

# ガウシアングラフィカルモデルを用いた ネットワーク交通状況の補間推定に関する研究

花岡 洋平<sup>1</sup>・安田 宗樹<sup>2</sup>・桑原 雅夫<sup>3</sup>・原 祐輔<sup>4</sup>

<sup>1</sup>学生非会員 東北大学大学院 情報科学研究科 (〒 980-8579 仙台市青葉区荒巻青葉 6-3-09)

E-mail: hanaoka@plan.civil.tohoku.ac.jp

<sup>2</sup>非会員 山形大学大学院准教授 理工学研究科 (〒 992-8510 山形県米沢市城南4丁目3-16)

E-mail: muneki@yz.yamagata-u.ac.jp

<sup>3</sup>正会員 東北大学大学院教授 情報科学研究科 (〒 980-8579 仙台市青葉区荒巻青葉 6-3-09)

E-mail: kuwahara@plan.civil.tohoku.ac.jp

<sup>4</sup>正会員 東北大学大学院助教 未来科学技術共同研究センター (〒 980-8579 仙台市青葉区荒巻青葉 6-3-09)

E-mail: hara@plan.civil.tohoku.ac.jp

近年、不確実性を持ったデータを扱う方法として、統計・確率的な手法が注目されている。これらの手法は、不確実性を持ったデータを扱うには非常に有効な手段であるが、交通流の物理的な性質がほとんど用いられていないため、交通流の本来満たすべき物理的性質を表現することが困難である。そこで本研究では、空間的に散在した交通データの補間推定について、これまでに提案されている統計・確率的な手法に交通流の基本的な性質である交通量保存則の制約を加えることにより、モデルの拡張を行った。また、テストネットワークを用いて本研究で定式化したモデルによる予測実験を行ったところ、良好な予測結果が得られ、補間推定モデルとしての妥当性が確認された。

**Key Words :** *interpolation estimation ,gaussian graphical model ,missing values*

## 1. はじめに

道路ネットワークの管理・運用計画や、道路交通の円滑化施策などを実施するためには、道路ネットワーク全体にわたって交通流の動的な挙動を的確に把握することが必要である。日々の交通状況の観測手段としては、プローブデータや車両感知器などが挙げられる。しかし、これらの交通データは時空間的に散在しており、交通状況の把握を行うためには未観測箇所の面的な補間推定を行う必要がある。

未観測リンクの交通量の推計には、これまでに様々なモデルが開発されている。外井・天本<sup>1)</sup>は、観測交通量とリンクのOD構成率とを用いて、未観測リンクの交通量の推計手法を示した。また朝倉ら<sup>2)</sup>は、観測リンク交通量の線形変換による、未観測リンクの交通量推計を行うモデルを開発した。

ところで、様々な観測手法を用いて観測される交通データは、日々変動し、また多くの欠損を含んでいる。このようなデータの中から必要なものだけを取り出し、また補間することが求められるが、近年、このような不確実性を持ったデータを扱う方法として、統計・確率的な手法が注目されている。統計・確率的な手法により未観測リンクの交通状況を推定した例として、熊谷ら<sup>3)</sup>による欠損値付き主成分分析を用いたプローブデータの未観測区間の旅行時間推定手法や、日々の交通現象の不確実性に着目した片岡ら<sup>4)</sup>による画像補修フィルタ技術を応用することによる交通状況の補間推定手法などが挙げられる。これら手法は、蓄積したデータを統計的に扱ったり、データの不確実性を分布としてモデル化するという点で、非常に有効な手段である。

しかしながら、これらの手法には、交通流の物理的な性質がほとんど用いられていないため、交通流の本来満たすべき物理的性質を表現することが困難である。そこで本研究では、統計・確率的な手法に交通流の物理的な制約を加えることにより、より現実の交通現象に立脚したモデリングを行う。具体的には、車両感知器データにより観測される交通台数や交通量をデータとして扱う片岡らが提案した交通状況の補間推定手法を基本モデルとし、この手法に交通流の基本的な物理特性である交通量保存則を導入することにより、モデルの拡張を行う。

本稿の構成は、以下の通りである。2.では、モデルの定式化を行う。3.では、仙台市の中心市街地をネットワークとして予測実験を行い、モデルの予測精度を調べる。最後に4.では、本研究のまとめと今後の課題を示す。

## 2. モデルの定式化

### (1) ガウシアングラフィカルモデル

グラフィカルモデルとは、確率的相关構造を図式的に表現することにより、確率モデルの構造を可視化したものである。グラフィカルモデルは、確率モデルが設計できれば様々な処理が可能であり、パターン認識やデータマイニング等、多くの分野で利用されている<sup>5)</sup>。特に、グラフィカルモデルの確率分布が正規分布であるとき、ガウシアングラフィカルモデル (Gaussian Graphical Model:GGM) とよばれる。GGMは、多次元ガウス分布の性質から解析性に優れており、とりわけ汎用性を持ったモデルとしてさまざまな分野に応用

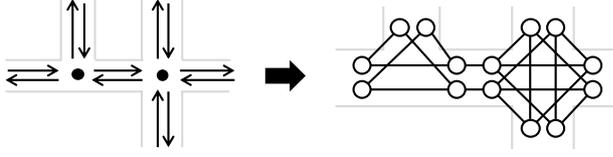


図-1 道路ネットワークと GGM のネットワーク構造との関係

されている。この GGM を交通状況の推定に応用したのが片岡らの手法である。次節では、片岡らの提案したモデルに、交通量保存則を導入したモデルの定式化を行う。

## (2) 設定

一般の道路ネットワークは、道路の結節点にノードをおき、ノードとノードの間の道路区間をリンクとして、ノードとリンクの組み合わせによって表現される。一方、本研究で用いるネットワークでは、各道路上の車両の流入地点、流出地点にそれぞれノード  $i \in V$  をおき、車両が直接移動可能なノード同士のペア  $(i, j)$  の集合を  $E$  で表し、これをリンクとする。道路の集合  $V$  とペアの集合  $E$  により、ネットワーク  $G(V, E)$  を定義する。図 (1) に実際の道路ネットワークと GGM のネットワーク構造の比較を示した。

また、各ノード  $i \in V$  上に交通量  $x_i$  を割り当てる。道路の流入部と流出部は車両感知器が主に設置されている位置であり、この位置に交通量を定義することにより、車両感知器から得られる交通データを効率よく利用することができる。

## (3) 交通量保存則を導入したモデルの定式化

ある道路の結節点  $J$  に接続する道路に流入する交通量の集合を  $J_{in}$ 、流出する交通量の集合を  $J_{out}$  を定義すると、全ての結節点において、それぞれ

$$\sum_{i \in J_{in}} x_i = \sum_{i \in J_{out}} x_i \quad (1)$$

とあらわせる。このグラフ上に、交通量保存則 (1) を導入したガウシアングラフィカルモデル

$$p(\mathbf{x}) \propto \exp\left(\beta^T \mathbf{x} - \frac{1}{2} \sum_{i \in V} \xi_i x_i^2 - \frac{\alpha}{2} \sum_{(i,j) \in J} \left(\sum_{i \in J_{in}} x_i - \sum_{j \in J_{out}} x_j\right)^2\right) \quad (2)$$

を定義する。 $\beta$  は各道路固有の交通量のバイアスであり、 $\alpha$  は交通量保存則の制約の強さを決めるパラメータである。 $\xi$  は各道路の交通量の分散に対応する。また、 $\mathbf{x} = (\mathbf{x}^{in}, \mathbf{x}^{out})$  とした。ここで、

$$\boldsymbol{\mu} := \mathbf{C}^{-1}\boldsymbol{\beta}, \quad C_{ij} = \begin{cases} \xi_i + 2\alpha, & i = j \\ -\alpha, & (i, j) \in A \\ \alpha, & (i, j) \in B \cup C \\ 0, & otherwise \end{cases}$$

とすると、式 (2) は、

$$\begin{aligned} & \exp\left(-\frac{1}{2} \sum_{i \in V} \xi_i x_i^2 - \frac{\alpha}{2} \sum_{(i,j) \in J} \left(\sum_{i \in J_{in}} x_i - \sum_{j \in J_{out}} x_j\right)^2\right) \\ &= \exp\left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \mathbf{C} (\mathbf{x} - \boldsymbol{\mu}) + \frac{1}{2} \boldsymbol{\beta}^T \mathbf{C}^{-1} \boldsymbol{\beta}\right) \end{aligned}$$

より、

$$p(\mathbf{x}|\boldsymbol{\beta}, \boldsymbol{\xi}, \alpha) = \sqrt{\frac{\det \mathbf{C}}{(2\pi)^D}} \exp\left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \mathbf{C} (\mathbf{x} - \boldsymbol{\mu})\right)$$

とあらわせる。ただし、各交差点において流入する道路  $i$  と流出する道路  $j$  のペア  $(i, j)$  の集合を  $\in A$ 、各交差点において流出する道路  $i$  と流出する道路  $j$  のペア  $(i, j)$  の集合を  $(i, j) \in B$ 、各交差点において流入する道路  $i$  と流入する道路  $j$  のペアの集合  $(i, j)$  を  $(i, j) \in C$  でそれぞれあらわす。

## (4) パラメータ推定方法

$M$  個のデータセット  $\mathcal{D}$  からなる経験分布  $Q(\mathbf{x})$  が与えられたとする。この経験分布に対して対数尤度関数は

$$\begin{aligned} \mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\xi}, \alpha) &:= \int_{-\infty}^{\infty} d\mathbf{x} Q(\mathbf{x}) \ln p(\mathbf{x}|\boldsymbol{\beta}, \boldsymbol{\xi}, \alpha) \\ &= \boldsymbol{\beta}^T \langle \mathbf{x} \rangle_{\mathcal{D}} - \frac{1}{2} \sum_{n=1}^{|\mathcal{D}|} (\xi_n + 2\alpha) \langle x_n^2 \rangle_{\mathcal{D}} \\ &\quad + \alpha \sum_{(n,m) \in A} \langle x_n x_m \rangle_{\mathcal{D}} - \alpha \sum_{(n,m) \in B} \langle x_n x_m \rangle_{\mathcal{D}} \\ &\quad - \alpha \sum_{(n,m) \in C} \langle x_n x_m \rangle_{\mathcal{D}} \\ &\quad - \frac{1}{2} \boldsymbol{\beta}^T \mathbf{C}^{-1} \boldsymbol{\beta} + \frac{1}{2} \ln \det \mathbf{C} + \text{const} \quad (3) \end{aligned}$$

となり、その勾配はそれぞれ

$$\Delta_{\boldsymbol{\beta}} \propto \langle \mathbf{x} \rangle_{\mathcal{D}} - \mathbf{C}^{-1} \boldsymbol{\beta} \quad (4a)$$

$$\Delta_{\xi_n} \propto -\frac{1}{2} \langle x_n^2 \rangle_{\mathcal{D}} + \frac{1}{2} (\mathbf{C}^{-1} \boldsymbol{\beta})_n^2 + \frac{1}{2} (\mathbf{C}^{-1})_{nn} \quad (4b)$$

$$\begin{aligned} \Delta_{\alpha} &\propto \frac{1}{|A| + |B| + |C|} \\ &\quad \left( \sum_{(n,m) \in A} (\langle x_n x_m \rangle_{\mathcal{D}} - \langle x_i \rangle_{\mathcal{D}} \langle x_j \rangle_{\mathcal{D}}) \right. \\ &\quad - \sum_{(n,m) \in B} (\langle x_n x_m \rangle_{\mathcal{D}} - \langle x_i \rangle_{\mathcal{D}} \langle x_j \rangle_{\mathcal{D}}) \\ &\quad \left. - \sum_{(n,m) \in C} (\langle x_n x_m \rangle_{\mathcal{D}} - \langle x_i \rangle_{\mathcal{D}} \langle x_j \rangle_{\mathcal{D}}) \right) \\ &\quad - \frac{1}{|A| + |B| + |C|} \left( \sum_{(n,m) \in A} (\mathbf{C}^{-1})_{nm} \right. \\ &\quad \left. - \sum_{(n,m) \in B} (\mathbf{C}^{-1})_{nm} - \sum_{(n,m) \in C} (\mathbf{C}^{-1})_{nm} \right) \quad (4c) \end{aligned}$$

となる。推定値は対数尤度関数を最大とするパラメータ  $\boldsymbol{\beta}, \boldsymbol{\xi}, \alpha$  であるから、尤度の最大点では以上の関係式を満たすことを用いて、パラメータ推定を行う。

### (5) 未観測道路の交通量推計方法

ネットワークにおける全道路  $i \in V$  は次の2つに場合分けできる.

- 道路状況が不明な (観測されていない) 道路  $i \in \mathcal{M} \subset V$
- 状況が判明している (観測されている) 道路  $i \in \mathcal{M}^* := V \setminus \mathcal{M}$

ここで,  $V \setminus \mathcal{M}$  とは,  $V$  から  $\mathcal{M}$  を除いた集合 (差集合) を表している. また, これらの道路について,

- 未観測道路の交通量  $\mathbf{x}^{\mathcal{M}} = \{x_i^{\mathcal{M}} | i \in \mathcal{M}\}$
- 観測されている道路の交通量  $\mathbf{y} := \{y_i \in (-\infty, \infty) | i \in \mathcal{M}^*\}$

をそれぞれ定義する. 道路交通ネットワークの交通量の推計とは, 状況の判明している道路  $i \in \mathcal{M}^*$  の交通量  $\mathbf{y}$  の観測から, 状況の判明していない道路  $i \in \mathcal{M}$  の交通量  $\mathbf{x}^{\mathcal{M}}$  を推計することである. いま, 交通量  $\mathbf{y}$  が与えられたときの尤度を,

$$p(\mathbf{y}|\mathbf{x}) := \prod_{i \in \mathcal{M}^*} \delta(y_i - x_i) \quad (5)$$

と設計することで, ベイズの定理より事後確率  $p(\mathbf{x}|\mathbf{y})$  は,

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x}, \mathbf{y})}{p(\mathbf{y})} = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x}|\beta, \xi, \alpha)}{p(\mathbf{y})} \\ \propto p(\mathbf{x}|\beta, \xi, \alpha) \prod_{i \in \mathcal{M}^*} \delta(y_i - x_i)$$

となる. 推計すべきは  $\mathcal{M}$  であるので, 未観測道路の交通量  $\mathbf{x}$  の推計はこの事後確率から算出される周辺事後確率の最大化, つまり,

$$\hat{\mathbf{x}}^{\mathcal{M}} = \arg \max_{\mathbf{x}^{\mathcal{M}}} \int d\mathbf{x}^{\mathcal{M}^*} p(\mathbf{x}|\beta, \xi, \alpha) \prod_{i \in \mathcal{M}^*} \delta(y_i - x_i) \\ = \arg \max_{\mathbf{x}^{\mathcal{M}}} p(\mathbf{x}^{\mathcal{M}}|\mathbf{y}, \beta, \xi, \alpha) \quad (6)$$

により与えられる. 式 (6) は, ガウス・ザイデル法

$$x_n^{(t+1)} \leftarrow \frac{1}{C_{nn}} \left( \beta_n - \sum_{m=1}^{n-1} C_{nm} z_m^{(t+1)} - \sum_{m=n+1}^{|V|+|E|} C_{nm} z_m^{(t)} \right) \\ z_m^{(t)} := \begin{cases} y_m & m \in \mathcal{M}^* \\ x_m^{(t)} & m \in \mathcal{M} \end{cases}$$

を利用することにより, 求めることができる.

## 3. 仙台市ネットワークによる予測実験

### (1) マルコフ連鎖モンテカルロ法 (MCMC) によるデータサンプリングとそのデータによる学習精度の検証

本章では, 仙台市の中心市街地の道路ネットワーク (道路リンク数: 322) を対象とした予測実験による検証結果を示す. 生成モデルは学習モデルと同様のガウシアングラフィカルモデル (2) であるとし, 生成モデルのパラメータをそれぞれ以下のように設定する.

$$\alpha = 0.3, \xi_i = 0.1, \beta_i \sim \mathcal{N}(50, 10^2) \quad (7)$$

またここでは, 解析の都合上, 行列  $\mathbf{C}$  を対角優位な行列に設計し直した. マルコフ連鎖モンテカルロ法 (MCMC) を用いて, この生成モデルから  $M=1000$  個のデータを

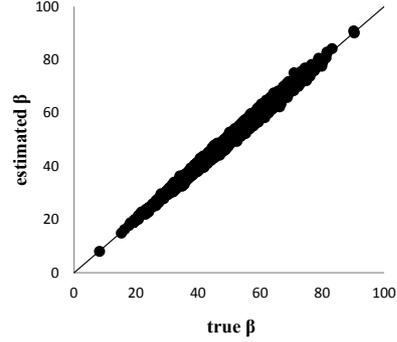


図-2 学習パラメータ  $\beta$  と真のパラメータ  $\beta$  の比較

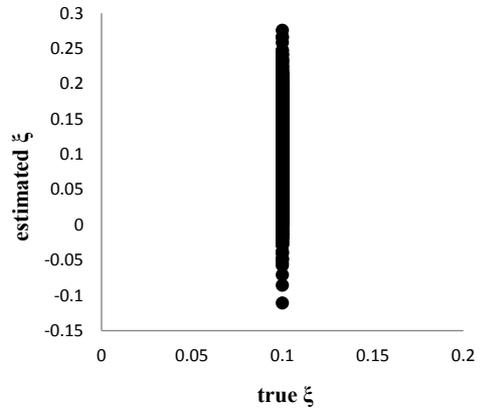


図-3 学習パラメータ  $\xi$  と真のパラメータ  $\xi$  の比較

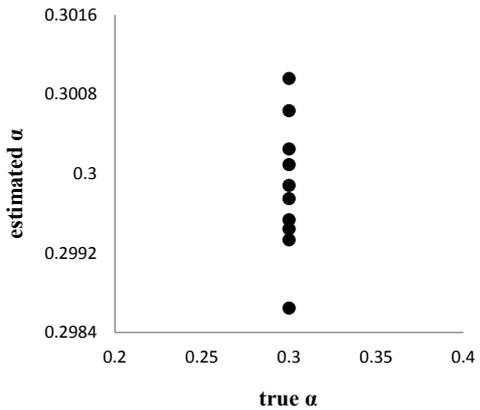


図-4 学習パラメータ  $\alpha$  と真のパラメータ  $\alpha$  の比較

生成し, 2.(4) 章の方法で学習する. 図 2~図 4 に学習パラメータと生成モデルの真のパラメータ (7) との比較を行った 5 回の試行の結果のプロットを示す.  $\beta, \alpha$  については, おおむね良好な学習が確認されている. しかし, 一方で,  $\xi$  に関してはそれほど高精度な学習が行われていない.

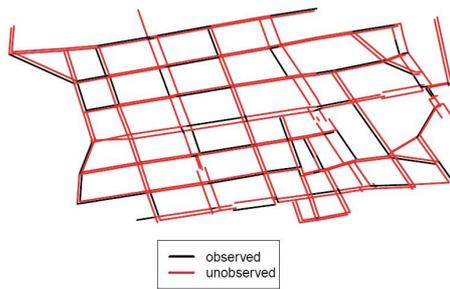


図-5 80%欠損させた例 (赤が欠損部)

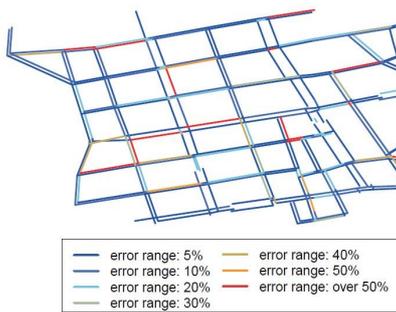


図-6 補間部の推定の結果

## (2) 交通量の推計

ここでは、人工的に生成した未観測データに対する交通量の推計を行う。データ生成には、前節で設定した生成モデルを用い、以下の方法によりデータを生成する。生成モデルの最大確率を与える配位(すなわち生成モデルの平均値  $\mu$ ) から、道路の交通量を1ステップのMCMCによって揺るがせたものをデータとする(従って、データは最大確率の点から若干ずれたものとなる)。得られたデータをランダムに80%欠損させて(つまり、残りの20%が  $y$  として得られている状況である)、2.(5)章の方法によりその補間部の推計を行う。図5は欠損させた道路の例である。

ただし、推計に用いるモデルのパラメータは、生成モデルにより  $M=10000$  個のデータをさせ、2.(4)の方法で学習したものを用いる。図6が、生成モデルにより生成したデータと推定値との相対誤差を、平均的な結果を与えるサンプルを用いて表した例である。ここでは、道路に流入する交通量と流出する交通量の差を各道路のとり値とし、生成モデルにより生成したデータ(真値)と推定値との誤差を算出した。

誤差が50%以上となった道路リンクの流入交通量と流出交通量は、真値および推定値ともにそれぞれ近い値をとっており、その差は平均1.04である。これに対して、誤差が10%台、5%以下の道路リンクの流入交通量と流出交通量の差はそれぞれ平均21.6、43.5と比較的大きな値であることから、一部の道路リンクで相対誤差が大きく算出された原因のひとつとして、道路リンクの流入交通量と流出交通量の差が小さいことが挙げられる。流入交通量と流出交通量の差が小さい道路リンクにおいて、流入交通量と流出交通量の差をとらずに算出した相対誤差は平均4.5%であり、高精度な推定が行われていることがわかる。以上から、このモデルによる推定精度はおおむね良好であるといえる。また、(1)にて学習を行った際、 $\xi$ の学習のみやや精度の低い結果が得られたが、予測実験ではおおむね良好な結果が得られたため、この $\xi$ の誤差が予測に及ぼす影響は低いものと考えられる。

## 4. おわりに

本稿では、空間的に散在した交通データの補間推定について、これまでに提案されている統計・確率的な手法に交通流の基本的な性質である交通量保存則の制約を加えることにより、モデルの拡張を行った。定式化されたモデルは、多次元ガウス分布の一般形に帰着することが確認された。また、テストネットワークにより予測実験を行ったところ、良好な結果が得られ、補間推定モデルとしての妥当性が確認された。今回の予測実験では、MCMCを用いてデータを生成したが、実際の交通データを用いたモデルの検証はまだ行っていないため、交通状況の補間推定手法としての適応可能性の確認はできていない。この検証結果は後日報告したい。

## 参考文献

- 1) 外井哲志, 天本徳浩: 非観測道路区間交通量推定のための交通量観測点の最適配置計画に関する研究, 土木計画学会・論文集, No.7, pp.251-258, 1989.
- 2) 朝倉康夫, 柏谷増男, 西山晶造: 観測リンク交通量を用いた道路網交通流の日変動推定とその信頼性分析への応用, 土木計画学会論文集, No.482/-22, pp.17-25, 1994.
- 3) 片岡駿, 安田宗樹, 田中和之: マルコフネットを用いた交通量の統計的予測, 日本物理学会講演概要集, Vol.67, No.2-2, pp.222, 2012.
- 4) 熊谷正俊, 伏木匠, 横田孝義, 君田和也: 特徴空間射影によるプローブカーデータのリアルタイム補完, 情報処理学会論文誌, Vol.47, No.7, pp.2133-2140, 2006.
- 5) C.M. ビショップ: パターン認識と機械学習 上, 丸善, 2007.

(2013.5.7 受付)

## A Study on the Interpolation Estimation of Network Traffic Conditions Using a Gaussian Graphical Model

Yohei HANAOKA, Muneki YASUDA, Masao KUWAHARA, Yusuke HARA