

# 購買時の候補商品推定システムの構築と初期評価

徳永 大空<sup>1</sup> 本松 大夢<sup>1</sup> Chen Bin<sup>1</sup> 中村 優吾<sup>1</sup> 荒川 豊<sup>1</sup>

**概要:** オンライン購買システムは、ページ遷移やクリックの履歴から購買者が商品の購買に至るまでの過程を記録することができる。これにより、興味がある商品や比較検討した商品がわかり、購買者のニーズに近く購買意欲を高めるような商品推薦が可能である。一方、実店舗においては商品の購買以外の情報は、取得と活用の両面における難しさからあまり記録されていない。そのため、購買者がどのような商品に興味を持っており何で悩んでいるのかがわからないという問題がある。こうした問題を解決するために、我々はLiDAR (Light Detection And Ranging) からの点群データと Depth (深度) カメラからの視線情報を用いた購買時における候補商品推定システムを構築し、その初期評価を行った。推定の上では、興味のある商品を購買候補商品とした。評価の結果、購買時の視線は候補商品と強く結びつき、視線の情報は候補商品を推定する上で非常に重要な特徴となり得ることがわかった。一方で、視線の評価を踏まえ点群データを使用した候補商品の予測を行った結果、予測対象の人物を含んだ状態で学習したモデルは49%と12クラス分類であることを考えると一定の精度を出すことができた。一方でLOPOでは精度が最大22%と、モデルの一般化はまだ課題が残ることがわかった。今後は、データ数を増やし、モデルをより適したものに变更・改善していく必要がある。

## Construction and initial evaluation of a system for estimating candidate products at the time of purchase

Taku Tokunaga<sup>1</sup> Hiromu Motomatsu<sup>1</sup> Chen Bin<sup>1</sup> Yugo Nakamura<sup>1</sup> Yutaka Arakawa<sup>1</sup>

### 1. はじめに

近年のオンライン購買システムは、購買履歴やページ遷移、サイトの滞在時間などを保存・収集することができ、それらのデータを分析することで購買者が商品の購入に至るまでの軌跡を追うことができる。これによって、購買者ごとに興味のある商品や実際に買った商品、比較検討した商品を把握することができ、購買者のニーズに近い、購買意欲を高めるような商品推薦が可能となっている。例えば、協調フィルタリングなどの技術は多くの商品推薦の文脈で使用されており、これは購買履歴と購買に至るまでの軌跡が記録されていることで可能となった。その一方で、実店舗において記録されているデータはPOS(Point of Sales)データと呼ばれる購買商品の情報以外は、取得と活用の両面における難しさからあまり記録されていない [1]。

そのため、オンラインの購買システムでは可能である、

購買者の興味の把握とそれに応じた適切な商品推薦ができないという問題がある。こういった問題を解決することができれば、オンラインとオフラインで購買者に提供する情報の質はなくなり、利便性という面で電子商取引 (EC) に劣ってきた実店舗での購買を促すことができると考えられる。これは近年注目されているO2O(Online to Offline)を促進することに繋がり社会的意義も大きい [2]。

実店舗での購買者の興味、すなわち購買時に候補に上がった商品群を得るために必要な購買時のデータは先行研究から、迷いなどに関連の深い購買時の行動データ (滞在時間や位置、商品を手取る動作など) と人の興味と関連の深い視線データであると考えられた。[3-7]。また実験環境としてそれらのデータに加えて、推定のための目的変数として、その購買によって何を候補商品とし、最終的に何を購入したのかを記録する必要がある。これらの情報を取得するために、我々はLiDAR (Light Detection And Ranging) と Depth (深度) カメラ、RFID (Radio Frequency Identifier)

<sup>1</sup> 九州大学, Kyushu University

タグを使用した購買時のデータを取得するためのシステムを構築した。LiDARによって人の購買時の点群データを、Depthカメラによって視線データをRFIDタグによって購買商品情報を収集することができる。今回Depthカメラに加え、LiDARによるデータ収集を採用した理由としては、Depthカメラの画角による取得データ範囲の問題や光の加減などの環境要因による取得データ精度のばらつきを無くすためである。LiDARを用いた点群データであれば、カメラの画角や向き、環境要因に左右されることなく、購買者の行動を記録できる。

システムは、購買開始から購買終了までの間でこれらのセンサを自動で動作させ、購買後にアンケートによって候補商品を収集、RFIDタグによって最終的に購買した商品を記録する構成となっている。

本稿では、このシステムの詳細について述べ、さらに初期評価として取得可能な一連のデータを可視化し、購買時の点群データと視線データの有用性を評価する。具体的にはまず、視線データを使用し、購買時の視線と興味のある購買候補商品が結びついているのか否かを評価する。次に、点群データを使用し、人の行動・姿勢のみから購買時に見ている商品カテゴリを推定できるか否かを評価する。

これらの評価の結果、購買時には興味のある商品を興味の無い商品にくらべて約2倍見ていることから、購買時の視線は候補商品と強く結びつき、視線の情報は候補商品を推定する上で非常に重要な特徴となり得ることがわかった。また、点群データを用いた推定では、もっともよく適合したエポックで49%の精度を得ることができ、正解ラベルが予測確率上位3つの中に含まれているかとした場合は、80%の精度で含まれていることがわかった。これにより、我々のアプローチの有効性を示すことができた。一方で、LOPO(Leave One Person Out)での評価では、精度が22%と課題が残る結果となった。

本稿の構成は以下のとおりである。まず2章にて、購買ログの収集や興味推定の関連研究を示し、候補商品推定のために必要だと考えられるデータを洗い出す。次に3章にて、研究室内に実際に構築した実験環境とシステムの説明を行い、得られたデータの詳細を述べる。その後4章にて、それらのデータの有効性の評価を行い、最後に5章でまとめについて述べる。

## 2. 関連研究

### 2.1 購買ログ収集に関する研究

これまでの研究においても、商品購買時のデータを収集し、活用する研究は多く行われている。例えば、電子商取引(EC)などに代表されるオンラインでの購買では、購買時の購買・閲覧履歴、閲覧時間、クリックストリームなどのデータは以前から積極的に収集され、商品推薦や購買商品の予測に活用されている。[8,9]

また近年、オフラインにおいてもスマート商品棚(商品棚にカメラやセンサ、ディスプレイを取り付け、購買時のデータ収集と購買者への情報提示などを行うもの)やRFIDなどを用いることで、購買後のPOSデータにとどまらず、購買時のデータを収集する研究が行われるようになってきている。過去には、カメラやセンサーが取り付けられたスマート商品棚を用いた、予め購入する商品が決まっている非計画購買者と予め購入する商品が決まっている計画購買者の分類を行う研究[3]や、カメラとRFIDタグを用いた、購買時における迷い行動を検出する研究[4]などが行われた。また、最近では、コロナ禍の影響で実店舗における試食販売が実施できないという問題に着目し、深度カメラに基づくユーザ購買行動センシング機構と、調理映像と香りを用いた情報提示機構からなるインタラクティブシステムを用いて生鮮食品の購買促進に取り組む研究が存在する[6,7]。これらの研究では、購買者の迷いに注目し、商品棚前での滞在時間、位置、商品へ手を伸ばした時間等の特徴が活用され、迷いとこれらの行動には関連があることが明らかとなった。

このように、商品棚にセンサを設置し、購買時のデータを収集することで迷い状態や購買者の特性を推定する手法は、我々の研究において非常に有用な知見となった。しかしこれらの研究では、どの商品間で悩んだのか、どのような商品に興味をもったのかといった、購買者の心理状態には言及していない。購買時の候補商品を明らかにするには、どの商品に興味を持ったのかという購買者の心理状態を推測する必要がある。これらは先行研究で使用している特徴だけでは難しいと考えられる。

### 2.2 興味を推定する研究

先述したように、購買者が買おうとしている商品や購買時に候補となっていた商品を判断するには、購買者の興味や迷いといった内部的心的状態を推測する必要がある。興味を推定する研究において、人の視線情報は広く活用されている。

Liらの研究[5]では、スマートフォン端末のカメラを使って視線を推定しGoogle PlayStoreの画面における視線とユーザーが申告した興味を示した箇所との比較を行っている。その結果、ユーザーが興味があると申告した部分のほうが、それ以外の部分より有意に見ている時間が長くなっていることがわかった。またそれらの視線情報やユーザーの好きなカテゴリといった特徴を使用して、興味がある・ないの2値分類を行った結果、90%以上のAUCスコアを実現している。

他にもJacobらの研究[10]では、メガネ型のアイトラッカーを用いて視線情報を取得し、その情報を用いてどの新聞記事に興味を持ったのかを推測しており、Karolusらの研究[11]では、複数の視線推定・注視情報メトリクスを用

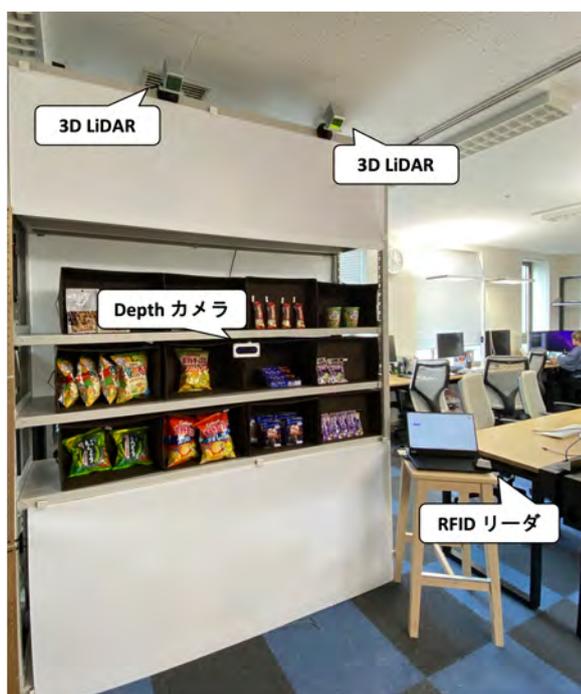


図 1 購買環境の様子

いて、提示される画像に対する興味推定を行い、どの推定においてもランダムより、良い精度を示している。このように、視線の情報から注意や興味を推定する研究は様々な分野において近年盛んになっている。

実店舗の購買においても商品パッケージや商品の棚内での位置と視線の関連を扱う研究 [12, 13] が行われているように、購買時における視線情報は人の興味を表す上で、とても重要な特徴であると考えられる。しかし、我々が作成した環境における購買時にも同様の傾向を示すのかについては検証が必要であると考え。以上を踏まえ、購買候補商品の推定のために購買時に取得すべきコンテキスト情報としては以下が挙げられる。

C1: 購買者の行動や滞在時間

C2: 購買者の視線情報

### 3. 購買行動ログ収集システム

2章にて、購買者の興味を推定するためには、行動や滞在時間、視線情報を収集する必要があることを示した。本章では、こうした購買者の行動ログを収集するために研究室内に構築したシステムについて説明を行う。本システムは、12種類の商品を設置した商品棚前での購買者の様々な購買行動ログを複数のセンサーを用いて収集することができる。収集する購買行動ログは以下の5つである。

D1: RFIDによるユーザID

D2: Depthカメラによる視線データ

D3: LiDARによる点群データ

D4: RFIDによる購買商品データ

D5: アンケートによる購買候補商品データ

#### 3.1 購買環境について

まず、本システムの購買環境について説明を行う。今回、実験的な購買環境を研究室(図1)に実装した。本購買環境における商品棚には1列に4つのボックスが3列、合計12個のボックスが置かれており、ボックスごとに異なる商品を陳列している。本論文では、1ボックスごとの商品を1カテゴリと定義し、12カテゴリのうち購買者がどのカテゴリに興味があるのかを推定することを目的としている。また、購買商品の登録や購買後のアンケートへの回答を行うためのセルフレジシステムを実装した、PCを商品棚近くに設置している。

#### 3.2 各種センサーについて

本システムでは、購買者の行動ログを収集するために3種類のセンサーを用いている。

1つ目のセンサーは、Depthカメラである。本カメラは、商品棚に購買者の目線と同じ高さに設置されており、購買者の顔の向きや眼球の動きを取得することができる。これにより、購買者がどの商品をどれくらいの時間見ていたかといった情報を収集することができる。本センサーを利用した理由としては、購買者にセンサーを取り付けることなく購買時行動情報を収集することができる点である。購買者がメガネ型のセンサーを取り付けて視線情報を収集する手法も考えられたが、購買者にセンサーを取り付けず購買時行動を記録することで実際の店舗での購買に近い状態でデータを集めることができ、購買者への負担が少なく本システムの継続的な利用も促すことができると考えた。

次に2つ目のセンサー、LiDAR [14] について説明を行う。本センサーは商品棚の上に2台設置されており、購買者の点群データを取得する事ができる。この点群データを用いることで、購買者がどの商品の前で立ち止まった、どの商品に手を伸ばした、どの商品の方に体が向いていたといった情報を収集することができる。本センサーを利用した理由は、点群データは画像データと比較して、処理が高速に行える点である。実店舗で候補商品推定を行い、候補商品に応じた商品推薦のような介入を行う場合、処理速度の早さは重要となってくる。また、Depthカメラでの録画データを確認したところ、画面が暗い、顔が隠れるなどの場合は視線推定を行うことができなかった。つまり、購買時における棚に近寄って手にとって確認するという場合の視線を取得することができない。しかしながら、このような行動こそ購買者の興味を強く反映していると考えられる。LiDARから得られる点群情報を用いて大まかな視線情報を推測することができれば、カメラではデータが取得できない暗い場合や顔が隠れる場合でも興味を推定できると考えられる。以上の理由から、今回はカメラでの録画ではなくLiDARでの購買行動の録画を採用した。

最後に3つ目のセンサー、RFIDリーダーについて説明

する。本システムでは、各商品と購買者には RFID タグが紐付けて登録されている。購買時に商品と購買者の RFID タグを本センサーに読み取らせることで、どの商品がどの購買者によって購入されたという情報を収集することができる。

### 3.3 実装したシステム

実装した購買ログ収集システムのフローを図 2 に示す。まず、購買開始時に購買者は自身と紐付いた RFID タグを RFID リーダーに読み取らせる。購買者タグを読み取ると LiDAR と Depth カメラの録画が開始する。次に、購買者は商品を選択する。商品選択が終わると購買者は商品の RFID タグをリーダーに読み取らせ購買を確定する。購買が確定したタイミングで LiDAR と Depth カメラの録画を止め、録画データを購買情報に紐づけて保存する。最後に、購買者は候補商品に関するアンケートに回答する。これにより、購買者の候補商品の正解データを収集することができ、候補商品推定システムの精度向上に役立てることができる。

### 3.4 システムから得られたデータ

本システムから得られたデータについて詳細を述べる。Depth カメラによる視線データは購買 1 回に対して 1 つの録画データが生成される。棚に取り付けられた Depth カメラから撮影するため、購買者と棚の距離が近すぎる場合は画角に入り切らない場合がある。LiDAR による点群データは図 3 に示したようなものである。この図は 4.2.2 項にて記述する前処理後の点群である。このように、購買商品の俯瞰、選択、確定までの様子が取得できていることが確認できる。棚とその周辺を棚上部に取り付けた LiDAR で撮影する。この点群データは LiDAR1 台につき 1 秒間で 10 万点取得できるため、今回は 1 秒で 20 万点の点群データが取得できる。購買 1 回に対して点群を 1 つの csv ファイルで保存する。RFID とアンケートによる、購買商品と購買候補商品のデータはすべて同じデータベース上に保存される。3.2 節で述べたように、RFID タグを用いて商品を管理しており、リーダーで読み取ることで最終的な購買商品を、3.3 節で述べた購買後のアンケートで興味のある購買候補商品を記録する。購買 1 回に対して 1 行のレコードとなる。

## 4. 評価

本章では、3 章のシステムを用いて得られたデータの評価を行う。まず、Depth カメラから得られた視線データが候補商品の推定に有用であるかを検証するために、視線データと候補商品の関係性の評価を行う。次に、視線での結果を踏まえ、LiDAR から得られた点群データを用いた候補商品の推定を行う。

表 1 学生ごとのフレーム数と興味のあるカテゴリ数

| ID | フレーム数 (Total frames) | カテゴリ数 ( $n$ ) |
|----|----------------------|---------------|
| 1  | 78                   | 3             |
| 2  | 138                  | 6             |
| 3  | 137                  | 3             |
| 4  | 84                   | 2             |
| 5  | 169                  | 3             |
| 6  | 136                  | 4             |
| 7  | 117                  | 2             |

### 4.1 Depth カメラから得られた視線データと、購買・候補商品の関係

2 章にて、視線情報と興味度の関連を示す先行研究を示した。これらの結果から視線と興味には関連があるものと考えられる。本節では、それらの先行研究を踏まえた上で、購買時における視線情報の有用性を検証するために、Depth カメラから得られた購買中の視線データと購買時の候補商品の関係性の評価を行う。

#### 4.1.1 視線の評価に使用するデータ

3 章にて説明した環境を用いて、同研究室の学生 10 人に実際に購買をしてもらいデータを収集した。この際に学生には、実験の詳細は伝えずに購買中のデータを取得する旨のみ伝えている。また、購買中の動作についても特に制限することなく、普段どおりの動作で購買を行ってもらった。これらの学生は、普段から研究室の購買環境を使用しており、商品の種類などはある程度認識していると考えられる。

表 1 に視線の評価に使用したデータを示す。これら 10 人の購買中のデータの内、研究室内の光の当たり具合などによって推定そのものが難しかった 3 人を除く 7 人のデータを使って評価を行う。7 人のデータは 1 秒間に 25 フレームで取得されており、より正確な推定を行うために購買中すべてのデータではなく、カメラに顔が写っているフレームのみを使用している。すなわち、顔がよく写っていないフレームや棚に近づいて商品を詳しく見る、カメラの範囲外で商品を手にとって見るなどの行動をしているフレームなどは取り除かれている。表 1 においてフレーム数が少なく、時間に戻すと 3-4 秒ほどになってしまう学生がいるのはこういったことが原因となっている。

#### 4.1.2 視線推定と評価指標

モデルの訓練のために、先述した購買時のデータを取得後に、それぞれの学生に指定の点を見てもらうように依頼し、訓練用のデータを取得した。そのデータで学生ごとにベースとなるモデルをファインチューニングして視線の推定を行う。ベースとなるモデルは Few-Shot Adaptive Gaze Estimation (FAZE) と呼ばれるフレームワークを使用している。このモデルを訓練時に使用したテストデータで検証した結果、それぞれ 5 度~8 度のズレであり十分に活用できると判断した。



図 2 購買行動ログ収集システムの動作フロー

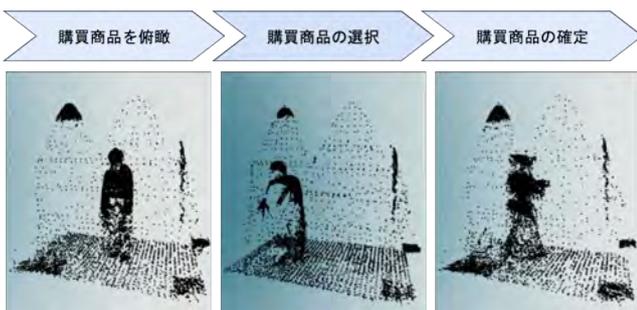


図 3 システムから得られた点群データ



図 4 視線情報の可視化

図 4, 5 はそれぞれ, ある学生の視線の推定結果を可視化したものと, その学生が興味があると回答したカテゴリである. 図 4 の可視化では, 赤い部分が特によく見ている部分であり, そこから青い部分にかけて見ているとモデルが判断した部分を表している. それ以外の部分は注視していないと判断された部分である. 図 5 の興味の図示では, 濃い青色の点が視線推定の結果の点を散布図で表したもので, 薄い青の範囲が興味があると回答したカテゴリの位置を表している. この位置は図 4 と一致している. この図で見ると, 1 番上の段の右上と真ん中の段の右から 2 列目が, 学生が興味があったと回答した部分であり, 視線はその周辺に特に集中していることが見て取れる. 他の学生においても同様の可視化を行い, 定量的に評価できるように以下のように指標を設定し, 評価を行った.

学生が後のアンケートで興味があると回答した商品カテゴリの範囲を  $I$ , 興味のあるカテゴリ数を  $n$  とする. 1 フレームあたりの視線情報  $g(\text{frame})$  が  $I$  の範囲に含まれている場合に 1, それ以外の場合には 0 となる関数を  $f(g, I)$  としたとき, 指標は

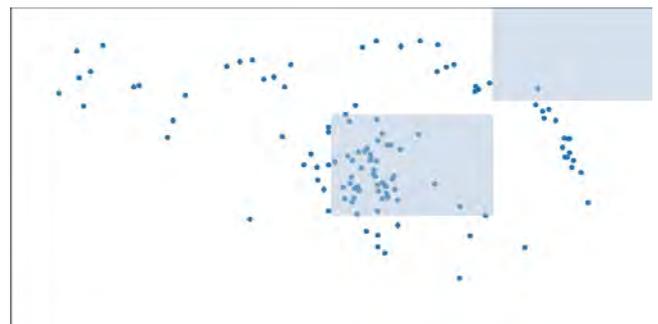


図 5 興味を示した商品の情報

$$G_{in} = \frac{1}{n} \sum_i f(g, I) \quad (1)$$

$$G_{out} = \frac{1}{12-n} (\text{Total frames} - \sum_i f(g, I)) \quad (2)$$

となる. つまり式 1 は, 図 5 の興味を示す薄い青色のエリアに入っている, 視線 (青色の点) の数を興味のあるカテゴリ数で平均したものとなっている. 式 2 は, 式 1 とは逆に, 興味を示す薄い青色のエリア外の視線の数を, 興味

表 2 点群データのフレーム数

|       | 訓練データ (4人の合計) | 評価用データ |
|-------|---------------|--------|
| フレーム数 | 4561          | 371    |

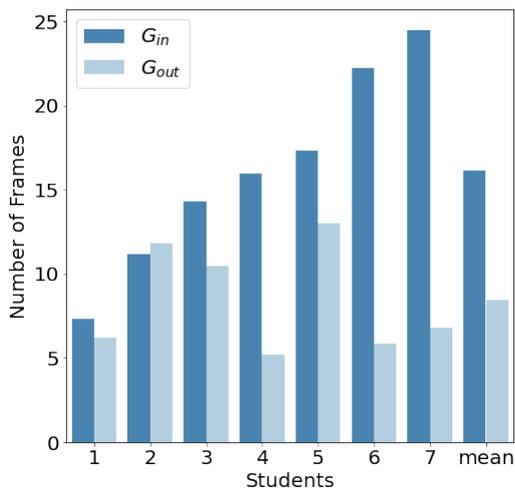


図 6 指標の計算結果

を示さなかったカテゴリ数で平均したものである。これにより、購買中、興味のある部分と興味のない部分をどの程度見ているのかを知ることができる。

#### 4.1.3 結果と考察

式 1, 2 の結果を図 6 に示す。濃い青色が  $G_{in}$ 、淡い青色が  $G_{out}$  を表しており、縦軸がフレーム数、横軸はそれぞれ被験者となった学生と全員分の平均値である。図を見ると、7人中6人で興味のあるものを見ている時間のほうが、それ以外のものを見ている時間よりも長いことがわかる。平均するとここから、視線は興味をよく表していると考えられる。

今回の視線の評価は 4.1.1 項でも述べたように、使用できるフレームが限られていた。カメラに写すことができるのは柵から一定以上の距離をとって俯瞰している場合なので、柵の前に来てから買うものを決定するまで段階の内、初期～中期の間のデータが多いと考えられる。早期の段階で購買候補との関連の深い視線の情報は、興味を持った購買候補商品を推定するだけでなく、購買中の情報提示などを見据えると、非常に重要な情報であるといえる。

## 4.2 LiDAR からの点群データを用いた候補商品推定

4.1 章にて、視線情報が興味のある候補商品と強く結びついていることが確認できた。しかし、上記のカメラを使った視線の推定には課題もある。その 1 つに、そもそもカメラにはっきりと映らないと正確な推定ができないことが挙げられる。今回の実装では推定が可能なように顔を撮るには、柵側から撮影する必要があり、単一のカメラでは画角の問題ですべての範囲をカバーすることは難しい。また、光の加減などの環境要因で精度が変動するなど、普通の購買で正確な推定をし続けるには現状では限界がある。

一方で LiDAR を用いた点群データであれば、カメラの

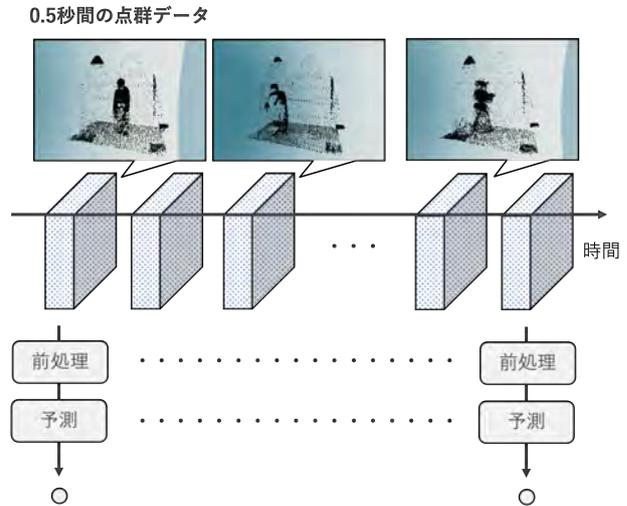


図 7 推定処理の概要

画角や向き、環境要因に左右されることなく、購買者の行動を記録できる。また今回のどの商品に興味を持っているかという候補商品を推定する状況において、詳細な視線情報は必要なく購買時にどのカテゴリの商品を見ているのかがわかれば良い。そこで、LiDAR から得られる点群データを使って、購買時にどの商品を見ているのかをクラス分類タスクとして推定を行う。これが可能であれば、カメラを使わずに興味を推定することができるようになる。

また LiDAR から得られる点群データは自動運転などの文脈で使用されることから分かる通り、画像や動画に比べて高速な処理が期待でき、今後見据えられる情報提示などへも活用しやすいと考えられる。

#### 4.2.1 訓練・評価用データの取得

学習、評価用のデータとして 4 人の学生に協力を依頼した。学生の身長や体型は研究室内でできるだけバリエーションをもたせるようにして選定している。これら 4 人の学生にそれぞれのカテゴリを姿勢を変えながら見てもらうという動作を 12 カテゴリ全てに対して行ってもらった。これを学習データとし、後日その中の 1 人に同様の指示をし評価用のデータを取得した。取得したデータをそれぞれ 0.5 秒ごとに切り分け、1 つの学習データ (0.5 秒の点群と目的変数のセット) とする。表 2 に訓練データと評価データのデータ数を示す。

#### 4.2.2 前処理とモデル

図 7 に今回の前処理とモデルの概要を示す。1 フレーム (0.5 秒ずつ) の点群に対して、前処理としてクロップ、ダウンサンプリング、外れ値処理を行う。クロップは、柵・柵前の床・人の 3 つの部分に分けて行い、それぞれの領域でサンプリングレートを変えてサンプリングを行う。柵や床

の点群は、モデルに位置関係などをうまく学習させることを目的として使用し、それぞれの領域でサンプリングレートを変えるのは、その後の処理の高速化を目指しつつ人の点群は密な状態で特徴抽出を行い、より詳細な表現を得るためである。その後、外れ値処理を行い別々にした領域をあわせてもとに戻す。外れ値処理を行った後、モデルへ入力するために点群数を揃える必要がある。そのためここまでの処理で点群数が 8000 点より多い場合はランダムに点群を削除し、8000 点になるように調整する。上記の処理を行った結果、図 3 のような点群データが得られる。この点群をモデルへの入力とする。

モデルに関しては PointNet [15] を使用する。PointNet の選定理由として、点群データをそのまま入力に取ることができるため初期評価として使いやすい点と、室内で取られた比較的密な点群データに対して良い性能を示すモデルであるという点が挙げられる。モデル訓練時のパラメータはバッチサイズ 64、エポック数 100 設定した。損失関数には NLLoss、最適化アルゴリズムには Adam を使用し、学習率 0.001 としている。この学習率は 20 エポックごとに 50% ずつ減衰する。つまり、20 エポックで 0.0005、40 エポックで 0.00025 となる。これらの数値はエポック数以外は PointNet 側で初期値として使用されているものである。

このモデルを収集した学習データと評価データでの訓練・推定と、Leave One Person Out(LOPO)での訓練・推定によって評価する。この評価で、モデルの精度と予測対象の人物の特徴が精度に影響を与えるか否かについて調査する。

#### 4.2.3 結果と考察

学習の経過を図 8、図 9 に示す。図 8 は、予測対象のデータを含んだ学習による結果、図 9 は LOPO の結果である。本人のデータを含んで学習させたモデルは、エポックごとの変動が大きくなっているものの、エポックを重ねるごとに評価データへ適合していることが見て取れる。エポックごとの変動はタスクに対する学習率が大きいこと、学習・評価のためのデータが少ないことが原因と考えられる。精度としては最高でエポック平均 49% の正解率となっていた。12 クラスのクラス分類であることを考えると、この精度は非常に高いと考えられる。また、予測確率の高い 3 つの内にどれだけ正解ラベルが入っているかという調査を行ったところ、最もよい適合を示したエポックでは 80% という精度を得ることができた。まだまだモデルや前処理に改善の余地があった上で高精度を実現することができたことは、我々のアプローチの有効性を示すものとなったと考えられる。このように正確な視線情報ではなく、商品棚のようなある程度同カテゴリがまとまっている環境では、それらを見る姿勢から分類タスクに落とし込んで視線の情報を得ることができると考えられる。

LOPO の結果では、予測対象のデータを含んだ結果より

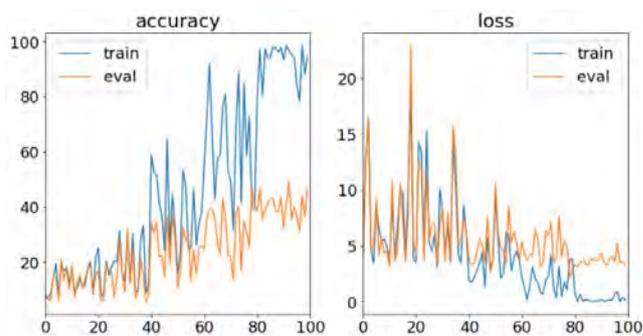


図 8 視線情報の可視化

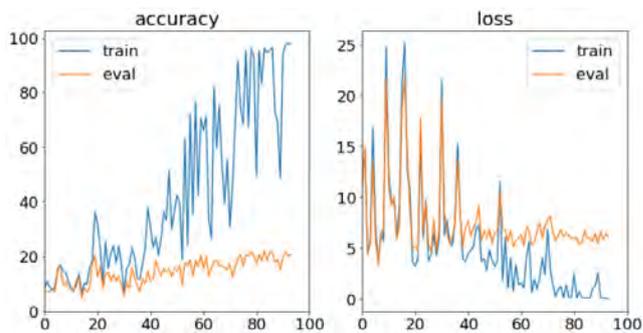


図 9 興味を示した商品の情報

精度が全体的に低くなっており、最も良い精度でも 22% だった。こちらでも精度、Loss の両方でエポックごとの変動が大きくなっている部分が見られるが、これも先述した原因と同じと考えられる。この結果から、この推論モデルの一般化にはまだまだ課題が残ることがわかる。精度が上がらない原因としては、データの不足がまず挙げられる。点群データは対象とする人物によっても大きく変わり、また棚を見る際の人の行動も身長などによって変化する。そういった個人の差に今回の推論モデルは対応できていないと考えられる。また、モデルに関しても PointNet は [15] にて、室内のセグメンテーションタスクや家具単位のクラス分類タスクで紹介されている。ここから同じ形状の人というクラスには分類できても、人の姿勢や頭の向きといったより細かい特徴を抽出し分類することは本来のモデルの用途とは違っている。そこで今後は、データ数を増やすことに加え、モデルをより局所的な特徴を抽出できるものに変更するなど、タスクに対して適したモデルにすることで精度向上を目指せすことができると考えられる。

## 5. おわりに

本稿では、実店舗での購買中に興味を持った購買候補商品の推測を目的とした、実験環境・システムの構築とその初期評価を行った。実験環境では、備え付けられたセンサやシステムによって、候補商品推定に必要と考えられた 4 つのデータを取得できる。Depth カメラによる視線データ、LiDAR による点群データにより、購買中の人の視線や行動を、RFID を用いたシステムによる購買商品データ、

購買後のアンケートによる候補商品データにより、行動や視線を用いた推定の目的変数となるデータを取得する。

それらのデータの有効性の評価を行うために、興味のある商品を候補商品とし、まずは視線情報と候補商品との関係性を分析した。その結果、視線と候補商品には深い関連性があることが示された。結果を踏まえて、次は制約の多いカメラではなく、比較的制約の少ないLiDARによる点群データを用いて、視線情報を推定することで、候補商品を推定できると考えられた。実際に推定を行ったところ、予測対象の人物のデータを含んだ状態では49%の精度で推定することができ、さらに正解ラベルが予測確率が上位3つの中に含まれる確率は80%と高精度であった。一方でLOPOの評価結果からは課題も見受けられた。データ不足やモデルの表現力、前処理方法など様々な箇所で、改善すべき点が多い。加えて、今回は点群データの特徴抽出をPointNetのネットワークでのみ行ったが、先行研究でも取得されていた滞在時間や商品を手に取る動作、移動時の加速度などのすでに重要だとわかっている特徴に関しては、手で抽出し明示的にモデルに与えることも精度向上に寄与すると考えられる。今後はこれらの課題を解決し精度向上を目指す。

また、今後の発展として、ユーザの関心に沿った購買行動変容を支援するべく、このシステムをベースに、購買者への適切な情報提示などの検討を進める予定である。具体的には、ゲーミフィケーションやナッジ理論に基づく情報システムに関する先行研究[16–18]を踏まえつつ、ユーザにとってより良い購買行動変容をもたらす為に、適切な介入タイミング、刺激する感覚モダリティ、提示すべき情報・フレームの検討、および、これらのパラメータを制御する際のアルゴリズムの設計開発を進めていく予定である。

#### 謝辞

本稿の研究を行うに際し、ご助言いただいたソニーグループ株式会社尾崎氏、安田氏に感謝の意を表す。また、本研究の一部は、JST さきがけ (JPMJPR21P7) の助成によって行われた。

#### 参考文献

- [1] 岩井将行, 森雅智, 徳田英幸. センサーノードを用いた商品の購買前注目度把握システム. Technical Report 18(2008-MBL-044), 慶應義塾大学大学院政策・メディア研究科, 慶應義塾大学大学院政策・メディア研究科, 慶應義塾大学環境情報学部, feb 2008.
- [2] Jee-Won Kang and Young Namkung. The information quality and source credibility matter in customers' evaluation toward food o2o commerce. *International Journal of Hospitality Management*, Vol. 78, pp. 189–198, 2019.
- [3] 杉山悠司ほか. 商品棚前における非計画購買者の行動特徴量の検討. 研究報告ユビキタスコンピューティングシステム (UBI), Vol. 2019, No. 10, pp. 1–8, 2019.
- [4] 荒木貴好ほか. 実店舗における商品購買時の迷い検出システムの構築. 情報処理学会第 71 回全国大会, 2009.

- [5] Yixuan Li, et al. Towards measuring and inferring user interest from gaze. *WWW*, p. 525–533.
- [6] 大輝真弓, 優吾中村, 裕貴松田, 智一松井, 慎也三崎, 慶一安本, 潤子野原. 香りナッジが実店舗の購買行動に及ぼす影響の調査. 行動変容と社会システム vol.08, 第 2022 巻, mar 2022.
- [7] Daiki Mayumi, Yugo Nakamura, Yuki Matsuda, Tomokazu Matsui, Shinya Misaki, Keiichi Yasumoto, and Junko Nohara. Aroma nudges: Exploring the effects on shopping behavior in a supermarket. In *International Workshop on Digital Nudging and Digital Persuasion (DNDP'22)*, pp. 1–11, 2022.
- [8] Hyunwoo Hwangbo, Yang Sok Kim, and Kyung Jin Cha. Recommendation system development for fashion retail e-commerce. *Electronic Commerce Research and Applications*, Vol. 28, pp. 94–101, 2018.
- [9] Congying Guan, Shengfeng Qin, Wessie Ling, and Guofu Ding. Apparel recommendation system evolution: an empirical review. *International Journal of Clothing Science and Technology*, 2016.
- [10] Soumy Jacob, Shoya Ishimaru, Syed Saqib Bukhari, and Andreas Dengel. Gaze-based interest detection on newspaper articles. In *Proceedings of the 7th Workshop on Pervasive Eye Tracking and Mobile Eye-Based Interaction*, PETMEI '18, New York, NY, USA, 2018. Association for Computing Machinery.
- [11] Jakob Karolus, Patrick Dabbert, and Pawel W. Wozniak. I know what you want: Using gaze metrics to predict personal interest. In *The 31st Annual ACM Symposium on User Interface Software and Technology Adjunct Proceedings*, UIST '18 Adjunct, p. 105–107, New York, NY, USA, 2018. Association for Computing Machinery.
- [12] A Selin Atalay, H Onur Bodur, and Dina Rasolofoaarison. Shining in the center: Central gaze cascade effect on product choice. *Journal of Consumer Research*, Vol. 39, No. 4, pp. 848–866, 2012.
- [13] Jesper Clement, Tore Kristensen, and Kjell Grønhaug. Understanding consumers' in-store visual perception: The influence of package design features on visual attention. *Journal of Retailing and Consumer Services*, Vol. 20, No. 2, pp. 234–239, 2013.
- [14] Xiaoye Liu. Airborne lidar for dem generation: some critical issues. *Progress in physical geography*, Vol. 32, No. 1, pp. 31–49, 2008.
- [15] Charles Ruizhongtai Qi, Hao Su, Kaichun Mo, and Leonidas J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. *CoRR*, Vol. abs/1612.00593, , 2016.
- [16] Ana Caraban, Evangelos Karapanos, Daniel Gonçalves, and Pedro Campos. 23 ways to nudge: A review of technology-mediated nudging in human-computer interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–15, 2019.
- [17] Yutaka Arakawa and Yuki Matsuda. Gamification mechanism for enhancing a participatory urban sensing: Survey and practical results. *Journal of Information Processing*, Vol. 24, No. 1, pp. 31–38, 2016.
- [18] Yugo Nakamura and Yuki Matsuda. Iot nudge: Iot data-driven nudging for health behavior change. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 51–53, 2021.