

HDP-HMMとLDAに基づく 保育園児の行動軌跡からの教師なし活動場面分類

Unsupervised Activity Classification from Behavioral Trajectory of Nursery School Children
Based on HDP-HMM and LDA

池田 佳那^{*1}
Kana Ikeda

張 斌^{*1}
Bin Zhang

中村 友昭^{*1}
Tomoaki Nakamura

長井 隆行^{*1}
Takayuki Nagai

大森 隆司^{*2}
Takashi Omori

岡夏樹^{*3}
Natsuki Oka

金子 正秀^{*1}
Masahide Kaneko

^{*1}電気通信大学

The University of Electro-Communications

^{*2}玉川大学

Tamagawa University

^{*3}京都工芸繊維大学

Kyoto Institute of Technology

In recent years, the demand for nursery school has increased with the increase of nuclear families and double-income households. However, there is a problem that the turnover rate of nursery teachers is high and the number of nursery teachers are not enough. Under these circumstances, the nursery teachers need to record child activities on the daily basis. We consider that the workloads on nursery teachers can be reduced by constructing a nursery teachers' support system that makes child activities record efficient. As a first step to construct such a system, in this paper, we propose a method to classify children behavioral trajectories, which are measured with Kinect sensors in a room of nursery school, into the activities in an unsupervised way. This classification consists of two phases. First, a continuous behavioral trajectory is discretized into local actions by Hierarchical Dirichlet Process - Hidden Markov Model, and then, these local actions are classified into activities by Latent Dirichlet Allocation.

1. はじめに

近年、核家族化・共働き世帯の増加に伴って保育園の需要が高まり、保育園での保育の質が重要視されている。一方で、保育士の離職率は高く、現状では保育士が不足している問題がある。その結果、保育士一人でも多くの子どもを見る必要がある。このような状況の中で、保育士は一日の活動を記録する保育記録を付ける必要がある。子どもたちの活動を振り返り、思い出す作業を考えると、負担の大きい業務の1つであると言える。特に、新人の保育士は慣れていないため、作業に時間がかかる可能性が高い。保育記録業務を効率化する保育士支援システムを構築することができれば、保育士の負担を軽減できると考えられる。

そのような保育支援システムとして、園児の行動軌跡から活動毎に自動分類された活動場面を保育士に提示することで、保育記録業務の効率化が可能なシステムを考えている。本稿では、そのようなシステムに必要な活動の自動分類手法を提案する。提案手法では、保育園の室内に設置した Kinect センサにより園児の行動軌跡を計測し、その結果を Hierarchical Dirichlet Process - Hidden Markov Model (HDP-HMM) [Beal 01] により個人の行動毎に分類し、さらにその行動の集合を Latent Dirichlet Allocation (LDA) [Blei 03] を用いて活動場面毎に分類する。実験において活動場面の分類性能を評価し、提案手法の有効性を検証する。

人の教師なし活動分類に関する関連研究として、Sun らが提案した手法がある [Sun 14]。しかし、この手法は個人の活動の分類であり、集団で活動している場面の分類は考えられていない。一方、本稿の提案手法では、個人の行動毎に分類した後、その行動の集合をさらに分類することで、活動毎に分類することが可能である。

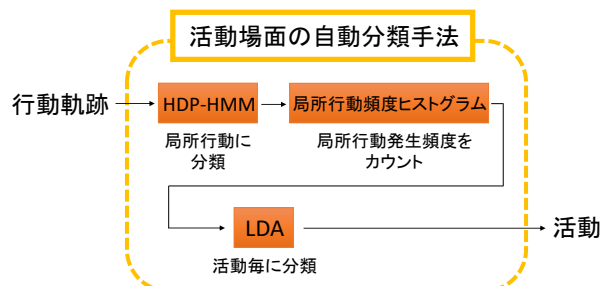


図 1: 活動場面の自動分類手法の概要



図 2: (a) 本実験で用いた場面の一例 (b) Kinect によって計測した1人の行動軌跡

2. 活動場面の自動分類手法

活動場面の自動分類手法の概要を図1に示す。本稿では、各個人がある瞬間に行っている動作を局所行動と定義し、活動は複数人の局所行動の集合によって構成されると考える。

2.1 行動軌跡データ

まず初めに子どもの行動軌跡を計測する。行動軌跡の情報としては位置および速度を用いる。本稿では、文献 [Zhang 16] で提案された手法を用いて、保育園の室内にマイクロソフト社

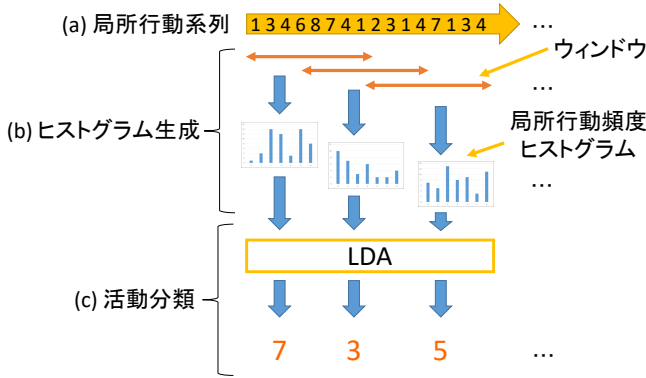


図 3: 局所行動系列を用いて活動分類を行うまでの流れ

製の Kinect v2 センサを設置し、園児の位置を計測する。なお、位置としては、地面に対し平行な平面上の 2 次元の座標を用いる。また、速度としては、各時刻における位置の差分を用いる。ある子どもの実際の行動軌跡の例を図 2 に示す。図 2(a) は実験で用いたシーンの一部である。このように、室内全体が映るように Kinect を設置し計測を行う。図 2(b) はある子どもの行動軌跡であり、連続的な時系列データとなっている。

2.2 Hierarchical Dirichlet Process - Hidden Markov Model による局所行動への分類

行動軌跡は、位置および速度の連続的な時系列データであり、単純には活動へと分類することはできない。そこで、まず各個人の行動軌跡から HDP-HMM を用いて分節化することで局所的な行動へ離散化する。HDP-HMM は、Hidden Markov Model に Hierarchical Dirichlet Process を導入することで、状態数の推定を可能としたモデルであり、局所行動の数が自動的に推定される。HDP-HMM のグラフィカルモデルを図 4 に示す。図 4 では、 t は時刻、 $(s_1, \dots, s_t)$ は隠れ状態、 $(x_1, \dots, x_t)$ は観測データを表す。状態 $(s_1, \dots, s_t)$ は無限個の状態 $k (= 0, \dots, \infty)$ をとることが可能である。 π_k は状態 k からそれぞれの状態へ遷移する確率を表している。また、 π_k は β , λ をパラメータとした Dirichlet Process (DP) により生成される。パラメータ β は、 γ をハイパーパラメータとした GEM 分布により生成される。 θ_k は観測データ $(x_1, \dots, x_t)$ を生成するガウス分布のパラメータであり、パラメータ α はその事前分布のパラメータである。時刻 t における状態 s_t は、直前の状態 s_{t-1} および遷移確率 π_k により決定される。また、観測データ x_t は θ_{s_t} をパラメータするガウス分布から生成される。

$$\pi_k \sim \text{DP}(\beta, \lambda) \quad (1)$$

$$\beta \sim \text{GEM}(\gamma) \quad (2)$$

$$\theta_k \sim P(\theta_k | \alpha) \quad (3)$$

$$s_t \sim P(s_t | s_{t-1}, \pi_k) \quad (4)$$

$$x_t \sim P(x_t | \theta_{s_t}) \quad (5)$$

本稿では、観測データ $(x_1, \dots, x_t)$ を園児の行動軌跡である位置および速度とし、 θ_k はその平均・分散とした。図 3 に局所行動系列を用いて活動分類するまでの流れを示す。HDP-HMM

により、各子どもの連続的な行動軌跡の系列が、図 3(a) のようにフレーム毎に局所行動を表す離散値へと変換される。

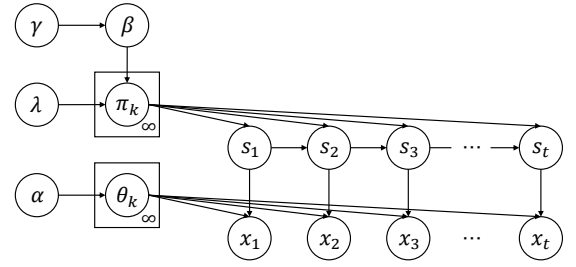


図 4: HDP-HMM のグラフィカルモデル

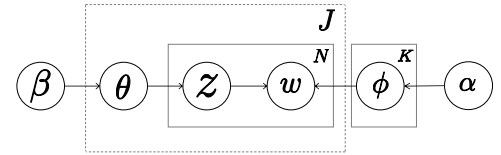


図 5: LDA のグラフィカルモデル

2.2.1 学習

パラメータの推定には、ギブスサンプリングを使用する。時刻 t の状態 s_t を除いた以下の条件付き確率により s_t をサンプリングする。

$$P(s_t | s_{-t}, \beta, \mathbf{X}, \lambda, \alpha) \propto P(s_t | s_{-t}, \beta, \lambda) P(x_t | s_t, s_{-t}, \mathbf{X}_{-t}, \alpha) \quad (6)$$

ただし、 $\mathbf{X}$ は全観測データの集合を表す。添字の $-t$ は時刻 t の情報を除いた残りを表しており、 s_{-t} は s_t を除いた全ての時刻の隠れ状態、 $\mathbf{X}_{-t}$ は $\mathbf{X}$ から x_t を除いた残りの観測データを示している。この式において、 $P(x_t | s_t, s_{-t}, \mathbf{X}_{-t}, \alpha)$ は、ベイズ推定より以下ようになる。

$$P(x_t | s_t, s_{-t}, \mathbf{X}_{-t}, \alpha) = \int P(x_t | s_t, \theta_{s_t}) P(\theta_{s_t} | s_{-t}, \mathbf{X}_{-t}, \alpha) d\theta_{s_t} \quad (7)$$

また、状態遷移確率である $P(s_t | s_{-t}, \beta, \lambda)$ は、 n_{ij} を状態 i から j へ遷移した回数とし、次式となる。

$$P(s_t = k | s_{-t}, \beta, \lambda) \propto (n_{s_{t-1}, k} + \lambda \beta_k) \frac{n_{k, s_{t+1}} + \lambda \beta_{s_{t+1}}}{n_k + \lambda} \quad \text{if } k \leq K, k \neq s_{t-1} \quad (8)$$

$$(n_{s_{t-1}, k} + \lambda \beta_k) \frac{n_{k, s_{t+1}} + 1 + \lambda \beta_{s_{t+1}}}{n_k + 1 + \lambda} \quad \text{if } k = s_{t-1} = s_{t+1} \quad (9)$$

$$(n_{s_{t-1}, k} + \lambda \beta_k) \frac{n_{k, s_{t+1}} + \lambda \beta_{s_{t+1}}}{n_k + 1 + \lambda} \quad \text{if } k = s_{t-1} \neq s_{t+1} \quad (10)$$

$$\lambda \beta_k \beta_{s_{t+1}} \quad \text{if } k = K + 1 \quad (11)$$

2.3 活動特微量

2.2節の処理によって計算した局所行動の集合が、その場で行われている活動の特徴を表していると考えられる。そこで、ある時間内の全ての子どもたちの局所行動の発生回数をカウントした局所行動の頻度ヒストグラムを、その時間内の活動の特徴量とする。このヒストグラムは、図3(b)のように、ある時間幅のウィンドウをオーバーラップさせながら、ウィンドウ内の局所行動の頻度をカウントすることで計算する。また、オーバーラップをさせることにより、連続した2つのウィンドウの局所行動頻度ヒストグラム同士が急激な変化をしない特徴量を作ることができる。

2.4 LDA による活動分類

活動特徴量である局所行動頻度ヒストグラムを分類することで、各ウィンドウを活動毎に分類することができる。ここでの分類には LDA を用いる。LDA は自然言語処理の分野で、文書分類のために提案されたモデルである。LDA では、文書を単語の集合と考え、単語の発生頻度ヒストグラムを文書の特徴量として用いて、単語の発生頻度の類似性から文書をトピック毎に分類することができる。

図5に LDA のグラフィカルモデルを示す。このグラフィカルモデルは、 J 個の文書の中に N 個の単語が存在するモデルとなっている。 w が単語、 z が文書のトピック、 θ がトピック z の出現確率分布を表すパラメータである。また、 θ はハイパーパラメータを β とするディリクレ分布に従って文書の数と同じ J 個生成される。 ϕ は単語 w の生起確率を表す多項分布のパラメータである。

$$\theta \sim \text{Dir}(\beta) \quad (12)$$

$$\phi \sim \text{Dir}(\alpha) \quad (13)$$

$$z_n \sim \text{Multinomial}(\theta) \quad (14)$$

$$w \sim p(w|z_n, \phi) \quad (15)$$

本稿では、単語 w を局所行動と捉えることで、局所行動頻度ヒストグラムの類似性を基に活動へと分類することが可能となる。

2.4.1 学習

パラメータの推定には、HDP-HMM 同様ギブスサンプリングを使用する [Griffiths 04]。文書 j の $i (= 1, \dots, N)$ 番目の単語に割り当てられたトピック z_{ji} を除いた以下の条件付き確率により z_{ji} をサンプリングする。

$$P(z_{ji} | \mathbf{z}_{-ji}, \beta, \mathbf{W}, \alpha) \propto P(z_{ji} | \mathbf{z}_{-ji}, \beta) P(w_{ji} | z_{ji}, \mathbf{z}_{-ji}, \mathbf{W}_{-ji}, \alpha) \quad (16)$$

ただし、 $\mathbf{W}$ は全文書の単語データの集合を表す。添字の $-ji$ は文書 j の i 番目の情報を除いた残りを表しており、 $\mathbf{z}_{-ji}$ は z_{ji} を除いた全てのトピック、 $\mathbf{W}_{-ji}$ は $\mathbf{W}$ から文書 j の i 番目の単語 w_{ji} を除いた残りの単語データを示している。この式において、 $P(z_{ji} | \mathbf{z}_{-ji}, \beta)$ 、 $P(w_{ji} | z_{ji}, \mathbf{z}_{-ji}, \mathbf{W}_{-ji}, \alpha)$ はそれぞれ、ベイズ推定より以下ようになる。

$$P(z_{ji} | \mathbf{z}_{-ji}, \beta) = \int P(z_{ji} | \theta) P(\mathbf{z}_{-ji} | \theta) P(\theta | \beta) d\theta \quad (17)$$

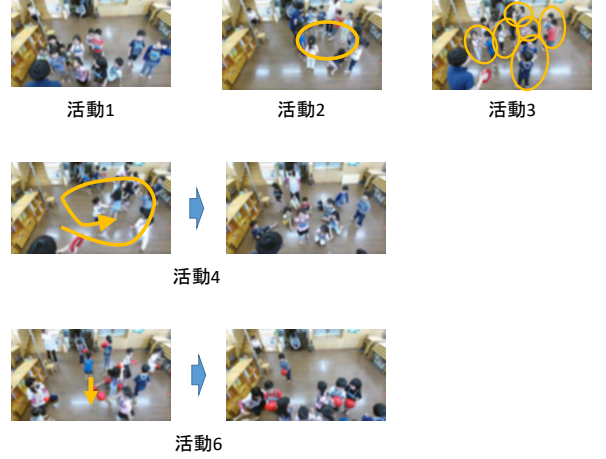


図 6: 各活動に分類されたシーンの一例

$$P(w_{ji} | z_{ji}, \mathbf{z}_{-ji}, \mathbf{W}_{-ji}, \alpha) = \int P(w_{ji} | \phi) P(\mathbf{W}_{-ji} | \phi) P(\phi | \alpha) d\phi \quad (18)$$

$P(z_{ji} | \theta)$ および $P(w_{ji} | \phi)$ は多項分布であり、 $P(\phi | \alpha)$ および $P(\theta | \beta)$ はディリクレ分布であるため、 $P(z_{ji} | \mathbf{z}_{-ji}, \beta)$ および $P(w_{ji} | z_{ji}, \mathbf{z}_{-ji}, \mathbf{W}_{-ji}, \alpha)$ は以下ようになる。

$$P(z_{ji} | \mathbf{z}_{-ji}, \beta) = \frac{n_{z_{ji}} + \beta}{n_j + Z\beta} \quad (19)$$

$$P(w_{ji} | z_{ji}, \mathbf{z}_{-ji}, \mathbf{W}_{-ji}, \alpha) = \frac{n_{w_{ji}} + \alpha}{n_{z_{ji}} + W\alpha} \quad (20)$$

式 (19) $n_{z_{ji}}$ は文書 j の単語でトピック z_{ji} に分類された単語数、 n_j は文書 j 中の全単語数、式 (20) の $n_{w_{ji}}$ は全文書で単語 w がトピック z_{ji} に分類された個数、 $n_{z_{ji}}$ はトピック z_{ji} に分類された全単語数を表している。また、 Z および W は、それぞれトピック数および単語の種類数を表している。

3. 活動場面分類実験

提案手法の有効性を検証するために実験を行った。行動軌跡データは、玉川さくら保育園の園児 13 名が先生と共にリトミック^{*1}をしている 16 分 40 秒の場面 (10fps, 合計 10,000 フレーム) を 5fps にダウンサンプリングして使用した。

本実験では局所行動頻度ヒストグラムを作成する際のウィンドウを 100 フレーム、オーバーラップを 60 フレームとした。その結果、局所行動 0 番の出現頻度が 3~4 桁なのに対し、他の局所行動番号の出現頻度が 1~2 桁となった。そのため、局所行動 0 番による分類結果への影響を抑えるために全ウィンドウ内の 0 番の出現頻度を 1/10 にした後に小数点以下を切り捨てた頻度を新たな出現頻度とした。

このようにして計算した局所行動頻度ヒストグラムを、LDA により 8 個の活動へと分類した。その結果、各活動には以下のようなシーンに分類された。

*1 リトミックとは、音楽を取り入れた身体的・感覚的・知的に優れた子どもを育てる教育法を指す

活動 1 主に画面の下側で活動しているシーン
 活動 2 輪になって活動しているシーン
 活動 3 2人組や輪になって活動しているシーン
 活動 4 ぐるぐる回る状態とその場に留まっている状態が交互に含まれて活動しているシーン
 活動 5 動き回っている状態とその場に留まっている状態が交互に含まれて活動しているシーン
 活動 6 画面全体で活動している場面と画面の端の方で活動している場面が含まれているシーン
 活動 7 その場に留まって活動しているシーン
 活動 8 ほぼ全員がその場に留まって活動しているが、数人は動いているシーン

各活動に分類されたシーンの一例を図 6 に示す。分類された結果を全体的に見ると、位置（画面のどの部分で活動しているか）による分類、速度（動きがあるか無いか）による分類、動き方の特徴（ぐるぐる回っている場面や動き回っている場面）を捉えた分類、子どもたちの相互の位置関係（2人組や輪になっている場面）による分類がされていることが分かる。このように、それぞれの活動の特徴を捉えて分類ができているといえる。

しかし、活動 2 には、各活動に分類される確率がほぼ一樣になるような、場面の切り替わりで活動の特徴が曖昧となるシーンが分類されてしまっている。これは活動分類数を増加させることで解決できる可能性があり、今後適切な活動分類数を自動的に推定することが可能なノンパラメトリックベイズへ拡張することを考えている。また、活動 5 には、ぐるぐる回る状態でもその場に留まっている状態でもないような特徴が表れているシーンが分類され、活動 6 には、終始画面全体で活動しており画面の端で活動をしている場面を含まないシーンが分類されてしまった。活動 5 および活動 6 の場面は、本稿の提案手法の特徴量では表現しきれない可能性がある。これは加速度などの別の特徴量を用いることでより精度の高い分類ができる可能性がある。

4. まとめ

本稿では、保育園児の行動軌跡から、教師なしで活動毎に分類する手法を提案した。実験の結果、位置、速度、動き方の特徴、子どもたちの相互の位置関係に基づき活動の特徴を捉えて分類をすることができた。今後、適切な活動分類数を自動的に推定することや、位置と速度以外の特徴も利用することで、より高精度な分類を実現することを考えている。そして、本実験で用いたリトミックの場面に活動のアノテーションを行い、アノテーションと共に学習させることにより、新たなシーンにおいてどのような活動が行われているかを推定できる手法へと拡張することを考えている。

謝辞

本研究は、文部科学省科学研究費補助金新学術領域研究「子ども＝大人インタラクションの認知科学的分析とモデル化」（課題番号：26118003）の助成を受けて行われた。実験を実施するに当たり多大なご協力をいただいたさくら保育園の関係者の皆様に心より感謝する。

参考文献

[Beal 01] Matthew J. Beal, Zoubin Ghahramani and Carl Edward Rasmussen : The infinite hidden Markov

model, Advances in Neural Information Processing Systems, pp. 577-584 (2001)

[Blei 03] David M. Blei, Andrew Y. Ng and Michael I. Jordan : Latent Dirichlet Allocation, Journal of Machine Learning Research, Vol. 3, pp. 993-1022 (2003)

[Sun 14] Feng-Tso Sun, Yi-Ting Yeh, Heng-Tze Cheng, Cynthia Kuo, Martin Griss : Nonparametric Discovery of Human Routines from Sensor Data, IEEE International Conference on Pervasive Computing and Communications (PerCom), pp. 11-19 (2014)

[Zhang 16] Bin Zhang, Tomoaki Nakamura, Rena Ushiogi, Takayuki Nagai, Kasumi Abe, Takashi Omori, Natsumi Oka and Masahide Kaneko: Simultaneous Children Recognition and Tracking for Childcare Assisting System by Using Kinect Sensors, Journal of Signal and Information Processing, Vol. 7, No.3, p.148 (2016)

[Griffiths 04] Thomas L. Griffiths and Mark Steyvers: Finding scientific topics, National Acad Sciences, Vol.101, pp. 5228-5235 (2004)