

# 混合ポアソンモデルによる道路状態推定

## Traffic State Estimation with Mixture Poisson Model

木下 僚<sup>†</sup>      高須 淳宏<sup>‡</sup>      安達 淳<sup>‡</sup>  
 Akira Kinoshita      Atsuhiko Takasu      Jun Adachi

### 1 はじめに

種々のセンサー類を備えた組み込みシステムが実世界に浸透し、膨大で多種多様なデータが収集可能となった。実世界のデータをサイバー空間で解析し、そこで得た情報を実世界に還元するという循環をもつサイバーフィジカルシステム (Cyber-Physical System, CPS) [1] によって、より効率的な社会が実現できると期待されている。CPS がサイバー空間から実世界への適切かつリアルタイムなフィードバックを実現するためには、第一に、観測データからセンサーを取り巻く実世界の状態を詳細に知り、あるいは予測する必要がある。

実世界から得られる具体的なセンサーデータとして、本稿では、自動車のセンサーデータであるプローブカーデータ (Probe Car Data, PCD) を取り扱う。PCD は車載センサーから得られる位置情報や走行速度などのデータである。自動車に関する観測データであるため、直接には各自動車の追跡が可能である。一方で道路に着目すれば、PCD は走るセンサーによって道路を間接的に観測したデータと捉えることもできる。

本稿は後者の視点からの道路状態推定を目標とする。まず PCD の観測値は道路と時間帯に影響され決まるという仮定のもと、データ集合の背後にある潜在的なトピックを仮定した確率モデルであるトピックモデル [2] を用いて道路状態をモデル化する。道路と時間帯によって PCD をセグメントに分け、その背後にあるトピックすなわち基本道路状態と、各セグメントでの基本道路状態の混ざり具合を求め、

### 2 混合ポアソンモデル

まず、道路区間や時間帯などによって定義されるセグメント集合があり、PCD は各セグメントごとに観測されると仮定する。PCD には時刻や位置情報の他、速度や進行方向など種々のセンサーデータが付随する場合がある。本稿ではプローブカーの観測データが  $D$  次元のベクトルであるとし、かつ、 $D$  次元ベクトルの各成分が非負整数である場合を対象として、ポアソン分布を用いたモデル化とパラメータの推定手法を検討する。表 1 に本稿で用いる記号の一覧をまとめた。

一次元のポアソン分布は平均および分散を与える正の実数  $\lambda$  で特徴づけられる。簡単のため  $D$  次元ベクトルの各成分が互いに独立であると仮定すると、ベクトル  $\mathbf{x} = (x_1 \cdots x_D)^T$  を観測する確率は、各次元の平均および分散  $\boldsymbol{\lambda} = (\lambda_1 \cdots \lambda_D)^T$  をパラメータとして、次式で与えられる。

$$\Pr(\mathbf{x}|\boldsymbol{\lambda}) = \prod_{d=1}^D \frac{\lambda_d^{x_d} e^{-\lambda_d}}{x_d!}$$

ここで、全ての道路で共通して存在する  $K$  個の「基本道路状態」を仮定し、各道路区間の詳細な状態をこれら基本状態の混合によって表現することを考える。基本道路状態は具体的にはパラメータの互いに異なるポアソン分布であり、 $K$  個のパラメータ  $\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_K$  で定まる。 $s$  番目のセグメントにおける  $k$  番目のポアソン分布の混合比率を  $\theta_{sk}$  と書くとき、このセグ

表 1 本稿で用いる記号の一覧

| 記号                       | 定義   |
|--------------------------|--|
| $S$                      | セグメントの数.   |
| $D$                      | PCD の次元.   |
| $K$                      | 混合する状態の数.  |
| $\mathbf{x}_{sn}$        | $s$ 番目のセグメントで $n$ 番目に観測された PCD.<br>$\mathbf{x}_{sn} = (x_{sn1} \cdots x_{snD})^T$ , $x_{snd}$ は非負整数.   |
| $N_s$                    | $s$ 番目のセグメントで観測された PCD の個数.  |
| $X_s$                    | $s$ 番目のセグメントで観測した PCD の列.  |
| $X$                      | 観測した全ての PCD.   |
| $\boldsymbol{\theta}_s$  | $s$ 番目のセグメントにおける $K$ 個の状態の混合比率.<br>$\boldsymbol{\theta}_s = (\theta_{s1} \cdots \theta_{sK})^T$ , $\sum_{k=1}^K \theta_{sk} = 1$ , $0 \leq \theta_{sk} \leq 1$ . |
| $\boldsymbol{\lambda}_k$ | $k$ 番目の状態を表すポアソン分布のパラメータ.<br>$\boldsymbol{\lambda}_k = (\lambda_{k1} \cdots \lambda_{kD})^T$ , $\lambda_{kd} > 0$ .  |
| $\Lambda$                | 全てのパラメータ. $\Lambda = \{\{\boldsymbol{\lambda}_k\}, \{\boldsymbol{\theta}_s\}\}$  |

メントでベクトル  $\mathbf{x}$  を観測する確率は次式で表される。

$$\Pr(\mathbf{x}|s) = \sum_{k=1}^K \theta_{sk} \Pr(\mathbf{x}|\boldsymbol{\lambda}_k)$$

以上の混合ポアソンモデルを仮定するとき、道路状態推定の目標は、観測データ  $X$  からパラメータ  $\Lambda = \{\{\boldsymbol{\lambda}_k\}, \{\boldsymbol{\theta}_s\}\}$  の値を推定することである。これは混合ガウスモデル [3] と同様に EM アルゴリズムを用いた最尤推定が可能である。EM アルゴリズムの E ステップでは、前回のイテレーションで得たパラメータ  $\hat{\Lambda}$  を用いて次式で定義される変数  $\{\gamma_{snk}\}$  の値を求め、

$$\gamma_{snk} = \frac{\hat{\theta}_{sk} \Pr(\mathbf{x}_{sn}|\hat{\boldsymbol{\lambda}}_k)}{\sum_{k=1}^K \hat{\theta}_{sk} \Pr(\mathbf{x}_{sn}|\hat{\boldsymbol{\lambda}}_k)}$$

また M ステップでは、E ステップで得た  $\{\gamma_{snk}\}$  を用いてパラメータ  $\Lambda$  を次式により更新する。

$$\lambda_{kd} = \frac{\sum_{s=1}^S \sum_{n=1}^{N_s} \gamma_{snk} x_{snd}}{\sum_{s=1}^S \sum_{n=1}^{N_s} \gamma_{snk}}, \quad \theta_{sk} = \frac{\sum_{n=1}^{N_s} \gamma_{snk}}{N_s}$$

これら更新式の導出については紙面の都合上省略する。この E ステップと M ステップを交互に繰り返すことで、混合ポアソン分布を推定することができる。

### 3 実験

前節で述べた混合ポアソンモデルとそのパラメータの最尤推定法を利用して、実際に PCD からの道路状態推定を試みた。実験には東京都千代田区一ツ橋付近の約 1km 四方の範囲において 2010 年の一年間に観測された PCD を用いた。なお、本稿では PCD の観測値のうち走行速度のみを利用した。すなわち  $D = 1$  である。

これらの PCD を、まずセグメントごとに分割した。セグメントは道路区間、進行方向、時間帯によって定義される。道路

<sup>†</sup> 東京大学, The University of Tokyo.

<sup>‡</sup> 国立情報学研究所, National Institute of Informatics.

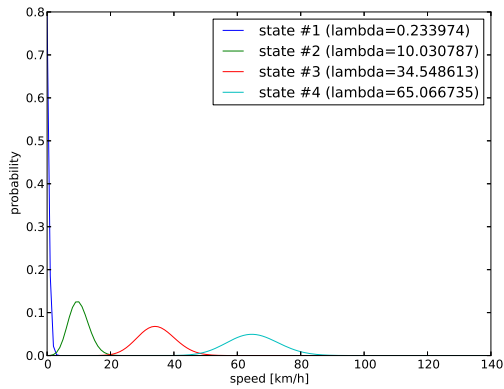
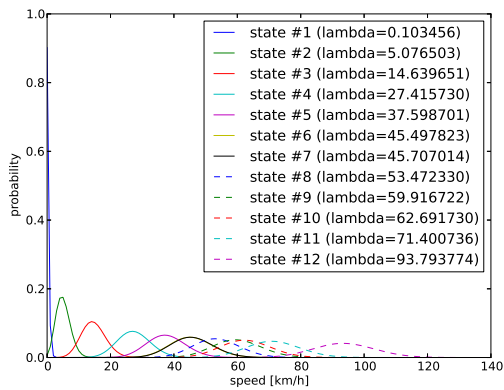
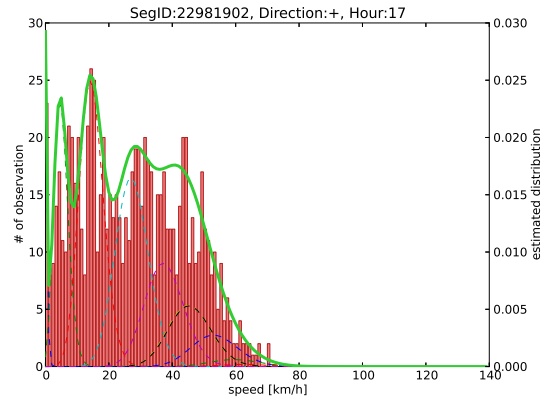
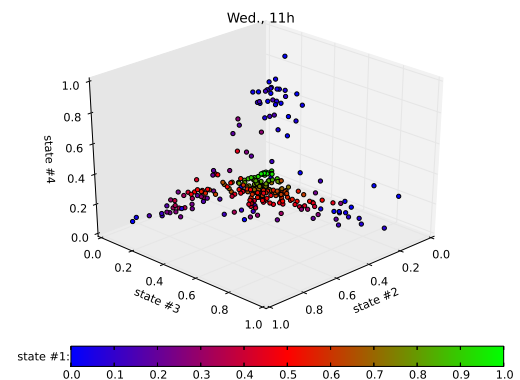
(a)  $K = 4$ (b)  $K = 12$ 

図1 推定された基本道路状態を表すポアソン分布

区間は電子地図データの情報を利用して決定し、各区間で観測された PCD を進行方向によって 2 分割した。さらに、時間帯は一週間を 1 時間ごと 168 分割することとした。PCD を各セグメントに分割した後、セグメント内の観測データが 100 個に満たないセグメントについてはサンプルが十分多くないと考えられるため、推定に利用せずにデータを破棄した。

図 1 は推定を行った結果得たポアソン分布である。このポアソン分布は全セグメントで共通して用いられる基本道路状態である。図 1(a) に示す  $K = 4$  での推定結果からは、PCD が典型的には「停止」「低速」「中速」「高速」の 4 状態をとり得て、毎回の観測値はそのいずれかから生成されていると考えることができる。図 1(b) のように  $K$  の値を大きくすると、各状態を示すポアソン分布の意味付けは困難になるが、推定の結果得られる混合ポアソン分布は実際の分布により近い確率分布となる。図 2 には、あるセグメントにおける実際の PCD のヒストグラムを赤で示し、 $K = 12$  として推定されたこのセグメントでの混合ポアソン分布を太い緑色の実線で示した。図中の破線は、この混合ポアソン分布を構成する各ポアソン分布に混合比率をかけた分布を示している。全てのセグメントで共通のポアソン分布を用い、セグメントごとに混合比率  $\theta_s$  のみを変える混合ポアソン分布を用いることで、観測データの近似的な分布を少ないデータ量で保持できる。

ポアソン分布の集合は全セグメントで共通であり、その混合具合だけが道路ごとに異なるとするモデルであるため、混合比率  $\theta_s$  を各セグメントの特徴ベクトルと見なすことができる。図 3 は、 $K = 4$  として推定された各  $\theta_s$  のうち水曜午前 11 時台のセグメントのみを抜粋して、 $\theta_s$  を 3 次元空間に射影してプ

図2 あるセグメントにおける PCD の実際のヒストグラムと、推定された混合ポアソン分布 ( $K = 12$ )図3 セグメントごとの状態混合比率 ( $K = 4$ )

ロットしたものである。各軸の state は図 1(a) の各状態の混合比率を表し、プロットの色は停止状態 (state #1) の比率を表している。この図では、セグメント集合が高速状態を強く取るクラスターとそうでないグループに分かれているように見える。

#### 4 おわりに

本稿ではプローブカーデータから道路状態を推定するためのモデルとして、共通のポアソン分布をセグメントごとに異なる比率で混合する混合ポアソンモデルを提案し、そのパラメータを推定する手法を説明した。過去のデータ全体を使って推定された各道路区間の状態は全期間を通じた平常の状態をコンパクトに表現している。随時送信される PCD の観測値が普通かどうかを判定することで早期の異常検知が可能になるほか、大多数の普通の観測値を破棄することで蓄積データ量が削減できると考えられる。今後はこれらの応用により即したりリアルタイム処理が可能となるモデルや推定手法を検討する予定である。

#### 参考文献

- [1] Edward A. Lee. Cyber physical systems: Design challenges. In *Proceedings of the 2008 11th IEEE Symposium on Object Oriented Real-Time Distributed Computing*, ISORC '08, pp. 363–369, Washington, DC, USA, 2008. IEEE Computer Society.
- [2] David M. Blei. Probabilistic topic models. *Commun. ACM*, Vol. 55, No. 4, pp. 77–84, April 2012.
- [3] Christopher M. Bishop. *Pattern Recognition and Machine Learning*, chapter 9. Springer, 2006.