

深層学習による経済指標動向推定

Estimating the Trend of Economic Indicators by Deep Learning

吉原 輝*¹ 藤川 和樹*² 関 和広*² 上原 邦昭*²
Akira Yoshihara Kazuki Fujikawa Kazuhiro Seki Kuniaki Uehara

*¹神戸大学工学部情報知能工学科

Department of Computer Science and Systems Engineering, Kobe University

*²神戸大学大学院システム情報学研究科

Graduate School of System Informatics, Kobe University

Investors make decisions based on various information sources, including consumer price index, price-earnings ratio, and miscellaneous events reported by news wires. To assist their decisions, many studies have been conducted in the last decades. However, the majority of the efforts was made for utilizing numerical information, partly due to the difficulty to analyze natural language texts and to make sense of their temporal properties. This study sheds light on this problem using deep learning, which has been attracting much attention in the pattern mining and machine learning communities for its ability to automatically extract useful features from a large amount of data. Specifically, this study proposes an approach to market trend prediction based on a deep recurrent network to model temporal effects of past events. The validity of the proposed approach is demonstrated on real-world data for ten Nikkei companies.

1. はじめに

近年、機械学習の手法を応用し、株価を含む膨大な金融情報を分析することで、投資家の判断を支援する技術が注目されている。投資を行う際、投資家は新聞記事やマイクロブログなどの言語情報から表出する有益な情報を基に、市場の動きを分析・予測する。しかし、発信される全ての言語情報を投資家自身が全て分析することは困難である。そこで、Lavrenkoら [Lavrenko 00] や Schumaker ら [Schumaker 09] の研究など、これらの言語情報を用いて株価の動きを分析する試みが盛んに行われてきた。

一方、深層学習を用いた手法が、自然言語処理や画像認識などの分野において高い精度を上げており、近年注目を浴びている [Dahl 12]。Socher ら [Socher 11] は、深層学習の手法の一つである Autoencoder を拡張したモデルを用いることで、文章の感情予測での精度の高さを示している。

このことから、株価動向推定に関しても、深層学習を利用することにより、精度が向上する可能性があると考えられる。しかし、深層学習を用いた手法のうち、言語情報に対する手法の大半が、時間的な変化を考慮していない。株価は、様々な事象によって時々刻々と変動する。また、ある株価に影響を与える事象が生じ、それが言語情報として現れたとき、その情報は、長期に渡って株価に影響を与え得る。例えば、2008年9月15日にリーマン・ショックが起きたとき、多くの株価が10月下旬まで下落している。同日、日本経済新聞朝刊にリーマン・ショックに関する記事が複数あり、これらの記事を考慮すれば、長期に渡る株価の下落を推定できる可能性がある。従って、時間的な変化を含む情報に対しては、それらを考慮したモ

デルを構築することが必要であると考えられる。

本研究では、Recurrent Neural Networks-Restricted Boltzmann Machine (RNN-RBM) [Boulanger-Lewandowski 12] を用いて、言語情報が株価に与える影響の時間的な変化を捉え、深層学習の枠組みで株価動向推定を行う手法を提案する。

2. 関連研究

本章では、まず株価動向推定に関連する研究を述べ、次に一般的な深層学習のモデルについて説明する。その後、時系列データを扱う深層学習のモデルの構造や学習アルゴリズムについて述べる。

2.1 言語情報を用いた株価動向推定

Lavrenko ら [Lavrenko 00] は、金融のニュース記事を用いて株価の変動の予測を行った。まず株価を区分的線形回帰 [Pavlidis 74] によって平滑化し、平滑後の各区間をトレンドとして定義する。次に、区間の長さや傾き等の素性を基にトレンドをクラスタリングする。同時に、各トレンドの発生する5時間前までのニュース記事がトレンドにつながるニュース記事であると定義し、Yahoo Finance から取得した127銘柄のタグの付いたニュース記事と結びつける。テスト期間に新たなニュース記事が登場すると、記事の bag-of-words からベイズの定理を用いて近い将来に各トレンドが発生する確率を求め、どのトレンドが次に発生するのかを予測する。実験では予測したトレンドを基に実際に各銘柄の売買を行うシミュレーションを行い、利益を得ることが可能であることを示した。

2.2 一般的な深層学習モデル

2.2.1 Restricted Boltzmann Machines

Restricted Boltzmann Machines (RBM) [Smolensky 86] は、確率的深層学習モデルのひとつである。RBM では、可視

連絡先: 吉原 輝, 神戸大学工学部情報知能工学科,
yoshihara@ai.cs.kobe-u.ac.jp

層 \vec{v} と隠れ層 \vec{h} の結合分布をエネルギー関数 E を用いて定義する．

$$P(\vec{v}, \vec{h}) = e^{-E(\vec{v}, \vec{h})} / Z \quad (1)$$

$$E(\vec{v}, \vec{h}) = -b_v^T \vec{v} - b_h^T \vec{h} - \vec{v}^T W \vec{h} \quad (2)$$

ここで、 b_v, b_h, W はモデルのパラメータであり、それぞれ、可視層のバイアス項、隠れ層のバイアス項、重み行列を表す． Z は正規化項である． \vec{v} が与えられたときの隠れ層のノード h_i 、 \vec{h} が与えられたときの可視層のノード v_i はそれぞれ以下のように計算される．

$$P(h_i = 1 | \vec{v}) = \sigma(b_h + Wv)_i \quad (3)$$

$$P(v_j = 1 | \vec{h}) = \sigma(b_v + W^T h)_j \quad (4)$$

ただし、 $\sigma(\cdot)$ はシグモイド関数を表す． \vec{v} の周辺確率 $P(\vec{v})$ は、free-energy $F(\vec{v})$ を用いて、

$$P(\vec{v}) = e^{-F(\vec{v})} / Z \quad (5)$$

$$F(\vec{v}) = -b_v^T \vec{v} - \sum_i \log(1 + e^{b_h + Wv}_i) \quad (6)$$

と表される．この周辺確率 $P(\vec{v})$ の対数尤度を最大化することにより、パラメータの学習を行う．対数尤度の勾配を求める際に登場する 2 つの項をそれぞれ正項、負項と呼ぶ．しかし、実際には負項の計算コストが高いため、ギブスサンプリング法を用いて、近似的に求める．このような学習法を Contrastive Divergence (CD) 法 [Hinton 02] と呼ぶ．

2.2.2 Deep Belief Networks

Deep Belief Networks (DBN) [Hinton 06] は、RBM を階層的に積み上げて構成される深層学習モデルの一種である．RBM と違い、DBN は隠れ層同士で結合しているため、異なる表現能力を持つことが期待される．DBN のパラメータの学習は、RBM を利用した事前学習 (pre-training) とファインチューニングに分けられる [Bengio 07]．事前学習では、階層的に積み上げられた RBM を個々に分け、各層の RBM ごとにパラメータの学習を行う．

ファインチューニングでは、事前学習で求められたパラメータを用いて、教師あり学習を行う．まず、出力層として DBN にロジスティック回帰を行う層を追加する．この層を追加することで、学習したパラメータを基に推定値が算出できる．次に、ラベルと算出された推定値の誤差を計算し、この誤差を最小とするように出力層のパラメータ W^o, \vec{b}^o を更新する．その後、 W^o, \vec{b}^o を用いて、隠れ層 \vec{h}^3 のパラメータ W^3, \vec{b}^3 を更新する．他の層のパラメータも同様にして更新される．

2.3 時系列データを扱う深層学習モデル

RTRBM (Recurrent Temporal Restricted Boltzmann Machine) [Sutskever 08] では、モデルを再帰的に構築することで、ある時刻 t の隠れ表現 \vec{h}_t は、 t 以前の時間的な変化を表現しているとし、時刻が 1 つ前の隠れ層の結合のみを考えている．このことにより、サンプリングの際に用いる条件付き確率を算出するコストが削減される．時刻 t において、 \vec{h}_{t-1} が与えられたときの \vec{v}_t と \vec{h}_t の条件付き結合分布は、次のよう

に表される．

$$P(\vec{v}_t, \vec{h}_t | \vec{h}_{t-1}) = \frac{\exp(\vec{v}_t^T b_v + \vec{h}_t^T W \vec{v}_t + \vec{h}_t^T (b_h + W' \vec{h}_{t-1}))}{Z(\vec{h}_{t-1})} \quad (7)$$

ここで、 b_v, b_h, W は式 (2) と同様であり、 W' は、 \vec{h}_{t-1} から \vec{h}_t への重み行列である．この式を用いて、RTRBM における結合分布 $P(v_1^T, h_1^T)$ は、

$$\begin{aligned} P(v_1^T, h_1^T) &= \prod_{t=1}^T \sum_{h'_t} P(\vec{v}_t, \vec{h}'_t | \vec{h}_{t-1}) P(\vec{h}_t | \vec{v}_t, \vec{h}_{t-1}) \\ &= \prod_{t=1}^T P(\vec{v}_t | \vec{h}_{t-1}) P(\vec{h}_t | \vec{v}_t, \vec{h}_{t-1}) \end{aligned} \quad (8)$$

である．

3. 提案手法

本研究では、新聞等の言語情報から時間的に変動する株価の上昇・下落を予測することを目的とする．言語情報が株価に与える影響の時間的な変化を捉えるために、ここでは Recurrent Neural Networks-Restricted Boltzmann Machine (RNN-RBM) を利用する．RNN-RBM は、時系列情報を考慮した深層学習のモデルであり、2.3 節で述べた RTRBM を拡張したモデルである．本章では、3.1 節で、RTRBM の問題点と RNN-RBM の概要について述べ、その後、本手法の概要及び本手法で用いる素性について述べる．

3.1 RNN-RBM の概要

RTRBM では、ある時刻の隠れ層が必ず次の時刻の隠れ層に影響を与えるという制約が生じている．時系列データにおいて、全てのデータが必ずしも未来に影響を与えるとは限らず、時刻によって、その影響は変動すると考えられる．実際に新聞記事においても、株価に影響を与える記事が数多く存在する．例えば、リーマンショックに関する記事が株価に与える影響は長期的である．一方で、株式市場に関する記事が株価に与える影響は短期的と言われており、これらの記事は未来の株価には影響しないと考えられる．もし、RTRBM を株価動向推定に用いると、ある記事が必ず未来の株価に影響するという制約が生まれ、結果として予測を誤る可能性がある．RNN-RBM は、そのような制約を解消するため、RTRBM を更に拡張した手法である．

RNN-RBM は、パラメータ $W, \vec{b}_h^{(t)}, \vec{b}_v^{(t)}$ を持つ RBM と、 $\vec{W}', W'', W_2, W_3, \vec{h}^{(t)}$ を持つ RNN で構成されている．RBM の隠れ層とは異なる隠れ層を加えることで、時間的な変化を表す隠れ表現 \vec{h} と観測データの隠れ表現 \vec{h} を区別し、先に述べた制約を解決している．

単純化のために、1 層のみの RNN-RBM を考えることにすると、隠れ表現 $\vec{h}^{(t)}$ は、次のように表される．

$$\vec{h}^{(t)} = \sigma(W_2 v^{(t)} + W_3 \vec{h}^{(t)} + \vec{b}_h) \quad (9)$$

また、各バイアス項 $\vec{b}_h^{(t)}, \vec{b}_v^{(t)}$ は、

$$\vec{b}_h^{(t)} = \vec{b}_h + W' \vec{h}^{(t-1)} \quad (10)$$

$$\vec{b}_v^{(t)} = \vec{b}_v + W'' \vec{h}^{(t-1)} \quad (11)$$

で求められる。これらのパラメータを用いて、このモデルの学習は行われる。学習は、2つのステップに分かれており、まずCD法によりRBMのパラメータにおける導関数を求め、その後、Backpropagation Through Time (BPTT) アルゴリズムを適用することによりRNNのパラメータにおける導関数を求める。

3.2 DBN への導入

提案手法の概要を図1に示す。本手法では、DBNを拡張したモデルを利用する。拡張するモデルにDBNを選択した理由は2点あり、まず1点目は、DBNが多層構造であるからである。深層学習では、層の数を増やすことによって、表現力が増すと考えられている。2点目は、RNN-RBMがRBMを拡張したモデルであるからである。DBNがRBMを階層的に積み上げて構成されていることから、その構成要素であるRBMをRNN-RBMに拡張することは容易である。このように拡張することで、DBNは多層構造であり時系列情報を考慮したモデルとなる。

本手法は、DBNの層のうち、入力層 \vec{v} と隠れ層 \vec{h}^1 からなるRBMの代わりに、RNN-RBMを用いる。このことにより、入力データの時間的な変化を考慮した隠れ表現が得られると考えられる。また、本研究は、株価の上昇あるいは下落を推定する2値分類問題を扱うため、出力層の次元は1となる。

本手法の学習は、DBNと同様、事前学習とファインチューニングに分けて行う。事前学習では、先に述べたRBMとRNN-RBMの学習法により、それぞれのパラメータを求める。ファインチューニングでは、それらのパラメータを更新する。

3.3 本手法で用いる素性

先に述べたように、本研究は言語情報が株価に与える影響の時間的な変化を捉えることが重要となるため、入力データは言語情報となり、モデルが扱えるようにこれらをベクトルで表現する必要がある。本手法では、日単位で記事を分け、それぞれの記事群をbag-of-wordsでベクトル表現する。

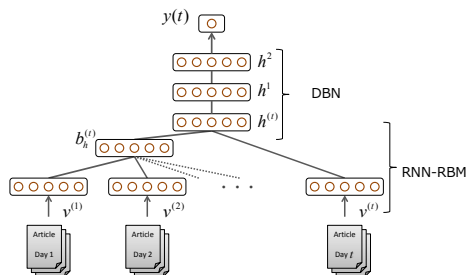


図1: 提案するモデルのグラフィカルモデル

4. 評価実験

4.1 実験設定

本実験で扱う言語情報は、日本経済新聞の本紙朝刊であり、1999年から2008年までの全1,033,277記事を用いた。このうち、1999年から2006年までの8年間の834,882記事を訓練データ、2007年の98,667記事を検証データ、2008年の99,728記事をテストデータとした。株価動向推定の対象とした銘柄は、日経平均に採用されている225銘柄のうち、銘柄名を含む記事が存在する日数が最も多い10銘柄と日経平均株価を用いた。実験結果の評価には、新聞記事の発行された日のMoving Average Convergence Divergence (MACD)*1と翌日のMACDに関する株価動向適合率 (Up Down Correct Rate; UDCR) を用いた。

4.2 実験結果

本手法により株価動向を推定し、MACDにより評価を行った結果を表1に示す。表1が示すそれぞれの数値は、テストデータにおけるエラー率 (%) である。本実験では、テストデータ中の株価上昇・下落を集計し、多い方を常に選択した方法をベースラインとする。比較手法には、ベースラインの他、SVM、時間的な変化を考慮しないDBNを用いた。ベースライン、SVMとの比較で、深層学習を導入することの有効性が確かめられ、DBNとの比較で、時系列情報を考慮する必要性が検証できる。

ベースラインやSVMと比較すると、全ての銘柄において提案手法のエラー率が下回っている。平均値においてもそれぞれ約8.2ポイント、約3.5ポイントのエラー率減少を実現した。また、DBNと比較すると、11銘柄中6銘柄において提案手法のエラー率が下回り、3銘柄においては同等のエラー率、平均値においては約1ポイントのエラー率減少を達成した。

表1: テストデータにおける株価動向推定のエラー率

brand	RNN-RBM			
	baseline	SVM	DBN	+DBN
Nikkei Stock Average	54.66	49.57	45.50	43.62
Hitachi	35.71	37.29	32.00	32.00
Toshiba	39.50	43.64	38.50	38.50
Fujitsu	40.00	41.53	32.00	34.00
Sharp	42.00	52.12	40.00	40.00
Sony	43.00	45.34	41.43	40.95
Nissan Motor	44.92	40.00	39.50	37.00
Toyota Motor	53.39	44.29	43.81	42.38
Canon	43.81	50.42	43.00	39.11
Mitsui	46.96	47.88	41.43	41.43
Mitsubishi	43.81	49.15	43.33	40.43
average	42.60	47.30	40.05	39.04

4.3 考察

表1に示す精度が妥当であるかどうか評価するため、 t 検定を行った。その結果、SVMとベースラインに対しては、有意水準1%において提案手法の有効性が確認できた。一方、DBNに対しては p 値が0.076となり、有意差が見られなかった。こ

*1 移動平均収束拡散手法とも呼ばれ、株価の将来の値動きを予想するテクニカル分析の指標の一つである。

の原因として、一年間を通して、長期的に株価へ影響を与える事象が非常に少ないということが考えられる。実際に、テストデータである 2008 年の新聞記事において、1ヶ月近くの株価の下落に影響を与えたと思われる事象は、リーマンショックのみであった。従って、その事象を考慮していたとしても、一年間全体の精度では差が現れにくいのではないかと推測される。

そこで、リーマンショックが長期的に株価へ与えた影響を考慮できているか検証するため、一年間全体ではなく、特定の期間において再度実験を行った。ここで、特定の期間とは、リーマンショックが生じた 2008 年 9 月 15 日から 10 月 28 日までの約 1ヶ月間である。その結果を表 2 に示す。

表 2: リーマンショックが生じてから 1ヶ月間でのエラー率の比較

brand	DBN	RNN-RBM+DBN
Nikkei Stock Average	51.61	38.70
Hitachi	61.29	32.25
Toshiba	54.83	38.70
Fujitsu	45.16	32.25
Sharp	58.06	45.16
Sony	41.93	41.93
Nissan Motor	29.03	35.48
Toyota Motor	48.38	45.16
Canon	54.83	54.83
Mitsui	41.93	38.70
Mitsubishi	29.03	25.80
average	46.92	39.00

表 2 から、特定の期間に注目した場合、表 1 と比べて、ほとんどの銘柄において、提案手法のエラー率が DBN を下回っていることがわかる。このことから、RNN-RBM を用いることで、リーマンショックという長期的に株価に影響を与える事象を捉えることができ、且つ、その事象を考慮することが株価動向の推定に有効であると判断できる。

5. 結論

本論文では、株価という時系列情報の特性に着目し、深層学習によって新聞記事が株価に与える影響の時間的な変化を捉え、株価動向推定を行う手法を提案した。本手法では、時系列情報を考慮した再帰的なモデルである RNN-RBM と RBM を階層的に積み上げて構成されたモデルである DBN を組み合わせたモデルを利用した。評価実験では、株価動向推定に深層学習を導入することの有効性と長期的に株価に影響を与える事象を考慮することの有効性を示した。

今後の課題としては、bag-of-words の他の文書の表現方法として、言語情報から必要な情報を抽出・統合するような文書の圧縮表現方法を検討していく。

参考文献

[Bengio 07] Bengio, Y., Lamblin, P., Popovici, D., and Larochelle, H.: Greedy layer-wise training of deep networks, in *Proceedings of the twenty-first international conference on Neural Information Processing Systems*, pp. 153–160 (2007)

[Boulanger-Lewandowski 12] Boulanger-Lewandowski, N., Bengio, Y., and Vincent, P.: Modeling temporal dependencies in high-dimensional sequences: Application to polyphonic music generation and transcription, in *Proceedings of the twenty-ninth International Conference on Machine Learning*, pp. 1159–1166 (2012)

[Dahl 12] Dahl, G. E., Yu, D., Deng, L., and Acero, A.: Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition, *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 20, No. 1, pp. 30–42 (2012)

[Hinton 02] Hinton, G. E.: Training products of experts by minimizing contrastive divergence, *Neural computation*, Vol. 14, No. 8, pp. 1771–1800 (2002)

[Hinton 06] Hinton, G. E., Osindero, S., and Teh, Y.-W.: A fast learning algorithm for deep belief nets, *Neural computation*, Vol. 18, No. 7, pp. 1527–1554 (2006)

[Lavrenko 00] Lavrenko, V., Schmill, M., Lawrie, D., Ogilvie, P., Jensen, D., and Allan, J.: Mining of concurrent text and time series, in *Proceedings of the KDD-2000 Workshop on Text Mining*, pp. 37–44 (2000)

[Pavlidis 74] Pavlidis, T. and Horowitz, S.: Segmentation of plane curves, *IEEE transactions on Computers*, Vol. 23, No. 8, pp. 860–870 (1974)

[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 (TOIS)*, Vol. 27, No. 2, pp. 12:1–12:19 (2009)

[Smolensky 86] Smolensky, P.: Information Processing in Dynamical Systems: Foundations of Harmony Theory, in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 1, pp. 194–281 MIT Press (1986)

[Socher 11] Socher, R., Pennington, J., Huang, E. H., Ng, A. Y., and Manning, C. D.: Semi-supervised recursive autoencoders for predicting sentiment distributions, in *Proceedings of the sixteenth Conference on Empirical Methods in Natural Language Processing*, pp. 151–161 (2011)

[Sutskever 08] Sutskever, I., Hinton, G. E., and Taylor, G. W.: The recurrent temporal restricted Boltzmann machine, in *Proceedings of the twenty-second international conference on Neural Information Processing Systems*, pp. 1601–1608 (2008)