 番目の学習データに対する誤差評価関数は式 (3.7) となる．

$$E = \frac{1}{2} \sum_{k=1}^K (t_k^n - o_k^n)^2 \quad (3.7)$$

出力ユニットの値と教師データの値に近いほど E の値が 0 に近づく．そのため，BP では式 (3.7) の誤差評価関数が 0 に近づく重みを求めることによって学習を行う．重みの更新には最急降下法が用いられる．

3.4 交差確認法

BP による学習はネットワークに与えられた入力ベクトル x に対して教師データに近い出力ベクトル y となる識別関数を近似している．関数近似の方法の代表としてフーリエ級数展開があるが，BP とフーリエ級数展開の最大の違いは基底関数の性質が異なることである．フーリエ級数展開で用いられる基底関数は形状の変化が無い固定基底であるのに対して，BP で用いられる基底関数はパラメータによって形状を変化させることができる可変基底である．固定基底の場合，入力次元が大きくなると，近似に必要な基底関数の数が膨大になってしまう．しかし，可変基底の場合は基底関数の形状を変化させることでこの問題を回避している [14] ．

3.3.3 Dropout 法

Dropout 法とは 2012 年に Hinton によって考案された過学習を防ぐための手法である [15] ．ニューラルネットワークは，隠れ層の数が多くなると BP において誤差のパラメータに対する勾配が小さくなり，入力層側に近い層の重みの更新が十分に行われなくなってしまう．また，過学習に陥りやすいという問題もある．この問題を解決するための手法が Dropout 法である．図 3.6 に Dropout 法の概略を示す．Dropout 法では毎回ランダムに選んだユニットをネットワークから一旦取り除く．ユニットを取り除かれたネットワークは一時的にユニット間の依存関係が減るため，正則化の効果があり，汎化性能の向上を期待することができる．

3.4 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．汎化能力とは未知データに対する識別能力のことである．学習によって得られた識別モデルが学習データに含まれていない未知のデータに対しても有効であるかは分からない．過学習と呼ばれる，訓練データに対して過剰に適合した識別モデルになってしまう可能性もある．そのため，データを訓練データと検証用の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

3.4 交差確認法

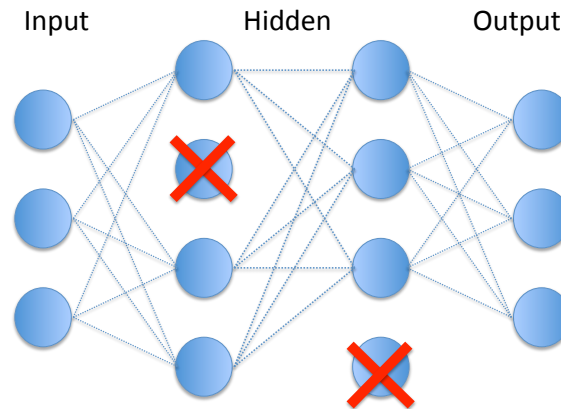


図 3.6 Dropout 法の概略図

能評価を行う必要がある．データを訓練データとテストデータに分割する代表的な手法にホールドアウト法がある．ホールドアウト法では手元のデータを 2 つに分割し，一方を訓練データ，もう一方をテストデータに使用する．そのため，訓練データを多くすればテストデータが少なくなり，性能評価が悪くなる．逆にテストデータを多くすれば訓練データが少なくなり，学習の精度が悪くなってしまふ．従って，データが大量に無ければ，十分な性能評価を行うことができないという欠点がある．ホールドアウト法の欠点を補う手法が交差確認法である [16]．交差確認法の概要を図 3.7 に示す．交差確認法ではデータを k 個に分割する． $k - 1$ 個のデータを訓練データとして学習を行い，残りの一個をテストデータとして評価を行う．これを k 回繰り返したときの平均で性能評価を行う．

3.4 交差確認法

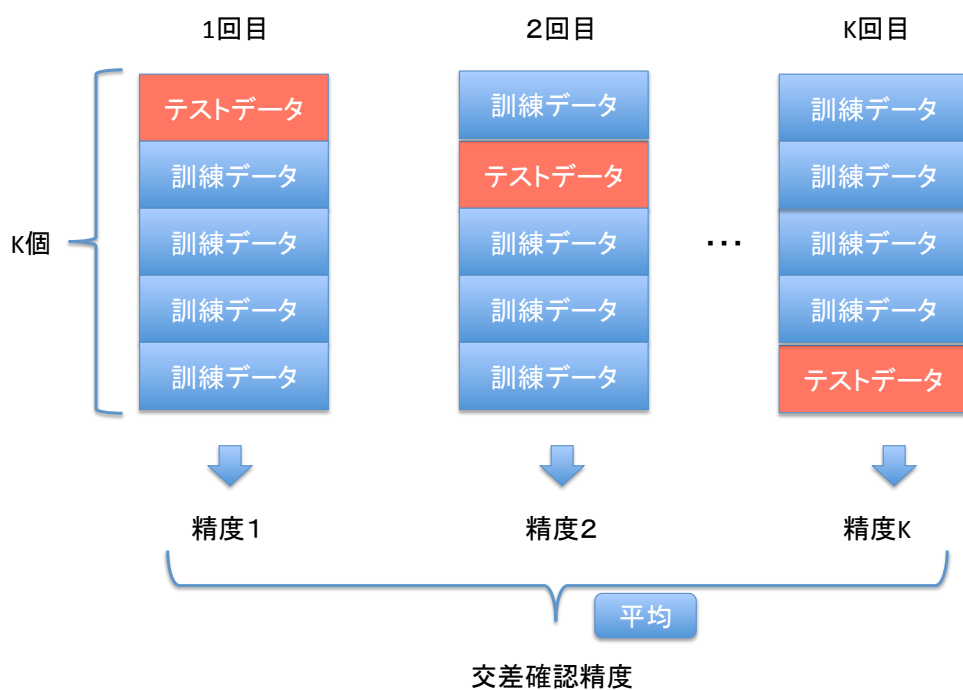


図 3.7 交差確認法

第 4 章

ニュース記事による予測実験

本章では学習用データの作成手順と実験手法についての説明を行う。

4.1 学習用データセットの作成

学習用データセットを以下の手順で作成する。

- 教師データの生成
- 単語ベクトルの生成
- 相関に基づく単語選出
- 学習データの標準化

4.1.1 教師データの生成

本研究では、当日の朝刊のニュース記事を用いて、その日の日経平均株価の値動きを予測する。株価の値動きは日経平均株価の終値が前日比+0.5%以上の増加を上昇、前日比-0.5%以上の現象を下降、その間を中間と定義し、3クラスの単純な予測を行う。表 4.1 に示すように教師データとして下降を-1、中間を 0、上昇を 1 とラベル化した。

4.1.2 単語ベクトルの生成

予測には経済記事テキストデータとして日本経済新聞の朝刊のニュース記事を使用する。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て、かつ日本経済

4.2 学習条件

表 4.1 教師データ

変動	ラベル
下降	-1
中間	0
上昇	1

新聞が休刊でない 234 日を対象とし、人事・訃報記事、スポーツ記事を除いた 64564 件の記事を扱う。これらの記事を形態素解析器 MeCab を用いて全てのテキストデータを単語に分割する。分割した単語の内、名詞、形容詞、動詞のみを取得する。この際、取得する単語は文面の文字列でなく原型を単語として取得する。全ての記事を単語に分割すると 54779 単語が取得でき、日付毎に全ての単語の出現回数を計測し 234 日分の単語ベクトルを生成する。生成した単語ベクトルを 234 行 54779 行の学習データとして CSV ファイルに保存した。

4.1.3 相関に基づく単語の選出

学習データセットは 54779 次元の高次元データとなるため、予測に効果的な単語を選出する必要がある。そのため、各単語出現回数と株価との相関係数を求める。算出した相関係数の絶対値の降順に単語ベクトルを並び替えた。

4.1.4 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を行う。標準化処理により各次元について平均が 0、分散が 1 となるようにデータを変換する。

4.2 学習条件

DBN と SVM の学習条件は以下の通りである。

4.3 性能評価

4.2.1 DBN

DBN の検証用のパラメータを表 4.2 に示す。DBN のユニット数の決め方には、明確な基準は無いため、隠れ層 1 層についてのユニット数は 100 個に設定し、3 層から 10 層までの間の検証を行う。learn-rate と epochs は Fine-tuning における学習係数と重みの更新回数、learn-rates-pretrain と epochs-pretrain は Pre-training における学習係数と重みの更新回数である。dropout は休止させるユニットの割合である。

表 4.2 検証用パラメータ

隠れ層	3~10 層
隠れユニット数	100
learn-rate	0.1
epochs	10, 100, 1000
learn-rates-pretrain	0.001, 0.0001
epochs-pretrain	100
dropout	0.1, 0.2, 0.3, 0.4, 0.5

4.2.2 SVM

SVM の学習には RBF(ガウス) カーネルと Linear(線形) カーネルの 2 種類を用いる。

4.3 性能評価

学習したモデルの評価には、汎化性能を検証するために k -分割交差検証を行う。 k の値は 5 に設定し、5 分割交差検証を行う。

4.4 実験

4.4 実験

学習手法として SVM(RBF カーネル, Linear カーネル), DBN の 3 つを用いて比較実験, 識別に用いる単語数を 1 から 3500 まで変化させつつ識別率と学習時間を計測する. 識別率は 5 分割交差確認法を用いて求める.

第 5 章

結果・考察

5.1 結果

サンプル数 234 のデータに対して識別に用いる単語数 (次元数) を 1 から 3500 個まで変化させた時の各手法の識別率の最大値とその時の学習時間および識別に用いた単語数を表 5.1 に示す。3 手法の中で DBN (4 層) の識別率が 97% で最も高い識別率を記録した。また、各種法の識別率をプロットしたグラフを図 5.1 に示す。図 5.1 の結果から SVM のカーネルは Linear カーネルよりも RBF カーネルの方が識別率が高いことが確認できる。DBN の精度は揺らぎが激しいが、単語数が 1000 から 3000 までの間では、RBF カーネルと比較しても同等かそれ以上の精度を記録している。それぞれの手法で単語数をさらに増やした場合の識別率も計測したが、精度が上昇することはなかった。

表 5.1 各手法の識別率 (\pm S.D)

学習手法	識別率	学習時間 [s]	単語数
SVM(線形)	0.88(\pm 0.03)	0.09	1763
SVM(RBF)	0.94(\pm 0.02)	0.21	3272
DBN(4 層)	0.97(\pm 0.01)	16.21	2588

各手法の単語数と学習時間を表 5.2 に示す。表 5.2 より SVM の Linear カーネルの学習時間が最も速いことが分かる。識別率が最も良かった DBN は、速度においては最も遅い結果となった。

学習に使用する単語数の数を固定し、DBN の層の数を 3 層から 10 層まで変化させた時

5.1 結果

表 5.2 各手法の単語数と学習時間 [s]

単語数	SVM(lineaar)	SVM(RBF)	DBN
500	0.02	0.03	8.08
1000	0.05	0.07	8.69
1500	0.08	0.10	11.32
2000	0.11	0.15	13.49
2500	0.13	0.17	15.71
3000	0.16	0.21	18.85
3500	0.19	0.25	20.68

の識別率の推移を図 5.2 に示す．4 層，5 層のときに識別率が 9 割を超えているが，層の数を増やしていくと，識別率が落ちていくことが確認できる．

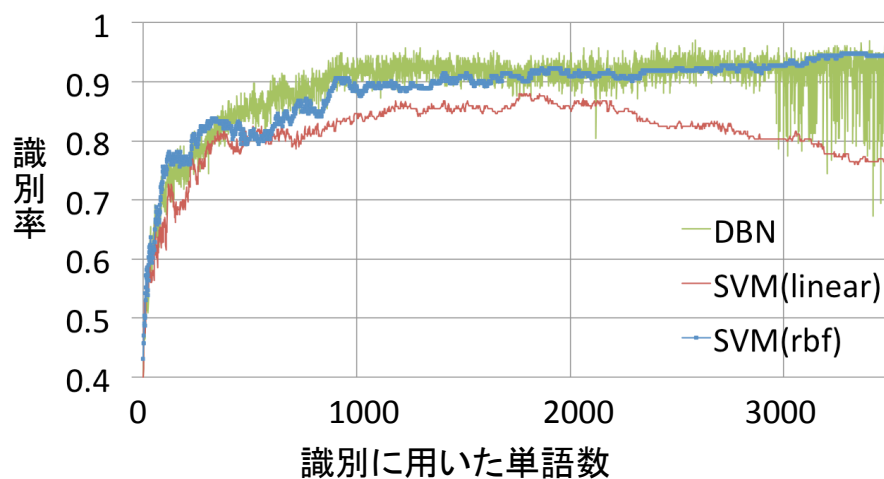


図 5.1 DBN の識別結果

5.2 考察

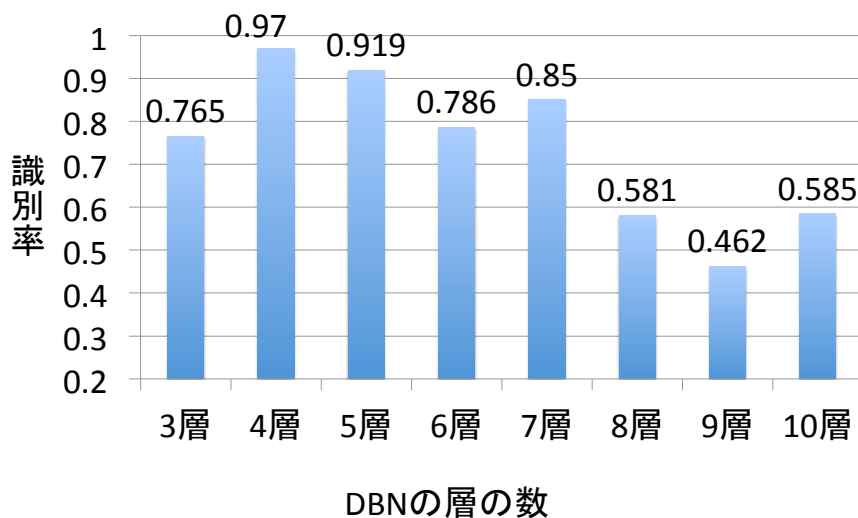


図 5.2 層の数と識別精度の推移

5.2 考察

結果として DBN が最も高い識別率となった。しかし学習にかかる時間が長く、識別率の揺らぎが大きいことが問題であることが分かった。識別率が揺らぐ理由としては、Fine-tuning で行っている誤差逆伝播法に原因があると考えられる。誤差逆伝播法では最急降下法に則って重みの更新を行う。最急降下法には局所解が存在し、最適解に辿り着けるかはネットワークの重みの初期値に依存するという特性がある。重みの初期値は乱数によって与えられるため、乱数の影響を受けて識別率が揺らいでいると考えられる。

識別に用いる単語数が増加すると、識別率の揺らぎが大きくなることの原因としては、次元の呪いが影響し、過学習を起こしている可能性が考えられる。今回の実験では、株価との相関係数の絶対値の高い順に単語の選出を行っているため、徐々に識別に有効でない単語が選出される。これらの単語が学習時のノイズになってしまい、過学習を起こしてしまっていると考えられる。また、識別に用いる単語数に対して学習データが少ないことも要因として考えられる。ノイズを含んだデータに対して過学習を防ぐには、Dropout させるユニット

5.2 考察

の割合の変更が考えられる。しかし，Dropout 法では識別に有効なユニットが休止状態になってしまう可能性がある。そのため，ユニットではなく，ユニット間の結合を休止させる Dropconnect 法 [17] についても検討する必要がある。

層の数を増やしたときに，精度が徐々に落ちる原因としては，パラメータの調整が十分でなかった可能性が考えられる。今回の実験では隠れ層 1 層につき，ユニットの数を 100 個に固定している。また，重みの更新回数についても Pre-training で 100 回，Fine-tuning で 1000 回までしか行っていない。これらのパラメータについても変更を加えて実験をすることが求められる。

今回の実験ではデータセット全てを対象にして教師データとの相関係数を求め，相関係数の絶対値の高い順に単語選出を行った。従って，評価用のデータも含めた相関係数の値に基づいて単語が選出されているため，本実験の識別率の結果を汎化性能の評価とすることには問題があると考えられる。また，本研究では，2014 年の新聞記事を基に単語の選出を行ったが，選出された単語を翌年の予測に適用した場合，精度が下がることが予想される。

5.2.1 追加実験

未知データに対する汎化能力の検証として 2014 年のデータセットから選出された単語を用いて 2015 年 1 月の株価の予測を行った。識別に用いる単語数を 1 から 1000 まで変化した時の最大の識別率とその時の学習時間および単語数を表 5.3 に各手法の識別率を示す。結果として DBN が最も高い識別精度となった。

表 5.3 各手法の識別率

学習手法	識別率	学習時間 [s]	単語数
SVM(Linear)	57.8%	0.007	95
SVM(RBF)	57.8%	0.004	17
DBN	63.1%	6.697	104

第 6 章

まとめ

結果として DBN は SVM より学習時間については劣るが、高い識別率を有していることが確認された。通常、株価予測においては、学習にかかる時間よりも、予測の精度の方が重視される場面が多いため、DBN は株価予測に有効であると結論づけられる。本研究では 2014 年の新聞記事を対象に単語の選出を行ったが、選出した単語を翌年の株価予測に適用した場合、識別精度が下がることも確認された。これは、選出単語を 2014 年の株価との相関に基づいて算出していることに起因している。その年の情勢によって掲載される記事の内容は大きく異なるため、2014 年の株価予測に有効な単語であっても、2015 年の記事に最悪の場合出現しないこともある。そのため、より識別の汎化性能を上げるには、2014 年より前の記事も対象にして予測に有効な単語の選出を行う必要がある。相関を用いた単語選出は、経済分野に依存しない方法であり、新聞記事だけでなく、ブログや SNS などのテキストデータへの適用も可能であると考えられる。

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、研究に行き詰まった際に相談に乗ってくださり、的確な助言をして頂きました。梗概や論文の執筆の際には、夜遅くまで残って添削をして頂きました。無事に論文の執筆を終えることができたのも先生のおかげだと思います。また、吉田先生からは、人に伝えるプレゼンの仕方など研究以外のことでも多くを学ばせて頂きました。今の自分の成長は吉田先生のご指導の賜物であると思います。重ねて御礼申し上げます。

また、副査を快く引き受けて頂いた、高知工科大学 情報学群 福本昌弘教授と高田喜朗准教授に深く感謝申し上げます。梗概の至らない部分を指摘して頂き、大変勉強になりました。両先生のご指摘のおかげで、より良い梗概を執筆することができました。

研究室の先輩方にも感謝申し上げます。修士の小池氏、藤森氏、松尾氏には、研究室での生活を送る上でのサポートをしてくださり、様々な面でお世話になりました。発表の練習やスライド作成の際に助言をして頂き、ありがとうございました。

同研究室の江木氏、奥山氏、高橋氏、竹中氏、中路氏、矢野氏、山本氏とは多くの時間を共に過ごし、皆さんから多くを学び、楽しい日々を送ることができたことに感謝します。江木氏からはよくお菓子を頂きました。研究で疲れている時や夜遅くまで残っている時の江木氏からの施しは、とてもありがたかったです。奥山氏とは1年次からの友人ということもあり、様々な場面でお世話になりました。ことあるごとに奥山氏の家に押しかけて鍋パーティを催したり、一緒にラーメンを食べに行ったり、楽しい時間を過ごすことができ、研究室生活のモチベーションを維持することができました。高橋氏とは趣味が近いということもあり、楽しい話をすることができました。学祭の時の高橋氏は輝いていたと思います。研究室のイベントの時には車に乗せて頂き、非常に助かりました。竹中氏はよく面白い話をしてくださり、研究室の皆を楽しませてくれました。また、皆が疑問に思っていることを率先して

謝辞

質問するなど、良い学習環境の構築に尽力していただいたおかげで、輪講の理解が深まったと思います。中路氏とは席が隣りということもあり、研究に関するだけでなく色々な話をすることができて、とても楽しかったです。中路氏の細やかな気配りや人当たりの良さのおかげで明るい雰囲気の研究室で過ごすことができたと思います。矢野氏には、よく励ましの言葉を掛けて貰いました。中間発表の時に緊張して弱音を吐く私の言葉を隣りで静かに聴いてくれて、励ましてくれたことを覚えています。山本氏の明るい性格には何度も元気づけられました。研究の合間の休憩している時などに山本氏を中心に話の輪が広がり、研究室一同で楽しい会話をすることができたと思います。

後輩の皆様には、研究室でのイベントの幹事などでお世話になりました。積極的に仕事をこなし、素直で明るい後輩のおかげで楽しい研究室生活を送ることが出来ました。

そして、遠くはなれた高知県の大学へ進学することを許して下さり、大学生活4年間で温かく見守り、生活を支えてくださった両親にも心から感謝いたします。

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] 中田 貴之, “対象の値動きの方向性に着目した経済時系列予測へのサポートベクターマシンの応用に関する研究,” 東京大学, 2008 年度修士論文.
- [2] 山口 和孝, “ニューラルネットと遺伝的アルゴリズムを用いた株式売買支援システム,” 東京大学院, 2002 年度修士論文
- [3] 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.
- [4] 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.
- [5] Deng, Li, et al., “Recent advances in deep learning for speech research at Microsoft,” Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp.8604-8608 , 2013.
- [6] Le, Quoc V, “Building high-level features using large scale unsupervised learning.” Proceedings of IEEE International Conference on Acoustics,Speech and Signal Processing (ICASSP), pp.8595-8598, 2013.
- [7] Chao, Jing, Furao Shen, and Jinxi Zhao, “Forecasting exchange rate with deep belief networks,” Proceedings of IEEE International Joint Conference on Neural Networks (IJCNN) , pp.1259-1266, 2011.
- [8] Bengio, Y, “Practical Recommendations for GradientBased Training of Deep Architecture,” arXiv:1206.5533, 2012.
- [9] Bengio, Y, “Learning deep architectures for AI,” Foundations and trends in Machine Learning 2.1, pp1-127, 2009.
- [10] Vladimir Naumovich, and Vlamimir Vapnik, “Statistical learning theory,” Vol. 1. New York: Wiley, 1998.

参考文献

- [11] 熊沢逸夫, “学習とニューラルネットワーク,” 森北出版株式会社 (1996)
- [12] Hinton, G. , Osindero, S. and Teh, Y. W, “ A fast learning algorithm for deep belief nets,” Neural Computation, Vol. 18,No. 7 , pp.1527-1554, 2006.
- [13] Tieleman, Tijmen, “ Training restricted Boltzmann machines using approximations to the likelihood gradient,” Proceedings of the 25th international conference on Machine learning, ACM, pp.1064-1071, 2008.
- [14] 萩原克幸, “ニューラルネットワークの基礎と理論的に重要な課題 (ニューラルネットワークの応用と今後の発展),” プラズマ・核融合学会誌 82(5), pp.282-286, 2006.
- [15] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton, “ Imagenet classification with deep convolutional neural networks,” Advances in neural information processing systems, pp.1097-1105, 2012.
- [16] 平井有三, “はじめてのパターン認識,” 森北出版株式会社, 2012.
- [17] Wan, Li, et al., “Regularization of neural networks using dropconnect,” Proceedings of the 30th International Conference on Machine Learning (ICML-13), pp.1058-1066, 2013.

付録 A

グリッドサーチ

グリッドサーチによるパラメータと精度の変化を以下に示す。

表 A.1: グリッドサーチによるパラメータと精度の変化

Accuracy	epochs	dropouts	layer_sizes
0.799	10	0.1	[2588, 100, 3]
0.795	10	0.2	[2588, 100, 3]
0.838	10	0.3	[2588, 100, 3]
0.782	10	0.4	[2588, 100, 3]
0.799	10	0.5	[2588, 100, 3]
0.795	10	0.6	[2588, 100, 3]
0.816	10	0.7	[2588, 100, 3]
0.752	10	0.8	[2588, 100, 3]
0.825	10	0.9	[2588, 100, 3]
0.859	10	0.1	[2588, 100, 100, 3]
0.88	10	0.2	[2588, 100, 100, 3]
0.893	10	0.3	[2588, 100, 100, 3]
0.94	10	0.4	[2588, 100, 100, 3]
0.889	10	0.5	[2588, 100, 100, 3]
0.85	10	0.6	[2588, 100, 100, 3]

表 A.1: グリッドサーチによるパラメータと精度の変化

Accuracy	epochs	dropouts	layer_sizes
0.88	10	0.7	[2588, 100, 100, 3]
0.936	10	0.8	[2588, 100, 100, 3]
0.915	10	0.9	[2588, 100, 100, 3]
0.735	10	0.1	[2588, 100, 100, 100, 3]
0.808	10	0.2	[2588, 100, 100, 100, 3]
0.795	10	0.3	[2588, 100, 100, 100, 3]
0.829	10	0.4	[2588, 100, 100, 100, 3]
0.838	10	0.5	[2588, 100, 100, 100, 3]
0.803	10	0.6	[2588, 100, 100, 100, 3]
0.816	10	0.7	[2588, 100, 100, 100, 3]
0.692	10	0.8	[2588, 100, 100, 100, 3]
0.744	10	0.9	[2588, 100, 100, 100, 3]
0.808	10	0.1	[2588, 100, 100, 100, 100, 3]
0.692	10	0.2	[2588, 100, 100, 100, 100, 3]
0.692	10	0.3	[2588, 100, 100, 100, 100, 3]
0.722	10	0.4	[2588, 100, 100, 100, 100, 3]
0.778	10	0.5	[2588, 100, 100, 100, 100, 3]
0.765	10	0.6	[2588, 100, 100, 100, 100, 3]
0.624	10	0.7	[2588, 100, 100, 100, 100, 3]
0.744	10	0.8	[2588, 100, 100, 100, 100, 3]
0.607	10	0.9	[2588, 100, 100, 100, 100, 3]
0.722	10	0.1	[2588, 100, 100, 100, 100, 100, 3]

表 A.1: グリッドサーチによるパラメータと精度の変化

Accuracy	epochs	dropouts	layer_sizes
0.56	10	0.2	[2588, 100, 100, 100, 100, 100, 3]
0.594	10	0.3	[2588, 100, 100, 100, 100, 100, 3]
0.598	10	0.4	[2588, 100, 100, 100, 100, 100, 3]
0.594	10	0.5	[2588, 100, 100, 100, 100, 100, 3]
0.568	10	0.6	[2588, 100, 100, 100, 100, 100, 3]
0.594	10	0.7	[2588, 100, 100, 100, 100, 100, 3]
0.628	10	0.8	[2588, 100, 100, 100, 100, 100, 3]
0.603	10	0.9	[2588, 100, 100, 100, 100, 100, 3]
0.585	10	0.1	[2588, 100, 100, 100, 100, 100, 100, 3]
0.423	10	0.2	[2588, 100, 100, 100, 100, 100, 100, 3]
0.483	10	0.3	[2588, 100, 100, 100, 100, 100, 100, 3]
0.573	10	0.4	[2588, 100, 100, 100, 100, 100, 100, 3]
0.509	10	0.5	[2588, 100, 100, 100, 100, 100, 100, 3]
0.517	10	0.6	[2588, 100, 100, 100, 100, 100, 100, 3]
0.368	10	0.7	[2588, 100, 100, 100, 100, 100, 100, 3]
0.423	10	0.8	[2588, 100, 100, 100, 100, 100, 100, 3]
0.496	10	0.9	[2588, 100, 100, 100, 100, 100, 100, 3]
0.47	10	0.1	[2588, 100, 100, 100, 100, 100, 100, 100, 3]
0.453	10	0.2	[2588, 100, 100, 100, 100, 100, 100, 100, 3]
0.47	10	0.3	[2588, 100, 100, 100, 100, 100, 100, 100, 3]
0.372	10	0.4	[2588, 100, 100, 100, 100, 100, 100, 100, 3]
0.457	10	0.5	[2588, 100, 100, 100, 100, 100, 100, 100, 3]

表 A.1: グリッドサーチによるパラメータと精度の変化

Accuracy	epochs	dropouts	layer_sizes
0.393	10	0.6	[2588, 100, 100, 100, 100, 100, 100, 100, 3]
0.419	10	0.7	[2588, 100, 100, 100, 100, 100, 100, 100, 3]
0.368	10	0.8	[2588, 100, 100, 100, 100, 100, 100, 100, 3]
0.53	10	0.9	[2588, 100, 100, 100, 100, 100, 100, 100, 3]
0.389	10	0.1	[2588, 100, 100, 100, 100, 100, 100, 100, 3]
0.393	10	0.2	[2588, 100, 100, 100, 100, 100, 100, 100, 3]
0.449	10	0.3	[2588, 100, 100, 100, 100, 100, 100, 100, 3]
0.444	10	0.4	[2588, 100, 100, 100, 100, 100, 100, 100, 3]
0.423	10	0.5	[2588, 100, 100, 100, 100, 100, 100, 100, 3]
0.474	10	0.6	[2588, 100, 100, 100, 100, 100, 100, 100, 3]
0.436	10	0.7	[2588, 100, 100, 100, 100, 100, 100, 100, 3]
0.462	10	0.8	[2588, 100, 100, 100, 100, 100, 100, 100, 3]
0.406	10	0.9	[2588, 100, 100, 100, 100, 100, 100, 100, 3]
0.774	100	0.1	[2588, 100, 3]
0.778	100	0.2	[2588, 100, 3]
0.791	100	0.3	[2588, 100, 3]
0.838	100	0.4	[2588, 100, 3]
0.825	100	0.5	[2588, 100, 3]
0.739	100	0.6	[2588, 100, 3]
0.803	100	0.7	[2588, 100, 3]
0.791	100	0.8	[2588, 100, 3]
0.846	100	0.9	[2588, 100, 3]

表 A.1: グリッドサーチによるパラメータと精度の変化

Accuracy	epochs	dropouts	layer_sizes
0.88	100	0.1	[2588, 100, 100, 3]
0.97	100	0.2	[2588, 100, 100, 3]
0.859	100	0.3	[2588, 100, 100, 3]
0.88	100	0.4	[2588, 100, 100, 3]
0.915	100	0.5	[2588, 100, 100, 3]
0.855	100	0.6	[2588, 100, 100, 3]
0.88	100	0.7	[2588, 100, 100, 3]
0.868	100	0.8	[2588, 100, 100, 3]
0.906	100	0.9	[2588, 100, 100, 3]
0.889	100	0.1	[2588, 100, 100, 100, 3]
0.769	100	0.2	[2588, 100, 100, 100, 3]
0.816	100	0.3	[2588, 100, 100, 100, 3]
0.846	100	0.4	[2588, 100, 100, 100, 3]
0.825	100	0.5	[2588, 100, 100, 100, 3]
0.795	100	0.6	[2588, 100, 100, 100, 3]
0.855	100	0.7	[2588, 100, 100, 100, 3]
0.855	100	0.8	[2588, 100, 100, 100, 3]
0.923	100	0.9	[2588, 100, 100, 100, 3]
0.581	100	0.1	[2588, 100, 100, 100, 100, 3]
0.799	100	0.2	[2588, 100, 100, 100, 100, 3]
0.859	100	0.3	[2588, 100, 100, 100, 100, 3]
0.598	100	0.4	[2588, 100, 100, 100, 100, 3]

表 A.1: グリッドサーチによるパラメータと精度の変化

Accuracy	epochs	dropouts	layer_sizes
0.825	100	0.5	[2588, 100, 100, 100, 100, 3]
0.662	100	0.6	[2588, 100, 100, 100, 100, 3]
0.778	100	0.7	[2588, 100, 100, 100, 100, 3]
0.897	100	0.8	[2588, 100, 100, 100, 100, 3]
0.658	100	0.9	[2588, 100, 100, 100, 100, 3]
0.577	100	0.1	[2588, 100, 100, 100, 100, 100, 3]
0.658	100	0.2	[2588, 100, 100, 100, 100, 100, 3]
0.662	100	0.3	[2588, 100, 100, 100, 100, 100, 3]
0.726	100	0.4	[2588, 100, 100, 100, 100, 100, 3]
0.577	100	0.5	[2588, 100, 100, 100, 100, 100, 3]
0.615	100	0.6	[2588, 100, 100, 100, 100, 100, 3]
0.838	100	0.7	[2588, 100, 100, 100, 100, 100, 3]
0.538	100	0.8	[2588, 100, 100, 100, 100, 100, 3]
0.722	100	0.9	[2588, 100, 100, 100, 100, 100, 3]
0.581	100	0.1	[2588, 100, 100, 100, 100, 100, 100, 3]
0.594	100	0.2	[2588, 100, 100, 100, 100, 100, 100, 3]
0.491	100	0.3	[2588, 100, 100, 100, 100, 100, 100, 3]
0.538	100	0.4	[2588, 100, 100, 100, 100, 100, 100, 3]
0.47	100	0.5	[2588, 100, 100, 100, 100, 100, 100, 3]
0.675	100	0.6	[2588, 100, 100, 100, 100, 100, 100, 3]
0.615	100	0.7	[2588, 100, 100, 100, 100, 100, 100, 3]
0.534	100	0.8	[2588, 100, 100, 100, 100, 100, 100, 3]

表 A.1: グリッドサーチによるパラメータと精度の変化

Accuracy	epochs	dropouts	layer_sizes
0.526	100	0.9	[2588, 100, 100, 100, 100, 100, 100, 3]
0.667	100	0.1	[2588, 100, 100, 100, 100, 100, 100, 3]
0.577	100	0.2	[2588, 100, 100, 100, 100, 100, 100, 3]
0.551	100	0.3	[2588, 100, 100, 100, 100, 100, 100, 3]
0.474	100	0.4	[2588, 100, 100, 100, 100, 100, 100, 3]
0.509	100	0.5	[2588, 100, 100, 100, 100, 100, 100, 3]
0.547	100	0.6	[2588, 100, 100, 100, 100, 100, 100, 3]
0.432	100	0.7	[2588, 100, 100, 100, 100, 100, 100, 3]
0.363	100	0.8	[2588, 100, 100, 100, 100, 100, 100, 3]
0.342	100	0.9	[2588, 100, 100, 100, 100, 100, 100, 3]
0.393	100	0.1	[2588, 100, 100, 100, 100, 100, 100, 100, 3]
0.402	100	0.2	[2588, 100, 100, 100, 100, 100, 100, 100, 3]
0.419	100	0.3	[2588, 100, 100, 100, 100, 100, 100, 100, 3]
0.419	100	0.4	[2588, 100, 100, 100, 100, 100, 100, 100, 3]
0.44	100	0.5	[2588, 100, 100, 100, 100, 100, 100, 100, 3]
0.329	100	0.6	[2588, 100, 100, 100, 100, 100, 100, 100, 3]
0.368	100	0.7	[2588, 100, 100, 100, 100, 100, 100, 100, 3]
0.496	100	0.8	[2588, 100, 100, 100, 100, 100, 100, 100, 3]
0.393	100	0.9	[2588, 100, 100, 100, 100, 100, 100, 100, 3]
0.821	1000	0.1	[2588, 100, 3]
0.786	1000	0.2	[2588, 100, 3]
0.774	1000	0.3	[2588, 100, 3]

表 A.1: グリッドサーチによるパラメータと精度の変化

Accuracy	epochs	dropouts	layer_sizes
0.821	1000	0.4	[2588, 100, 3]
0.846	1000	0.5	[2588, 100, 3]
0.803	1000	0.6	[2588, 100, 3]
0.863	1000	0.7	[2588, 100, 3]
0.808	1000	0.8	[2588, 100, 3]
0.791	1000	0.9	[2588, 100, 3]
0.808	1000	0.1	[2588, 100, 100, 3]
0.876	1000	0.2	[2588, 100, 100, 3]
0.897	1000	0.3	[2588, 100, 100, 3]
0.893	1000	0.4	[2588, 100, 100, 3]
0.906	1000	0.5	[2588, 100, 100, 3]
0.897	1000	0.6	[2588, 100, 100, 3]
0.893	1000	0.7	[2588, 100, 100, 3]
0.846	1000	0.8	[2588, 100, 100, 3]
0.833	1000	0.9	[2588, 100, 100, 3]
0.859	1000	0.1	[2588, 100, 100, 100, 3]
0.923	1000	0.2	[2588, 100, 100, 100, 3]
0.855	1000	0.3	[2588, 100, 100, 100, 3]
0.769	1000	0.4	[2588, 100, 100, 100, 3]
0.889	1000	0.5	[2588, 100, 100, 100, 3]
0.825	1000	0.6	[2588, 100, 100, 100, 3]
0.85	1000	0.7	[2588, 100, 100, 100, 3]

表 A.1: グリッドサーチによるパラメータと精度の変化

Accuracy	epochs	dropouts	layer_sizes
0.919	1000	0.8	[2588, 100, 100, 100, 3]
0.816	1000	0.9	[2588, 100, 100, 100, 3]
0.774	1000	0.1	[2588, 100, 100, 100, 100, 3]
0.791	1000	0.2	[2588, 100, 100, 100, 100, 3]
0.863	1000	0.3	[2588, 100, 100, 100, 100, 3]
0.915	1000	0.4	[2588, 100, 100, 100, 100, 3]
0.748	1000	0.5	[2588, 100, 100, 100, 100, 3]
0.765	1000	0.6	[2588, 100, 100, 100, 100, 3]
0.654	1000	0.7	[2588, 100, 100, 100, 100, 3]
0.714	1000	0.8	[2588, 100, 100, 100, 100, 3]
0.735	1000	0.9	[2588, 100, 100, 100, 100, 3]
0.402	1000	0.1	[2588, 100, 100, 100, 100, 100, 3]
0.697	1000	0.2	[2588, 100, 100, 100, 100, 100, 3]
0.564	1000	0.3	[2588, 100, 100, 100, 100, 100, 3]
0.59	1000	0.4	[2588, 100, 100, 100, 100, 100, 3]
0.744	1000	0.5	[2588, 100, 100, 100, 100, 100, 3]
0.697	1000	0.6	[2588, 100, 100, 100, 100, 100, 3]
0.722	1000	0.7	[2588, 100, 100, 100, 100, 100, 3]
0.594	1000	0.8	[2588, 100, 100, 100, 100, 100, 3]
0.547	1000	0.9	[2588, 100, 100, 100, 100, 100, 3]
0.483	1000	0.1	[2588, 100, 100, 100, 100, 100, 100, 3]
0.577	1000	0.2	[2588, 100, 100, 100, 100, 100, 100, 3]

表 A.1: グリッドサーチによるパラメータと精度の変化

Accuracy	epochs	dropouts	layer_sizes
0.803	1000	0.3	[2588, 100, 100, 100, 100, 100, 100, 3]
0.423	1000	0.4	[2588, 100, 100, 100, 100, 100, 100, 3]
0.419	1000	0.5	[2588, 100, 100, 100, 100, 100, 100, 3]
0.598	1000	0.6	[2588, 100, 100, 100, 100, 100, 100, 3]
0.466	1000	0.7	[2588, 100, 100, 100, 100, 100, 100, 3]
0.372	1000	0.8	[2588, 100, 100, 100, 100, 100, 100, 3]
0.359	1000	0.9	[2588, 100, 100, 100, 100, 100, 100, 3]
0.483	1000	0.1	[2588, 100, 100, 100, 100, 100, 100, 3]
0.342	1000	0.2	[2588, 100, 100, 100, 100, 100, 100, 3]
0.329	1000	0.3	[2588, 100, 100, 100, 100, 100, 100, 3]
0.415	1000	0.4	[2588, 100, 100, 100, 100, 100, 100, 3]
0.483	1000	0.5	[2588, 100, 100, 100, 100, 100, 100, 3]
0.389	1000	0.6	[2588, 100, 100, 100, 100, 100, 100, 3]
0.479	1000	0.7	[2588, 100, 100, 100, 100, 100, 100, 3]
0.376	1000	0.8	[2588, 100, 100, 100, 100, 100, 100, 3]
0.415	1000	0.9	[2588, 100, 100, 100, 100, 100, 100, 3]
0.432	1000	0.1	[2588, 100, 100, 100, 100, 100, 100, 3]
0.329	1000	0.2	[2588, 100, 100, 100, 100, 100, 100, 3]
0.53	1000	0.3	[2588, 100, 100, 100, 100, 100, 100, 3]
0.406	1000	0.4	[2588, 100, 100, 100, 100, 100, 100, 3]
0.483	1000	0.5	[2588, 100, 100, 100, 100, 100, 100, 3]
0.393	1000	0.6	[2588, 100, 100, 100, 100, 100, 100, 3]

表 A.1: グリッドサーチによるパラメータと精度の変化

Accuracy	epochs	dropouts	layer_sizes
0.363	1000	0.7	[2588, 100, 100, 100, 100, 100, 100, 100, 100, 3]
0.474	1000	0.8	[2588, 100, 100, 100, 100, 100, 100, 100, 100, 3]
0.419	1000	0.9	[2588, 100, 100, 100, 100, 100, 100, 100, 100, 3]

付録 B

選出単語

壮観に基づく単語選出によって選出された単語 (上位 1000 個) を以下に示す。

表 B.1 選出単語一覧

使用	画質	コンサルタ ント	ホット	並	ちりばめる	済み	学歴
盛り	宮田	課長	境遇	保つ	強める	占拠	さらす
協和	造反	設備	退去	減産	魅せる	余暇	収れん
太刀打ち	安原	振り返る	いたす	鵬	分社	楽しみ	遊園
札幌	甲信越	うえ	孝行	換算	再現	ドミニク	貧しい
深沢	海面	参与	柘植	丸い	缶詰	刺す	魔よけ
建て替える	L T	つきまとう	俊夫	外す	お祝い	立て直せる	飛び込み
賄賂	鳴き声	修三	間際	患う	新聞	号機	目新しい
凍結	若い	引き渡す	リズム	かたどる	アジア	女流	近鉄百貨 店
差し引き	八重洲ブッ クセンター	オーシャン	大きい	枚数	懇親	檜原	雁
実父	しょく	げいこ	遠慮ない	車中	姿勢	仕向け	彩色
最低限	チャンネル	もみじ	憤慨	有機物	就職	呼び出す	井
実装	おふくろ	レート	年初	トライ	巻き返す	越	標的
稲野	指摘	勇	担い手	基軸	衰退	力国	作成
狩野	動態	はやい	先生	河南	お待ち	長野	三井
刊	方式	智絵	尺度	芸妓	北海道大	さぼる	原島
北青山	カ年	盗み出す	高品	プレゼンス	滞る	広東	ほか
西方	カナダ	久門	酔	革命	乗り継ぐ	無謀	帝政
ドライバー	新鮮	アーカイブ	子孫	たか	徳之島	登山	綱渡り
チャンピオ ンズ	アンバラン ス	マイナス	社会党	僅差	そごう	見方	踏み込む
これら	尻目	短	出来	貴文	愛知製鋼	千駄ヶ谷	けやき坂
グラフィカ ル	隆平	日東	歓声	開閉	披露宴	決選	東武スト ア
秋口	清家	おん	内	意中	琢也	アイスラン ド	本
松屋フーズ	たどれる	陰陽	外回り	新参	トータル	浩二	ナンバー
大豊	抜け落ちる	トレー	完璧	曇らせる	攻撃	ボーダーレ ス	地層

おなじみ	頂く	ゆかり	未踏	信奉	由香	山地	脆弱
清潔	F G	インフラ	有名	建てる	エルニーニョ	ファンド	触れ込み
祝福	する	適否	心臓	シネマ	末尾	隊	神体
ダーバン	手痛い	取り立て	択捉島	中浜	洗髪	水洗	八角
放り込む	そっけ	しんぼ	隆太	安打	容姿	振付	党紀
中村	商事	和久	気質	地裁	はまる	本人	イメージ
高価	肉類	通信	バーバラ	孟	政則	参道	粗末
変心	堀江	セリ	乗り切れる	ブラック	庄司	英断	紋
ふじい	新進	うなずける	合点	かれる	煙突	強者	割引
物体	請求	求める	課	官房	主将	国保	テルアビブ
交易	当地	エッセンス	メンツ	パークレー	人肌	市庁舎	水田
矢口	石室	各所	気体	発想	武川	ニス	海風
アトピー	淡い	ステンレス	見過ごす	ジェブ	応え	飛び立つ	返納
ばね	あい	頭文字	美女	ダルトン	累計	サロン	秩序
改廃	取れる	原子	のぞみ	別人	最高検	長靴	検事
養う	どっち	セネガル	惑う	アブドラ	まき	覇	リハビリ
マザー	壁画	観艦式	減衰	依存	跳ね上がる	採取	無駄遣い
フリー	執筆	版	間口	アイルランド	皆さん	毎年	内定
貼り紙	鷲	熱処理	カビ	隠し	言い分	当たり前	当分
書道	松永	浅い	自習	水産庁	燃料	あべの	落札
水平	マイホーム	並べる	住友	暗闇	グリーンズ	活発	空き
音楽	症	誤報	支柱	言明	介す	名古屋大学	指揮
架ける	おこす	見聞き	明く	日本経済新聞社	前橋	代償	銘じる
停滞	見送り	ラード	J R東日本	物質	香月	封鎖	移り変わる
なじめる	日比野	頂上	蘇州	ハイテク	自公民	よみ	首尾

しかめる	コンベンション	美的	焦げつく	クジラ	復活	四捨五入	賃
建株	ワンマン	速まる	宏之	喜び	われわれ	景気動向指数	辺り
お話し	年初来	薬師	断念	翌年	買い付ける	ゼンショー	キロ
幹線	メルシャン	勝ち抜く	ため	八千代銀行	有珠山	不祥事	杉並
大学	政財界	開場	道中	和菓子	双竜	プール	小川
足首	廃れる	祐二	祥司	出没	戻せる	胃酸	わいしょう
交ざる	矮小	羨望	竜一	来社	ペンション	綴	楠本
雷門	張り付ける	治彦	言い残す	輝夫	常々	華道	後ろ向き
中近東	聡	あっせん	汚れ	くださる	参事官	結核	喜光
猛	繊細	主眼	隣る	響かせる	窮状	決議	リマ
発行	上京	焼き物	ヒアリング	逆襲	づらい	ガラス	川
弁当	年月	中2	フセイン	自社	一時	国家	あしらう
高層	新店	忠志	勝浦	隆明	勸	構成	スロープ
校	繰る	骨抜き	格納	伊丹	紀伊国屋	巡る	払拭
乗務	光一	楽団	政男	はん	増長	豪雪	対決
将軍家	無線	ナショナル	売り	併せる	行事	支出	サーベイ
搬入	貧困	旗艦	点	出願	主席	図書	手っ取り早い
東京電力	テラス	じん	俊之	那須塩原	決めつける	ドバイ	素っ気
隣県	移り変わり	カモ	見逃せる	隆司	もらえる	対極	暗転
ハンド	書く	国名	貢献	P G	日本ケンタッキー・フライド・チキン	朝鮮中央通信	開花
捷	集落	並び	パース	すりつぶす	フリ	コーディネット	間違う
銀色	健闘	めし	上村	倉本	弘文	長谷部	駆使
読解	混入	刃物	月曜日	駅伝	建屋	旅行	塗
人目	住友信託銀行	くら替え	武史	温存	伯母	美智雄	悟り
息遣い	坊や	寝返る	内張り	くい止める	修次	チェックポイント	よしの

クォー リー	タ	朝日大	ひろまる	てきめん	さいし	斉田	新湊	達男
東端		東京外国語 大	金沢工業大	押出	ふじみ野	甲高い	設け	博喜
善文		天満	羽入	拘留	八丈島	エフ・ディ・ シィ・プロダ クツ	ひめる	邦昭
正朗		はげ	暴騰	邦題	亡国	危なげない	備え付け	長子
夕闇		瓜	経典	卓哉	さとり	書院	譲り合う	東札幌
社葬		紆余曲折	中層	敬吾	手近	成人の日	製氷	大発会
フローリン グ		西田	番付	補強	食い違い	強豪	廃校	ジャンル
エイ		アタック	近付く	めかす	もつれこむ	オリ	あけぼの	浄
でっち上げ		めくれる	区民	益雄	核燃	不可分	偶数	幻滅
智恵		見開き	あかね	跡形	夕日	たたずむ	書き込める	煙たがる
森内		NEWS	思想	応募	インスタ ト	習い事	高揚	真保
災		足元	快適	出頭	しゅう	送受信	果実	精力
ユニセフ		トランク	クレーン	半日	塩素	くぎ	茅野	事典
一真		知至	芥田	襲来	港北	定職	とりざたす	比喩
利息		塾長	使いこなせ る	完走	内幸町	所帯	リフト	クアラルン プール
麻奈		押しつぶす	せり	四半世紀	媛	師傅	ゆるい	木野内
関門海峡		楽園	能美	浅草	アクリル	京都大学	同派	鎖
仏滅		兄弟	鉦区	モール	人民大会堂	塩	スマ	ウエスコ
令状		モーグル	冷やす	応援	膝	メインバン ク	官邸	取り戻せる
見受ける		動かせる	利回り	河合	脳裏	働きかけ	なれる	害する
にぎやか		Web	色彩	見捨てる	真央	コンテナ	妥協	打者
シンボル		電光	彩	賠償	東洋紡	付属	ページ	関西学院大
麻原		宝飾	野神	国際線	山木	祝杯	巻頭	クリ
税別		昭和女子大	アドリブ	低空	ポピュラー	轟音	紀彦	そろい踏み
フュージョ ン		米村	尾畑	状態	境内	珈琲	揺るぐ	生かせる

める	精子	乱調	埋める	メンテナン ス	名城大	コカ・コー ラ	経費
武美	メンタル	はずれる	三菱	偉い	ワイヤ	ファ	希
帝国データ バンク	国史	柏木	喜べる	旦那	上関原発	十六銀行	付け方
のり子	映り	育める	徳丸	火気	暴れ	はじき返す	燃やせる
せわ	全数	兼備	直系	廃品	重病	ふとい	メラニン
紀尾井	みちお	やぶさか	青天井	目張り	あつみ	禁城	万葉
金納	樺太	泰成	信重	長常	枕草子	撃ち合う	怠ける
交点	石飛	行き帰り	片りん	ナース	すそ野	九谷焼	御幸
併任	嶽	直進	金髪	ベネトン	公国	校風	黒み
立ち姿	売り出せる	用材	興ざめる	おかあさん	理想郷	無色	し尿
薄手	表側	慈愛	早川書房	豊雄	勝ち負け	光敏	パーティ
でんぷん	ホ	パラマウン トベッド	栃木銀行	喉頭	禁煙	エイト	連携
千客万来	ひろみ	うめく	撮像	病死	興銀	陸運	上回る
今どき	付き合い	東北大学	不便	橋	克彦	勝る	グルメ
赤松	橋口	能美防災	政春	サインペン	窓辺	きよみ	シリウス
裕司	けんじる	日本経済新 聞	カシオ計算 機	身	倍	京子	作者
館	衛	長ける	胆力	三唱	俵	腫れる	小栗
みつる	ダグラス	美德	ディン	打ち寄せる	マークシー ト	高階	プラモデル
空転	勸業	ぬかるむ	申しあげる	風圧	花街	パワー	雅典
船団	理系	パイオニア	和成	紙巻き	暴落	白黒	たたき込む
流す	一生懸命	トレーラー	シアター	コーセー	動かす	量	放射線
史跡	圧巻	演技	注水	阿南	牙城	外	持ち
広げる	鐘	帰京	引き締まる	久米島	ウォール街	かぎ	パイプライ ン
いただく	打つ	鍋	中値	自慢	省内	理不尽	館内
朝刊	筆	買い戻す	坪井	かさ	下がる	伊東	小出し