

# 確率モデルを用いた大規模イベントの来場者回遊行動分析

古田真理<sup>1</sup> 山下和也<sup>1</sup> 本村陽一<sup>1</sup>

**概要:** 近年実社会ビッグデータ収集やそのデータ分析の重要性が高まっている。本研究では実社会ビッグデータの例として、サイエンスアゴラとして開催されている大規模科学イベントでの心理・行動データに焦点を当て、実際に数千人規模のデータを収集する実験を行った。また、収集したデータから PLSA やベイジアンネットワークを使用して確率モデルを構築した結果の分析を行い、そこからイベント来場者・出展者の心理的価値を高めるイベントのサービス改善に応用する方法を考える。

**キーワード:** PLSA, ベイジアンネットワーク, 実社会ビッグデータ

## Analysis of Visitors' Migration Behavior of Large-scale Event Using Probabilistic Model

MARI FURUTA<sup>†1</sup> KAZUYA YAMASHITA<sup>†1</sup>  
YOICHI MOTOMURA<sup>†1</sup>

**Abstract:** In recent years, the importance of real-world big data collection and its data analysis has been increasing. In this research, as an example of real-world big data, we focused on psychological and behavioral data at a large-scale scientific event held as Science Agora. An experiment was conducted to collect data for thousands of people. In addition, we analyzed the results of constructing a probabilistic model from the collected data using PLSA and Bayesian networks, and from that, we analyzed the psychology of event visitors and exhibitors. We thought about how to apply it to improving the service of events that increase the value.

**Keywords:** PLSA, Bayesian Network, Real-World Big Data

### 1. はじめに

サービス現場におけるデータの収集、例えば購買行動に関する人の ID と商品 ID、場所(店舗)や時刻などのデータ(ID-POS データ)収集などはサービスの改善への応用事例も多い。ここで言うサービスとは、価値(利用者が認識でき、提供者が自覚しているベネフィット)の提供・利用が行われることを指す。また、本稿ではある特定の空間・時間を共有して価値の提供・利用を行う一般化したシステムをサービスシステムと言う。サービスシステムの 1 つであるフェスなどのイベントは現在日本でも多数開催されているが、実際にデータを収集し、サービスの改善に役立っている例はまだ少ない。本稿では、サイエンスアゴラ 2019 にて、先行研究[1], [2], [3], [4], [5], [6]で開発した、イベントにおいて来場者のブース回遊行動や各ブースを訪問時の感情をデータとして取得するシステムである AI タッチラリーを利用し、実際に来場者についてのデータを収集し、その解析を行った結果を報告する。サイエンスアゴラ 2019 とは、2019 年 11 月 16, 17 日に行われた、科学に関する展示や体験のできる大規模イベントで、およそ 2816 人を動員した。また、得られたデータの解析だけでなく、サイエン

スアゴラ 2018 との比較を行いながら、イベントにおけるサービス向上やこのイベント支援システムのさらなる活用方法についても検討する。

### 2. 大規模イベントにおけるデータ収集

#### 2.1 データ収集システム

AI タッチラリーとは来場者行動データ収集システムの 1 つで先行研究にて様々な実験が行われてきた。受付にて来場者に RF-ID 付きのカードを配布し、来場者は受付、周遊したブース、振り返りにて ID 認証端末にカードをタッチすることで来場者の行動履歴を取るとともに、各地点でアンケートを取ることで来場者の属性についてのデータも収集した。今回はこのシステムを利用して来場者のデータを収集した。なお、過去の実験では RF-ID 付きのカードを用いたが、今回はその代わりに RF-ID 付きのリストバンドを用いた。

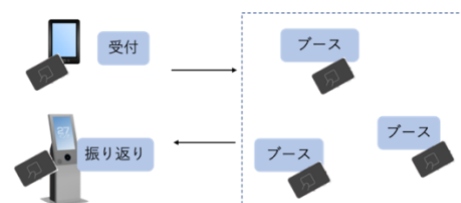


図 1. AI タッチラリーの概要図

<sup>1</sup> 産業技術総合研究所 人工知能研究センター  
AIST Artificial Intelligence Research Center

## 2.2 実験環境

サイエンスアゴラ 2019 にて AI タッチラリーを導入し、イベントの来場者のデータを収集した。イベントの開催場所は日本科学未来館、シンボルプロムナード公園、テレコムセンター1F, 3~5F でセッション、ブース合わせて 140 企画以上が開催された。そのうちテレコムセンター1F, 3F~5F に設置された 88 のブースにて AI タッチラリーを導入し、データ収集を行った。図 2 にブース配置図を示す。来場者はまず 1F の受付にて RFID シールつきリストバンドを受け取り、ID 認証端末にそのリストバンドをタッチし、アンケートに答える。その後、各ブースを体験し、各ブースに設置されている ID 認証端末にタッチすることで回遊行動とアンケート、記録時間とリストバンドの ID のデータを対応付けたデータが収集される。ブース回遊後は 1 階の振り返り地点に設置されている ID 認証端末にタッチすると、来場者が回遊したブースが提示され、最終アンケートに答える。以上が AI タッチラリーの導線になる。今回は受付はサイエンスアゴラ 2019 に入場するために必須としたが、振り返りは任意とした。また、3 階と 5 階の案内所では、リストバンドを ID 認証端末にタッチすると受付で答えたアンケートを元にオススメのブースを提示した（オススメシステム）。図 3 に受付、各ブース、振り返りにて行ったアンケート一覧を示す。

また、2018 年 11 月 10, 11 日に行われたサイエンスアゴラ 2018 では 5 階などの上の階まで上がる人が少ないという課題を解決するために、上の階をオススメする確率の高いアルゴリズムであるオススメシステムを導入し、来場者の利用率が高まるよう積極的に呼び込みを行った。その結果、来場者が比較的上の階まで足を伸ばすようになった。また、サイエンスアゴラ 2018 ではパンフレットで注目企画と書かれていたブースの集客率が高いこともわかった。以上を元に、サイエンスアゴラ 2019 では、2018 で行った上層階に誘導するオススメシステムのアルゴリズムではなく、どの階のブースも均等にオススメするアルゴリズムを用いることで、オススメシステムでの上層階への誘導を廃止するとともに、パンフレットにある注目企画のブースを上層階に多く配置することで上層階へ行く人が増えるよう誘導するようになった。

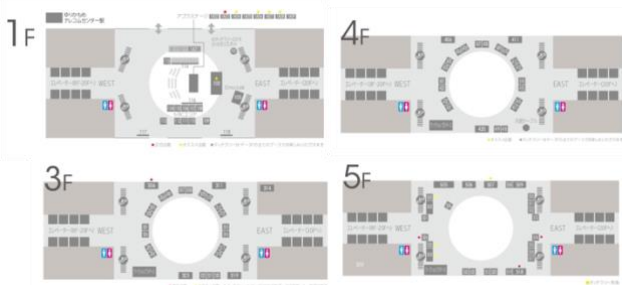


図 2. ブース配置図

受付	あなたの年齢を教えてください	10歳未満 10代 20代 30代 40代 50代 60代 70代以上
受付	あなたの職業を選択してください	小学生未満 小学生 中学生 高校生 高専生 大学・大学院生 専門学校生 研究職(人文・社会学系) 研究職(科学・技術系) 専門・技術職 事務職 教職 教職 パート・アルバイト 専業主婦・主夫 定年退職・無職 その他
受付	今日はどなたと来場しましたか	1人 家族と 友人と その他
受付	関心のある科学分野を選択して下さい(複数回答可)	暮らしと健康 教育 医療 情報 機械 日本・世界の未来 生物・環境・宇宙 算数・数学 化学・エネルギー 選択肢には興味のあるものがない
各ブース	今の気持ちを教えてください	up (顔画像) mid (顔画像) down (顔画像)
各ブース	このブースのイメージを色で表すと? (直感で!)	赤 黄 緑 青 紫 黒
振り返り	サイエンスアゴラはいかがでしたか(10点満点)	10点 9点 8点 7点 6点 5点 4点 3点 2点 1点
振り返り	サイエンスアゴラを他の人にどのくらい勧めたいですか (10点満点)	10点 9点 8点 7点 6点 5点 4点 3点 2点 1点
振り返り	新たに興味を持った分野はありますか	暮らしと健康 教育 医療 情報 機械 日本・世界の未来 生物・環境・宇宙 算数・数学 化学・エネルギー 選択肢には興味のあるものがない
振り返り	活発で、人とよく関わると思う	強く思う やや思う どちらとも言えない あまり思う 全く思わない
振り返り	人に気を遣う、優しい人間だと思う	強く思う やや思う どちらとも言えない あまり思う 全く思わない
振り返り	しっかりしていて、自分に厳しい人間だと思う	強く思う やや思う どちらとも言えない あまり思う 全く思わない
振り返り	心配性で、おろおろしやすいと思う	強く思う やや思う どちらとも言えない あまり思う 全く思わない
振り返り	新しいことが好きで、変わった考えを持つと思う	強く思う やや思う どちらとも言えない あまり思う 全く思わない

図 3. アンケート一覧

### 3. データセットと分析手法

2の実験により得られたデータで分析を行った. その手法について述べる.

#### 3.1 データセット

本実験では来場者に配布したリストバンドのRF-ID, ブース番号, ID 認証端末にタッチした時間, 質問の回答を記録し, それらを統合したデータを分析対象とした. ブース番号は図2に記載されている番号と同じものを使用し, 例えば101のような1から始まるブース番号は1階のブース, といったようにブース番号の初めの数字が階数を示している. ただし今回の実験では, 1階のブースで使用する機材の設置がきちんと行われていなかったことより, 正確にデータを取ることができず, 他の階のブースに比べて圧倒的に回遊した来場者の人数が少なくなっている. また, イベント時間外にタッチしているRF-ID, 回遊を受付から始めていないRF-IDを運営側のRF-IDとし, データセットから除外した.

#### 3.2 PLSA を用いたクラスタリング

確率的潜在意味解析 (Probabilistic Latent Semantic Analysis, PLSA) [7] はトピックモデルの1つで, 2つの変数  $x, y$  の背後に共通特徴となる潜在変数があると仮定し,  $x, y$  を同時にクラスタリングする手法である. [8] 今回の解析では,  $x$  を来場者のRF-ID,  $y$  をその来場者の回遊ブースIDとし, 来場者と回遊ブースのクラスタリングを行った.

#### 3.3 ベイジアンネットワークを用いた可視化

ベイジアンネットワークは確率変数をノードとし, 変数同士の依存関係を確率的なネットワークとしてモデル化したものである. ベイジアンネットワークによって構築したモデルはグラフとして可視化することが容易であり, 変数の意味を視覚的にもとらえやすいという利点があり先行研究で用いられている. [8] 今回の解析ではデータの可視化にベイジアンネットワークを用いた.

### 4. 大規模イベントにおけるデータ分析結果

受付, 振り返りで行ったアンケートの集計結果を図4に示す. 受付では2816人, 振り返りでは785人のアンケートを収集した. 今回はクラスタ分析, 人気のあるブースに関する分析の2点に着目して述べる.

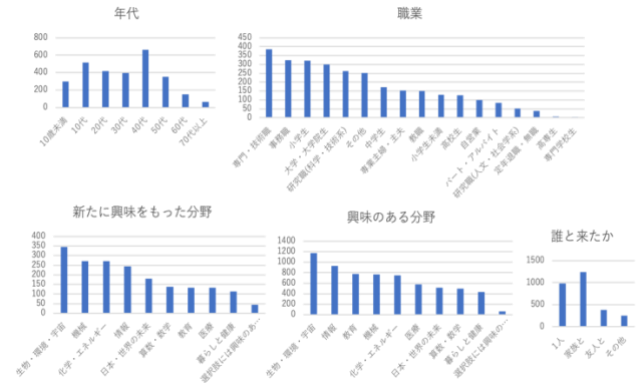


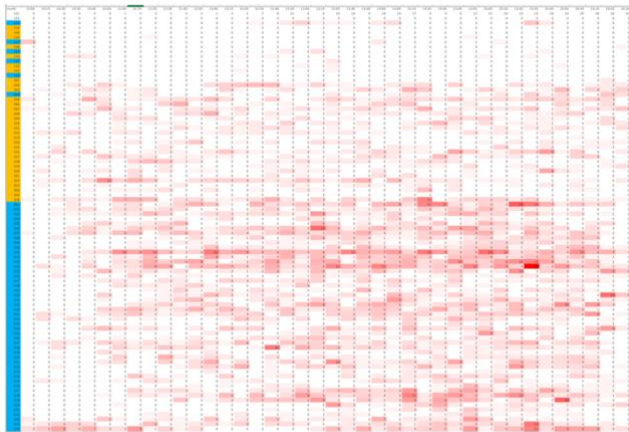
図4. 受付, 振り返りのアンケート集計

#### 4.1 クラスタ分析

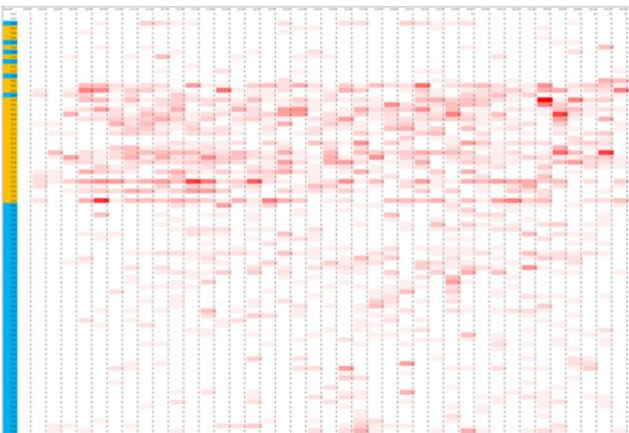
前記の通り,  $x$  を来場者のRF-ID,  $y$  をその来場者が周遊したブースとしてクラスタリングを行った. クラスタ数を変化させてPLSAを実行したところ, クラスタ数が2の時にAICが最小となったため, 2つのクラスタでクラスタ比較を行った. 本論文では2つのクラスタをそれぞれ2Z1, 2Z2といった名前をつけて扱う. なお, クラスタリング対象者は1つ以上のブースを周遊した者とし, 2Z1が771人, 2Z2が473人, 計1244人である. クラスタリング対象ブースはテレコムセンター1F, 3F~5Fにあるブースから振り返り, 案内所2つを抜いたブースとし, 2Z1が54個, 2Z2が31個, 計85個である.

2Z1, 2Z2それぞれの来場者が回遊したブースの時間帯ごとの集計表を図5に, 立ち寄ったフロアの集計を図6に示す. 図5では, 横軸を時間帯, 縦軸をブース番号とし, 赤色が濃いほど立ち寄った人数が多いことを示す. 縦軸において, 青色のブースは2Z1に, 橙色のブースは2Z2に所属している. この2つの図から, サイエンスアゴラ2019に来た人は4,5階をメインに様々な階に立ち寄るグループ(2Z1)と, 3階をメインにまわっていてあまり上の階に行かないグループ(2Z2)があることが分かる. また, 2Z1の方が多くの階に立ち寄っている確率が高い一方, 平均滞在時間は2Z1が3時間51分, 2Z2が3時間49分とさほど差がないことより, 2Z2の方が1フロアでの滞在時間が長いことが分かる. 2Z1と2Z2にはそれぞれどのような人が所属しているのか見てみると, 2Z1には化学・エネルギーに興味がある人や回遊後に暮らしと健康に興味を持った人, ブースを11個以上周遊している人が多く, 2Z2には算数・数学に興味がある人が多いことが分かる. 特に化学・エネルギーに興味のある家族・友達と来た人は2Z1に所属する確率が非常に高く, 数学・算数に興味のある1人で来た人は2Z2に所属する確率が高い. 来場者全員の回遊したブースの数の第3四位数が11個であることより, 2Z1にはたくさん回遊した人が2Z2より多くいることが分かる.





2Z1



2Z2

図 5. 回遊ブースの時間帯ごとの集計表

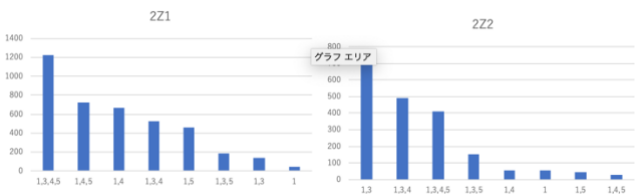


図 6. 立ち寄ったフロアの集計

#### 4.2 人気のあるブースに関する分析

“人気がある”という指標について、今回は 2 つの物差しで測ることにした。1 つ目が単純に周遊した人数が多いブース、2 つ目は振り返りで投票を行い、投票数が多かったブースである。はじめにブースごとの周遊した人数について述べる。図 7 にブースごとの周遊した人数を記したフロア図を示す。各ブースに記してある人数が周遊人数で、橙色が濃いほど人数が多く、緑色が濃いほど人数が少ないことを示し、赤い星は注目企画であることを示す。今回、開催時間のほとんどでブース番号 322 と 325 の回遊人数を誤って統合してしまっていたため、正しい値よりも多くなっていることと、先にも述べたように 1 階では正確にデータが取れておらず全体的に少なくなっていることより、1 階とブース番号 322 と 325 の周遊データを分析データから除外した。図から、必ずしも注目企画の周遊人数が多い

わけではないこと、人数の多いブースは比較的かたまっていること、西側（図左側）のエレベーターの近くほどのフロアでも人数が多くなっていることなどが分かる。フロアごとの集計では 4 階が一番周遊人数が多く、次いで 3 階 5 階である。



図 7. 周遊人数のマッピング図

次に振り返りでの投票について述べる。今回振り返りでは、1 人につき 0~3 個の良かったブースに投票を行った。投票人数は 785 人、投票総数は 1193 票であった。15 票以上の表が集まったブースの投票数のマッピング図を図 8 に示す。やはり周遊人数の多いブースは投票数の多い割合が高くなっているが、投票数が多いにも関わらず周遊人数が少ないブースが少しあることが分かる。また、投票という視点から見ても注目企画が多いわけではないことも分かる。フロアごとの比較では、3 階が一番投票数が多いブースが多く、次いで 4 階 5 階である。また属性を見てみると、専業主婦・夫向けのブースは投票数が多くなっていることが分かる。これは家族連れの来場者が多かったことも関係していると思われる。

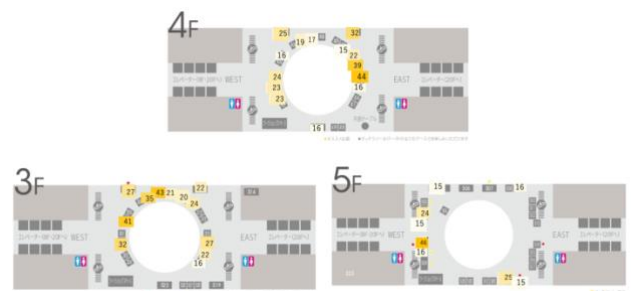


図 8. ブースの投票数のマッピング図（15 票以上）

また、各ブースでは回遊後にその時の気分を上・中・下の 3 択で尋ねた。これは気分であるので直接的なブースの評価ではないが、回遊直後の気分であるのでブースの評価とも関係があるのではないかと考え取り入れたアンケートである。単純な集計結果でいうと上をつけた来場者が多いため、上に対する下の割合を出し分析を行った。結果、上の割合が高いまたは低いブースと周遊人数、投票数とは相関がなく、フロアごとの偏りもなかった。よって今回は気分を聞いたが、直接的にブースの感想を聞くことで、ブースの回遊人数に関わらないブースの評価が可能である可能

性がある。また、上に対する下の割合が高いブースには、算数・数学に関するブースであるという特徴があった。

## 5. 考察

2でも述べたようにサイエンスアゴラ 2019 では、サイエンスアゴラ 2018 で採用したオスメシステムでの上層階への誘導を廃止するとともに、パンフレットにあるオスメのブースを上層階に多く配置することで上層階へ誘導するようにした。すると、3,4階に行く人に比べて5階に行く人の割合は少ないという結果になった。注目企画のブースに関して、投票数や周遊人数が一律して多いと言えないことから、サイエンスアゴラ 2019 では注目企画を上層階に固める施行の効果があったとは言にくい。この点サイエンスアゴラ 2018 とは違った結論が出た理由の1つとして、サイエンスアゴラ 2018 では出展していた全てのブースではなく一部のブースのみ AI タッチラリーに参加していたことが挙げられると思われる。次年度に開催されるサイエンスアゴラ 2020 でオスメシステムを使わずに上層階に人を誘導する施策として考えられるのが、4-1で述べた3階をメインにまわっていてあまり上の階に行かないグループ(2Z2)に人気であるブースを上層階に配置することである。さらに、サイエンスアゴラ 2018 で導入し、効果が見られたオスメシステム[5]を活用し、どのような関心を持つ人に対しても、適切なブースの内容と場所を紹介することで上層階への回遊を促す方法が考えられる。また、この際、オスメシステムの利用率を向上することも合わせて必要となるが、そのためにオスメシステムの利用を積極的に促せるような擬人化エージェントを導入することで、サイエンスアゴラ 2018 では必要だった呼び込み要員を代替することを考えている。これらの効果を実証することが今後の課題である。

## 6. おわりに

本研究では大規模イベントにおいて AI タッチラリーを用いることで 2816 名の来場者のデータの取得を行い、イベント改善に反映するための分析結果を示した。運用に関する改善点として、今回の実験では、来場者への AI タッチラリーの説明が十分でなかったために、本来はさらに多数のデータが収集可能であると考えられる。次回の実験に向けて、来場者や運営側への AI タッチラリーの説明やスタッフ間の連携を進めるサービスシステムの改善によってデータ収集だけでなく、価値の提供や来場者満足度の向上についても改善していくことが必要である。そのために今回、収集した投票や気分のような価値を表す目的変数に関する制御モデル化と、その目的変数を改善する持続的な制御手法の検討も重要である。

**謝辞** 本研究は国立研究開発法人科学技術振興機構 (JST)、ソニーモバイルコミュニケーションズ株式会社、SFI リーシング株式会社、新エネルギー・産業技術総合開発機構 NEDO 事業「次世代人工知能・ロボット中核技術開発/次世代人工知能技術分野 /人間と相互理解できる次世代人工知能技術の研究開発」、及び「次世代人工知能・ロボットの中核となるインテグレート技術開発/人工知能技術の社会実装に向けた研究開発・実証及び人工知能技術の適用領域を広げる研究開発/サイバーフィジカルバリューチェーンの構築・AI 導入加速技術の研究開発」の支援を受け、産業技術総合研究所人工知能技術コンソーシアムの社会課題解決ワーキンググループのサイエンスアゴラプロジェクトチームの活動として実施を行いました。

## 参考文献

- 1) 近藤那央, 竹内理人, 櫻井瑛一, 本村陽一: ID カードAI 対話システムを用いたイベント空間における行動データの収集と行動支援技術, 人工知能学会合同研究会, 2016
- 2) 近藤那央, 竹内理人, 山下和也, 櫻井瑛一, 本村陽一: 大規模イベントにおける人の行動履歴情報収集とそれを利用したインタラクティブシステムの開発に向けて, 人工知能学会全国大会, 2017
- 3) 近藤那央, 原田奈弥, 山下和也, 大前智嵩, 本村陽一: 大規模イベントにおける来場者回遊行動分析, 人工知能学会全国大会, 2018
- 4) 大和田智之, 山下和也, 大前智嵩, 本村陽一: イベント参加者の属性情報および行動履歴データの分析と活用, 人工知能学会全国大会, 2018
- 5) 大和田智之, 對間悠一, 山下和也, 高松倫芳, 櫻井瑛一, 高岡昂太, 大塚芳嵩, 澤谷真澄, 齊藤裕一郎, 中庭伊織, 長谷篤拓, 潤間励子, 本村陽一: イベント来場者への行動推薦による変容の分析, サービス学会, 2019
- 6) 山下和也, 對間悠一, 大和田智之, 高松倫芳, 櫻井瑛一, 高岡昂太, 大塚芳嵩, 澤谷真澄, 齊藤裕一郎, 中庭伊織, 長谷篤拓, 潤間励子, 本村陽一: イベント来場者と企画者双方へ最適化されたサービスを提供するAI対話型行動支援システム開発に向けて~サイエンスアゴラ2018での実証実験~, サービス学会, 2019
- 7) T.Hofmann and J.Puzicha, "Latent class models for collaborative filtering", *Proc. 16th international joint conference on Artificial intelligence*, 1999
- 8) 古田真理, 山下和也, 碓井舞, 内藤まゆこ, 本村陽一: 実社会ビッグデータと確率モデルを用いた施設内サービス空間でのイベント来場者の行動・感情分析, 2019