

ベイジアンネットワークに基づく ODデータの地域間依存関係の視覚化

布施 孝志¹・渡邊 拓也²

¹正会員 東京大学大学院 工学系研究科社会基盤学専攻 (〒113-8656 東京都文京区本郷7-3-1)
E-mail: fuse@civil.t.u-tokyo.ac.jp

²正会員 公益財団法人 鉄道総合技術研究所 (元 東京大学大学院工学系研究科社会基盤学専攻)
(〒185-8540 東京都国分寺市光町2-8-38)
E-mail: wttk@rtri.or.jp

人の移動などにおける地域間の依存関係は、その全体構造を把握する上で重要である。これらの動態を知るために、各種統計データが整備されている。近年では、ベイジアンネットワークなどにより、変数間の依存関係を抽出する手法が発展している。本研究では、ベイジアンネットワークに基づくグラフ構造推定により、ODデータから、地域間の依存関係を視覚化する手法の構築を目的とする。ベイジアンネットワークでは、変数間の依存関係を、条件付き確率を伴う有向グラフで表現する。この関係をB-スプライン非線形回帰モデルでモデル化し、事後確率最大化により、グラフ構造を推定する。有向グラフの組み合わせは多数に及ぶため、情報量規準の下、グラフ構造を決定する。提案手法を、全国幹線旅客純流動調査に適用し、地域間の依存関係の視覚化を試みた。

Key Words : Bayesian network, graph structure inference, inter-regional dependence, OD table

1. はじめに

人の移動や財・サービス・情報などの流通は、一つの地域において独立して生じるものではない。そのため、人口、旅客数、あるいは経済活動など、ある地域の変動が他地域の変動を直接、あるいは間接的に惹起するといった地域間の依存関係を視覚化することは、その全体構造を把握する上で重要である。一般的に、地域における上記の動態を知るためのデータとして、国や各自治体で整備されている各種統計データが存在する。近年、それらの統計データの整備・公開が積極的に進められており、誰でも手軽に、デジタルデータとして整備された膨大な量の政府統計情報にアクセスすることが可能になった。

行政で整備されている統計情報の多くは、行列形式のデータである。このような地域に関する統計情報から、地域間の関係を視覚化する試みは以前から行われてきた。最も簡単なものは、地域間の関係量を希望線図により、表現するものである。例えば、トリップ数に比例した太さの線により、地域間をつなげて表現することは、これまでも多く用いられてきた。より詳細な関係を把握するため、クラスター分析や主成分分析により、類似度に基づいて地域をいくつかのグループに分類する方法^{1,2)}も

挙げられる。これらによれば、地域の全体的な傾向の把握が可能であるものの、階層的な表現や相関関係の把握に留まるという限界がある。あるいは、空間的相互作用モデルなどの重回帰分析により、地域間関係のモデリングを行うことができるが、局所的なモデルが一般的であり、全体の最適性を考慮することが困難である。

一方で、近年の計算機の進展から、膨大なデータに対するデータマイニング技術が発達し、グラフィカルモデリングなどの変数間の依存関係を抽出する手法も登場してきた^{3,4)}。グラフィカルモデリングによれば、確率的なモデリングにより、階層構造や相関関係に留まらない依存関係などの自由な表現が可能であり、変数間の局所的な関係性だけでなく変数集合全体でモデルを最適化することができる。そのため、行列形式のデータから、地域間の依存関係を抽出し、それを視覚的に表現することが可能になると考えられる。

以上の背景の下、本研究では、行列形式で表現された地域に関する統計データ、特に人の移動に着目し、ODデータから、地域間の依存関係を視覚化する手法を構築することを目的とする。そして、観光目的の集中度に着目し、地域間の集客に関する依存関係を把握することの適用可能性を検証する。

2. 依存関係視覚化の基本的手法

(1) グラフィカルモデリング

近年、複数の変数間における依存関係を、グラフによって表すことにより、変数間の構造を視覚的に表す手法が多数登場してきた⁹⁾。これらの手法はグラフィカルモデリングと呼ばれる。一般的に依存関係は、変数をグラフにおけるノードとし、リンクを有向グラフまたは無向グラフとして表現する。そして、変数間の依存関係は、説明変数と被説明変数の関係、あるいは条件付き確率の関係で捉えられる。代表的なものとして、構造方程式モデリング、グラフィカルガウシアンモデル、ベイジアンネットワークなどが挙げられる。

構造方程式モデリング^{3), 6)}は、グラフ構造（パスダイアグラム）により統計モデルを表現する。一般的には、仮説に基づき、因子分析などにより、潜在変数を含めたグラフ構造を構築した後、適合度指標などから、モデル説明力などの検証を行う。構造方程式モデリングでは、既存知識から手動により依存関係をモデル化するため、データから任意のグラフ構造に対して、網羅的に最適なモデルを探索することには向かない。また、多くの構造方程式モデリングでは、変数間の関係に線形性を仮定する。非線形の関係性を仮定した構造方程式モデリング手法⁷⁾も提案されているが、まだ一般的に用いられるまでには至っていない⁸⁾。

グラフィカルガウシアンモデル^{3), 9)}（離散変数の場合、グラフィカル対数線形モデル）は、データが多変量正規分布に従うという仮定の下、変数間の条件付き独立性を探索し、無向グラフにより表現する方法である。具体的には、偏相関係数が0となるモデルを仮定し、その制約の下、標本相関係数行列とのズレが最も小さくなるように母相関係数行列を推定する。ズレの評価基準を母相関係数行列と標本相関係数行列に依存した逸脱度とし、これを用いて適合度検定を行う。しかしながら、前述の通り、相関関係を無向グラフとして表現するため、依存関係を表現することはできない。

ベイジアンネットワーク^{6), 10)}は、複数変数間の依存関係を条件付き確率で表現したネットワークモデルである。グラフ構造は既存の知識により事前に作成されることが多く、このグラフ構造に基づき、確率推論を行う。確率推論では、確率変数間の条件付き確率の積から、全ての確率変数の生起確率を計算することができる。変数間の依存関係を条件付き確率で表現することにより、柔軟なモデルを作成することが可能である。ただし、グラフ構造は、非閉路有向グラフ、すなわち、任意ノードから出発してリンクをたどっても、出発ノードに戻ることはないものに限定される。近年は、グラフ構造の推定法も提案され、依存関係の抽出法として、着目されている。

本研究の目的を鑑みると、地域間の依存関係を表現することができ、さらにその関係を抽出することができ、より自由な表現を可能とする条件付き確率に基づく、ベイジアンネットワークに着目する。

(2) ベイジアンネットワーク

ベイジアンネットワークでは、変数 X から変数 Y への有向グラフが存在した場合、 X を Y の親変数とよび、 Y を X の子変数とよぶ。変数間の依存関係には、1次のマルコフ性を仮定する。すなわち、各変数は与えられたグラフ構造 G において、親変数にのみ依存する。例えば、 p 個の確率変数の同時確率は

$$P(X_1, \dots, X_p) = \prod_{i=1}^p P(X_i | pa(X_i)) \quad (1)$$

と分解できる。ここで、 $pa(X_i)$ は、 G における X_i の親変数集合である。確率推論においては、親変数集合が既知であるため、上記の確率分解から、任意の変数の生起確率が計算可能である。例えば、遺伝子間の発現の依存関係をベイジアンネットワークによって記述した遺伝子制御ネットワーク研究が2000年頃から確立され、その成果が創薬や治療などで効果をあげている⁹⁾。その他、多数の実現象の記述に対しても適用例が報告されており、その有用性が評価されている。

しかしながら、本研究で取り扱う地域間の依存関係ではグラフ構造が未知であるため、その推定が必要である。ベイジアンネットワークのグラフ構造推定では、全ての変数間に対して、情報量規準などに基づき、最適な依存関係を同時に推定する。複数の地域間には、複雑に依存関係が存在していることが想定されるため、地域間の依存関係を表現するモデルとして適していると考えられる。また、一度、依存関係を構築すれば、確率推論の考え方をを用いることができ、その応用可能性も高い。

3. 地域間依存関係の視覚化手法の構築

(1) 問題の設定

いま、 p 個の地域があり、確率変数ベクトルとして、各地域の分布交通量によるベクトルを考える。本研究では、到着地に着目し、OD表の列ベクトルを、この確率ベクトルとし（図-1）、 $\{x_1, \dots, x_p\}$ で表わす。また、観測

	目的地						計	
	A	B	C	...	X	...		Z
出発地	A	0	$q_{A \rightarrow B}$	$q_{A \rightarrow C}$...	$q_{A \rightarrow X}$...	$q_{A \rightarrow Z}$
	B	$q_{B \rightarrow A}$	0	$q_{B \rightarrow C}$...	$q_{B \rightarrow X}$...	$q_{B \rightarrow Z}$
	C	$q_{C \rightarrow A}$	$q_{C \rightarrow B}$	0	...	$q_{C \rightarrow X}$...	$q_{C \rightarrow Z}$

Z	$q_{Z \rightarrow X}$...	0	
計	他県からX県に向かう交通量						集中交通量	

図-1 確率変数ベクトルの定義

値は変数間の依存関係を表したグラフ構造 G からの実現値とみなす。この確率変数ベクトルにより構成される行列を、データ行列 $D = (x_{ij})_{1 \leq i \leq p, 1 \leq j \leq p}$ とする。なお、グラフ構造 G は直接観測できないものとする。

観測できないグラフ構造 G の推定グラフ \hat{G} を、観測データ行列 D から求めることになる。ベイジアンネットワークの推定問題では、最小記述長MDL(Minimum Description Length)や、BDM(Bayesian Dirichlet Metric)など、様々な規準が用いられてきたが、現在最も一般的なグラフ選択の規準となっているものは事後確率(MAP解)である⁹⁾。この場合、データ行列 D が与えられた下でのグラフ構造 G の事後確率

$$P(G|D) = \frac{1}{Z} P(D|G) \cdot P(G) \quad (2)$$

の最大化に基づいてグラフ構造を選択する。なお、 Z は規格化乗数である。ここで、尤度の計算のため、条件付き確率の関数形を定義する。親変数と子変数を、それぞれ、説明変数と被説明変数としてモデル化を行う¹¹⁾。このモデル化に関しては、次節で説明する。また、多くの場合、事前確率に一様分布を仮定し、尤度最大化によりグラフ構造を推定しているが、本研究では、事前知識を可能な限り利用するべく、事前確率も導入する。

(2) 尤度関数の設定

尤度計算において、説明変数(親変数)と被説明変数(子変数)の関係を、条件付き確率で表現するため、確率モデルの設定を行う。モデルとしては、より複雑な関係性を表現できるよう、非線形回帰モデルを用いる。特に、遺伝子ネットワーク推定で実績のある、B-スプラインによる非線形回帰モデル¹²⁾を用いる。

a) B-スプライン基底関数

B-スプラインは、複数の基底関数の線形結合として、滑らかな関数を表現したものである。それぞれの基底関数は、制御点と呼ばれる点で滑らかに連結している。一般的には、制御点の間隔は一定とする。各基底関数は、説明変数とするデータ点と制御点・基底関数の個数をそれぞれ与えれば、一意に定まる関数である。例えば、3次多項式で構成される3次B-スプラインは、図-2の通り模式的に示すことができる。以下、3次B-スプラインを例に記述する。

基底関数は、データ $\mathbf{x} = [x_1, \dots, x_p]^T$ に対して、その存在区間 $[\min x_i, \max x_i]$ を等間隔に分割し、各区間を4つの基底関数で覆うように構成する。 i 番目の制御点の値を t_i とした場合、 m 番目の基底関数 $b_m(x)$ は、 t_m から t_{m+4} までの5つの制御点で滑らかに連結されており、制御点の幅 d 、および基底関数の対称中心 $c_m (= t_{m+2})$ を用いて定義すれば

$$b_m(x) = \begin{cases} \frac{1}{6d} \left\{ \left(2 - \frac{|x - c_m|}{d} \right)^3 - 4 \left(1 - \frac{|x - c_m|}{d} \right) \right\}, & |x - c_m| \leq d \\ \frac{1}{6d} \left(2 - \frac{|x - c_m|}{d} \right)^3, & d < |x - c_m| \leq 2d \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

と表わされる。

b) モデルの作成

前述のB-スプライン基底関数を用いて、非線形回帰モデルを作成する。回帰関数 m_i は、

$$m_i(x_{ij}) = \sum_{m=1}^{M_i} \gamma_{im} b_{im}(x_{ij}) \quad (4)$$

と表すことができる。ここで、 γ_{im} は回帰係数であり、

M_i は説明変数 x_i に関する基底関数の個数である。また、確定項 $\mu(\mathbf{x}_i)$ は全ての説明変数の回帰関数の和を用いて、加法モデルにより表される。したがって、

$$\mu(\mathbf{x}_j) = \sum_{i=1}^p m_i(x_{ij}) \quad (5)$$

となり、被説明変数 y_j は

$$y_j = \sum_{i=1}^p \sum_{m=1}^{M_i} \gamma_{im} b_{im}(x_{ij}) + \varepsilon_j \quad (6)$$

とモデル化できる。ここで、誤差項を $\varepsilon_j \sim N(0, \sigma^2)$ と仮定する。被説明変数の確率密度関数は、

$$f(y_j | \mathbf{x}_j; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2\sigma^2} \left(y_j - \sum_{i=1}^p \sum_{m=1}^{M_i} \gamma_{im} b_{im}(x_{ij}) \right)^2 \right] \quad (7)$$

となる。ここで、 $\boldsymbol{\theta}$ はパラメータベクトルである。これから尤度関数が得られる。が、データが p 個取得されたとすると、その対数尤度関数を最大化すればよい。

本手法では、基底関数が多く用いられており、モデルが柔軟になり過ぎ、推定パラメータはデータに強く依存する。そのため、回帰曲線の局所変動の程度を考慮に入れた、罰則付き対数尤度を最大化することを考える。

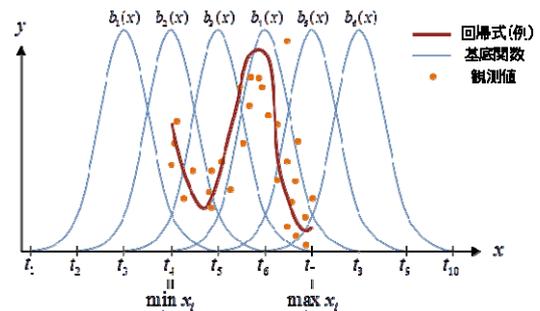


図-2 3次B-スプライン関数を用いた非線形回帰

$$l_{\lambda}(\boldsymbol{\gamma}, \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \left(\mathbf{y} - \sum_{i=1}^p \mathbf{b}_i \gamma_i \right)^2 - \frac{n\lambda}{2} k(\boldsymbol{\gamma}) \quad (8)$$

ここで、目的関数 y_i 、およびパラメータ γ_{im} と b_m はベクトル表示をしており、 λ は平滑化パラメータ、 $k(\boldsymbol{\gamma})$ は罰則項である。なお、尤度計算は、親子関係にある変数間のみで行うものとする。

(3) 事前確率の導入

本研究では、地域間の直線距離を事前確率に反映させる。これは、より近い地域の影響が大きいことを考慮するものである。地域 i, j 間の基準化距離を、

$$d(i, j, \alpha) = \frac{d(i, j)^{\alpha}}{\sum d(i, j)^{\alpha}} \quad (9)$$

とし、事前確率を

$$P(G) = \frac{1}{Z} \exp \left[-w \sum_{(i,j) \in E} d(i, j, \alpha) \right] \quad (10)$$

と定義できる¹³⁾。ここで w は重み係数である。

(4) グラフ全体における事後確率計算

これまでみてきたように、尤度関数および事前確率から、事後確率を計算することができる。全地域に対する同時確率は、式(1)の通りであり、尤度関数の積と等しくなる。よって、グラフ構造 G の対数事後確率は、

$$\begin{aligned} \ln P(G|D) &= \ln P(D|G) + \ln P(G) + const. \\ &= \sum_{j \in \text{全都市}} l_{\lambda j} - w \sum_{(i,j) \in E} d(i, j, \alpha) + const. \end{aligned} \quad (11)$$

となる。

(5) グラフ構造推定アルゴリズム

上記の事後確率最大化を適用するにあたって、非閉路有向グラフを網羅的に探索するのは、その膨大な組み合わせ数から事実上困難である。そのため、近似的に最適解を探索する。ここでは、greedy hill-climbing algorithm¹²⁾を用いて逐次探索を行う。

具体的には、現在のグラフ構造の事後確率と、その近傍のグラフ構造の事後確率を比較し、高い事後確率のグラフ構造を保存することにより、逐次、最適なグラフ構造に近づけていく。近傍のグラフ構造とは、現在のグラフ構造から、ある変数に対して、(a) 新しい親変数を一つ付け加える、(b) いま繋がっている親変数を一つ取り除く、(c) いま繋がっている子変数とのリンクを反転させる、(d) 何もしない、の4つの操作のいずれか一つを行ったグラフ構造のことである。これらのグラフ構造のうち、情報量基準に基づき、最適な操作を採択する。

4. 適用

(1) 適用データ

提案手法を、全国幹線旅客純流動調査に適用する。本調査における都道府県間のODデータのうち、観光目的(全機関)の到着地、すなわち、旅客の集客に着目して、都道府県間の依存関係の視覚化を試みる。対象とする調査年は、1990年、1995年、2000年、2005年とし、この各データに提案手法を適用する。なお、適用対象データにおいて、三大都市圏については、都市圏内の流動は調査対象外となっているため、首都圏(埼玉県、千葉県、東京都、神奈川県)、中京圏(岐阜県、愛知県、三重県)、関西圏(京都府、大阪府、兵庫県、奈良県)としてまとめて扱う。また、北海道については、道央、道南、道北、道東に区分されており、これをそのまま適用する。この結果、42県・地域が適用対象数となる。

前述の通り、提案手法を適用するにあたり、到着地別の列ベクトルを確率変数ベクトルとする(図-1)。このデータから、都道府県をノードとし、依存関係を有向グラフで表現することが目的となる。

(2) 適用結果と考察

提案手法を適用するに当たり、いくつかのパラメータを試行錯誤により設定した。B-スプライン非線形回帰モデルに関する、基底関数の数を $M=6$ 、平滑化パラメータを $\lambda=10$ とした。また、事前確率に用いる基準化距離、また重み係数に対しては、 $\alpha=1$ 、 $w=250$ とした。

適用結果のうち、2005年のものを図-3に示す。

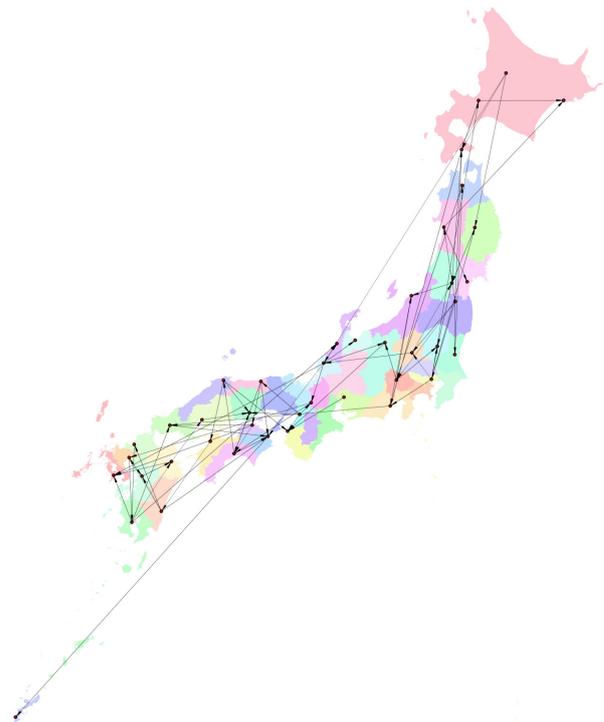


図-3 提案手法の適用結果(2005年)

ベイジアンネットワークに基づけば、有向グラフが存在する2地域間（A→B）には、全国からBに向かう観光目的の分布交通量は、全国からAに向かう分布交通量により説明されると解釈することができる。

1990年から2005年まで、全ての期間において有向グラフが現われたリンクは、「関西圏→滋賀」、「佐賀→鹿児島」であった。これらの県・地域間では、その結びつきが強いリンクであると言える。

他の特徴的なリンクについても、考察を加える。なお、以下で示す図は、判読性のため、地理的位置を考慮せず、対象ノードのみを示すこととする。

まず、中京圏、滋賀、和歌山、関西圏の関係に着目する（図-4）。前述の通り、「関西圏→滋賀」は、全ての期間において、依存関係が確認される。一方、「中京圏→和歌山」は、1990年、1995年には依存関係が見られなかったが、2000年、2005年には依存関係が現われている。和歌山県内では、熊野三山への参詣道（いわゆる、熊野古道）が、「紀伊山地の霊場と参詣道」として、2000年に国の史跡に指定され、2004年には世界遺産に認定された。以上のような観光地指定によるブームによって、近隣から熊野周辺に向かう観光客が有意に増え、中京圏への観光集客が、和歌山県への観光集客に影響を及ぼすようになったと考えられる。ここで、単純な流動数の変動と比較を行う。表-1に、「関西圏→滋賀」と「関西圏→和歌山」の分布交通量、および全国からの集中交通量に対する比率を示す。実際の分布交通量の変動やそれぞれの地域を目的地とした集中交通量に対する比率からは、依存関係の出現傾向に違いがあることは読み取れない。和歌山も滋賀も、全ての期間で最多流動相手は関西圏であり、滋賀と同様に、和歌山も強い依存関係があるように見える。一方で、提案手法によれば、これまでの、分布交通量のみでの比較だけでは明らかにならなかった依存関係の抽出が可能になる。

同様に「関西圏→岡山」、「広島→岡山」、「高知→岡山」の三つのリンクも、1990年、1995年には依存関係は確認できないが、2000年、2005年には依存関係が出現した。岡山県内では、1997年に岡山自動車道（岡山総社IC・北房JCT間）が開通したことにより、中国自動車道と山陽自動車道の接続高速道路ができ、関西方面・広島方面とのアクセス状況に変化が生じたと思われる。また、1998年には二番目の本州四国連絡橋である神戸・鳴門ルート（明石海峡大橋）が開通したことにより、一番目の本州四国連絡橋であった瀬戸大橋の、本州側の玄関口である岡山県にも四国各県との交通量の依存関係に何らかの変化が生じたかもしれない。

「福岡→鹿児島」では、1990年から2000年までは依存関係が見受けられなかったにもかかわらず、2005年では依存関係が生じたリンクも存在する。このような変化が

生じたのは、2004年3月の九州新幹線（新八代・鹿児島中央間）の開業によって、北九州と南九州の結びつきが強くなり、福岡県の観光の集客数が、鹿児島県の観光の集客数に有意に影響を及ぼすようになったという解釈も可能である。

一方で依存関係が消滅した例として、青森、岩手、秋田、山形に着目する（図-5）。「秋田→山形」のように、1990年では依存関係が出現していたにもかかわらず、1995年以降は依存関係が認められない。山形新幹線や秋田新幹線が開業する以前は、上野から福島・山形を經由して秋田に向かう特急列車なども存在していたが、1992年に山形新幹線（福島・山形間）が開業したことで、山形はそれ以降、福島や首都圏などの結びつきが強くなり、秋田との結びつきが弱くなったことで、依存関係が見られなくなったと解釈することができる。

また、「佐賀→宮崎」では、1990年から2000年までは依存関係が存在していたにもかかわらず、2005年の結果からは依存関係が反転し、「宮崎→佐賀」の依存関係が見られた。依存関係の反転が生じた明確な原因は不明であるが、上述の九州新幹線の開業により、九州内の観光による集客数のバランスが変化して、生じたものであると考えられる。

以上の通り、分布交通量の比較のみからでは行うことのできない、依存関係の把握を行うことができた。ただし、ここでの結果の解釈は、その一例に過ぎず、全てを説明できているわけではない。解釈の妥当性の判断が介入することは、提案手法の限界である。



図-4 中京圏、滋賀、和歌山、関西圏の依存関係の変化

表-1 分布交通量と集中交通量の比率の変化

	関西圏→滋賀県 (全期間依存関係あり)		関西圏→和歌山 (2005年のみ依存関係あり)	
	分布交通量 (人/日)	関西圏→滋賀 /全国→滋賀	分布交通量 (人/日)	関西圏→和歌山 /全国→和歌山
1990	13740	75.2%	7963	77.2%
1995	7884	54.9%	6953	67.2%
2000	17367	71.2%	12668	73.2%
2005	14015	67.7%	11001	78.3%



図-5 青森、岩手、秋田、山形の依存関係の変化

5. おわりに

本研究では、ベイジアンネットワークに基づき、行列形式の統計データ、特にODデータから、地域間の依存関係を視覚化する手法を構築した。ここでは、到着地の地域に着目し、その分布交通量を特徴ベクトルとして捉え、特徴ベクトル間の親子関係を、それぞれ説明変数、被説明変数とするB-スプライン非線形回帰モデルで表現し、事後確率最大化の下、そのグラフ構造を推定した。さらに、他の情報を事前確率として扱うことにより、複数データの統合を可能としている。提案手法を、全国幹線旅客純流動調査の観光目的のODデータに適用し、観光による集客数の県や地域間の依存関係の視覚化を行った。調査期間ごとの適用結果を比較することにより、依存関係の変化（依存関係の出現・消滅・反転など）を把握することができた。

提案手法により、多様なデータに対して依存関係の視覚化を行うことが期待されるが、今後の課題としては、以下の通りである。今回は、時間断面ごとに手法の適用を行ったが、時系列データに対して、逐次グラフ構造を推定することや、全データをまとめて時空間でグラフ構造を推定することなど、時系列データに対する分析手法の拡張が挙げられる。また、クラスター分析などの手法と併用することにより、事前に地域をグループ化することも考えられる。さらに、今回は、事前確率は直線距離としたが、より複雑な事前確率の検討も望まれる。最後に、複雑な依存関係に対応できるよう非線形回帰モデルを尤度関数として設定したが、線形回帰モデルを用いることも視野に含めるべきである。線形の関係性のみに限定されるものの、推定結果の解釈の方法が広がるためである。回帰モデルの選定は、分析目的により異なるため、目的別の回帰モデルの整理を行うことが必要である。

参考文献

- 1) 例えば、湯川創太郎: 地方都市交通の現状把握と類型化 —多変量解析による都市・交通指標の作成と都市の分類—, 交通学研究, Vol.54, pp.195-204, 2010.
- 2) 例えば、足永靖信, 尹聖暎: 東京都 GIS を用いた都市構造の統計分析と気温観測値との関連性, 日本建築学会環境系論文集, Vol.581, pp.81-86, 2004.
- 3) 宮川雅巳: グラフィカルモデリング, 朝倉書店, 1997.
- 4) C.M.Bishop: *Pattern Recognition and Machine Learning*, Springer, New York, 2006.
- 5) J.パール, 黒木学訳: 統計的因果推論 —モデル・推論・推測—, 共立出版, 2009.
- 6) 宮川雅巳: 統計的因果推論—回帰分析の新しい枠組み—, 朝倉書店, 2004.
- 7) 例えば, A.J.Bauer: A semiparametric approach to modeling nonlinear relations among latent variables, *Structural Equation Modeling*, Vol.12, No.4, pp.513-535, 2005.
- 8) 豊田秀樹: 共分散構造分析 [実践編] —構造方程式モデリング—, 朝倉書店, 2009.
- 9) 日本バイオインフォマティクス学会編: バイオインフォマティクス事典, 共立出版, 2006.
- 10) 本村陽一, 岩崎弘利: バ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.
- 11) N.Friedman, M.Linial, I.Nachman and D.Peer: Using Bayesian networks to analyze expression data, *Journal of Computational Biology*, Vol.7, No.3-4, pp.601-620, 2000.
- 12) S.Imoto, K.Sunyoung, T.Goto, S.Aburatani, K.Tashiro, S.Kuhara and S.Miyano: Bayesian network and nonparametric heteroscedastic regression for nonlinear modeling of genetic network, *Proceeding of the First IEEE Computer Society Bioinformatics Conference*, pp.219-227, 2002.
- 13) Y.Tamada, S.Y.Kim, H.Bannai, S.Imoto, K.Tashiro, S.Kuhara and S.Miyano: Estimating gene networks from gene expression data by combining bayesian network model with promoter element detection, *Bioinformatics*, Vol.19(Suppl.2), pp.227-236, 2003.

(2013.?? 受付)

VISUALIZATION OF INTER-REGIONAL DEPENDENCE BASED ON OD DATA BY USING BAYESIAN NETWORK

Takashi FUSE and Takuya WATANABE

Inter-regional dependence of traveling is important to understand the regional characteristics. So far, various statistics have been prepared for grasping the regions. Recently, extraction methods of dependence between variables using Bayesian network have been developed. This study proposes visualization method of inter-regional dependence based on OD data by using the Bayesian network. The Bayesian network expresses the dependence between variables as directed graph with conditional probability. The conditional probability is modeled by B-spline non-linear regression model. The graph structure is inferred based on maximum a posteriori probability. We apply the proposed method to the Inter-Regional Travel Survey in Japan, and attempt to visualize inter-regional dependence.