

ベイズ推定の景気動向指数予測への応用について

北 栄輔^{1,a)} 中村 典雅¹

概要：景気とは、商売や取引など、社会全体の経済活動の状態を示す。経済状況が好調の場合は「好景気」、不調の場合は「不景気」などと表現する。景気動向を数値化するために景気動向指数が用いられる。景気動向指標にはディフュージョン・インデックス (Diffusion Index, DI) とコンポジット・インデックス (Composite Index, CI) がある。本研究では、過去の DI 値と CI 値を用いて景気判断を予測する方法について述べる。予測する数理モデルとしてベ이지アンネットワークを用いる。ベ이지アンネットワークでは確率変数をノードとして、変数間の依存関係を有向グラフで表現する。目的変数に景気を、説明変数に DI 値と CI 値を用いて予測を行う。実験データとして、1985 年から 2020 年までの日本の景気動向、景気動向指数を用いる。1985 年から 2017 年までのデータを学習データとして予測モデルを決定し、2018 年から 2020 年までの景気動向を予測する。予測精度を線形時系列モデルによる予測結果と比較した結果、ベ이지アンネットワークを用いる方がよい計算性を示したので、提案手法の有効性を評価することができた。

Application of Bayesian Prediction for Indexes of Business Condition

1. 緒論

景気とは、商売や取引など、社会全体の経済活動の状態を示し、経済状況が好調の場合を「好景気」、不調の場合を「不景気」などとする。景気の状態を景気動向とよび、景気の拡張期、後退期に分けられる [1]。景気動向を数値化するために景気動向指数が用いられる。景気動向指標にはディフュージョン・インデックス (Diffusion Index, DI) とコンポジット・インデックス (Composite Index, CI) が用いられる [2]。DI は、景気の各経済部門への波及の度合いを測定することを主な目的とし、CI は景気変動の大きさや量感を測定することを主な目的とする。DI と CI には、景気に対し先行して動く先行指数、ほぼ一致して動く一致指数、遅れて動く遅行指数の 3 つがある。DI 一致指数は、景気拡張局面では 50% を上回り、後退局面では下回る傾向がある。CI 一致指数が上昇している時は景気の拡張局面、低下している時は後退局面である。

これまでに景気動向指数の予測手法はいくつか研究されている。景気動向指数は時系列データであることから、これまで時系列データ分析手法が用いられている [3], [4], [5], [6], [7], [8]。これらの中には、自己回帰モデ

ル (Autoregressive model) や自己回帰和分移動平均モデル (ARIMA model) 等がある。これに対して、近年では機械学習を用いる方法が提案されている。佐々木らは、重回帰モデルとランダムフォレストを用いる手法を提案している [9]。また、坂地らは自然言語処理を用いて、景気動向に対する社会の印象を推定する方法を提案している [10]。これらに対して、本研究では、ベ이지アンネットワークを用いた景気動向の予測について述べる [11], [12], [13]。ベ이지アンネットワークでは、確率変数をノードで表し、変数間の因果関係を非循環有向グラフにより表現する。そして、変数間の依存関係の強度を条件付き確率によって評価する。提案手法では、説明変数として DI 値と CI 値の一致指数と先行指数をとる。なお、CI 値の一致指数と先行指数については、値そのものの代わりに、3 ヶ月前からの変動値をとる。目的変数としては、次期における景気動向が拡張期であるか、後退期であるかをとる。説明変数は連続地であるため、Ward 法を用いて離散値に変換し、これらと目的変数の関係についての数理モデルをベ이지アンネットワークによって構築する。実験データとして、1985 年から 2020 年までの日本の景気動向、景気動向指数を用いる。1985 年から 2017 年までのデータを学習データとして予測モデルを決定し、2018 年から 2020 年までの景気動向を予測する。提案手法を時系列分析手法と比較し、有効性を検

¹ 名古屋大学大学院情報学研究科
Nagoya University, Nagoya, 464-8601, Japan

^{a)} kita@i.nagoya-u.ac.jp

討する。

2. 研究の背景

2.1 景気動向指数

景気動向とは景気の状態であり、景気の拡張期、後退期 [10] に分けられる。景気動向を数値化するために景気動向指数が用いられる。景気動向指数には、コンポジット・インデックス (CI) とディフュージョン・インデックス (DI) がある [2]。

CI は景気変動の大きさや量感を測定することを主な目的とする。DI は構成する指標のうち、改善している指標の割合を算出することで景気の高経済部門への波及の度合い (波及度) を測定することを主な目的とする。CI と DI には、景気に対し先行して動く先行指数、ほぼ一致して動く一致指数、遅れて動く遅行指数の 3 つがある。

CI 一致指数が上昇している時は景気の拡張局面、低下している時は後退局面であり、CI 一致指数の動きと景気の転換点は概ね一致する。DI 一致指数は、景気拡張局面では 50% を上回り、後退局面では下回る傾向がある。

2.2 ベイジアンネットワーク

2.2.1 概要

ベイジアンネットワークでは、確率変数をノードとし、確率変数間の条件付依存関係を有向グラフによって表現する [11], [12]。変数間の依存関係の強度は、有向グラフに付された条件付確率で表現する。リンクの先にあるノードを子ノード、リンクの元にあるノードを親ノードとよぶ。

子ノード x_i に対して親ノードが n 個あり、それらが $x_{i1}, \dots, x_{in}$ であるとき、子ノード x_i に対する親ノード集合 $Pa(x_i)$ は次式で与えられる。

$$Pa(x_i) = \{x_{i1}, \dots, x_{in}\} \quad (1)$$

2.2.2 条件付き確率

条件付確率とは、親ノードがある値を取った時に、子ノードがある値を取る確率のことである。 $Pa(x_i)$ がある値をとったときに、 x_i が取る確率は $P(x_i|Pa(x_i))$ とする。子ノードと親ノードが取るすべての状態は、それぞれにおける確率値を列挙し、条件付確率表 (CPT) に保存する。

2.2.3 K2 Metric

ベイジアンネットワークの構造学習では、現実的な時間でネットワークの構造を評価し、学習することが必要である。そのため大量なネットワークの評価に対して、効率的な計算手法が必要である。

クーパーは、ベイジアンネットワークにおける予測分布を、式 (2) のように導いている。

$$Pr(D|B) = \prod_{i=1}^n \prod_{j=1}^{q_i} \frac{(r_i - 1)}{(N_{ij} + r_i - 1)} \prod_{k=0}^{r_i-1} N_{ijk}! \quad (2)$$

K2 Metric は値が大きいほどネットワークとしての信頼性

が高くなる。

2.2.4 ネットワーク構造探索

ベイジアンネットワークの構造探索には K2 アルゴリズムを用いる。K2 アルゴリズムにおいては、あらかじめ変数間の全順序関係が必要となる。そこで、 n 個の変数 $x_1, x_2, \dots, x_n$ に次の全順序関係があるとする。

$$x_1 > x_2 > \dots > x_n \quad (3)$$

まず、変数ノード x_i は 1 から n までの親ノード集合 $Pa(x_i)$ を初期化する。そして、1 から n までのループでネットワークの探索と学習を行う。変数ノード i を x_i 、 x_i の親ノードの集合を $Pa(x_i)$ とする。次に、変数ノード x_j は $j = i + 1$ から n までのループでノード x_j を $Pa(x_i)$ に追加する。この新しいネットワークを評価し、評価結果が良くなれば、それを新しいネットワークとする。もし良くなければ、 $Pa(x_i)$ からノード x_j を削除する。このように繰り返して、最適なネットワークの探索を行う。

探索アルゴリズムを以下に示す。

- (1) $i = 1$
 - (2) ノード x_i に対する親ノード集合 $Pa(x_i)$ を空集合 ϕ として定義する。
 - (3) x_i と $Pa(x_i)$ から構成されるネットワークについて K2Metric を評価し、これを S_{best} とする。
 - (4) $j = i + 1, \dots, N$ について以下の操作を行う。
 - (a) x_j を $Pa(x_i)$ に加える。
 - (b) x_i と $Pa(x_i)$ から構成されるネットワークについて K2Metric を評価し、これを S とする。
 - (c) $S > S_{best}$ でないならば、 x_j を $Pa(x_i)$ から除外する。
 - (5) $i = i + 1$ とし、 $i \leq N$ ならばステップ 2 へ戻る。
- ここで、ノード総数を N とする。

2.2.5 確率推論

ベイジアンネットワークの確率推論では、確率変数の確定値から、知りたい確率変数の事後確率を求め、これを用いて期待値などを計算する。

条件付き依存関係が成立しているネットワークにおいて、確率変数の確定値 e に対する知りたい確率変数 x_i の事後確率を $P(x_i|e)$ とすると、これは以下の手順で行われる。

- (1) 観測された変数の値 e をノードにセットする。
 - (2) 知りたい確率変数 x_i の条件付き確率 $P(x_i|e)$ を求める。
- 知りたい確率変数 x_i の各値についての事後確率 $P(x_i|e)$ は、あり得るすべての状態で平均化する周辺化によって求める。

確率変数 x_i が取り合える状態に $X^1, X^2, \dots, X^L$ の L 個の場合があり、そのうちで確率変数 x_i が状態 X^l をとりえる場合、確率 $P(x_i = X^l|e)$ は周辺化によって次式で与えられる。

$$P(x_i = X^l|e)$$

$$= \frac{\sum_{j=1, j \neq i}^N \sum_{x_j=X^1}^{X^L} P(x_1, \dots, X^L, \dots, x_N, e)}{\sum_{j=1}^N \sum_{x_j=X^1}^{X^L} P(x_1, \dots, x_N, e)} \quad (4)$$

ここで、 $\sum_{x_j=X^1}^{X^L}$ は確率変数 x_j の取り得るすべての場合 $X^1, X^2, \dots, X^L$ について総和を取ることを意味する。

3. 提案手法

3.1 目的変数

目的変数として t 期の景気動向を x_t とする。 x_t の取り得る値は、拡張期、後退期の2つであるから、それらを1および0とする。つまり、

$$x_t = \begin{cases} 1 & \text{拡張期} \\ 0 & \text{後退期} \end{cases} \quad (5)$$

3.2 説明変数

DI 値と CI 値には景気に対し先行して動く先行指数、ほぼ一致して動く一致指数、遅れて動く遅行指数の3つがあることはすでに述べた。本研究では景気予測を目的としていることから、説明変数として一致指数と先行指数を用いることとする。また、CI は前期の値との相対的な量として定義されているので、CI については、値そのものではなく、その前期からの変化量を説明変数に用いる。

t 期における DI の一致指数と先行指数の値をそれぞれ d_t, dl_t とする。また、 t 期における CI の一致指数と先行指数の前期からの変化量を $\Delta c_t, \Delta cl_t$ とする。つまり、

$$\Delta c_t = c_t - c_{t-1} \quad (6)$$

$$\Delta cl_t = cl_t - cl_{t-1} \quad (7)$$

これらは連続値であるが、Ward 法により DI 値、CI 値の分布をそれぞれ離散値に変換する。離散値を、それぞれ $d_t^j, dl_t^j, \Delta c_t^j, \Delta cl_t^j$ とする。また、それぞれの総数を、 N_1, N_2, N_3, N_4 とする。このとき、以下ようになる。

$$d_t \simeq \{d_t^j\} = \{d_t^1, d_t^2, \dots, d_t^{N_1}\}^T \quad (8)$$

$$\Delta c_t \simeq \{\Delta c_t^j\} = \{\Delta c_t^1, \Delta c_t^2, \dots, \Delta c_t^{N_3}\}^T \quad (9)$$

$$dl_t \simeq \{dl_t^j\} = \{dl_t^1, dl_t^2, \dots, dl_t^{N_2}\}^T \quad (10)$$

$$\Delta cl_t \simeq \{\Delta cl_t^j\} = \{\Delta cl_t^1, \Delta cl_t^2, \dots, \Delta cl_t^{N_4}\}^T \quad (11)$$

3.3 提案手法のアルゴリズム

景気動向の拡張期と後退期を予測するアルゴリズムは以下のようなになる。

- (1) 説明変数を離散化する。
- (2) 説明変数から目的変数を予測するベイジアンネットワーク B を決定する。
- (3) B を用いて予測期間の景気動向を予測する。

3.4 ベイジアンネットワーク B の決定

K2Metric をネットワークの評価指標として採用し、K2

アルゴリズムを用いてネットワークのグラフ構造を決定する。目的変数として景気動向 x_t を、説明変数として、過去 M 期の DI 値 d_t と CI 値 c_t とする。

K2 アルゴリズムを利用するためには、確率変数間の全順序関係が必要である。DI 値と CI 値には時系列に基づく順序がある。そこで、次の全順序関係を仮定する。

$$d_1, \dots, d_i, dl_i, c_i, cl_i, \dots, cl_M \quad (12)$$

この全順序関係に基づいてベイジアンネットワーク B を決定する。

3.5 景気動向の予測

予測用ネットワーク B を用いて景気動向 $x_t = x_t^j$ となる確率を $P(x_t^j|B)$ と定義する。 t 期の景気動向として、 $P(x_t^j|B)$ が最大となる x_t^j を選択する。つまり、

$$x_t = \arg \max_{x_t^j} (P(x_t^j|B)) \quad (13)$$

4. 解析例

実験データとして、1985 年から 2020 年までの日本の景気動向、景気動向指数を用いる。1985 年から 2017 年までのデータを学習データとして予測モデルを決定し、2018 年から 2020 年までの景気動向を予測する。精度比較のために、AR, ARIMA, VAR モデルを用いる。これらのモデルによって DI 値の時系列予測を行い、その値が 50 を超える場合は拡張期、下回る場合は後退期とする。

予測精度は、推定した回数 K に対する正解した回数 k で定義した正解率 η で評価する。

$$\eta = \frac{k}{K} \quad (14)$$

表 1 変数の離散値

Table 1 Discretized values of variables (Four explanatory variables)

d_t^1	d_t^2	d_t^3	d_t^4	d_t^5	d_t^6
7.237	24.68	45.94	66.16	80.00	92.26

dl_t^1	dl_t^2	dl_t^3	dl_t^4	dl_t^5
20.21	42.11	58.40	76.56	92.89

Δc_t^1	Δc_t^2	Δc_t^3	Δc_t^4
-6.450	-2.217	-0.2746	1.110

Δcl_t^1	Δcl_t^2	Δcl_t^3	Δcl_t^4	Δcl_t^5
-5.367	-1.583	-0.5733	0.4483	1.840

説明変数として、 n 期前の DI 値の一致指数 d_{t-n} と先行指数 dl_{t-n} CI 値の一致指数の変化量 Δc_{t-n} と先行指数の変化量 Δcl_{t-n} を用いる。また、期間 $n = 1, 2, 3$ とする。

まず、説明変数の分布を異なる分割数にて分割し、AIC

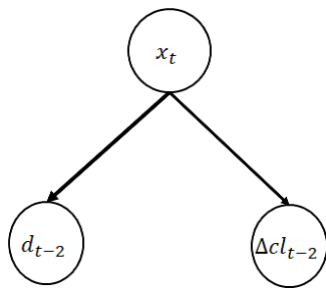


図 1 ベイジアンネットワーク
Fig. 1 Bayesian network

表 2 離散化された説明変数の比較

Table 2 Comparison of discrete values of explanatory variables

予測モデル	正解率
ベイジアンネットワーク	75.0%
AR モデル	55.6%
ARIMA モデル	58.3%
VAR モデル	66.7%

が最も良い値を示したのは、DI 値の一致指数において 6、先行指数において 5、CI 値の一致指数において 4、先行指数において 5 である。つまり、 $N_1 = 6, N_2 = 5, N_3 = 4, N_4 = 5$ となる。変数の離散値を表 1 にまとめる。この場合について 3 に示された方法で作成した予測用ベイジアンネットワークを図 1 に示す。このベイジアンネットワークでの予測の正答率は 75.0% である。説明変数として選択されたのは、2 期前の DI 値の一致指数と CI 値の先行指数の変動値である。

実験結果を他の手法と比較する。比較に用いるのは、時系列分析で用いる AR モデル、ARIMA モデル、VAR モデルである。これらの手法で予測したい期の DI 値を推定し、その値が 50 より大きい小さいかによって拡張期と後退期を判定する。

精度の比較結果を表 2 に示す。AR モデル、ARIMA モデル、VAR モデルによる予測精度は、それぞれ 55.6%、58.3%、66.7% である。ベイジアンネットワークを用いる方法は、これらよりも精度が高いことがわかる。

5. 結論

景気動向は、景気の拡張期と後退期に分けられる。これを評価するために DI と CI が用いられる。DI とは指数に採用している経済指標のうちで景気の拡大（拡張）を示している指標の割合を示し、CI とは景気拡大や後退の速度や程度を表す。

本研究では、ベイジアンネットワークを用いた景気動向予測について述べた。DI と CI の一致指数と先行指数を説明変数、景気動向を目的変数として、変数間の関係性をベイジアンネットワークによってモデル化した。まず、説明変数は連続変数であるため、Ward 法によって離散値に変

換した。変数の全順序関係によって目的変数を説明するベイジアンネットワークを決定した。

1985 年から 2017 年までのデータを学習データとして予測モデルを決定し、2018 年から 2020 年までの景気動向を予測した。精度比較のために、時系列予測手法である AR、ARIMA、VAR モデルを用いた。これらの手法の正解率は、それぞれ 55.6%、58.3%、66.7% であった。これに対して、提案手法の正解率は 75.0% であった。このことより、提案手法に有効性が確認できた。

今後は、日本以外の国における景気動向予測などへの応用について検討を進めることを考えている。

参考文献

- [1] R.W.Robinson. Counting Unlabeled Acyclic Digraphs, *Lecture Notes in Mathematics*, Vol.622, pp.28-43, 1977
- [2] A.Klenke. *Probability Theory: A Comprehensive Course*, Springer, 2008
- [3] R. Bookstaber, *The Complete Investment Book: Trading Stocks, Bonds, and Options With Computer Applications*, Scott Foresman Trade, 1985
- [4] 渡部敏明, 日本の株式市場におけるボラティリティの変動: ARCH 型モデルによる分析, 三菱経済研究所, 1997
- [5] G.Box, G.M.Jenkins and G.Reinsel, *Time Series Analysis: Forecasting and Control*, Prentice-Hall, 1994
- [6] T.Terasvirta, D.Tjostheim and C.W.J.Granger : *Modelling Nonlinear Economic Time Series*, Oxford Univ Pr, 2011
- [7] P. J. Brockwell and R.A.Davis, *Time Series: Theory and Methods*, Springer, 2009
- [8] 郭 旭東, 根木 佐一 株価指数による環太平洋諸国の景気相互連関分析, 東海大学紀要情報通信学部, Vol.7, No.2, 44-51, 2014.
- [9] 佐々木 秀彰, 浦野 昌一, 重回帰モデルとランダムフォレストを用いた景気動向指数予測, 人工知能学会全国大会論文集, Vol.JSAI2020, 2IIGS202-2IIGS202, 2020.
- [10] 坂地 泰紀, 酒井 浩之, 増山 繁, 景気動向を示す根拠表現の抽出と分析, 情報処理学会研究報告. NL, 自然言語処理研究会報告, 0919-6072, Vol.180, 151-156, 2007.
- [11] 繁樹算男, 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.
- [12] 本村陽一, 岩崎弘利. ベイジアンネットワーク技術ユーザ・顧客のモデル化と不確実性推論. 東京電機大学出版局, 2006.
- [13] Heckerman, D., Geiger, D., and Chickering, D. Learning Bayesian Networks : The Combination of Knowledge and Statistical Data, *Machine Learning*, Vol.20, pp. 197-243 (1995).