

統計的モデル選択による TASEP と ZRP の比較

山崎 啓介

東京工業大学大学院知能システム科学専攻

概要

交通流データを的確に表現するモデルを見つけるために複数のセルオートマトンモデルを比較することは重要である。統計学では異なるモデルの比較はモデル選択問題として知られており様々な定量的指標が提案されている。我々は構造の最適化に用いられるベイズ周辺尤度に着目し、多種粒子 TASEP と多種粒子 ZRP における値を時空図から計算し比較した。

Comparison between TASEP and ZRP Based on Statistical Model Selection

Keisuke Yamazaki

Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology

Abstract

Comparison of some cellular automata is important for finding the model which appropriately expresses the given traffic flow data. In statistics, this comparison is known as the model selection and many criteria have been proposed. We focus on the Bayesian marginal likelihood, which optimizes the structure of the model, and investigate the automaton comparison based on the likelihood values of multi-species TASEP and ZRP from real spacio-temporal diagram.

1 はじめに

交通流を表現するセルオートマトンモデルは様々なものが提案されており、所望のデータを的確に表すためにはパラメータの最適化のみならず、モデル間の比較が重要である。TASEP は平均速度をパラメータとする最も単純なモデルであるが、ZRP のような複雑な速度決定を行うモデルに対する優劣は対象とする交通流データの特性に強く依存する。つまり採集されたデータが単純な規則による流れの場合は TASEP で十分である可能性がある。その場合、必要以上に複雑なモデルを用いるとパラメータ推定の精度が低下することが知られており [1]、モデルの複雑さを適切に設定する必要がある。これが統計学におけるモデル選択問題であり、これまでに様々な定量的指標が提案されている [2, 3]。本稿ではベイズ統計においてモデル構造を決定する指標である

「周辺尤度」に基づきセルオートマトンの比較を行う。現実の交通流では多くの車両が混在しており運転規則が多様化していると考えられる。このような状況に対処するため多種粒子の TASEP と ZRP を対象とし、実車走行による時空図から周辺尤度を計算する。

2 多種粒子 TASEP と ZRP の統計的表現

多種粒子セルオートマトンのパラメータ推定には車両のグループ化を行う必要がある [4, 5]。ここでは時空図からの特徴量を基にグループ化およびパラメータ推定を行うためのモデル表現（統計的表現と呼ぶ）を示す。

時空図において i 番目の車両が前方の車両との車間距離 m となった回数を x_{im} とする。またその次

の時刻において前に進んだ回数を y_{im} とする. この 2 つがパラメータ推定における特徴量である. 多種粒子 ZRP では複数の最適速度関数 (OV 関数) が存在するが, k 番目の OV 関数における車間距離 m の前進確率を p_{km} とする. この OV 関数に従う車両について, 前進した回数 y_{im} に関する確率は

$$p(y_{im}|x_{im}, p_{km}) = \binom{x_{im}}{y_{im}} p_{km}^{y_{im}} (1 - p_{km})^{x_{im} - y_{im}}$$

で与えられる. 車間距離 1 から M についてまとめた特徴量を $x_i = (x_{i1}, \dots, x_{iM})$, $y_i = (y_{i1}, \dots, y_{iM})$ とすると

$$p(y_i|x_i, p_k) = \prod_{m=1}^M p(y_{im}|x_{im}, p_{km}).$$

ここで $p_k = (p_{k1}, \dots, p_{kM})$ とした. k 番目の OV 関数に従う車両の割合を a_k で表す. つまり $a_k \geq 0$, $\sum_k a_k = 1$ が制約条件となる. 同じグループに属する車両は共通の OV 関数に従うと仮定する. 通常, 時空図からは車両が従うグループは不明である. その場合, 全ての OV 関数の可能性を考慮し

$$p(y_i|x_i, w) = \sum_{k=1}^K a_k p(y_i|x_i, p_k)$$

となる. ここで $w = \{a_1, \dots, a_K, p_1, \dots, p_K\}$ とした. これを多種粒子 ZRP の統計的表現と呼ぶ.

全ての車両の特徴量を $x^n = \{x_1, \dots, x_n\}$, $y^n = \{y_1, \dots, y_n\}$ とすると

$$p(y^n|x^n, w) = \prod_{i=1}^n p(y_i|x_i, w)$$

となる. これを w の関数とみなしたものが尤度である.

多種粒子 TASEP においては車間距離による前進確率に違いがないため, $x_i = \sum_m x_{im}$, $y_i = \sum_m y_{im}$ と定義し, k 番目のグループの前進確率を p_k として

$$p(y_i|x_i, p_k) = \binom{x_i}{y_i} p_k^{y_i} (1 - p_k)^{x_i - y_i}$$

となる. 統計的表現は

$$p(y_i|x_i, w) = \sum_{k=1}^K a_k p(y_i|x_i, p_k)$$

となり, 尤度は

$$p(y^n|x^n, w) = \prod_{i=1}^n p(y_i|x_i, w)$$

で表される.

3 周辺尤度によるモデル比較方法

まずベイズ法におけるグループ分けの方法を示す. i 番目の車両が k 番目のグループに属することを $z_i = k$ と表す. 全車両についてまとめて $z^n = \{z_1, \dots, z_n\}$ とおく. このとき z^n の確率は形式的に

$$p(z^n|x^n, y^n) = \frac{p(x^n, y^n, z^n)}{p(x^n, y^n)}$$

となる. ベイズ法ではパラメータの事前確率 $\varphi(w)$ を用いて

$$p(x^n, y^n, z^n) = \int \prod_{i=1}^n a_{z_i} p(y_i|x_i, p_{z_i}) \varphi(w) dw,$$

$$p(x^n, y^n) = \int p(y^n|x^n, w) \varphi(w) dw$$

とする. 各車両に所属するグループを割り当てることは z^n の値を決めることに他ならない. そこで $p(z^n|x^n, y^n)$ から MCMC 法により z^n をサンプリングすることによりグループの割り当てを得る. $p(z^n|x^n, y^n) \propto p(x^n, y^n, z^n)$ より同時確率の値が計算できればサンプリングが可能である. 今回のモデルでは a_k の事前分布にディリクレ分布を, p_k の事前分布にベータ分布を与えることで同時分布の積分が解析的に行える [4].

次にモデル比較のための指標を示す. ベイズ推定において上記の $p(x^n, y^n, z^n)$ と $p(x^n, y^n)$ はそれぞれ完全データの周辺尤度と不完全データの周辺尤度に相当する. 本研究では計算が比較的容易な完全データの周辺尤度を選択の規準とする. MCMC 法により割り当て z^n を得ることができるため, 多種粒子 TASEP と多種粒子 ZRP において周辺尤度 $p(x^n, y^n, z^n)$ が計算できる. これは与えられた時空図から得られる特徴量とグループ割り当ての同時確率なので, その値が大きいほどデータをより良く表現しているとみなすことができる.

4 実データへの適用例

周辺尤度を実際の交通流データに対し計算した例を示す. 用いたデータは Sugiyama らが計測した円状コースを走行する 23 台の車両による時空図である [6]. 1/3[s] 毎の各車両の位置情報 (円上の位置座標) をセルオートマトンによる表現に合致するよう離散化した. その際, 離散化された領域 (セル) に車両が最大一台存在するよう, また 1 時刻ステップあたり 2 つ以上先のセルに進まないようセルのサイ

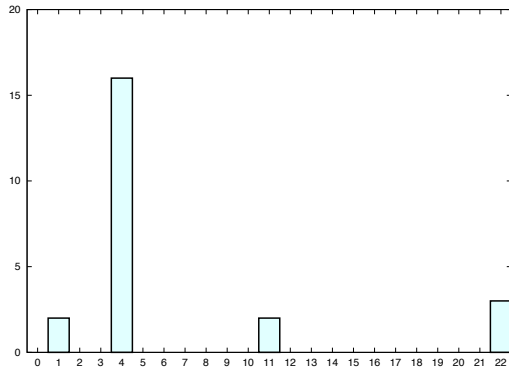


図 1: 多種粒子 TASEP におけるグループ分けの結果

グループ番号	1	4	11	22
前進確率 p_k	0.5850	0.4199	0.7164	0.5076

表 1: 各グループにおける前進確率

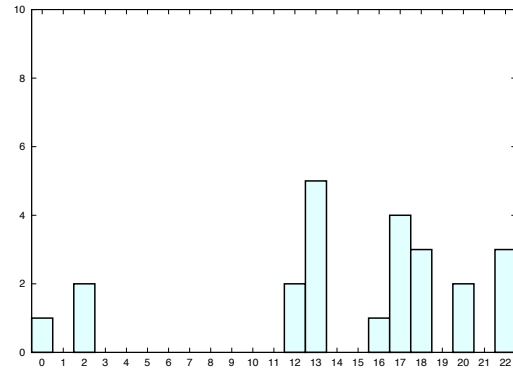


図 2: 多種粒子 ZRP におけるグループ分けの結果

モデル	TASEP	ZRP
対数周辺尤度	-19048	-17268

表 2: 周辺尤度によるモデルの比較

ズを調整した。結果として 230[m] の円周が 60 個のセルに分割された。

上記の要領で得られた時空図に対し、特徴量 x_{im}, y_{im} を数え TASEP と ZRP においてグループ分けを行った。バイズ法では a_k の事前分布のディリクレ分布の作用により不必要なグループが消去されることが知られている [7, 8]。このため双方のモデルにおいてグループ数を車両の台数と同じ $K = 23$ とした。また推定に考慮する車間距離の最大値を $M = 3$ とした。グループ分けには z^n に対する MCMC 法を用いた。サンプリングにはギブスサンプラーを用い、200 回の繰り返しの後に得られた z^n をグループ分けの結果として採用した。

TASEP におけるグループ分けの結果を図 1 に示す。横軸はグループの番号 $k = 0, \dots, 22$ を表し、縦軸はグループに属する車両の数を表す。モデルは 23 個のグループを持つが 4 つのグループのみが使用されている。

また各グループにおける前進確率を表 1 に示す。多くの車両が属するグループ 4 に比べ、他の 3 グループに属する車両が速いことがわかる。

ZRP におけるグループ分けの結果を図 2 に示す。TASEP と比較し OV 関数に多様性が増すため、9 つのグループが使用されている。

各グループにおける OV 関数を図 3 にまとめる。グループ 12 と 18 が車間距離を空けない走行をしており、グループ 20, 22 とあわせて目標速度が高い

車両と推測できる。またグループ 0 の車両は車間距離が十分でも低速で走行したとみられる。

次に TASEP と ZRP の比較を行う。表 2 は周辺尤度の対数値を示す。多種粒子 TASEP に比べ多種粒子 ZRP が大きい値を持ち、このデータの表現により適していることがわかる。ZRP における OV 関数が多様であることからドライバーが車間距離に応じた運転を行っていたことがうかがえる。

5 今後の展望

本稿では所望のデータを表現するのにふさわしい多種粒子セルオートマトンモデルの選択問題を扱った。今回は車間距離の最大値 M やグループ数 K を予め決めてから比較を行ったが、これらの量も周辺尤度で最適化することができる。それぞれのモデル変数を最良のものにした後に比較を行うことは今後の課題である。

また機械学習や自然言語処理の分野ではモデルの階層化による拡張が広く研究されている [9]。多種粒子セルオートマトンの統計的表現の多くは混合分布であるため同様の拡張が可能である。階層化による時系列の反映 [10] など時空図の特徴に即した改良を行いたい。

謝辞 本研究の一部は SCAT 研究助成、リバネス研究費 AZAPA 賞、デンソー基礎研究所共同研究費、科研費 (15K00299) の助成を受けたものである。デー

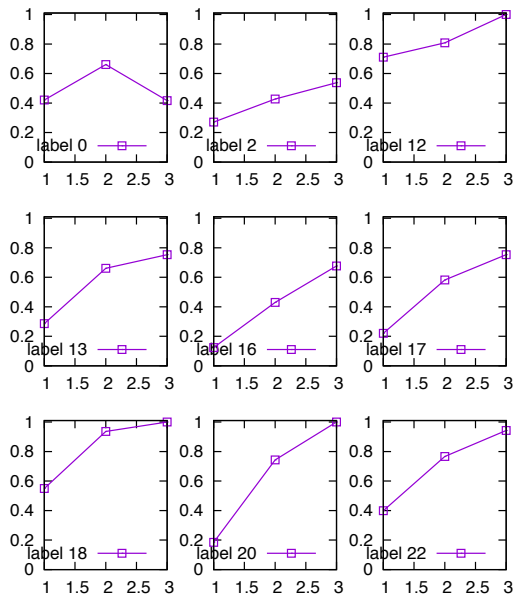


図3: 各グループのOV関数, 左上から右に向かって順にグループ0,2,12,13,16,17,18,20,22

タを公開してくださった杉山雄規先生をはじめとする研究グループの皆様, および前処理を行ってくれた中村文士氏に感謝いたします。

参考文献

- [1] H. Akaike, “A new look at the statistical model identification,” *IEEE Trans. on Automatic Control*, vol.19, pp.716–723, 1974.
- [2] K. Takeuchi, “Distribution of informational statistics and a criterion of model fitting,” *Suri-Kagaku (Mathematica Sciences)*, vol.153, pp.12–18, 1976. (in Japanese).
- [3] G. Schwarz, “Estimating the dimension of a model,” *Annals of Statistics*, vol.6 (2), pp.461–464, 1978.
- [4] 山崎 啓介, “多種粒子 TASEP を表現する混合分布モデルと統計的粒子クラスタリングについて (応用),” *日本応用数学会論文誌*, vol.24, no.4, pp.357–372, Dec. 2014.
- [5] 中村文士, 山崎啓介, “交通流映像からの速度決定則のモデル化とグループ分け,” *信学技報*,

第114巻, pp.155–160, ITS2014-53, 北海道, 2月2015.

- [6] Y. Sugiyama, M. Fukui, M. Kikuchi, K. Hasebe, A. Nakayama, K. Nishinari, S.-i. Tadaki, and S. Yukawa, “Traffic jams without bottlenecks-experimental evidence for the physical mechanism of the formation of a jam,” *New Journal of Physics*, vol.10, no.3, p.033001, March 2008. <http://stacks.iop.org/1367-2630/10/i=3/a=033001>
- [7] K. Yamazaki, “Asymptotic accuracy of distribution-based estimation for latent variables,” *Journal of Machine Learning Research*, vol.13, pp.3541–3562, 2014.
- [8] K. Yamazaki, “Asymptotic accuracy of Bayes estimation for latent variables with redundancy,” doi:10.1007/s10994-015-5482-3, 2015.
- [9] D.M. Blei, A.Y. Ng, and M.I. Jordan, “Latent dirichlet allocation,” *J. Mach. Learn. Res.*, vol.3, pp.993–1022, March 2003.
- [10] D.M. Blei and J.D. Lafferty, “Dynamic topic models,” *Proceedings of the 23rd International Conference on Machine Learning*, pp.113–120, ICML ’06, ACM, New York, NY, USA, 2006.