

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

Analysis and utilization of the action history of the event participants

大和田智之*¹
Tomoyuki Owada

山下和也*¹
Kazuya Yamashita

大前智嵩*¹
Tomotaka Omae

本村陽一*¹
Yoichi Motomura

*¹ 産業技術総合研究所

National Institute of Advanced Industrial Science and Technology

There's more application in the field of data collection and data analysis recently. POS data for marketing, search or access log in web, image data with GPS to track human action history, are the remarkable examples. In this research, we collected the action history of the event participants by using ID card. We report the result of clustering using pLSA, the investigation for the causal relationship between participants' movement and booths arrangement, and also consider more application in the future of this system.

1. はじめに

実社会におけるビッグデータの収集, 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)収集と, これを用いたデータ分析はサービスの改善への応用事例も多い[関 03, 石垣 11]. こうしたマーケティングの分野における商品の購買履歴(POS データ)だけでなく, web における検索履歴やアクセス履歴の活用, さらに実世界の位置情報を付与した GPS データ付き画像データなどを用いた人の行動分析など, その対象はさらに広がってきている.

本研究では新たに大規模なイベントにおいて, 来場者全員に ID 付きカードを付与することで来場者の属性(年齢や性別など)を登録し, さらに来場者が訪問した展示ブースでカードを端末にタッチすることで得られる回遊データの分析を行った結果を報告する. また, このイベント支援システムの今後やさらなる活用方法についても検討する.

2. データ収集

2.1 収集方法

人の行動履歴を収集する場合, まず検討しなければならないのがデータの収集方法である. 方法として, GPS を利用する方法, カメラによる人認識を利用する方法, ID カード等の IC タグを利用する方法などが提案されている.

GPS を利用した場合, 同じ建物内などの細かい移動には十分な精度を発揮しづらいという問題がある. カメラを利用する場合, プライバシーの問題があり, また, 技術的にも設備的にも高コストである.

本実験ではシール型の IC タグをイベント来場者に配布し, 各ブースでタッチしてもらうことによりデータ収集を行った. これにより, 行動履歴の収集と同時に各ブースの満足度を来場者に入力してもらうことが可能になった.

2.2 イベント概要

実験は 2017 年 11 月に行われたサイエンスアゴラ 2017 というイベントで行った. サイエンスアゴラは 2006 年から毎年行われ

ている科学と社会をつなぐ日本最大級のオープンフォーラムで, 例年の出展団体数は 100~200 団体, 参加者数は 4000 人~1 万人に及ぶ. サイエンスアゴラ 2017 では, ブースは 1 階, 3~5 階の 4 つのフロアで実施され, フロアごとでそれぞれテーマが分かれていた. セッションは 8 階と 20 階で実施された. このイベントにおいてデータ収集を兼ねたタッチラリー企画を行い, 参加者に各ブースやセッションを回ってもらい, 参加賞として飲料を配布した. 3 日間の開催のうち 1 日分, 分析に利用可能な約 1000 人の来場者データについて分析を行った.

2.3 タッチラリーの流れ

イベント来場者はまず 1 階の受付でシール型 IC タグを貼付したカードを受け取り, 受付横にあるサイネージ(図 1)で年齢や性別などの簡単な属性情報を入力する. その後来場者は, 各ブースを見学した後, 設置された端末に IC タグをタッチし, 満足度を入力する. 各ブースに置かれた端末の様子を図 2 に示す.

最後に, 5 つ以上のブースを見学した参加者は, 出口に設置されたサイネージ体型の自販機(図 3)に IC タグをタッチすることにより, ジュースを受け取ることができる. また同時に, 参加者がイベント中に回ったブースが画面上に経路として表示される.

また企画提供者は, イベント開催中からイベント開催後に開催したブースの年齢別, 男女別の来場者数や, 平均評価を見ることができる. これにより, 各々の提供者が企画の質の向上に役立てることができる.



図 1: 受付に設置されたサイネージ端末

連絡先: 大和田智之, 産業技術総合研究所人工知能研究センター, 〒135-0064 東京都江東区青海 2-4-7, t.owada@aist.go.jp

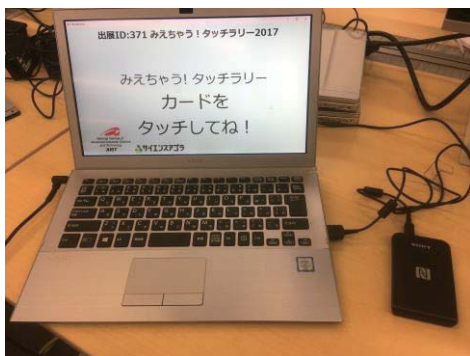


図 2: 各ブースに設置された端末



図 3: サイネージ体型の自販機

3. 分析手法

3.1 pLSA を用いたクラスタリング

pLSA[Hofmann 99, Hofmann 01]は、2 変数 x, y の間に共通の潜在変数 $z \in Z = \{z_1, z_2, \dots, z_k\}$ があると仮定し、 x, y に関する確率が最も高いような z を選ぶことにより x, y を同時にクラスタリングする手法である。

pLSA は、元々は自然言語処理などの分野で活用されてきた手法であるが、最近では本研究に関連した分野も含む多くの分野で応用されている。pLSA を用いた POS データの活用事例として、石垣らは、顧客と商品の同時クラスタリングを行っている[石垣 10]。

前章で示した方法により得られたデータを用い、来場者を x 、各ブースを y として pLSA によるクラスタリングを行った。使用したデータにおける来場者の総数は 976 人、ブースの総数は 51 個、来場者によるブース訪問数の合計は 6440 件であった。

3.2 ベイジアンネットワークの構築

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る。また、ある変数 x とその変数に影響を及ぼ

す変数 z_1, z_2, \dots, z_k の関係は条件付確率 $P(x|z_1, z_2, \dots, z_k)$ として得られるため、因果関係の分析が容易である。

4.2 節では、参加者の属性情報、行動履歴(各ブースの滞在時間、評価等)、および pLSA による同時クラスタリングの結果を利用したベイジアンネットワークによるモデルの構築結果とその考察を示す。

4. 結果と考察

4.1 ブースのクラスタリング結果と階層の関係

pLSA を用いて来場者とブースを4クラスタに分けた場合のクラスタリング結果を、ブースごとにまとめたものを表1に示す。表の各列はそれぞれのクラスタに属する確率を示し、各行が各ブースに対応する。表から、各クラスタが各ブースの位置するフロアとかなりの精度で対応していることが分かる。来場者は同じ階のブースを優先して回る傾向があり、逆に言えばブースの位置関係は来場者の行動傾向に極めて大きい影響を与えているといえる。

また、2 クラスタに分けた場合では、1 階と 3 階のブースがクラスタ 1、4 階と 5 階のブースがクラスタ 2 に属する傾向があり、3 クラスタに分けた場合では 4 階と 5 階のブースがさらに 2 つに分かれることが分かった。いずれも表 1 に示した結果と同様、クラスタと位置するフロアは高い相関を示した。

4.2 参加者の属性情報と行動履歴の関係

参加者の属性情報、行動履歴、および前節の結果から得られたクラスタをベイジアンネットワークによりモデル化し、グラフを可視化した結果を図 4 に示す。ここで、前節の考察に基づき、クラスタ名には各フロアを回っている確率が高いという意味で、それぞれ floor1like, floor3like, floor4like, floor5like と名付けた。

ここで Discharge は出口で飲み物を受け取ったかどうか、また受け取ったならばその時間帯、TotalResidence はイベント滞在時間、AveResidence はブース一つあたりの平均滞在時間、AveEvaluation はブース一つあたりの平均評価を表す。

得られたモデルから、floor1like に属する参加者はブース訪問数が少なくブース一つあたりの滞在時間が長い傾向にあることが分かった。このため、1 階を回って満足したもしくは疲れてほかの階を回らなかったことが考えられる。また floor5like に属する参加者は比較的早い時間からきて長く滞在している傾向にあった。



図 4: 属性情報、行動履歴から構築したベイジアンネットワーク

表 1: ブースの位置するフロアとクラスタの関係

floor	Z1	Z2	Z3	Z4
1	1.00	0.00	0.00	0.00
1	1.00	0.00	0.00	0.00
1	1.00	0.00	0.00	0.00
1	1.00	0.00	0.00	0.00
1	1.00	0.00	0.00	0.00
3	0.00	1.00	0.00	0.00
3	0.00	1.00	0.00	0.00
3	0.00	1.00	0.00	0.00
3	0.00	1.00	0.00	0.00
3	0.00	1.00	0.00	0.00
3	0.00	1.00	0.00	0.00
3	0.00	1.00	0.00	0.00
3	0.00	1.00	0.00	0.00
3	0.00	1.00	0.00	0.00
3	0.00	1.00	0.00	0.00
3	1.00	0.00	0.00	0.00
3	0.30	0.54	0.00	0.16
3	0.00	0.81	0.19	0.00
3	0.00	1.00	0.00	0.00
3	0.00	1.00	0.00	0.00
4	0.18	0.00	0.82	0.00
4	0.06	0.00	0.94	0.00
4	0.00	0.00	1.00	0.00
4	0.00	0.00	1.00	0.00
4	0.00	0.00	1.00	0.00
4	0.00	0.00	1.00	0.00
4	0.00	0.00	1.00	0.00
4	0.00	0.00	1.00	0.00
4	0.00	0.00	1.00	0.00
4	0.00	0.00	1.00	0.00
4	0.00	0.00	1.00	0.00
4	0.19	0.00	0.81	0.00
4	0.00	0.00	1.00	0.00
4	0.15	0.00	0.85	0.00
4	0.00	0.00	1.00	0.00
5	0.00	0.00	0.00	1.00
5	0.00	0.00	0.00	1.00
5	0.00	0.00	0.00	1.00
5	0.00	0.00	0.00	1.00
5	0.00	0.00	0.00	1.00
5	0.00	0.00	0.00	1.00
5	0.00	0.00	0.00	1.00
5	0.00	0.00	0.00	1.00
5	0.00	0.00	0.26	0.74
5	0.00	0.00	0.00	1.00
5	0.00	0.00	0.00	1.00
5	0.00	0.00	0.29	0.71
5	0.00	0.00	0.00	1.00
5	0.00	0.00	0.00	1.00
5	0.00	0.00	0.00	1.00
8	1.00	0.00	0.00	0.00
20	0.00	0.00	0.00	1.00

4.3 ブースの位置と参加者の興味に関する考察

ブースのクラスタリング結果およびユーザーの行動を分析した結果はいずれもブースの位置が参加者の行動に強い影響を与えていることを示している。2.2 節で記したように分析対象としたイベントではフロアごとのテーマ分けがなされており、ブースの位置が及ぼす影響とテーマ分けが及ぼした影響を完全に切り分けることはできない。しかしながら、ペイジアンネットワークの結果から、近いフロアから順に回っていき、時間の都合やイベントへの満足度から遠くのブースまで回るかを考える参加者が多いと考えられる。

今後のブース配置を検討するにあたっては、時間がない人でも楽しめるように低層階に様々なジャンルのコンテンツをちりばめる一方で、各階にも目玉となるようなブースを用意することで参加者に各フロアを回ってもらうといったことが考えられる。

5. まとめ

本稿で示したデータ収集のシステム、データ解析のプロセスはイベントの価値を高めるにあたって一定の成果を出せたといえる。今後は、参加者の回遊情報からリアルタイムにブースを提案するシステムや、AI を用いたインタラクティブなイベント支援など、さらなる応用が期待される。

謝辞

本研究は NEDO 委託事業「人間と相互理解できる 次世代人工知能技術の研究開発」の支援による。国立研究開発法人科学技術振興機構の主催するサイエンスアゴラ 2017 において、公式イベントとして実験を実施した。サイネージ体型自動販売機により来場者に提供した飲料は、麒麟株式会社様より提供を受けた。

参考文献

- [関 03] 関 庸一: ID 付き POS データからの顧客行動パタンの抽出, オペレーションズ・リサーチ: 経営の科学, Vol.48, No.2, pp.75-82, 2003.
- [石垣 11] 石垣司, 竹中毅, 本村陽一: 日常購買行動に関する大規模データの融合による顧客行動予測システム, 人工知能学会論文誌, Vol.26, No.6, pp.670-681, 2011.
- [Hofmann 99] T.Hofmann and J.Puzicha, "Latent class models for collaborative filtering", *Proc. 16th international joint conference on Artificial intelligence*, 1999.
- [Hofmann 01] T.Hofmann, "Unsupervised Learning by Probabilistic Latent Semantic Analysis", *Machine Learning*, Vol.42, No.1-2, pp.177-196, 2001.
- [石垣 10] 石垣司, 竹中毅, 本村陽一: 確率的潜在意味解析を用いた大規模 ID-POS と顧客アンケートの統合利用による顧客-商品の同時カテゴリ分類, 信学技報, Vol.109, No.461, NC2009-160, pp.425-430, 2010.