

# ベイジアンネットワークによる複数深層学習器からの データ適合型学習器選択法

## A Selection Method of Data Adaptive Learner from Multiple Deep Learners Using Bayesian Networks

小林 秀輔<sup>1\*</sup> 白山 晋<sup>2†</sup>

Shusuke KOBAYASHI<sup>1</sup> and Susumu SHIRAYAMA<sup>2</sup>

<sup>1</sup> 千葉大学医学部医学科

Department of Medicine, Chiba University

<sup>2</sup> 東京大学工学系研究科システム創成学専攻

Graduate School of Engineering, the University of Tokyo

**Abstract:** This paper proposes a new method of time series prediction, using multiple deep learners and a Bayesian network. We firstly suggests two approaches. The former is a method in which explanatory variables of inputs data are nodes of a Bayesian network and are associated with learners. On the other hand, the latter method is a method in which the outputs of all the learners are made to nodes of the Bayesian network and the outputs are integrated. In this paper, the former method will be proposed in detail. Training data is divided into some clusters with K-means clustering and the multiple deep learners are trained, depending on each clusters. A Bayesian network is used to determine which the deep learner is in charge of predicting a time series. Our proposed method is applied to financial time series data, and the predicted results for the return of Nikkei 225 is demonstrated.

### 1 はじめに

我々は、K-means 法によって教師データを分割し、分割された教師データによる学習から複数の深層学習器を作成するという方法を提案している。本稿では、その方法の高度化を試みる。

ディープラーニングの発展の一方で適切な教師データの取得・生成、学習に要する長い計算時間、パラメータ選択の難しさなどが指摘されることも少なくない。それらの問題を解決するための研究も進んでいるが [1],[2],[3],[4],[5]、なぜうまくいかなかったかが十分に説明されていない。しかしながら、近年、ニューラルネットワークが適用されてきた分野、領域、あるいは対象を中心として、ディープラーニングへの置き換えが進み、効率化、高精度化に寄与したという報告が多くなされている [6],[7],[8]。

中でも、アンサンブル学習や相補的学習のような複

数の学習器の効率化や高精度化への寄与が大きいとされる。アンサンブル学習では、重み付き平均や投票法などによって学習器からの出力を統合する [9],[10],[11]。また、相補的学習では、お互いの欠点を補うように学習器を組み合わせる [12]。これらで用いられる学習器は、主として弱学習器である。一方、多くのタスクはサブタスクに分けられることに注目し、階層型制御機構が採用される。階層型制御機構を実現するために、複数の学習器が利用される場合がある。

高橋と浅田は、同一構造の複数の学習器を階層的に構築することによるロボットの行動獲得法を提案している [13]。下位の学習器を、それぞれ異なったサブタスクを担当し低レベルな行為を学習させるものとし、上位の学習器を下位の学習器を利用し、より高いレベルの行為を学習させるものとするという学習器である。

我々は、複数の深層学習器と学習器の選択法にベイジアンネットワークの組み合わせによる学習器を提案した [14]。K-means 法によって教師データを分割し、分割された教師データによる学習から複数の深層学習器を作成するという方法である。この方法は、教師データの分割と分割データを独立して学習するという部分では、アンサンブル学習のバギング法と同じ考え方に

\*連絡先：千葉大学医学部医学科  
〒 260-0856 千葉県千葉市中央区亥鼻 1-8-1  
E-mail: shuu.kbys@gmail.com

†連絡先：東京大学工学系研究科システム創成学専攻  
〒 113-8656 東京都文京区本郷 7-3-1  
E-mail: sirayama@sys.t.u-tokyo.ac.jp

基づいている。分割法、複数の学習器の統合法（学習器の選択法）、学習後も学習器を高精度化するという考えが異なる。また、高橋と浅田の考え方にも近いが、タスクがサブタスクに分けられるという前提は必要としない。学習器の選択法は異なる。また、彼らが学習器として Q 学習を連続状態行動空間に拡張したもの [15] を用いるのに対して、深層学習を利用する点も異なる。

しかしながら、我々の方法は、教師データの分割によって複数の学習器を作成するという単純なものである。また、データの分割には K-means 法という単純なものを用いており、分割法による差違について検討できていない。さらに、ベイジアンネットワークを予測に適した深層学習器の選択に用いるという方法に複数の学習器の統合を試みたが、最も単純なネットワークを用いたため、次に述べるベイジアンネットワーク自体の問題が解決できていない。ベイジアンネットワークでは、確率変数をノードで表し、変数間の定量的な依存関係を条件付き確率で表現する。この際に、ネットワークの構造を事前に定める必要があるが、定め方には多義性があるという点が問題である。

本稿では、複数深層学習器とベイジアンネットワークによる学習器の選択に対するより一般的な枠組みを示すことから始める。具体的には、複数深層学習器は異なる種類のデータから生成されるものとし、データに適合するようにベイジアンネットワークのネットワーク構造を決定し、学習器の選択にそのベイジアンネットワークを用いるというものである。教師データに適合した構造を持つベイジアンネットワークを作成し、そのベイジアンネットワークによって学習器を選択する。用いる事例も先行研究との比較のため、複数の株式市場の影響を考慮した日経平均株価予測とする。具体的には、今期の日経平均株価を、1 期前の日経平均株価、NY ダウや FTSE100 などの海外の主要株価指標から推定する。推定結果の精度から提案手法の妥当性を評価する。

## 2 提案手法

### 2.1 概念

本節では、本稿の提案手法の概念を示す。教師データ  $\mathbf{S} \in \mathbb{R}^{n \times d}$  を  $K$  個にクラスタリングし、それぞれのデータで  $K$  個の深層学習器の学習を行う。  $K$  個の深層学習器  $l_1, \dots, l_K$  で、テストデータ  $\mathbf{S}_{test} \in \mathbb{R}^{m \times d}$  から、予測  $\hat{\mathbf{y}} \in \mathbb{R}^m$  を得る問題を考える。

ベイジアンネットワークを複数深層学習器と組み合わせるには、2つのアプローチがある。1つは、複数学習器にデータを入力する前にベイジアンネットワーク

を適用するアプローチであり、もう1つは、複数学習器からの出力を統合する際にベイジアンネットワークを適用するアプローチである。本稿では、前者のアプローチを手法 A、後者のアプローチを手法 B とする。

手法 A は、入力  $\mathbf{x} = (x_1, \dots, x_d)^T$  にベイジアンネットワークを適用し、入力  $\mathbf{x}$  と予測に用いる学習器  $l_k$  を対応づける手法である。入力  $\mathbf{x} = (x_1, \dots, x_d)^T$  は対応づけられた学習器のみに入力される。入力データの次元を  $d$  とすると、ベイジアンネットワークは、入力データの各変数  $x_1, \dots, x_d$  を示す  $d$  個のノードと、どの学習器を採用するかを示すノードの合計  $d+1$  個のノードをもつ。教師データとそのクラスタリング結果からベイジアンネットワークは学習される。

この手法では、ベイジアンネットワークにより対応づけられた学習器  $l_k$  からの出力が予測  $\hat{y} \in \mathbb{R}$  となる。以下に手法 A のプロセスを示す。

1. 教師データ  $\mathbf{S} \in \mathbb{R}^{n \times d}$ 、正解データ  $\mathbf{y} \in \mathbb{R}^n$
2. 教師データ  $\mathbf{S}$  を  $K$  個にクラスタリングする。クラスタ毎の教師データと対応する正解データ  $(\mathbf{S}_i, \mathbf{y}_i)_{i=1}^K$ 、 $\mathbf{S}$  の各行が属するクラスタ番号を格納したベクトル  $\mathbf{c}$  を得る。
3.  $(\mathbf{S}_i, \mathbf{y}_i)$  で  $l_i (i = 1, \dots, K)$  を学習する。
4.  $(\mathbf{S}, \mathbf{c})$  でベイジアンネットワークを学習する。
5. テストデータ  $\mathbf{S}_{test} \in \mathbb{R}^{m \times d}$
6.  $\mathbf{S}_{test}$  をベイジアンネットワークに入力し、出力  $\mathbf{x}_{bn} \in \mathbb{Z}^m$  を得る。
7.  $\mathbf{S}_{test}$  を  $\mathbf{x}_{bn}$  の対応する学習器に入力し、予測  $\hat{\mathbf{y}} \in \mathbb{R}^m$  をえる。

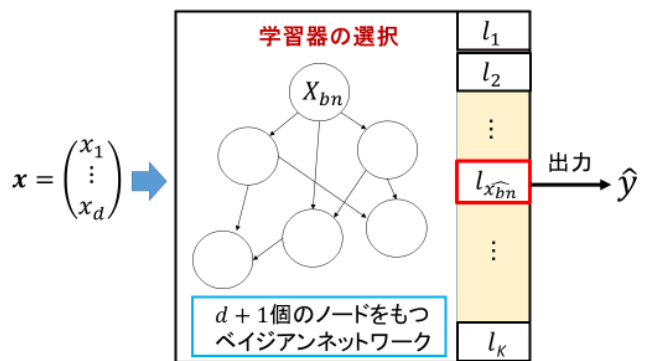


図 1: 手法 A

図 1 に手法 A の概略を示す。ここで、 $X_{bn}$  は学習器の選択を表す確率変数であり、 $\hat{x}_{bn} \in \{1, \dots, K\}$  は  $X_{bn}$

からの推定値である。学習済みのベイジアンネットワークに  $\mathbf{x} = (x_1, \dots, x_d)^T$  を入力することで、 $x_{\hat{b}_n}$  を得る。 $x_{\hat{b}_n}$  に対応する学習器から、予測  $\hat{y}$  を得る。手法 A のプロセスでは、6, 7 に対応している。

手法 B は、 $K$  個の学習器の出力  $\mathbf{u} \in \mathbb{R}^K$  をベイジアンネットワークにより統合する手法である。

ベイジアンネットワークは、 $K$  個の学習器ノードと最終出力ノードの合計  $K+1$  のノードをもつネットワークとなる。教師データを  $K$  個全ての学習器に入力した場合の出力と正解ラベルからベイジアンネットワークを学習させる。

学習器への入力  $\mathbf{x} = (x_1, \dots, x_d)^T$  が  $K$  個全ての学習器に入力され、出力  $\mathbf{u} \in \mathbb{R}^K$  が得られる。この手法では、複数学習器からの出力  $\mathbf{u} \in \mathbb{R}^K$  をベイジアンネットワークに入力することで得られる最終出力ノードの推定値が、予測  $\hat{y} \in \mathbb{Z}$  となる。

以下に手法 B のプロセスを示す。

1. 教師データを  $\mathbf{S} \in \mathbb{R}^{n \times d}$  とする。  
正解データを  $\mathbf{y} \in \mathbb{R}^n$  とし、複数学習器の学習に用いる。正解ラベルを  $\mathbf{y}_{label} \in \mathbb{Z}^n$  とし、ベイジアンネットワークの学習に用いる。
2. 教師データ  $\mathbf{S}$  を  $K$  個にクラスタリングする。  
クラスタ毎の教師データと対応する正解データ  $(\mathbf{S}_i, \mathbf{y}_i)_{i=1}^K$  を得る。
3.  $(\mathbf{S}_i, \mathbf{y}_i)$  で  $l_i (i = 1, \dots, K)$  を学習する。
4.  $\mathbf{S}$  をすべての学習器  $l_1, \dots, l_K$  に入力し、出力  $\mathbf{U} \in \mathbb{R}^{n \times K}$  をえる。
5.  $\mathbf{U} \in \mathbb{R}^{n \times K}$  と  $\mathbf{y}_{label} \in \mathbb{Z}^n$  でベイジアンネットワークを学習する。
6. テストデータ  $\mathbf{S}_{test} \in \mathbb{R}^{m \times d}$
7.  $\mathbf{S}_{test}$  を  $K$  個の学習器  $l_1, \dots, l_K$  に入力する。  
学習器からの出力をベイジアンネットワークで統合し、予測  $\hat{\mathbf{y}} \in \mathbb{Z}^m$  をえる。

図 2 に手法 B の概略を示す。ここで、 $\hat{Y}$  は最終出力を表す確率変数である。学習済みの複数学習器に  $\mathbf{x} = (x_1, \dots, x_d)^T$  を入力し、複数学習器からの出力  $u_1, \dots, u_K$  をベイジアンネットワークで統合する。手法 B のプロセスでは 7 に対応している。

ベイジアンネットワークによる複数深層学習器からの選択的データ学習法には、以上の 2 つの手法が考えられる。手法 A と手法 B を組み合わせることで、データに応じた学習器選択と選択された学習器を用いた予測が可能である。

本稿では、手法 A を用いる。

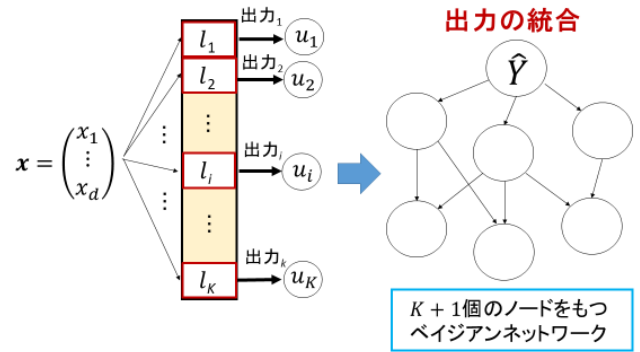


図 2: 手法 B

## 2.2 手法 A

本節では、手法 A について説明を行う。

前節で述べたように、手法 A は、複数深層学習器を学習させたのち、データをどの学習器に入力すべきかをベイジアンネットワークで対応付ける手法である。ベイズ分類器はベイズの定理を利用して、分類問題を解く手法である。我々の先行研究 [14] では、ナイーブベイズ分類器をベイズ分類器として採用している。ナイーブベイズ分類器は、特徴量間の条件付き独立性を仮定したモデルであり、最も簡単な構造を有するベイジアンネットワークであるが、構造が簡単な構造のみに限定されており、ノード間の関係性が十分に表現されない可能性がある。従って、本稿では特にネットワークの構造を限定せず、データからネットワーク構造を構築したベイジアンネットワークをベイズ分類器として用いる。

### 2.2.1 クラスタリングと深層学習器の訓練

教師データを  $\mathbf{S} \in \mathbb{R}^{n \times d}$ 、正解データを  $\mathbf{y} \in \{0, 1\}^n$  とする。まず、教師データ  $\mathbf{S}$  を  $K$  個にクラスタリングする。クラスタ毎の教師データと対応する正解  $(\mathbf{S}_i, \mathbf{y}_i)_{i=1}^K$ 、 $\mathbf{S}$  の各行が属するクラスタ番号を格納したベクトル  $\mathbf{c} \in \{1, 2, \dots, K\}^n$  を得る。本稿では、クラスタリングには K-means 法を用いた。また、K-means 法の分割数は X-means [16] によって定めた。クラスタリングされた  $K$  個のデータ  $(\mathbf{S}_i, \mathbf{y}_i)_{i=1}^K$  で、 $K$  個の深層学習器の学習を行う。

### 2.2.2 ベイジアンネットワーク

ベイジアンネットワークの学習は、 $(\mathbf{S}, \mathbf{c})$  のデータで行う。 $d$  次元データ  $(x_1, \dots, x_d)^T$  が与えられた際に、 $K$

個の学習器のうち、どの学習器に  $(x_1, \dots, x_d)^T$  を入力すればよいかを確率的に対応付ける分類器を構築する。

ベイジアンネットワークはノード間の条件付き確率に正規分布を仮定することで、連続値も扱うことができる [17]。本研究では連続値を含むベイジアンネットワークを構築しているが、ここでは説明のため、各変数は離散的な値をとることを仮定する。

$\mathbf{X} = (X_1, X_2, \dots, X_d, X_{d+1})$  を  $d+1$  個の確率変数の集合とする。  $X_1, \dots, X_d$  は入力に対応する確率変数、  $X_{d+1}$  は学習器選択に対応する確率変数である。つまり、  $X_{d+1}$  は図 1 における  $X_{bn}$  である。各変数は  $r_i$  個の状態集合  $\{0, 1, \dots, r_i-1\}$  ( $i = 1, \dots, d+1$ ) の中から 1 つの値をとる。ベイジアンネットワークの構造  $B_s$  を所与とし、変数  $i$  の親ノードを  $pa(X_i)$  と表すと、同時確率分布は式 (1) で表すことができる。

$$p(X_1, X_2, \dots, X_{d+1}|B_s) = \prod_{i=1}^{d+1} P(X_i|pa(X_i), B_s) \quad (1)$$

ここで、データは欠損値がない完全データと仮定する。ベイジアンネットワークは、(i) 構造決定 (ii) 条件付き確率の推定 (iii) 確率伝播法による周辺化、のプロセスで扱われる。(i) と (ii) でベイジアンネットワークの学習が完了する。観測値やテストデータから予測を行う際は、バイズの定理と (ii) で推定した条件付き確率に基づいて周辺化が行われるが、確率伝播法のアルゴリズムは周辺化を効率的に実行する。

まず、ベイジアンネットワークの構造を決定する。ベイジアンネットワークの構造は、ノード数に対して組み合わせ数が爆発的に増えていくことが一般に知られている。例えば、ノード数 2 のときは組み合わせ数は 3、ノード数 3 のときは組み合わせ数は 25、ノード数 5 のときは組み合わせ数が 29,000 になる [18]。つまり、ベイジアンネットワークの構造を全探索で求めることには限界がある。ベイジアンネットワークの構造は、変数間の事前知識、情報量基準、予測分布などから決定される。本研究では、情報量基準を用いて、データから構造を決定した。

情報量基準を  $I_B$  とし、式 (2) とする [19]。

$$I_B = -2 \sum_{i=1}^{d+1} \sum_{j=1}^{q_i} \sum_{k=0}^{r_i-1} n_{ijk} \ln \frac{n_{ijk}}{n_{ij}} + c_B \sum_{i=1}^{d+1} q_i (r_i - 1) \quad (2)$$

ただし、確率変数  $X_i$  がとりうる状態数を  $r_i$ 、 $X_i$  の親変数の状態総数を  $q_i$  とする。 $n_{ijk}$  は  $pa(X_i) = j$  のとき  $X_i = k$  となるデータ数を示している。

式 (2) の第一項は対数尤度を、第二項はパラメタの数を示しており、 $c_B$  の部分を定めることで情報量基準をつくることができる。たとえば、 $c_B = 2$  のとき、式 (2) は AIC になり、 $c_B = \log(d+1)$  のとき BIC になる。ベイジアンネットワークにおいても、これらの情

報量基準を最適化する構造を見つければよい。実際は、ヒューリスティックな探索手法の指標として情報量基準が使われることが多く、本研究では、hill climbing algorithm や tubu search で構造を探索する際の指標として用いた。

つぎに、各ノードの条件付き確率  $P(X_i|pa(X_j))$  をデータから推定する必要がある。推定すべきパラメタを  $\theta_{ijk} = P(X_i = k|pa(X_i) = j)$  とする。 $\Theta_{B_s} = (\theta_{ijk}) (i = 1, \dots, d+1, j = 1, \dots, q_i, k = 0, \dots, r_i-1)$

このときデータ  $\mathbf{S}$  を所与とした尤度  $L(\Theta_{B_s}|\mathbf{S})$  は、式 (3) でかける。

$$L(\Theta_{B_s}|\mathbf{S}) \propto \prod_{i=1}^{d+1} \prod_{j=1}^{q_i} \prod_{k=0}^{r_i-1} \theta_{ijk}^{n_{ijk}} \quad (3)$$

ベイジアンネットワークの尤度分布は多項分布に従うため、共役事前分布としてディレクレ分布を設定する。

$$P(\Theta_{B_s}) = \prod_{i=1}^{d+1} \prod_{j=1}^{q_i} \prod_{k=0}^{r_i-1} \frac{\Gamma(\sum_{k=0}^{r_i-1} n'_{ijk})}{\prod_{k=0}^{r_i-1} \Gamma(n'_{ijk})} \prod_{k=0}^{r_i-1} \theta_{ijk}^{n'_{ijk}-1} \quad (4)$$

ここで、 $\Gamma()$  はガンマ関数を示す。 $n'_{ijk}$  は、 $n_{ijk}$  に対応する事前分布のハイパーパラメタを示し、 $n'_{ijk} > 0$  ( $k = 0, \dots, r_i-1$ ) である。

事後分布は事前分布と尤度の積だから、事後分布は式 (5) で与えられる。

$$P(\Theta_{B_s}|\mathbf{S}) \propto \prod_{i=1}^{d+1} \prod_{j=1}^{q_i} \prod_{k=0}^{r_i-1} \theta_{ijk}^{n'_{ijk} + n_{ijk} - 1} \quad (5)$$

条件付き確率の推定値  $\hat{\theta}_{ijk}$  は、式 (5) から求めることができる。式 (5) の MAP 推定値  $\theta_{MAP}$  は、式 (6) となる [19]。

$$\theta_{MAP} = \frac{n'_{ijk} + n_{ijk} - 1}{n'_{ij} + n_{ij} - r_i} \quad (k = 0, \dots, r_i - 2) \quad (6)$$

ただし、 $n'_{ij} = \sum_{k=0}^{r_i-1} n'_{ijk}$ 、 $n_{ij} = \sum_{k=0}^{r_i-1} n_{ijk}$  である。

各ノードに条件付き確率が推定されると、具体的な観測値やテストデータから注目するノードの事後確率を計算し、比較することで予測を行うことができる。この際、確率伝播法を使うことで効率的に周辺化を行うことができる。 $X_1, \dots, X_d$  は入力の各変数を表す確率変数、 $X_{d+1}$  は入力をどの学習器に対応付けるかを表す確率変数であった。いま、注目するノードを  $X_{d+1}$  とし、それ以外のノード  $X_1, \dots, X_d$  に観測値  $\mathbf{e}$  が与えられたとき、 $X_{d+1}$  の推定値  $x_{\hat{d+1}}$  は、式 (7) で与えられる。

$$x_{\hat{d+1}} = \arg \max_{x_{d+1}=1,2,\dots,K} P(X_{d+1} = x_{d+1}|\mathbf{e}) \quad (7)$$

$X_{d+1}$  は図 1 の  $X_{bn}$  に対応しているから、 $x_{\hat{d+1}} = \hat{x}_{bn}$  となる。

### 2.2.3 予測

手法 A では、ベイジアンネットワークでテストデータ  $\mathbf{S}_{test} \in \mathbb{R}^{m \times d}$  と  $K$  個の学習器の対応付けを行う。テストデータは  $d$  次元の変数をもつ  $m$  個のデータである。  $\mathbf{S}_{test}$  をベイジアンネットワークに入力すると、出力  $\hat{\mathbf{x}}_{bn} \in \mathbb{Z}^m$  を得る。  $\hat{\mathbf{x}}_{bn}$  は、  $m$  個のデータがそれぞれの学習器に入力されるかを示す。  $\mathbf{S}_{test}$  を  $\hat{\mathbf{x}}_{bn}$  の対応する学習器に入力することで、予測  $\hat{\mathbf{y}} \in \mathbb{R}^m$  を得ることができる。

## 3 実験

提案手法を評価するために、数値実験を行う。本稿では、手法 A の実験を行った結果を示す。先行研究 [14] との比較のため、同じデータセットを用いて実験を行った。

### 3.1 データセット

本節では、実験で用いたデータセットについて説明する。

6 指標の金融時系列データを利用し、日経平均株価の平均リターンからの乖離を予測した。データの内訳は、2000 年 1 月 1 日から 2014 年 12 月 31 日までの日経平均株価、NY ダウ、NASDAQ、S & P500、FTSE100、DAX の日足データの終値であり、Yahoo finance!、Federal Reserve Bank of St. Louis[20]、[21] から取得した。NY ダウ、NASDAQ、S & P500 はアメリカの株式銘柄指標であり、FTSE100、DAX はヨーロッパの株式銘柄指標である。

各国の祝日の関係で、いずれかの指標が欠損しており、6 指標すべてのデータが揃わない日がある。その場合、データがない指標は前日から変化なしと仮定し、前日の指標の値を採用した。一般に、金融時系列の生データは非定常性が強い傾向を有するため、適切な変形を施す必要性がある。本稿では、株式データをリターンに変形して利用した。

時系列データを  $\mathbf{p}(t)(0 \leq t \leq T)$  とする。式 (8) に従って、 $\mathbf{p}(t-1)$ 、 $\mathbf{p}(t)$  をリターン  $\mathbf{r}(t)$  に変形する。

$$\mathbf{r}(t) = \{\log \mathbf{p}(t) - \log \mathbf{p}(t-1)\} \times 100 \quad (8)$$

リターン  $\mathbf{r}(t)(0 \leq t \leq T)$  は、1 期ずつデータをシフトさせることで得られる。

また、リターン  $\mathbf{r}(t)$  の定常性を確認するために、ディッキーフルー検定を行った。帰無仮説は、単位根が存在することである。一方、対立仮説は検定した時系列が定常過程であることである。ディッキーフルー検定の結果、帰無仮説が優位水準 5% で棄却されたため、リターン  $\mathbf{r}(t)$  は定常な時系列であると仮定し、日経平均

株価の平均リターンからの乖離を予測した。定常性の仮定より、平均リターンは全ての時間にわたり一定である。

### 3.2 実験結果

本節では数値実験の結果を示す。

$t$  期のリターン  $\mathbf{r}(t)$  から、 $t+1$  期の日経平均株価のリターン  $\mathbf{r}(t+1)$  について、平均リターンからの乖離を正負の 2 値で予測する。データの総数は、3912 個である。2000 年から 2013 年までの 3652 個のデータを教師データ、2014 年の 260 個のデータをテストデータに分割する。つまり、教師データ  $\mathbf{S} \in \mathbb{R}^{3652 \times 6}$ 、正解データ  $\mathbf{y} \in \{1, -1\}^{3652}$ 、テストデータ  $\mathbf{S}_{test} \in \mathbb{R}^{260 \times 6}$ 、テストデータの正解  $\mathbf{y}_{test} \in \{1, -1\}^{260}$  である。

ベイジアンネットワークの構造は、情報量基準をもとにデータからヒューリスティックに決定する。本研究では情報量基準として、対数尤度、AIC、BIC の 3 種類を利用し、hill climbing algorithm と tabu search を用いてベイジアンネットワークを作成した。

ベイジアンネットワークの作り方によって、提案手法は 6 つに分かれる。

提案手法 1 では、対数尤度を基準に、hill climbing algorithm を用いてベイジアンネットワークを作成する。提案手法 2 では、AIC を基準に、hill climbing algorithm を用いてベイジアンネットワークを作成する。提案手法 3 では、BIC を基準に、hill climbing algorithm を用いてベイジアンネットワークを作成する。

提案手法 4 では、対数尤度を基準に、tabu search を用いてベイジアンネットワークを作成する。提案手法 5 では、AIC を基準に、tabu search を用いてベイジアンネットワークを作成する。提案手法 6 では、BIC を基準に、tabu search を用いてベイジアンネットワークを作成する。

単一の深層学習器を用いる予測、ナイーブベイズ分類器を使った複数学習器による予測 [14] の 2 つを対比手法とした。

複数深層学習器には、DNN、RNN、LSTM の 3 種類の深層学習器を用いた。隠れ層はユニット数 6 の層を 2 層とし、epochs 数 100 回、dropout rate を 0.2 として学習を行った。ミニバッチの数は訓練データ数の 5 分の 1 とした。

各々 10 回試行し、精度と  $f$  値の平均と標準誤差、最大値を示す。

表 1 と表 2 に DNN を深層学習器に用いた場合の結果を示す。次に、表 3 と表 4 に RNN を深層学習器に用いた場合の結果を示す。最後に、表 5 と表 6 に LSTM を深層学習器に用いた場合の結果を示す。

表 1: Accuracy and F-value(DNN)

	Accuracy	F-value
DNN	0.6658 $\pm$ 0.0143	<b>0.6524 <math>\pm</math> 0.0194</b>
DNN Naive Bayes[14]	<b>0.6788 <math>\pm</math> 0.00714</b>	0.5664 $\pm$ 0.00601
Proposed Method 1	0.6623 $\pm$ 0.00821	0.5497 $\pm$ 0.02959
Proposed Method 2	0.6627 $\pm$ 0.00826	0.5454 $\pm$ 0.01991
Proposed Method 3	0.6588 $\pm$ 0.00958	0.5525 $\pm$ 0.03059
Proposed Method 4	0.6619 $\pm$ 0.01024	0.5610 $\pm$ 0.02935
Proposed Method 5	0.6608 $\pm$ 0.00803	0.5507 $\pm$ 0.01772
Proposed Method 6	0.6642 $\pm$ 0.01377	0.5563 $\pm$ 0.04158

表 5: Accuracy and F-value(LSTM)

	Accuracy	F-value
LSTM	0.6604 $\pm$ 0.0198	<b>0.6502 <math>\pm</math> 0.0114</b>
LSTM Naive Bayes[14]	<b>0.6827 <math>\pm</math> 0.0065</b>	0.5609 $\pm$ 0.0042
Proposed Method 1	0.6654 $\pm$ 0.0082	0.5373 $\pm$ 0.0079
Proposed Method 2	0.6654 $\pm$ 0.0094	0.5440 $\pm$ 0.0199
Proposed Method 3	0.6665 $\pm$ 0.0039	0.5376 $\pm$ 0.0021
Proposed Method 4	0.6665 $\pm$ 0.0057	0.5381 $\pm$ 0.0066
Proposed Method 5	0.6646 $\pm$ 0.0064	0.5412 $\pm$ 0.0140
Proposed Method 6	0.6673 $\pm$ 0.0058	0.5388 $\pm$ 0.0062

表 2: Maximum value of Accuracy and F-value(DNN)

	Accuracy	F-value
DNN	0.6885	<b>0.6754</b>
DNN Naive Bayes[14]	0.6885	0.5763
Proposed Method 1	0.6769	0.6304
Proposed Method 2	0.6808	0.5926
Proposed Method 3	0.6731	0.6064
Proposed Method 4	0.6769	0.6050
Proposed Method 5	0.6769	0.5778
Proposed Method 6	<b>0.6962</b>	0.6664

表 6: Maximum value of Accuracy and F-value(LSTM)

	Accuracy	F-value
LSTM	0.6885	<b>0.6628</b>
LSTM Naive Bayes[14]	<b>0.6923</b>	0.5689
Proposed Method 1	0.6808	0.5542
Proposed Method 2	0.6846	0.5967
Proposed Method 3	0.6692	0.5395
Proposed Method 4	0.6731	0.5454
Proposed Method 5	0.6731	0.5782
Proposed Method 6	0.6731	0.5508

表 3: Accuracy and F-value(RNN)

	Accuracy	F-value
RNN	0.6765 $\pm$ 0.0146	<b>0.6714 <math>\pm</math> 0.0138</b>
RNN Naive Bayes[14]	<b>0.6819 <math>\pm</math> 0.0102</b>	0.6011 $\pm$ 0.0395
Proposed Method 1	0.6700 $\pm$ 0.0091	0.5649 $\pm$ 0.0355
Proposed Method 2	0.6638 $\pm$ 0.0058	0.5723 $\pm$ 0.0356
Proposed Method 3	0.6650 $\pm$ 0.0105	0.5687 $\pm$ 0.0356
Proposed Method 4	0.6700 $\pm$ 0.0099	0.5907 $\pm$ 0.0362
Proposed Method 5	0.6642 $\pm$ 0.0157	0.5653 $\pm$ 0.0448
Proposed Method 6	0.6685 $\pm$ 0.0204	0.5681 $\pm$ 0.0485

表 4: Maximum value of Accuracy and F-value(RNN)

	Accuracy	F-value
RNN	0.7038	<b>0.6928</b>
RNN Naive Bayes[14]	0.7000	0.6664
Proposed Method 1	0.6885	0.6355
Proposed Method 2	0.6769	0.6314
Proposed Method 3	0.6885	0.6479
Proposed Method 4	0.6846	0.6425
Proposed Method 5	0.6885	0.5450
Proposed Method 6	<b>0.7115</b>	0.6723

## 4 考察

表 1, 表 3 より, 深層学習器として DNN や RNN を用いた場合, 単一学習器を用いた場合や先行研究 [14] に比べて, 提案手法の精度と  $f$  値は平均的には向上しなかった。

しかし, 表 2, 表 4 にみられるように, BIC を基準に tabu search でベイジアンネットワークを構築した提案手法では, 精度の最大値について既存手法より 1% 程度の向上が見られた。提案手法の  $f$  値の最大値についても, 我々の先行研究 [14] からの向上が見られた。

ベイジアンネットワークはナイーブベイズと比べ, 複雑な構造を表現できるかわりに, 最適な構造を探索する必要があるという難しさを抱えている。本研究では実験において, 10 回試行を行いその平均値を求めたが, その都度ベイジアンネットワークを構成しており, 毎回最適なネットワークが得られていたとは限らない。つまり, 提案手法では結果のばらつきが既存手法と比較して大きくなる。

一方, 最適構造を探索できた場合, その結果は既存手法より良くなることが期待される。精度と  $f$  値の最大値を示した表 2, 表 4 は, 最適構造を探索できた場合の実験結果を示していると考えられ, 提案手法 6 の精度は既存手法より良い結果を示したことが確認できる。

深層学習器として LSTM を用いた場合、表 5 より、提案手法の精度は単一の LSTM を用いた場合と比較し、平均的には向上しているが、その最大値は既存手法より悪い。これは LSTM という深層学習器の性質や本研究で用いたデータの性質が関係していると思われる。

## 5 結論と課題

本研究では、複数深層学習器とベイジアンネットワークによる学習器の選択に対する、より一般的な枠組みを示したのちに我々の先行研究 [14] を発展させた手法を提案した。具体的には、先行研究で用いたベイズ分類器をベイジアンネットワークに変更した。

ベイジアンネットワークはナイーブベイズ分類器と比較して、複雑な構造を表現できる一方で最適構造を探索する必要性があるという難しさがある。従って、深層学習器に複数の DNN, RNN を用いた場合において、精度と  $f$  値の平均値は既存手法よりも悪い。しかし、最適なベイジアンネットワークを構築できた場合、提案手法は先行研究よりも良い結果を示した。例えば、BIC を基準に tabu search で作成したベイジアンネットワークによる提案手法では、精度の最大値は、単一の学習器で予測した場合や先行研究よりも 1% 程度、高い値を示した。 $f$  値の最大値についても、先行研究から向上しているケースが確認できた。

一方、深層学習器に複数の LSTM を用いた場合、提案手法の平均精度は、学習器に単一の LSTM を用いた場合よりも向上したが、我々の先行研究と比べると悪い結果となった。

本研究の今後の課題であるが、以下の 3 点を挙げる。

第 1 点目は、 $f$  値の向上である。提案手法はクラスタリングを行い  $K$  個にデータを分割し、各々のデータで  $K$  個の複数学習器を構築する。単一の学習器で訓練を行う場合に比べ、類似したパターンを有するデータのみで学習器の訓練が行えるため、精度は一般に向上するが、過学習の傾向が生じるため  $f$  値が既存手法に比べ小さくなってしまう。クラスタリングに含めるデータ、分割数  $K$  の決定法、深層学習器のパラメタ調整などを改善することによって、この問題は解決ができると思われる。

2 点目の課題は、ベイジアンネットワークの構築の仕方についてである。ベイジアンネットワークの作成に関して、本研究では対数尤度、AIC、BIC を指標にして hill climbing algorithm と tabu search という 2 種類の方法でヒューリスティックに作成した。情報量基準に関して、BIC を採用したケースにおいて、既存手法より 1% 程度高い精度が得られているものの、対数尤度や AIC と比べて絶対的に優れているとは言えない。これは hill climbing algorithm と tabu search につい

ても同様である。今後、別のデータに提案手法を適用した場合の精度や  $f$  値などの比較を行うことで、この課題についても検討をしていきたいと考えている。

また、3 点目の課題として、ベイジアンネットワークを複数学習器からの出力の統合に用いる手法の確立を挙げる。提案手法の章で、ベイジアンネットワークと複数学習器の組み合わせ方には、2 つのアプローチがあることを述べた。入力の変数をノードにして予測に用いる学習器をベイジアンネットワークで選択する手法 A と、複数学習器からの出力をベイジアンネットワークで統合する手法 B の 2 つの手法である。本稿では、2 つのアプローチの概念、そして手法 A の詳細について述べたが、手法 B も深層学習器とクラスタリング、そしてベイジアンネットワークを使って設計することが可能である。最終的には 2 つの手法を組み合わせることで、データに応じた学習器選択と選択された学習器を用いた予測を行う手法の提案をしたいと考えている。

## 参考文献

- [1] J.Bergstra and Y.Bengio, Random search for hyperparameter optimization, Journal of Machine Learning Research, Vol.13,pp.2817-305, 2012.
- [2] I.Loshchilov and F.Hutter., CMA-ES for hyperparameter optimization of deep neural networks, CoRR, vol. abs/1604.07269, 2016.
- [3] P. R.Lorenzo, J.Nalepa, M.Kawulok, L.S.Ramos and J.R.Pastor, Particle swarm optimization for hyperparameter selection in deep neural networks, In Proceedings of the Genetic and Evolutionary Computation Conference, pp.481-488, ACM, 2017
- [4] J.Snoek, H.Larochelle and R. P.Adams, Practical bayesian optimization of machine learning algorithms, In Advances in neural information processing systems, pp.2951-2959, 2012
- [5] T.Kuremoto, S.Kimura, K.Kobayashi and M.Obayashi, Time series forecasting using a deep belief network with restricted Boltzmann machines. Neurocomputing, vol.137, pp.47-56, 2014
- [6] Y.Bengio, P.Lamblin, D.Popovici, and H.Larochelle, Greedy layer-wise training of deep networks, In Advances in neural information processing systems, pp.153-160, 2007
- [7] A.Krizhevsky, I.Sutskever, and G. E. Hinton, Imagenet classification with deep convolutional neural networks, In Advances in neural information processing systems, pp.1097-1105, 2012
- [8] G.E.Dahl, D.Yu, L.Deng and A.Acerio, Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition, IEEE Transactions on audio, speech, and language processing, vol.20, Issue.1, pp.30-42, 2012

- [9] H. Z. Wang, G. Q. Li, G. B. Wang, J. C. Peng, H. Jiang and Y. T. Liu, Deep learning based ensemble approach for probabilistic wind power forecasting, *Applied Energy*, vol.188, pp.56-70, 2017
- [10] H. I. Suk, S. W. Lee, D. Shen, and Alzheimer 's Disease Neuroimaging Initiative, Deep ensemble learning of sparse regression models for brain disease diagnosis. *Medical image analysis*, vol.37, pp.101-113, 2017
- [11] Y. Zhao, J. Li and L. Yu, A deep learning ensemble approach for crude oil price forecasting. *Energy Economics*, 2017
- [12] 野宮浩揮, 上原邦昭, 相補的な視覚的学習による複数の認識手法の統合, *電子情報通信学会論文誌 D*, Vol.J90-D, No.11, pp.3043-3054, 2007
- [13] 高橋泰岳, 浅田稔, 複数の学習器の階層的構築による行動獲得, *日本ロボット学会誌*, Vol.18, No.7, pp.1040-1046, 2000
- [14] S.Kobayashi and S.Shirayama, Time Series Forecasting with Multiple Deep Learners: Selection from a Bayesian Network. *Journal of Data Analysis and Information Processing*, 5(03), 115, 2017
- [15] Y.Takahashi, M.Takeda and M. Asada, Continuous Valued Q-learning for Vision-Guided Behavior Acquisition, in *Proceedings of 1999 IEEE/SICE/RSJ International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pp.255-260, 1999
- [16] D. Pelleg and A. Moore, X-means: Extending K-means with Efficient Estimation of the Number of Clusters, in *Proceedings of 7th International Conference on Machine Learning*, pp.727-734, 2000
- [17] D. Geiger and D. Heckerman, Learning Gaussian Networks, in *Tenth Conference on Uncertainty in Artificial Intelligence*, pp. 235-243, 1994
- [18] 繁柘算男, 本村陽一, 植野真臣, *ベイジアンネットワーク概説*, 培風館, 2006
- [19] 植野真臣, *ベイジアンネットワークの統計的学習*, *人工知能学会誌*, vol.25, No.6, pp.803-810, 2010
- [20] Yahoo! Finance, < <http://finance.yahoo.com/> > (2016/8/18 確認)
- [21] Economic Reseach Federal Researve Bank of St. Lois, < <http://research.stloisfed.org/> > (2016/8/18 確認)