

機械学習を用いた子どもの遊び行動解析のための 多人数同時位置計測システムの構築

Development of a Multi-Person Simultaneous Position Measurement System for Analyzing Children's Play Behavior Using Machine Learning

金井 康二*, 張山 昌論*, 小柴 満美子*†, 小林 康浩**

Koji Kanai*, Masanori Hariyama*, Mamiko Koshiba*†, Yasuhiro Kobayashi **

*東北大学, †山口大学, **小山工業高等専門学校,

*Tohoku University, †Yamaguchi University,

**National Institute of Technology, Oyama College

キーワード : 個別最適 (specific optimization), 無線通信 (wireless communication),
MQTT (message queuing telemetry transport), ウォード法 (Ward's method),
決定木分析 (decision tree analysis)

連絡先 : 〒 980-8579 仙台市青葉区荒巻字青葉 6-3-09
東北大学大学院 情報科学研究科 張山・ウィッデヤスーリヤ研究室 金井康二
Tel.: (022)795-7155, E-mail: koji.kanai.t4@dc.tohoku.ac.jp

1. はじめに

子どもは発達段階や興味・関心の対象が一人一人異なる。近年、保育現場ではそれぞれの発達段階や興味・関心に応じた保育内容を提供する「個別最適化」の重要性が高まっている¹⁾。そのためには、子ども一人ひとりに寄り添い、個性を理解する必要がある。個性の推定の手法の一つとして、行動解析があり、現状では保育士や教員の観察や子どもとの会話を通じた記録が主な手段である。しかし、これらの方法では、すべての子どもに目が届かず、客観的な記録が困難であるとともに、すべての子どもの手動記録は負荷が高い。

そこで、本研究では、ICT 技術を活用し、子ど

もの行動を多人数同時に記録することで、観察可能な子どもの数を増やすとともに、客観的な記録及び手動記録の負荷低減を目指す。特に、本システムでは子どもの位置情報(“いつ”, “誰が”, “どこにいるか”)に着目し、その情報から、遊具の滞在時間・割合、使用頻度、交友関係などを把握できる。本稿では、そのような計測を可能にした、小型無線タグを用いた多くの子どもの同時位置計測システムを提案するとともに、先行研究で得られたデータを用いて、機械学習による子どもの遊び行動の解析結果を例示する。

2. 本システムの機器・及び構成

2.1 TWELITE モジュール

本研究では, 子どもの位置情報を計測する機器として, MONOWIRELESS 社製の TWELITE モジュールを用いた. TWELITE モジュールは, サイズ・消費電力がともに小さく, 電池で長時間稼働が可能な無線通信可能なマイコンモジュールである. 本研究では, TWELITE モジュールの中の小型無線タグ TWELITE CUE (Fig.3) 及び受信機 MONOSTICK (Fig.4) を用いている. また, 諸元を Table 1 に示す.

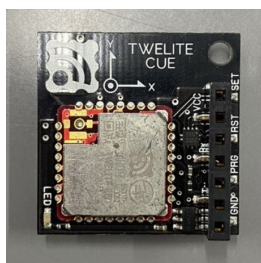


Fig. 1: 基盤部

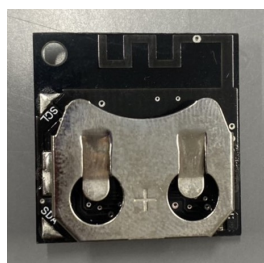


Fig. 2: 電池装着部

Fig. 3: TWELITE CUE



Fig. 4: MONOSTICK

まず, TWELITE CUE について説明する. TWELITE CUE は 1 円玉サイズのコイン型電池 CR2032 で長時間稼働する. また, 加速度セン

サー及び高性能な小型の基盤アンテナ (MW-A-P1934) を内蔵しており, 数十メートルから数百メートルの通信ができる.

次に, MONOSTICK は, 基盤に USB シリアルポート接続部が装着された構造をしており, コンピュータ等に接続し, 被接続機器の給電を得て動作する. また, パソコンやボード PC 上のアプリで TWELITE を操作したり, センサデータ情報を無線で収集したりできる. TWELITE CUE 及び MONOSTICK はともに基盤に搭載されているマイコンに動作プログラムを書き込むことで, 機能をカスタマイズすることができる.

2.2 システムの構成

本システムは, TWELITE CUE は規定された情報を電波で定期的に発信する子機として, MONOSTICK は複数の子機から受信したデータを USB シリアル通信を用いて PC に記録する親機としての役割を担う. なお, MONOSTICK は USB シリアル通信を可能にするコンピュータ等が必要であるが, 本研究では, Raspberry Pi 4 でシリアル通信を読み取る Python スクリプトを実行することで情報の記録を可能にした. 今後, 子どもに装着するために TWELITE CUE をケースに入れたものを「タグ」, 子機から受信した情報を記録するための Raspberry Pi 4 と MONOSTICK を接続しまとめたものを「基地局」と呼ぶ. 基地局は各遊具に設置し, タグを付けた子どもの情報を記録する. 最後に, 複数の基地局に記録されている子どものタグの情報について分析を行い, その結果をもとに現在

Table 1: 各機器の諸元

機器	たて/よこ/高さ [mm]	重さ [g]	実勢価格 [円/個]
TWELITE CUE 基板部	25/25/10(最厚部)	4	3000
TWELITE CUE ケース	30/30/15	6	
MONOSTICK	50/22/8(USB 端子部除く)	10	3000

最も近い基地局, すなわち現在遊んでいる遊具を特定する位置推定を行う (次節参照). そのようにして得られる位置推定データの例を Table 2 に示す. 以下の通り, 位置情報が各子ども別で記録された時系列データを得ることができる. 値は受信時に得られた電波強度である.

2.3 最近傍基地局の推定

TWELITE CUE により発信される電波には, 一般的な電波同様に, 距離に対して単調に減衰するという性質がある (Fig.5). また, Table2 のように, 得られるデータは, 各時刻, 各子供, 各基地局に得られる電波強度の値となっている. その電波の性質及び得られるデータより, 電波強度が最も高い基地局を最近傍基地局と推定する. なお, 図中の LQ 値は dB に比例した電波強度指標である.

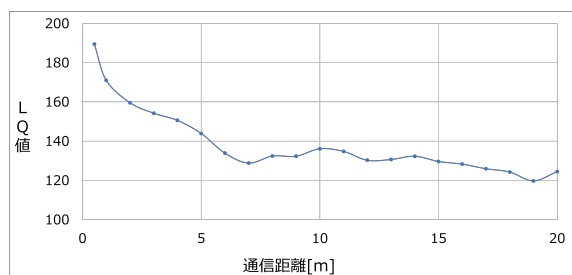


Fig. 5: 実測による距離と電波強度の関係

3. 本システムの通信

3.1 MQTT プロトコルを活用したリアルタイム通信

MQTT プロトコルは, HTTP に比べて軽量かつ柔軟な特徴を持つことから, IoT の分野では標準的な通信プロトコルとして扱われている²⁾. 送信側が送るデータをデータ領域に一度保持しながら, 受信側の処理が完了するのを待たずに次の処理へに移る方式である. また, 構成コンポーネントは「パブリッシャ」, 「サブスクライバ」, 「ブローカー」からなる. メッセージを確実に送信することを目的としているため, プロトコルヘッダが小さいことが特徴であり, ヘッダサイズは HTTP が 50 バイトからであるのに対し, MQTT は 2 バイトからである. さらに, 1 対多や多対多の通信ができることも特徴である.

接続・データ通信はパブリッシャがブローカーにメッセージを送り, サブスクライバはブローカーからメッセージを受信することを宣言することで確立される. この機能により, パブリッシャ・サブスクライバそれぞれがブローカーとのみ接続を行うため, 接続相手が頻繁に変わる場合でも接続先を意識せずに通信ができたり, 接続相手に自局の情報を送らずに通信で

Table 2: 位置推定データ例

時刻	子ども	基地局 1	基地局 2	基地局 3	位置推定結果
10:00:00	子ども A	87	66	160	基地局 3
	子ども B	155	95	210	基地局 3
	子ども C	148	57	210	基地局 3
10:00:01	子ども A	210	148	128	基地局 1
	子ども B	95	142	78	基地局 2
	子ども C	168	78	225	基地局 3
10:00:02	子ども A	39	135	89	基地局 2
	子ども B	79	178	65	基地局 2
	子ども C	93	165	42	基地局 2
⋮	⋮	⋮	⋮	⋮	⋮

きるなどのメリットがある。また、ブローカーはメッセージの送受信を管理し、パブリッシャから受信したメッセージを必要としているサブスクライバにメッセージを配信する。本システムにおけるメッセージのやり取りの図を Fig.6に示す。

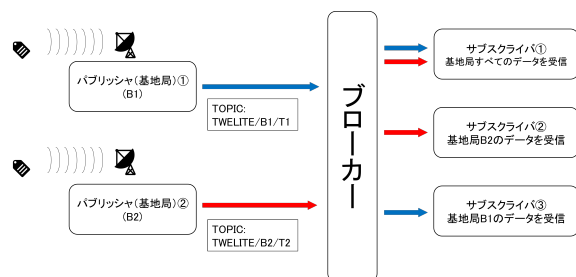


Fig. 6: 本システムにおける通信イメージ

3.2 本研究での通信構成

本研究では、各基地局をパブリッシャ、ホスト PC 1 台をブローカー兼サブスクライバとする。そして、基地局（各 Raspberry Pi 4）及びホスト PC を Wifi を用いた同一のネットワークにつなぐ。このようにすることで、タグから送信された情報を各基地局が受信し記録、その後、ホスト PC（ブローカー）に情報を転送し、ホスト PC（サブスクライバ）が必要な情報のみを取得することができる。

4. 機械学習による遊び行動の解析

本章では、前章までで得られるデータを基にした遊び行動の解析方法を提案するとともに、先行研究³⁾のデータを用いた解析結果を例示する。

4.1 遊び行動の解析の概要

本研究での解析の流れを説明する。まず、前章までで得られる「子どもの最近傍基地局の時系列データ (Table 2)」を「各子供の各基地局に対する滞在割合 (Table 3)」に変換する。

Table 3: 各子供の各基地局に対する滞在割合

子ども	基地局 1	基地局 2	基地局 3
子ども A	0.40	0.15	0.45
子ども B	0.30	0.35	0.35
子ども C	0.10	0.10	0.80
⋮	⋮	⋮	⋮

その滞在割合のデータを機械学習のクラスタリング手法を用いて複数のグループに分類する。ここで、各グループは何らかの類似度をもって分類される。次に、各子供に対するグループを教師ラベルとして、遊具の滞在割合とともに機械学習の決定木分析を行う。それにより、各グループがどの遊具の滞在割合が高い（低い）のか、すなわち、各グループに属する子どもの好みの遊具が可視化されるのである。

4.2 クラスタリング

クラスタリングとは、データ間の類似度に基づいて、データをグループ分けする手法である。また、クラスタリングによってできた類似度が高いものの集合をクラスと呼ぶ。クラスタリングには、大きく分けて2つの方法があり、階層的手法と非階層的手法に分けられる。階層的手法とは、初期状態を一個のサンプルに一個のクラスが割り振られており、距離が近いものから段階的に結合する手法である⁴⁾。分割クラスタ数を決めなくて良いことが利点として挙げられる。また、デンドログラムを作成することでクラスタの階層が視覚的にわかりやすい。非階層的手法とは、分割クラスタ数が固定で、最適となる分割を探索する手法である。

本研究、及び先行研究では分割数が事前にわからないこと、また、クラスタの特徴を得るためにクラスタの階層を視覚的に理解する必要があることから、階層的手法を選択した。また、その中でも最も外れ値に強いことから ward 法を使用している。

ward 法は,2つのクラスターを統合する際に、クラスター内の偏差平方和の増分が最小となるようにクラスターを形成する手法である。クラスター C_i と C_j が融合されたときの偏差平方和の増分は

$$\Delta E_{ij} = \frac{n_i n_j}{n_i + n_j} \sum_{l=1}^p (\bar{n}_l^{(i)} - \bar{n}_l^{(j)})^2 \quad (1)$$

となり,これをクラスター C_i と C_j との非類似度とみなすとき, $C_i \cup C_j$ と他のクラスター C_k を融合したときの増分は,

$$\Delta E_{(ij)k} = \frac{1}{n_i + n_j + n_k} \cdot \{(n_i + n_k)E_{ik}^2 + (n_j + n_k)E_{jk}^2 - n_k E_{ij}^2\} \quad (2)$$

となる。これを最小にするクラスターを融合する⁵⁾。

4.3 決定木分析

決定木は,統計学,データマイニング,機械学習で使用される予測モデリングアプローチの一種であり,いくつかの判断経路とその結果とを木構造を使って表現したものである⁶⁾。決定木は一般的に分類木と,回帰木に分かれる。決定木の木構造は,葉はクラスラベルを表し,枝はその葉へ接続の過程を示す。本研究・先行研究では,CART アルゴリズムを採用している。

CART アルゴリズムは,分類および回帰のための決定木モデルを構築する手法である⁷⁾。CART はツリー構造を使用し,各ノードでデータを分割する際に,「不純度」が最も低くなる特徴量,閾値を選択することを繰り返すことで最終的な予測を行う。その不純度の基準として,Gini 係数を用いた。Gini 係数は, L を全ラベルの集合, $l \in L$ としたとき,全学習データに対する,ラベル l の要素を持つ割合を p_l として以下のように表せる。

$$Gini = 1 - \sum_{l \in L} p_l^2 \quad (3)$$

4.4 先行研究の概要とデータ

先行研究と本研究の大きく違うところは,計測方法である。本研究は,2章,3章での同時位置計測システムを用いて計測するのに対し,先行研究ではビデオカメラを設置し,その映像データから子どもの最近傍基地局の時系列データを手動で記録し,滞在割合のデータに変換している。先行研究の複数の実験のうち,本稿では,2019/5/11 に行われた神原公園での実験を用いて解析手法を例示していく。

この実験では,神原公園に遊び場を用意し,公園の遊びエリアには子どもが自由に出入りして遊ぶことができる。また,用意した遊びには,運動能力を必要とする「一本ロープ」「二本ロープ」「やぐら」,創造性を必要とする「ダンボール遊び」「ものづくり」「泥遊び」,料理などに関する遊びである「七輪」,そのほか公園に用意されている「既存遊具」がある。解析方法は4.1節で述べたように,まず映像記録から10秒ごとに各子供の最近傍基地局を手動で記録し,各子供の基地局に対する滞在割合のデータに変換した。次に,ward 法を用いたクラスタリングを行い,複数のクラスターに分割した。そのクラスタリング結果を教師ラベルとして,遊具の滞在割合のデータとともに決定木分析を行い,クラスターの特徴抽出を行った。

4.5 解析結果

神原での実験の参加人数,計測時間を Table 4に示す。

Table 4: 神原公園での実験の概要

	参加人数	計測時間 [h]
2019/5/11 神原	11	3

また,ward 法を用いたクラスタリングのデンドログラム及び決定木分析の結果を Fig.7, Fig.8に示す.

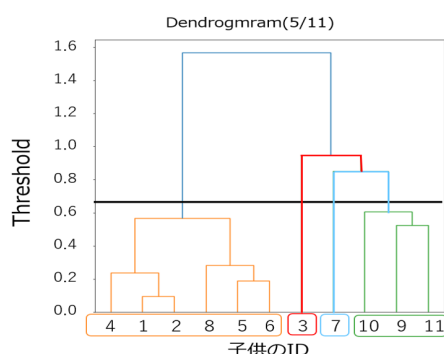


Fig. 7: クラスタリング結果とデンドログラム

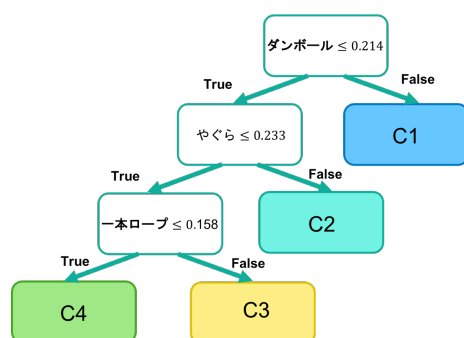


Fig. 8: 決定木分析の結果

Fig. 7より, クラスタリングによるクラスタの人数は6:1:1:3となった. また, 閾値はクラスタが4つに分けられるように設定している. Fig. 8より, 今回, 重要度が高い遊びは「ダンボール」, 「やぐら」, 「一本ロープ」であることがわかった. C1はダンボールの割合が高い子どものクラスタであるため, 創造性を好む子どもたちの可能性が高い. C2はダンボールの割合が低くやぐらの割合が高いので, 運動を好むクラスタである可能性が高い. C3においては, 「ダンボール」「やぐら」の割合が低く, 「一本ロープ」の割合が高いクラスタであるため, このクラスタも運動を好むクラスタの可能性が高い. 最後にC4はダンボール, やぐら, 一本ロープの割合がすべて低いクラスタである. この条件ではどのような遊びを好むクラスタかの判断は難しい. このように子どもの遊具の滞在割合のデータ

に対してクラスタリング手法を用いることで, 子どもを各クラスタに分類し, その結果に対して決定木分析を行う事で, クラスタリングの過程を可視化できることがわかった. 先行研究のデータは, 実験対象の人数が少なく, 時間も短いため, 分類精度及び分類の可視化の正確性が低いと考えられる. また, 映像データを基に記録するのも工数が多い. そこで, 本研究の同時位置計測システムを用いることで, 長時間, 多くの子どもの計測を低い工数で行う事ができる. そのようにすることで, 多くのデータを自動的に集めることができ, 解析の精度向上へつながり, 子どもの個性把握へ近づくと考えている.

5. おわりに

本稿では, 子どもの「個別最適化」のために, 子どもの個性の定量的な理解・把握を目的とした多人数同時位置計測システムの説明および先行研究を用いた遊び行動の解析の例示を行った. 第2章および第3章では, 本システムの構成および計測方法, 受信方法を説明した. 具体的には, TWELITE モジュールの「TWELITE CUE」(タグ) を子どもの一部に装着し, 「Raspberry Pi 4 に接続した MONOSTICK」(基地局) がタグからの情報を受信および記録する. その受信情報を, 同一のネットワークにつないだホスト PC に, MQTT プロトコルを用いて転送することでリアルタイム通信を可能にした. 第4章では先行研究のデータを用いて遊び行動の解析の例を示し, 解析方法および解析結果・考察を示した.

先行研究では, ビデオデータによる手動記録であったため, 時間および人数の増加とともに, 子どもの遊具に対する滞在データの取得に多くの工数を必要とした. そのため, 本研究のシステムを用いることで, 工数の低減だけでなく, 取得データの増加により遊び行動の解析精度

の向上が見込める。

本研究の展望としては、実際に幼稚園・保育園に導入し、正確な計測ができるかを調べる必要がある。また、実際に設置および運用したときに生じる問題を検討しなければならない。例えば、子どもの人数が増えたときの通信の輻輳制御や、継続的な Raspberry Pi 4 の給電方法、子どもへの負荷のかからないタグの装着方法などが挙げられる。また、タグによる加速度情報の抽出を行い、子どもの運動量の計測なども行うことで、位置情報からはわからない解析もできるだろう。

参考文献

- 1) 中央教育審議会. ”「令和の日本型学校教育」の構築を目指して”. 文部科学省. https://www.mext.go.jp/content/210330-mxt_kyoiku01-000013731_09.pdf, (参照 2024-07-05).
- 2) 山崎祥司. MQTT リファレンスブック. Interface, CQ 出版社, 2022 年 11 月号.
- 3) 張山昌論, ”子どもの「遊び」を見える化するための AI・データ駆動型アプローチ”. 第 64 回日本小児神経学会学術集会, セッション：こころの発達を支援する IT 療育システム (2022-6)
- 4) 有賀友紀, 大橋俊介. R と Python で学ぶ 実践的データサイエンス & 機械学習. 技術評論社, 2022.
- 5) 佐藤義治. 多変量データの分類-判別分析・クラスター分析-. 朝倉書店, 2009, p.114-115.
- 6) Joel Grus. ゼロから始めるデータサイエンス. オーム社, 2017.
- 7) 柴原一友, 築地毅, 古宮嘉那子, 宮武孝尚, 小谷善行. 機会学習教本. 森北出版, 2019, p.140-152.