DEIM Forum 2018 F3-5

# 時系列金融ネットワークの深層表現による金利相関分析

川上　雄大† 　江口　浩二††

† 神戸大学工学部情報知能工学科　〒 657–8501 兵庫県神戸市灘区六甲台町 1-1
†† 神戸大学大学院システム情報学研究科情報科学専攻　〒 657–8501 兵庫県神戸市灘区六甲台町 1-1
E-mail: †yuta@cs25.scitec.kobe-u.ac.jp, ††eguchi@port.kobe-u.ac.jp

**あらまし**　現実世界には, ノード (頂点) とリンク (辺) で構成されるネットワークで表現可能な関係データが数多く存在する. これらを解析しネットワークの潜在的な表現などを獲得することが課題となっている. 本稿では, 生成的確率ネットワークに基づいたネットワーク分析手法に改善を加え, さらに, 深層表現を用いたリンク間の属性値の予測モデルを新たに提案する. 生成的確率ネットワークは自己符号化器を拡張したモデルであり, 入力データの特徴を捉えつつ次元が圧縮された深層表現を獲得できる. ネットワークが有向グラフであるとき, リンクを隣接行列で表すとする. この隣接行列の深層表現を獲得することを試みたとき, 行もしくは, 列に着目した次元圧縮で異なる深層表現が獲得できる. これらを用いてリンク間の属性値を予測する. 本稿では, 時系列ネットワークである銀行間の取引を表した金融データを用い, 取引の貸し手側と借り手側に着目した深層表現を月ごとに獲得する. 銀行間で取引を行うとき, 片方が貸し手側, もう一方が借り手側であることから, 2 種類の深層表現を用いて取引における金利を予測するモデルを構築する. 構築したモデルの予測精度を平均絶対パーセント誤差などで評価した. 結果として元データを扱うより深層表現を用いた方が予測精度が良いことが確認できた.
**キーワード**　自己符号化器, 金融ネットワーク

## 1. はじめに

関係データをネットワークで表現し, 計算的な解析を行うことでコミュニティの検出やノード間の属性値の予測などが可能となる. ネットワークの構造を分析する手法は様々であるが, 本稿では自己符号化器 (Autoencoder: AE) [1] の一種である生成的確率ネットワーク (Generative Stochastic Networks: GSN) [2] に着目したネットワーク分析を行う. 自己符号化器はデータの次元圧縮を行うモデルであり, 本来のネットワークの表現から次元を削減した潜在表現を介して, 本来のネットワーク表現を再構築する. したがって, 得られた潜在表現はネットワークの潜在的な特徴を表しているといえる. ノード間の属性値の予測などを行う際, 本来のネットワークの表現よりも潜在表現を扱うことで精度の改善が見込めると考える.

筆者らの先行研究 [3] に, 銀行間の取引データを扱った GSN によるネットワーク分析を行ったものがある. これは, 銀行間の取引を時系列的に変化するネットワークと捉え, リーマンショックが起こった 2008 年の欧州の銀行間の取引データを用いて分析を行い, 再構築評価とリンク予測評価という 2 つの評価尺度からモデルを評価している. 再構築評価とは GSN によって再構築された表現が入力表現とどれだけ異なっているか, リンク予測評価とは予め隠しておいたデータをどれだけ予測できるかを評価している.

本稿では, 先行研究をベースに, 再構築評価, リンク予測評価の改善を目指す. 再構築評価における評価値が改善することは, 入力データの特徴をよく捉えた潜在表現を獲得できたことを表す. そこで, 先行研究のモデルの問題点として 3 点あげ, 解決案を提案する. 次に, 獲得した潜在表現を用いて銀行間の取引における金利を予測するモデルを提案する. 金利は, 銀行間の取引時に決められるもので, 両銀行の財務指標を測る尺度と考えられる. 例として, A 銀行が B 銀行からお金を借りたとする. A 銀行の財務状況が良いとき, 返済が十分に見込めることから金利が低く設定される. また, 財務状況が悪いとき, 返済のリスクが高いと判断され金利が高く設定されるといったことが起こりうる. したがって, 金利が予測できることは, 銀行の財務状況を予測できることにつながる. 更に, 過去の傾向を捉えた予測が行えると, 未知の取引の金利や, 未来の銀行の財務状況などを把握できることになる.

本稿では, これらを踏まえて深層表現の獲得と深層表現からの金利予測実験を行った.

## 2. 関連研究

### 2.1 自己符号化器と深層学習

自己符号化器は深層学習 [4] によりデータの有用な表現を学習するためにしばしば用いられる. 本稿で深層表現の獲得に扱うモデルは自己符号化器の拡張モデルである GSN をベースとしたものである. 最も簡単な自己符号化器から始めて GSN までの概要を以下に示す.

#### 2.1.1 自己符号化器

自己符号化器 (Autoencoder: AE) [1] は教師なし学習の一種で, データの次元圧縮を行い, 潜在表現 (Latent representation) を獲得するモデルである. AE の一番簡単なモデルは入力層, 潜在層, 出力層の 3 層から構成されるニューラルネットワークで, 目標出力を入力データ $X$ 自身とする. 構造例を図 1 に示す.

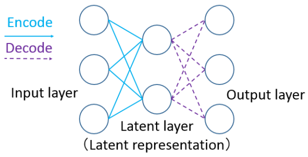


図 1　Structure of Autoencoder.

一般的に, 潜在層の次元数が入力層の次元数以上であると, 恒等写像を学習する. したがって, 潜在層の次元数を入力層の次元数未満に設定することで, 入力データを次元圧縮した潜在表現が獲得できる.

入力層 $X$ に対して, 潜在層 $Y$ は $Y = f_\theta(X) = \phi(WX + b)$, 出力層 $X'$ は $X' = f'_{\theta'}(Y) = \phi'(W'Y + b')$ と定義される. ここで, $W, W'$ は重みパラメータ, $b, b'$ はバイアスパラメータであり, $\theta = (W, b)$, $\theta' = (W', b')$ を表す. また, 関数 $\phi, \phi'$ は活性化関数と呼ばれ, シグモイド関数やランプ関数 (Rectified Linear Unit: ReLU) などが用いられる. 入力層の表現から潜在層の表現を得ることをエンコード, 潜在層の表現から出力層の表現を得ることをデコードと呼ぶ.

パラメータ学習の際の損失関数を入力層と出力層の表現の誤差として, $L = (X, X')$ で定義する. ここでいう学習とは, パラメータの更新を行い, 損失関数 $L$ を最小化することである. パラメータの更新には確率的勾配降下法 [5] などが用いられ, 損失関数には二乗誤差や交差エントロピーなどが用いられる. 損失関数の値が収束し最小化されたときを学習の終了とし, このときの潜在層の表現が入力データを圧縮した潜在表現となる.

自己符号化器の潜在層を増やしたものを深層自己符号化器 (Deep Autoencoder) [1] と呼ぶ. 層を増やしていくことで, ニューラルネットワークの柔軟性が増し, 複雑な特徴を捉えることができる可能性がある. しかし, 層の増加とともにパラメータ数も増加するので, 学習コストが高くなることや学習自体が困難になるといった問題がある. したがって, 入力データに応じて, 適切な層数を設定する必要がある. また, 潜在層が多層のとき, 獲得した潜在表現は深層の表現であることから深層表現とも呼ぶ.

### 2.1.2　雑音除去自己符号化器

自己符号化器の一種に雑音除去自己符号化器 (Denoising Autoencoder: DAE) [6] がある. このモデルは, 予め扱うデータの表現 $X$ に対してノイズを印加することで $\tilde{X}$ を得る. このとき印加するノイズにはガウシアンノイズや Salt and Pepper ノイズなどが用いられる. ノイズが印加された $\tilde{X}$ を入力データの表現とし, エンコードとデコードを介して出力層で再構築された表現 $\tilde{X}'$ を得る. DAE では, 損失関数 $L = (X, \tilde{X}')$ を最小化する. 本来のデータの表現 $X$ と再構築された出力層の表現 $\tilde{X}'$ の誤差を最小化することは, 印加されたノイズを除去する学習を行っているといえる. これにより, ノイズに頑健なモデルが学習できる.

次に, DAE を一般的な確率モデルとして捉える.

$$X_{t+1} \sim P_{\theta_1}(X|\tilde{X}_t), \quad \tilde{X}_{t+1} \sim P_{\theta_2}(\tilde{X}|X_{t+1}) \tag{1}$$

式 (1) は $X$ と $\tilde{X}$ を交互に生成するマルコフ連鎖であり, 生成の流れを図 2 に示す. パラメータセット $\theta_1$ を持つ $P_{\theta_1}$ はノイズが印加された $\tilde{X}_t$ を入力の表現とする DAE のモデル, パラメータセット $\theta_2$ を持つ $P_{\theta_2}$ は入力の表現 $X_{t+1}$ にノイズを印加する破損分布を表す. 式 (1) に入力データとして $X_0$ を与えると, $\tilde{X}_0$, $X_1$, $\tilde{X}_1$, ..., と連鎖的に生成される. このマルコフ連鎖の漸近分布が存在するとき, 漸近分布は入力データの真の生成分布 $P(X)$ に収束する. したがって, $P_{\theta_1}(X|\tilde{X})$ を正しく推定することができれば, 入力データの真の生成分布を得ることができると文献 [2] で述べられている.

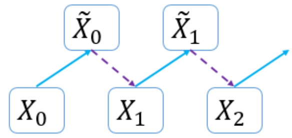


図 2　Markov chain in Denoiding Autoencoder.

このマルコフ連鎖は, 生成を繰り返していくに連れ誤ったモードを推定する可能性がある. これを回避するために, マルコフ連鎖の過程で得られた $\tilde{X}_t$ も訓練データとして用いることで誤ったモードの推定を元のモードに引き戻す Walkback アルゴリズムが提案されている. Walkback アルゴリズムを用いた DAE は元の DAE よりも性能が良いことが文献 [2] で示されている.

### 2.1.3　生成的確率ネットワーク

DAE を更に拡張したモデルに生成的確率ネットワーク (Generative Stochastic Networks: GSN) [2] がある. これは, DAE に潜在層を加えることで, より柔軟で複雑な構造をもてるニューラルネットワークを構築している. GSN のマルコフ連鎖は潜在層の表現を $H_t$ とすると式 (2), 図 3 のように示される.

$$H_{t+1} \sim P_{\theta_1}(H|H_t, X_t), \quad X_{t+1} \sim P_{\theta_2}(X|H_{t+1}) \tag{2}$$

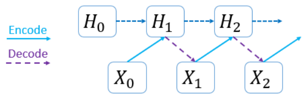


図 3　Markov chain in Generative Stochastic Networks.

次に本稿のモデルのベースとなっている潜在層が 2 層である GSN を図 4 に示す. ここで, $W_1, W_2$ は重み, $b_0, b_1, b_2$ はバイアスのパラメータである. ただし, $W_1$, $W_2$ はそれぞれ 0 層目から 1 層目, 1 層目から 2 層目への重み, $b_i$ は $i$ 層目でのバイアスを表す. GSN の学習では Walkback アルゴリズムが適用され, 図 4 では Walkback 回数 $T$ を 3 回としている. 各層の更新は式 (3)〜(5) で示される.

$$X_t^0 = \phi(W_1^T H_t^1 + b_0) \tag{3}$$

$$H_t^1 = \phi(W_1 X_{t-1}^0 + W_2^T H_t^2 + b_1) \tag{4}$$

$$H_t^2 = \phi(W_2 H_{t-1}^1 + b_2) \tag{5}$$

式 (3)〜(5) に従って学習を重ね, 再構築された表現が元のデータの表現に近づくようにパラメータを更新していく. GSN は DAE と同様に元のデータの表現にノイズを印加したものを入力データの表現としている. また, 潜在層の表現は活性化関数に代入する前後にノイズを印加する [7].

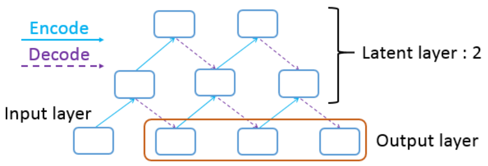


図 4　Structure of GSN with multiple layers.

以上をまとめて, GSN の学習アルゴリズムを簡単に示す.

(1) 元のデータの表現を $X$ とし, これにノイズを印加して得られる表現を $\tilde{X}$ とする. (2) $\tilde{X}$ を GSN の入力とし, $X_0^0 = \tilde{X}$ とする. (3) 入力層のある一番下の層を 0 層目とし, 奇数番目の層を偶数番目の層からサンプリングすることで更新する. (4) 偶数番目の層を奇数番目の層からサンプリングすることで更新する. (5) 手順 3 と 4 を合わせたものを層の更新 1 回分とし, これを Walkback 回数の $T$ 回分行う. (6) $X$ と再構築された表現 $X_t^0$ $(t \in T)$ の損失関数を $L$ とすると, 目的関数は $\frac{1}{T}\Sigma_{t=1}^{T} L(X, X_t^0)$ となる. これが最小化されるようにパラメータを更新する. (7) 目的関数の値が収束するまで手順 3 ～ 6 を繰り返す.

### 2.2　回帰モデル

本稿では, 金利を目的変数, 深層表現を説明変数とした線形回帰モデルの学習を行う. ここでは, 線形回帰モデルと, モデルの学習に関連する手法を紹介する.

#### 2.2.1　線形回帰モデル

回帰とは, ある 2 つの変数 $X$, $Y'$ 間に $Y' = f(X)$ となる式 (モデル) を当てはめることである. $Y'$ を目的変数 (従属変数), $X$ を説明変数 (独立変数) と呼ぶ. $X = [x_0, x_1, ..., x_{n-1}]$ と表されるとき, 各説明変数の重みを $W = [w_0, w_1, ..., w_{n-1}]$, バイアス項を $b$ とおく. このとき, $Y' = WX^T + b = w_0x_0 + w_1x_1 + \cdots + w_{n-1}x_{n_1} + b$ となる. これを線形回帰モデルと呼ぶ.

次に, 線形回帰モデルを学習する方法を説明する. 真値 $Y_i$ と説明変数のデータ $X_i$ からなる組が $N$ 個あるとき, $\Sigma_{i=0}^{N-1} L(Y_i, Y_i')$ が最小化されるようにパラメータを更新する. ここで, 損失関数 $L$ には二乗誤差を用いる. また, 過学習とよばれる訓練データに過剰に適合したモデルとなることを避けるために, 目的関数に $L_2$ 正則化項を加えた学習を行う. このときの目的関数は式 (6) と定義され, $\gamma_i$ は各項の比重を決める超パラメータである. ここで, $||\cdot||_F$ はフロベニウスノルムであり, 行列 $A$ の各要素を $a_{i,j}$ としたとき, $||A||_F = \sqrt{\Sigma a_{i,j}^2}$ となる.

$$COST = \gamma_0 \sum_{i=0}^{N} L(Y_i, Y_i') + \gamma_1 ||W||_F^2 + \gamma_2 ||b||_F^2 \qquad (6)$$

#### 2.2.2　交 差 検 証

超パラメータを格子探索で求める際に, 交差検証と呼ばれる手法が用いられることがある. 訓練データによる学習を行ったモデルは未知データに対しても正しい値を予測することが求められる. そこで, 訓練データを $K$ 分割し, $k_i (i \in K)$ のデータセットを valdation set(モデルの精度確認用の未知データの集合), それ以外を training set(学習に扱う訓練データの集合) と分割し, 組み合わせを変えてモデルの学習を $K$ 回行う. モデルの評価を各 validation set の評価値の平均とすることで, 未知データに対しても精度がよい超パラメータを決定できる.

## 3.　GSN によるネットワーク分析

筆者らの先行研究 [3] で提案されている時系列データを扱う際の深層表現学習のモデルについて紹介する.

### 3.1　モデルの構造

まず, 先行研究の時系列データである銀行間の取引データを扱った GSN をベースとするモデルを紹介する. モデルの構造を図 5 に示す. 潜在層 $K$ が 2 層, Walkback 回数 $T$ が 5 回, 印加ノイズにガウシアンノイズと Salt and Pepper ノイズを用いている. 活性化関数にはすべてシグモイド関数を用い, パラメータの更新には確率的勾配降下法を適用している. 取引量を重みとしたリンクを持つ金融ネットワークのノード数 (銀行数) を $N$ としたとき, 隣接行列を $S = \{s_{i,j}\}$, $i, j \in N$, 入力層の表現を $X_t^k = \{X_{it}^0\}$ $(i \in N)$, 再構築された表現を $X_t^{k\prime} = \{X_{it}^{0\prime}\}$ $(i \in N)$ とする. ここで, $s_{i,j}$ は銀行 $i$ から銀行 $j$ に対して取引を行った量を表す. 入力層, 潜在層 1 層目, 潜在層 2 層目の次元数を $N, N_1, N_2$ $(N > N_1 > N_2)$ とすると, $N$ 次元の表現を $N_1$ 次元に圧縮する重みパラメータ $W_{0,1}$ とバイアスパラメータ $b_1$ が定義でき, 同様に $W_{1,2}, b_0, b_2$ も定義できる. また, 重みパラメータに関しては重み共有の考えを用いており, $N_1$ 次元から $N$ 次元へ再構築する際は重みパラメータとして $W_{0,1}^T$ を扱う. 最小化を行う目的関数 $C$ は, 式 (7) である.

$$C = \frac{1}{T} \sum_{t=1}^{T} ||(X_t^{0\prime} - X_t^0) \odot B||_F^2 + \gamma \sum_{k=0}^{K-1} ||W_{k,k+1}||_F^2 \quad (7)$$

ここで, $B$ とは $s_{i,j} = 0$ ならば $b_{i,j} = 1$, $s_{i,j} > 0$ ならば $b_{i,j} = \beta(> 1)$, $s_i = \Sigma_j s_{i,j} = 0$ ならば $b_i = 0$ とし, 隣接行列で表されるリンクに比重を持たせた学習を行う. $\odot$ はアダマール積の演算子で同サイズの行列に対して, 要素ごとに積をとるものである. $A, B$ の各要素を $a_{i,j}, b_{i,j}$ とすると, $(A \odot B)_{i,j} = a_{i,j} \cdot b_{i,j}$ となる.

### 3.2　時系列プレトレーニング

次に, 文献 [3] で提案されている時系列プレトレーニングについて説明する. 深層学習において, 重みパラメータなどの初期値の設定は重要である. そこで, 時系列データを扱う際に, ある時刻 $(t-1)$ で学習したモデルのパラメータセットを, 次の時刻 $t$ での学習を行う際のパラメータセットの初期値として設定する. これを時系列プレトレーニングと呼ぶ. これにより, 時刻 $t$ 以前の傾向を捉えたパラメータを初期値として用いることができ, 時系列データの時区間の依存性や過去の傾向をより詳細に捉えた学習が行えると考えられる.

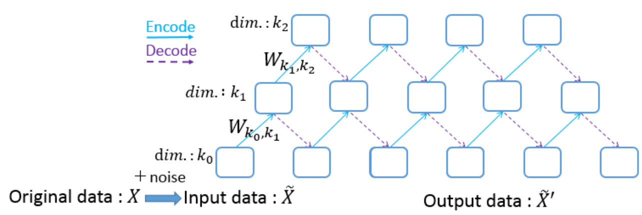


図 5　Structure of the model based on GSN.

## 4. 提 案 手 法

### 4.1 深層表現学習の提案モデル

#### 4.1.1 既存モデルの問題点

先行研究として前章にて紹介したモデルを既存モデルとし，調整や変更を行うことで評価値の改善を試みた．既存モデルの問題点を解決したものを本稿での深層表現学習の提案モデルとする．まず，既存モデルの問題点を述べていく．

第一に，既存モデルでは，超パラメータの調整などを行っているが，式 (7) の目的関数が各項ともに総和をとっているため，扱うネットワークデータのノード数に依存した超パラメータ設定であるといえる．実際に，扱うデータの期間やデータ自体が変わるとネットワークのノード数 (銀行数) が変化する．例として，$A_1$ を $N \times N$ 行列の入力表現，$B_1$ を $N \times N_1$ 行列の重みパラメータとすると，目的関数 $COST_1 = \alpha||A_1||_F^2 + \beta||B_1||_F^2$ が定義できる．次に，$A_1$ を $2N \times 2N$ 行列の入力表現，$B_2$ を $2N \times N_1$ 行列の重みパラメータとすると，目的関数 $COST_2 = \alpha||A_2||_F^2 + \beta||B_2||_F^2$ が定義できる．この 2 式より，扱うネットワークのノード数によって，第 1 項と第 2 項の目的関数への比率が異なることがわかる．

第二に，既存モデルでは印加ノイズにガウシアンノイズと Salt and Pepper ノイズを併用している．ガウシアンノイズはガウシアン分布からサンプリングした値をノイズとして印加するものである．Salt and Pepper ノイズは二値データにノイズを印加するときなどによく用いられ，破損確率 $p$ をもとに，行列 $A$ の各要素 $a_{i,j}$ を破損させるか決める．破損させるなら，$a_{i,j} \in \{0, 1\}$ とすることでノイズを印加している．したがって，Salt and Pepper ノイズはガウシアンノイズより影響力が高いノイズ印加といえる．本稿で扱うデータは二値で表されるデータではなく，活性化関数もシグモイド関数であるため，影響力の高い Salt and Pepper ノイズは不適切なノイズではないかと考える．

第三に，既存モデルでのバイアス項についてだ．まず，ある $n$ 次元データ $X_i = [x_{i,0}, x_{i,1}, ..., x_{i,n-1}]$ について考える．これを次元削減して $n_1$ に落とし込むとする．このとき，$n \times n_1$ の重みパラメータ $W$ を設定する．$Y_i = [y_{i,0}, y_{i,1}, ..., y_{i,n_1-1}]$ を出力の $n_1$ 次元データとしたとき，$Y^T = W^T X^T$ となる．そして，説明変数 $X_i$ に依存しないグローバルな関係性をバイアス項で捉えようとする．バイアス項を $b = [b_0, b_1, ..., b_{n_1-1}]$ と設定することで，$Y_i^T = W^T X_i^T + b^T$ となる．既存モデルでは，$X_i$ で表されるデータが $n$ 個あるとして，$Y = [Y_0, Y_1, ..., Y_{n-1}]$，$X = [X_0, X_1, ..., X_{n-1}]$ としたとき，$Y = XW + Bias$ となる．このときの $Bias$ は $n$ 個のデータに共通するバイアス項となっており $Bias = [b, b, ..., b]$ である．これは $n$ 個のデータの背景にある関係性を捉えたものといえる．しかし，扱う取引データはリンク密度が小さく，取引が存在しないというデータが多い．そのため，実際に取引が行われたというデータが軽視される可能性がある．再構築を目指した学習を行うのであれば，バイアス項をより柔軟なものにするべきであると考える．

#### 4.1.2 既存モデルの問題点の解決案

第一の問題の解決案として，既存モデルの式 (7) の目的関数 $C$ を以下の式 (8) に変更する．

$$C = \frac{1}{T}\sum_{t=1}^{T} M(||(X_t^{0\prime} - X_t^0) \odot B||_F^2) + \gamma \sum_{k=0}^{K-1} M(||W_{k,k+1}||_F^2) \quad (8)$$

ここで，$M(\cdot)$ とは行列 $A$ が $N_1 \times N_2$ 行列であるとき，$M(||A||_F^2) = \frac{1}{N_1 * N_2}||A||_F^2 = \frac{1}{N_1 * N_2}\Sigma a_{i,j}^2$ と行列の要素数でフロベニウスノルムを割り，要素の平均値を求める関数とする．これにより，入力データのノード数に影響されることがなく，各項の比重を表した超パラメータの設定が行えるのではないかと考える．

第二の問題の解決案として，銀行の取引データは非常にスパースである．期間全体で min-max normalization を用いて正規化を行ったとき，取引量の分布は，図 6 のようになる．ただし，図 6 では取引が存在しない，すなわち取引量が 0 であるリンクは除いている．非常に分布が偏っており，印加するノイズを大きくしすぎると取引データの原型を留めないものとなってしまう可能性がある．そこで，Salt and Pepper ノイズを印加ノイズとして扱うことを止め，ガウシアンノイズのみを用いることにする．これにより，適度なノイズを印加することができると考える．

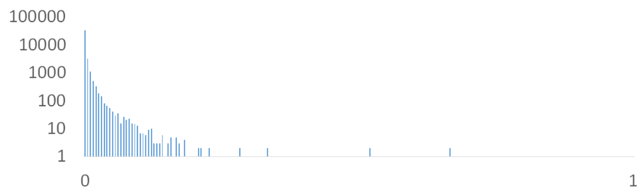


図 6 Histgram of the transaction amount.

第三の問題の解決案として，より柔軟な表現が可能なバイアス項を設定することである．そこで，$B = [b_0, b_1, ..., b_{n_1-1}]$ をデータ $X_i$ ごとに用意して，$Bias = [B_0, B_1, ..., B_{n-1}]$ とする．これを用いると，$Y = XW + Bias$ が定義できる．これは説明変数 $X_i$ に依存しないバイアス項が各要素ごとに求められることになる．したがって，バイアス項は表現力の高いものとなり，既存モデルで求められていた共通するバイアスも各要素ごとに含んだものとなると考える．

以上の 3 点を既存モデルに対する改善案として提案する．

### 4.2 2 種類の深層表現を用いた予測モデル

ネットワークが無向グラフであるとき，ネットワーク構造を表した隣接行列は対称行列となるが，有向グラフでは非対称行列となる．非対称行列は列に着目した場合と行に着目した場合で，表現するものが異なる．行に着目すると，あるノードから他のノードへのリンクが存在するかを，列に着目すると，あるノードに他のノードからのリンクが存在するかを表している．したがって，行に着目した次元圧縮を行うと，他のノードへのリンクが存在するかという深層表現を，列に着目した次元圧縮を行うと，他のノードからのリンクが存在するかという深層表現を獲得できる．ここでは，この 2 種類の深層表現を用いて，リンク間の属性値を予測する線形回帰モデルを提案する．本稿では予測する属性値が金利となっている．

**4.2.1　金利予測モデル**

貸し手着目の学習で得られた融資パターンと考えられる深層表現と，借り手着目の学習で得られた借金パターンと考えられる深層表現から，取引における金利を予測する．なお，貸し手着目，借り手着目については 5.1 節にて詳しく述べる．具体的には，$A-B$ 間の金利を予測するとき，$A$ 銀行の融資パターンと $B$ 銀行の借金パターンを組み合わせて金利を予測する．融資パターン，借金パターンが互いに $m$ 次元で表されているとき，ある銀行 $i$ の融資パターンをベクトル $I$，ある銀行 $j$ の借金パターンをベクトル $J$ とし，$X_{i,j}=[IJ]$ という $n=2m$ 次元からなる深層表現ベクトル $X_{i,j}$ を作成する．重みパラメータ $W$ は取引における金利を予測するため $n\times 1$ の行列とする．バイアスパラメータは深層表現 (説明変数) では説明できないものを捉えるものとして，$Bias=b$ と設定する．予測値 $Y'$ は，$Y'_{i,j}=X_{i,j}W+Bias$ で表される．

このモデルの学習方法について説明する．学習データを $K$ 分割して，$K$ 分割交差検証を行うことで超パラメータを設定する．パラメータの更新には勾配降下法などを用いる．training set が $n$ 個のデータから成るとき，式 (9) を目的関数として学習を行う．

$$COST=\alpha\cdot M(||Y'-Y||_F^2)+\beta\cdot M(||W||_F^2)+\gamma\cdot M(||Bias||_F^2) \quad (9)$$

ここで，$Y'$ は予測値，$Y$ は真値，$W$ は重みパラメータ，$Bias$ はバイアスパラメータで，第 2 項と第 3 項は正則化項である．

## 5．実　　験

この章では，深層表現学習の提案モデルの性能評価実験と金利予測モデルの超パラメータの決定と性能評価実験を行う．深層表現学習は再構築評価とリンク予測評価という 2 つの観点からケンドール順位相関係数を用いて評価する．金利予測モデルの超パラメータは 5 分割交差検証を用いた格子探索で決定し，モデルの性能評価には 3 つの評価尺度を用いる．最後に，各結果について考察する．

### 5.1　データセット

本稿では，時系列ネットワークの一つである金融データに着目した実験を行った．実験に用いた金融データのデータセットは欧州債務危機と呼ばれる時期を含む，2009 年 7 月 1 日から 2012 年 12 月 31 日の欧州の銀行間での取引データである．このデータセットは，取引の契約日，取引を持ちかけられた側の銀行名 (Quoter)，取引を持ちかけた側の銀行名 (Agressor)，取引金額，取引の状態，取引の金利などから構成される．この期間の内，取引が行われた銀行は全部で 153 行，総取引件数は 162,075 件となる．取引金額は百万ユーロ単位で示されており，取引の状態は”Sell”と”Buy”の 2 種類がある．

$A$ 銀行から $B$ 銀行へ取引による金額が動いたとき，ネットワークの $A-B$ 間のリンクに取引金額分を重みとしてのせる．本稿では，各月ごとに集計し，$153\times 153$ の隣接行列を作成した．以下，この隣接行列を $S$ とし，$i$ 行 $j$ 列の要素を $S_{i,j}$ で表す．また，この隣接行列に対して，期間全体で min-max normalization を用いて正規化を行い，$S_{i,j}\in[0,1]$ とした．各月ごとのリンク密度を図 7 に示す．リンク密度 $d$ はリンク数を $e$，ノード数を $N$ とすると，$d=e/(N\times N)$ で定義される．リンク密度が 5% を下回っていることから，非常にスパースな隣接行列であることがわかり，2011 年 9 月頃から減少傾向がみられる．また，未知のリンクを予測できるかというリンク予測評価を行うために，実験では隣接行列 $S$ のリンクを予め 15%削除したものを深層表現学習の入力データとして扱う．これに応じて，式 (8) の $B$ を次のように設定した．$s_{i,j}=0$ ならば $b_{i,j}=1$，$s_{i,j}>0$ ならば $b_{i,j}=\beta(>1)$，そして，未知のリンクとして削除したリンクであれば $s_{i,j}=0$ とした．バイアス項を柔軟なものにしたため，$B$ の制約を既存モデルよりも緩めた．

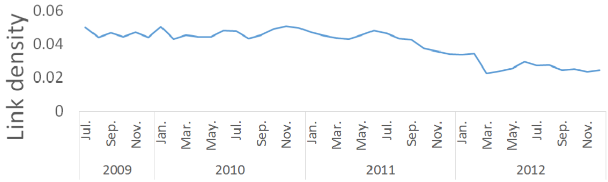


図 7　Link density in data set.

次に，隣接行列の行に着目する．隣接行列の行は取引の貸し手側を示しており，ある銀行が他の銀行に融資を行ったということを表している．これは，融資パターンを表した表現と捉えることができる．この表現を GSN で次元圧縮することで，融資パターンの深層表現を獲得できる．また，取引の借り手側である列に着目すると，ある銀行が他の銀行から融資を受けた，すなわち，借金パターンを表した表現と捉えることができる．この表現も同様に GSN で次元圧縮することで，借金パターンの深層表現が獲得できる．融資パターンの深層表現を獲得する実験を貸し手着目 (lender-focused)，借金パターンの深層表現を獲得する実験をを借り手着目 (borrower-focused) と呼ぶ．

取引における金利の隣接行列も同様に作成する．ただし，期間内にあるノード間で複数回の取引を行っていたとき，ノード間の金利は平均値とする．この隣接行列を $R$ とし，各要素を $R_{i,j}$ で表す．また，各月ごとの金利の平均値を図 8 に，期間全体の取引における金利の分布をヒストグラムで図 9 に示す．

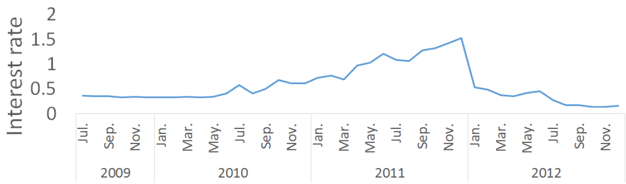


図 8　Average of monthly interest rates.

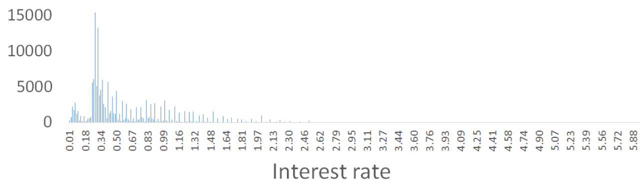


図 9　Relative frequency distribution of interest rates.

### 5.2 評　　価

本稿で用いた評価尺度について紹介する．深層表現学習の評価にケンドール順位相関係数, 金利予測モデルの評価に回帰誤差を測る尺度を用いた．

#### 5.2.1 ケンドール順位相関係数

ケンドール順位相関係数 $\tau$2 つのランキングデータ $X = [x_0, x_1, ..., x_{n-1}]$, $Y = [y_0, y_1, ..., y_{n-1}]$ の相関を測る尺度である． データ対 $(x_i, y_i), (x_j, y_j)$ を取り出す． これは, ${}_nC_2 = n(n-1)/2$ 通り存在する．$(x_i - x_j)(y_i - y_j) > 0$ のとき $P_s$ に 1 追加, $(x_i - x_j)(y_i - y_j) < 0$ のとき $P_r$ に 1 追加する．また, $X$ において同順位の数を $U_X$, $Y$ において同順位の数を $U_Y$ とし, $X$ と $Y$ 両方において同順位であるデータ対は $U_X$ にも $U_Y$ にも加えない．ケンドール順位相関係数は式 (10) で定義される．$\tau \in [-1, 1]$ で, $|\tau|$ が大きいほど強い相関があることを示す．

$$\tau = \frac{P_s - P_r}{\sqrt{P_s + P_r + U_X} \cdot \sqrt{P_s + P_r + U_Y}} \tag{10}$$

#### 5.2.2 回帰誤差を測る尺度

回帰モデルによる予測値と真値の誤差を測る尺度を紹介する．真値を $y_i$, 予測値を $y'_i$, データ数が $N$ とする．平均二乗誤差 (Mean Squared Error: MSE), 平均平方二乗誤差率 (Root Mean Squared Percentage Error: RMSPE), 平均絶対パーセント誤差 (Mean Absolute Percentage Error: MAPE) は, それぞれ式 (11), (12), (13) で定義される．

$$MSE(Y, Y') = \frac{1}{N} \sum_{i=0}^{N-1} (y_i - y'_i)^2 \tag{11}$$

$$RMSPE(Y, Y') = 100 \sqrt{\frac{1}{N} \sum_{i=0}^{N-1} \left( \frac{y_i - y'_i}{y_i} \right)^2} \tag{12}$$

$$MAPE(Y, Y') = \frac{100}{N} \sum_{i=0}^{N-1} \left| \frac{y_i - y'_i}{y_i} \right| \tag{13}$$

#### 5.2.3 深層表現学習モデルの評価

深層表現学習のモデルは再構築評価と, リンク予測評価の 2 つの観点から評価する．再構築評価とは, 入力表現から再構築した表現がどれだけランキングを反映させつつ再構築されているかを測るものである．本稿では, 深層表現学習のモデルへの入力データであるリンクを 15%を削除した表現と, 再構築された表現の順位相関をみている．リンク予測評価とは, 予め隠しておいた 15%のデータを未知データとして予測できるかを測るものである．時系列プレトレーニングにより, 過去の傾向も捉えながら再構築できると考え, 次の月の隠しておいた 15%のデータを未知データとして, 当月のパラメータで次の月のデータを再構築し, 再構築表現と未知データの 15%の順位相関をみる．この評価値が 1 に近いほど, 未知データを過去の傾向から予測できることを表している．

#### 5.2.4 金利予測モデルの評価

はじめに, 式 (9) の金利予測モデルの目的関数の超パラメータ $\alpha, \beta, \gamma$ を決定する．ある $t$ 月のデータに対して, 5 分割交差検証を行う．validation set の MSE の値の更新率が 0.01%を下回った時点で学習を終了とする．また, 収束速度もモデルの評価には重要と考え, MSE と収束速度を用いた評価を行う．MSE は小さいほどよいので, 格子探索での MSE の最小値から誤差が 5%以内である超パラメータ設定を候補とする．そして, 収束速度は速いほどよいので, エポック数が小さいほどよい．この 2 点から, MSE × エポック数を格子探索における評価値とし, 評価値が最小である超パラメータ設定を採用し, 実験を行う．この評価尺度により, validation set(未知データ) に対しても予測ができ, 特徴を捉えた学習が効率よく行えているモデルを選ぶことができる．次に, 採用した超パラメータでのモデルの性能を評価する．モデルの学習を $t$ 月で行ったとき, 評価に用いる予測対象は $(t+1)$ 月の実際に取引の行われたデータとする．金利予測モデルの入力データによる予測精度の比較として, 次の 4 つのパターンの貸し手着目, 借り手着目の表現を用いた．学習データで金利予測モデルの学習を行い, テストデータを入力したときの $(t+1)$ 月の金利の予測精度を MSE, RMSPE, MAPE で評価した．(1) 学習データ：$t$ 月の深層表現学習に用いた 85%データ．テストデータ：$(t+1)$ 月の深層表現学習に用いた 85%データ．(2) 学習データ：$t$ 月の学習で得られた深層表現．テストデータ：$t$ 月の深層表現学習で得られたパラメータセットを用いて $(t+1)$ 月の 85%データを次元削減した深層表現．(3) 学習データ：$t$ 月の学習で得られた再構築表現．テストデータ：$t$ 月の深層表現学習で得られたパラメータセットを用いて $(t+1)$ 月の 85%データを再構築した再構築表現．(4) 学習データ：0~1 の乱数データ $(153 \times 153)$．テストデータ：学習データとは異なる乱数データ．

### 5.3 実験内容と結果

深層表現を獲得する実験は, 貸し手着目, 借り手着目を同様に行った．各月のデータを提案モデルの入力として与え, モデルの学習を行い, 各月のデータを再構築するためのパラメータセット $(W_{0,1}, W_{1,2}, b_0, b_1, b_2)$ を獲得する．このパラメータセットを用いて, 入力データに対する深層表現と, 再構築表現を獲得する．これを, 時系列プレトレーニング有り (with pretraining) と時系列プレトレーニング無し (without pretraining) で行う．ただし, 実験の一月目に該当する 2009 年 7 月は時系列プレトレーニングが行えないのでプレトレーニング無しの結果と同値としている．深層表現学習の提案モデルの超パラメータの設定は, 潜在層 $K = 2$, 各潜在層の次元数は $(100, 50)$, Walkback 回数 $T = 5$, $\beta = 40$, $\gamma = 10^{-3}$, エポック数 100 万回, 学習率 0.3, ガウシアンノイズの分散パラメータ $\sigma^2 = 0.01$ である．

貸し手着目, 借り手着目の再構築評価の結果を図 10 に, リンク予測評価の結果を図 11 に示す．それぞれ, 破線が時系列プレトレーニング無しのグラフとなっている．また, 各評価の平均値と標準偏差を表 3 にまとめた．なお, 既存モデルの評価値は文献 [3] を参照した．

再構築評価では, ベースラインである既存モデルよりも評価値が改善していることから提案モデルの有効性が確認できた．また, 平均的に見ると時系列プレトレーニングの有効性も確認できた．事前学習有りが事前学習なしにくらべ評価値が悪くなる原因は取引情報が過去の傾向から大きく変化した月であるこ

とが考えられる．

リンク予測評価でも，ベースラインである既存モデルよりも評価値の改善が見られ，提案モデルの有効性が確認できた．こちらは，どの月も時系列プレトレーニングを行ったほうがよいことがわかり，時系列プレトレーニングの有効性も確認できた．

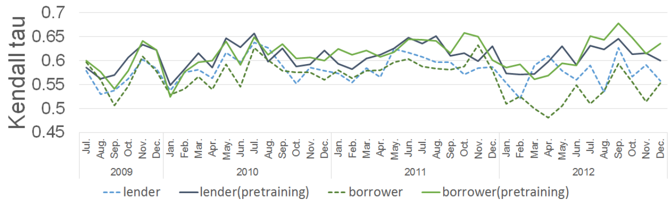


図 10　Reconstruction performance in time-series plots.

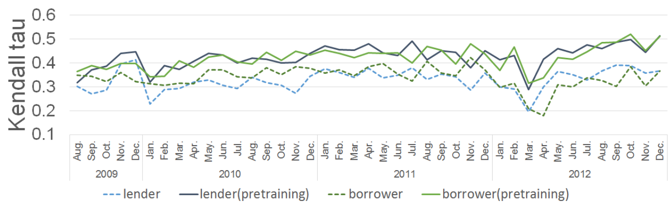


図 11　Link prediction performance in time-series plots.

表 1　Average of performance.

| | Reconstruction | Link prediction |
|---|---|---|
| lender-focused(pretraining) | 0.608 ± 0.026 | 0.426 ± 0.047 |
| lender-focused | 0.580 ± 0.028 | 0.329 ± 0.045 |
| borrower-focused(pretraining) | 0.612 ± 0.032 | 0.422 ± 0.046 |
| borrower-focused | 0.562 ± 0.035 | 0.340 ± 0.045 |
| previous model(pretraining) | 0.433 ± 0.033 | 0.345 ± 0.043 |
| previous model | 0.396 ± 0.065 | 0.308 ± 0.066 |

次に，金利予測モデルの式 (9) の目的関数の超パラメータ $\alpha, \beta, \gamma$ を決定する． $\alpha \in \{1, 10, 100, 500, 1000, 5000, 10000\}$, $\beta \in \{1, 10, 100, 1000\}$, $\gamma \in \{1\}$ で格子探索を行った．実際は，比率を維持して $\alpha + \beta + \gamma = 1$ と総和を 1 にした超パラメータで学習を行った．格子探索に用いたデータセットは，2009 年 8 月の貸し手着目と借り手着目の深層表現と，予測対象である 2009 年 8 月の金利の隣接行列 $R$ である．validation set の MSE の値の更新率が 0.01%を下回った時点で学習を終了している．これは，更新率が 0.01%を下回ったとき，すでに収束し特徴を捉えた学習が十分にできていると考えられるからである．また，有効なパラメータ設定は格子探索における MSE の最小値から悪化度が $-5\%$ 以内のものとする．今回の格子探索の評価方法である，COST = MSE × エポック が最小となったのは，$\alpha = 500, \beta = 100, \gamma = 1$ である．このパラメータ設定を用いて金利予測モデルの性能評価実験を行った．

MSE，RMSPE，MAPE の結果をそれぞれ，図 12，図 13，図 14 に示す．そして，全期間で平均を取ったものを表 2 に示す．乱数データは金利との相関がないため，どの評価尺度でも最も悪い．評価のベースラインは深層表現を学習しなくても予測ができる最低限の精度として 85%データとする．どの評価尺度でも，ベースラインより深層表現による予測性能が優位であるといえる．また，線形回帰のモデルを考えたとき，85%データと再構築データは 306 次元 (貸し手着目の 153 次元 + 借り手着目の 153 次元) の説明変数をもって予測しているのに対し，深層表現は 100 次元 (貸し手着目の 50 次元 + 借り手着目の 50 次元) で予測を行っている．説明変数が多いほど線形回帰モデルは複雑な特徴を捉えた表現力豊かなモデルとなることが考えられるため，この優位性は非常に良い結果だといえる．

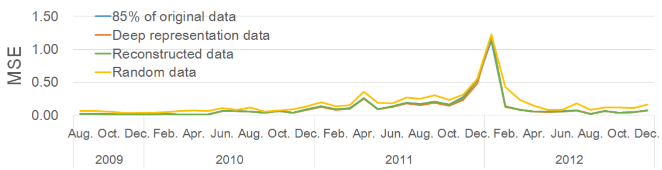


図 12　MSE in time-series plots.

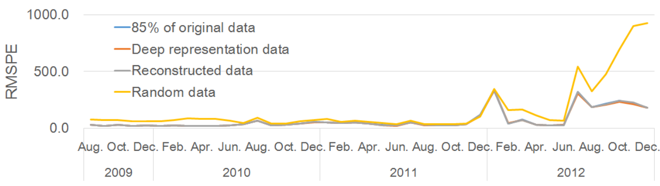


図 13　RMSPE in time-series plots.

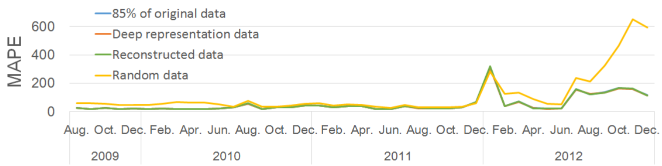


図 14　MAPE in time-series plots.

表 2　Average of each evaluation.

| | MSE | RMSPE | MAPE |
|---|---|---|---|
| 85% of original | 0.1159 | 77.2 | 52.97 |
| Deep representation | **0.1120** | **70.2** | **51.78** |
| Reconstructed | 0.1170 | 72.2 | 53.02 |
| Random | 0.1782 | 161.0 | 112.37 |

## 5.4　考　　察

まず，深層表現学習の結果についての考察を行う．貸し手着目の再構築評価において，2010 年 8 月，2012 年の 3 月，4 月において時系列プレトレーニングによる改善がみられなかった．これは，過去の傾向も反映させた学習が効果的でなかったことを表し，銀行間の取引関係などが大きく変動した月であると考えられる．そこで，データセットの分析を行った．

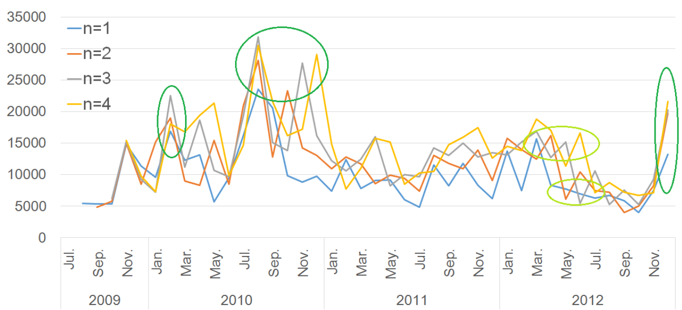


図 15　Frobenius norm of transaction amount for every n-month.

はじめに, $t$ 月と $(t-n)$ 月のデータ $D_t$ と $D_{t-n}$ に対して, $||D_t - D_{t-n}||_F$ を求めた. $n = 1, 2, 3, 4$ とし, 結果を図 15 に示す. 緑色の円は取引データが大きく変化していると思われる月であり, 2010 年 2 月と 8 月, 2012 年 12 月である. 特に 2010 年 8 月のデータは $(t-n)$ が 8 月となる点でも値が増加していることから変化が顕著だといえる. 薄い緑色の円は増加のち減少という傾向が強く見られる. これは, 数ヶ月に渡り状態が安定しなかったと考えられる. 次に, 各月の取引量の平均の変化率を求めた. 結果を図 16 に示す.

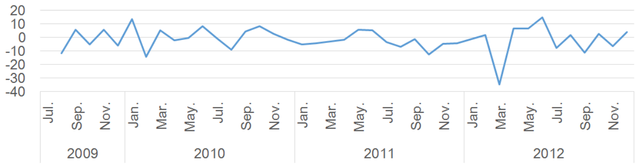


図 16　Change rate of average in transaction amount.

大きな変化として見て取れるのは 2012 年 3 月の急激な減少である. 実際にギリシャでは 2011 年 11 月に政権交代し, 2012 年 6 月には再び選挙を行うなど政治状況が大きく動いた時期で, 他の欧州の数ヶ国でも国債金利の上昇などが確認できた. これらの結果から, モデルの性能は向上したものの, 扱うデータの急激な変化への対応が不十分であるといえる.

次に, 金利予測モデルの結果についての考察を行う. どの評価尺度でも 2012 年 1 月の予測精度が芳しくないことから, 2011 年 12 月 ~2012 年 1 月に着目して考える. 深層表現学習において, 評価尺度や時系列プレトレーニングの有無に関わらず評価値が減少していることが確認できた. また, 図 8 から分かるように金利の平均が 12 月から 1 月にかけて 1.0 以上の減少で, 3 年半を通して最大の減少率である. これにより, 金利予測が 2011 年 12 月 ~2012 年 1 月間で困難になったと考えられる.

また, データ期間を情勢を基準に以下の 4 期に分け深層表現による予測精度の期待値を図 3 に示す. (1) 第 1 期 : 2009/07 ~ 2010/06 ギリシャの財政赤字が発覚するも比較的安定期 (2) 第 2 期 : 2010/07 ~ 2011/12 欧州債務危機ともいわれる変動期 (3) 第 3 期 : 2012/01 ~ 2012/07 沈静化へ向かった時期 (4) 第 4 期 : 2012/08 ~ 2012/12 再び安定期

表 3　Average of performance.

| | MSE | RMSPE | MAPE |
|---|---|---|---|
| First term | 0.019 | 23.5 | 19.26 |
| Second term | 0.138 | 42.3 | 32.43 |
| Third term | 0.235 | 119.9 | 90.58 |
| Fourth term | 0.049 | 203.7 | 138.66 |

これより, 安定期である第 1 期は他の期より精度が良いことが確認でき, データが安定であるならば MAPE より, ±20%以内の精度で予測できることが分かった. また, 第 4 期で MSE だけ改善してるようにみえるのは, 金利の期待値が低いことから二乗しても僅かな誤差であるからと考えられる. 一方で, RMSPE や MAPE は時期における金利の期待値の影響を受けない評価尺度である. したがって, これらを式 (9) の第一項に用いることで予測精度が改善する可能性が考えられる.

## 6. おわりに

本稿では, GSN を用いた深層表現学習のモデルの改善と深層表現学習により得られた表現から金利を予測するモデルを提案した. 深層表現学習では, 既存手法の問題点であった 3 点を解決でき, 再構築評価, リンク予測評価という 2 つの評価手法において提案モデルの有効性を示すことができた. また, 金利予測モデルでは, 入力データを次元削減した深層表現による予測精度が最も良いことが確認できた.

本稿の今後の課題として, 以下の点が挙げられる. まず, 深層表現学習においては, 過去の傾向から大きく変化した際も精度が劣化しないモデルを構築することと, 入力データを再構築できるだけでなく金利予測を行うに適した深層表現を獲得することである. 現段階では, 入力データを深層表現から再構築できたかを評価する目的関数を用いて学習を行っているが, この目的関数に金利との誤差などを含む項を追加した学習を行うことで, 金利予測に適した深層表現を獲得できる可能性がある. 以上に述べた方策により, 安定期と変動期を問わず高い予測性能を備えた深層表現学習を実現することが今後の課題である.


## 謝　　辞

本研究を行うにあたり，有益な助言を頂いた神戸大学大学院経済学研究科の羽森茂之教授と金京拓司教授に感謝する．本研究に使用したデータを提供して頂いた神戸大学大学院システム情報学研究科谷口隆晴准教授に感謝する．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．


## 文　　献


[1] Geoffrey Hinton and Ruslan Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, Vol. 313, No. 5786, pp. 504 – 507, 2006.

[2] Yoshua Bengio, Li Yao, Guillaume Alain, and Pascal Vincent. Generalized denoising auto-encoders as generative models. *CoRR*, Vol. abs/1305.6663, , 2013.

[3] 円道滉一郎, 江口浩二, 羽森茂之, 金京拓司. 深層生成モデルによる時系列ネットワークの低次元埋め込み. 第 18 回人工知能学会金融情報学研究会 (SIG-FIN) 予稿集, pp. pp.120–127, 2017.

[4] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. The MIT Press, 2016.

[5] Quoc V. Le, Jiquan Ngiam, Adam Coates, Abhik Lahiri, Bobby Prochnow, and Andrew Y. Ng. On optimization methods for deep learning. In *Proceedings of the 28th International Conference on International Conference on Machine Learning*, ICML'11, pp. 265–272, USA, 2011. Omnipress.

[6] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th International Conference on Machine Learning*, ICML '08, pp. 1096–1103, New York, NY, USA, 2008. ACM.

[7] Yoshua Bengio, Nicholas Léonard, and Aaron C. Courville. Estimating or propagating gradients through stochastic neurons for conditional computation. *CoRR*, Vol. abs/1308.3432, , 2013.