

集団生活環境における生活パターンの推定手法

Estimating Regular Patterns of life in Group Living Environment

櫻井 隆平*¹ 李 周浩*¹
Ryuhei Sakurai Joo-Ho Lee

*¹立命館大学
Ritsumeikan University

We propose a new method for human activity recognition in our everyday life. In particular, we focused on the activity of people in group living environment such as office or laboratory. Monitoring the people and when each individual entered or left a room, we estimate patterns of what time period they are in the room or away. We considered such pattern as latent event by which people's schedules of daytime are determined. In group living situation, those latent events can be shared among multiple persons. For example, some of them participate a same regular meeting. We developed an unsupervised learning method to extract such patterns from the collection of enter/leave data, which is based on Bayesian nonparametrics. In particular, we modeled the schedule of activity by the beta-Bernoulli process based sparse activation of latent events and the noisy-or observation model. Experimental results show that our method can estimate the time period of each latent event behind our everyday activity, along with the number of them.

1. はじめに

人々は日常生活を営むうえで、その時間の大きな部分をオフィスや研究室などの集団生活環境で過ごしている。我々はそのような、固定的なメンバーが暮らす環境において、様々な支援を可能にするために空間の機能を拡張すべく、人々の生活行動を認識するための研究を進めてきた。本研究は、集団生活環境における人々の生活パターンを自動的に認識することが目的である。特に、人々はどの時間帯に部屋にいる、あるいはいないのかという在室・不在の典型的パターンの抽出を目指す。具体的には、集団生活環境におけるメンバーの在室・不在データから、背後に存在する不在を引き起こす要因的イベントを抽出し、その開始・終了時刻を推定する。また同時に、各人物がどのイベントに参加しているのかを推定し、生活パターンの規則性を推定する。生活パターンを把握することができれば、行動を予測することでより状況に適した支援が可能になったり、在室状況の推定による人物識別の性能向上など、様々な応用が期待できる。

富永らは自宅に設置したセンサから得た長日数の在宅・不在データを用いて、クラスタリングにより典型的な外出パターンを抽出し、1日の行動を予測する研究を行った [1]。毎日の行動は複数存在するパターンのうちいずれかに属するが、それらが具体的に何個必要なのかを事前に知ることは難しい。そこで、ノンパラメトリックベイズの手法を用いることで、具体的な数を指定することなく適切なパターン数を推定できることが示された。しかし、このモデルでは1日分の観測全体を1つの典型パターンにより表現しているため、1日に含まれる複数のイベントに起因する要素的外出パターンを個別に表現することができない。そのため、要素的パターンの個数に対して、全日的パターンの個数が指数的に必要となってしまう可能性がある。これは、個人の生活パターンをモデル化する際にはあまり問題とはならないが、集団生活環境において複数人を考慮する場合には、個々人の属する要素的パターンの組み合わせが異なることがあるため問題となる。そこで本研究では、複数人物間

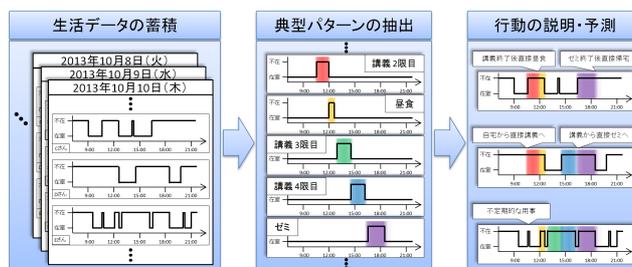


図 1: 提案手法のイメージ図。左が入力例、中と右が出力例。

で共有される要素的パターンの抽出と、その組み合わせの割り当てにより個々人の生活パターンを表現するための、ノンパラメトリックベイズの手法を提案する。提案手法のイメージ図を図 1 に示す。

2. データセット

本研究では、我々の所属する大学の研究室を集団生活環境とみなし、生活パターン分析の対象はその所属メンバーである。ある人物の1日における在室・不在を二値により表現する時系列データを1サンプルとし、これを研究室内の複数人について長期間収集したものをデータセットとして用いる。研究室の出入口とメンバー個人のデスクとを結ぶ通路を据え付けカメラにより観察しており、人物が通過した時刻とその入退室方向を、画像とともに自動的に記録する。さらに通過した人物が誰であるかを画像の目視により判別してラベル付けを行うことで、個人毎のタイムスタンプ付き入退室記録を作成する。この入退室記録にもとづき、入室してから次に退室するまでの期間を在室、退出から入室までの期間を不在とする。最後に、1日の時間を等間隔に分割し、各区間において在室ならば0、不在ならば1として前述の二値時系列データ形式へと変換した。このようにして作成したデータセットを $X = \{x_1, \dots, x_N\}$ と表す。ここで、 N はサンプルの総数である。また、各サンプルを $x_i \in \{0, 1\}^T$ と表す。 T は分割数であり、例えばサンプル

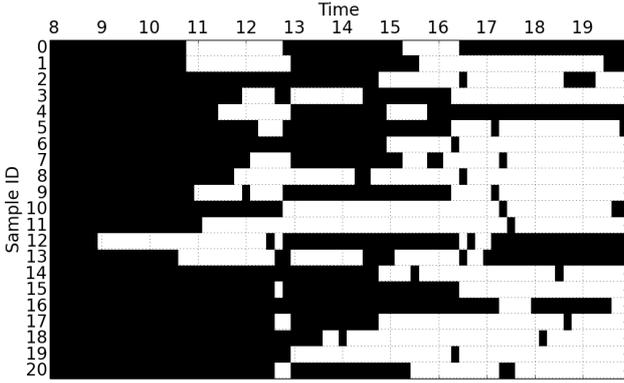


図 2: 在室・不在データの例. 横軸は 8 時から 20 時までの時間, 縦軸はサンプル番号を表す. 白・黒はそれぞれ在室・不在を表す.

i において時刻 t に不在であれば $x_{it} = 1$ となる. 図 2 はデータセットの一部を抜粋したものである. 白色が在室, 黒色が不在を表しており, 散在する短い黒帯からスケジュール上にないちよとした外出の存在を読み取ることができる.

3. 生成モデル

本研究では, 人々は日常生活において, 自らが割り当てられたイベントに起因するスケジュールに従って行動していると仮定する. つまり大局的には, イベントの無い時間帯は在室しており, イベントに参加している間は不在となる. ただし, 局所的にはイベントによる外出時刻はいつも同じとは限らず, またイベントが無いときに外出することもある. そのため, 観測 x_i は確率変数としてモデル化し, またその大局的スケジュールを確率的潜在変数 $\mu_i \in (0, 1)^T$ により表現する. ここで μ_i はサンプル i の各時刻における不在確率を表すパラメータであり, 観測 x_{it} がパラメータ μ_{it} をもつベルヌーイ分布に従って生成されるとする.

スケジュールのモデル化にあたり, 個人が 1 日に参加するイベントの数は高々少数であると仮定する. すなわち, 各サンプルがいずれのイベントに参加するかは, スパースな二値ベクトルにより表現することができる. さらに, 複数人物が同一のイベントに参加する可能性があるものとする. 本研究では, 人々の行動スケジュールにおける以上のような性質を表現するために, スパース符号化的アプローチをとる.

3.1 Noisy-or 観測モデルとベータ-ベルヌーイ過程

μ_i の確率的生成過程を考える. そのために, イベントの典型パターンと, 各サンプルにおけるイベントの割り当てを表す潜在変数を導入する. K 個の潜在的イベントが存在するとして, $d_k \in \{0, 1\}^T (k = 1, \dots, K)$ により表す. d_k は, イベント k に参加した場合に不在となる時間帯を表現しており, 1 日のスケジュールは複数の d_k を組み合わせることによって決まる. さらに, サンプル i へのイベント割り当てを $z_i \in \{0, 1\}^K$ により表し, サンプル i がイベント k に参加するならば $z_{ik} = 1$ とする.

全てのイベント変数 $D = \{d_1, \dots, d_K\}$ と z_i からスケジュール μ_i を合成する必要がある. しかし, d_k, z_i が共に二値変数であり, また μ_i には区間 $(0, 1)$ に値をとるという制限があるため, 通常のスパース符号化で用いられる加法的モデルは採用することができない. そのため, 本研究ではここで noisy-or

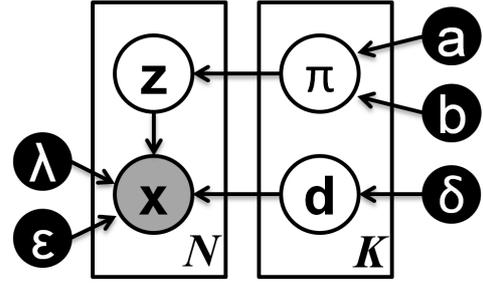


図 3: 提案手法のグラフィカルモデル. 白丸は潜在変数, 灰色は観測変数, 黒はハイパーパラメータ.

モデルを用いる [2]. すなわち,

$$\mu_{it} = 1 - (1 - \epsilon)(1 - \lambda)^{\sigma_{it}} \quad (1)$$

である. ここで, $\sigma_{it} = \sum_k z_{ik} d_{kt}$ である. $\lambda \in (0, 1)$ は割り当てられたイベント参加中の不在確率を調整するハイパーパラメータである. λ を大きな値に設定すると, σ_{it} が 1 以上のとき, すなわちその時刻に何らかのイベントに参加しているとき, μ_{it} は 1 に近い値となる. また, $\epsilon \in (0, 1)$ はいずれのイベントにも割り当てられていない時刻, すなわち σ_{it} が 0 のときにおける不在確率を調整するハイパーパラメータであり, 通常小さな値に設定する.

スパース符号化と辞書学習の方法としてベータ-ベルヌーイ過程を用いる Zhou らの手法 [3] は, あらかじめ辞書のサイズ, ここでは潜在イベントの個数を具体的に与える必要がないという利点をもつ. ベータ-ベルヌーイ過程に基底測度として多変量ベルヌーイ分布を与えることで, D と $Z = \{z_1, \dots, z_N\}$ の生成モデルとする. すなわち, d_{kt} には事前分布としてベルヌーイ分布を考え, Z_k はベータ-ベルヌーイ階層分布によりモデル化される.

以上をまとめると, モデル全体の生成過程は次のようになる.

$$d_{kt} \sim \text{Bernoulli}(\delta) \quad (2)$$

$$\pi_k \sim \text{Beta}(a/K, b(K-1)/K) \quad (3)$$

$$z_{ik} \sim \text{Bernoulli}(\pi_k) \quad (4)$$

$$\mu_{it} = 1 - (1 - \epsilon)(1 - \lambda)^{\sigma_{it}} \quad (5)$$

$$x_{it} \sim \text{Bernoulli}(\mu_{it}) \quad (6)$$

ここで, δ, a, b はハイパーパラメータである. 図 3 にグラフィカルモデルを示す.

3.2 モデルの学習

モデルパラメータの推定は, 観測データを所与とするパラメータ事後分布からのサンプリングにより行う. 以下の更新式を用いてギブスサンプリングを行うことでモデルパラメータが得られる.

$$\bar{\epsilon} = 1 - \epsilon, \quad \bar{\lambda} = 1 - \lambda \text{ として,}$$

$$p(d_{kt} | -) \propto \delta^{d_{kt}} (1 - \delta)^{1 - d_{kt}} \times \prod_{i=1}^N (1 - \bar{\epsilon} \bar{\lambda}^{\sigma_{it}})^{x_{it}} (\bar{\epsilon} \bar{\lambda}^{\sigma_{it}})^{1 - x_{it}} \quad (7)$$

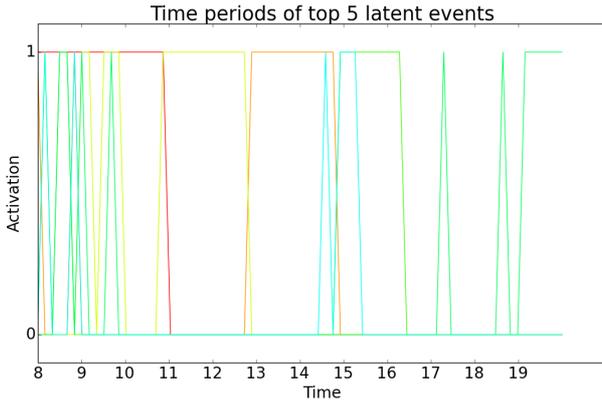


図 4: 上位 5 個の潜在イベント. 色は図 6 に対応.

$$\pi_k \sim \text{Beta} \left(\frac{a}{K} + \sum_{i=1}^N z_{ik}, \frac{b(K-1)}{K} + N - \sum_{i=1}^N z_{ik} \right) \quad (8)$$

$$p(z_{ik} | -) \propto \pi_k^{z_{ik}} (1 - \pi_k)^{1 - z_{ik}} \times \prod_{t=1}^T (1 - \bar{\epsilon} \lambda^{\sigma_{it}})^{x_{it}} (\bar{\epsilon} \lambda^{\sigma_{it}})^{1 - x_{it}} \quad (9)$$

4. 実験

提案手法の有効性を確認するために、実データを用いて実験を行った。2 節で述べた形式に従い、7 人の人物の平日 2 週間分の在室・不在時系列データからなるデータセットを構築した。ただし、1 日中不在であったサンプルは除外しており、サンプル数は延べ 59 日分である。さらに、本実験では、夜間から早朝の時間帯を除外し、8 時から 20 時の区間のみを対象とした。また分割数 T は、各区間が 10 分となるように 72 に設定した。

4.1 潜在イベントの抽出結果

潜在イベントの最大数 K を 10 に設定して実験を行ったところ、5 個の有意な潜在イベントが抽出され、残りの 5 個はアクティベーション頻度が低く実質的にほとんど使用されなかった。図 4 は抽出された 5 個の潜在イベントを示しており、学生の講義・TA による不在によく対応する結果が得られている。ここで、オレンジ色のイベントの終了時刻は時間割上の終了時刻よりも約 30 分後ろに伸びているが、これは TA 終了後に皆で昼食をとる習慣があったことに起因している。

4.2 観測データに対する潜在イベントの推論結果

観測データに対してどのイベントに参加しているのかを推論した結果の例を図 5 に示す。観測データのうち、おおむね不在の時間帯にはそれに対応するイベントに参加していることが、正しく推論できている。また、観測データに含まれる多少のスパイク状ノイズへの頑健性も認められる。

4.3 潜在イベントのアクティブ率

図 6 に示すのは、各潜在イベントがデータセット全体のうちどれだけ生じたかという割合である。人々の活動の大部分は非常に少ない種類のイベントによって説明されていることがわかる。

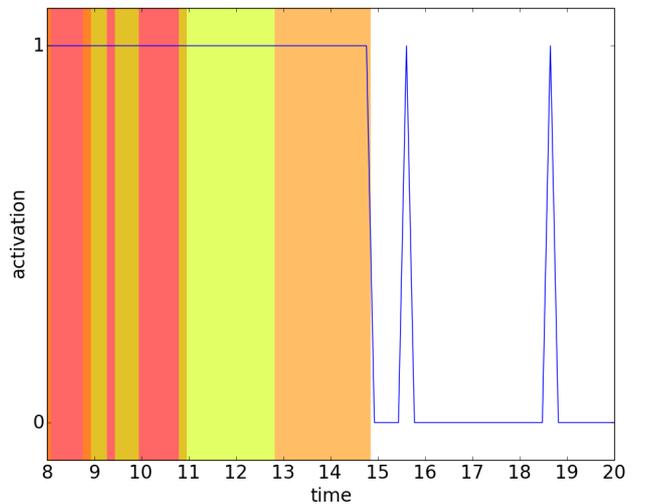
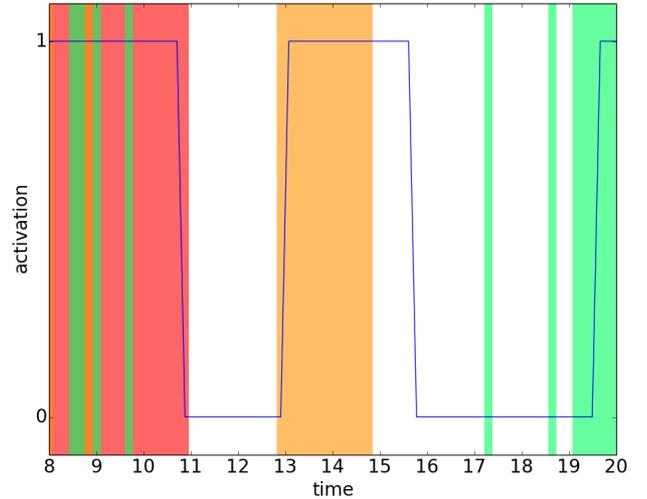


図 5: 観測データに対する潜在イベント割り当て推論の結果.

5. まとめ

本研究では、集団生活環境における典型的行動パターンの把握のために、潜在的イベントにもとづく在室・不在パターンの抽出を行った。そのために、noisy-or による観測モデルとベータ-ベルヌーイ過程を用いる教師なし学習モデルを提案した。実験により、容易に解釈できる潜在イベントが獲得され、またその個数も自動的に推定されることが確かめられた。より大規模なデータを用いた実験が今後の課題である。

参考文献

- [1] Shoji Tominaga, Masamichi Shimosaka, Rui Fukui, and Tomomasa Sato, "A Unified Framework for Modeling and Predicting Going-out Behavior", Pervasive 2012, Vol. 7319, LNCS, pp. 73-90. Springer, Heidelberg, 2012.
- [2] F. Wood, T. L. Griffiths, and Z. Ghahramani, "A Non-parametric Bayesian Method for Inferring Hidden

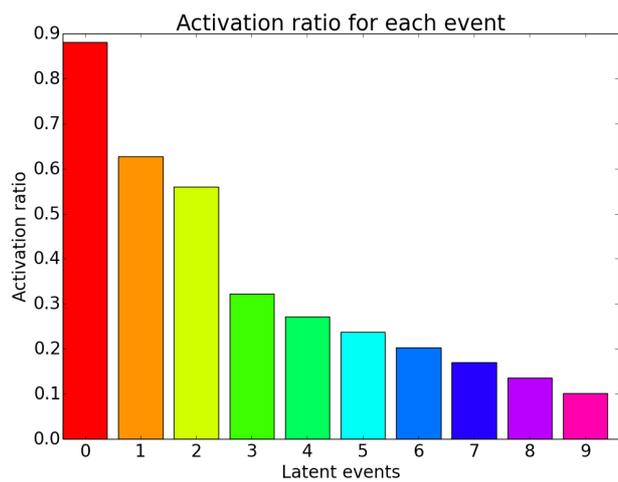


図 6: 全サンプルに対して各潜在イベントがアクティブであった比率

Causes”, Proceeding of the 22nd Conference on Uncertainty in Artificial Intelligence, 2006.

- [3] Mingyuan Zhou, Haojun Chen, John Paisley, Lu Ren, Lingbo Li, Zhengming Xing, David Dunson, Guillermo Sapiro and Lawrence Carin, ”Nonparametric Bayesian Dictionary Learning for Analysis of Noisy and Incomplete Images,” IEEE Trans. Image Processing, Vol. 21, pp. 130-144, Jan. 2012.