

多相アテンション LSTM を用いた財務時系列データの予測 Forecasting Corporate Financial Time Series using Multi-phase Attention Recurrent Neural Networks

吉見 脩平^{1*} 江口 浩二^{2†} 金京 拓司³ 羽森 茂之³
Shuhei Yoshimi¹ Koji Eguchi² Takuji Kinkyo³ Shigeyuki Hamori³

¹ 神戸大学大学院システム情報学研究科

¹ Graduated School of System Informatics, Kobe University

² 広島大学情報科学部

² School of Informatics and Data Science, Hiroshima University

³ 神戸大学大学院経済学研究科

³ Graduated School of Economics, Kobe University

Abstract: Recently, attention-based RNNs have been studied to represent multivariate temporal or *spatio-temporal* structure underlying multivariate time series. One recent study has achieved improved performance by employing attention structure that simultaneously capture the *spatial* relationships among multivariate time series and the temporal structure of those time series. That method assumes a single time-series sample of multivariate explanatory variables, and thus, no prediction method was designed for multiple time-series samples of multivariate explanatory variables. Moreover, such previous studies have not explored on financial time series incorporating macroeconomic time series, such as Gross Domestic Product (GDP) and stock market indexes, to our knowledge. Also, no neural network structure has been designed for focusing a specific industry. We aim in this paper to achieve effective forecasting of corporate financial time series from multiple time-series samples of multivariate explanatory variables. We propose a new industry specific model that appropriately captures corporate financial time series, incorporating the industry trends and macroeconomic time series as side information. We demonstrate the performance of our model through experiments with Japanese corporate financial time series in the task of predicting the return on assets (ROA) for each company.

1 はじめに

近年、情報技術の発達に伴い、日々多くの情報が収集、蓄積されている。その分野は多岐にわたり、枚挙に暇がないが、その一つに、経済・金融分野における企業財務時系列データが挙げられる。これらを活かして新たな知見を得ることについて当該分野において関心が高まりつつある。企業財務時系列データは、企業の経営状況や業界の動向、さらには社会の景況感などの影響を受けた複雑な時系列データとなっており、その予測のためには複数の時系列の分析を必要とする。従来の時系列分析や最近の深層学習の研究において、時系列予測

の問題に広く取り組まれてきたものの、企業の財務時系列などの複雑な多変量時系列に関しては新たな研究の余地が十分にある。広く認知された時系列予測手法の一つに自己回帰統合移動平均モデル (Autoregressive integrated moving average: ARIMA) があり、時系列における時間的パターンの 1 つの側面の分析に焦点が当てられているものの、多変量時系列の的確な長期予測を達成することは容易ではないとされる [1]。再帰型ニューラルネットワーク (Recurrent neural network: RNN)、および、RNN の欠点を改善して長期にわたった時間依存性を考慮できる長・短期記憶モデル (Long short-term memory: LSTM) は、株価予測などの金融時系列分野における有用な手法の一つとして注目されつつある [8]。しかしながら、多変量時系列の適格な長期予測を達成するには解決すべき課題が依然として少なくない。具体的には、多変量の時系列を考慮するこ

*連絡先： 神戸大学大学院システム情報学研究科
〒 657-8501 兵庫県神戸市灘区六甲台町 1-1
E-mail:shuhei@cs25.scitec.kobe-u.ac.jp

†連絡先： 広島大学情報科学部
〒 739-8527 広島県東広島市鏡山 1-4-1
E-mail:eguchi@acm.org

とにより、一部の説明変数時系列が目的変数時系列の予測への寄与が小さいときに、その予測性能に悪影響を与え得る。このような問題意識のもと、複数時系列に対して予測に関する重要度の分布を学習して複数時系列の特徴表現能力をさらに強化する目的で、アテンションメカニズムに基づく RNN を用いて、時系列間の相関（以下、空間的相関）を捉えつつ時間的関連性を抽出する予測モデルが提案された [2]。また、Qin ら [11] と Yuxuan ら [9] は、アテンションメカニズムとエンコーダ・デコーダモデルを組み合わせ、単一ステップおよび短期時系列予測での予測性能を向上させた。さらに、Liu ら [10] は、多段階のアテンションメカニズムを介して空間的相関へのアテンションを強化しつつ、目的変数時系列の埋め込みを介して時間的関連性へのアテンションを強化するような、2 段 2 相 (dual-stage two-phase: DSTP) のアテンションメカニズムに基づく RNN モデルを開発した。しかしながら、それらの研究では多変量説明変数を伴う複数の時系列サンプルが与えられた状況での分析と予測に関して報告されていない。また、国内総生産 (GDP) や日経平均株価などのマクロ経済時系列を組み込んだり、業種ごとの傾向に着目するような RNN モデルによる財務時系列分析の研究は、著者の知る限りない。本論文では、企業財務時系列の予測に対する有用な手法を確立するため、アテンションメカニズムに基づく RNN モデルを拡張し、多変量説明変数を伴う複数時系列サンプルに対する的確な予測を目指す。また、マクロ経済時系列を適切に捉えたモデル化を行う。さらに、特定の業種の動向を適切に反映したモデルを提案する。日本の企業財務時系列を用いて各企業の総資本経常利益率 (ROA) を予測する実験を行い、提案モデルが有効に機能することを示す。

2 関連研究

本章は次の 3 つの話題からなる。最初の話題としては、時系列データを予測するための基本的な深層学習手法の 1 つとして RNN を概説する。次に 2 番目の話題として、RNN における長期依存性を考慮できない問題に対処すべく拡張された LSTM を紹介する。さらに 3 番目の話題として、近年注目を集めつつあるアテンションメカニズムを取りあげる。

2.1 再帰型ニューラルネットワーク (RNN)

系列データは、サンプルの順序が重要であるようなデータの総称である。特に、その順序が時間に基づいている場合は時系列データと呼ばれる。過去のサンプルと現在のサンプルとの間の依存関係をモデル化する研

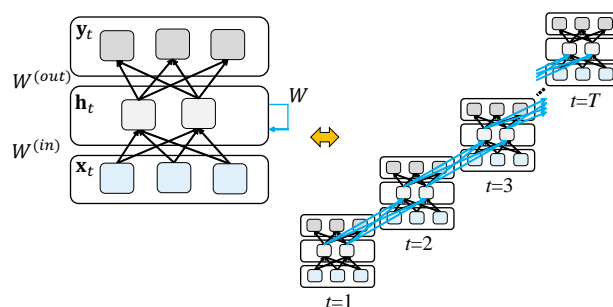


図 1: RNN の構造.

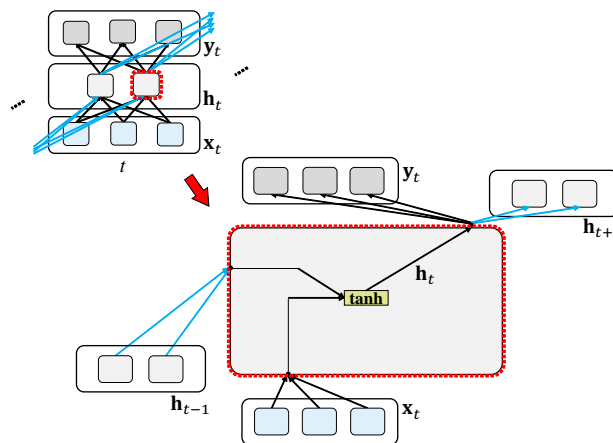


図 2: RNN の中間層のユニット.

究が、時系列予測の問題に対して取り組まれてきた。なかでも広く受け入れられた手法の 1 つに再帰型ニューラルネットワーク (Recurrent Neural Network: RNN) があり、順伝播型ニューラルネットワークを系列データの処理のために拡張したモデルとして位置づけられる [3]。ここで、RNN は各時間 t の入力系列 \mathbf{x}_t を受け取り、時間 t で 1 つの出力系列 \mathbf{y}_t を返すように定式化する¹。時間 $t = n$ において出力系列 \mathbf{y}_n が入力系列 $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ から正常に生成されたと仮定できる。これは RNN が「リターンパス」と呼ばれる有向閉ループを使用しているためである。この構造により、時間情報の保存と更新が可能になる。図 1 に、RNN の構造と、時間次元で展開された構造を示す。図中で、 \mathbf{x}_t は入力、 \mathbf{h}_t は隠れ状態、 \mathbf{y}_t は出力を示し、それぞれの層を入力層、中間層、出力層と呼ぶ。 $W^{(in)}$ は入力層から中間層への重み、 $W^{(out)}$ は中間層から出力層への重みを示している。RNN では時間 $t-1$ の隠れ状態から時間 t の隠れ状態への接続が想定され、時間 $t-1$ における任意の隠れ状態と時間 t の任意の隠れ状態の間に重み $w \in W$ が割り当てられる。図 2 に RNN の中間層の 1 つのユニットを示す。

¹後述する提案モデルでは、各時間 t の多変量説明変数系列 \mathbf{X}_t を受け取り、時間 $t+1$ で 1 つの目的変数系列 \mathbf{y}_{t+1} を返すように定式化する。

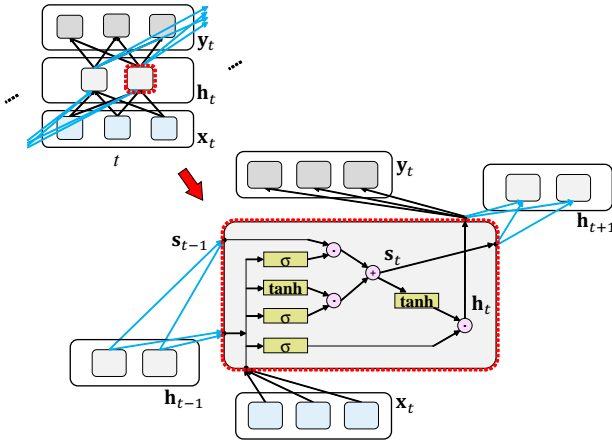


図 3: LSTM の中間層のユニット。

2.2 長・短期記憶モデル (LSTM)

RNNにより系列データのコンテキストを捉えることができるが、そのとき、現在の時刻からどれだけ過去の入力を反映して出力を予測するかを考えることが重要である。RNNの未知パラメータを学習する際に目的関数の勾配を用いて更新を繰り返すのが基本となるが、一定数以上の繰り返して勾配が消失（または発散）することが知られており、勾配消失問題と呼ばれる [4, 5]。この問題に対処するために、長期記憶を実現する長・短期記憶モデル (Long Short Term Memory: LSTM) が提案されている [6]。図3に LSTM の中間層の1つのユニットを示す。LSTM では中間層の各ユニットにおいて情報を記憶するためのセル (s_t) と情報の流れを制御するためのゲートが配置され、その他の構造は基本的に RNN と同じである。一般に LSTM は RNN の一種と位置づけられる。

2.3 アテンションメカニズム

アテンションメカニズムは LSTM に使用され、その有効性が示されている [11, 10]。LSTM に基づく時系列予測において、アテンションメカニズムにより多変量時系列の間の相関とそれらの時間的構造を同時に捉えることができる。本節の残りでは、文献 [11, 10] において LSTM のために開発されたアテンションメカニズムを概説する。

空間的アテンション LSTM 空間的アテンションメカニズム (Spatial attention mechanism) は、多変量時系列の間の相関 (以下、空間的相関) を抽出することを目的としている [11, 10]。時刻 t までに観測された多変量時系列 (長さ T , 変量数 n) における k 番目の変量の系列を $\mathbf{x}_t^k = (x_{t-T+1}^k, \dots, x_t^k) \in \mathbb{R}^T$ とするとき、

次のようなアテンションメカニズムを考える。

$$a_t^k = \mathbf{v}_a^T \tanh(W_a[\mathbf{h}_{t-1}; \mathbf{s}_{t-1}] + U_s \mathbf{x}_t^k + \mathbf{b}_a) \quad (1)$$

$$\alpha_t^k = \frac{\exp(a_t^k)}{\sum_{j=1}^n \exp(a_t^j)} \quad (2)$$

ここで、 $[\cdot; \cdot]$ は接続演算を示し、 $\mathbf{h}_{t-1} \in \mathbb{R}^m$ および $\mathbf{s}_{t-1} \in \mathbb{R}^m$ はそれぞれ時間 $t-1$ に対応する LSTM ユニットの隠れ状態およびセル状態を示し、 m は隠れ状態数を示す。また、 $\mathbf{v}_a, \mathbf{b}_a \in \mathbb{R}^T$, $W_a \in \mathbb{R}^{T \times 2m}$, $U_s \in \mathbb{R}^{T \times T}$ は未知パラメータを示す。このように時間 $t-1$ の隠れ状態と時間 t までの多変量入力時系列によって、時間 t における各変量の重要性の度合い (以下、空間的アテンション重み) $\alpha_t = (\alpha_t^1, \dots, \alpha_t^n)$ を求める。この重みを用いて、時間 t の時点での多変量入力 $\mathbf{x}_t = (x_t^1, \dots, x_t^n)$ を次式のように重みづける。

$$\tilde{\mathbf{x}}_t = (\alpha_t^1 x_t^1, \alpha_t^2 x_t^2, \dots, \alpha_t^n x_t^n)^T \quad (3)$$

以上に述べたような空間的アテンション LSTM ユニットの $f_{spatial}$ で表すと次式を得る。

$$(\mathbf{h}_t, \mathbf{s}_t) = f_{spatial}(\mathbf{h}_{t-1}, \mathbf{s}_{t-1}, \tilde{\mathbf{x}}_t) \quad (4)$$

時間的アテンション LSTM 時間的アテンションメカニズム (Temporal attention mechanism) は、前項の空間的アテンションの時間依存性を捉えることを目的としている [11, 10]。長さ T の時間ウィンドウにおける時空間構造を抽出するために、前項で述べた多変量時系列の空間的相関を反映させた隠れ状態ベクトルを用いる。固定幅ウィンドウで時間的関連性を抽出するだけでは十分ではないため、隠れ状態を選択するためのアテンションメカニズムを適用する。具体的には、予測対象の目的変数に関連性が高い隠れ状態を選択することにより長期依存性を反映させる。アテンションメカニズムにより、 i 番目の隠れ状態に関して各時間の重要性の度合い (以下、時間的アテンション重み) $\beta_i = (\beta_i^1, \dots, \beta_i^T)$ を次のように求める。

$$b_t^i = \mathbf{v}_b^T \tanh(W_b[\mathbf{d}_{t-1}; \mathbf{s}'_{t-1}] + U_b \mathbf{h}_i + \mathbf{b}_b) \quad (5)$$

$$\beta_t^i = \frac{\exp(b_t^i)}{\sum_{j=1}^T \exp(b_t^j)} \quad (6)$$

ここで、 \mathbf{h}_i は、前項で述べた空間的相関を表す隠れ状態の i 番目のベクトルを示す。 $\mathbf{d}_{t-1} \in \mathbb{R}^p$ および $\mathbf{s}'_{t-1} \in \mathbb{R}^p$ はそれぞれ時間 $t-1$ に対応する LSTM ユニットの隠れ状態およびセル状態を示し、 p は隠れ状態数を示す。 $\mathbf{v}_b, \mathbf{b}_b \in \mathbb{R}^p$, $W_b \in \mathbb{R}^{p \times 2p}$, $U_b \in \mathbb{R}^{p \times m}$ は未知パラメータを示す。次に、コンテキストベクトル \mathbf{c}_t を次のように定義する。

$$\mathbf{c}_t = \sum_{j=1}^T \beta_t^j \mathbf{h}_j \quad (7)$$

コンテキストベクトル \mathbf{c}_t はすべての隠れ状態の重み付き和を表し、時間ウィンドウ内における時間的関連性を表す。

このようにして求めたコンテキストベクトル \mathbf{c}_t と目的変数 y_t を次式のようにして組み合わせる。

$$\tilde{y}_t = \tilde{\mathbf{w}}^T[y_t; \mathbf{c}_t] + \tilde{b} \quad (8)$$

ここで、 $\tilde{\mathbf{w}} \in \mathbb{R}^{m+1}$ および $\tilde{b} \in \mathbb{R}$ は接続ベクトルを目的変数に写像するパラメータである。目的変数時系列をコンテキストベクトルに対応付けることにより時間的関連性を管理しやすくなり、その結果を使用して隠れ状態を更新することができる。以上に述べたような LSTM ユニットの $f_{temporal}$ で表すと次式を得る。

$$(\mathbf{d}_t, \mathbf{s}'_t) = f_{temporal}(\mathbf{d}_{t-1}, \mathbf{s}'_{t-1}, \tilde{y}_{t-1}) \quad (9)$$

最後に、長さ T の多変量説明変数時系列 $X_T = (\mathbf{x}_1, \dots, \mathbf{x}_t, \dots, \mathbf{x}_T)$ (ただし $\mathbf{x}_t = (x_t^1, \dots, x_t^n)$) および目的変数時系列 $\mathbf{y}_T = (y_1, \dots, y_t, \dots, y_T)$ から時間 $T+1$ の目的変数 y_{T+1} を予測するため、隠れ状態 \mathbf{d}_T とコンテキストベクトル \mathbf{c}_T を接続し、次式を用いる。

$$\hat{y}_{T+1} = F(X_T, \hat{\mathbf{y}}_T) \quad (10)$$

$$= \mathbf{v}_y^T(W_y[\mathbf{d}_T; \mathbf{c}_T] + \mathbf{b}_y) + b'_y \quad (11)$$

ここで、 F は予測器を示す。 $W_y \in \mathbb{R}^{p \times (p+m)}$ および $\mathbf{b}_y \in \mathbb{R}^p$ は接続ベクトル $[\mathbf{h}_T; \mathbf{c}_T] \in \mathbb{R}^{p+m}$ を p 次元潜在空間に写像している。重み $\mathbf{v}_y \in \mathbb{R}^p$ およびバイアス $b'_y \in \mathbb{R}$ をともなう線形変換により、最終的な目的変数の予測値 \hat{y}_{T+1} を求めている。

3 データ

3.1 財務時系列データ

本論文では、統計法に基づく基幹統計として「法人企業統計調査規則」(昭和45年大蔵省令48号)に基づいて財務省により収集された「法人企業統計調査」を使用する²。2003年の第1四半期から2016年の第4四半期までの14年間(56四半期)の調査結果を使用する。なお、本調査は年次調査と四半期調査で構成されているが、四半期調査を主に使用し、年次調査は補完処理のために補助的に用いる。本調査に収録された企業の総数は四半期調査で57,775、年次調査で60,516であった。調査項目を表1に示す。

四半期調査による財務諸表に基づいて求めた種々の財務比率を目的変数または説明変数として用いる。前処理として、まず、短期間のみ記録された企業などのデータを除外する(除外処理)。次に、財務比率を

²調査対象のうち、「金融および保険」業界の企業は除外した。

計算する前に財務諸表における欠損値を補完する処理を行う(補完処理)。最後に、補完済みの財務諸表を用いて財務比率の計算を行う(計算処理)。以上の手順によって得られた四半期次の各種財務比率の時系列を分析に用いる。それぞれの詳細について以下に示す。

除外処理 次のいずれかの条件に該当する企業データを除外した。すなわち、除外の対象は、(1)財務諸表の分析対象期間(56期)の一部において全調査項目が欠損している企業データ、(2)財務諸表におけるある調査項目に関して分析対象の全期間にわたって欠損している企業データである。この除外処理の後、分析に使用される企業数は2296となった。なお、後述する業種特化型モデルのために着目する卸売業の企業数は237となった。

補完処理 本論文における補完処理は、年次データを利用した補完と線形補間に基づく。詳細については稿を改めて報告する。

計算処理 本論文では種々の財務比率を説明変数または目的変数として使用する。各財務比率は表2で定義される通りであり、これらは基本的に「法人企業統計調査」において使用されているものである。

3.2 マクロ経済時系列データ

本論文では、マクロ経済時系列データとして日経平均株価終値(以下、日経平均)および日本のGDPを用いる。ただし、日経平均は2003年1月から2016年12月までの月次データ³を用いる。また、日本のGDPは2003年第1四半期から2016年第4四半期までの四半期次データ⁴を用いる。

4 提案手法

本章で述べる提案モデルは3つの特徴を持つ。1つ目は多変量時系列が複数与えられる状況すなわち企業ごとに複数の財務比率で表現された時系列が企業数だけ与えられる状況を仮定したモデルである点であり、2つ目は財務時系列に加えてマクロ経済時系列を反映したモデルである点、3つ目は対象業種の動向を反映した業種特化型モデルである点である。また、提案モデルは目的変数時系列の1期先を予測するようなモデルである。提案モデルにおける入力された多変量時系列から予測出力に至るまでの流れを図4に示す。第1フェー

³<https://indexes.nikkei.co.jp/nkave/index/profile?idx=nk225>

⁴https://www.esri.cao.go.jp/jp/sna/data/data_list/sokuhou/files/2019/toukei.2019.html

表 1: 財務諸表. ただし, 「↑」は指定項目が期首に記録されることを示し, 「↓」は期末に記録されることを示す.

分類	四半期調査	年次調査
資産	支払手形 ↑, ↓	支払手形 ↑, ↓
	買掛金 ↑, ↓	買掛金 ↑, ↓
固定資産	土地 ↑, ↓	土地 ↑, ↓
	建設仮勘定 ↑, ↓	建設仮勘定 ↑, ↓
	その他有形固定資産 ↑, ↓	その他有形固定資産 ↑, ↓
	無形固定資産 ↑, ↓	無形固定資産 (ソフトウェア以外) ↑, ↓
総資本 ↑, ↓	総資本 ↑, ↓	
人員	従業員数 ↓	従業員数 ↓
損益	減価償却費 ↓	減価償却費 ↓
		特別減価償却費 ↓
	売上 ↓	売上 ↓
	売上原価 ↓	売上原価 ↓
	営業利益 ↓	営業利益 ↓
	経常利益 ↓	経常利益 ↓

表 2: 財務比率. ただし, 「*」は指定項目について各四半期の期首と期末の平均 (すなわち期中平均) が用いられることを示し, 「**」は指定項目について各四半期の期首から期末への増減量が用いられることを示す.

記号	財務比率	計算式
X_0	総資本営業利益率	$\frac{\text{営業利益}}{\text{総資本}^*}$
X_1	総資本経常利益率 (ROA)	$\frac{\text{経常利益}}{\text{総資本}^*}$
X_2	売上高営業利益率	$\frac{\text{営業利益}}{\text{売上}}$
X_3	売上高経常利益率	$\frac{\text{経常利益}}{\text{売上}}$
X_4	総資本回転率	$\frac{\text{売上}}{\text{総資本}^*}$
X_5	有形固定資産回転率	$\frac{\text{売上}}{(\text{土地} + \text{その他有形固定資産})^*}$
X_6	買掛金回転率	$\frac{\text{売上}}{\text{支払手形} + \text{買掛金}}$
X_7	減価償却率	$\frac{\text{減価償却費}}{\text{その他有形固定資産} + \text{有形固定資産} + \text{減価償却費}}$
X_8	労働装備率	$\frac{\text{従業員数}}{(\text{土地} + \text{その他有形固定資産})^*}$
X_9	キャッシュフロー率	$\frac{\text{経常利益} + \text{減価償却費}}{\text{総資本}^*}$
X_{10}	設備投資率	$\frac{\text{建設仮勘定}^{**} + \text{その他有形固定資産}^{**} + \text{無形固定資産}^{**} + \text{減価償却費}}{\text{総資本}^*}$
X_{11}	粗利率	$\frac{\text{売上} - \text{売上原価}}{\text{売上}}$

ズは対象業種の多変量時系列における空間的相関を抽出することを目的とする. 第 2 フェーズは説明変数時系列と既知の目的変数時系列における空間的相関を抽出することを目的とする. 第 3 フェーズは時間的関連性を抽出し, 時間 $1, \dots, T$ の多変量時系列が与えられたとき, 時間 $T + 1$ の目的変数を予測することを目的とする. 以下に各フェーズについて詳述する.

第 1 フェーズ このフェーズでは, 長さ L の多変量時系列における空間的相関を抽出する. 多変量時系列データから固定幅ウィンドウの部分時系列を切り出す前に, その時系列データ全体に対してアテンションメカニズムを適用することによって, 多変量時系列間の空間的相関をより適切に抽出する点に狙いがある. 多変量時系列が複数与えられた状況では, 固定幅の部分時系列を切り出した後にアテンションメカニズムを適用するならば, 同一企業の複数の部分時系列と他企業の部分時系列が同列に扱われて学習されるため, 企業という単位を考慮した空間的相関を抽出するのに適切でないと考えられるためである. そこでまず企業を単位とした複

数の多変量時系列に対して空間的アテンション LSTM を適用し, 時系列間の空間的相関を抽出する. J を企業数, L を学習データの時系列長, K を多変量時系列の変量数とすると, 多変量時系列 A は $J \times L \times K$ のテンソル (以下, (J, L, K) と表記する) で表現できる. これに空間的アテンション LSTM を適用する. これに対して式 (1), (2), (3) と同様に空間的アテンション重みによる重みづけ (以下, $F_{spatial}$ と表記) を適用する.

$$\hat{A} = F_{spatial}(A) \quad (12)$$

その後, W を重み, b をバイアスとする線形変換により $(1, L, K)$ の \tilde{A} へと集約する.

$$\tilde{A} = W\hat{A} + b \quad (13)$$

また, 日経平均 N は月次データを用いるとき時系列長は $3L$ であるが, アテンションメカニズムを適用するとともに四半期次集約する. このとき, $(1, 3L, 1)$ の N から $(1, L, 1)$ の \tilde{N} へと集約される. 具体的には各四半期 t に対して 1 月めから 3 月めまでを $i \in \{1, 2, 3\}$ で

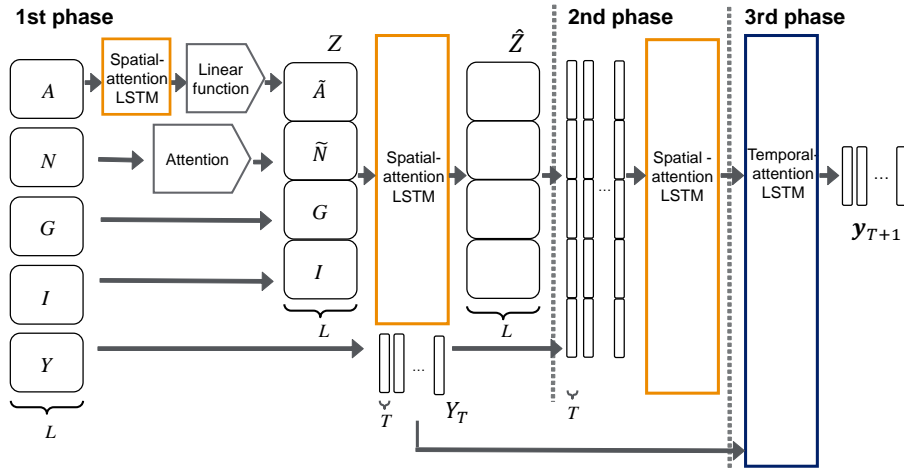


図 4: 提案モデルの構造.

示すとき次式を用いる.

$$\gamma_t^i = \frac{\exp(N_t^i)}{\sum_{j=1}^3 \exp(N_t^j)} \quad (14)$$

$$\tilde{N}_t = (\gamma_t^1 N_t^1, \gamma_t^2 N_t^2, \gamma_t^3 N_t^3)^T \quad (15)$$

以上のようにして全企業の財務時系列が集約された \tilde{A} に対して、四半期次を集約された日経平均の時系列 \tilde{N} 、日本の GDP の時系列 G 、および、対象業種における説明変数時系列 I (目的変数時系列を含まない) を接続する. なお、 G は四半期次データであるため、そのサイズは $(1, L, 1)$ である.

$$Z = [\tilde{A}; \tilde{N}; G; I] \quad (16)$$

ここで、学習時のミニバッチ数を B とすると、 I のサイズは $(B, L, K-1)$ である. したがって、接続後の Z のサイズは $(B, L, 2K+1)$ となる. これに対して空間的アテンション重みで重みづけることにより、時系列間の空間的相関を抽出する.

$$\hat{Z} = F_{spatial}(Z) \quad (17)$$

なお、式 (12) と (17) の $F_{spatial}$ は互いに独立に、それぞれ A と Z から求める.

その後、長さ L の時系列 \hat{Z} の開始期から固定長 T の部分時系列を切り出し、1期ずつずらして固定長 T の部分時系列の切り出しを順次繰り返し、それらを次のフェーズの入力とする. したがって、次のフェーズの入力サンプル数は $B \times (L-T)$ となり、そのサイズは $(B(L-T), T, 2K+1)$ となる.

第 2 フェーズ 第 2 フェーズでは、まず、第 1 フェーズで得た長さ T の \hat{Z} と目的変数時系列 Y_T を、各期

$t \in \{1, \dots, T\}$ で対応づけて接続する. ここで、第 1 フェーズの最後に述べた \hat{Z} を長さ T で分割する要領で、 Y_T も長さ T で分割した部分時系列を用いる.

$$\dot{Z} = [\hat{Z}; Y_T] \quad (18)$$

ここで、接続後の \dot{Z} のサイズは $(B(L-T), T, 2K+2)$ となる. これに対して空間的アテンション LSTM を適用し、目的変数時系列と説明変数時系列の間の空間的相関を抽出する.

$$\hat{\dot{Z}} = F_{spatial}(\dot{Z}) \quad (19)$$

第 3 フェーズ このフェーズでは、時間的アテンション LSTM を適用し、第 2 フェーズで得た空間的相関の時間的関連性を捉える. つまり、異なる時間の異なる (部分) 時系列の間の時空間関連性を抽出し、式 (10) と同様に時間的アテンション LSTM に基づく予測器 (以下、 $F_{temporal}$ と表記) で 1 期先の目的変数を予測する.

$$\hat{y}_{T+1} = F_{temporal}(\hat{\dot{Z}}, Y_T) \quad (20)$$

この \hat{y}_{T+1} が最終的な予測結果である. \hat{y}_{T+1} は予測対象の企業数だけの長さを持つベクトルであり、時間 $T+1$ での各企業に対する目的変数予測値を要素にとる.

提案モデルの学習には誤差逆伝播アルゴリズムを用いる. 学習過程で Adam オプティマイザ [7] によって、予測ベクトル \hat{y}_{T+1} と正解ベクトル y_{T+1} の間の平均二乗誤差 (MSE) を最小化する.

モデル構造 第 1 フェーズにおいて、部分時系列の切り出しを行う前に空間的相関を抽出し、企業を単位とした相関関係を捉えるようなモデル構造を設計した. その上で、前述の第 1~第 3 フェーズの機能を実現させ

て、多変量説明変数による時系列が複数与えられた状況での予測をより適切に行うことを狙いとした。それに対して、Liu ら [10] の DSTP モデルの研究においては、多変量説明変数による時系列が一つ与えられた状況が想定されている。この DSTP モデルを本論文で想定する多変量説明変数による時系列が複数与えられた状況下で用いる場合、同一企業の複数の部分時系列と他企業の部分時系列が同列に扱われて学習されることになる。提案モデルは、企業を単位とした時系列によって学習される点が特徴的であると言える。

5 実験

5.1 実験設定

第 4 章で述べた提案モデルの有効性を評価するため、従来の DSTP モデル [10] および単純な LSTM モデルと比較する。本実験では対象業種として「卸売業」(Whole sale and trade) を仮定する。また、予測対象の目的変数を総資本経常利益率 (ROA) すなわち $Y = X_1$ とし、説明変数を $X = \{X_0, X_2, X_3, \dots, X_{11}\}$ とする。本実験は検証実験とテスト実験からなる。いずれの場合も、実験に用いるデータセットの時系列において開始期から固定幅ウィンドウの部分時系列を抽出し、次に 1 期先にスライドさせて同じく固定幅ウィンドウの部分時系列を抽出することを順次繰り返すことにより、合計 15 の部分時系列を作成する。

検証実験 40 期からなる部分時系列セット ($L = 40$) によりモデルを学習し、それぞれの 1 期先を予測する際の誤差に基づいて、モデルのハイパーパラメータを最適化する。このとき、ハイパーパラメータとして特に重要な中間ユニット数 (隠れ状態数) U に着目する。その探索範囲は $U \in \{16, 32, 64, 128\}$ とする。

テスト実験 検証実験で選択したハイパーパラメータを用いて、41 期からなる部分時系列セット ($L = 41$) によりモデルを学習し、それぞれの 1 期先を予測する際の誤差の平均に基づいて予測性能を評価する。

その他の実験設定と評価指標 中間ユニット数 U 以外のハイパーパラメータについては、時間ウィンドウの長さを $T = 12$ 、エポック数を 500、学習率を 0.001、バッチサイズを 64 とした。また、単純な LSTM モデルにおいて中間層の数は 2 とした。これらは検証実験の過程で経験的に選択した。

従来の DSTP モデルと単純な LSTM モデルでは、すべての入力時系列が四半期次である必要があるため、マクロ経済時系列として用いる日経平均の月次データについて 3ヶ月分の平均をとることで四半期次とした。

本実験における評価指標には平均二乗誤差 (mean squared error: MSE) を用いた。

5.2 結果と考察

表 3 に、各種設定で提案モデル、DSTP モデル、および単純な LSTM モデルを用いた平均二乗誤差 (MSE) および標本標準偏差 (SD) による評価結果を示す。表中の「Data used」には各モデルに用いた説明変数時系列の種類を示す。なお、これらに加えてすべてのモデルに (既知の) 目的変数時系列を用いている。全業種の財務時系列 A と対象業種の財務時系列 I は、 $I \subset A$ の関係にあることに注意する。また、「# of units」には検証実験を経て選択された各モデルの中間ユニット数を示す。以下では、表 3 の実験結果に基づいて 3 つの観点から考察と貢献を述べる。なお、比較評価の際はデータ量に関して公平を期している。評価の結果、以下に示す比較のすべてにおいて、ウィルコクソンの符号付き順位検定による統計的な有意差がみられた (有意水準 0.05)。

マクロ経済時系列を反映したモデル構造による効果 マクロ経済時系列として日経平均 (N) と日本の GDP (G) を反映した Model 1 はそうでない Model 3 を上回っている。これは提案モデルが適切にマクロ経済時系列を活用できていることを示唆する。

業種に特化した構造による効果 まず初めに、単純 LSTM モデルにおいて、対象業種の財務時系列を用いた Model 9 の予測性能が、全業種の財務時系列を用いた Model 8 のそれを上回った。これは LSTM において業種の動向に着目することが有効であるという仮説の妥当性を示唆する。一方、DSTP モデルにおいては、対象業種の財務時系列を用いた Model 5 は全業種の財務時系列を用いた Model 4 を下回った。これは対象業種の財務時系列に限定した場合に学習のためのデータ量が十分でなくなり、そのことが単純 LSTM よりも複雑な構造をそなえた DSTP モデルにおいて顕在化したであろうことが理由として考えられる。また、全業種の財務時系列 A を含めたすべての時系列データを用いた提案モデルの Model 1 は、 A によらない Model 2 を上回った。これは提案モデルにおいて式 (12) および (13) による全業種の集約が、すべての時系列データの特徴をモデルに組み込む上で有効であることを示唆する。

多変量説明変数による時系列が複数与えられた状況を考慮してモデル化することの効果 最後に、すべての時系列データを用いた提案モデルの Model 1 は DSTP モデルの Model 4 および単純 LSTM モデルの Model 7

表 3: 種々の条件下での提案モデルと比較対象モデルによる平均二乗誤差 (MSE) と標本標準偏差 (SD) .

Model IDs	Models	Data used	# of units	MSE	SD
Model 1	Proposed model	(I), A, N, G	16	2.08×10^{-4}	8.48×10^{-5}
Model 2	Proposed model (w/o 'A')	I, N, G	32	2.11×10^{-4}	7.43×10^{-5}
Model 3	Proposed model (w/o 'N' and 'G')	(I), A	128	2.28×10^{-4}	7.64×10^{-5}
Model 4	DSTP-based model	A, N, G	16	2.12×10^{-4}	6.47×10^{-5}
Model 5	DSTP-based model	I, N, G	16	2.26×10^{-4}	9.40×10^{-5}
Model 6	DSTP-based model	A	16	2.11×10^{-4}	7.43×10^{-5}
Model 7	Simple LSTM model	A, N, G	16	3.86×10^{-4}	6.07×10^{-4}
Model 8	Simple LSTM model	A	16	3.10×10^{-4}	3.72×10^{-4}
Model 9	Simple LSTM model	I	16	2.95×10^{-4}	3.43×10^{-4}

を上回った。その一方で、マクロ経済時系列を用いない状況下では、提案モデルである Model 3 が DSTP モデルである Model 6 の性能を下回る結果となった。多変量説明変数からなる時系列が複数与えられる状況すなわち企業ごとに複数の財務比率で表現された時系列が企業数だけ与えられる状況において、適切なモデル構造についてさらなる検討の余地がある。

6 むすび

企業財務時系列を包括的に分析・予測するための有用な手段を確立するため、本論文では多変量説明変数で表現された時系列が複数与えられた状況を仮定したモデルを設計し、目的変数の 1 期先の予測を行う問題を想定した検討を行った。この目的の下で、マクロ経済時系列を適切に組み込んだ LSTM に基づく新たなモデルを提案した。また、対象業種に固有の動向を反映しつつ、全業種の企業に関する財務時系列の特徴を組みこむことができるモデル構造を設計した。提案モデルの有効性に関して、次に示す 3 つの貢献が得られた。一つ目に、提案モデルは GDP などのマクロ経済時系列を組み込みつつ対象業種の動向を反映した場合において、従来の DSTP モデルおよび LSTM モデルよりも多変量説明変数からなる複数時系列から目的変数時系列の 1 期先を適切に予測できることを実験的に示した。しかしながら、財務時系列のみを用いて予測を行った場合においては、DSTP モデルの方がより適切に予測できたことから、モデル構造にさらなる改善の余地があると言える。二つ目に、提案モデルにおいてはマクロ経済時系列を適切に反映して予測に活用することができた。三つ目に、提案モデルでは、対象業種に固有の動向を反映しつつ、全業種の企業財務時系列の特徴を組みこむことができた。なお、本論文の実験では対象業種として卸売業を取りあげ、目的変数時系列として ROA に着目したが、他の業種や他の目的変数についての検討は今後の課題である。

謝辞 本研究の一部は科学研究費補助金基盤研究 (B) (15H02703) の援助による。

参考文献

- [1] Amini, M. H., Kargarian, A. and Karabasoglu, O.: ARIMA-based decoupled time series forecasting of electric vehicle charging demand for stochastic power system operation, *Electric Power Systems Research*, Vol. 140 (2016), 378–390.
- [2] Choi, E., Bahadori, M. T., Schuetz, A., Stewart, W. F. and Sun, J.: RETAIN: Interpretable predictive model in healthcare using reverse time attention mechanism, *Advances in Neural Information Processing Systems*, Vol. 29 (2016), 3504–3512.
- [3] Goodfellow, I., Bengio, Y. and Courville, A.: *Deep Learning*, MIT Press, 2016.
- [4] Hochreiter, S.: Untersuchungen zu dynamischen neuronalen Netzen, Diploma thesis, Institut für Informatik, Technische Universität München, 1991.
- [5] Hochreiter, S., Bengio, Y., Frasconi, P. and Schmidhuber, J.: Gradient flow in recurrent nets: the difficulty of learning long-term dependencies, in Kremer, S. C. and Kolen, J. F. eds., *A Field Guide to Dynamical Recurrent Neural Networks*, IEEE Press, 2001.
- [6] Hochreiter, S. and Schmidhuber, J.: Long Short-Term Memory, *Neural Computation*, Vol. 9 (1997), 1735–1780.
- [7] Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, in *Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015)*, 2015.
- [8] Li, H., Shen, Y. and Zhu, Y.: Stock Price Prediction Using Attention-based Multi-Input LSTM, *Proceedings of Machine Learning Research*, Vol. 95 (2018), 454–469.
- [9] Liang, Y., Ke, S., Zhang, J., Yi, X. and Zheng, Y.: GeoMAN: Multi-level attention networks for geosensory time series prediction, in *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI-18)*, 2018.
- [10] Liu, Y., Gong, C., Yang, L., Chen, Y.: DSTP-RNN: A dual-stage two-phase attention-based recurrent neural network for long-term and multivariate time series prediction, *Expert Systems with Applications*, Vol. 143 (2019).
- [11] Qin, Y., Song, D., Cheng, H., Cheng, W., Jiang, G. and Cottrell, G. W.: A dual-stage attention-based recurrent neural network for time series prediction, in *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI-17)*, 2017.