

戦略的創造研究推進事業 CREST  
研究領域「イノベーション創発に資する人工知能基  
盤技術の創出と統合化」  
研究課題「異種ドメインユーザの行動予測を可能に  
するペルソナモデルの転移技術」

## 研究終了報告書

研究期間 2018年 10月～2021年 3月

研究代表者:原 隆浩  
(大阪大学大学院情報科学研究科 教授)

## § 1 研究実施の概要

### (1) 実施概要

本研究では、異なるドメイン(サービス業者やデータ所有者)で構築されるユーザ(ペルソナ)モデルと行動予測技術の有効活用を目的とし、ドメイン間でペルソナモデル上の行動予測技術を転移利用するための技術群を開発した。本研究は、プライバシーや権益の問題を考慮して、IDなどの個人情報を用いずに、異種ドメイン間で同一・類似ユーザやユーザグループの同定や、行動予測技術の転移利用を実現するものであり、世界で初めての試みである。また、サービス運用時に収集可能なユーザのオンライン行動に加えて、通信・移動ログなどの実空間行動データ、ソーシャルメディア(SNS)データなどを活用し、既存技術では実現できないような細粒度なペルソナモデルを構築した。

本研究の全体像を図1に示す。以下に各研究項目の成果の概要を述べる。

### **研究項目1：多様なデータソースからのペルソナモデルの構築**

本研究項目では、オンラインの購買行動ログや、実空間における位置ログ・店舗訪問ログ、SNSやWebなどの多様なデータから、ペルソナモデルを構築した。さらに、構築したモデルに対して、状況に応じて効率的かつ効果的にデータ処理を行う機能群を提供した。

#### **1-1: ペルソナモデルの構築**

(A) オンライン行動に基づくモデリングとして、EコマースドメインとWeb広告ドメインのオンラインデータを用いたペルソナモデリングを行った。それぞれのドメインにおいて、ドメイン内でのユーザの行動モデルを構築し、行動予測の精度向上を軸にして、構築したペルソナモデルの改良を行った。(B) 空間行動に基づくモデリングでは、WiFiデータとビーコンデータを利用し、実空間(オフライン)行動の予測(特定店舗への訪問など)を軸に、これらのデータを用いたペルソナモデリングを行った。さらに、ペルソナモデルの生成につなげるため、滞在場所がどう使われるかの特徴に関する分散表現を人々の位置情報から生成し、場所の使われ方をもとにユーザの特徴を抽出する検討を行った。また、(C) 環境(外的要因)のモデリングとして、SNS(Twitter)から得られる情報が、Eコマースドメインにおける商材のセールの成否の予想に有効か否かを検証した。チーム全体で頻りに意見交換を行いつつ、(A)は大阪大学グループ、KDDI総研グループ、(B)は名古屋大学グループ、(C)は大阪大学グループを中心に実施した。

#### **1-2: ペルソナモデルに対する処理機能群の実現**

大阪大学グループを中心に、ペルソナモデルなど複雑なデータ構造に対するデータ処理機能群として、大規模な空間データを高速にクラスタリングしたり、時系列データ集合に対して類似集合をモニターしたりする効率的な新技術を開発した。これらの技術は、既存技術や単純なアプローチと比較して、数十倍から百倍程度の高速化を実現した。その技術的貢献が国際的に評価され、データベース分野で著名な国際会議などに論文が採択された。

### **研究項目2：複数ドメイン間の個人を特定しないペルソナマッピング**

大阪大学グループとKDDI総研グループで連携し、研究項目1で構築した複数ドメイン(EコマースとWeb広告)のペルソナモデルにおいて、同一ユーザをマッピングする技術を開発した。具体的には、数万人規模の許諾済みデータ(両ドメインでIDマッチングを行ったデータ)を教師情報として、同一ユーザのマッチングを行うために、正準相関分析(CCA)や深層学習に基づくものなど様々な機械学習技術を開発した。実データを用いた実験により、考案技術により、マッチング候補の上位に同一ユーザが登場することを確認した。

### **研究項目3：ドメイン間の行動予測の転移技術の構築**

あるドメインにおいて構築された行動予測技術を、別のドメインに転移するための技術開発を以下のように実施した。

#### **3-1: 異種ドメインの行動モデルを横断的に扱う統計モデルの構築**

大阪大学グループとKDDI総研グループが連携し、EコマースドメインとWeb広告ドメインにおいて、行動予測モデルを転移する技術の開発を行った。特に、異なるペルソナモデリングの方法が行動予測の転移に与える影響を検証するために、いくつかの高度なモデリ

ング技術を実装し、評価実験等を行った。さらに、ペルソナモデリングにおける特徴表現の転移を敵対的学習を用いて実行する際の学習安定化技術について検証を行った。

### 3-2: オンライン行動および実空間行動情報を用いた検証実験

名古屋大学グループを中心に、実空間における行動情報を用いてのペルソナモデルの構築に向けて、「ある場所がどのように使われているか」という滞在場所の特徴 (Location Usage (LU)) を定義し、抽出した。GPS の位置情報データをもとにユーザの場所ごとの滞在時間を計測し、各々の場所に対して平日・休日や時間帯ごとの滞在時間の多寡の情報を用いて分散表現を生成した。得た分散表現をクラスタリングし、時間帯や滞在時間で似た特徴を持つ場所を抽出することで、ユーザの移動特性に「一般にどのように使用されている場所を通過しているか」の情報を付加できる可能性を検証した。この場所の使い方の一般性と個々人の使い方の特徴とを比較することで、多種多様な応用が可能になり、GPS データ以外の位置情報にも転移利用が可能なペルソナモデルが構築できることが期待される。

### 研究項目 4: コロナ禍などの非常時におけるユーザ行動の変動分析 (新設定項目)

昨今の新型コロナの影響で、オンラインおよびオフライン (実空間) のユーザの行動パターンが大きく変動している。このような非常時の状況下では、常時に作成した行動予測モデルが有効に機能しない可能性が高い。その影響を分析するために、名古屋大学グループと KDDI 総研グループを中心に、準備的なデータ解析を実施した。その結果、オンラインと実空間の両方において、ユーザの行動が大きく変動し、その変動のパターンが非常に複雑であることを確認した。この分析結果は、非常時におけるドメインを跨ったペルソナモデリングや行動予測の可能性を検証し、今後の研究の方向性を検討するうえで非常に貴重な知見を与えた。

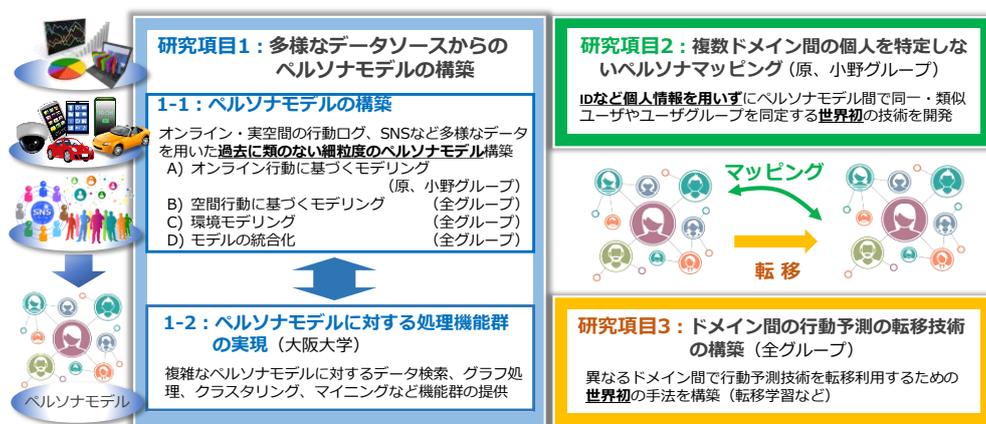


図 1: 本研究課題の全体像

## (2) 顕著な成果

<優れた基礎研究としての成果>

### 1. 複雑な構造をもつデータを効率的に処理する技術の確立

概要:

ペルソナモデルは多次元データや時系列データ、グラフデータなど非常に複雑な構造をもつため、複雑な構造をもつデータを効率的に処理するためのデータ処理技術を開発した。開発した技術は、従来技術の数十倍から数百倍の高速性を実現した。これらの成果は、データ管理分野で世界最高峰の国際会議 (IEEE ICDE など) に複数の論文が掲載されるなど、国際的に高い評価を得ている。

### 2. ドメインを跨るデータ解析による行動予測技術の確立

概要:

Web 広告業者(ドメイン)とE コマース業者(ドメイン)など、複数のドメインで収集・管理されているデータ(ユーザの行動履歴など)を統合的に解析・処理することで、ドメインを跨ってユーザの行動を予測する技術を確立した。従来のドメインを跨る行動予測の研究とは大きく異なり、本研究では特徴が大きく異なる複数のドメインを想定し、最新の人工知能技術を駆使して技術開発を行う点で技術的な重要性が高い。これらの成果は、産業情報処理に関する国際論文誌に採択された。

### 3. オンラインと実空間におけるペルソナモデリング技術の確立

概要:

Web やE コマースなどオンラインのドメインだけではなく、WiFi や位置情報などの実空間のデータを解析することでユーザのモデリング(ペルソナモデリング)を行う技術を開発した。その一例として、実空間の場所の位置(緯度経度)だけではなく意味を解釈し、ユーザの移動履歴を考慮したモデリングをより詳細に行う技術を開発した。この技術では、オンラインのドメインでのモデリングとの統合利用を可能とするために、モデリング結果をベクトルなどの分散表現として実現している。

< 科学技術イノベーションに大きく寄与する成果 >

#### 1. クロスドメインアプローチによる効果的なデジタルマーケティング技術の構築

概要:

ドメインを跨ってビッグデータを人工知能技術で解析することにより、従来技術よりもはるかに高い精度でユーザの行動予測を行い、デジタルマーケティングに大きく寄与する技術を開発した。その一例として、1500 商材からユーザが実際に購入した1商材を予測するタスクにおいて、10 商材の推薦で 26%以上の予測精度を達成した。この技術は、ユーザや商材をベクトルで表現しコサイン類似度に基づいて商材を推薦する一般的な手法と比較して、数十倍程度高い精度を実現した。この技術により、事業者は推薦(広告)のコストを大幅に削減できるとともに、ユーザは的を得ない大量の推薦に煩わされることなく、適切な商材の推薦を受けることが可能となる。

#### 2. 実空間の場所の意味を考慮したペルソナモデリング技術の構築

概要:

実空間におけるユーザの行動をモデリングするために、従来技術では訪問場所の時間的遷移のパターンを学習していたが、場所を単なる緯度経度の位置ではなく、その場所の意味(住宅街、ビジネス街、飲食街など)までを考慮してモデリングする技術を構築した。訪問人数・頻度、時刻、時間などの履歴情報をもとに、まず場所の意味をモデリングし、その後、各ユーザの行動履歴をもとに、ユーザの行動パターンをモデリングする。これにより、従来技術よりも詳細なペルソナモデリングが可能となり、ユーザの行動予測の精度が大幅に向上することが期待される。

#### 3. 新型コロナなど非常時におけるユーザの行動への影響の分析

概要:

本研究で開発するドメインをまたがる行動予測技術は、新型コロナ禍などの非常時におけるユーザの行動予測にも貢献できるのではないかと考え、準備的な分析を行った。その結果、新型コロナに関する非常事態宣言前後で、オンラインと実空間(オフライン)の両方でユーザの行動パターンに大きな変化があり、その変化が複雑であることを確認した。具体的な研究開発のためにはより詳細な検討が必要だが、単純な技術では到底対応できないことを確認できたことは、今後のイノベーションに向けた重要な一歩と考える。

< 代表的な論文 >

1. Daichi Amagata and Takahiro Hara, Identifying the Most Interactive Object in Spatial Databases, Proc. International Conference on Data Engineering, pages 1286-1297, 2019.

概要:

この論文では、位置情報付きデータの集合を一つのオブジェクトとして捉えて、最多数の他のオブジェクトとインタラクション可能(最近傍データが閾値以下のオブジェクト同士をインタラクション可能と定義)なオブジェクトを効率的に発見する技術を考案している。このような処理は複雑な構造をもつデータから、他のデータとの関連性が最も高いハブ的なデータ発見するものであり、ビッグデータ処理において重要な基礎技術の一つである。考案技術は、サイズの異なる二つのグリッドを構築し、各グリッドセルに各オブジェクトのデータが含まれるかどうかを示す 2 値の情報から構成されるインデックスを構築し、インタラクション可能なオブジェクト数の上限値と下限値を高速に計算する。これにより、解とはなりえない大多数のオブジェクトに対する詳細な計算を省くことができるため、処理を劇的に高速化できる。性能評価の結果、本論文の提案手法は既存技術と比較して、数十倍から数百倍の高速性を実現できることを確認した。

2. Daichi Amagata, Takahiro Hara, and Chuan Xiao, Dynamic Set kNN Self-Join, Proc. International Conference on Data Engineering, pages 818-829, 2019

概要:

この論文では、動的に更新し得る集合データが大量に与えられた際、それぞれの k 最近傍(最も距離が近い上位 k 個のデータ)を効率的にモニタリングするアルゴリズムを初めて提案している。このアルゴリズムは、商材推薦などでよく用いられる協調フィルタリング等に応用できる基幹技術である。実データを用いた性能評価の結果、提案アルゴリズムは既存技術を導入したベースラインと比較して常に大幅に高速であることを確認した。

3. Hanxin Wang, Daichi Amagata, Takuya Maekawa, Takahiro Hara, Niu Hao, Kei Yonekawa, and Mori Kurokawa, A DNN-based Cross-Domain Recommender System for Alleviating Cold-Start Problem in E-commerce, IEEE Open Journal of the Industrial Electronics Society, Vol. 1, pages 194-206, 2020.

概要:

この論文では、Web 広告ドメイン(業者)と E コマースドメイン(業者)の実データを用いて、E コマースドメインの商材推薦においてコールドスタート問題を解決する技術を考案している。E コマースでは、多くの商材が短期間に入れ替わり、多くのユーザは購買記録をもたない(商品情報を閲覧したりクーポンを得るだけ)という独特の特徴があるため、購買記録に基づく商材推薦がうまく機能しない。この問題を解決するために Web 広告ドメインのデータ(Web 閲覧既読)を用いてユーザをモデリングする技術を実現した。その結果、従来技術の数十倍の推薦制度を実現した。

## § 2 研究実施体制

### (1) 研究チームの体制について

#### ① 大阪大学グループ

研究代表者:原 隆浩 (大阪大学大学院情報科学研究科 教授)

研究項目

- ・ペルソナモデルの構築(オンライン行動に基づくモデリング)
- ・ペルソナモデルの構築(環境モデリング)
- ・ペルソナモデルの構築(統合)
- ・ペルソナモデルに対する処理機能群の実現
- ・データの特性に適応したペルソナマッピング
- ・ユーザグループのマッピング
- ・異種ドメインの行動モデルを横断的に扱う統計モデルの構築
- ・ペルソナ間社会的ネットワークの推定技術
- ・オンライン行動および実空間行動情報を用いた検証実験

#### ② 名古屋大学グループ

主たる共同研究者:河口 信夫 (名古屋大学未来社会創造機構 教授)

研究項目

- ・ペルソナモデルの構築(実空間行動に基づくモデリング)
- ・ペルソナモデルの構築(環境モデリング)
- ・ペルソナモデルの構築(統合)
- ・異種ドメインの行動モデルを横断的に扱う統計モデルの構築
- ・ペルソナ間社会的ネットワークの推定技術
- ・オンライン行動および実空間行動情報を用いた検証実験

#### ③ KDDI 総研グループ

主たる共同研究者:小野 智弘 (KDDI 総合研究所 部門長)

研究項目

- ・ペルソナモデルの構築(オンライン行動に基づくモデリング)
- ・ペルソナモデルの構築(環境モデリング)
- ・ペルソナモデルの構築(統合)
- ・データの特性に適応したペルソナマッピング
- ・ユーザグループのマッピング
- ・異種ドメインの行動モデルを横断的に扱う統計モデルの構築
- ・ペルソナ間社会的ネットワークの推定技術
- ・オンライン行動および実空間行動情報を用いた検証実験

### (2) 国内外の研究者や産業界等との連携によるネットワーク形成の状況について

KDDI 総合研究所が連携・出資している企業や他のいくつかの企業から、実サービスのデータの提供を受け、積極的な協力を得ている。本研究で開発したデジタルマーケティング技術やモデリング技術は、これらの協力企業に積極的に提供し、実サービスへの展開などに貢献できる体制を構築している。