

# 交通渋滞形成実験における統計的モデル選択による TASEP と ZRP の比較

石原 卓弥, 山崎啓介

産業技術総合研究所 人工知能研究センター

## 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。我々は以前の研究において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。本研究ではよりナゴヤドームで行われた渋滞形成実験データにこの手法を適用し比較を行った結果を報告する。

## Comparison between TASEP and ZRP Based on Statistical Model Selection on Simulation of Traffic Flow

Takuya Ishihara, Keisuke Yamazaki

AI Research Center, National Institute of Advanced Industrial Science and 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. Previous study focused on the Bayesian marginal likelihood, which optimizes the structure of the model, and investigated the automaton comparison based on the likelihood values of multi-species TASEP and ZRP from real spacio-temporal diagram. The present study applies this method to larger-scale simulation data taken in Nagoya Dome and reports the results of comparison.

## 1 はじめに

交通流を表現するセルオートマトンモデルに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。モデルを複雑にするとデータの表現力が上がるが、パラメータの数も増えるため推定精度が低下する [1]。このため一般にはデータに最適な複雑度が存在し、それを特定することが課題となる。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]。セルオートマトンを例に説明すると、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り、ZRP は車間距離により速度決定を行う複雑なモデルである。どちらのモデルが適しているか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には TASEP で十分である可能性がある。

現実の交通流においては、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。したがって、我々は過去の研究において、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、多種粒子の TASEP と ZRP を対象として、セルオートマトンの比較を行った [4]。比較は

実車走行による時空図から周辺尤度を計算することで行われている。本研究ではこの手法をナゴヤドームで行われた大規模な渋滞形成実証実験のデータに適用し、様々な密度での走行に対し TASEP と ZRP の比較を行う。

## 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 [5, 6]。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 [5]。

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と、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

周辺尤度を実際に交通流データに対し計算した例を示す。用いたデータは、只木らがナゴヤドームで計測した半径 50m の円上コースを走行する車両による時空図である [7]。実験は 19 セッション行われ、それぞれ異なる台数の車両が実験に参加した。本研究では、このうち 8 セッションのデータを使用する。これらのデータに対して、約 0.2 秒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に離散化した。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なセルの数を求めるために、隣り合う車両同士の距離を約 0.2 秒毎に計算した。それにより、実験を通して最も車両同士が近づいたときの距離が導出でき、セルの数を求めることが出来る。今回はセルの数を実験セッション毎に調整した。表 1 に使用した実験のセッション、それぞれの実験に参加した車両数、セルの数を示す。

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。ベイズ法では  $a_k$  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 [8, 9]。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。

各実験セッションのデータにおいて TASEP と ZRP の適用を行い、比較を行った結果を表 2 に示す。2, 3 列目はそれぞれのモデルの周辺尤度の対数値を表し、4 列目は BayerFactor を示す。BayesFactor はそれぞれのモデルの対数周辺尤度の差を表す。いずれの結果においても、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

表 1: 使用した実験セッションとセルの数

セッション番号	車両の台数	セルの数
1010	10	50
1210	10	52
1310	10	48
1412	12	70
1520	20	61
1625	25	78
1730	30	80
1835	35	80

表 2: 対数周辺尤度によるモデルの比較。BF は BayesFactor を表し、それぞれのモデルの対数周辺尤度の差である。

セッション番号	TASEP	ZRP	BF
1010	-11589	-11136	453
1210	-9738	-8937	801
1310	-10606	-10045	561
1412	-16159	-15465	694
1520	-34584	-31147	3437
1625	-48116	-44755	3361
1730	-55832	-50425	5407
1835	-69776	-64914	4862

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。

次に、表 2 において、ZRP が最も優位であったセッション 1730 の結果と、TASEP との結果に最も差がなかったセッション 1010 の結果の OV 関数を分析する。実験セッション番号 1730, 1010 での各グループの OV 関数を、それぞれ図 1, 図 2 に示す。

図 1 より、セッション 1730 では、グループ 9 の車両は目標速度が高い車両であったと推測できる。また、グループ 1 は車間距離を空ける走行をしており、車間距離が十分でも低速で走行したと見られる。また、セッション 1010 に比べてグループ数が多く、多様な運転行動が行われていたことがわかる。

図 2 より、セッション 1010 では、全てのグループの車両が車間距離に余裕がある場合でも前進していないことが多く、終始十分な車間距離を取って走行していたと考えられる。これは表 1 のセルの数が 50 となっていることから裏付けられる。つまり、実験を通して、隣り合う車両同士が最大でも約 7.2m の距離までしか近づかなかったことを示しており、空間的に余裕のある走行をしていたことがわかる。

これらの分析から、セッション 1730 は平均速度だけをパラメータにもつ TASEP では適用が難しい走行データとなっていたことが考えられる。それとは異なり、セッション 1010 では、他のセッションに比べて交通流は安定的に流れ、TASEP の表現力でも概ね適用可能な走行データとなっていたため、ZRP との対数周辺尤度の差が最も小さくなっていたと考えられる。

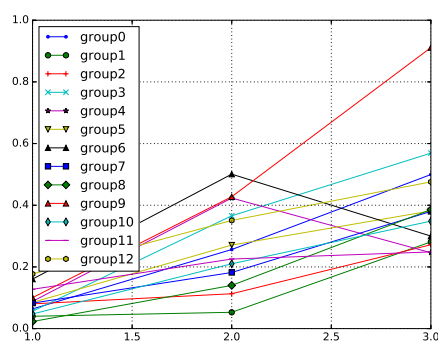


図 1: セッション番号 1730 での各グループの OV 関数

## 5 まとめと今後の展望

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取り扱った。本稿により、車両台数が多いセッション 1730 の場合に加えて、少ない車両台数が一見安定して走行しているセッション 1010 のような場合でも、ZRP モデルが優位であることが示された。

今後の展望として、より現実社会に近い状態でのセルオートマトンの選択問題を扱うことが考えられる。実際の道路を複数のセルオートマトンが連結されている状態であると見立て、道路を走る車両をモデル化して分析を行うことで、より一般的なモデル選択の法則性が見いだせる可能性がある。

**謝辞** 本研究の成果は科研費 (15K00299) の助成、および国立研究開発法人新エネルギー・産業技術総合開発機構 (NEDO) の委託業務の結果得られたものである。デー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 Auto-*

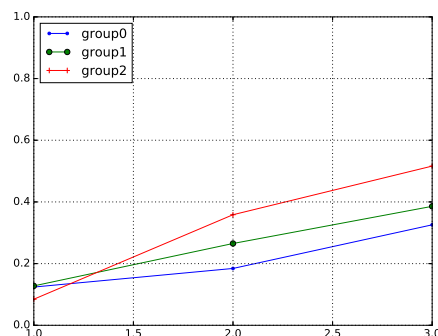


図 2: セッション番号 1010 での各グループの OV 関数

matic Control, vol.19, pp.716723, 1974.

- [2] K. Takeuchi, “Distribution of informational statistics and a criterion of model fitting,” *Suri-Kagaku (Mathematica Sciences)*, vol.153, pp.1218, 1976. (in Japanese).
- [3] G. Schwarz, “Estimating the dimension of a model,” *Annals of Statistics*, vol.6 (2), pp.461464, 1978.
- [4] 山崎啓介, “統計的モデル選択による TASEP と ZRP の比較”, 第 21 回交通流と自己駆動粒子系のシンポジウム, vol.21, pp.35-38, 2015
- [5] 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),” *日本応用数理学会論文誌*, vol.24, no.4, pp.357372, Dec. 2014.
- [6] 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.155160, ITS2014-53, 北海道, 2 月 2015.
- [7] 只木進一, 菊池誠, 福井稔, 中山章宏, 西成活裕, 柴田章博, 杉山雄規, 吉田立, 湯川論, “交通渋滞形成実験”, <http://traffic.phys.cs.is.nagoya-u.ac.jp/mstf/pdf/mstf2013-17.pdf>, 2013
- [8] K. Yamazaki, “Asymptotic accuracy of distribution-based estimation for latent variables,” *Journal of Machine Learning Research*, vol.13, pp.35413562, 2014.
- [9] K. Yamazaki, “Asymptotic accuracy of Bayes estimation for latent variables with redundancy,” doi:10.1007/s10994-015-5482-3, 2015.