

サービスパーソナライズの実現に向けたユーザ心理 属性推定に関する研究

石川, 雄一

<https://hdl.handle.net/2324/7182491>

出版情報 : Kyushu University, 2023, 博士 (工学), 課程博士
バージョン :
権利関係 :



サービスパーソナライズの実現に向けた ユーザ心理属性推定に関する研究

石川 雄一

2024 年 1 月

概要

ユーザの様々な心理属性のうち、性格、感情などの心理属性は、食品や衣服の好みなど特定の生活ドメインの行動にのみ影響を与えるドメイン固有の心理属性と異なり、衣食住や映画・音楽の視聴など幅広い生活ドメインの行動に影響を与えるドメイン非依存の心理属性である。こうした心理属性は、E コマース (EC) サイトや動画・音楽配信サイトにおける商品・コンテンツのレコメンドや、広告のターゲティングなど様々なサービスのパーソナライズに有効であることが知られている。これまでに、行動ログや生体ログなどの観測可能なユーザデータからユーザの性格、感情を推定する研究が盛んに行われてきた。しかし、従来研究には 1. 推定対象とする心理属性の Ground truth の選択と 2. 行動ログや生体ログからの特徴量の抽出に課題があった。

1. については、多くの従来研究で質問紙の測定結果を心理属性の Ground truth としてきたが、同じ心理属性（例えば性格の主要 5 因子、Big Five）を測定する複数の質問紙（Big Five では 20 種類以上存在）について、質問紙の違いによって、測定結果を用いたサービスパーソナライズの効果に差があるのか、その場合どの質問紙の測定結果を用いると効果が高いのか、が明らかにされてこなかった。また、似た性質を持つ心理属性（例えば、ともにドメイン非依存で長期間安定している性格と価値観）について、サービスパーソナライズにおいて独立した効果を持つのか明らかになっていなかった。

また、2. については、性格、感情のドメイン非依存性と Cross-modal 性が見落とされてきた。すなわち、これらの心理属性は衣食住にわたる様々な生活ドメインでのユーザの行動に影響することや（ドメイン非依存性）、脳波、心拍、皮膚電位など多様なモダリティの生体活動に影響を与えること（Cross-modal 性）が知られている。従来方式ではユーザの行動ログや生体ログから特徴量を抽出する際、ドメイン非依存または Cross-modal の成分と、特定のドメインもしくはモダリティにのみ影響を与えるドメイン/モダリティ特有の成分（例えば食べ物の好みや衣服の好み）を区別していなかった。このため、従来方式では性格、感情といった心理属性を正しく推定できず、推定結果を用いたサービスパーソナライズの効果は限定的であった。

本研究では上記の課題 1, 2 を解決し、行動ログ、生体ログを用いて、サービスのパーソ

ナライズ効果の高い心理属性を推定する方式を提案することを目指した。1. については、広告デザインのパーソナライズを想定し、質問数と質問方法が異なる 3 種類の Big Five の質問紙について、それぞれの測定結果を用いた際のパーソナライズの効果を比較した。また、ともにドメイン非依存で長期間安定している心理属性、すなわち性格と価値観について、質問紙測定結果のパーソナライズの効果を比較した。

2. については、複数の生活ドメインの行動ログから複数レイヤの RNN を用いてドメイン非依存の特徴量を抽出する新規な方式を提案した。提案方式により EC サイトの購買行動ログから抽出したユーザ特徴量が、どの程度性格を反映するのか、サービスのパーソナライズにどの程度有効かを評価した。また、提案方式を生体ログ（心拍、脳波等）が入力できるように拡張して、動画・音楽を視聴中の複数モダリティの生体ログから Cross-modal な特徴量を抽出できるようにし、提案方式により抽出した特徴量がどの程度感情を反映するのか評価した。

1. については質問紙によって広告デザインのパーソナライズの効果に有意な差があることが分かった。また、文章ではなく、形容詞単語による質問での測定結果を用いたパーソナライズが広告対象の商材によらず一貫して効果が高いことを確認した。さらに、性格と価値観が持つサービスパーソナライズの効果の一部は互いに独立していること、組み合わせて用いることでよりパーソナライズの効果が高まることを確認した。

2. については提案方式により EC サイトの購買行動ログから抽出したユーザ特徴量が従来方式と比べて、より正確に性格を反映すること、サービスパーソナライズの効果が高いことを確認した。また、生体ログから提案方式で抽出した特徴量は、従来方式に比べて、より正確に感情を反映することを確認した。

目次

第 1 章	序論	1
1.1	研究背景	1
1.2	研究の目的	2
1.3	本研究の貢献	4
1.4	本論文の構成	5
第 2 章	関連研究	7
2.1	質問紙を用いた心理属性の測定	7
2.2	心理属性と生活習慣, 生体活動との関係	17
2.3	心理属性を用いたサービスのパーソナライズ	23
第 3 章	心理属性・質問紙によるサービスパーソナライズ効果の違い	29
3.1	背景と目的	29
3.2	関連研究	30
3.3	データセット	31
3.4	方法	36
3.5	モデル比較結果	41
3.6	広告反応の予測精度評価	48
3.7	考察	56
3.8	結論	58
第 4 章	購買行動データを用いた心理属性の推定	61
4.1	背景と目的	61
4.2	関連研究	63
4.3	提案方式	65
4.4	実験	69
4.5	結果	78

4.6	考察	81
4.7	結論	83
第 5 章	生体データを用いた心理属性の推定	85
5.1	背景と目的	85
5.2	関連研究	87
5.3	提案方式	88
5.4	実験	92
5.5	結果と考察	97
5.6	結論	100
第 6 章	結論	101
6.1	本研究の貢献	101
6.2	実用上の意義	103
6.3	本研究の課題と今後の研究の方向性	104
参考文献		109

第 1 章

序論

1.1 研究背景

食べ物や衣服，映画，音楽の好みから，性格や価値観，さらには感情に至るまで人間の持つ心理的な属性は多岐にわたる．こうした様々な心理属性のうち，性格，価値観，感情などの心理属性は，食品や衣服の好みといった「食」や「衣」などの特定の生活ドメインの行動にのみ影響を与えるドメイン固有の心理属性と異なり，衣食住や映画・音楽の視聴など幅広い生活ドメインの行動に影響を与えるドメイン非依存の心理属性である．ドメイン非依存の心理属性を対象とする研究は心理学にとどまらず Human-computer interaction (HCI) の分野を中心に工学領域でも盛んに研究が行われ，その結果，E コマース (EC) サイトや，動画・音楽配信サイトにおける商品・コンテンツのレコメンドや，広告のターゲティング，カップルのマッチングなど様々なサービスのパーソナライズに有効であることが明らかにされてきた．

こうした研究が可能となった要因の一つに，心理学において心理属性を質問紙によって測定する手法（質問紙測定法）が確立されたことが挙げられる．直接観測できない潜在的な心理属性を，複数の質問項目に対する主観的な回答を集計することによって，型や数値などで表現する手法である．これにより，心理属性を統計モデルの中で変数として扱うことが可能となり，心理属性が，日常行動や生活習慣，学業や職務の成績，婚姻や就職，寿命などどのように関係するかを分析することが可能となった．質問紙測定法による心理属性の測定は，研究，ビジネス問わず幅広い目的で広く利用されている．

近年では，質問紙測定法を発展させ，一部のユーザを対象に質問紙測定法で測定した心理属性を Ground truth として，直接観測可能な行動ログや生体ログから心理属性を推定するモデルを構築し，これによって他のユーザの心理属性を推定する方式の研究が行われている．特に大規模なサービスでは全てのユーザに対して質問紙調査を実施することはコストの観点から非現実的であること，さらに動的に変化する感情を質問紙によってリアル

タイムに測定することは不可能であることから、研究が盛んに行われている。

しかし、従来研究には 1. 推定対象とする心理属性の Ground truth の選択と 2. 行動ログや生体ログからの特徴量の抽出に課題があった。1. については、同じ心理属性を測定する複数の質問紙について、質問紙の違いによって測定結果を用いたサービスパーソナライズの効果に差があるのか、その場合どの質問紙の測定結果を用いると効果が高いのか、が明らかにされてこなかった。性格については、性格が 5 つの因子^{*1}で構成されるとする主要 5 因子モデル（Big Five モデル）が今日広く受け入れられており、各因子のスコアを測定するための質問紙が、英語版に限っても 20 種類以上存在する。一方で、これらの質問紙のうち、どの質問紙の測定結果がサービスのパーソナライズ効果が高いのかは明らかになっていない。また、似た性質を持つ心理属性について、サービスパーソナライズにおいて独立した効果を持つのか明らかになっていなかった。性格と価値観はともにドメイン非依存であり、かつ感情とは異なり長期にわたって安定する心理属性である。サービスをパーソナライズする際に、これらの心理属性を両方とも用いることでパーソナライズの効果を高められるか否かは明らかになっていない。

2. については、性格、感情のドメイン非依存性と Cross-modal 性が見落とされてきた。性格と感情は、ともに衣食住にわたる様々な生活ドメインでのユーザの行動に影響を与えること（ドメイン非依存性）や、脳波、心拍、皮膚電位など多様なモダリティの生体活動に影響を与えること（Cross-modal 性）が知られている。従来方式では、ユーザの行動ログや生体ログから特徴量を抽出する際、ドメイン非依存の成分と特定ドメインにのみ影響を与えるドメイン特有の成分（例えば食べ物の好みや衣服の好み）を区別したり、Cross-modal な成分と特定の生体モダリティにのみ影響を与えるモダリティ特有の成分（例えば心拍にのみ影響する心室壁の厚さや心室・心房のサイズ）を区別したりしていなかった。このため、従来方式では性格、感情を正しく推定できず、推定結果を用いたサービスパーソナライズの効果は限定的であった。

1.2 研究の目的

本研究では、ドメイン非依存の心理属性を用いたサービスのパーソナライズにおいて、上記 1, 2 の課題を解決することで、より高いパーソナライズ効果を得ることを目指した。

図 1.1 に本研究で想定するサービスパーソナライズのフローと例を示す。サービスパーソナライズは、I. ユーザ反応予測と II. ユーザ心理属性推定の二つの処理で実現される。I. ユーザ反応予測では、ユーザのドメイン非依存の心理属性を基に、サービス候補を当該ユーザに提供した際に、所望のユーザ反応が得られる可能性を予測する。この予測を複数

^{*1} 知的好奇心、誠実性、外向性、協調性、神経症傾向の 5 因子

のサービス候補について実行し、最終的に所望のユーザ反応が得られる可能性が高いサービスをユーザへ提供する。例えば、バナー広告配信における広告デザインのパーソナライズでは、複数の広告デザイン（サービス候補）の中から、広告クリックや対象商品の購入（所望のユーザ反応）に至る可能性が最も高い広告デザインの配信を行う。また、II. ユーザ心理属性推定では、ユーザの行動ログや生体ログなど、観測可能なログから特徴量を抽出し、抽出した特徴量を用いて心理属性の推定を行う。

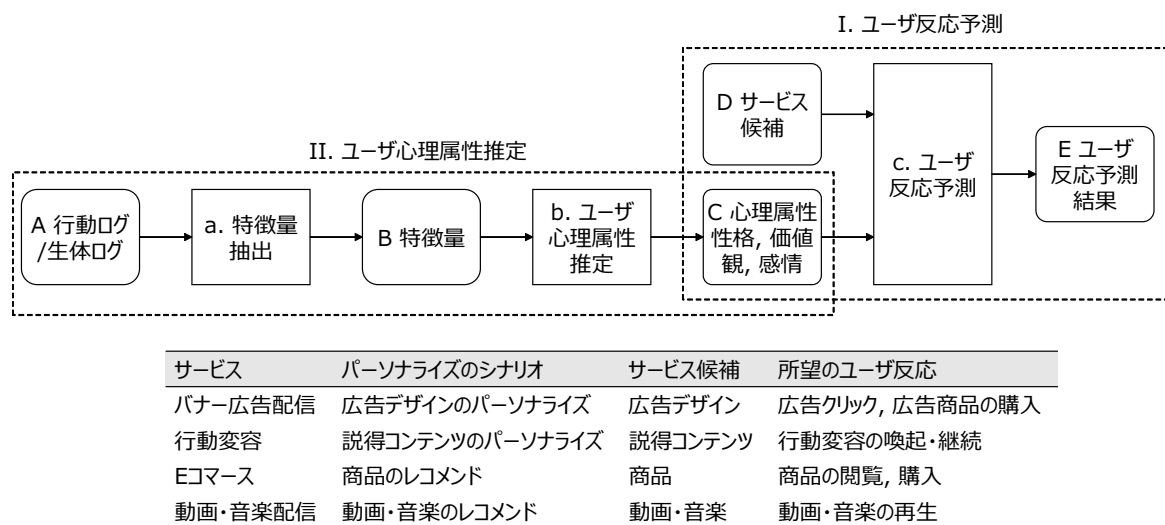


図 1.1 本研究で想定するサービスパーソナライズのフロー（上図）とサービスパーソナライズの例（下表）

こうしたフローを経てユーザへ提供するサービスを選択する際、ユーザから所望の反応を引き出すには（すなわち高いパーソナライズ効果を得るためには）、I においてはユーザ反応をより精度よく予測する必要がある、II においては心理属性をより精度よく推定する必要がある。

以上を踏まえ、本研究では I, II それぞれに関して以下を目的として設定した。

- I-1 複数の質問紙が存在する「性格」について、測定に用いる質問紙によってユーザ反応の予測精度が異なるかを明らかにする。
- I-2 (I-1 で予測精度が異なる場合) 高精度にユーザの反応を予測するには、「性格」の測定にどのような質問紙を用いるべきかを明らかにする。
- I-3 とともにドメイン非依存かつ長期安定の心理属性である「性格」と「価値観」は、ユーザ反応予測において独立した予測力を持つのかを明らかにする。
- II 観測可能な行動ログおよび生体ログから、性格及び感情を高精度に推定可能な特徴量を抽出する方式を考案する。

1.3 本研究の貢献

本研究の第一の貢献は、サービスパーソナライズのためにドメイン非依存の心理属性の推定を行う場合、高いパーソナライズ効果を得るためにはどのような Ground truth を用いるべきかを明らかにしたことである。具体的には性格と価値観に基づいたパーソナライズについて比較的多くの研究が行われている広告デザインのパーソナライズについて実験を行い、以下を示した。

- I-A 人間の性格を表現するモデルとして研究やビジネスにおいて最も広く利用されている Big Five を対象に、複数の質問紙の測定結果を比較し、質問紙によってユーザ反応の予測精度が有意に異なることを明らかにした。
- I-B 広告デザインのパーソナライズにおいて、ユーザ反応の予測精度を高めるためには、多数の質問項目を含み、かつ、文章ではなく形容詞単語で質問項目を表現する質問紙を用いることが望ましいことを明らかにした。
- I-C 性格と価値観は広告デザインに対するユーザ反応について、互いに独立した予測力を持つ成分を含み、共に用いることでそれぞれ単独で用いるよりも、より精度良くユーザ反応を予測できることを明らかにした。

本研究の第二の貢献は、性格、感情を推定するために、行動ログからはドメイン非依存な成分のみを、生体ログからは Cross-modal な成分のみを特徴量として抽出する方式を新規に考案し、実世界のデータでその有効性を検証したことである。具体的な貢献は以下のとおりである。

- II-A 複数の生活ドメインの商品を扱う EC ポータルにおける時系列の購買行動ログを入力とし、そこから複数の生活ドメインに共通して現れる成分（すなわちドメイン非依存の成分）のみを特徴量として抽出する多層 Recurrent Neural Network (RNN) を新規に考案した。
- II-B 実際の EC ポータルの購買行動ログから提案方式を用いて特徴量を抽出し、提案方式で抽出した特徴量が、従来方式で抽出した特徴量よりも性格をより強く反映していることを確認した。加えて、提案方式の特徴量が従来方式の特徴量に比べ、学習元とは異なるドメインの購買行動を精度良く予測できることを確認した。
- II-C 上記の提案方式をベースに、脳波や心電図、皮膚電位など複数の生体モダリティの生体ログから、Cross-modal な成分のみを特徴量として抽出する方式を提案し、提案方式により抽出した特徴量が、従来方式と比べて精度よく感情を推定できることを確認した。

1.4 本論文の構成

本論文の構成は次のとおりである。まず、2章では本研究の関連研究について整理する。心理属性の測定方法として科学研究で広く用いられている質問紙測定法について、その成立背景と代表的な手法を説明する。また、性格を中心にドメイン非依存の心理属性と行動習慣や生体活動との関係について既存研究の知見を整理し、性格や価値観、感情といったドメイン非依存の心理属性が、様々な生活ドメインにおける行動習慣や、生体モダリティにおける生体活動に影響を与えることを示す。さらに、ドメイン非依存の心理属性を用いたサービスパーソナライズに関する既存研究を紹介する。なお、ドメイン非依存の心理属性を用いた広告デザインのパーソナライズや、行動ログや生体ログを用いたドメイン非依存の心理属性の推定については、3章以降の各章（「関連研究」の節）で、関連研究を紹介し、課題を整理する。

3～5章では、心理属性を用いたサービスパーソナライズにおける心理属性の Ground truth の選択と、心理属性推定のための行動ログ及び生体ログからの特徴量抽出について検討する。図 1.2 に 3～5 章のサービスパーソナライズにおける位置づけを示す。

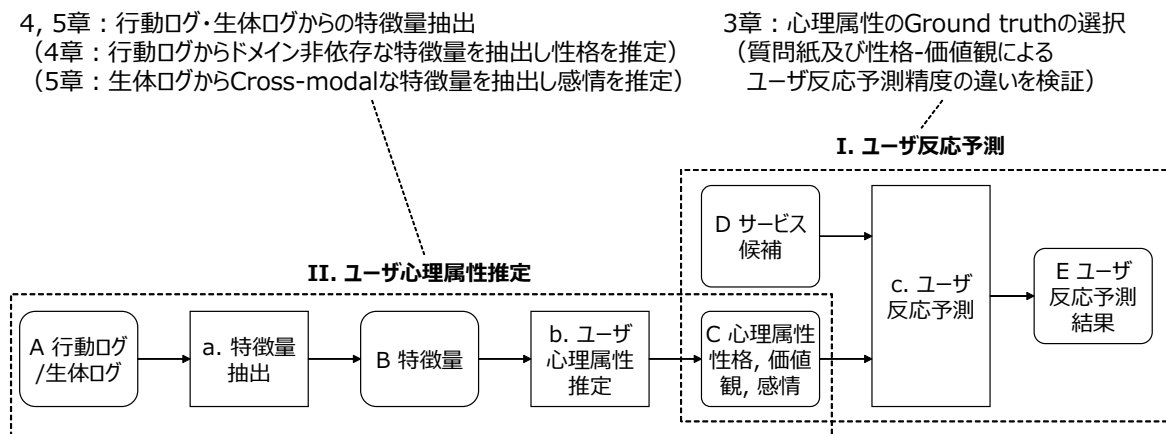


図 1.2 3～5 章の位置づけ

3章では、心理属性の Ground truth の選択にフォーカスし、ドメイン非依存の心理属性に基づくパーソナライズについて、比較的多くの研究が行われている広告デザインのパーソナライズを対象に、測定に用いる質問紙や測定対象とする心理属性によるユーザ反応の予測精度の違い（すなわちサービスパーソナライズの効果の違い）を検証する。

4章と5章では、性格と感情の推定に向け、行動ログ、生体ログからドメイン非依存、Cross-modal な成分のみを特徴量として抽出する多層 RNN を提案し、その実験結果を報告する。4章では、ユースケースとして EC ポータルにおけるサービスのパーソナライズ

を想定し，ユーザの購買行動ログを用いて提案方式で抽出した特徴量が性格をどの程度反映していたか，購買行動をどの程度予測可能だったかについて報告する．

5 章では，ユースケースとしてウェアラブルデバイスを用いた音楽や動画の視聴を想定し，視聴中の脳波や心拍といった生体ログを用いて提案方式で抽出した特徴量がどの程度正確に感情を推定できたかについて報告する．

最後に 6 章で，本研究で得られた知見と課題を整理し，将来の研究発展の方向性について述べる．

第 2 章

関連研究

2.1 質問紙を用いた心理属性の測定

性格や価値観、感情などの心理属性は直接観測できない潜在的なものであり、これらを科学研究の対象とするためには、型（タイプ）や数値などの変数として表現するモデル（表現モデル）が必要となる。また、変数化するための測定方法も必要となる。以下では、性格、価値観、感情を対象に、代表的な表現モデルについて説明するとともに、今日広く利用されている質問紙を用いた測定法（質問紙測定法）について説明する。

2.1.1 性格の測定

人の性格を如何に表現するか、測定するかは長い間研究の対象として扱われており、今日でもなお多くの研究が続けられている。特に性格の表現モデルについては過去多くの研究者から様々なモデルが提案されてきたが、研究者の間で意見の一致を見ることなく、各々の研究者が好みの表現モデルを利用する状態が長い間続いた。この状況は、20 世紀後半に提案された「Big Five」[46] によって解消し、ようやく研究者間で一定の合意が取れる表現モデルが確立されるに至った [99]。性格の表現モデルは今日なお複数のモデルが提案され続けているが、Big Five は研究及びビジネスの双方で最も広く受け入れられている表現モデルと言える。以下で、Big Five の概要と成立背景、測定方法を説明した後、Big Five 以外の表現モデルについて説明する。

Big Five

Big Five は、人間の性格が図 2.1 に示す 5 つの因子（特性、Trait と呼ばれる）によって構成されるとする性格の表現モデルである。質問紙の回答結果を集計することで各特性のスコアが得られ、人間の性格が 5 次元のベクトルで表現される。

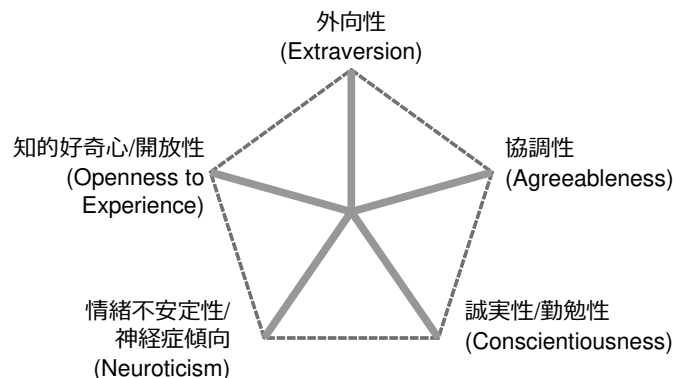


図 2.1 Big Five

以下，Nettle の著書 [99] および小塩の著書 [160] を参照し，各特性の概要を説明する．

■**外向性** 報酬または快の刺激に対する情動反応の強さを示す．例えば，仲間や興奮，出世，賛美，ロマンスと言ったものが報酬や快の刺激として挙げられ，外向性スコアが高い場合はこれらの刺激をとっても魅力的に感じる（強く反応する）のに対し，低い場合はこれらの刺激に対してあまり魅力を感じない（反応が低い）．

外向性は人付き合いや仕事・プライベートにおける行動，日常感じる感情などとの関係が強く，人付き合いにおいてスコアが高い人は低い人に比べて，社交に多くの時間を費やす，話好き，関心の的になることを好む，やすやすと新しい人間関係を形成する，などの傾向にある．また，仕事・プライベートでは，ステータスを手に入れることに喜びを見出し，名声や富を得るために猛烈に働く，活発なスポーツを好むなどの傾向がある．また，日常生活ではポジティブな情動（喜び，欲望，熱中，興奮など）を多く持つ傾向にある．

■**情緒不安定性/神経症傾向** 外向性とは逆に脅威への反応の強さを示す．本特性のスコアが高い人は低い人に比べて，脅威への反応として恐怖や不安，嫌悪，恥や罪悪感，悲哀といったネガティブな感情を持つことが多い．これらの感情は脅威を前にしたとき，自身を守るために喚起されるものであり，例えば「恐怖」の場合は潜在的な危険を警戒し，恐れのもととなった事柄に用心するなどの反応が引き起こされるが，それに伴いスコアが高い人は日常生活の苦労や厄介事について，スコアの低い人よりも強い影響を受けることになる．

結果として，本特性のスコアは鬱や不安障害，恐怖症，摂食障害，強迫性障害などと密接に関係しており，スコアの高い人は低い人に比べて，少ない量のストレスでこれらの疾患を抱える可能性が高い．

■**誠実性/勤勉性** 目先の報酬に対する自分の欲求を抑制し，目標や先送りされた満足を追求しようとする傾向を示す．Big Five の 5 つの特性のうち，学業成績や職務評価と最

も相関が高いのが本特性であり、特に職務評価については、専門職、管理職、営業職、警察、熟練工など職種を問わず、熟達度評価や昇進スピード、収入などと正の相関を持つことが知られている。

一方で、ギャンブルやアルコール、薬物（コカインやマリファナ）への依存、さらには反社会的行動を取るリスクは本特性のスコアと負の相関にあることが知られている。

■**協調性** メンタライジングと共感の高さを示す。メンタライジングとは他者の信念や欲望などの心的状態を類推したり、理解したりする能力を指し、共感とは他者と喜怒哀楽の感情を共有することを指す。例えば他人がおかれている状況や表情、声色などから「つらい思いをしているのだ」ということを理解するのがメンタライジングであり、その場合に自分も相手と同様につらい感情を持つのが共感である。

本特性のスコアが高い人は、他人と争ったり他人を侮辱したりすることは稀であり、何かあってもすぐ許し、実際に相手が悪くてもあまり怒ることはない。他人からは協力的で信頼でき、共感性があるという印象を持たれる。一方でスコアが低い場合には、他人から冷淡で敵意があり、不服従という印象を持たれ、実際に他の人を信頼したり助けたりせず、冷淡だったり敵対的になる傾向が強いとされる。

■**知的好奇心** 文化的、芸術的活動にどれほどかかわっているかを強く予測する。本特性のスコアが高い人は低い人に比べて、霊的で超自然的な信念への傾倒や規範への反抗、異常体験などの傾向が強いとされる。また、本特性のスコアと拡散的思考課題^{*1}の得点は正の相関にあることが知られている。このことから本特性は、思考の拡散や心の連想の広がり、すなわち一見何の関係もないように思えることを、意味を持たせて結びつける傾向を示すと解釈されている。

Big Five 以前の性格表現モデル

性格の表現モデルは「類型論」に基づくものと「特性論」に基づくものに大別されるが、Big Five は「特性論」に基づいて考案された。一方で Big Five が登場するまで、多くの性格の表現モデルは類型論に基づいていた。類型論は、人の性格はいくつかの類型に分類できるとする考え方であり、この考えに基づく性格表現モデルは、古いもので紀元前 100 年頃にギリシャの医学者ガレノスが提唱した四気質説にまでさかのぼる。ガレノスは同じくギリシャの医学者であるヒポクラテスの四体液説^{*2}にヒントを得て、人間の性格が四体液のいずれが優勢かによって、多血質（血液が優勢、快活、明朗、社交的など）、胆汁質

^{*1} 無関係な複数の名詞を共通の名詞で結びつける課題（例えば widow, bite, monkey の場合 spider）や日常使われているものについてできるだけ多くの用途を考えだす課題、無関係な単語を見せ、各単語の意味がどのくらい近いか評価させる課題など

^{*2} 人間の身体は血液、黄胆汁、黒胆汁、粘液の四種類の体液で構成されるとする説

(黄胆汁が優勢、せっかち、短気、積極的など)、憂鬱質(黒胆汁が優勢、用心深い、消極的、敏感、など)、粘液質(粘液が優勢、冷静、勤勉、冷淡、など)の四つの類型に分類できるとした。

近世以降、四気質説が支持されることはほぼなくなったが、20世紀になってからも多くの研究者は類型論に基づいた表現モデルを提案していた。例えばドイツの精神科医、クレッチマー(1888-1964)は精神科の入院患者の体格調査の結果を踏まえ、人の性格が体格に基づいて細長型(分裂気質・静かで控えめかつ真面目)、闘士型(粘着気質・几帳面でかつ熱中しやすく頑固)、肥満型(躁鬱気質・社交的で親切)に分類できると主張し、一方ドイツの教育学者、シュプリング(1882-1963)は生活の中でどの領域にもっとも価値を置くかで人の性格を次の六つに分類できると主張した: 権力志向、経済志向、社会志向、審美志向、理論志向、宗教志向。また、分析心理学を創始したことで有名なスイスの精神科医ユング(1875-1961)は、心のエネルギー(リビドー)が個人の外部に向かうか、内部に向かうかで人間の性格が外向型(外部の刺激に影響されやすい)と内向型(自分自身のことに関心が集中する)に分類できるとした。そのほか、最近のものでは、1962年に提案されたMBTI(Myers-Briggs Type Indicator)が注目を集めた。MBTIは人の性格を四つの気質それぞれで二分することで(1. 内向的 [I]/外向的 [E], 2. 感覚的 [S]/直観的 [N], 3. 思考的 [T]/感情的 [F], 4. 判断的 [J]/認知的 [P]), 16の類型で表現する(例えば、ISTJやENFPなど)。

このように、今日に至るまで類型論に基づいて多数の表現モデルが提案されてきたものの、筆者の知る限り、少なくとも科学研究の領域で、Big Fiveのように多くの研究者からの支持を集めている類型論ベースの表現モデルは存在しない。

Big Fiveの成立経緯

Big Five成立のきっかけは「語彙仮説」にあるとされている。すなわち、人間の性格に関する表現は普段使用している言葉の中に反映されており、性格の重要な表現はそれぞれの単語の中に組み込まれている、とする仮説である。この仮説にしたがって、イギリスの人類学者ゴールトン(1822-1911)やアメリカの心理学者オールポート(1897-1967)は、辞書の中にある性格を表現する単語(性格表現単語)をくまなくリストアップし、ゴールトンは約1,000語、オールポートは約4,500語の単語が存在することを見出した。

これらの調査をベースとして、性格表現単語を、アンケートを利用して分類する研究が複数の研究者によって行われた。すなわち、いくつかの性格表現単語をピックアップして「それぞれの単語はあなた(や周りの人)にどのくらい当てはまりますか」と質問した。このアンケートを多くの被験者に対して実施し、結果を因子分析によって分析したところ5つの因子(すなわち Big Five)が見出された。特に注目すべき点は、以下の2点である:

- 複数の研究者で実施した調査・分析（すなわちアンケートで利用した性格表現単語が異なっていたり，質問方法が異なっていたりする）の結果で，概念の似た 5 つの因子が見出された．
- 言語や文化が異なる 50 以上の国・地域で調査・分析を実施した結果，似た概念の 5 因子が見出された．

こうした結果から，性格研究に携わる研究者の注目を受けるに至り，質問紙による測定法が開発され，測定結果の信頼性や日常行動，生活習慣，生体活動などとの関連について研究が行われることとなった．信頼性の観点では，スコアが長期にわたって安定していること（同じ被験者に対して 6 年間隔で 3 回にわたって実施した質問紙調査の結果は相関係数が 0.68~0.85 であったとする報告がある），また行動・生活習慣との相関では，「外向性のスコアが高い人は実際に良く話す」，「神経症傾向の高い人は不快な事柄について見たり考えたりすると動揺しやすい」，「協調性の高い人に物語を聞かせると低い人よりも登場人物の心の状態に多くの注意を払う」など，実際の行動との相関が見出されること，などが複数の研究で確認された．こうした経緯から Big Five は，今日では科学研究において最も広く利用される性格表現モデルとなった．

Big Five の測定方法

Big Five の各特性のスコアは質問紙調査によって測定されることが一般的であり，これまでに英語のものに限っても 20 以上の質問紙が開発されている．これらの質問紙は，質問項目の数や表現方法（形容詞単語での質問か，文章形式の質問か），回答方法などが異なる．表 2.1.1 に英語と日本の質問紙のうち代表的なものを列挙する（引用件数が多いものを抽出した）．

質問紙は一つの特性につき 2 問（TIPI, TIPI-J）から 48 問（NEO-PI-R, 日本語版 NEO-PI-R）の質問で構成されており，それぞれの質問に対して n 件法または「はい/いいえ」により回答した結果を集計して特性毎にスコアを求める．また，質問紙によっては 5 つの特性のさらに下位の構成要素毎に集計できる．例えば NEO-PI-R（質問項目数は 240 問）では各特性が 6 つの“Facet”で構成されると想定し（例えば外向性は 1. 暖かさ，2. 群居性，3. 断行性，4. 活動性，5. 刺激希求性，6. よい感情），各 Facet のスコアを計算できる．また，各特性が 2 つの“Aspect”で構成されるとする BFAS（100 問）でも，Aspect ごとのスコアを計算できる．ただし，こうした下位の構成要素のスコアが計算できるのは多数の質問項目で構成される質問紙に限られ，例えば，TIPI（10 問）や Big Five 尺度短縮版（29 問）では下位の構成要素を計算できない．

Big Five の質問紙は「信頼性」(Reliability) と「妥当性」(Validity) が検証されることで，測定結果の品質が担保されている．信頼性は測定結果の値がどの程度一貫している

かを示す指標であり、内的一貫性、再検査信頼性、評定者間信頼性などで構成される。内的一貫性とは、一つの特性を複数の質問項目で測定する場合、それぞれの項目が同一の項目を測定しようとしているかを示す程度であり、その検証には通常クーロンバックの α 係数が用いられ、 α が大きいほど一貫性が高いとされる。また、再検査信頼性は、一定の期間をおいた再検査で前回とどの程度近い測定結果が得られるか、異なる評定者から同一の人物についてどの程度近い測定結果が得られるかを示す。一方で妥当性は、質問紙が測定したいものをどの程度適切に測定できているかを示す指標である。妥当性の検証については質問紙ごとに様々な方法が採られる。例えば、既存の他の質問紙の測定結果と比較するなどして妥当性が検証されるケースが多い。

表 2.1 代表的な Big Five の質問紙の一覧

言語	名称	発表年	項目数	質問方法	回答法
英語	NEO-PI-R ^a [21]	1985	240	文章	5 件法
	NEO-FFI ^b [22]	1989	60	文章	5 件法
	BFI ^c [60, 7, 61]	1991	44	文章	5 件法
	Big Five Mini-Markers [121]	1994	40	単語	9 件法
	TIPI ^d [48]	2003	10	単語	7 件法
	BFAS ^e [30]	2007	100	文章	5 件法
	BFI-2 [133]	2017	60	文章	5 件法
日本語	BFS ^f [163]	1996	60	単語	7 件法
	主要 5 因子性格検査 [161]	1997	70	文章	はい/いいえ
	FFPQ ^g [38]	1998	150	文章	5 件法
	日本語 NEO-PI-R [156]	1999	240	文章	5 件法
	日本語版 TIPI [159, 104]	2012	10	文章	7 件法
	Big Five 尺度短縮版 [162]	2012	29	単語	5 件法

^a Neuroticism, Extraversion, Openness-Personality Inventory-Revised^b Neuroticism, Extraversion, Openness-Five Factor Inventory^c Big Five Inventory^d Ten-Item Personality Inventory^e Big Five Aspects Scales^f Big Five Scales^g Five-Factor Personality Questionnaire

Big Five 以外の性格表現モデル

Big Five は人の性格を包括的に表現するモデルとして既に多くの研究者に受け入れられているが、Big Five 以外の性格表現モデルを探索する研究も続けられている。一つの方向性は Big Five 以外の特性の存在を探索する研究であり、もう一つの方向性としては Big Five のさらに上位の構成要素を探索する研究である。

前者の代表例として Ashton らが提案した HEXACO モデルが挙げられる [4]。本モデルは Big Five の 5 特性に、正直さ、謙虚さを示す特性 (Honesty) を加えた 6 つの特性で構成される。Big Five と同様に語彙仮説を元に開発され、著者らは 7 つの言語で共通する 6 つの特性が見出されたと主張している。HEXACO モデルが提案された背景には、Big Five では人間の持つ悪の気質 (精神病質^{*3}やマキャベリズム^{*4}、ナルシシズム^{*5}) を十分表現しきれないとする課題意識がある。実際に悪の気質の表現モデルであるダークトライアド [106] のスコアと Big Five のスコアは相関しないとする報告もある [106]。一方で、HEXACO モデルで新たに追加された正直さ、謙虚さの特性はダークトライアドと有意な負の相関にあることが確認されている。

Big Five の上位の構成要素を提唱するモデルの代表例としては Deyong らによる二因子モデル [29] と Musek による性格の g 因子モデル [95, 96] が挙げられる。両モデルとも、Big Five の特性同士に弱いものの有意な相関が認められることを起点に提案された。二因子モデルは、Big Five の上位に安定性 (Stability) と可塑性 (Plasticity) の二つの上位因子があるとし、安定性は神経症傾向、協調性、誠実性により、可塑性は外向性と知的好奇心により、それぞれ構成されたとした。また、g 因子モデルは、Big Five の全特性には相関があり、究極的には 1 つの因子 (g 因子) で表現できるとするモデルである。g 因子のスコアは外向性、協調性、誠実性、知的好奇心の各スコアと、神経症傾向を反転させたスコアを合計することで得られ、高いほど社会的に望ましく (飽くまで現代社会においてではあるが)、また主観的幸福感と正の、犯罪率やうつ病の罹患率と負の相関にあると言われている [158]。g 因子モデルを批判する研究も複数存在するものの [94, 25, 57]、行動遺伝学分野ではこれを支持する報告もある [158]。

2.1.2 価値観の測定

性格と並びドメイン非依存かつ長期間安定する心理的構成要素として価値観が挙げられる。性格と価値観が決定的に異なるのは、性格がその人間が時間や状況を通してどのよう

^{*3} 継続的な反社会的行動や衝動性、利己主義的な傾向を示す

^{*4} 対人操作や搾取を行う傾向や道徳観の無視、自己中心や欺瞞の行動を取る傾向を示す

^{*5} 誇大性や自尊心などの強さや共感の欠如を示す

に感じ、考え、振る舞うかをコントロールするのに対して、価値観は、その人が何を肯定的あるいは否定的にとらえるかをコントロールする点である。例えば、勤勉であること、周囲と協調的であること、は性格と価値観のどちらにも当てはまる特質だが、勤勉な人が勤勉であることを肯定的にとらえている（重要だと考えている）とは限らないし、協調性の高い人だからといって他人と協調的であることを肯定的にとらえているとも限らない。このように性格と価値観は少なくとも概念的には独立した構成要素である。

こうした背景から、価値観は性格とは別個に表現モデルと質問紙が開発されてきた。代表的なモデルに Rokeach の価値観モデル [119] と Schwartz の価値観モデル [122, 124] がある。共に人間の価値観が複数の次元で構成され、個人の価値観が各次元のスコアの高低で表現できるとするモデルである（特性論に基づく性格表現モデルと類似する）。以下では国連の World Value Survey^{*6}で利用され、サービスのパーソナライズにおいて複数の活用事例が見られる Schwartz の価値観モデルについて説明する。

Schwartz の価値観モデル

Schwartz の価値観モデルを構成する 10 の価値観次元を以下に示す。なお、各次元の日本語の定訳は無いため早稲田大学小塩教授の訳^{*7}を引用した。

- 自己決定 (Self-Direction)** 思考と行動の独立性（創造性、自由、独立心、好奇心）を求めること
- 刺激 (Stimulation)** 刺激的な経験（大胆さ、刺激的な人生）を求めること
- 快楽 (Hedonism)** 官能的な快楽（楽しさ、人生を楽しむこと）を求めること
- 達成 (Achievement)** 社会的に認められた成功（野心、能力）を求めること
- 権力 (Power)** 人や資源を管理し、金銭（社会的勢力、富、権威）を求めること
- 安全 (Security)** 自分、家族、国家の安全や安心（家族の安全、社会秩序、清潔さ）を求めること
- 調和 (Conformity)** 他人の期待に応えるために自らの衝動をコントロールすること（自己規律、服従）を求めること
- 伝統 (Tradition)** 伝統を守ること（節度を守ること、伝統を重んじること、信心深さ）を求めること
- 慈悲 (Benevolence)** 周りの人々の繁栄や幸福（親切、忠誠、正直、寛容）を求めること
- 普遍主義 (Universalism)** 人と自然の繁栄や幸福（平等、社会的正義、環境保護）を求めること

^{*6} <https://www.worldvaluessurvey.org/>

^{*7} <https://note.com/atnote/n/n0fc3dcc1e517#ea5ba2a2-4860-4c67-958a-c1311ce35a7a> 参照

これらの価値観次元の関係は図 2.2 に示す円環モデルで表現される。円環モデル上で隣り合う次元同士は概念が近いのに対して、正対する次元は概念的に対立する。10 の次元はその円環モデル上の近さにより、変化への開放性（Openness to change, 自己決定, 刺激, 快楽）、自己高揚（Self-enhancement, 快楽, 達成, 権力）、保守主義（Conservatism, 安全, 調和, 伝統）、自己超越（Self-transcendence, 慈悲, 普遍主義）の 4 つの高位の次元にまとめられる。

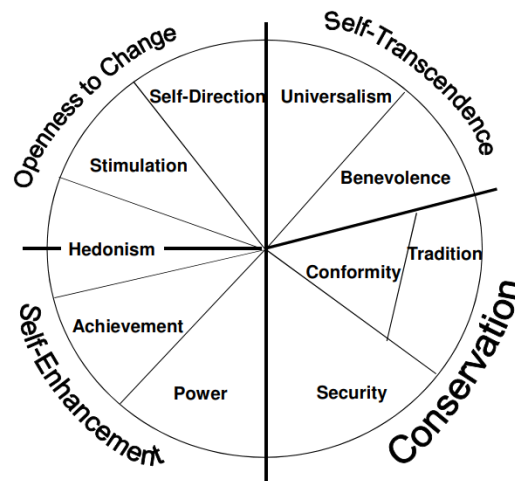


図 2.2 シュワルツの価値観 [124, p.9] より引用

Schwartz の価値観を測定する質問紙としては Schwartz Value Survey (SVS, 全 30 問) [122] およびその改訂版（全 26 問）と SVS を簡略化し子供から大人まで回答できるようにした Portrait Values Questionnaire (PVQ, 全 21 問) [123] が存在する。筆者の知る限り SVS, PVQ とともに日本語版開発に関する研究論文は存在せず, PVQ について広岡ら [157] の日本語訳（信頼性や妥当性の検証等を行われていない）が存在するのみである。

2.1.3 感情の測定

感情の表現モデルは性格と同様に感情を離散的なカテゴリとして表現する方法と複数次元の連続値のスコアで表現する方法に大別できる。前者の代表例としては Ekman の基本 6 感情 [35] が挙げられる。すなわち人間の感情は怒り, 嫌悪, 恐怖, 喜び, 悲しみ, 驚きのいずれかにカテゴライズされるとするモデルである。他方, 後者の代表例としては Russell の円環モデル [120] と PANAS (Positive and Negative Affect Schedule) [148] が挙げられる。

Russell の円環モデルは, 図 2.3 に示すように, 人間の感情が感情価 (Valence, 感情のポジティブ [+]/ネガティブ [-] の度合いを表す) と覚醒度 (Arousal, 覚醒度の強さを表す) の二次元により表現できるとするモデルである。本モデルに人間が感情をどの程度コ

ントールできているか/感情にどの程度支配されているかを表す支配度 (Dominance) を加えた PAD (Pleasure, Arousal, and Dominance) モデルも提案されている [85].

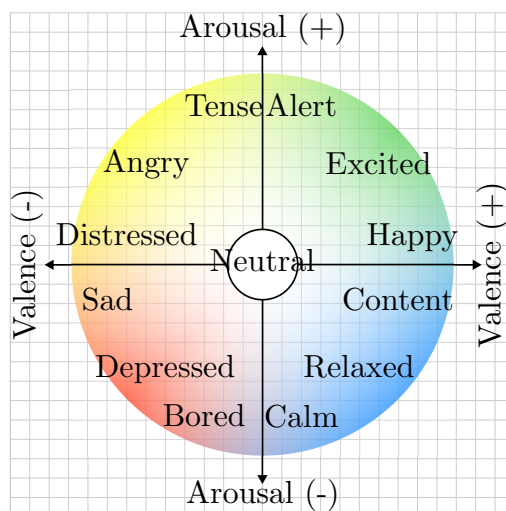


図 2.3 Russell の円環モデル (Wikimedia Commons より引用)

Russell の円環モデル, PAD モデルに基づいた感情の測定を行う場合, Self-Assessment Manikins (SAM) [92] を用いる研究事例が多くみられる (例えば, [87, 68]). SAM の例を図 2.4 に示す. 上の 3 段が上段から覚醒度, 感情価, 支配度を測定する領域であり, 人型の図を示すことで回答者が選択項目を視覚的に理解できる工夫が施されている. 回答者は各段の下部に設けられたスライダーを左右に移動させることで, 自身の感情を回答する. なお, スライダーではなく, 自身の感情に最も近い人型の図を選択させるケースもある (この場合各次元 1~5 のスコアとなる).

他方, PANAS は人間の感情をポジティブ感情 (Positive Affect) とネガティブ感情 (Negative Affect) の二次元で表現するモデルであり, 測定のための質問紙が開発されている. 質問紙では各次元 10 個ずつの感情を表す単語が列挙され (例えばポジティブ感情では, やる気がわいた-Inspired, 活気のある-Active, 熱狂した-Enthusiastic, ネガティブ感情では恐れた-Afraid, 怯えた-Scared, うろたえた-Upset, など), それぞれに対して「非常によく当てはまる」(6 点) から「全く当てはまらない」(1 点) の 6 件法で採点される. なお, SAM が動画や音楽などを視聴後の短期的な感情を測定するケースで用いられることが多いのに対して, PANAS は気分やムードなど SAM と比較して長期間持続する感情を測定するケースで用いられることが多い.

図 2.4 Self-assessment manikins の使用例 ([87] より引用)

2.2 心理属性と生活習慣，生体活動との関係

性格は、音楽・映画の嗜好性や食生活と運動、スマホや SNS の利用やリスクの取り方など、様々な生活ドメインにおける日常行動に影響を与えることが知られている。また、職業上の評価や学業成績、健康や夫婦関係とも相関する。さらに、脳活動や心拍、皮膚電位など生体活動とも関係することも知られている。以下では、性格を中心に、心理属性と生活習慣、ライフイベント、生体活動の関係について、従来研究で得られている知見を整理する。

2.2.1 心理属性と生活習慣・ライフイベントの関係

学業成績と職務パフォーマンス

従来研究から誠実性と神経症傾向は、学業成績と関係があることが明らかにされている。Chamorro-Premuzi ら [13] は、イギリスの大学生 70 名を対象に縦断的調査を行い、定期試験の点数と Big Five の特性スコアとの関係を分析した。その結果、定期試験の点数に対して誠実性は正の、神経症傾向は負の相関をそれぞれ持ち、これら 2 つの特性のスコアによって試験の点数の分散の 10% を説明できることを見出した。また、Komarraju ら [69] は、172 名の学部生について学業に対するモチベーションを質問紙により測定し

(Academic Motivations Inventory [89] を用いた), モチベーションを構成する因子のうち, Engagement は知的好奇心及び外向性と, Achievement は誠実性, 神経症傾向及び知的好奇心と, Avoidance は神経症傾向及び外向性と, それぞれ正の相関を持ち, また Avoidance は誠実性及び知的好奇心と負の相関を持つことを見出した。

誠実性は職務上のパフォーマンスとも有意な相関を持つことが複数の研究で確認されている。Barrick ら [6] は職務及び訓練の熟達度と Big Five の関係について, 専門職, 管理職, 営業職, 警察など複数の職業グループを対象にメタ分析を行い, 調査した全ての職業について, 誠実性が職務及び訓練の熟達度と正の相関を持つことを見出した。他の特性については職業によって, 職務・訓練の熟達度との関係が異なり, 外向性については管理職と営業職の職務・訓練の熟達度と, 外向性と知的好奇心は全職業グループについて訓練の熟達度と, それぞれ正の相関を持つことを示した。外向性が管理職に求められるリーダーシップと相関することは他の研究でも示されている。Lim ら [76] らは変革的リーダーシップ (自らの配下にいるチームメンバーに行動変容を促す能力) と Big Five の関係について調査を行い (アジアの軍人 276 サンプル, 39 チームを調査した), 外向性が変革的リーダーシップと正の相関を持つことを確認した。

健康と寿命

複数の研究で誠実性と寿命は正の相関にあることが示されている [41, 82]。例えば, Friedman ら [41] は誠実性が高いほど飲酒, 喫煙を控える傾向が高いため癌や心臓疾患にかかる確率が低く, 結果として長い寿命につながっていることを示した。Bogg ら [9] は, 誠実性と寿命との関係をより深く検討するため, 寿命に影響し得る様々な生活習慣と誠実性の関係をメタ分析によって調査し, 誠実性が健康に害を及ぼし得る様々な生活習慣, すなわち過度の飲酒, 喫煙, 薬物使用, 危険運転, 不健康な食事, 暴力行為などと負の相関にあることを確認した。また, Courneya ら [23] は, 264 名の学部学生を対象に Big Five と運動習慣の関係を調査し, 誠実性と運動習慣が正の相関を持つことを示した。このほかの特性では, 運動習慣に対して外向性は正の, 神経症傾向は負の相関を持つことも確認した。

リスク行動の傾向

性格は人生におけるリスクの取り方, 例えば危険性の高いスポーツへの参加傾向や, 高リスクの投資行動などと相関があることが知られている。Tok[143] は, スキューバダイビングやパラグライダー, ラフティング, ロッククライミングなど危険性の高いスポーツに定期的に参加している被験者 ($n = 85$) とそうしたスポーツへの参加経験がない被験者 ($n = 173$) で Big Five のスコアを比較し, 前者は後者に比べて有意に外向性, 知的好奇心のスコアが高く, また誠実性と神経症傾向のスコアが低いことを示した。また,

Lauriola ら [71] は，次の実験を行うことで投資行動におけるリスクの取り方と Big Five の関係を調査した．すなわち，76 人の被験者に対して (A) 一定確率で損または得をする契約と (B) 必ず損または得をする契約（ただし損または得する額は A よりも小さい）の 2 種類の契約を示し，(A)，(B) のどちらかを選択させた．その結果，(A)，(B) とともに得をするパターンでは知的好奇心が高く，神経症傾向が高いほど (B) すなわちリスクの低い契約を選択する傾向が強かった．逆に (A)，(B) とともに損をするパターンでは，神経症傾向が高いほど (A) を選択する傾向が強かった．すなわち，神経症傾向とリスク行動の関係は得をする場合と損をする場合で逆になることを見出した．

スマートフォンや SNS の利用

スマートフォンの利用については，主に通話（着信頻度や着信相手の数，通話時間など）や SMS 送受信，アプリ利用などと Big Five が相関関係を持つことが知られている．Chittaranjan ら [16] は，117 人の被験者を対象に，Big Five と 17 カ月間にわたり収集したスマートフォンの利用傾向との関係を調査し，以下の関係があることを見出した．

- 外向性 外向性が高い被験者は他者から通話を受ける回数が多く，かつ 1 回あたりの通話の時間も長い傾向にある．また，被験者へ通話するユニークユーザ数も平均よりも多い．外向性のスコアとインターネット，ゲーム，カメラアプリとの利用頻度には負の相関があるが，一方でオフィスやカレンダーアプリの利用頻度とは正の相関がある．
- 協調性 協調性のスコアと通話のログに有意な相関は見られなかったものの，ショートメッセージ一通あたりの長さは協調性と正の相関にある．また，アプリ利用については，インターネット，動画や音楽の再生アプリ，メール，カレンダーといったアプリの利用頻度と協調性のスコアが負の相関にある．
- 誠実性 誠実性の高い被験者は他者からの着信における通話時間が短く，また，着信を逃す回数が少ない．アプリの利用に関しては動画や音楽の再生アプリの利用頻度と誠実性のスコアは負の相関にある．
- 神経症傾向 神経症傾向の高い被験者は他者からの着信における通話時間が長く，また，ショートメッセージ一通あたりのメッセージが短い傾向にある．アプリの利用に関してはオフィスやカレンダーアプリの利用頻度が高い傾向にある．
- 知的好奇心 知的好奇心の高い被験者は発着信先のユニークユーザの数が多い傾向にある．ショートメッセージ一通あたりの長さは短く，また送受信の回数も少ない．アプリ利用に関してはオフィスやカレンダーアプリの利用頻度が低い傾向にある．

また，SNS の利用については，Facebook や Twitter における行動パターンも Big Five

表 2.2 Big Five と Facebook 利用傾向の相関関係 ([91] より抜粋. $> |.14|$ の場合 $p < .05$)

	利用時間	利用頻度	友人の数	写真投稿数	自身の投稿数
外向性	.15	-.30	.36	.22	.11
協調性	.09	.20	.23	.22	.21
誠実性	.02	.02	.10	.08	.03
神経症傾向	-.20	-.16	-.03	-.01	-.02
知的好奇心	.02	.06	.05	.09	.13

と相関関係があることが知られている。Moore ら [91] は、学部学生 219 名を対象に、Big Five と Facebook の利用傾向の関係を調査した。被験者全員を対象に質問紙調査で Facebook の利用傾向を調査するとともに、うち 143 人については許可を得て友人や写真、メッセージ投稿の数を調査し、その結果、表 2.2 に示す相関関係を確認した。また、Big Five は Facebook の利用傾向について性年代以上の説明力を持つことを明らかにした。

このほか、Qiu ら [111] は、Big Five と Tweet の関係について調査を行った。142 人の被験者を対象に 1 か月間 Tweet を収集し、Tweet の言語的な特徴が誠実性を除く Big Five の各特性と有意な相関関係を持つことを確認した。

消費行動

従来研究から、性格は映画や音楽などのコンテンツの好みや商品の選定基準（デザインに対する好みや環境への配慮）、支払い方法など、消費行動の様々な側面に影響を与えることが知られている。Weaver[149] はテレビ番組や映画、音楽のジャンルに対する嗜好性と性格の関係について調査を行った。119 人の男性学部生を対象に質問紙調査を行い、神経症傾向が高い被験者ほど、Action や Comedy のテレビ番組は好まず、逆に情報番組を好む傾向にあること、また映画について Adventure を好まず、Tragedy を好む傾向にあること、音楽については Upbeat よりも Downbeat を好む傾向にあること、などを確認した。また、Golbek ら [43] は、73 人の Netflix ユーザを対象に、性格とレコメンドシステムの利用傾向について調査を行い、誠実性のスコアが高いユーザほどレコメンドシステムへの信頼が強く、推薦された映画を鑑賞する頻度が高いことを確認した。

商品の選定基準、なかでもデザイン性については、Myszkowski ら [98] が調査を行い、知的好奇心が高いユーザほどデザイン性を重視した商品選びを行うことを明らかにしている。また、Westfall ら [150] は特に自家用車の購入行動に着目し、協調性、誠実性が高い人はコンパクトカーを好む傾向が強いのにに対して、知的好奇心が高い人はオープンカーを好む傾向が強いことを明らかにした。

性格は商品を選ぶ際の環境への配慮にも影響を与える．Fraj ら [39, 40] は，573 人の被験者に対してアンケート調査を行い，外向性，協調性，誠実性が高い消費者は環境への配慮を訴求した商品を選択する傾向が強いことを明らかにした．また，Fraj らは価値観についても調査を行い，自己実現を重視する消費者ほど，エコなライフスタイルを心掛け，環境への関心が高く，結果として環境への配慮を訴求した商品を選択する傾向が強いことも明らかにした．

Pirog ら [110] は，性格が消費における支払い方法の選択に与える影響を調査した．254 人の大学生に調査を行った結果，神経症傾向が高い被験者ほどクレジットカードを使用する傾向が強く，過剰利用の頻度も高いことを明らかにした．一方で誠実性が高い学生は，支払い方法としてデビットカードや現金を使用する頻度が高いことを明らかにした．

2.2.2 心理属性と生体活動の関係

性格や感情は複数のモダリティの生体活動と関係することが知られている．特に，感情については生体活動との関係性の強さから，関係性のみを対象とする研究よりも，生体信号から感情を推定する研究が盛んに行われている．このため，感情については 5.2 節でまとめて説明することとし，以下では性格と生体活動の関係に関する従来研究を説明する．

脳活動

性格と生体活動との関係に関する研究では，何らかの刺激（特定の感情を喚起することを意図した視聴覚の刺激など）を被験者へ提示した際の生体信号と性格の関係を調査する研究が多い．Canli ら [12] は，外部刺激に対する脳反応は性格によって調整（Moderate）されるはず，との仮説を立て，これを実験によって確認した．ポジティブな感情，ネガティブな感情を喚起する写真を各 20 枚ずつ用意し（ポジティブな写真としては幸せそうなカップル，子犬，アイスクリームやケーキなど，ネガティブな写真としては怒ったり泣いたりしている人物，蜘蛛，銃，墓などを用意），これらの刺激を 14 人の女性被験者に提示した際の脳活動を fMRI（Functional magnetic resonance imaging）で観察する実験を行った．その結果，ポジティブな写真を提示した際の，扁桃体，尾状核，中前回頭，被殻といった部位の活動レベルが外向性のスコアと有意な正の相関（ $r = .79 \sim .86$ ）を示すことを確認した．また，ネガティブな写真を提示した際の，中前回頭，中側回頭といった部位の活動レベルが神経症傾向のスコアと有意な負の相関（ $r = -.79 \sim -.75$ ）を示すことを確認した．

同様の実験は被験者の性年代や提示する刺激を変えて複数の研究者によって行われてお

り（例えば [28, 11, 20, 26, 88] など）、外向性が高いほどドーパミン^{*8}系神経系（腹側被蓋野や側坐核などのエリアが含まれる）の反応性が高いこと、一方で神経症傾向は扁桃体や右背外側前頭前野の活動と密接に関係していること、などが共通の知見として得られている。

心拍活動

心拍活動と性格の関係を調査した研究でも、刺激を提示した際の心拍変化の個人差を調査するアプローチが取られている。Harvey ら [52] は、人間が暴力的に殺害されるシーンのスライドを被験者に提示し、その際の心拍変化を観察する実験を行った。実験の結果、心拍変化は外向性と神経症傾向と強い関係性を持ち、例えば外向性が低い被験者、及び神経症傾向が高い被験者は、刺激提示時の初期段階で心拍が加速度的に早まっていくこと、一方で外向性が高い被験者や神経症傾向が低い被験者ではこうした心拍変化が見られなかったことを確認した。

また、Lü ら [79] と O'Riordan ら [103] は、ストレス環境における心拍活動の変化と外向性の関係について調査した。このうち、Lü らは 166 人の学部学生を被験者として、社会的ストレスを与えられた環境（公衆の前でスピーチを行うが、実験条件によって聴衆からのフィードバックが異なる）の被験者の心拍変化と外向性の関係を調査した。外向性が高い群は中程度のストレス環境下では外向性が低い群と比べて大きな心拍変化は見られなかったものの、高いストレス環境下では逆に外向性が低い群と比べて大きな心拍変化を示すことが確認された。

皮膚電位

皮膚電位とパーソナリティの関係について調査した研究は、筆者の知る範囲では脳活動、心拍活動ほど多くない。しかし、皮膚電位はストレスや不安などネガティブな感情に対する生理的な反応を測る指標として用いられるケースがあり、これと性格の関連について調査した研究が存在する。Binboga ら [8] らは 30 人のテコンドー選手を対象に、性格と試合を前にした不安や緊張を、皮膚電位を用いて調査した。その結果、神経症傾向と試合 1 日前の EDA（Electrodermal activity）が正の相関にあること、一方で試合 1 時間前の EDA と相関したのは Big Five のうち協調性のみであったこと、などを示した。

^{*8} 報酬に関わる神経伝達物質であり、環境内に報酬のキューがある時に放出され、快や多幸福感を引き起こす

2.3 心理属性を用いたサービスのパーソナライズ

前節までで説明したように、性格、価値観、感情は様々な生活ドメインにおける日常行動に影響を与える。そのため、性格、価値観、感情に基づいてサービスのパーソナライズを行う研究は、様々なサービスを対象に行われてきた。本節ではこうしたサービスパーソナライズに関する研究のうち、特に性格、価値観、感情に基づくパーソナライズの効果が高い広告配信、人と人とのマッチング、行動変容のための説得、商品やコンテンツのレコメンドを対象に既存研究の概要を説明する。

2.3.1 広告配信

心理属性を用いた広告配信のパーソナライズに関しては、特に配信する広告デザインのパーソナライズ（同じ商品やサービスについて複数の広告デザインを用意し、これをターゲットの心理属性に合わせて使い分ける）に関する研究が多数行われてきた。Myers ら [97] は広告のデザインが持つ訴求力は消費者の性格によって調整 (moderate) されるとの仮説を提起した。Myers らは、広告デザインを Transformational な広告（感情に訴えかけることにフォーカスした広告）と Informational な広告（客観的な情報提供にフォーカスした広告）、及び比較広告と比較無しの広告などに分類し、Transformational な広告や比較広告は外向性が高い消費者に好まれるのに対し、Informational な広告や比較無しの広告は誠実性の高い消費者に好まれるであろうと主張した。

Myers らの研究は仮説の提示にとどまったが、Hirsh ら [54] はこの仮説が正しいこと、すなわち広告のデザインが持つ訴求力は消費者の性格によって調整されることを実験によって示した。Hirsh らは同一の携帯電話について、商品の持つ複数の特徴のうち特定の特徴を訴求する広告メッセージを複数用意し、被験者に提示する実験を行った。その結果、神経症傾向の高い被験者は携帯電話の物理的な堅牢さとセキュリティ機能を訴求する広告により魅力を感じる一方、外向性の高い被験者は携帯電話が想定するアウトドアスポーツなどの利用シーンを訴求する広告に、より魅力を感じることを確認した。Ding ら [33] は同様の実験を、自家用車を対象に行い（安全性、燃費、デザインなど同一の自家用車の異なる側面を訴求する広告を用意）、それぞれの広告に対して感じる訴求力は消費者の Big Five スコアと相関することを確認した。

Sofia ら [132] と、小林と筆者ら [66] は、広告に用いられる説得戦略及び認知バイアスの効果と Big Five の関係を調査し、説得戦略及び認知バイアスの効果が消費者の Big Five スコアと相関することを見出した。

Matz ら [83] は広告の視覚的なデザインに注目した研究を行った。Matz らは同一の商

品に対して対象とする消費者の性格に合わせて異なる写真を用いた広告を用意し、実際に Facebook 上で広告キャンペーンを行った。具体的には図 2.5 左に示すとおり、化粧品とパズルゲームの 2 種類の商品を対象に、前者については外向性が高い消費者と低い消費者（すなわち内向的な消費者）をターゲットとした広告を（図の A）、後者については知的好奇心の高い消費者と低い消費者をターゲットとした広告を（図の B）、それぞれ用意した。その結果、図 2.5 右に示すように、消費者は自身の性格に合わせてデザインされた広告を示された方が、商品の購入に至る確率（すなわち Conversion rate）が高くなることを確認した。

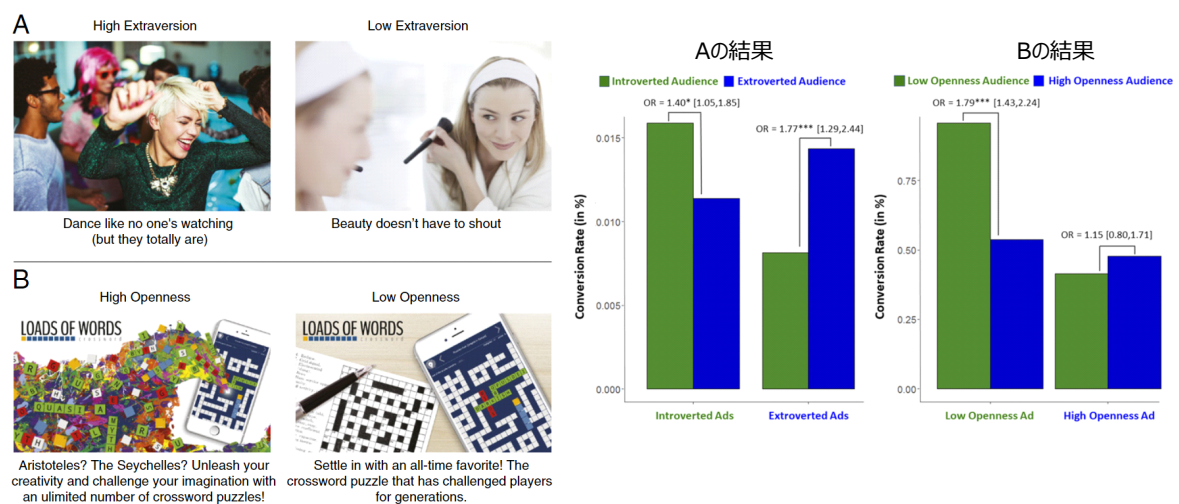


図 2.5 Matz ら [83] の実験で用いられた広告とその効果 ([83] より抜粋)

上述の研究は全て性格によるパーソナライズを対象とした研究だが、価値観に基づいて広告デザインをパーソナライズする研究も行われており、性格同様に価値観による広告デザインのパーソナライズにも効果があることが示されている。前述の Ding ら [33] は Big Five に加え、シュワルツの価値観を基に測定した被験者の価値観によって自家用車の広告デザインの効果が異なることを確認した。

これらの従来研究を要約すると、広告の視覚的および言語的なデザインが持つ訴求力は、Big Five とシュワルツの価値観で測定可能な性格と価値観によって調整される、ということになる。すなわち、全ての消費者に対して同一のデザインの広告を配信するのに比べて、消費者の性格に合わせて異なるデザインの広告を配信することで、消費者が広告のクリックや商品の購入など所望の行動を取る可能性が高まると言える。

2.3.2 人と人とのマッチング

様々な場面を対象に、心理属性を利用して人と人とのマッチングを行う研究が行われている。第一の例は、企業側エージェントと顧客のマッチングである。Conway ら [135] は、類型論にしたがって、コールセンターに架電してきた顧客とコールセンターの顧客サポートエージェントの性格を分類し（顧客の性格は過去の通話記録の会話内容を基に推測した）、同一の性格に分類された顧客とエージェントをマッチングした場合、通話の 90% 以上が 5 分以内で終了したのに対して、性格が正反対の顧客とエージェントでは、通話の 50% 近くが 5 分以上継続したと報告している。同様に Herzig ら [53] や、小林と筆者ら [67] は、それぞれ顧客サポートと電話営業（テレマーケティング）において Big Five と Schwartz の価値観の類似度を用いて顧客とエージェントのマッチングを行うことで、顧客の満足度が高まること、電話営業の成功率が高まること、をそれぞれ確認した。

心理属性を利用したマッチングは、組織におけるチーム形成の場面でも有効性が確認されている。企業や軍隊において、類似する性格のメンバー同士でチームを作ることでチームのパフォーマンスが上がることや [107]、性格、価値観に基づいて学生を班分けすることで協働学習（Collaborative learning）の学習効果が高められること [51, 65] などが報告されている。

また、恋愛や結婚におけるマッチングと心理属性の関係についても研究が行われており、Montoya ら [90] は、短期間での付き合いでは、自分と性格・価値観が類似する相手ほど好意を抱きやすいものの、付き合いが長時間になるにつれ、類似性と好意の強さの相関は弱くなっていくことを報告している。

2.3.3 行動変容のための説得

心理属性を用いて説得戦略をパーソナライズすることで、行動変容の効果が高まるということが既存研究において明らかにされている。Oyibo ら [105] は、説得戦略を類型として広く用いられている Cialdini の 6 つの類型 [18]、すなわち 1. 返報性 (Reciprocity), 2. 一貫性 (Commitment and consistency), 3. 社会的証明 (Social proof), 4. 権威 (Authority), 5. 好意 (Liking), 6. 希少性 (Scarcity) を訴求する戦略類型*⁹、を対象に、被説得者の Big Five と、各戦略の説得力の評価（どの戦略にどの程度の説得力を感じるか）の関係を質問紙により調査した。その結果、誠実性が高い場合は、1., 2. が、協調性が高い場合は、2., 4., 5. が、知的好奇心が低い場合は、3., 4., 5. が、それぞれ説得力が強いと感じることなどを確認した。

*⁹ これに 7. 連帯 (Unity) を加えた 7 つの類型とする場合もある。

Oyibo らの研究が質問紙調査によるものだったのに対して、Ciocarlan[19] らは、PCゲームを利用した実験により、どの説得戦略が実際に行動変容を喚起しやすいかを調査した。具体的には被験者に対して、ゲームの進行に合わせて Cialdini の戦略類型に対応した選択肢を複数回提示し（例を図 2.6 に示す）、被験者の選択結果と Big Five の関係を調査した。その結果、協調性と希少性、外向性と権威の間の有意な負の相関や、知的好奇心と一貫性の間の有意な正の相関などを確認した。また、これらの性格と説得戦略の関係は実験を通じてある程度一貫していることを確認した。なお、Ciocarlan らの実験で確認された性格と説得戦略の関係は、Oyibo らの質問紙調査結果と必ずしも一致しておらず、Ciocarlan らは質問紙調査で確認した各戦略の説得力の評価と、各戦略が実際の行動に影響を与えるかは別のものとして評価すべきだと主張した。

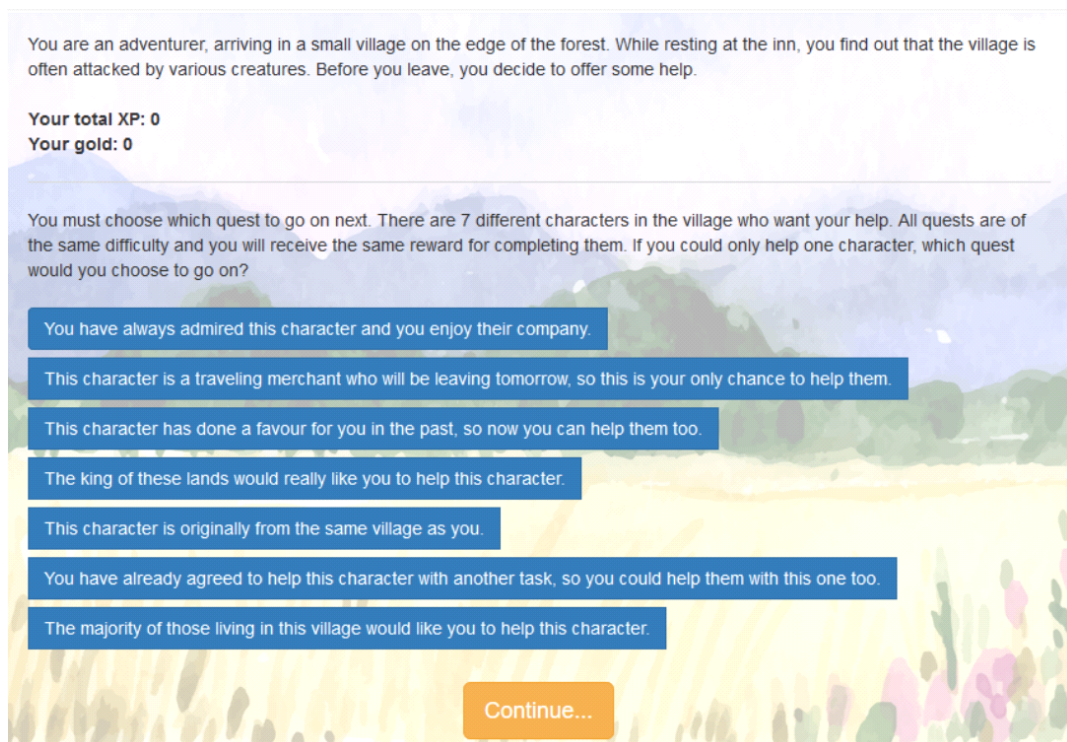


図 2.6 Ciocarlan ら [19] が実験で用いた PC ゲームのスクリーンショット ([19] より抜粋)。被験者に Cialdini の説得戦略に対応した選択肢を提示した。図の 7 つの選択肢は上から好意、希少性、返報性、権威、連帯、一貫性、社会的証明にそれぞれ対応する。

2.3.4 レコメンド

2.2.1 節で述べたように、心理属性は様々な生活ドメインにおける日常行動に影響を与えることから、ユーザへ商品や映画・音楽などのコンテンツ、さらにはレストランや観光スポットなどの POI (Point of interest) のレコメンドなど、多岐にわたるサービス領域

で心理属性を利用したレコメンドに関する研究が行われている [31]. 心理属性（特に性格）がレコメンダーシステムにおいて注目を集めている背景には主に、1) 新規ユーザ（いわゆる Cold start problem への対応）に心理属性が有効であることと、2) レコメンドの多様性に対する選好性と心理属性に相関があること、の 2 点がある [142].

まず、1) については、ユーザに対してレコメンドを行う際には、レコメンドを行うサービスにおける当該ユーザの選好性等に関する情報が必要となる（例えば、音楽のレコメンドを行う場合には音楽の好みに関する情報が必要）。通常、こうした情報は、ユーザの当該サービスにおける過去のログ（音楽の場合は楽曲の鑑賞履歴など）から把握することになるが、ユーザが当該サービスを利用するのが初めての場合、または利用を開始して間もない場合など、ユーザの選好性を把握するのに十分な量のログが蓄積されていない場合があり、こうした状況では通常のアプローチではレコメンドを行えない（これを Cold start problem と呼ぶ）。

心理属性はこれを解決する手段の一つになり得る。サービスを利用するのが初めてのユーザであったとしても、そのユーザの心理属性が把握できている場合、他のユーザのログから得られた心理属性とレコメンド対象の関係に基づいて（例えば外向性の高いユーザはアップテンポの音楽を好むなど）、レコメンドを行うことが可能となる。

次に、2) については、通常ユーザに対してレコメンド結果を提示する場合、単一ではなく複数のアイテムやコンテンツを提示するが、提示されるレコメンド結果の多様性に対する好みとユーザの心理属性の間に関連性があることが知られている。このため、提示するレコメンド結果の多様性をユーザの心理属性によってコントロールする研究が行われている [142]. 例えば、Chen ら [14] は、映画のレコメンドについて、提示するレコメンドに含める映画の各属性（俳優・監督やジャンル、制作年代や制作国など）の多様性に対する好みと性格の関係を調査した。その結果、神経症傾向が高いユーザは、多様な監督が含まれていることを、協調性が低いユーザは、制作国が多様であることを、知的好奇心が高いユーザは俳優が多様であることを、それぞれ好む傾向にあることを明らかにした。また、どの映画属性についても、誠実性とレコメンド結果の多様性の好みは負の相関であることを明らかにした。

第3章

心理属性・質問紙によるサービス パーソナライズ効果の違い

3.1 背景と目的

2章で述べたようにユーザの心理属性に基づいたサービスのパーソナライズは様々なサービスで広く研究されており、所望のユーザ反応を得る確率を高めるのに有効であることが確認されている。実サービスのパーソナライズにおいて心理属性を用いる場合、多数のユーザー一人ひとりの心理属性を質問紙によって測定するのは非現実的なため、心理属性を行動データや生体データから推定する研究も幅広く行われてきた。実サービスではこれらの研究の知見を組み合わせ、ユーザの心理属性を推定した上で、推定した心理属性によってサービスをパーソナライズすることになると考えられる。このため、ユーザの心理属性を推定する際は、よりパーソナライズの効果が高い心理属性をターゲットとして推定を行うことが望ましい。すなわち、サービスパーソナライズのために心理属性推定を行う場合、パーソナライズの効果が高いものを Ground truth として選択する必要がある。

そこで、本章では心理属性に基づくパーソナライズの研究が比較的多く行われている広告デザインのパーソナライズを対象に、Ground truth の選択について検討する。広告デザインのパーソナライズのシナリオは以下を想定した。

1. 同一の商材について予め複数の広告デザインを用意する
2. 配信先の各ユーザの心理属性と各広告デザインの特徴量を用いて、ユーザ-広告デザインの組合せ毎にクリックまたはコンバージョン（購買）の確率を予測する
3. 予測した確率が最も高い広告デザインを各ユーザに配信する

上記シナリオを想定し、従来研究では未解決であった以下のリサーチクエスション（RQ）を検討する。括弧内は想定シナリオにおいて検討する具体的な RQ である。

- RQ1 Big Five には複数の質問紙が存在するが、質問紙によって測定結果が持つサービスパーソナライズの効果は異なるか (Big Five の質問紙によって、クリック・コンバージョンの予測力は異なるか)。
- RQ2 長期安定性とドメイン非依存性の二つの点で共通の特徴を持つ「性格」と「価値観」はどちらの方がパーソナライズの効果が高いのか (Big Five とシュワルツの価値観ではどちらの方がクリック・コンバージョンの予測力が高いか)。
- RQ3 性格と価値観のサービスパーソナライズの効果は独立か、組み合わせて用いることでパーソナライズの効果は高まるか (Big Five とシュワルツの価値観のクリック・コンバージョンの予測力は独立か)。
- RQ4 上記 RQ の答えは広告対象の商材によって変わるか。

3.2 関連研究

2.3.1 節で述べたように、従来研究により性格や価値観を用いて広告デザインをパーソナライズすることで、クリック率や購買率を高められることが明らかにされていたものの、性格・価値観に合わせた広告デザインの作成は専門家の手に委ねられていた。実際、Hirsh ら [54] や Matz ら [83] の研究で用いられた広告デザインは、心理学の専門家によるレビューを経て作成されていた。このため、消費者の反応を見ながら頻繁に広告デザインの差し替えを行うケースでは、コストが課題となる。

こうした課題を踏まえ、Matz ら [84] と筆者ら [59] は、広告デザインから機械的に抽出した特徴量（配色に関する統計量など）と、配信先の消費者の心理属性を用いて（Matz らは Big Five を、筆者らは Big Five とシュワルツの価値観を、それぞれ用いた）、消費者の広告反応（クリックや購買に至る確率）を、統計モデルにより予測する手法を検討した。双方の研究とも、予測結果を用いて広告デザインの配信を行い、最もクリック・購買に至る確率の高い広告デザインを各消費者へ配信することで、専門家の介入なしに広告のクリック率や購買率を高められることを実験により示した。

しかし、双方の研究において、未解決の課題が残されている。第一は、Big Five とシュワルツの価値観の違いに関するものである。すなわち、広告反応の予測において、Big Five とシュワルツの価値観のどちらがより高い予測力を持つのか、また、その予測力は独立しているのか、という点が明らかにされてこなかった。言い換えると、人間が広告デザインから感じる訴求力を調整 (Moderate) する効果は、Big Five とシュワルツの価値観のどちらが大きいのか、両者の調整効果 (Moderating effect) は独立しているのか、という点が明らかにされてこなかった。2.1.2 節で述べたように、性格と価値観は概念の異なる心理的構成要素だが、両者の間には語彙的な重複が見られる。例えば、Big Five の特性

の1つである「外向性」には、刺激を求める傾向が含まれる一方、「刺激」自体はシュワルツの価値次元の1つである。同様に、Big Fiveの「協調性」には優しさや暖かさ、同情といった要素が含まれているが、これらはすべてシュワルツの価値観の「慈悲」にも含まれる概念である。これらを踏まえると、広告デザインの訴求力の調整効果について、両者がどの程度異なるかは必ずしも明らかではなく、両者を用いても、どちらか一方を用いた場合と比較してモデルの予測力が向上しない可能性がある。

もう1つの課題は、Big Fiveの測定に使用する質問紙の違いである。Matzら[84]も筆者ら[59]も、質問紙を用いてBig Fiveの測定を行ったが、測定に用いた質問紙の測定結果が他の質問紙の測定結果と比べて、広告デザインのパーソナライズに適していたかについては明らかにしていない。この点について、Big Fiveを異なる質問紙で測定した二つの研究[42, 139]が、Big Fiveと人間が好むデザインの特徴の関係について、互いに一貫性のない結果を報告していることは、注目に値する。両研究は、抽象芸術の嗜好性とBig Fiveの関係を調査するため、NEO-FFI[22]とTIPI[48]をそれぞれ用いてBig Fiveの測定を行った。NEO-FFIを用いた研究では、抽象芸術に対する嗜好性と誠実性及び協調性の間に有意な負の相関を確認したものの、TIPIを用いた研究ではそのような相関は認められなかった。このような矛盾が生じる理由の一つとして、NEO-FFIは抽象芸術の嗜好に影響する性格の領域を測定できるが、TIPIではそのような領域が測定できないことが考えられる。同じことが広告の訴求力の認知にも当てはまる可能性がある。つまり、一方の質問紙では、広告デザインが訴求力認知に与える影響を大きく調整する性格の領域を測定できるが、他方の質問紙で測定できる性格領域には、その調整効果はほとんどないという可能性がある。

3.3 データセット

被験者のBig Five、シュワルツの価値観のスコア、広告デザインに対する反応、および広告デザインの特徴量で構成されるデータセットを構築した。これらのデータを収集するため、100個の広告デザインを用意し、8,245人の日本人被験者に対してWeb上でアンケート調査を実施した。アンケート調査の概要を図3.1に示す。

3.3.1 広告反応

4つの架空の商品/サービスそれぞれについて、5人のプロのグラフィックデザイナー(Ds1~Ds5)が5個のWebバナー広告を作成した(つまり、各デザイナーは合計20個のバナー広告を作成した)。商品/サービスは以下のとおりである:

- PL: モバイルネットワークサービスの料金プラン

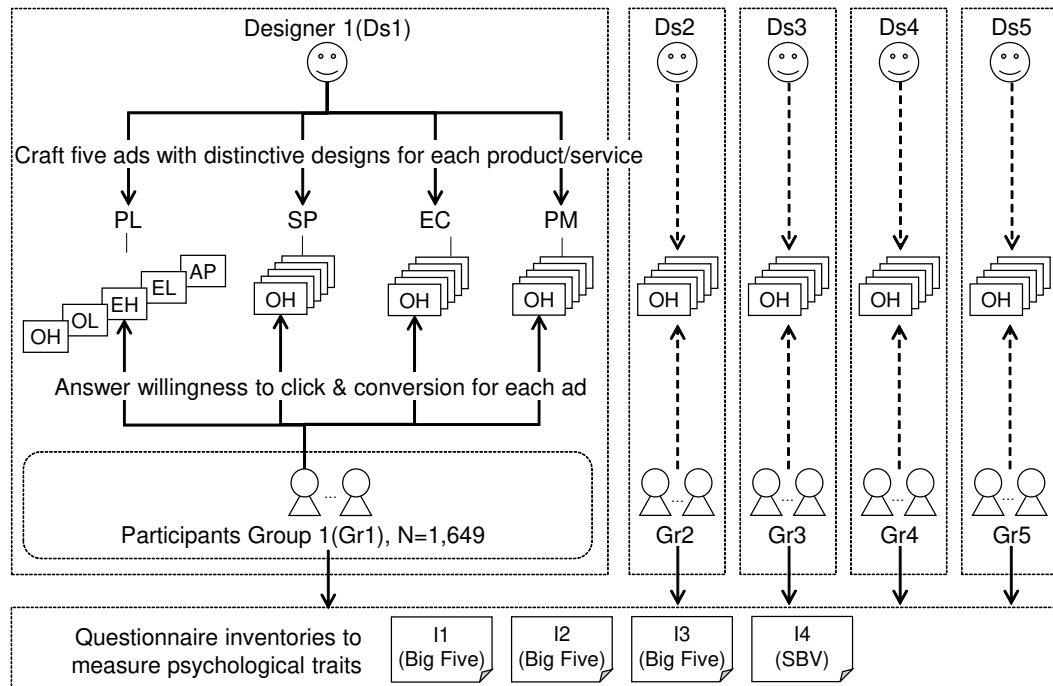


図 3.1 アンケート調査の概要

- SP: スマートフォン
- EC: EC サイト
- PM: モバイルネットワークキャリアのプレミアムメンバーサービス

可能な限り多様な商品/サービスをカバーできるように上記を選んだ。PL と PM の広告はサービスへの加入を促すためのもので、PL と PM では料金レンジが異なる（PL は月額 2,000 円程度なのに対して PM は月額 500 円程度）。SP は商品購入を促す広告であり、EC はボーナスポイントを付与するキャンペーンを告知するための広告とした。

各商品/サービスについて異なるデザインのバナー広告を準備するため、デザイナーには 5 つの異なる性格をターゲットとしたバナー広告をデザインするよう依頼した。ターゲットとする性格は Big Five を用いて以下のように定義した:

- OH: 知的好奇心が高い
- OL: 知的好奇心が低い
- EH: 外向性が高い
- EL: 外向性が低い
- AP: 平均的な性格

このようにターゲットとする性格を定義したのは、各デザイナーが同一商品/サービスについて可能な限り異なるデザインの広告を準備できるようにするためである。デザインさ

れた広告が対象とする性格にどの程度適しているかは配慮せず，したがって筆者らからデザイナーに対してデザインについて助言を与えたり，デザインと性格の適合性についてレビューを行ったりもしなかった．各デザイナーにはターゲットとする性格の簡単な説明を与えるのみにとどめた．図 3.2 にバナー広告の例を示す．バナー広告のサイズは Web 上でよく利用される 300 × 250 ピクセルのサイズとし，日本語によるメッセージを含めた．



図 3.2 バナー広告の例 (PL)

被験者の負荷を抑制し，不誠実な回答を避けるために，被験者をランダムに 5 つのグループに分割し (Gr1~5; 各グループは 1,649 人で構成)，各グループに対して同一デザイナーが作成した 20 個のバナー広告を提示した．すなわち各被験者には 4 つの商品/サービスそれぞれについて 5 個のバナー広告を提示し，各バナー広告に対してクリックと商品/サービスの購入/加入 (コンバージョン) の意思を 5 件法で質問した (質問-以下の広告バナーをご覧ください，クリック/購入/加入してみたいと思いましたか？その度合いをそれぞれ 5 段階で教えてください．回答-5: とても思う，4: やや思う，3: どちらとも言えない，2: あまり思わなかった，1: 全く思わなかった)．図 3.3 にアンケート調査の質問フォームの例を示す．アンケート調査では，バナー広告と商品/サービスの提示順序を被験者ごとにランダム化し，可能な限り回答の品質に偏りがでないようにした．

データ分析では，各回答のうち，4 以上の回答をクリックおよびコンバージョンと見なし，クリックした/しなかった，コンバージョンした/しなかった，に 2 値化して被験者の広告反応と見なした．なお，データ分析を行う前，95% 以上のバナー広告に対して同一のスコアを付与している被験者は不誠実な回答を行っている被験者と見なし除外した．8,245 人中 7,857 人の被験者 (95.3%) の回答をデータ分析に利用した．

3.3.2 心理属性

被験者の Big Five とシュワルツの価値観は表 3.1 に示す質問紙を用いて測定した (表中 SBV はシュワルツの価値観を意味する)．

Big Five については，質問紙を選択するに当たり最初に日本語の質問紙を列挙し (表 2.1.1 参照)，被験者の回答負担を考慮して 200 問以上の質問項目で構成される質問紙を排除した．その上で残った質問紙を質問項目数 (大-20 問以上，少) および質問方法 (形容

以下の広告バナーをご覧ください、クリックしてみたいと思いましたが？
その度合いをそれぞれ5段階で教えて下さい。

① 第一回答 ★ 必須回答 ▶ とじる 1/20

スマートフォンの新しい携帯料金プランについての広告バナーです。

1



① 全く思わなかった
② **あまり思わなかった**
③ どちらともいえない
④ やや思う
⑤ とても思う

2



3



図 3.3 アンケート調査における質問フォームの例

表 3.1 心理属性の測定に用いた質問紙（SBV はシュワルツの価値観を表す）

	心理属性	項目数	質問方法	回答方法	質問紙名
I1	Big Five	10	形容詞単語	7 件法	日本語版 TIPI[159, 104]
I2	Big Five	70	文章	はい/いいえ	主要 5 因子性格検査 [161]
I3	Big Five	29	形容詞単語	5 件法	Big Five 尺度短縮版 [162]
I4	SBV	21	文章	6 件法	PVQ[123]

詞単語、文章）で 4 つのカテゴリに分類し、各カテゴリから引用件数の多い質問紙を選択した。なお、質問項目数-少、質問方法-文章に該当する日本語質問紙は筆者が調査した範囲では存在しなかった。

シュワルツの価値観については日本語訳が唯一利用可能な PVQ (Portrait Value Questionnaire) [123] を利用した*1。

被験者に対して I1~I4 の質問紙を提示する際、その順序はランダムイズし、質問紙間で回答の品質に偏りが出ないように配慮した。

回答結果を用いて、Big Five については質問紙ごとに知的好奇心 (O)、誠実性 (C)、外向性 (E)、協調性 (A)、神経症傾向 (N) の各スコアを、シュワルツの価値観については、自己決定 (SD)、刺激 (ST)、快楽 (HE)、達成 (AC)、権力 (PO)、安全 (SE)、調

*1 広岡ら [157] の日本語訳を利用した。

表 3.2 I1~I4 による測定結果の α 係数と I1~I3 で測定した Big Five 質問紙間の相関係数 (r)

	Big Five						シュワルツの価値観			
	α			r			α			
	I1	I2	I3	I1 - I2	I1 - I3	I2 - I3	I4			
O	.39	.79	.83	.63	.50	.63	SD	.51	ST	.53
C	.46	.82	.82	.67	.56	.68	PO	.45	CO	.52
E	.54	.84	.89	.70	.70	.78	UN	.58	TR	.38
A	.25	.76	.72	.58	.40	.44	AC	.71	HE	.31
N	.39	.83	.90	.64	.58	.73	SE	.40	BE	.52

和 (CO), 伝統 (TR), 慈悲 (BE), 普遍主義 (UN) の各スコアを計算した。

表 3.2 に各質問紙の信頼性 (内部一貫性, クーロンバックの α 係数により定量化した) と質問紙間の Big Five スコアの相関係数 (r) を示す. I1 および I4 で測定したいくつかのスコアについては α 係数が低かった ($\alpha < .40$; 例えば I1-A, I4-HE). これらの質問紙では各スコアを 2, 3 個の質問で測定しているためである. また, 質問紙間の Big Five スコアの相関係数も比較的低い結果となった (例えば I1 と I3 で協調性のスコアの相関係数は $r = .40$). この結果は, 同一の特性 (例えば協調性) であっても質問紙によって測定している領域が異なることを示唆している。

3.3.3 広告のデザイン特徴量

100 個のバナー広告それぞれについて, 視覚的および言語的なデザイン特徴量を抽出した. 視覚デザインについては, 計 64 個の特徴量を Matz ら [84] が開発したプログラム^{*2}を用いて抽出した. 同プログラムによって抽出した特徴量のうち, 画像サイズと縦横比, および画像の内容に関する 22 個の特徴量 (画像中の人の顔や自動車, 飛行機などの検出数と検出エリア) については使用しなかった. 前者については本研究のバナー広告では全て同じ値となるため, 後者については本研究のバナー広告では検出精度が極めて低かったためである. 分析に用いた視覚デザインの特徴量の概要は以下のとおりである。

Color 全 20 次元, HSV (Hue, Saturation, Value) 統計量, PAD (Pleasure, Arousal, Dominance) の測定値, 11 の色の多様性と量

Composition 全 8 次元, 画像全体の視覚的要素の構成に関する特徴量

^{*2} <https://tinyurl.com/OSFmaterial>

Texture 全 29 次元, 強度と色の空間的配置に関する特徴量

Content 全 7 次元, 画像の Busyness に関する特徴量 (2 次元) とコンピューターグラフィックと自然画像の判別に関する特徴量 (5 次元)

言語デザインの特徴量については, 筆者らの過去の研究 [59] と同様のアプローチを採用し, 広告メッセージの言語的意味と表現の両方を捉えるため, 日本語版 Wikipedia の記事で事前学習した単語ベクトル (50 次元)*³を利用し, 広告メッセージに含まれる全ての単語の重心を計算した.

3.4 方法

RQ1~RQ4 に答えるため, 前節で述べたデータセットを用いてモデル比較分析を行った. 各 RQ について, バナー広告のデザイン特徴量に加え, 異なる心理属性の測定結果を説明変数とするモデルを複数用意し, 各モデルがどの程度正確に被験者のクリック, コンバージョンの意思を予測できるかを比較した. 例えば, RQ1 については I1~I3 を説明変数とする三つのモデルを用意し, いずれのモデルが真のモデル (すなわち被験者のクリック, コンバージョンの意思を誤りなく予測するモデル) に近いかを評価した.

3.4.1 モデルの定式化

被験者が広告デザインから感じる訴求力は, (1) 個人の心理属性によって他者とは感じ方が異なる訴求力 (Personal appeal) と (2) 広告デザインに対して誰しも共通して感じる訴求力 (Universal appeal) の二つの要素で構成されると仮定した. この仮定に基づいて, Personal appeal を広告デザイン特徴量と心理属性の調整効果 (Moderating effect) として定式化するとともに, Universal appeal については, 広告デザイン特徴量の主効果として定式化した. モデル式を以下に示す.

$$P_{\text{react}}(A_m, R_n) = \text{logistic}(Z_{mn}) \quad (3.1)$$

$$Z_{mn} = \text{main_effect}_m + \text{moderating_effect}_{mn} + b \quad (3.2)$$

$$\text{main_effect}_m = \sum_i (B_i D_{mi}) \quad (3.3)$$

$$\text{moderating_effect}_{mn} = \sum_i \sum_j (B_{ij} D_{mi} T_{nj}) \quad (3.4)$$

*³ <https://github.com/chakki-works/chakin>

表 3.3 モデルに用いたデザイン特徴量セットと心理属性スコアセット

デザイン特徴量の組合せ		心理属性スコアの組合せ	
D1	V_m	T1	BF_I1_n
D2	V_m, L_m	T2	BF_I2_n
D3	V_m, V_m^2	T3	BF_I3_n
D4	V_m, V_m^2, L_m	T4	$BF_I1_n, BF_I2_n, BF_I3_n$
D5	V_m, V_m^2, L_m, L_m^2	T5	SV_n
		T6	$BF_I1_n, BF_I2_n, BF_I3_n, SV_n$

$P_{\text{react}}(A_m, R_n)$ は被験者 R_n がバナー広告 A_m に対してポジティブな反応（すなわちクリックまたはコンバージョン）を示す確率を表す．モデルにおいて，各被験者がバナー広告から感じる訴求力は $P_{\text{react}}(A_m, R_n)$ で表現可能と考えた．バナー広告から感じる訴求力が高いほど，クリックまたはコンバージョンする確率も高まるためである． Z_{mn} はロジスティック回帰式の線形予測子であり， main_effect_m と $\text{moderating_effect}_{mn}$ で構成される． b , B_* , D_{mi} , 及び T_{nj} はそれぞれ切片，偏回帰係数， A_m の i 番目のデザイン特徴量， R_n の j 番目の心理属性のスコアを表す．

クリックとコンバージョンを予測するモデルは別々に構築し，それぞれ目的変数は $P_{\text{click}}(A_m, R_n)$ と $P_{\text{conversion}}(A_m, R_n)$ とした．

3.4.2 デザイン特徴量と心理属性

デザイン特徴量と心理属性スコアについて表 3.3 に示す複数の組合せ（セット）を準備し，これらをモデルで用いた．表において V_m と L_m はバナー広告 A_m の視覚的デザインの特徴量と言語的デザインの特徴量ベクトルである．これらのベクトルの要素は A_m の i 番目のデザイン特徴量 V_{mi} , L_{mi} である（すなわち， $V_m = (V_{m1}, \dots, V_{mi}, \dots)$, $L_m = (L_{m1}, \dots, L_{mi}, \dots)$ ）．また， V_m^2 と L_m^2 は各特徴量の二乗値を要素とするベクトルである（ $V_m^2 = (V_{m1}^2, \dots, V_{mi}^2, \dots)$, $L_m^2 = (L_{m1}^2, \dots, L_{mi}^2, \dots)$ ）．例えば，デザイン特徴量セットとして D2（つまり V_m と L_m ）をモデルで用いる場合， main_effect_m は以下のように表現される．

$$\text{main_effect}_m = \sum_{i_V} (B_{i_V} V_{mi_V}) + \sum_{i_L} (B_{i_L} L_{mi_L}) \quad (3.5)$$

BF_I1_n , BF_I2_n , BF_I3_n はそれぞれ I1, I2, I3 で測定した Big Five スコアを表し，次の数式で表される． $BF_I*_n = (BF_I*_{n1}, \dots, BF_I*_{nj}, \dots)$ ．ここで BF_I*_{nj}

は質問紙 I*で測定した R_n の j 番目の性格特性のスコアを表す。また、 $\mathbf{SV}_n$ は R_n のシュワルツの価値観スコアを表し、次の数式で表される。 $\mathbf{SV}_n = (SV_{n1}, \dots, SV_{nj}, \dots)$ 。ここで SV_{nj} は R_n の j 番目の価値観次元のスコアを表す。例えば、モデルに D2 と T2 (すなわち $\mathbf{BF_I2}_n$) を用いた場合、 $\text{moderating_effect}_{mn}$ は次の数式で表される。

$$\begin{aligned} \text{moderating_effect}_{mn} = & \sum_{i_V} \sum_j (B_{i_V j} V_{mi_V} \mathbf{BF_I2}_j) \\ & + \sum_{i_L} \sum_j (B_{i_L j} L_{mi_L} \mathbf{BF_I2}_j) \end{aligned} \quad (3.6)$$

3.4.3 モデルの構築

モデルの構築に当たって、まず特徴量選択を実施した。例えば、D5 と T6 をモデルに利用した場合、説明変数の数が 5,000 を超えるなど、モデル内で使用する説明変数が膨大な数になるケースがあり、オーバーフィッティングのリスクを考慮する必要があったためである。オーバーフィッティングのリスクを軽減するため、LASSO (binomial LASSO) [141] を用いて特徴量選択を行い、被験者の反応予測に有用でない変数を除外した。最初に LASSO において正則化の強さを制御するパラメタ λ のチューニングを行った。 λ の値を変更しながら 10-fold の交差検証を実施し、被験者の実際の広告反応 (0 または 1) とモデルの予測値 (0 ~ 1 の値をとる) の平均二乗誤差が最小となる λ の値を特定した。その上で、特定した λ の値を用いて LASSO 回帰を再度実施し、偏回帰係数がゼロとなる説明変数を除外した。

特徴量選択を実施した後、ロジスティック回帰を実施した。いくつかのモデルでは LASSO による特徴量選択を行っても、変数間の多重共線性が原因となりロジスティック回帰が収束しなかった。このようなケースでは、VIF (Variance Inflation Factor) を用いて多重共線性のリスクが高い説明変数を排除した。VIF の値を全ての説明変数について計算し、最大の値を取る説明変数を除外、再度残りの全説明変数について VIF を計算する、というステップを VIF の最大値が閾値未満 ($\text{VIF} < 300$) になるまで繰り返し、多重共線性のリスクを抑制した。

3.4.4 モデル比較の指標

モデルの比較には赤池情報量規準 (AIC) [2] とベイズアン情報量規準 (BIC) [125] を用いた。これらの情報量規準は、実世界の事象をモデリングする過程で失われた情報量を見積もることができる指標として、多くの研究でモデルの比較に用いられている。したがって、AIC, BIC を用いることで、どのモデルがより正確に「個人が広告デザインから

訴求力を知覚するプロセス」を再現しているかを比較できる。また、AIC, BIC はモデルで利用される変数の数に応じてペナルティを課すため、オーバーフィッティングのリスクも含めてモデルを比較できる。前述のとおりモデル構築時には特徴量選択を行ったものの、モデルに最初に入力される変数の数が膨大であるため、オーバーフィッティングのリスクは残っており、AIC, BIC を用いることでこうしたリスクを考慮した評価を行うことができる。

比較結果の信頼性向上のため、AIC と BIC は両方利用した。AIC, BIC は両方とも、比較するモデルのいずれが真のモデル（実世界の事象に完全にフィットするモデル）に近いかを「決定」する指標ではなく、比較するモデルのいずれが真のモデルに近い、その尤度を見積もる指標に過ぎない（例：モデル A よりもモデル B の方が真のモデルに近いとする尤もらしさは 0.01, つまりモデル A の方が B よりも真のモデルに近い、とする判断は妥当）。したがって、あるモデルの AIC と BIC が両方とも他のモデルよりも小さい場合、その結果は AIC または BIC のどちらかを用いて得られた結果よりも信頼できる。

しかし、AIC と BIC で矛盾する結果となるケースもある（例えば、モデル m_1 の AIC は m_2 の AIC より小さいのに対し、 m_1 の BIC は m_2 より大きい）。このような場合は、AIC の結果を採用することとした。AIC と BIC の比較については研究者の間でも議論があるが、その中で「真のモデルが候補モデルセットにない場合、AIC は漸近的に予測・推定の平均二乗誤差を最小にするモデルを選ぶので効率的である。このような状況では BIC は効率的ではない」とする意見がある [146]。本研究におけるモデルでは心理属性以外に広告の訴求力知覚に影響を与えるであろう個人の属性（家族構成や地理的属性、類似商品/サービスの購買履歴など）を使用していないため、今回の評価で比較するモデルの中には真のモデルは存在し得ない。これらは広告訴求の知覚に影響を与える可能性があるが、本研究の範囲外である。こうしたことから AIC と BIC で矛盾する結果となったケースでは AIC の結果を採用することとした。

3.4.5 モデルの比較

各 RQ についてモデルの比較を実施した。RQ4 を除きモデルは全被験者、全商品/サービスのデータを用いて構築した。

RQ1: Big Five の質問紙によって、クリック・コンバージョンの予測力は異なるか

異なる質問紙 I1~I3 による Big Five の測定結果 T1~T3 を利用する次の三つのモデルについてモデル比較 (C1) を実施した: $m_{Dk,T1}$, $m_{Dk,T2}$, $m_{Dk,T3}$ 。この三つのモデルから AIC/BIC が最小となるモデル m_{min} (真のモデルに最も近いモデル) を特定した。以

下式の中で $IC(m)$ はモデル m の AIC または BIC の値を表す。

$$\begin{aligned} m_{min} &= \arg \min_{m \in M} IC(m) \\ \text{where : } M &= \{M_{BF}\} \\ M_{BF} &= \{m_{Dk,T1}, m_{Dk,T2}, m_{Dk,T3}\} \\ Dk &\in \{D1, \dots, D5\} \end{aligned} \quad (C1)$$

モデル比較はデザイン特徴量セット (Dk) 毎に行った。 m_{min} で使用した Big Five の測定結果は、 Dk に対して T1~T3 の中で最大の調整効果を持つと言える。例えば D1 について $m_{min} = m_{D1,T1}$ の場合、 T1 が D1 に対して最大の調整効果を持つことを意味する。 D1~D5 のそれぞれのデザイン特徴量セットについて、こうしたモデル比較を行うことで、最大の調整効果を持つ Big Five の測定結果を特定した。 D1~D5 の全てについて、同一の質問紙から得られた Big Five 測定結果が最大の調整効果を持つ場合、そのような質問紙は他の質問紙で得られた Big Five の測定結果よりも広告デザインのパーソナライズ効果が高いと言える。

RQ2: Big Five とシュワルツの価値観ではどちらの方がクリック・コンバージョンの予測力が高いか

Big Five の測定結果を用いたモデルとシュワルツの価値観 (SV) を用いたモデルで以下のモデル比較 (C2) を行った。

$$\begin{aligned} m_{min} &= \arg \min_{m \in M} IC(m) \\ \text{where : } M &= \{M_{BF+}, M_{SV}\} \\ M_{BF+} &= \{M_{BF}, m_{Dk,T4}\}, M_{SV} = \{m_{Dk,T5}\} \end{aligned} \quad (C2)$$

Big Five を用いたモデルのグループ M_{BF+} には、Big Five の3つの質問紙全ての測定結果を使用したモデル $m_{Dk,T4}$ を追加した。質問紙 I1~I3 の間で、測定結果の相関が強くなかったことを踏まえると (表 3.2 参照)、各質問紙で測定可能な性格の領域の重複は大きくないと考えられる。このため $m_{Dk,T4}$ によって個別の質問紙の測定結果を用いたモデル ($m_{Dk,T1} \sim m_{Dk,T3}$) よりも、性格の広い領域をカバーできると考えた。

C1 と同様、それぞれのデザイン特徴量セット Dk について、 $m_{Dk,T1} \sim m_{Dk,T5}$ の比較を行い、 m_{min} で使用した心理属性の測定結果が T1~T5 のいずれなのかを特定した。また、全てのデザイン特徴量セット D1 ~ D5 について、 m_{min} で用いられる心理属性測定結果が一致するかについても確認した。

RQ3: Big Five とシュワルツの価値観のクリック・コンバージョンの予測力は独立か

次の 3 タイプのモデルを用意してモデル比較 C3 を実施した: M_{BF+} - Big Five を用いたモデル, M_{SV} - シュワルツの価値観を用いたモデル, $M_{BF\&SV}$ - Big Five とシュワルツの価値観の両方の測定結果 (T6) を用いたモデル. モデル比較の式を以下に示す.

$$m_{min} = \arg \min_{m \in M} IC(m) \quad (C3)$$

where : $M = \{M_{BF+}, M_{SV}, M_{BF\&SV}\}$
 $M_{BF\&SV} = \{m_{Dk, T6}\}$

C1, C2 と同様に, それぞれのデザイン特徴量セット Dk について, $m_{Dk, T1} \sim m_{Dk, T6}$ の比較を行い, m_{min} で使用した心理属性の測定結果が T1~T6 のいずれなのかを特定した. T6 だった場合, Dk については Big Five とシュワルツの価値観の測定結果は, 互いに独立した予測力を持つ成分を含むと言える. また, 全てのデザイン特徴量セット $D1 \sim D5$ について, m_{min} で用いられる心理属性測定結果が一致するかについても確認した.

RQ4: RQ1~RQ3 の答えは広告対象の商材によって変わるか

RQ4 については, 複数の商品/サービスの全データを用いるのではなく, 商品/サービス毎 (PL, SP, EC, PM) のデータを用いてモデルを構築した. そして, 商品/サービスごとのモデルを用いて, RQ1~RQ3 の答えが商品/サービスによって変化するか否かを確認した. モデルの比較方法は各 RQ の答えに依存するため, 次章で説明する.

3.5 モデル比較結果

図 3.4 と図 3.5 に RQ1~RQ3 のモデル比較の結果を示す. 図 3.4, 3.5 はそれぞれ目的変数がクリック, コンバージョンのモデルの比較結果である. 図中, グラフは合計 30 個のモデルの AIC と BIC を示しており, 各モデルはそれぞれ異なる広告デザイン特徴量セットと心理属性測定結果のセットを説明変数として用いている. グラフでは, 同一の広告デザイン特徴量セットを使用する 6 つのモデルをグループ化して示した. 例えば, 図 3.4, 3.5 とともに, 最も左のグループは広告デザイン特徴量セットとして $D1$ を用いたモデルの結果を示しており, 表示されているモデルは左から順に $m_{D1, T1}, m_{D1, T2}, \dots, m_{D1, T6}$ の 6 モデルである. グラフ下の表はモデル比較 C1, C2, C3 の結果を示している. AIC と BIC で C1, C2, C3 の結果が同じであったため, 結果は 1 つの表にまとめた.

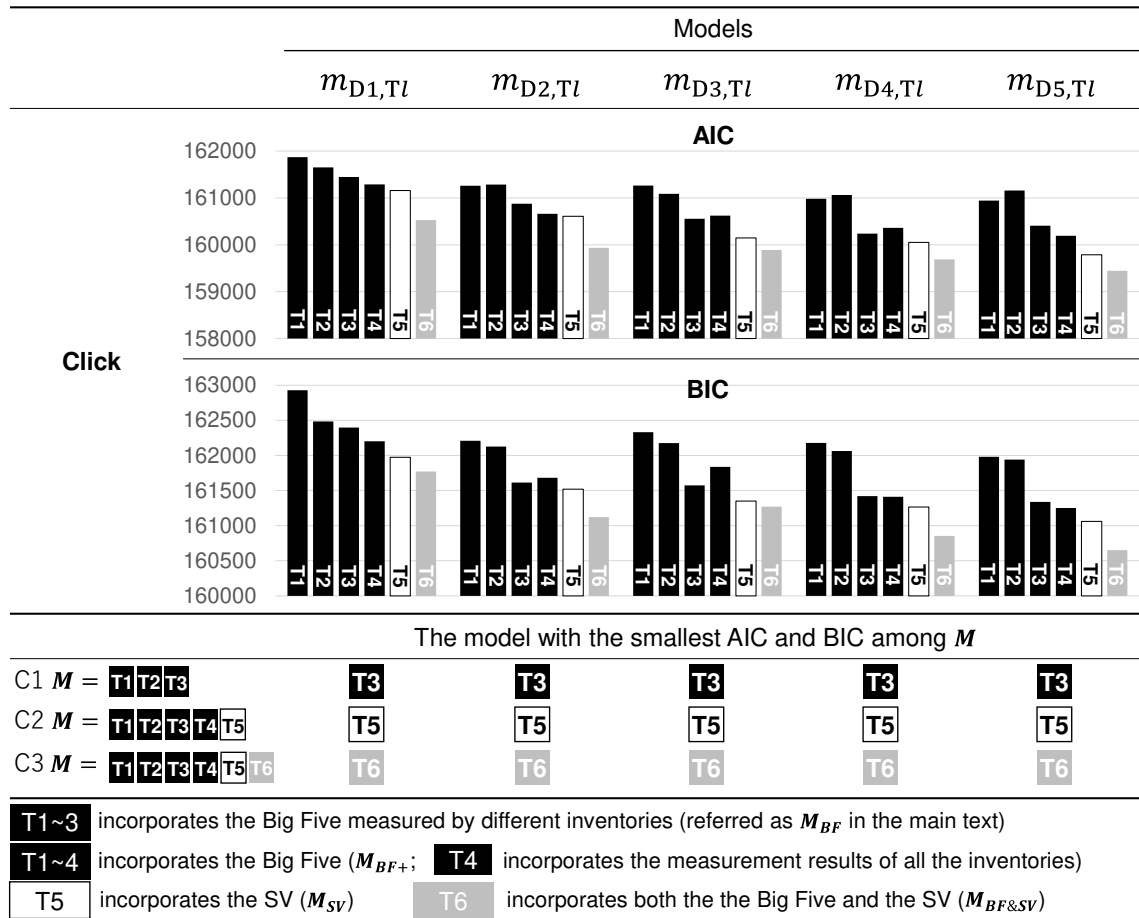


図 3.4 モデル比較の結果（目的変数はクリック）

3.5.1 RQ1: Big Five の質問紙によって、クリック・コンバージョンの予測力は異なるか

モデル比較 C1 の結果, $m_{Dk,T3}$ の AIC, BIC が Big Five の測定結果を用いたモデル $M_{BF} = \{m_{Dk,T1}, m_{Dk,T2}, m_{Dk,T3}\}$ の中で、目的変数や使用する広告デザイン特徴量によらず最小となった。この結果は、質問紙 I1~I3 の中で、I3 で測定した Big Five スコアがクリック・コンバージョンの予測力が最も高いことを意味する。C1 の結果、Big Five の質問紙によって Big Five のスコアが持つクリック・コンバージョンの予測力が異なることが分かった。

以上の結果を踏まえると、Big Five を用いて広告デザインのパーソナライズを行う際には、Big Five の測定に用いる質問紙を選択することが極めて重要であると言える。今回の結果から、広告デザインのパーソナライズに用いるべき Big Five の質問紙の要件をすべて引き出すことは難しいが、十分な質問項目数を持つことは要件の一つであると考え

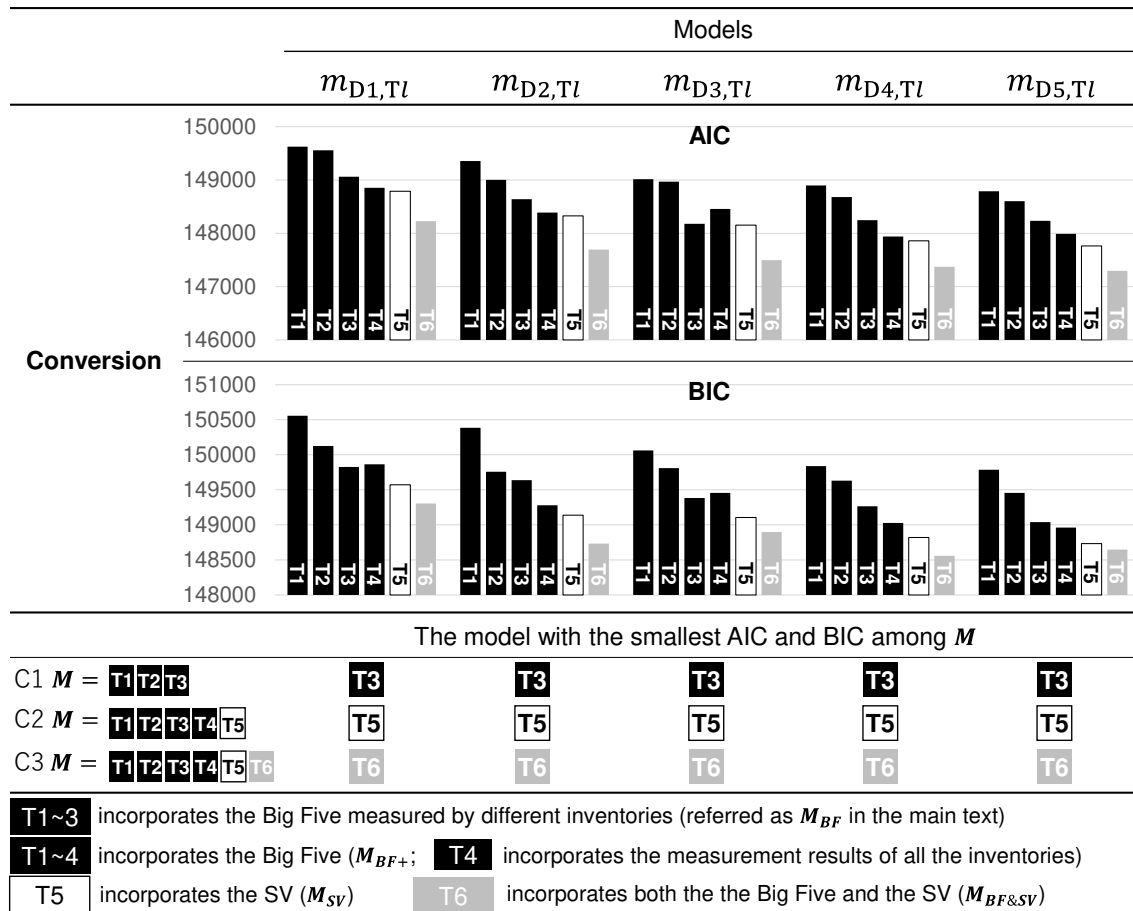


図 3.5 モデル比較の結果（目的変数はコンバージョン）

られる。C1 の結果において、I1（日本語版 TIPI）で測定した Big Five スコアからなる T1 は、ほぼすべての条件下で、I1~I3 の中で広告デザインの調整効果が最も小さかった。その理由の一つとして、I1 がわずか 10 項目の質問で構成され、5 つの特性をそれぞれ 2 つの質問のみで測定している点が挙げられる。Big Five の各特性がさらに複数のサブ特性で構成されとするモデルも提唱されていることを踏まえれば（例えば NEO-PI-R[21] は各特性が 6 つのサブ特性、Facet で構成されるとしている）、各特性の性格領域を測定するのに 2 つの質問では少なすぎる。TIPI の提案者である Gosling は、「少数の質問項目での測定に伴うリスクは必ずしも懸念されているほど大きくない」[48] と述べているが、少なくとも広告デザインのパーソナライズにおいては、このリスクは無視できるものではなく、各特性のスコアの測定は十分な数の質問項目が必要と言える。

一方、I2 と I3 でそれぞれ測定されたスコアからなる T2 と T3 の違いから、広告デザインのパーソナライズに用いるべき質問紙の要件として何を引き出せるかは、あまり明確ではない。I2 と I3 の大きな違いは、回答方法（I2：はい/いいえ、I3：5 件法）と項目の表現方法（I2：文章、I3：形容単語）の 2 つである。前者の違いについては、どちらの回

答尺度にも長所と短所があるため（例えば、I2 には同意バイアス^{*4}のリスクがあり、I3 には中央化バイアス^{*5}のリスクがある）、どちらの尺度が優れていると結論づけるのは難しい。このため、後者の違いが T2 と T3 の違いの原因である可能性が高く、そこから質問紙に対するもう 1 つの要件を引き出せる。すなわち、広告デザインのパーソナライズにおいては、形容詞単語によって質問を表現することが望ましい。ただし、これを確認するためにはさらなる研究が必要である。

3.5.2 RQ2: Big Five とシュワルツの価値観ではどちらの方がクリック・コンバージョンの予測力が高いか

モデル比較 C2 では、I1~I3 の Big Five の質問紙全ての測定結果（T4）を用いたモデルを用意したものの、目的変数や広告デザイン特徴量セットによらず、AIC、BIC が最小となったのはシュワルツの価値観を用いたモデル（ $M_{SV} = \{m_{Dk, T5}\}$ ）であった。この結果は、シュワルツの価値観は Big Five よりも広告デザインから知覚する訴求力に強い影響を与え、したがって個人が知覚する広告訴求力の予測において、シュワルツの価値観の方が Big Five よりも高い予測力を持つことを意味する。

Big Five の測定に用いた質問紙によらず同じ結果となったことから、広告デザインの訴求力予測における Big Five とシュワルツの価値観の違いは、測定方法によるものではなく、性格と価値観という二つの心理属性の本質的な違いによってもたらされたものと考えられる。Schwartz は [124] において、価値観が「個人の行動の基準（Standards or criteria）となるもの」であり、「行動の選択や評価の指針となるもの」であるのに対して、性格は「人が何を重要視するかよりも、その人がどのような人であるか」であり、「ある行動特徴を示す人は対応する目標に価値を見出さないかもしれないし、ある特徴を欠く人は対応する目標に高い価値を見出すかもしれない」（例えば Big Five の誠実性が低く遅刻を繰り返す人物であったとしても、規律を重んじるという価値観が重要だと考えているケースがある）と述べている。こうした違いが今回の結果の主な原因であると考えられる。すなわち、人は広告を見たとき、無意識のうち平素の行動傾向にしたがってクリック/コンバージョンの判断を下すのではなく、自分が何を重要視するか（価値観）を意識的に考慮して、クリック/コンバージョンの判断を下すと考えられる。性格もある程度はこの訴求力知覚に影響を与えるが、その程度は価値観に比べれば小さいと言える。

^{*4} 回答者は、答えが不確かな質問にも肯定的な回答をしがちな傾向がある

^{*5} 3 段階以上の選択肢が示された場合、回答者は真ん中の選択肢を回答しがちな傾向にある

3.5.3 RQ3: Big Five とシュワルツの価値観のクリック・コンバージョンの予測力は独立か

C3 の結果において、広告デザインの特徴量や目的変数に依らず、Big Five とシュワルツの価値観の両方を用いたモデル $M_{BF\&SV} = \{m_{Dk,T6}\}$ が、 M_{BF+} と M_{SV} のどのモデルよりも AIC と BIC が小さかった。この結果から、Big Five とシュワルツの価値観は、広告デザインに対する訴求力知覚に互いに独立して影響を与える成分を含み、したがって Big Five とシュワルツの価値観はクリック・コンバージョンに独立した予測力を持つと言える。すなわち、広告デザインのパーソナライズにおいては、Big Five とシュワルツの価値観の両方を説明変数としてモデルに組み込むことで、どちらか一方を用いたモデルよりも予測力が向上する。

3.5.4 RQ4: RQ1~RQ3 の答えは広告商品/サービスによって変わるか

C1~C3 のモデル比較結果を、ステートメント（命題）S1~S3 として定義し、広告対象の商品/サービスごとに真偽を確認した。単純化のため、前節までと異なり広告デザイン特徴量セットごとにモデルを比較するのではなく、全ての広告デザイン特徴量セットを対象に AIC および BIC が最小となる心理属性セットの特定を行った。例えば、 $m_{D1,T3}$ を $m_{D1,T1}$ や $m_{D1,T2}$ だけでなく $m_{D2,T1}$ や $m_{D2,T2}$ と比較した。表 3.4 に結果を示す。

以下に検証したステートメントとその結果を述べる。 $M_{D,T}$ はモデルのグループを表し、 $M_{D,T} = \{m_{D,T}|D \in \mathbf{D}, T \in \mathbf{T}\}$ である。 $D \in \mathbf{D}$ と $T \in \mathbf{T}$ は広告デザイン特徴量セットと心理属性の測定結果セットを表し、 $M_{D,T}$ は $\mathbf{D}$ と $\mathbf{T}$ の全ての組合せをカバーする。例えば $\mathbf{D} = \{D1, D2\}$ 、 $\mathbf{T} = \{T1, T2\}$ の場合、 $M_{D,T} = \{m_{D1,T1}, m_{D1,T2}, m_{D2,T1}, m_{D2,T2}\}$ である。

- S1 I1~I3 の Big Five の測定結果 (T1~T3) のうち、ユーザ反応 (クリック・コンバージョン) の予測力が最も高いのは I3 による測定結果である。

ステートメント:

$$\arg \min_{m \in M_{D,T}} IC(m) \in M_{D,\{T3\}} \quad (S1)$$

where : $\mathbf{D} = \{D1, \dots, D5\}$, $\mathbf{T} = \{T1, T2, T3\}$

結果: 全ての条件において、S1 が真 (T) となったため、ステートメントは全ての商品/サービスについて有効である。

- S2 シュワルツの価値観のユーザ反応の予測力は Big Five よりも高い。

表 3.4 広告対象の商品/サービスごとのモデル比較結果

ステートメント	AIC/BIC	目的変数	Product/Service			
			PL	SP	EC	PM
S1	AIC	クリック	T	T	T	T
		コンバージョン	T	T	T	T
	BIC	クリック	T	T	T	T
		コンバージョン	T	T	T	T
S2	AIC	クリック	T	T	T	T
		コンバージョン	T	T	T	T
	BIC	クリック	F	T	T	T
		コンバージョン	T	F	F	T
S3	AIC	クリック	T	T	T	T
		コンバージョン	T	T	T	T
	BIC	クリック	F	T	F	T
		コンバージョン	T	T	T	T

ステートメント:

$$\arg \min_{m \in M_{D,T}} IC(m) \in M_{D,\{T5\}} \quad (S2)$$

where : $D = \{D1, \dots, D5\}$, $T = \{T1, \dots, T5\}$

結果: BIC の結果は一部の条件で偽 (F) となったが, AIC の結果は全条件で真となったため, このステートメントはすべての商品/サービスで有効である.

S3 シュワルツの価値観と Big Five のユーザ反応の予測力は独立である.

ステートメント:

$$\arg \min_{m \in M_{D,T}} IC(m) \in M_{D,\{T6\}} \quad (S3)$$

where : $D = \{D1, \dots, D5\}$, $T = \{T1, \dots, T6\}$

結果: S2 と同様, BIC の結果は一部の条件で偽 (F) となったものの, AIC の結果は全条件で真となったため, このステートメントはすべての商品/サービスに対して有効である.

表 3.5 提案モデル

AIC/BIC	目的変数	広告対象商品/サービス				
		All	PL	SP	EC	PM
AIC	クリック	$m_{D5, T7}$	$m_{D4, T6}$	$m_{D3, T6}$	$m_{D4, T6}$	$m_{D3, T6}$
	コンバージョン	$m_{D5, T7}$	$m_{D4, T6}$	$m_{D4, T6}$	$m_{D3, T6}$	$m_{D5, T6}$
BIC	クリック	$m_{D5, T7}$	$m_{D3, T7}$	$m_{D5, T7}$	$m_{D3, T7}$	$m_{D5, T6}$
	コンバージョン	$m_{D5, T7}$	$m_{D3, T7}$	$m_{D3, T7}$	$m_{D4, T7}$	$m_{D3, T6}$

以上のように、S1～S3 の全てのステートメントが商品/サービスに依らず有効であることが確認できた。したがって、S1～S3 を取り入れた予測モデルを構築することで、より効果の高い（より多くのユーザのクリック・コンバージョンを促すことが可能な）広告デザインのパーソナライズが可能となる。

3.5.5 従来研究のモデルとの比較

最後に S1～S3 を取り入れたモデル（提案モデル）とベースラインモデルとの比較を行った。提案モデルについては T1～T6 に加え、I3 で測定した Big Five スコアとシュワルツの価値観スコアの組合せである T7 を用意した。T7 を用意したのは、Big Five スコア I1～I3 の測定結果の組合せである T4 が必ずしも T3（I3 の測定結果）の予測力を上回る結果にならなかったためである。例えば、図 3.4, 3.5 に示したように、広告デザイン特徴量セットとして D3 を用いた場合、クリックとコンバージョン双方において $IC(m_{D3, T4}) > IC(m_{D3, T3})$ であった。

提案モデルは以下に示す式により選択した。選択結果を表 3.5 に示す。

$$\begin{aligned} \text{our model} &= \arg \min_{m \in M_{\mathbf{D}, \mathbf{T}}} IC(m) \\ \text{where : } \mathbf{D} &= \{D3, D4, D5\}, \mathbf{T} = \{T6, T7\} \end{aligned} \quad (3.7)$$

ベースラインモデルは、Matz らの先行研究 [84] を参照して構築した。Matz らは広告デザイン特徴量セットとして、視覚的なデザイン特徴量のみを取り入れ、Big Five も単一の質問紙による測定結果のみを利用した。これを踏まえベースラインモデルとして $m_{D1, T3}$ を採用した。T3 を選択したのは T3 の予測力が T1～T3 の中で最も高かったためである。提案モデルとベースラインモデルの比較結果を表 3.6 に示す。

表に示すとおり、提案モデルは広告対象の商品/サービスや目的変数によらず、ベース

表 3.6 提案モデルとベースラインの比較結果

			広告対象商品/サービス				
	目的変数	モデル	All	PL	SP	EC	PM
AIC	クリック	提案モデル	159,362	39,252	39,491	37,672	42,290
		ベースライン	161,446	39,481	39,804	37,906	42,620
	コンバージョン	提案モデル	147,290	36,170	35,043	34,463	40,945
		ベースライン	149,061	36,472	35,291	34,692	41,271
BIC	クリック	提案モデル	160,645	39,838	39,935	38,163	42,867
		ベースライン	162,396	39,909	40,107	38,335	42,982
	コンバージョン	提案モデル	148,407	36,651	35,436	34,892	41,415
		ベースライン	149,825	36,917	35,602	35,112	41,684

ラインに比べて AIC, BIC とともに低い値となった。つまり、提案モデルはベースラインよりも真のモデルに近く、提案モデルに基づく予測結果を利用した広告デザインのパーソナライズの方がベースラインよりも効果が高いと考えられる。

3.6 広告反応の予測精度評価

本節では、前節までに述べた知見が広告反応の予測精度の向上にどの程度貢献するかを評価する。評価には、3.5 節と同様に、3.3 節で説明したデータセットを用いた。評価は実際の広告運用シナリオを想定して行った。モデル構築に用いるデータの範囲はシナリオによって複数の選択肢がある。例えば、商品 X の広告への反応を予測する場合、X のみのデータを用いてモデルを構築することもできるし、他の商品のデータも用いてモデルを構築することもできる。そこで、予測精度を評価する前に、異なるデータ範囲から構築されたモデルを比較し、どのデータを活用してモデルを構築すべきかを検討した。そして、ベースラインと比較して前節までの知見によって予測精度がどの程度向上するかを評価した。

3.6.1 評価シナリオ

評価は以下のシナリオで行った（図 3.6 参照）。

- I. 広告対象の商品/サービスに関する広告反応データは利用できない。
対象の商品/サービスについて初めて広告を配信するケースを想定した。

- II. 広告対象の商品/サービスの過去の広告反応データは利用できるが、配信予定の広告の広告反応データは利用できない。
ビジネス上の理由で広告デザインを更新するシナリオを想定した（例：同じ広告デザインを長時間使用しており、消費者の関心を惹くために更新する必要がある）。
- III. 配信予定の広告について広告反応データ（テスト配信の結果）を利用できる。
今日の広告配信では、配信対象者全体に広告を配信する前に、対象者のごく一部に対して広告デザインのテスト配信を行うのが一般的である。テスト配信の目的は、最も多くのクリックまたはコンバージョンを生み出す広告デザインを特定することにある。本シナリオでは、そのような状況を想定した。

3.6.2 評価の流れ

評価では、図 3.6 の表に示すセル毎に予測モデルを構築した。クリックとコンバージョンを予測するモデルは別々に構築した。例えば、PL-Gr1 と PL-Gr2 のクリック予測モデルは別々に構築し、PL-Gr1 のクリック予測モデルとコンバージョン予測モデルも別々に構築した。モデルの構築は、モデル構築に利用したデータと特徴量選択を除き、3.4 と同様の手順で実施した。

LASSO による特徴量選択は比較的時間がかかるためモデル毎には行わず、特徴量選択結果を可能な限り再利用した。具体的には、モデル構築に利用するデータの範囲が一つの商品/サービス（シナリオ II-1 と III）であれば、その商品/サービスの全セルを用いて特徴量選択を行い、その結果を再利用してその商品/サービスの全セルのモデルを構築した。また、モデル構築に利用するデータの範囲が複数の商品/サービスにまたがる場合は（シナリオ I, II-2）は、表中の全てのデータ（セル）を用いて特徴量選択を行い、結果を再利用した。

構築したモデルを用いて被験者の広告反応を予測した。図 3.7 に予測の流れを示す。まず、モデルにより被験者と広告デザインのペアごとにクリック/コンバージョンの確率を予測し、被験者ごとにセル内の 5 つの広告デザインの中で最も確率の高い広告デザインを選択する。次に、各被験者のアンケート回答を確認し、選択した広告デザインに対してクリック/コンバージョンを行ったかどうかを判断した。予測精度は以下の数式で定義されるクリック率（CTR）とコンバージョン率（CVR）で評価した。

$$\text{CTR} = \frac{\text{クリック数}}{\text{グループ内の被験者数}} \quad (3.8)$$

$$\text{CVR} = \frac{\text{コンバージョン数}}{\text{グループ内の被験者数}} \quad (3.9)$$

シナリオの説明	モデル構築に使用したデータの範囲	データ範囲の例ー灰色セル (PL-Gr1の広告反応を予測する ケースを想定)																														
I 広告対象の商品/サービスに関する広告反応データは利用できない	広告対象の商品/サービス以外の広告反応データを利用	<table><tr><td></td><td>Gr1</td><td>Gr2</td><td>Gr3</td><td>Gr4</td><td>Gr5</td></tr><tr><td>PL</td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>SP</td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>EC</td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>PM</td><td></td><td></td><td></td><td></td><td></td></tr></table>		Gr1	Gr2	Gr3	Gr4	Gr5	PL						SP						EC						PM					
	Gr1	Gr2	Gr3	Gr4	Gr5																											
PL																																
SP																																
EC																																
PM																																
II 広告対象の商品/サービスの広告反応データは利用できるが、配信予定の広告の広告反応データは利用できない	(II-1) 広告対象の商品/サービスの広告反応データのみを利用する	<table><tr><td></td><td>Gr1</td><td>Gr2</td><td>Gr3</td><td>Gr4</td><td>Gr5</td></tr><tr><td>PL</td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>SP</td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>EC</td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>PM</td><td></td><td></td><td></td><td></td><td></td></tr></table>		Gr1	Gr2	Gr3	Gr4	Gr5	PL						SP						EC						PM					
	Gr1	Gr2	Gr3	Gr4	Gr5																											
PL																																
SP																																
EC																																
PM																																
	(II-2) 上記に加え、他の商品/サービスの広告反応データを全て利用する	<table><tr><td></td><td>Gr1</td><td>Gr2</td><td>Gr3</td><td>Gr4</td><td>Gr5</td></tr><tr><td>PL</td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>SP</td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>EC</td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>PM</td><td></td><td></td><td></td><td></td><td></td></tr></table>		Gr1	Gr2	Gr3	Gr4	Gr5	PL						SP						EC						PM					
	Gr1	Gr2	Gr3	Gr4	Gr5																											
PL																																
SP																																
EC																																
PM																																
III 配信予定の広告について広告反応データ(テスト配信の結果)を利用できる	(III-1) テスト配信の結果のみを利用する	<table><tr><td></td><td>Gr1</td><td>Gr2</td><td>Gr3</td><td>Gr4</td><td>Gr5</td></tr><tr><td>PL</td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>SP</td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>EC</td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>PM</td><td></td><td></td><td></td><td></td><td></td></tr></table>		Gr1	Gr2	Gr3	Gr4	Gr5	PL						SP						EC						PM					
	Gr1	Gr2	Gr3	Gr4	Gr5																											
PL																																
SP																																
EC																																
PM																																
	(III-2) 広告対象の商品/サービスの広告反応データ全てを利用する	<table><tr><td></td><td>Gr1</td><td>Gr2</td><td>Gr3</td><td>Gr4</td><td>Gr5</td></tr><tr><td>PL</td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>SP</td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>EC</td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>PM</td><td></td><td></td><td></td><td></td><td></td></tr></table>		Gr1	Gr2	Gr3	Gr4	Gr5	PL						SP						EC						PM					
	Gr1	Gr2	Gr3	Gr4	Gr5																											
PL																																
SP																																
EC																																
PM																																

図 3.6 広告反応予測の評価シナリオの例

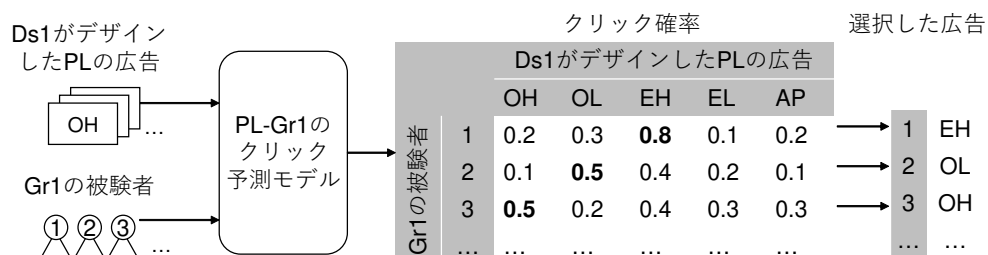


図 3.7 評価における広告反応の予測～広告デザインの選択の流れ

3.6.3 モデル構築に用いたデータ範囲によるモデル比較

図 3.6 に示したとおり，シナリオ II と III については，モデル構築に用いるデータの範囲について二つの選択肢がある．これらを比較するためモデルを構築し CTR と CVR の評価を行った．

シナリオ II

本シナリオでは，広告対象の商品/サービスに関する広告反応データのみを用いるオプション (II-1) と，他の商品/サービスを含め利用可能な全ての広告反応データを用いるオプションの二つが存在する．モデルの構築に当たっては表 3.5 を参照し，モデルに使用する広告デザイン特徴量セットと心理属性セットを選択した．II-1 については，各商品/サービスの結果を参照し，最良の AIC を観測したモデルを元を選択した（例えば PL の広告のクリックを予測するケースでは， $m_{D4,T6}$ が最良のモデルなので，D4 と T6 を選択した）．また，II-2 については，“All” の列を参照し，最良の AIC となったモデルを選択した（例えばクリックを予測する場合，最良の AIC となったのは $m_{D5,T7}$ のため，D5 と T7 を選択した）．表 3.7 に II-2 をベースとした II-1 の CTR と CVR のリフトを示す．

結果から，シナリオ II においては，II-1 の方が優れた選択肢であることが分かる．平均リフトもリフトの 90% 信頼区間もともに 1.00 を上回っていた．したがって，広告対象の商品/サービスについて十分な広告反応データが確保できる場合，モデルは当該データのみを利用して構築すべきだと言える．

シナリオ III

本シナリオでは，配信予定の広告デザインのテスト配信結果のみを利用するオプション (III-1) と，広告対象の商品/サービスの全ての広告デザインの広告反応データを利用するオプション (III-2) を取り得る．

これらのオプションを比較するため，まずターゲットとなる被験者グループ（図 3.6 内の表のセル，例えば Gr1）をランダムに 5 分割し，そのうちの 1 つを評価サブグループ，他の 4 グループを訓練サブグループとし，訓練サブグループにテスト配信が行われたと仮定した．すなわち訓練サブグループの広告反応データを訓練データとしてモデル構築に利用し，評価サブグループの広告反応データを評価に用いた．これを，評価サブグループを変えて 5 回繰り返し，平均 CTR と CVR を評価した．表 3.8 に III-2 をベースとした II-1 の CTR と CVR のリフトを示す．

表に示したとおり，全ての被験者グループ，商品/サービスにおいて，III-1 と III-2 に有意な差は見られなかった．すなわち，配信予定の広告デザインのテスト配信データが利用

表 3.7 II-2 をベースとした II-1 の CTR と CVR のリフト

CTR lift					
	Gr1	Gr2	Gr3	Gr4	Gr5
PL	1.14*	0.97	1.00	1.20**	1.26**
SP	1.08	1.35**	0.85**	0.98	1.07
EC	0.99	1.29**	1.00	0.90	0.90
PM	1.24**	1.05	1.02	1.13**	0.99
Average lift: 1.07 (90% confidence interval: 1.02~1.12)					
CVR lift					
PL	0.98	1.14*	1.01	1.00	1.05
SP	1.31**	1.40**	1.27**	1.08	0.99
EC	1.04	1.00	1.00	0.91	1.10
PM	1.36**	1.04	1.18**	0.92	0.71**
Average lift: 1.08 (90% confidence interval: 1.01~1.14)					

*: $p < .05$, **: $p < .01$

可能な状況では、他の広告デザインの広告反応データを用いても広告反応の予測精度の向上には寄与しないと言える。

3.6.4 ベースラインとの比較

提案モデルと二つのベースラインを比較した。ベースライン 1 は Matz ら [84] が検討したモデルを元にしたモデル (3.5.5 節で評価したベースラインと同じモデル) である。ベースライン 2 は 3.6.1 で述べた今日の広告配信で一般的に利用されるモデルである。

ベースライン 1 との比較

3.5.5 で述べたように提案モデルはベースライン 1 よりも小さな AIC, BIC であることが分かっているため、AIC, BIC の差分が CTR, CVR の改善にどの程度反映されるかを確認する目的で評価を行った。この目的を踏まえ、最もシンプルなシナリオ、すなわちシナリオ I で評価を行った。3.5.5 と同様に、提案モデルの広告デザイン特徴量セットと心理属性セットは表 3.6 に示したものをを用い (“All” 列に示したものをを用いた)、ベースライン 1 については、 $m_{D1, T3}$ を採用した。表 3.9 に結果を示す。

表 3.8 III-2 をベースとした III-1 の CTR と CVR のリフト

CTR lift					
	Gr1	Gr2	Gr3	Gr4	Gr5
PL	0.98	0.99	0.99	1.03	1.03
SP	1.00	1.00	1.00	1.03	1.00
EC	1.00	1.01	1.01	1.02	1.01
PM	1.00	1.01	1.00	1.01	0.99
Average lift: 1.01 (90% confidence interval: 1.00~1.01)					
CVR lift					
PL	1.07	0.99	0.98	1.00	1.01
SP	1.00	1.00	0.99	0.98	0.99
EC	1.00	1.04	0.98	1.03	1.04
PM	0.99	1.00	1.04	0.99	1.00
Average lift: 1.01 (90% confidence interval: 1.00~1.01)					

*: $p < .05$, **: $p < .01$

クリックについては、平均リフトおよびリフトの信頼区間がともに 1.00 を上回った。加えて、有意な正のリフトが確認できたセルの数は、ベースラインに比べて有意に CTR が低下したセル（すなわち有意な負のリフトを観測したセル）の数を上回った。これらの結果から、提案モデルを用いて広告配信を行うことでベースライン 1 よりも高い CTR を期待できる（ベースライン 1 よりも高い広告デザインのパーソナライズ効果を得られる）。コンバージョンについては、クリックよりも多くのセルで負のリフトが観測されたものの、提案モデルによる平均リフトは 1.00 を上回った。したがって、提案モデルはコンバージョンについてもベースラインよりも良好な結果（高い CVR）を期待できる。

ベースライン 2 との比較

ベースライン 2 との比較は、シナリオ III で実施した。ベースライン 2 はシナリオ III のみ使用可能なためである。

比較を行う前に、図 3.6 に示した表の各セル（被験者グループ/デザイナーと商品/サービスの組合せ）から、CTR/CVR の差分が最も小さい二つの広告デザインを選択し、これら二つの広告デザインを用いて評価を行った。仮に二つの広告デザインの差が比較的大きい場合、提案モデルにおいても大多数の被験者に対して CTR/CVR が高い広告デザイ

表 3.9 ベースライン 1 からの提案モデルの CTR, CVR のリフト

CTR lift					
	Gr1	Gr2	Gr3	Gr4	Gr5
PL	1.08	1.00	1.00	1.14*	1.00
SP	0.88	0.78**	1.40**	1.06	0.94
EC	1.07	1.70**	1.25**	1.00	0.91
PM	1.50**	1.00	1.05	1.82**	1.15**
Average lift: 1.14 (90% confidence interval: 1.03~1.24)					
CVR lift					
PL	1.09	0.94	0.86*	1.09	0.84**
SP	0.74**	0.83*	0.85*	0.53**	1.24**
EC	1.49**	2.50**	1.93**	0.97	1.03
PM	1.38**	0.99	0.78**	2.19**	1.03
Average lift: 1.16 (90% confidence interval: 0.97~1.36)					

*: $p < .05$, **: $p < .01$

ンを選択すると想定したためである（例えば広告デザイン A,B について、ほとんど消費者にとって A の広告デザインの方が優れ、B が劣っている場合、パーソナライズの効果は出にくい）。すなわち、提案モデルは評価の分かれる複数の広告デザインが配信候補として存在する場合に有効であると想定した。実際の広告配信においても、配信者は複数の広告デザインについてテスト配信の結果を用いて、配信する候補を少数に絞り込んでいく。提案モデルはこのように絞り込まれた広告デザインに対して適用されるべきだと考えた。

3.6.3 節と同様に、セルにおける広告反応データを評価データと訓練データに分割し評価を行った。3.6.3 節の結果を踏まえ、予測モデルは III-1 により構築し、III 節と同じ広告デザイン特徴量セットと心理属性セットを用いた。ベースラインの評価については、訓練データを用いて最良の広告デザインを選択した。結果を表 3.10 に示す。

表に示すとおり、提案モデルによって CTR/CVR が有意に上昇したセルはなかった。信頼区間は CTR と CVR の両方で 1.00 を超えたが、ベースライン 2 からのリフトの平均は小さい。この結果から提案モデルがベースライン 2 よりも多くのクリックまたはコンバージョンをもたらすとは言えない。

他方で、リフトと評価に用いた二つの広告デザインの CTR/CVR の差には有意な負の相関があった（CTR は $r = -.57, p < .01$, CVR は $r = -.50, p < .05$ ）。これは、配信候

表 3.10 ベースライン 2 からの提案モデルの CTR/CVR のリフトと評価に用いた二つの広告デザインの CTR/CVR の差

CTR lift					
	Gr1	Gr2	Gr3	Gr4	Gr5
PL	1.04	1.02	1.03	0.98	1.04
	1.3%	1.9%	0.1%	2.0%	0.4%
SP	1.00	1.05	1.10	1.01	1.02
	1.5%	0.7%	0.5%	2.0%	1.4%
EC	0.99	1.02	1.07	1.04	0.97
	0.9%	0.8%	0.7%	0.1%	1.1%
PM	1.00	0.98	0.99	0.99	1.00
	2.9%	2.6%	1.8%	4.4%	2.2%
Average lift: 1.02 (90% confidence interval: 1.00~1.03)					
CVR lift					
PL	1.11	0.98	1.03	1.02	1.03
	0.2%	1.2%	0.1%	0.5%	0.2%
SP	0.98	1.06	1.04	1.05	1.06
	0.5%	0.6%	1.4%	0.7%	0.1%
EC	0.96	1.06	1.03	1.06	0.98
	1.4%	0.5%	1.7%	0.2%	2.1%
PM	0.99	0.99	1.01	1.00	0.98
	4.1%	1.9%	0.2%	3.9%	0.8%
Average lift: 1.02 (90% confidence interval: 1.01~1.04)					

*: $p < .05$, **: $p < .01$

補の広告デザインの CTR/CVR が近づくにつれて、ベースライン 2 からの正のリフトが提案モデルにおいて期待できることを示唆する。さらに、有意な負のリフトとなるセルが存在しないことから、負のリフトのリスクなしに正のリフトを期待できる。したがって、ベースライン 2 からの提案モデルによる正のリフトの可能性が示されたと言える。

3.7 考察

3.7.1 実際の広告配信における知見の有用性

本章の評価結果はアンケート調査のデータを使用して得られたものである。人々は必ずしもアンケートの回答と一致した行動をとるとは限らないので、アンケート調査結果の「クリック」や「コンバージョン」は、被験者が実際の環境でクリックやコンバージョンを行うことを保証するものではない。しかし、結果を統計的に分析した場合、被験者の回答は広告がどれだけ魅力的であるか（つまり、広告がクリックやコンバージョンを誘発する可能性がどれだけ高いか）を反映していると考えられる。

本章の研究と同様に、広告デザインの訴求力をアンケートで評価することは学術研究でもマーケティングビジネスでも用いられている一般的な手法である。例えば、Hirsh ら [54] や Matz ら [84] は、広告デザインの訴求力を評価するためにアンケート調査を行っている。Hirsh らは複数の広告デザインに関するアンケート回答から特定の性格タイプに最も訴求力のあるデザインを抽出し、また、Matz らはアンケート結果を用いて、広告デザインの訴求力を予測する複数のモデルを構築・評価している。

同様に、ビジネスにおいてもマーケティング企業がアンケート調査を利用した「広告テスト」サービスを提供している。これは、実際の広告キャンペーンに備えて、潜在的な顧客にアンケート調査を行い、候補の中から魅力的な広告デザインを特定するものである*6。

以上を踏まえると、本章で述べた実験におけるクリックやコンバージョンが、必ずしも実際のクリックやコンバージョンを意味するわけではないものの、アンケート調査で観測されるクリックやコンバージョンの量は広告デザインがいかに魅力的であるかを反映するものであり、このようなデータによって導出された知見は、実世界の広告キャンペーンにも適用可能であると考えられる。

3.7.2 知見の有用性が特に期待できる広告配信シナリオ

これまで述べてきたように、本章で述べた知見は、個人が知覚する広告訴求力の予測精度向上に寄与する。すなわち実世界における広告キャンペーンの費用対効果の向上に寄与する。同一の広告デザインを対象者全員に配信する、一般的な広告配信とは対照的に、本章の研究では、Matz ら [84] や筆者らの過去の研究 [59] で想定した広告デザインのパー

*6 例: <https://www.qualtrics.com/experience-management/brand/how-to-run-a-successful-ad-testing-program/>, <https://marketlensresearch.com/solutions/creative-testing>, <https://blog.constructionmarketingassociation.org/practical-tips-for-advertising-testing/>

ソナライズを対象とした。すなわち、(i) 企業が同じ商品やサービスに対して異なるデザインの広告を複数用意し、(ii) 用意された広告と広告配信先のペアごとに広告配信先の個人が知覚する広告訴求力を予測し、(iii) 最も訴求力の高い広告デザインを各個人に配信する。このようなパーソナライズされた広告配信では、各個人が知覚するであろう広告訴求力の予測精度が非常に重要である。予測精度が高いほど、配信された広告が所望のユーザ反応を誘発する可能性が高くなる。本章の研究で得た知見をステップ (ii) に適用することで、企業は予測精度を向上でき、その結果、同じ広告配信数で（つまり同じコストで）より多くのクリックやコンバージョンを期待できる。

また、本章の研究における知見は、上述の広告デザインのパーソナライズに加え、以下に述べる今日一般的な広告配信シナリオにおける費用対効果の改善にも貢献できる。すなわち企業は複数の広告デザイン候補を用意し、テスト配信を行う。テスト配信では、広告デザイン候補を一部の対象者に配信し、最も訴求力のある広告デザインを特定する。その後、特定した広告デザインを全対象者に配信する。テスト配信は一部の対象者に対してのみ行われるがコストがかかる。特に、候補となる広告デザインが多数あり、候補を絞り込むためにテストを複数回実施する必要がある場合は顕著である。また、テスト配信を実施する時間がないケースもある（競合他社が新商品・新サービスをリリースし、自社商品・サービスの訴求力を維持するために早急に広告を配信して対抗する必要があるケースなど）。企業が各広告デザイン候補の訴求力を十分正確に予測できれば、テスト配信を省略でき、コストと時間を節約できる。広告訴求力の予測に本章の知見を応用することで、予測精度が向上し、絞り込まれた候補の中に最も訴求力の高い広告が含まれる可能性が高くなる。

3.7.3 課題と今後の研究の方向性

本章の研究には今後の研究で解決すべき課題が複数ある。第一の課題は、本章の知見が質問紙によって測定された心理属性を用いて得られている点である。質問紙測定法は、心理属性の測定に最も広く使われている方法であるが、実際の広告配信では配信対象者に質問紙調査を行うことは、コストの観点で現実的ではないし、質問紙調査に参加し回答するのは配信対象者のごく一部に限られる可能性が高い。したがって、本章の発見を現実のシナリオに適用するためには配信対象者の心理属性を、質問紙調査を行うことなく把握する必要がある。すなわち配信対象者の日常行動や生体データなどから心理属性を推定する技術が必要となる。Big Five については推定技術について多くの研究が行われており、ソーシャルメディアの利用ログ [144, 126] や、スマートフォンの利用履歴 [16, 24, 134]、自動車の運転ログ [59] などから推定する方式が検討されている。

しかし、このような推定手法の限界について言及する研究もある。Roberts ら [117] は、

人々の性格と日常行動の関係の強さは 0.30 から 0.40（ピアソンの相関関係）であると主張している。また、Azucar ら [5] は、複数の研究にまたがるメタ分析を行った結果、ソーシャルメディア上の行動によって推定された Big Five スコアと、質問紙によって測定されたスコアとの間には、弱い相関から中程度の相関（0.29 から 0.40 の範囲）しかなかったと報告している。こうした心理属性の推定の難しさを踏まえると、今後の研究ではこれらの手法で推定された心理属性を用いて本研究の知見を検証することが必要となる。

第二の課題は、3.5.1 節において、広告デザインのパーソナライズで使用されるべき Big Five の質問紙の要件について仮説を提起したが、調査した質問紙の数が限られていたため、この仮説を検証できなかったことである。今後さらに研究を進め、仮説を検証するとともに他の要件を探索することが期待される。

第三の課題は、調査した広告デザイン特徴量が限定的であったことである。バナー広告から抽出可能な多様なデザイン特徴量のうち本章では検討しなかったものの、広告訴求力に影響を与える可能性があるデザイン特徴量が存在する。視覚的デザインの特徴量について、本章では Segalin ら [126] が検討したデザイン特徴量を用いた。同研究において、Segalin らは画像から当該画像を好む人物の Big Five を推定するための特徴量を抽出する手法を検討している。しかし、こうした目的で画像から特徴量を抽出する方法は他にも複数検討されている。例えば Ferwada ら [36, 37] は Instagram に投稿された写真から投稿した人物の Big Five を推定するための特徴量を抽出する手法を検討しており、その中には写真に写っている人物や建物など、本章の研究では検討しなかった特徴量が含まれている。こうした特徴量が広告の訴求力に影響している可能性がある。

また、深層学習技術の普及により、近年では画像から多様な視覚的特徴量を自動抽出できるようになっている。これらの特徴量は人間が解釈できるものではないが、写真に写っている人物の検出や表情認識など様々なタスクで有効性が実証されているため、このような特徴量を予測モデルに組み込むことで、広告訴求力の予測精度が向上する可能性がある。言語デザインの特徴量についても同様である。本章では、広告の言語デザイン特徴量を抽出するために単語ベクトルを使用した。文や文書全体から特徴を抽出できる深層学習技術もある。こうした技術を利用することで、広告訴求力の予測精度を向上できる可能性がある。

3.8 結論

本章では、広告デザインのパーソナライズを対象に、心理属性や質問紙によるサービスパーソナライズ効果の違いについて検討した。本章で述べた研究による主な貢献は以下のとおりである。

1. 個人が広告デザインから知覚する広告訴求力には、視覚的なデザイン、言語的なデザイン双方とも Big Five よりもシュワルツの価値観の方が大きな影響を与えることを示した。加えて、Big Five とシュワルツの価値観は個人の広告訴求力の知覚に互いに独立した影響を与える成分を含むことを示した。広告デザインのパーソナライズを行う際には、Big Five とシュワルツの価値観の双方を用いてクリックやコンバージョンの予測を行うことで、単独で使用するよりも高いパーソナライズの効果を得られる。
2. 複数の Big Five 質問紙について調査を行い、異なる質問紙による測定結果は個人の広告訴求力の知覚について有意に異なる予測力を持つことを示した。また、視覚的なデザイン、言語的なデザインを問わず、同じ質問紙が広告訴求力知覚に最も高い予測力を持つことを確認した。こうした質問紙が備える条件について検討を行い、各特性について十分な数の質問項目を備えていること、また質問を形容詞単語で表現すること、の2条件が要件であるとの仮説を提起した。
3. 上記の知見が異なる複数の広告対象商品/サービスで有効であることを確認し、本章の研究で得られた知見は、幅広い商品/サービスを対象とした広告のデザインパーソナライズに活用できる可能性を示した。
4. 上記の知見を取り入れた予測モデル（提案モデル）は、従来研究で検討されてきたベースラインよりも、広告反応の予測精度が高く、現在一般的に利用されている広告配信アプローチに比べ、クリック率やコンバージョン率を改善できる可能性があることを示した。

第 4 章

購買行動データを用いた心理属性の推定

4.1 背景と目的

オンラインサービスを提供する企業にとって、ユーザ属性の学習は欠かせない。ユーザ属性は個々のユーザ向けにサービスをパーソナライズするための基礎となり、ユーザと良好な関係を築くための鍵となる。アマゾンやアリババに代表される EC ポータルは、衣料品から食品、家電製品、音楽・動画配信から写真保存まで、日常生活の様々なドメインをカバーする多様な商品・サービスを扱っている。サービス利用者数は 2014 年に 10 億人を突破し^{*1}、COVID-19 の世界的な感染拡大により、その成長スピードは加速している [1]。

様々なユーザ属性の中でも、年齢、性別、性格など、次の 2 つの特徴を持つ属性を学習することは、EC ポータルにとって特に有益である。1 つ目は、「ドメイン非依存性」である。この特徴を持つ属性は、日常生活の様々なドメインにわたって人々の行動に影響を与える。例えば、一般的に年齢や性別によってライフスタイルが異なるため、食事や衣服、音楽、映画などに対するニーズや嗜好が異なり、購買行動に影響を与える。同様に、既存の研究で、人の性格が食品 [128]、ブランド [93, 152]、音楽 [58]、映画 [44] など多岐にわたる嗜好に影響することが確認されている。したがって、これらのドメイン非依存の属性は、EC ポータルがカバーする様々なドメインでサービスをパーソナライズするために使用できる。ドメイン非依存な属性をいったん学習すれば、ユーザが過去にそのドメインのアイテムを閲覧したり購入したりしたことがなくても、そのドメインのアイテムを推薦するために使用できる。ドメイン非依存な属性が持つこのようなドメインに対する適応性 (Domain adaptability) は特定のドメインでの購買行動に影響を与えるドメイン固有の

^{*1} <https://www.statista.com/statistics/251666/number-of-digital-buyers-worldwide/>

属性との大きな違いである。2つ目は、「安定性」である。属性が短期間で大きく変化することがないことを意味する。

従来、研究者はユーザ属性の表現（User representation, 以下 UR）の学習を目的として、多数の方式を研究してきた [74, 70, 115, 34, 75, 140, 63, 155, 78, 112, 49, 114]。これらの方式は、ユーザ属性の Ground truth を用いることなく、ユーザとアイテムのインタラクション（ユーザによるアイテムの閲覧や購入、レビューなど）からユーザ属性の高次元ベクトル表現（UR）を学習する。ユーザに属性を手動登録するよう求める方法（例えば、EC ポータルにサインアップする際に年齢や性別、興味のある商品ジャンルを登録するよう求める）では、限られた情報量しか収集できず、かつ一部のユーザが意図的／非意図的に虚偽の情報を登録することが避けられない。一方で、自動学習を行う方式であれば、EC ポータルを含むサービスプロバイダがユーザ属性に関するより豊富で信頼性の高い情報を学習することが可能となる。

従来研究は、前述の2つの重要な特徴である「ドメイン非依存性」と「安定性」のうち、後者の特徴を持つ UR を学習することに焦点が当てていた。ユーザ-アイテムのインタラクションの時系列データを活用するアプローチ（例えば RNN を用いたアプローチ）[34, 114, 140, 78, 63, 49, 155, 112] は、短期的なユーザ属性と長期的なユーザ属性を区別して学習できる。一方で、ドメイン非依存性は、様々なドメインでサービスをパーソナライズする上で重要であるにもかかわらず、従来研究ではほとんど注目されてこなかった。従来方式により学習した UR は、UR を学習したドメインに固有のユーザ属性のみを反映するか、ドメイン固有の属性とドメイン非依存の属性を区別しない状態で反映するかのどちらかであった。このため、従来方式で学習した UR では、UR の学習元のドメインとは異なるドメインでサービスを効果的にパーソナライズできなかった。

以上を踏まえ、本章ではドメイン非依存かつ長期的に安定するユーザ属性を表現する UR を、ユーザとアイテムのインタラクションデータから学習する方式を提案する。既存の UR 学習方式と同様に、提案方式はユーザ-アイテムの時系列インタラクションデータから教師無しで UR を学習する。提案方式では、ユーザ属性を長期安定/短期変動、ドメイン固有/ドメイン非依存、の2軸で区別し、互いに異なる4種類のユーザ属性を学習する。RNN を採用することで、長期属性と短期属性を別々に学習する。提案方式の新規な点は、ドメイン固有とドメイン非依存の属性を区別して学習するためにマルチレイヤの RNN を導入した点である。提案方式の RNN は、個々のドメインにおけるユーザの行動をモデリングする複数の Domain specific layer (DSL) とドメイン非依存の属性をモデリングする単一の Domain independent layer (DIL) で構成される。提案方式の RNN は、ユーザが、例えば衣服、食料品、家電などの複数のドメインで、閲覧、購入、レビューなどのインタラクションを行ったアイテムの時系列データを入力として受け取る。それぞれの DSL は対応するドメインにおいてユーザが過去インタラクションしたアイテムの情

報を受け取り（例えば DSL1 は衣服, DSL2 は食料品, DSL3 は家電）, ドメイン内のアイテムに対するユーザの将来のインタラクションの意図（例えば DSL1 の場合, 個々の衣服アイテムに対する購入意図）を隠れ状態に反映する. DSL は, 隠れ状態を更新する際, DSL の隠れ状態だけでなく, DIL の隠れ状態を使用する. 全ての DSL が隠れ状態の更新に DIL の各状態を用いるため, DSL が対応する全てのドメインにおいて, DIL の隠れ状態がユーザの意図に影響を与えることになる. このプロセスによって, DIL は複数のドメインにまたがるユーザの行動に影響を与える属性, すなわちドメイン非依存のユーザ属性を学習できる.

EC ポータルで収集したデータを格納した 2 つのオープンデータセット, Amazon (Am) [100] と RetailRocket (RR) [116] を用いて UR を学習し, ドメイン非依存性の評価を行った. UR がドメイン非依存である場合, 1) UR の学習を行ったドメインに依らず, 同一ユーザの UR は類似する, 2) UR の学習元とは異なるドメインでの行動予測に有用である, と考えられる. これを検証するため, 以下のリサーチクエスチョン (RQ) を設定し, 実験を行った.

Q1 異なるドメインで学習した同一ユーザの UR はどの程度似ているか

Q2 UR の学習元とは異なるドメインにおける行動予測に, UR はどの程度有用か

UR のドメイン非依存性に関する RQ に加え, ドメイン非依存なユーザ属性である性格に焦点を当て, UR がこれをどの程度反映するか評価するための RQ を設定した.

Q3 UR はユーザの性格をどの程度反映しているか

データセットにはユーザの性格を直接示す情報（例えば Big Five スコア）は含まれていなかったが, データセットの 1 つ (Am) にはユーザが書いたアイテムレビューの文章が含まれていた. 既存研究で, ユーザが書いた文章から抽出した特徴量は, ユーザの Big Five と有意な相関があることが確認されている. そこで, 性格と有意な相関を持つテキスト特徴量を UR がどの程度正確に予測できるかを評価することで, UR が性格をどの程度反映しているか検討できると考えた.

4.2 関連研究

既存の UR 学習方式は, Non-sequential と Sequential な方式に分類できる [74]. Matrix factorization (MF) [70] は Non-sequential の代表的な方式であり [74], ユーザとアイテムのインタラクションを表現する行列をユーザ行列とアイテム行列に分解することで UR を学習する. MF がユーザとアイテムの ID 情報しか利用できないのに対して, Factorization machine (FM) [115] はユーザとアイテムの付随情報（例えば, ユーザのデ

バイスタップや映画のジャンルなどの情報)も利用できる。また、FM をさらに発展させた方式も考案されている。FM が考慮できるのは入力情報の二次の相互作用のみ(例えば ユーザ ID とアイテム ID, ユーザ ID とアイテムのジャンル, など)だったのに対して、より高次の相互作用(例えば ユーザ ID とアイテム ID とアイテムのジャンル, など)を取り込めるようにしたアプローチがいくつか提案されている [75, 15, 50, 113, 147]。その中でも、xDeepFM [75] は他の方式と異なり、Compressed Interaction Network (CIN) を導入して、高次の相互作用を明示的にモデリングできる。ここまでで述べた Non-sequential な方式は、その単純さと有効性から多くの商用システムに広く導入されている。しかし、ユーザ属性を長期的/短期的な属性に区別して学習することはできない。

一方で、Sequential な方式では、ユーザがインタラクションしたアイテムのシーケンスから、長期的/短期的なユーザ属性を明示的に反映した UR を学習できる。このカテゴリの方式の一部は、RNN を使用する。提案方式もこのカテゴリに属し、具体的には、Sequential user-based RNN (SURNN) [34] が提案方式のベースとなっている。提案方式のマルチレイヤ構造とは対照的に、SURNN は RNN セルが持つユーザ行列に UR が格納される単層の RNN で構成される。RNN セルにアイテムデータ(例えば、視聴した映画の ID や購入した服の ID)が入力されると、RNN セルはユーザ行列から UR を取り出し、入力されたアイテムデータと共に用いて隠れ状態を更新する。隠れ状態は入力されたアイテムデータと共に逐次更新されるが、ユーザ行列(つまり UR)は変わらない。これにより、長期的な属性を UR に、短期的な属性を隠れ状態に、それぞれ反映させることが可能となる。

他の Sequential な方式として、複数の研究者が検討を進めているカテゴリとして、Pretrain-finetune のコンセプトを UR 学習に取り入れた Universal UR learning がある [49, 112, 130, 138, 101, 155]。このカテゴリの方式は、時系列のユーザ-アイテムのインタラクションデータを用いて、様々な目的(例えばアイテムの推薦やユーザのプロファイリング)で利用可能な UR を Pretrain しておき、個別のタスクで利用する際に UR を Finetune する。特定のタスクのために UR を学習する他の方式(例えば SURNN など)とは対照的な方式である。このカテゴリのうち、Self-supervised User Modeling Network (SUMN) [49] は、テキストで表現されたユーザ-アイテムのインタラクションデータ(ユーザが購入した商品の名称やユーザの検索ログなど)を用いて、そこからアテンション機構によって UR を抽出する。過去の行動から抽出した UR と将来の行動パターンを比較し、両者の損失が最小になるようにモデルを学習することで、長期的な属性表現を得られる。同じ目的で、U-BERT[112] はレビューテキストを、AutoEncoder-coupled transformer network (AETN) [155] は携帯電話のアプリケーションのインストール・アンインストールログを利用する。

Sequential なアプローチは、長期的/短期的なユーザ属性を区別して学習できるが、筆

者の知る限り，ドメイン非依存の属性とドメイン固有の属性を区別できる方式は存在しない．上記の全ての方式は，複数のドメインにまたがるインタラクショndataを入力できるが，そうしたデータから学習した UR はドメイン非依存のユーザ属性だけでなく，ドメイン固有の属性も反映している可能性が高い．ドメイン固有の属性は，UR の学習元ではない他のドメインでの行動予測にはほとんど役に立たない．そのため，そのようなドメインではサービスパーソナライズの有効性が低下する．

4.3 提案方式

図 4.1 に提案方式の概要を示す．図の左側は提案方式のマルチレイヤ RNN の例を示しており，図の右側には 4 種類のユーザ属性がマルチレイヤ RNN の変数にどのようにマッピングされているかを示している．提案した RNN において，DSL は個々のドメインに対応する．例えば，入力データに 2 つのドメインのアイテムが含まれている場合には図に示すように，対応する DSL が 2 つ存在することになる．各 DSL は，対応するドメインにおけるユーザ-アイテムのインタラクションの時系列データを入力データとして受け取る．それぞれの入力データはユーザ ID とアイテム ID の 1 hot ベクトルの組で表される．DSL の RNN セル（DSL セル）はこの入力を用いて，ユーザ行列とアイテム行列からそれぞれ UR と Item representation を取り出し，これらの Representation を用いて隠れ状態を更新する．隠れ状態は入力アイテムによって逐次更新されるが，ユーザ行列（UR）は変わらない．これにより，DSL は短期的なユーザ属性を隠れ状態で表現し，長期的なユーザ属性を UR で表現できる．また，DSL はこれらの属性に対応するドメインにおけるユーザの行動のみから学習するため，図右側の象限 I と II に示すように，DSL はドメイン固有のユーザ属性を学習することになる．

上記のフローは SURNN と同様のものであるが，DSL が SURNN と異なる点は，図左側のリンク (A) に示すように，隠れ状態の更新に，自身の隠れ状態だけでなく，DIL の隠れ状態も用いる点である．更新された隠れ状態は，対応するドメインにおける次のアイテムを予測するのに使われるため，ドメインにおける行動履歴を表すと同時に次の行動に対するユーザの意図を表現しているとみなせる．このような DSL の隠れ状態を DIL の隠れ状態を用いて更新することは，DIL の隠れ状態が個々のドメインにおけるユーザの行動に影響を与えることを意味する．すべての DSL がこのように隠れ状態を更新するため，DIL は複数のドメインにまたがってユーザの行動に影響を与えるユーザ属性，すなわちドメイン非依存のユーザ属性（象限 III と IV）を学習することになる．DSL と同様に，DIL もその隠れ状態に短期的なユーザ属性を，UR に長期的なユーザ属性を反映するが，DIL はドメイン非依存の属性を学習するという点で DSL と異なる．

なお，図に示す通り提案方式では，上述のリンク (A) に加えてリンク (B) も追加した．

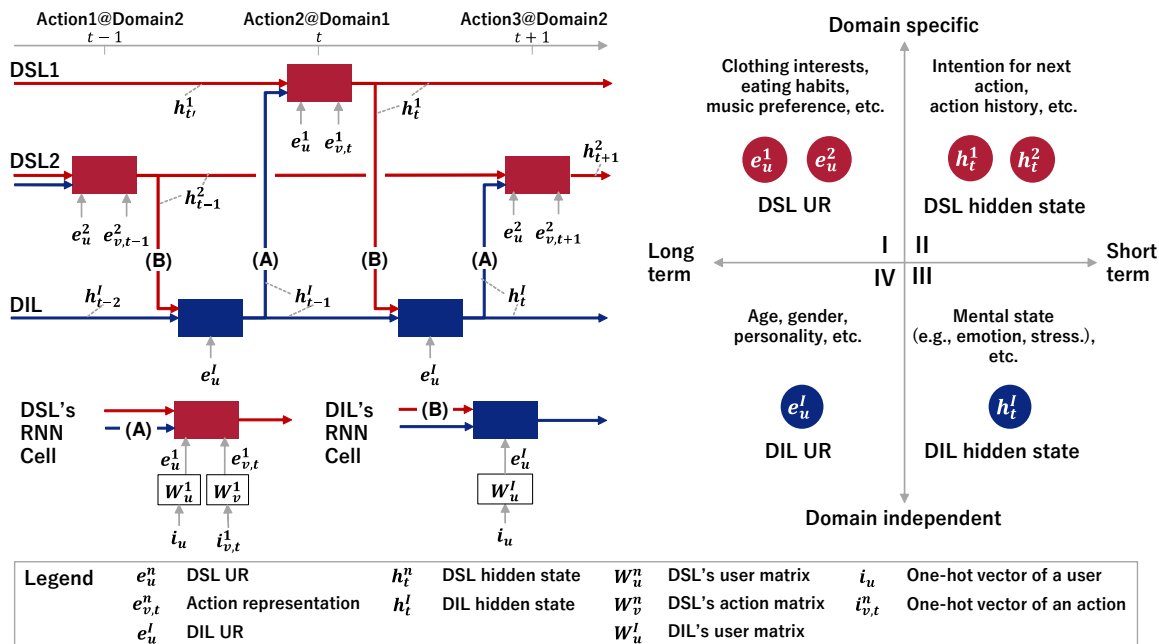


図 4.1 提案方式のマルチレイヤ RNN (左) と RNN 内の変数とユーザ属性の関係 (右)。

すなわち、DIL も DSL の隠れ状態を使用して自身の隠れ状態を更新する。リンク (B) を追加した目的は、ユーザの行動がドメイン非依存の属性、特に短期的な属性 (象限 III) に影響を与えるプロセスを RNN でモデリングすることである。象限 III の属性、すなわち精神状態 (感情やストレスなど) は、ユーザが次に何をするかに影響を与えるが (これはリンク (A) でモデリング)、同時に過去の行動によって影響を受ける。リンク (B) によってこれをモデリングすることを狙った。

リンク (B) のもう一つの狙いは、ドメイン非依存の属性、特に性格が、行動と精神状態の関係に及ぼす調整効果 (Moderating effect) を RNN でモデリングすることである。例えば、Big Five の神経症傾向が高い人間がホラー映画を見て不安や恐怖を強く感じる一方、神経症傾向が低い人間は同じホラー映画を見ても不安や恐怖を感じる可能性は低い。このように、同じ行動をとったとしても、性格が異なれば異なる精神状態が引き起こされる。このような調整効果を、DIL が自身の UR とリンク (B) で受け取る DSL の隠れ状態を用いて、隠れ状態を更新することで、モデリングできると考えた。

次節以降、隠れ状態の更新と RNN の訓練について説明する。説明に用いる記号表記については表 4.1 にまとめた。

表 4.1 記号表記の一覧

Notation	Meaning	Subscript	Superscript
N_u, N_v^n	# of users/items	u : # of users, v : # of items	n : domain n
$\mathbf{i}_u \in \mathbb{R}^{N_u}, \mathbf{i}_{v,t}^n \in \mathbb{R}^{N_v^n}$	One hot vector of ...	u : user ID, v, t : item ID of t -th action	
$d_{ue}^l, d_{ue}^n, d_{ve}^n, d_h^l, d_h^n$ ($d_{ve}^n = d_h^n$)	# of dimensions of ...	ue: UR, ve: item representation, h : hidden state	
$\mathbf{W}_u^l \in \mathbb{R}^{d_{ue}^l \times N_u}, \mathbf{W}_u^n \in \mathbb{R}^{d_{ue}^n \times N_u}, \mathbf{W}_v^n \in \mathbb{R}^{d_{ve}^n \times N_v^n}$	User/item matrix	u : user matrix, v : item matrix	l : DIL n : DSL n
$\mathbf{e}_u^l \in \mathbb{R}^{d_{ue}^l}, \mathbf{e}_u^n \in \mathbb{R}^{d_{ue}^n}, \mathbf{e}_{v,t}^n \in \mathbb{R}^{d_{ve}^n}$	UR/item representation	u : UR, v, t : item representation of t -th action	
$\mathbf{h}_t^l \in \mathbb{R}^{d_h^l}, \mathbf{h}_t^n \in \mathbb{R}^{d_h^n}$	Hidden state	t : hidden state updated by t -th action	

4.3.1 隠れ状態の更新

提案方式の入力は以下の形式をとる.

$$\text{data}_a = [x_{a,1}, x_{a,2}, \dots, x_{a,t}, \dots, x_{a,T}]. \quad (4.1)$$

ここで, $x_{a,t} = (\mathbf{i}_u, \mathbf{i}_{v,t}^n)$ はユーザ a の t 番目の行動を表す. $x_{a,t}$ が対応する DSL (以下, DSL n とする) に入力されると, DSL は最初に UR と Item representation をユーザ行列, アイテム行列から次のように取得する. $\mathbf{e}_u^n = \mathbf{W}_u^n \mathbf{i}_u$, $\mathbf{e}_{v,t}^n = \mathbf{W}_v^n \mathbf{i}_{v,t}^n$. その後, DSL は, $\mathbf{e}_u^n, \mathbf{e}_{v,t}^n$, 自身の前時刻の隠れ状態 $\mathbf{h}_{t'}^n$ (t' はドメイン n でユーザが前回行動をとった時刻), 及びリンク (A) により受け取った DIL の隠れ状態 $\mathbf{h}_{t-1}^I$ を用いて隠れ状態 $\mathbf{h}_t^n$ を更新する. DSL の隠れ状態の更新は以下の式で表される.

$$\mathbf{h}_t^n = f_{\text{RNN}}^n(\mathbf{h}_{t'}^n + \mathbf{W}_I^n \mathbf{h}_{t-1}^I, \mathbf{e}_{v,t}^n, \mathbf{e}_u^n). \quad (4.2)$$

ここで, f_{RNN}^n は隠れ状態を更新するための関数であり, Gated recurrent unit (GRU) [17] や Long short-term memory (LSTM) [56] で実装される. 式 (4.2) に示したように, $\mathbf{h}_{t-1}^I$ は $\mathbf{W}_I^n \in \mathbb{R}^{d_h^n \times d_h^I}$ によって線形変換したものを使用する. 線形変換を行う理由は二つある. 第一は, $\mathbf{h}_{t-1}^I$ の次元を $\mathbf{h}_t^n$ と整合させるためである. 第二は, $\mathbf{h}_{t-1}^I$ の影響を個々のドメイン毎に調整するためである. 既に述べたように, $\mathbf{h}_{t-1}^I$ は精神状態を反映していると仮定しているが, 精神状態が次の行動に影響を与える度合いは, ユーザが次の行動をとるドメインに応じて異なると考えられる. そのため, ドメイン毎の調整が必要となる.

隠れ状態の更新後, DSL は隠れ状態 $\mathbf{h}_t^n$ を, リンク (B) を通じて DIL に送信する. これは DSL が隠れ状態を更新する度に行われる. DIL は $\mathbf{h}_t^n$ を受け取ると, UR をユーザ行列から取り出し ($\mathbf{e}_u^I = \mathbf{W}_u^I \mathbf{i}_u$), 自身の隠れ状態 $\mathbf{h}_t^I$ を $\mathbf{h}_t^n$, $\mathbf{e}_u^I$, 及び自身の前時刻の隠れ状態 ($\mathbf{h}_{t-1}^I$) を用いて更新する. DIL における隠れ状態の更新は次の式で表される.

$$\mathbf{h}_t^I = f_{\text{RNN}}^I(\mathbf{h}_{t-1}^I + \mathbf{W}_n^I \mathbf{h}_t^n, \mathbf{e}_u^I). \quad (4.3)$$

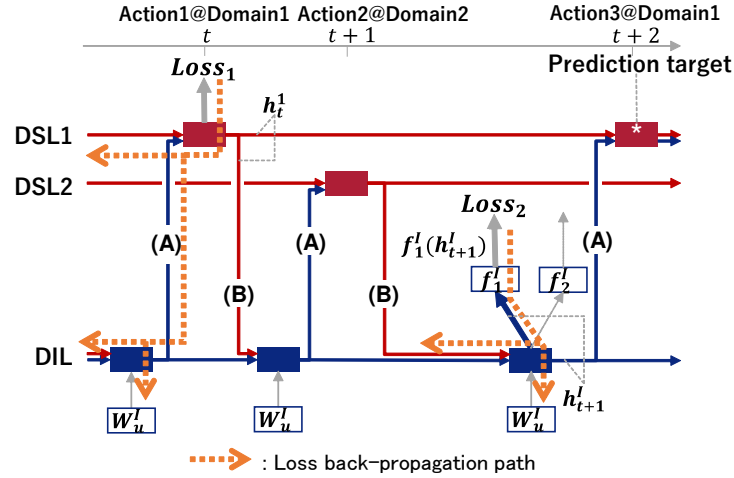


図 4.2 ロス計算の二つのオプション, $Loss_1$ と $Loss_2$.

ここで, f_{RNN}^I は隠れ状態を更新するための関数であり, GRU や LSTM で実装される. DSL と同様に, DIL も h_t^n を $W_n^I \in \mathbb{R}^{d_h \times d_h}$ によって線形変換したものを使用する. 前時刻の行動が精神状態に与える影響は, 行動がとられたドメインによって異なると考えられるため, W_n^I は DSL 毎に用意した.

4.3.2 モデルの訓練

提案方式を訓練する際は, 最初に入力データ $data_a$ にウィンドウサイズ w , スライドサイズ s のスライディングウィンドウを適用し (例: $win_1 = [x_{a,t}, x_{a,t+1}, \dots, x_{a,t+w-1}]$, $win_2 = [x_{a,t+s}, x_{a,t+s+1}, \dots, x_{a,t+s+w-1}], \dots$), 各ウィンドウを RNN に入力する. ウィンドウが入力されると, RNN は各入力データについて次のアイテムを予測する. 例えば, 入力ウィンドウが win_1 の場合, 出力は $[\hat{x}_{a,t+1}, \hat{x}_{a,t+2}, \dots, \hat{x}_{a,t+w}]$ となる. 予測したアイテムは実際のアイテムと比較され, ロスが計算される. ロスは DSL, DIL セル内のパラメタ (ゲートの重みやバイアスなど) や, ユーザ行列, アイテム行列 (W_u^n, W_u^I, W_v^n), 線形変換用の行列 (W_I^n, W_n^I) の学習に使用される.

図 4.2 に示すように, ロスの計算方法として, 二つのオプション, $Loss_1$ 及び $Loss_2$ を検討した (図中灰色矢印で表示). $Loss_1$ は, DSL の隠れ状態が対応するドメインにおけるユーザの次の行動に対する意図を反映している, とする考えに基づいており, DSL は対応ドメインにおける次のアイテムを自身の隠れ状態を使用して予測する. 例えば, 図 4.2 において, DSL1 は Action3@Domain1 におけるアイテムを Action1@Domain1 で更新した隠れ状態 h_t^1 を用いて予測している. しかし, $Loss_1$ では Action3@Domain1 より前に別のドメインでユーザが行動を起こした場合 (図の Action2@Domain2), Action3@Domain1 の予測にそうした行動を考慮できない. Action2@Domain2 もユーザの精神状態に影響

を与えるため、結果として Domain1 におけるユーザの行動に影響を与える可能性がある。また、 $Loss_1$ では、DIL のユーザ行列 ($\mathbf{W}_u^I$) に到達する前に勾配消失を起こす関数を含む二つの RNN セルを経由する必要がある。

以上の課題を踏まえ、DIL がロス計算を行う $Loss_2$ を検討した。 $Loss_2$ において、次の行動の予測は、予測対象の直前に更新された DIL の隠れ状態 $\mathbf{h}_{t+1}^I$ を用いて行う。 $\mathbf{h}_{t+1}^I$ をドメイン毎に用意した非線形関数 f_n^I (例えば活性化関数に ReLu を用いる多層パーセプトロン) に入力し、得られた出力 $f_n^I(\mathbf{h}_{t+1}^I)$ を用いて次の行動の予測を行う。 $Loss_2$ では、過去の全ての行動が考慮されることに加え、DIL のユーザ行列に到達するまでに通過するのは単一の RNN セルのみである。これにより $Loss_1$ の課題を解決できる。

ロス計算

$Loss_1$, $Loss_2$ とともに、次のアイテムの予測は以下のように行う。

$$\hat{\mathbf{y}}_t = \mathbf{W}_v^n \mathbf{h} + \mathbf{b}_v. \quad (4.4)$$

$\hat{\mathbf{y}}_t \in \mathbb{R}^{N_v^n}$ はドメイン n におけるアイテム数 (N_v^n) と同じ次元数のベクトルであり、各次元にはアイテムの予測スコアが格納されている。 $\mathbf{h}$ は $Loss_1$ と $Loss_2$ で異なる形式をとる。 $Loss_1$ では $\mathbf{h} = \mathbf{h}_v^n$ であり、 $Loss_2$ では $\mathbf{h} = f_n^I(\mathbf{h}_{t-1}^I)$ である。なお、 $\mathbf{b}_v$ はバイアスである。

DIL または DSL セルは、 $\hat{\mathbf{y}}_t$ を用いて Weighted approximate rank pairwise (WARP) ロス [151] を計算する。 WARP ロスの計算では、 $\hat{\mathbf{y}}_t$ から予測スコアが正例 (ユーザが実際に閲覧や購入したアイテム) よりも高い負例 (実際には閲覧や購入しなかったアイテム) が見つかるまでランダムにアイテムをサンプリングする。 WARP ロスの計算は次の式で行われる。

$$Loss_{WARP} = \ln\left(\frac{N_v^n - 1}{N}\right)(r^- - r^+). \quad (4.5)$$

ここで、 N は RNN セルがサンプリングしたアイテムの数である。すなわち、サンプリングしたアイテムの数が多いほどロスは小さくなり、予測が正確であることを意味する。 r^- と r^+ は、負例 (予測スコアが正例よりも高いアイテムのうち最初にサンプリングしたアイテム) と正例のスコアをそれぞれ表す。

4.4 実験

提案方式 ($Loss_1$ 及び $Loss_2$) とベースラインにより UR の学習を行い、Q1~Q3 に答えるための評価実験を行った。

4.4.1 データセット

実験では EC ポータルで収集した実際のユーザ-アイテムのインタラクションデータを格納した “Amazon Review Dataset” (Am) [100] と “RetailRocket” (RR) [116] の二つのオープンデータセットを用いた。これらのデータセットは様々な種類のアイテムをカバーしている。例えば、Am のアイテムは衣料品から食料品、電化製品からアウトドア用品まで多岐にわたる。この他にも映画の鑑賞、音楽の視聴、ニュースの閲覧などの実データを格納した様々なデータセットが公開されているが、こうしたデータセットは、日常生活の限られたドメインにおける行動のみを対象とした「ドメイン固有」のデータセットであり、これらのデータセットを用いてもドメイン固有の属性（例えば、映画や音楽の好み）しか学習できない。対照的に、Am と RR には、日常生活の様々なドメインでユーザがどのように行動するか情報が含まれている。例えば、衣料品のログは普段どのような服を着ているかを、食料品のログには食習慣を、電子機器のログには IT の使い方を、それぞれ示唆する情報を含んでいる。こうしたことから、Am や RR を用いることでドメイン特有のデータセットよりも、ドメイン非依存のユーザ属性を学習できる可能性が高いと考えた。

実験で用いたデータセットの概要を表 4.2 に示す。Am には、アイテムのレビュー行動のシーケンスが含まれる。各ログは、ユーザ ID、商品名、カテゴリ、ジャンル、レビューテキストで構成される。元のデータセットに含まれる 17 の商品ジャンルのうち、ログの数に基づいて表に示す 6 つのジャンルを選択し、各ジャンルをドメインとみなした。そして、“C”、“E”、“H”、“T” の 4 つのドメインそれぞれについて、14 以上のログを持つユーザを抽出した（“G” と “S” はユーザ数が極端に減るので除外した）。各ドメインには数百から数千のアイテムカテゴリがあり、それらは複数のカテゴリラベルを連結したもので表現される（例えば、“男性”+“靴”+“運動靴”、“女性”+“服”+“ドレス”）。アイテムカテゴリに ID を割り当て（カテゴリラベルには割り当てない）、データのスパース性を抑制するために、アイテム ID ではなく、ユーザのレビュー行動を表現するためにこれらのカテゴリ ID を使用した。

RR には、アイテムの閲覧・購入行動のシーケンスが含まれる。各ログは、ユーザ ID、アイテム ID、カテゴリとジャンル、行動種別（閲覧/購入）で構成される。アイテムのジャンルは 258 あり、それぞれが数十のカテゴリを持っている。各ジャンルは少数のログしか持たないため、ジャンルを 4 つのグループに統合し、それらをドメインと見なした（すなわち、G1~G4）。元のデータセットでは、ジャンルは乱数（例えば、213, 169）で表現されているため、ジャンル間の関係に基づいてマージできず、ドメイン間でログ数がバランスするようにマージすることしかできなかった。したがって、RR では、無関係な

表 4.2 データセットの概要

Amazon (Am)			
# of users (N_u)	Domains	# of categories (N_v^n)	# of logs
691	Clothing, shoes, and jewelry (C)	4,625	26,978
	Electronics (E)	719	47,990
	Home and kitchen(H)	1,195	50,686
	Tools and home improvement (T)	736	21,133
	Grocery and gourmet food (G)	587	13,566
	Sports and outdoors (S)	1,038	15,581
RetailRocket (RR)			
# of users (N_u)	Domains	# of categories (N_v^n)	# of logs*
632	Group1 (G1)	238	43,515
	Group2 (G2)	248	30,685
	Group3 (G3)	222	34,076
	Group4 (G4)	200	36,752

* # of browse logs

ジャンルのアイテムが同じドメインにマージされた可能性がある。しかし、このような場合において、提案方式がどのように機能するかを検証するために、RR を使用した。2 つのドメインの 6 つの組み合わせ (C_2^4) のそれぞれにおいて、9 つ以上の閲覧ログを持つユーザを選択した。また、Am と同様に、ユーザの行動をカテゴリ ID で表現した。

4.4.2 ベースライン

ベースラインとして、UR の学習を多様なドメインで行うことが可能な方式を提案方式と比較するため、次の二つの条件を満たす方式を選択した: (1) ID で表現されたユーザ行動から UR を学習できる, (2) 特定のドメインでのみ実行可能な特殊なタスクを行うことなく UR を学習できる。Universal UR learning の方式は、ユーザの行動をテキストで表現する必要があったり（ユーザが購入したりレビューした商品名の情報が必要 [49, 130, 138], 商品レビューのテキストが必要 [112]), 店舗や値段の好みの予測タスクなど特殊なタスクが必要であったりするため [101], (1), (2) の条件に合致しない。このため、ベースラインからは除外した。

Sequential な方式からは、提案方式の多層構造の有効性を確認するため、SURNN[34] を選択した。Non-sequential な方式からは、シンプルであることと、多数の研究で比較対象に用いられていることから MF[70] と FM[115] の二つの方式を選択した。また、性能面で優れていることから xDeepFM[75] も比較対象として選択した。

4.4.3 UR の学習

まず、UR の学習に用いるドメインの組合せを複数用意した。ベースラインでは、ドメイン非依存のユーザ属性とドメイン固有の属性が同じ UR の中に混在しているのに対し、提案方式で学習する UR (DIL UR) には、組合せに含まれる全てのドメインにおけるユーザ行動に影響を与えるドメイン非依存の属性が明示的に反映されると期待した。Am では、 C_2^6 , C_3^6 , C_4^6 , C_5^6 の組合せ、すなわち、2 つ、3 つ、4 つ、5 つのドメインの組合せを作成した (例えば、2 つのドメインでは 'CE', 3 つのドメインでは 'CEH', ..., 合計で 55 の組合せを作成した^{*2})。RR については、 C_2^4 個の組合せを作成した (例えば 'G1G2', 合計 6 個の組合せ)。ドメインの組合せを作成した後、ドメインの組合せすべてについて、一つ一つの組合せを用いて UR の学習を行った。このように様々なドメインの組合せについて UR の学習を行ったのは、1) UR の性能が学習するドメインにどのように依存するか、2) ドメインの数を増やすと性能がどのように変化するか、を調べるのが目的である。各ドメインの組合せについて、80% のログを訓練に、20% のログを検証に使用した。アイテムのジャンルやカテゴリは ID で表現し、テキスト情報は使用しなかった。また、RR では購入ログは用いず閲覧ログのみを使用して UR の学習を行った。

提案方式と SURNN (Sequential な方式) では、LSTM を用いて RNN セルを実装し、ユーザがレビューする (Am) / 閲覧する (RR) アイテムのカテゴリ ID を予測してモデルを学習した。提案方式では、その後の評価に DIL UR を用いた。一方、MF, FM, xDeepFM (Non-sequential な方式) については、ユーザ ID とアイテムカテゴリ ID のペア毎に予測を行った。すなわち Am ではユーザがそのカテゴリの商品をレビューしているか否かを、RR ではユーザがそのカテゴリの商品を閲覧した回数を、それぞれ予測した。

提案方式とベースラインの実装詳細

以下、実験における提案方式とベースラインの実装の詳細について説明する。実験において UR の次元数 (d_{ue}^*) は 16 とした。

■提案方式と SURNN (Sequential な方式) 提案方式と SURNN については、 $(i_u, i_{v,t})$ の形式のデータを入力した。 $i_{v,t}$ は Am においてはユーザがレビューした、RR においてはユーザが閲覧したアイテムのカテゴリ ID を表す 1-hot ベクトルである。提案方式、SURNN ともユーザがレビュー/閲覧するアイテムのカテゴリ ID を予測した。

入力データに適用するスライディングウィンドウはウィンドウサイズを 15、スライド

^{*2} ログの数が少ないため、'GS' の組合せは UR 学習に用いなかった。したがって二つのドメインの組合せについては $C_2^6 - 1 = 14$ 個作成した。

幅を 5 とした。ロス関数は、学習率を 0.005 に設定し、Adam optimizer により最適化した。また、バッチサイズは 96 とした。Validation データにおいてロスが収束したタイミングでモデルの訓練を停止した。Item representation, DSL, DIL, SURNN の隠れ状態の次元数は UR の次元数と同一とした ($d_{\text{ve}}^*, d_h^* = d_{\text{ue}}^*$)。提案方式 (Loss_2) の f_n^I は、ReLU を活性化関数とし、第一、第二のレイヤがそれぞれ $2 \times d_h^I$ 個、 $2 \times d_{\text{ve}}^n$ 個のパーセプトロンで構成される、2 層パーセプトロンにより実装した。

DSL (f_{RNN}^n), DIL (f_{RNN}^I), SURNN の RNN セルは LSTM により実装した。すなわち、隠れ状態 $\mathbf{h}_t$ は以下の式で表される Input, Forget, Output ゲート ($\mathbf{I}_t, \mathbf{F}_t, \mathbf{O}_t \in \mathbb{R}^{d_h}$) 及び Candidate memory, Present memory ($\tilde{\mathbf{C}}_t, \mathbf{C}_t \in \mathbb{R}^{d_h}$) を用いて更新した。

$$\mathbf{I}_t = \sigma(\mathbf{W}_{ei}\mathbf{e} + \mathbf{W}_{hi}\mathbf{h} + \mathbf{b}_i), \quad (4.6)$$

$$\mathbf{F}_t = \sigma(\mathbf{W}_{ef}\mathbf{e} + \mathbf{W}_{hf}\mathbf{h} + \mathbf{b}_f), \quad (4.7)$$

$$\mathbf{O}_t = \sigma(\mathbf{W}_{eo}\mathbf{e} + \mathbf{W}_{ho}\mathbf{h} + \mathbf{b}_o), \quad (4.8)$$

$$\tilde{\mathbf{C}}_t = \tanh(\mathbf{W}_{ec}\mathbf{e} + \mathbf{W}_{hc}\mathbf{h} + \mathbf{b}_c), \quad (4.9)$$

$$\mathbf{C}_t = \mathbf{F}_t \odot \mathbf{C}_{t-1} + \mathbf{I}_t \odot \tilde{\mathbf{C}}_t, \quad (4.10)$$

$$\mathbf{h}_t = \mathbf{O}_t \odot \tanh(\mathbf{C}_t). \quad (4.11)$$

ここで、 σ と $\odot$ はシグモイド関数とエレメント毎の積をそれぞれ表す。 $\mathbf{e} \in \mathbb{R}^{d_e}$ と $\mathbf{h} \in \mathbb{R}^{d_h}$ は RNN セルへの入力であり、 $\mathbf{W}_{e*} \in \mathbb{R}^{d_h \times d_e}$, $\mathbf{W}_{h*} \in \mathbb{R}^{d_h \times d_h}$, $\mathbf{b}_* \in \mathbb{R}^{d_h}$ は学習可能なパラメタ (learnable parameter) である。上述の式は DSL, DIL, SURNN で共通だが、 $\mathbf{e}$ 及び $\mathbf{h}$ は Table 4.3 に示すとおり異なる形式をとる。

表 4.3 DSL, DIL, SURNN の RNN セルへの入力 (括弧内に次元数を示す)

	$\mathbf{e} (d_e)$	$\mathbf{h} (d_h)$
DSL	$\mathbf{e}_u^n \oplus \mathbf{e}_{v,t}^n (d_{\text{ue}}^n + d_{\text{ve}}^n)$	$\mathbf{h}_t^n + \mathbf{W}_I^n \mathbf{h}_{t-1}^I (d_h^n)$
DIL	$\mathbf{e}_u^I (d_{\text{ue}}^I)$	$\mathbf{h}_{t-1}^I + \mathbf{W}_n^I \mathbf{h}_t^n (d_h^I)$
SURNN	$\mathbf{e}_u^S \oplus \mathbf{e}_{v,t}^S (d_{\text{ue}}^S + d_{\text{ve}}^S)$	$\mathbf{h}_{t-1}^S (d_h^S)$

SURNN について、 $\mathbf{e}_u^S \in \mathbb{R}^{d_{\text{ue}}^S}$ と $\mathbf{e}_{v,t}^S \in \mathbb{R}^{d_{\text{ve}}^S}$ は、それぞれ SURNN の UR と Item representation であり、ユーザ行列 $\mathbf{W}_u^S \in \mathbb{R}^{d_{\text{ue}}^S \times N_u}$ と アイテム行列 $\mathbf{W}_v^S \in \mathbb{R}^{d_{\text{ve}}^S \times N_v}$ (N_v は全ドメインにおけるアイテムカテゴリ数の合計を表す) から $\mathbf{e}_u^S = \mathbf{W}_u^S \mathbf{i}_u$, $\mathbf{e}_{v,t}^S = \mathbf{W}_v^S \mathbf{i}_{v,t}$ のように取り出される。以下、SURNN のモデル訓練について説明する。

提案方式と同様に、SURNN は以下のように予測を行った。

$$\hat{\mathbf{y}}_t = \mathbf{W}_v^{S^\top} \mathbf{h}_{t-1}^S + \mathbf{b}_v. \quad (4.12)$$

ここで、 $\hat{\mathbf{y}}_t \in \mathbb{R}^{N_v}$ は対応する次元にアイテムカテゴリ毎の予測スコアを含んでいる。ロス提案方式と同様、WARP ロスによって計算した。訓練後、 $\mathbf{W}_u^S$ を取り出し、UR として評価に用いた。

■MF, FM, xDeepFM (Non-sequential な方式) MF, FM, xDeepFM については、ユーザ ID と アイテムカテゴリの各ペアについて予測を行った。以下、 $\hat{\mathbf{y}}_{a,b}$ はユーザ a と アイテムカテゴリ b の予測結果を表す。Am では、モデルにはユーザがアイテムカテゴリのレビューを行ったかどうかを予測させた（すなわち、 $\hat{\mathbf{y}}_{a,b} \in [0...1]$ ）。一方で、RR では、モデルにユーザがアイテムカテゴリの閲覧を何回行ったかを予測させた（すなわち、 $\hat{\mathbf{y}}_{a,b} \in \mathbb{R}$ ）。

全ての方式について、Am では Binary cross entropy を、RR では RMSE をロス関数として用いた。

- MF

Am ではロジスティック MF[62] を用いて以下のように予測を行った。

$$\hat{y}_{a,b} = \sigma(\mathbf{e}_u^\top \mathbf{e}_v + b_u + b_v). \quad (4.13)$$

$\mathbf{e}_u \in \mathbb{R}^{d_{ue}}$, $\mathbf{e}_v \in \mathbb{R}^{d_{ve}}$, b_u , b_v はそれぞれ UR, Item representation, ユーザ及びアイテムバイアスを表す。なお、MF では $d_{ue} = d_{ve}$ である。

RR では通常の MF[70] を用いて以下のように予測を行った。

$$\hat{\mathbf{Y}} = \mathbf{W}_u^\top \mathbf{W}_v. \quad (4.14)$$

$\hat{\mathbf{Y}} \in \mathbb{R}^{N_u \times N_v}$ の a 番目の行、 b 番目の列には $\hat{y}_{a,b}$ が格納される。 $\mathbf{W}_u \in \mathbb{R}^{d_{ue} \times N_u}$ 及び $\mathbf{W}_v \in \mathbb{R}^{d_{ve} \times N_v}$ は各列がそれぞれ $\mathbf{e}_u$ 及び $\mathbf{e}_v$ に対応する行列である。

- FM

ユーザ ID, アイテムカテゴリ, ドメインを表す 3 つの 1-hot ベクトルを連結した $\mathbf{x} \in \mathbb{R}^d$ を入力し ($d = N_u + N_v + N_g$, N_g はドメイン数を表す), 以下のように予測をおこなった。

$$\text{Am, } \hat{y}_{a,b} = \sigma(f_{\text{FM}}(\mathbf{x})) \quad (4.15)$$

$$\text{RR, } \hat{y}_{a,b} = f_{\text{FM}}(\mathbf{x}), \quad (4.16)$$

$$f_{\text{FM}}(\mathbf{x}) = w_0 + \sum_{i=1}^d w_1 x_i + \sum_{i=1}^d \sum_{j=i+1}^d \mathbf{g}_i^\top \mathbf{g}_j x_i x_j. \quad (4.17)$$

x_i は $\mathbf{x}$ の i 番目の要素, w_0 , w_1 及び $\mathbf{g}_i \in \mathbb{R}^{d_{ue}}$ は学習可能なパラメタである。UR の評価にはユーザ ID の 1-hot ベクトルに対応する $\mathbf{g}_i$ を UR として利用した。

- xDeepFM

FM と同様, $\mathbf{x} \in \mathbb{R}^d$ を入力し, 以下のように予測を行った.

$$\text{Am}, \hat{y}_{a,b} = \sigma(f_{\text{xDFM}}(\mathbf{x})) \quad (4.18)$$

$$\text{RR}, \hat{y}_{a,b} = f_{\text{xDFM}}(\mathbf{x}), \quad (4.19)$$

$$\begin{aligned} f_{\text{xDFM}}(\mathbf{x}) = & \mathbf{W}_{\text{linear}}\mathbf{x} + \mathbf{W}_{\text{DNN}}f_{\text{DNN}}(\mathbf{e}_u, \mathbf{e}_v, \mathbf{e}_g) \\ & + \mathbf{W}_{\text{CIN}}f_{\text{CIN}}(\mathbf{e}_u, \mathbf{e}_v, \mathbf{e}_g) + b. \end{aligned} \quad (4.20)$$

$\mathbf{e}_u, \mathbf{e}_v, \mathbf{e}_g \in \mathbb{R}^{d_{\text{ge}}}$ は, それぞれ UR, Item representation, Domain representation を表す ($d_{\text{ve}}, d_{\text{ge}} = d_{\text{ue}}$). これらは xDeepFM の Embedding layer において 1-hot ベクトルから抽出される. 評価では $\mathbf{e}_u$ を UR として用いた. $\mathbf{W}_*$ と b は学習可能なパラメタであり f_{DNN} と f_{CIN} は, Deep neural network (DNN) と Compressed interaction network (CIN) の計算処理を表す. f_{DNN} と f_{CIN} の詳細については [75] を参照した. バッチサイズ及び DNN の隠れユニット, CIN のレイヤサイズは 128, (256, 128, 64), 及び (128, 128) とした.

4.4.4 UR の評価

同一ユーザの UR の類似性 (Q1)

ドメインの組合せのペアを作成し, 各ペアについて, 同一のユーザの UR の類似性を評価した. 例えば, ペアが ['CE', 'HT'] の場合, UR_{CE}^i ('CE' で学習したユーザ i の UR) と UR_{HT}^i ('HT' で学習したユーザ i の UR) を比較した. なお, 異なるドメインの組合せで学習した UR はベクトル空間自体が異なるため (例えば, UR_{CE}^i と UR_{HT}^i の第 1 次元は異なる意味合いを持つ), 直接は比較せず, 一方の UR を他方のベクトル空間に写像した上で比較を行った. すなわち, $\mathbf{W}_{\text{CE} \rightarrow \text{HT}}^{\text{prj}} UR_{\text{CE}}^i$ と UR_{HT}^i 及び $\mathbf{W}_{\text{HT} \rightarrow \text{CE}}^{\text{prj}} UR_{\text{HT}}^i$ と UR_{CE}^i を比較した. ここで, $\mathbf{W}_{C_S \rightarrow C_T}^{\text{prj}} \in \mathbb{R}^{d_{\text{ue}} \times d_{\text{ue}}}$ は一方の空間 (C_S) から他方の空間 (C_T) への写像行列を表す.

なお, 評価からは例えば ['CE', 'CH'] のように, 共通のドメインを持つペアは除外した. 共通のドメイン固有のユーザ属性がドメイン非依存性に関係なく類似性を高めてしまうと考えたためである. このようなペアを除外した後, 例えば ['CE', 'HTS'] や ['CE', 'HTGS'] などドメイン数が異なるペア同士も含め, 全てのペアについて評価を行った.

最初に, ユーザをランダムに 5 グループに分割し, うち 4 グループを $\mathbf{W}_{C_S \rightarrow C_T}^{\text{prj}}$ の訓練に, 残りのグループを評価に利用した. 評価するグループを変更してこれを 5 回繰り返し, 全てのユーザについて UR の類似性を評価した (すなわち, 5-fold cross validation を実施した).

類似性の指標としては, ユークリッド距離の Mean reciprocal rank (MRR, MRR_{EUC})

を用いた.

$$MRR_{\text{EUC}}(C_S, C_T) = \frac{1}{N} \sum_{i=1}^N \frac{1}{\text{rank}_i}. \quad (4.21)$$

$MRR_{\text{EUC}}(C_S, C_T)$ は UR_{C_S} を UR_{C_T} へ写像した際の MRR_{EUC} を表す. $\mathbf{W}_{C_S \rightarrow C_T}^{\text{prj}} UR_{C_S}^i$ と C_T で学習した全ての UR の距離を調べ, 距離により昇順ソートを行った. rank_i は $\mathbf{W}_{C_S \rightarrow C_T}^{\text{prj}} UR_{C_S}^i$ と $UR_{C_T}^i$ の間の距離 (すなわちユーザ i の UR 同士の距離) の評価グループにおける順位を表す. N は, 評価グループ内のユーザ数を表す. 同一ユーザの UR が近いほど, rank_i は小さい値となるため, MRR_{EUC} の値は大きくなる. 異なるペアで結果を比較するため (例えば ['CE', 'HT'] と ['CEG', 'HTS']) で結果を比較するため), ユークリッド距離に代えて本指標を用いた.

行動予測 (Q2)

ユーザの行動を予測するためには, 予測対象のドメインと同じドメインのログからユーザの特徴量を抽出するのが一般的である (例えば, ユーザの音楽視聴履歴から抽出した音楽の好みを用いて, ユーザが聴くであろう曲を予測する). このようなシナリオでは, 異なるドメインで学習した UR がドメイン非依存の属性を反映している場合, ターゲットドメインから抽出されたユーザ特徴 (このようなユーザ特徴をドメイン固有特徴, DF と呼ぶ) に加えて使用することで, 予測精度が向上すると考えられる. そこで, DF に加え, 異なるドメインで学習した UR を用いて, 対象ドメインにおけるユーザの行動を予測し, その予測精度を評価した. 具体的には, Am では特定のカテゴリラベルのアイテムをレビューしたかどうか, RR では特定のジャンルのアイテムを購入したかどうかを予測した.

なお, 本評価では UR のみを利用した予測実験は行わなかった. そうした予測では UR のドメイン非依存性を評価できないためである. 例えば, “Video games(V)” と “Software(W)” において UR を学習し, “Electronics(E)” の行動予測を行った場合, 仮に UR_{VW} がドメイン特有の属性のみを反映していたとしても, 3つのドメインの類似性から, ‘E’ において高い予測力を持つ可能性がある. UR_{VW} はユーザが日常 IT にどのように触れているかを反映しているはずであり, それが ‘E’ における行動予測に寄与する可能性があるためである. UR_{VW} に加え ‘E’ で学習した DF を利用することで, UR_{VW} に含まれるドメイン特有の属性 (日常 IT にどう触れているか) は3つのドメインの類似性により冗長なものとなる. これにより, ドメイン非依存な属性が反映していない UR を利用した予測が高精度となることが防がれる.

最初に, SURNN を用いて, 予測対象のドメインにおいて UR を学習し, これを DF として用いた (DF の次元数は 16 とした). そして予測対象ドメインにおいて正例数の基

準^{*3}を満たすカテゴリラベル (A_m) / 商品ジャンル (RR) について、ロジスティック回帰 (σ) による予測モデルを構築した。例えば、予測対象のドメインが A_m の ‘E’ の場合、 $UR_{C \setminus E}$ を用いた。ここで $C \setminus E$ は ‘E’ を含まないドメインの組合せ (‘CH’, ‘CHT’ など) を表す。以下のように予測を行った。

$$\hat{y}_{label=1} = \sigma(b_0 + \mathbf{b}_1 DF_E + \mathbf{b}_2 UR_{C \setminus E}). \quad (4.22)$$

$\hat{y}_{label=1}$ はユーザがカテゴリラベル 1 のアイテムをレビューする確率を表す。 b_0 は切片、 $\mathbf{b}_1$ と $\mathbf{b}_2$ は DF, UR の偏回帰係数ベクトルを表す。 $C \setminus E$ に含まれる全てのドメインの組合せについて評価を行った。ロジスティック回帰モデルは 5-fold cross validation により訓練と評価を行い、予測精度の指標としては ROC-AUC を用いた。また、比較のため、DF のみを用いたモデル $\hat{y}_{label=1} = \sigma(b_0 + \mathbf{b}_1 DF_E)$ の評価も行った。

テキスト特徴量の推定 (Q3)

最後に、UR と性格との関連性を評価した。まず、アイテムのレビューテキストから、従来研究により Big Five と相関があると報告されているテキスト特徴量を抽出した。そして、UR を用いてテキスト特徴量を予測した際の精度を評価した。本評価は、アイテムのレビューテキストが利用可能な A_m でのみ行った。

これまでに多くの研究者が、性格が文章の書き方 (例：エッセイやツイートにおける言葉の使い方) に影響を与えることを報告している。そのような研究の中で、テキスト特徴量と Big Five スコアとの相関を報告した文献を参照し、Big Five と有意な相関を持つテキスト特徴量を全て抽出した (合計 40 の特徴量)。特徴量の抽出には、[45, 137, 55] で使用されたテキスト特徴量抽出ツール LIWC[109] を使用した。テキスト特徴量には、テキストに関する一般的な情報 (単語数など) のほか、心理的構成要素に関するカテゴリ (否定的／肯定的感情など)、個人的関心カテゴリ (仕事、家庭など)、単語クラスカテゴリ (冠詞、助動詞など) など、テキスト内で使用されている単語カテゴリの頻度が含まれる。こうしたテキスト特徴量についてユーザ毎に平均を取ってベクトル化し ($\mathbf{P} \in \mathbb{R}^{40}$)、主成分分析により次元圧縮を行った。具体的には、 $\sim k$ 番目の主成分で構成されるベクトル ($\mathbf{P}' \in \mathbb{R}^k$) を生成した。 k については $\arg \min_k \sum_{i=1}^k loading_i \geq Th$ ($loading_i$ は i 番目の主成分の因子負荷量を表す) とした ($Th = 0.9$ と 0.8 について $k = 21$ と 15 の値を得た)。そして、線形回帰により、UR を用いて $\mathbf{P}'$ を予測した。 $\hat{\mathbf{P}}' = \mathbf{W}^P \mathbf{UR}$ ($\mathbf{W}^P \in \mathbb{R}^{k \times d_{ue}}$)。予測評価は 5-fold cross-validation により行い、評価指標としては $\hat{\mathbf{P}}'$ と $\mathbf{P}'$ のユークリッド距離を用いた。

^{*3} $0.05N_u \leq n \leq 0.5N_u$; n は正例の数 (商品レビュー/購入したユーザの数) を表す。

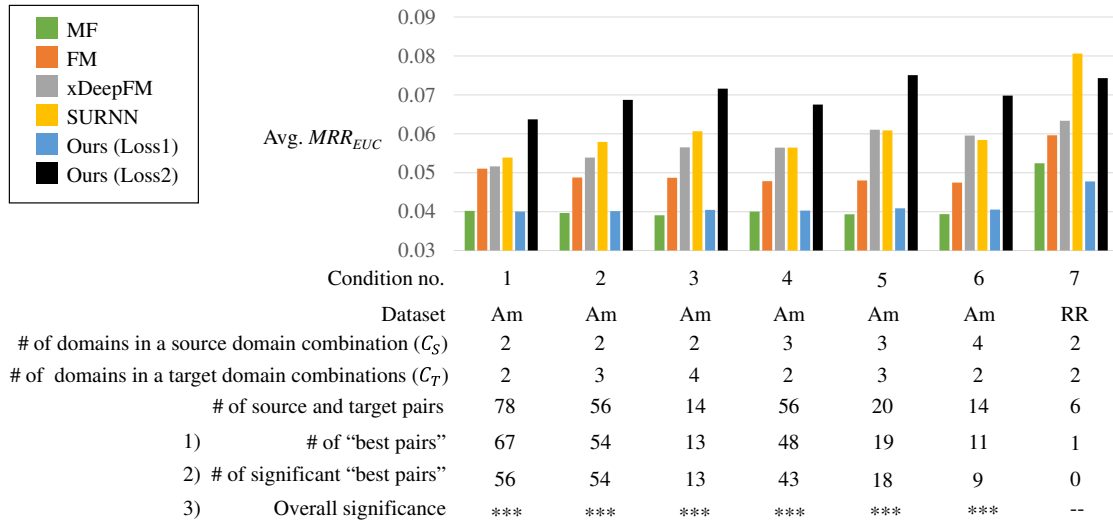


図 4.3 Q1 - 同一ユーザの UR の類似性. 棒グラフの各グループは同一条件における MRR_{EUC} の平均を示す（例えば、条件 1 は 78 個の MRR_{EUC} の数値の平均を示す）。グラフ下の表において、1) は提案方式 ($Loss_2$) の MRR_{EUC} が最良の結果（同一ユーザの類似性が最も高かった）となったペアの数を示す。2) はそうしたペアのうち、提案方式 ($Loss_2$) の MRR_{EUC} が他の全方式よりも有意に高かった ($p < .05$) ペアの数を示す。3) は全ペアの MRR_{EUC} の平均について、提案方式 ($Loss_2$) と他の方式の差分の有意性を示す。

4.5 結果

4.5.1 同一ユーザの UR の類似性 (Q1)

図 4.3 に結果を示す。Am では、提案方式で学習した UR ($Loss_2$) が全ての条件で MRR_{EUC} の平均が最も高い結果（すなわち最も類似する結果）となった。図中の 3) に示すように、ベースラインとの差は Am の全ての条件で統計的に有意であった。これらの結果から、提案方式で学習した同一ユーザの UR が最も類似していることが確認され、提案方式の UR が評価対象の中で最も高いドメイン非依存性を持つことが示された。

結果から、提案方式がほとんどのドメインの組合せにおいて、最もドメイン非依存性の高い UR を学習したことが分かる。図の 1) に示すように、全 238 のペアのうち、212 ペアにおいて提案方式 ($Loss_2$) の UR が最も良い性能を示し、2) に示すように、そのうち 193 ペアにおいてベースラインとの差が統計的に有意であった。 $Loss_1$ と $Loss_2$ を比較すると、前者の課題を解決するために設計した後者が想定通り前者を上回った。

Am におけるもう一つの注目すべき結果は、 MRR_{EUC} の平均と、ソースとターゲットのドメインの組み合わせ (C_S と C_T) におけるドメイン数との関係である。グラフの条

件 1~3 において (C_S で 2 つのドメインを使用) では提案方式 ($Loss_2$) の MRR_{EUC} が C_T の増加に伴い向上している. 同様に, 条件 1, 4, 6 (C_T で 2 ドメイン使用) の結果について C_S のドメイン数の増加に伴い向上している. これらの結果は, 学習に使用するドメインを増やすほど, 提案方式で学習できるユーザ属性のドメイン非依存性が向上することを示している.

Am の結果とは対照的に, 提案方式は RR において SURNN を上回らなかった. これは, 複数のアイテムジャンルを, それらの関係性を考慮することなく 1 つのグループに統合し, 同じ DSL にまとめたことが原因と考えられる. このような場合, DSL はドメイン固有のユーザ属性を学習できないため, 提案方式ではドメイン固有のユーザ属性とドメイン非依存のユーザ属性を区別できない.

4.5.2 行動予測 (Q2)

図 4.4 に結果を示す. Am において, 提案方式 ($Loss_2$) の ROC-AUC は, ドメイン数の増加に伴い向上している. グラフに示すとおり, 2 つ以上のドメインを用いて UR を学習した場合, 提案方式 ($Loss_2$) は, UR を用いないドメイン特徴量 (DF) のみによる予測 (グラフ中破線で示す) を含むすべての方式の中で, 最も高い平均 ROC-AUC を示した. 図中の 1) に示すように, 条件 6 の 2 つのドメインの組合せを除く, ドメイン数が 2 つ以上の全てのドメインの組合せにおいて, 提案方式 ($Loss_2$) が最良の結果となった. さらに, 図中の 2) に示すように, ドメイン数が 3 以上の場合, 条件 7 の 1 つのドメインの組合せを除くすべてのドメインの組合せにおいて, 提案方式の結果はベースラインの方式に比べ有意に高かった. これらの結果は Q1 の結果と一致する. すなわち, 提案方式で学習した UR はベースラインで学習した UR と比べてドメイン非依存のユーザ属性をより強く反映している.

しかし, 提案方式 ($Loss_2$) の結果はドメイン数が増えるのに伴い改善するものの, 改善の度合いは減少することに注意する必要がある. これについては次章で考察する.

ベースラインの結果を見ると, その ROC-AUC は DF による結果よりも有意に低いか, DF による予測とほぼ同じレベルに留まる (条件 3, 4, 12 の SURNN の UR を除く). 4.2 節で述べた通り, ベースラインの UR に含まれるドメイン固有のユーザ属性が, 他のドメインにおける行動についてほとんど予測力を持たず, 逆にノイズとなって予測精度を劣化させていることを示していると考えられる.

RR については, Q1 における結果と同様, 提案方式は $Loss_1$, $Loss_2$ ともにベースラインを上回る結果を示さなかった.

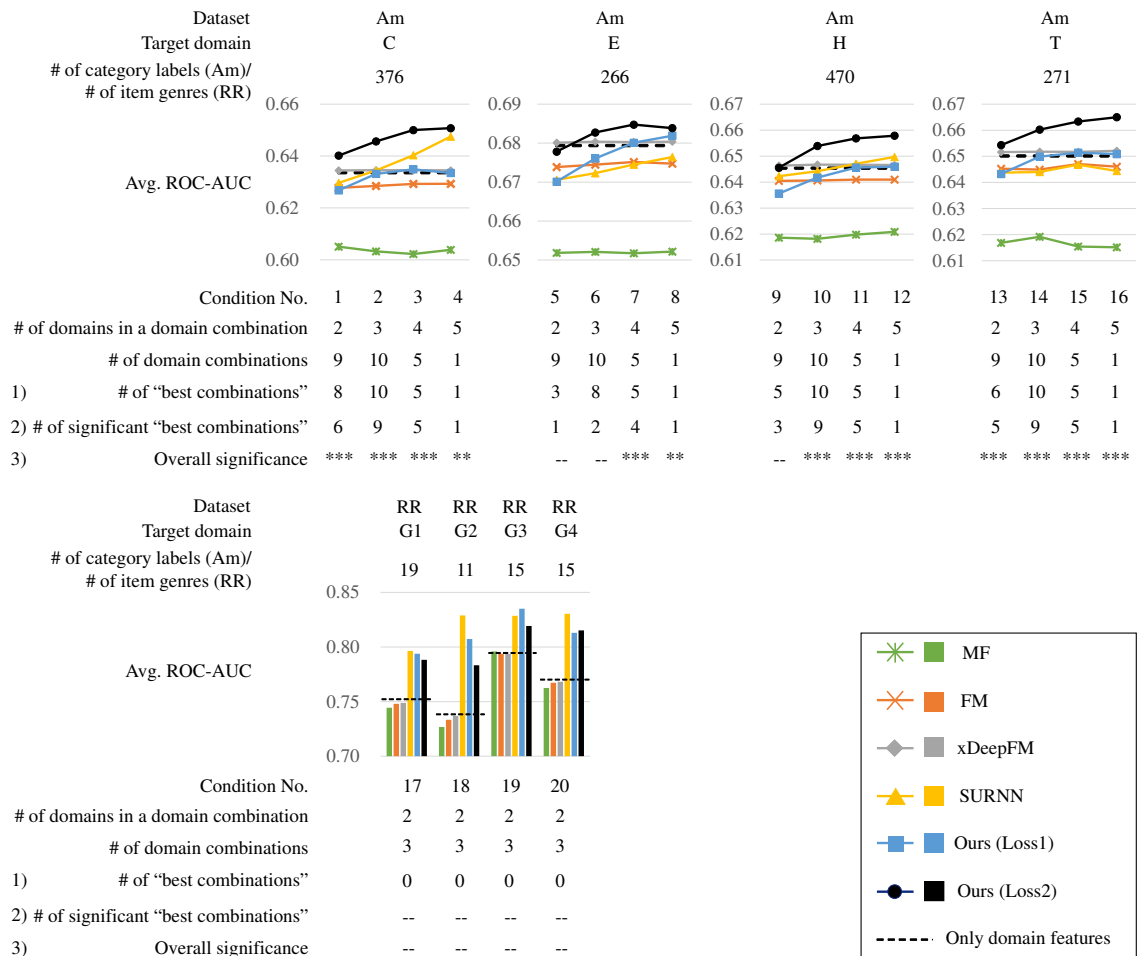


図 4.4 Q2 - 行動予測結果. グラフ中の各点/棒は、「カテゴリラベルの数 (Am) / 商品ジャンル数 (RR) × ドメインの組合せ数」個の結果の平均を示す (例えば, 条件 1) における各点は $376 \times 9 = 3,384$ の ROC-AUC の平均である). 破線は DF のみを用いた場合の予測精度を示す. グラフ下の表において, 1) は提案方式 ($Loss_2$) で学習した UR が最良の ROC-AUC となったドメインの組合せ数を示す. 2) は提案方式 ($Loss_2$) とベースラインの結果をドメインの組合せ毎に比較し, その差分が統計的に有意 ($p < .05$) となったドメインの組合せ数を示す. 3) は, 全てのドメインの組合せの ROC-AUC の平均を, 提案方式 ($Loss_2$) とベースラインで比較し, その差分の統計的有意性を示す (** $p < .05$, *** $p < .01$).

4.5.3 テキスト特徴量の推定 (Q3)

図 4.5 に結果を示す. $k = 21$ と 15 の両方において, ドメイン数が 5 未満の場合, 提案方式 ($Loss_2$) は図の 3) に示すように, 全てのベースラインよりも有意に短い平均距離となった. また, ドメイン数が 5 未満の 49 のドメインの組合せのうち, 1) に示すように, $k = 21$ の場合 43, 15 の場合 39 の組合せにおいて, 提案方式 ($Loss_2$) が最も良好な推定

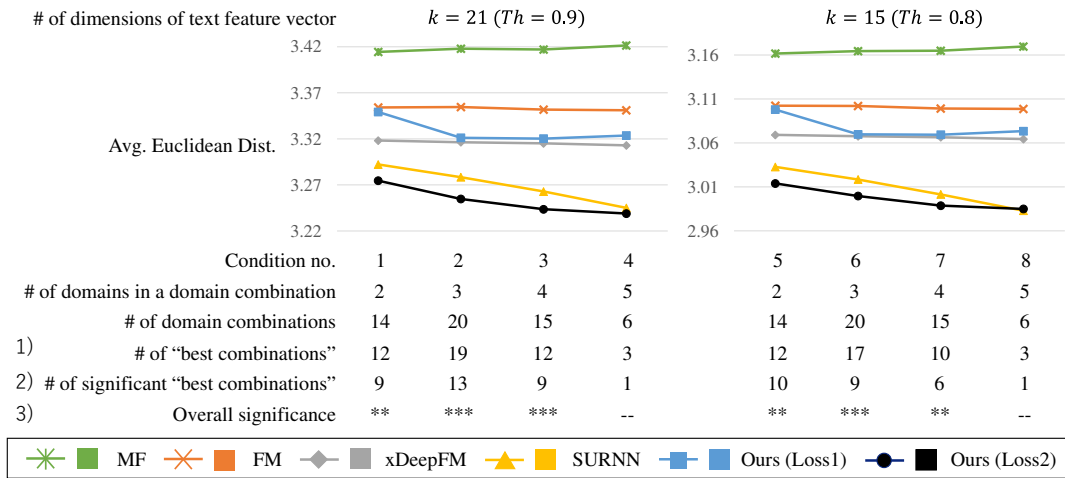


図 4.5 Q3 - テキスト特徴量推定結果. グラフ中の各点は、テキスト特徴量ベクトル ($\mathbf{P}'$) と推定結果のベクトル ($\mathbf{W}^P \mathbf{UR}$) のユークリッド距離の平均を示す. 例えば、条件 1 の各点は、 14×691 (ドメインの組合せ数が 14, ユーザ数が 691) のユークリッド距離の平均である. グラフ下の表において、1) は提案方式 ($Loss_2$) の UR が最短の平均距離となったドメインの組合せの数を示す. 2) は 1) のドメインの組合せにおいて、提案方式 ($Loss_2$) とベースラインの差分が統計的に有意 ($p < .05$) であったドメインの組合せ数を示す. 3) は各条件における全ユークリッド距離の平均について、提案方式 ($Loss_2$) とベースラインの差分の優位性を示す (** $p < .05$, *** $p < .01$).

精度を示した. これらの結果から提案方式 ($Loss_2$) で学習した UR がベースラインに比べて Big Five との関連性が高いことを示唆する.

一方で、ドメイン数の増加に伴い推定精度は向上するものの、Q2 の結果で見られたように、向上の度合いは減少する. また、5 つのドメインで UR を学習した場合、SURNN に対する提案方式 ($Loss_2$) の優位性は低下する. Q1, Q2 の結果と同様に、この評価でも $Loss_2$ が $Loss_1$ を上回る結果となった.

4.6 考察

Am の Q1 と Q2 の評価結果から、ベースラインと比較して、提案方式はよりドメイン非依存性の高い UR を学習できることが確認できた. Q2 の評価で示したように、提案方式の UR を用いることで、UR の学習元ではないドメインにおけるタスクパフォーマンスが向上する. これは特に、レコメンドなどのタスクを実行する際に、必ずしも対象ドメインのアイテムについてユーザのログが存在しない、あるいは、対象ドメインにおけるログが限られている、といった EC ポータルにおいて有益である. EC ポータルは、提案方式により他のドメインで学習しておいた UR を利用することで、このようなドメインにおけるタスクのパフォーマンスを向上できる.

また、Q3 の評価において、提案方式の UR は性格との関連性が強いことが示された。性格はドメイン非依存の属性の一つであり、EC サービスを含むサービスのパーソナライズの基盤として多くの研究が注目している [73, 102, 32, 10]。もし提案方式の UR が実際に性格を反映していれば（後述するようにこれを確認するためにはさらなる研究が必要であるが）、既存の性格推定手法の 2 つの重要な問題点、すなわち (1) 情報量の少なさ、(2) 信頼性の低さ、を解決できる可能性がある。

2.1.1 節で述べたように、心理学では特性論に基づいて人間の性格を測定するのが一般的である。性格をいくつかの特性（例えば、Big Five における「外向性」、「神経症傾向」、「協調性」、「誠実性」、「知的好奇心」）から構成されるとみなし、質問紙への回答によって各特性のスコアを測定するものである。測定結果は解釈しやすいが、次元数が少ないこと、すなわち性格が少数の特性で表現されていること、またスコアの粒度が粗いこと、すなわちスコアが X-point リッカート尺度や 2 値尺度への回答を合計して算出されるため、連続的ではなく離散的であることから、限られた量の情報しか得られない。加えて、人は時に偏った（例えば社会的に望まれる）回答をしたり、質問文をよく読まずに回答したりすることがあり、これが測定値の信頼性を損なう。日常行動から性格を自動的に推定する研究は数多くあるが（例えば、[145, 86]）、それらはすべて教師あり学習の方式を採用している。つまり、依然として質問紙による測定を Ground truth としているため、(1) と (2) の制約にさらされている。

こうした方式とは対照的に、提案方式はアンケート回答を用いることなく、ユーザとアイテムのインタラクションのシーケンスからドメイン非依存のユーザ属性の高次元表現を学習できる。したがって (1) と (2) の解決策となる可能性があり、本研究はその実現に向けて大きな一歩を踏み出したと考えられる。

課題と今後の研究の方向性

提案方式には今後の研究で解決すべきいくつかの課題がある。一つは DSL への入力データの構築である。RR の結果において、提案方式で学習した UR がベースラインのパフォーマンスを下回ったことが示すように、データに適切なドメインのラベル付けがなされていない場合、提案方式は期待通りに機能しない。人間のアノテーターがデータにラベルを付けるのはコストがかかるし、必ずしも実行可能ではない。データを自動的にラベル付けする機能は将来研究されるべきである。

もう一つの課題は、ドメインの数が増えるにつれて、提案方式の UR のパフォーマンス向上の度合いが減少することである。これは、複数ドメインから学習したドメイン非依存の属性と、そこに新しくドメインを追加することで学習できるドメイン非依存属性との重複が原因であると考えられる。ドメインの数が増えるにつれて、ドメイン非依存の属性が UR に反映されることが、Q1 と Q2 の評価で観察された。すでに多くのドメイン非依存

属性を学習している場合、新しくドメインを追加しても、それ以上学習できることは少ない。一方で、ドメインを追加すると、対応する DSL の追加に伴い、新たに RNN セルの重みとバイアスを学習する必要が生じる（つまり学習可能パラメタの数が増加する）。このように、ドメインの数が増えるにつれて、利得とコストのバランスが悪化する。この解決策の 1 つは、複数の「関連する」ドメインを 1 つの DSL にまとめることである。課題はドメイン間の関係をどのように自動的に決定するかである。これは将来的に研究されるべきである。

最後に、提案方式の UR を長期的かつドメイン非依存の属性の Ground truth と比較し、どのような属性がどの程度反映されているかを検証する必要がある。提案方式の UR が反映するドメイン非依存な属性のうち一部はユーザ自身によって提供されるかもしれない（例えば、ユーザがサービスに登録した際、性別と年齢は手動登録する）。このような場合、UR にユーザ自身が提供した情報を持たせるのは冗長であり、UR の代わりにユーザが提供する情報を利用すれば十分である。したがって、UR に反映される属性を決定し、その属性ごとに UR を学習するように提案方式を拡張する必要がある。これはまた、UR の解釈可能性を向上させ、サービスをパーソナライズするために UR がどのように機能するか解釈性を向上させることになる。

4.7 結論

本章では、EC ポータルにおけるユーザの購買行動から長期的かつドメイン非依存のユーザ属性を反映した UR を学習する方式を提案した。実際の EC ポータルの商品レビューや閲覧ログを用いた実験により、提案方式で学習した UR が既存の方式よりも高いドメイン非依存性を持ち、様々なドメインへの適応性に優れることを示した。また、提案方式の UR が Big Five の性格特性を既存方式の UR よりも強く反映している可能性を確認した。

第 5 章

生体データを用いた心理属性の推定

5.1 背景と目的

ユーザの感情を理解することは、対話 [80, 108], コンテンツ生成 [3], 音楽や動画といったコンテンツのレコメンド [27] など, 様々な Human-AI interaction (HAI) シナリオにおいて非常に重要である. 多くの研究では, ユーザに明示的に感情を尋ねることによって, 感情の Ground truth を収集しているが, 実世界のシナリオでそれを行うことはユーザエクスペリエンスを低下させるため, 非現実的である. そのため, ユーザが生成する各種データからユーザの感情を自動的に認識することが強く求められている.

ユーザが生成する様々なデータの中で, 筆者は脳波 (Electroencephalography; EEG), 心電図 (Electrocardiography; ECG), 皮膚電位 (Galvanic skin response; GSR) などのユーザの生体データに注目した. こうした生体データはウェアラブルデバイス (リストバンド型デバイスやイヤホン, VR ゴーグルなど) を使用することで収集できるため, テキストや声色, 顔の表情など, 収集するに当たってユーザが何かを書いたり話したり, カメラの前にいたりする必要がある他のデータに比べ, 収集に伴う制約が少ない. さらに, これらのデータとは異なり, 生体データは, ユーザが意図的に感情を隠そうとしているときでも, 感情に関する多くの情報を提供することが知られている [64].

これまで, 生体データから感情を認識する手法が数多く研究され, 生体データが感情認識に有用であることが確認されてきた [80, 131]. しかし, 生体データが感情だけでなく, 感情とはほとんど関連性を持たない様々な要因 (ファクター) の影響を受けることについては, あまり注目されてこなかった. その中には, 単一の生体モダリティ (例えば心臓の動き) のみに影響を与えるファクター, すなわちモダリティ特有のファクター (Modality-specific factor) がある. 例えば, 心臓の筋肉の強さは EEG に影響を与えるが, 脳活動や皮膚電位など心臓の活動以外の生体モダリティにはほとんど影響しない. これに対して, 感情は様々な生体モダリティに影響するモダリティ横断的なファクタ

(Cross-modal factor) であり、複数の生体モダリティにまたがる反応を引き起こす。例えば、怒りは心拍数と皮膚電位レベルを同時に増加させることが知られている。また、生体活動に影響する他のファクターとしては年齢、体格や性別といったファクターもある。これらのファクターは非常にゆっくりと変化するか、全く変化しない（すなわち Long-term factor である）。一方で感情は短時間で変化するファクターである（すなわち Short-term factor である）。

このように、感情は Cross-modal かつ Short-term なファクターである一方、生体データは Modality-specific および/または Long-term なファクターの影響を受け、これらを反映する。これらのファクターは感情認識にはほとんど役に立たないが、既存のアプローチはこれらのファクターを区別することなく生体データから特徴量を抽出し、利用してきた。これにより感情認識の精度が劣化すると考えられる。

以上を踏まえ、生体データから Cross-modal かつ Short-term なファクターのみを反映した特徴量を抽出する方式を提案する。生体データに反映されるファクターを Long-term/Short-term, Modality-specific/Cross-modal の2軸で区別し、異なる4つのファクターを区別して抽出する。提案方式は RNN を採用することで、Long-term/Short-term なファクターを別々に学習する。

提案方式の特徴的な点は、Cross-modal, Modality-specific なファクターを区別して学習するために、2種類のレイヤで構成されるマルチレイヤ RNN を導入した点である。すなわち、提案方式のマルチレイヤ RNN は、個々の生体モダリティの活動をモデル化する複数の“Modality-specific Layer” (MSL) と、感情などの Cross-modal factor をモデル化する単一の“Cross-modal Layer” (CML) で構成される。提案方式は、複数の生体モダリティ（例えば、ECG と GSR 信号）について、時系列の生体データを入力として受け取る。各 MSL には対応するモダリティの生体データを入力し（例えば、MSL1 には ECG を、MSL2 には GSR を入力する）、隠れ状態に生体モダリティの状態を反映させる（MSL1 の隠れ状態は心臓の状態を、MSL2 の隠れ状態は皮膚の状態を、それぞれ反映）。隠れ状態を更新する際、MSL は自身の隠れ状態だけでなく、CML の隠れ状態も利用する。CML の隠れ状態を利用した隠れ状態の更新はすべての MSL で行われるため、CML の隠れ状態は、MSL が対応するすべての生体モダリティの状態に影響を与えることになる。これにより、CML は複数のモダリティにまたがる生体活動に影響を与えるファクター、すなわち Cross-modal factor を学習することが可能になる。

提案方式を評価するため、被験者を募集し、音楽と動画クリップ（すなわち刺激）を提示しながら、EEG, ECG, GSR を測定した。これらの生体信号によって RNN を訓練し、CML の隠れ状態を用いて、各刺激を提示した後に被験者が報告した感情をどれだけ正確に認識できるかを評価した。

5.2 関連研究

提案方式と同様に、生体データから感情認識を行う既存方式の多くは、複数モダリティの生体データを利用している。Subramanian ら [136] と Miranda ら [87] は、ECG, GSR, EEG を用いた方式を提案した。今日広く利用されている特徴量抽出技術を用いて、各モダリティから特徴量（例えば、ECG からは心拍間隔の標準偏差、GSR から皮膚コンダクタンスレベルの平均値など）を抽出し、これらの生体特徴量を連結（concatenate）して、分類器（SVM など）に入力した。しかし、生体特徴量に反映される Modality-specific factor は、単純な連結では除去できないため、感情認識精度には限界がある。また、彼らの提案方式では、Long-term factor と Short-term factor の区別ができない。Subramanian と Miranda は、それぞれの生体モダリティ単独で感情認識を行い、その結果を統合して最終的な結果を導く評価も行ったが、Modality-specific factor によって各モダリティの感情認識精度が制約されるという課題が残った。

ディープラーニング技術を採用した方式も提案されているが、上述の方式と同様の課題がある。Liu ら [77] と Yin ら [153] は、Deep autoencoder を用いて、複数のモダリティ（EEG や Electrooculogram など）の生体特徴量の Shared representation を学習し、Shared representation を分類器（SVM や多層パーセプトロンなど）に入力することで感情を認識する方式を提案した。これらの方式では、各モダリティの生体特徴量を Shared representation から再現できるように、Autoencoder を訓練した。これにより、Shared representation には Cross-modal factor が反映されるものの、同時に Modality-specific factor も反映されてしまう。また、Autoencoder を用いても、Long-term factor と Short-term factor を区別することはできない。

一方で、Li ら [72] が提案したアプローチは、生体データから Short-term factor のみを反映する特徴量を抽出できる。Li らは Miranda ら [87] が構築したデータセットを用いて、生体特徴量の時系列シーケンスを LSTM に入力し、その隠れ状態を Attention network に入力した。これらのステップにより、Li らの方式では感情に特に関連性の強い時系列区間を特定することが可能となる。そこから隠れ状態を抽出して多層パーセプトロンに入力し、感情を認識した。しかし、Li らは生体モダリティごとにこれらの処理を実行し、個々の生体モダリティの結果を組み合わせる最終的な結果を導き出した。そのため、[136, 87] と同様に、個々のモダリティにおける感情認識は、Modality-specific factor による制約を受けていた。

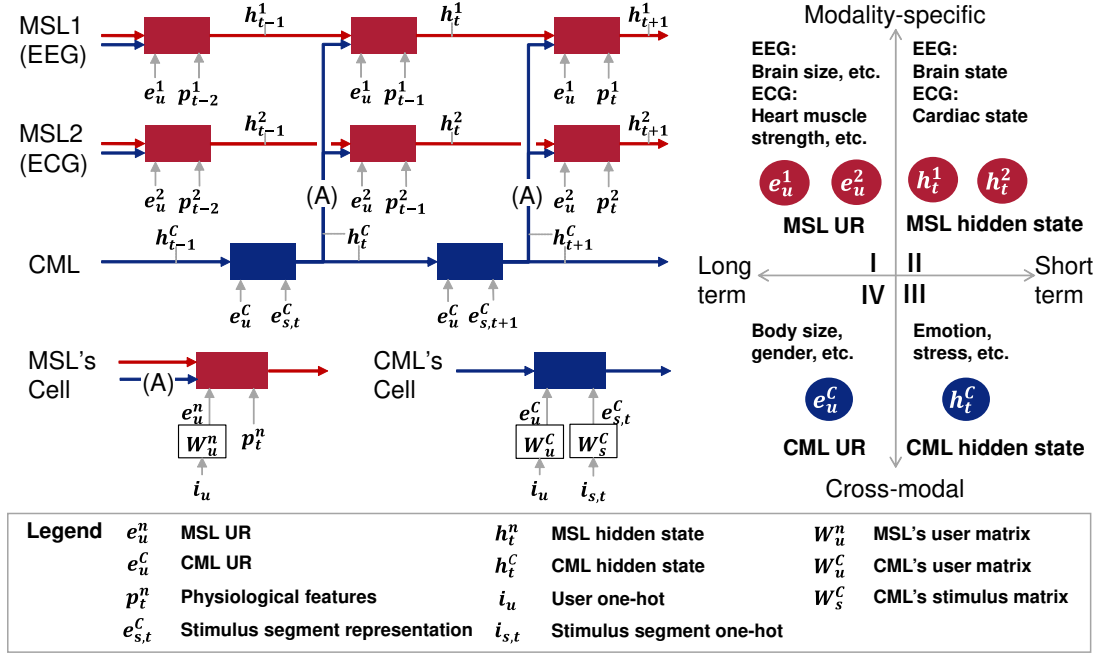


図 5.1 提案方式の概要

5.3 提案方式

既存方式とは対照的に、提案方式は生体活動に影響を与える 4 種類のファクターを明確に区別してモデリングする。図 5.1 に提案方式の概要を示す。図の左側に示しているのはマルチレイヤ RNN の例であり（図では MSL が二つのみの構成を示しているが、入力する生体モダリティの数分だけ MSL が設けられる）、図右側には生体活動に影響を与える 4 つのファクター、I~IV、がマルチレイヤ RNN の変数にどのようにマッピングされるかを示している。Modality-specific factor である I と II は MSL によってモデリングされる。各 MSL は単一の生体モダリティに対応し、例えば MSL1 は脳波、MSL2 は心電図に対応する。MSL は以下を入力として受け取る：

- 対応するモダリティの生体特徴量
既存方式 ([87, 136] など) と同じ手法で抽出される。
- ユーザ ID を示す One-hot vector
これによってユーザ情報を格納した行列 (User matrix) からユーザの特徴表現 (User representation; UR) が取り出される。

MSL に入力される生体特徴量は当該 MSL が対応する生体モダリティに限定されるため、UR と隠れ状態は入力された生体モダリティの Modality-specific factor を反映する

表 5.1 記号表記の一覧

Notation	Meaning	Subscript	Superscript
N_u, N_s	# of users/SS	u : # of users, s : # of stimulus segments (SS)	-
$\mathbf{i}_u \in \mathbb{R}^{N_u}, \mathbf{i}_{s,t} \in \mathbb{R}^{N_s}$	One hot vector of ...	u : user ID, s, t : SS ID of t -th action	-
$d_{ue}^c, d_{ue}^n, d_{se}^c, d_h^c, d_h^n, d_p^n$	# of dimensions of ...	ue : UR, se : SS Representation (SR), h : hidden state, p : physiological features	-
$\mathbf{W}_u^c \in \mathbb{R}^{d_{ue}^c \times N_u}, \mathbf{W}_u^n \in \mathbb{R}^{d_{ue}^n \times N_u}$	User matrix	u : user matrix	-
$\mathbf{W}_s^c \in \mathbb{R}^{d_{se}^c \times N_s}$	SS matrix	s : SS matrix	C : CML n : MSL n
$\mathbf{e}_u^c \in \mathbb{R}^{d_{ue}^c}, \mathbf{e}_u^n \in \mathbb{R}^{d_{ue}^n}, \mathbf{e}_{s,t}^c \in \mathbb{R}^{d_{se}^c}$	UR/SR	u : UR, s, t : SR of t -th action	-
$\mathbf{h}_t^c \in \mathbb{R}^{d_h^c}, \mathbf{h}_t^n \in \mathbb{R}^{d_h^n}$	Hidden state	t : hidden state after t -th action	-
$\mathbf{p}_t^n \in \mathbb{R}^{d_p^n}$	Physiological features	t : physiological features in t -th action	-

(I と II). さらに、隠れ状態は時系列の生体特徴量を入力する度に逐次更新されるが、User matrix (UR のセット) は入力によらず同じ値を保つ. これにより、MSL の UR は Long-term factor (I) を反映し、隠れ状態は Short-term factor (II) を反映する.

一方、Cross-modal factor である III と IV は CML によってモデリングされる. 図中のリンク (A) に示すように、CML の隠れ状態は MSL に入力される. MSL のセルは入力された CML の隠れ状態を、入力データ (生体特徴量と UR) および自身の前時刻の隠れ状態とともに用いて、自身の隠れ状態を更新する. 更新された MSL の隠れ状態は、次時刻の生体特徴量を予測するために使用されるので、生体活動の状態を表しているとみなすことができる. このように MSL の隠れ状態を CML の隠れ状態を用いて更新することで、CML の隠れ状態が各モダリティの生体活動に影響を与えるプロセスがモデリングされる. すべての MSL がこのように隠れ状態を更新するため、CML は複数のモダリティの生体活動に影響を与えるファクター、すなわち Cross-modal factor (III と IV) を学習する. CML においても MSL と同様に、Short-term factor (III) が隠れ状態に、Long-term factor (IV) が UR に反映されるが、MSL と異なり、これらは Cross-modal である.

I~IV を分離してモデリングするだけでなく、提案方式は個人間の生理的な違い (臓器の大きさや筋量の違いなど) が感情と生体活動の関係を調整する (Moderate する) プロセスもモデリングする. 例えば、心臓の筋力の強さが異なるユーザは、仮に全く同じ感情になったとしても出力する ECG は異なると考えられる. 提案方式は、CML の隠れ状態 (感情) と MSL の UR (個人差、例えば心臓の筋力の強さの違い) の両方を用いて MSL の隠れ状態 (生体活動を反映) を更新することで、このような個人差の調整効果 (Moderating effect) をモデリングする. この点も、5.2 節で説明した既存方式と異なる点である.

次節では、提案方式の隠れ状態の更新および訓練プロセスについて詳細に述べる. 次節で利用する記号表記については表 5.1 を参照されたい.

5.3.1 隠れ状態の更新

CML と MSL には次の形式のデータが入力される.

$$\text{CML: } data_a^C = [x_{a,1}^C, x_{a,2}^C, \dots, x_{a,t}^C, \dots, x_{a,T}^C] \text{ and} \quad (5.1)$$

$$\text{MSL } n: data_a^n = [x_{a,1}^n, x_{a,2}^n, \dots, x_{a,t}^n, \dots, x_{a,T}^n]. \quad (5.2)$$

ここで, $x_{a,t}^C = (i_u, i_{s,t})$ はユーザ a の t 番目の行動 (例えば動画クリップ M の t 番目のセグメントを視聴した) を表し, $x_{a,t}^n = (i_u, p_t^n)$ はユーザ a が t 番目の行動をとっている間に観測された生体信号から抽出した生体特徴量を表す. $x_{a,t}^C$ が CML に入力されると, CML は最初に UR と刺激 (動画や音楽など) セグメント (Stimulus segment; SS) の特徴表現 (SS representation; SR) をユーザ行列 (W_u^C), SS 行列 (W_s^C) からそれぞれ次のように取り出す. $e_u^C = W_u^C i_u$, $e_{s,t}^C = W_s^C i_{s,t}$. その後, 自身の隠れ状態 h_t^C を以下のように更新する.

$$h_t^C = f^C(h_{t-1}^C, e_u^C, e_{s,t}^C). \quad (5.3)$$

ここで, f^C は LSTM で実装される関数であり, 詳細は後述する.

CML は隠れ状態を更新すると, 更新後の隠れ状態を, リンク (A) を通じて全ての MSL に送信する. これは CML が隠れ状態を更新する度に行われる. MSL は h_t^C を受け取ると, UR をユーザ行列から取り出し ($e_u^n = W_u^n i_u$), 入力された生体特徴量 (p_{t-1}^n) とこれらを用いて隠れ状態 h_t^n を以下のように更新する.

$$h_t^n = f^n(h_{t-1}^n, h_t^C, e_u^n, p_{t-1}^n). \quad (5.4)$$

ここで, f^n は LSTM で実装される関数であり, 詳細は後述する.

CML, MSL の隠れ状態の更新関数 (f^C, f^n)

以下で, CML, MSL の LSTM セルにおける隠れ状態の更新プロセスの詳細について説明する. 隠れ状態の更新は通常の LSTM セルと同様に行われる. 更新プロセスは入力, 忘却, 出力ゲート (I_t, F_t, O_t), および Candidate memory, Present memory ($\tilde{C}_t, C_t$) を用いて以下のように行われる.

$$I_t = \sigma(W_{ei}e + W_{hi}h + b_i), \quad (5.5)$$

$$F_t = \sigma(W_{ef}e + W_{hf}h + b_f), \quad (5.6)$$

$$O_t = \sigma(W_{eo}e + W_{ho}h + b_o), \quad (5.7)$$

$$\tilde{C}_t = \tanh(W_{ec}e + W_{hc}h + b_c), \quad (5.8)$$

$$C_t = F_t \odot C_{t-1} + I_t \odot \tilde{C}_t, \quad (5.9)$$

$$h_t = O_t \odot \tanh(C_t). \quad (5.10)$$

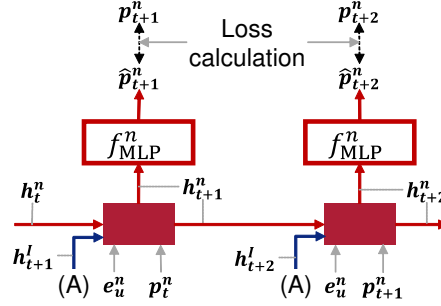


図 5.2 ロスの計算

上記のモデル式において、 σ と $\odot$ は、それぞれシグモイド関数、要素毎の積 (Element-wise product) を表す。 e と h は LSTM セルへの入力であり、 W と b は学習可能なパラメタである。上記モデル式は CML と MSL のセルで共通だが、 e と h が CML と MSL では異なる形式をとる。

CML (つまり f^C) において、 e と h は以下で表される。

$$e = e_u^C \oplus e_{s,t}^C, \quad (5.11)$$

$$h = h_{t-1}^C. \quad (5.12)$$

ここで、 $\oplus$ は二つベクトルの連結を表す。 f^C の出力 (h_t) は h_t^C である。

一方で、MSL n (つまり f^n) では、 e と h は以下で表される。

$$e = e_u^n \oplus p_{t-1}^n, \quad (5.13)$$

$$h = h_{t-1}^n \oplus h_t^C. \quad (5.14)$$

f^n の出力 (h_t) は h_t^n である。

5.3.2 マルチレイヤ RNN の訓練

$data_a^n$ が入力されると、MSL は各時刻における生体特徴量の予測を行う。例えば、入力が $data_a^n = [x_{a,t}^n, x_{a,t+1}^n, \dots, x_{a,T-1}^n]$ の場合、予測出力は $[\hat{p}_{t+1}^n, \hat{p}_{t+2}^n, \dots, \hat{p}_T^n]$ となる。生体特徴量の予測結果は実際の生体特徴量と比較され、ロスが計算される。ロスは MSL 及び CML のセル内のパラメタと、ユーザ行列、SS 行列 (W_u^n, W_u^C, W_s^C) の学習に用いられる。図 5.2 に予測とロス計算の概要を示す。MSL は自身の隠れ状態を用いて、生体特徴量の予測を次のように行う。 $\hat{p}_{t+1}^n = f_{MLP}^n(h_{t+1}^n)$ 。なお、ここで f_{MLP}^n は ReLu を活性化関数とする多層パーセプトロンである。MSL はその後生体特徴量の実際の値と実測値の残差平方和 (Residual sum of square) を計算し、これをロスとして用いる。

5.4 実験

データセットを構築し、CMLの隠れ状態が感情をどの程度反映しているのか評価した。評価は以下の3ステップで実施した。

ステップ1 感情特徴量の抽出。 データセットに格納された生体特徴量からさらに感情認識のための特徴量（以下、感情特徴量と呼ぶ）を抽出した。提案方式では、CMLの隠れ状態を感情特徴量として使用した。

ステップ2 感情特徴量の選択。 LASSOにより感情特徴量の選択を行った。

ステップ3 線形回帰。 選択した感情特徴量を用いて感情を識別するためのモデルを構築し、モデルフィットと識別精度を評価した。

データセットは二つ構築し、それぞれのデータセットについて上記1~3を実施した。また、実際のシナリオにおいて、感情認識に利用可能な生体モダリティはユーザが使用しているデバイスに依存するため、1~3は取り得る全ての生体モダリティの組合せ、すなわち A) EEG+ECG+GSR, B) EEG+ECG, C) EEG+GSR, D) ECG+GSR すべてについて実験を行った。

5.4.1 データセット

今日、感情と生体データを格納したデータセットが、一般に利用可能な状態で複数公開されているものの（例えば [87, 136]），データセットは独自のものを構築した。これらのデータセットの問合せ先が筆者からのコンタクトに応答しなかったという理由もあるが、他の理由として、これらのデータセットが被験者に提示する刺激として動画のみを用いていた点も挙げられる。音楽は仕事や勉強など他の活動を行いながら楽しむ刺激としては動画よりもよく利用されているため、データセットを構築する際、被験者には動画に加え音楽も提示した。

被験者とデータ収集

正常な聴力と視力を持つ日本人 54 名（平均年齢：男性-38.2 歳，女性-37.3 歳）を募集した。データ収集は実験施設で行った。EEG, ECG, GSR を測定しながら、PC ソフトを使って音楽と動画（音声付き）を提示した。認知的・肉体的疲労を避けるため、音楽と動画は異なる日に提示した（1 日目は音楽，2 日目は動画）。なお，1 日目のデータ収集が終了した後に 2 名が脱落したため，2 日目は 52 名からデータ収集を行った。

各日のデータ収集セッションは 2 時間で，その間にソフトウェアによって，用意された

刺激セットからランダムに刺激が選択された。各刺激を提示した後、被験者はソフトウェアを通じて自分の感情を報告した。

被験者がソフトウェアを操作する速さによって、提示される刺激の数は異なった。また、刺激の選択は無作為に行われたため、提示された刺激は被験者によって異なった。音楽、動画それぞれで、のべ 2,336 回、2,119 回の試行（刺激を一つ視聴し、感情を回答する行為を 1 試行とカウント）が行われた。

データ収集は、国立研究開発法人 情報通信研究機構の倫理委員会の承認を得て実施した。被験者には事前に実験内容を十分に説明し、同意を得た上でデータ収集を行った。

刺激

被験者に提示する刺激を準備するために、まず著作権フリーのコンテンツ集を参照し、約 100 曲の音楽と動画クリップをランダムに選んだ。次に、それぞれの音楽や動画クリップから誘発される感情を評価するため、3 人のスタッフに後述の 6 つの感情軸に従って感情を報告させた。その回答に基づいて、刺激の平均感情スコアができるだけ分散するように、70 曲の音楽と 42 本の動画を選択した。最後に、各音楽／動画から最初の 60 秒間を抜き出し、データ収集に利用した。

感情測定

感情は、次の 6 つの軸で 0~15 点の間で回答させた。

- (a) 悲しい-嬉しい
- (b) 不安-安心
- (c) 怖い-落ち着く
- (d) 退屈-興奮
- (e) せつない-胸躍る
- (f) 怒り-優しさ

感情測定には Russel の円環モデル [120] を利用する研究が多いものの、今回のデータ収集ではこれを用いなかった。同モデルで使用される“arousal（覚醒度）”と“valence（感情価）”の概念が日本人の被験者には理解しづらいためである。上述の (a) ~ (f) の 6 軸は覚醒度と感情価の 2 軸で構成される感情平面を可能な限りカバーするように選択した。

生体データの収集

生体データの測定には、横田ら [154] が開発した 11 個の電極を備えたデバイスを利用した。電極の構成は次のとおりである。EEG-5 個、ECG-1 個、GSR-1 個、EOG (Electrooculogram, EEG から眼球の動きや瞬きによるノイズを除去するために使用) -2

表 5.2 抽出した生体特徴量（太字は次元数を表す）

Modality	Window features	Stimulus features
EEG	29 : Power within individual frequency bands ($\delta, \theta, \alpha, \beta$) and composite score (S) at each electrode; power asymmetry of individual frequency bands for one pair of electrodes	116 : Statistics (mean, min, max, and standard deviation) of 29 window features
ECG	7 : SDNN, pNN50, RMSSD, and HR statistics (mean, min, max, and standard deviation)	10 : SDNN, pNN50, RMSSD, and HR statistics (mean, min, max, and standard deviation) and power within HF, LF and power ratio (HF/LF)
GSR	8 : SCL related features (slope, mean) and SCR related features (peak, skew, kurtosis, max, min, mean)	11 : SCL related features (slope and mean) and SCR related features (the number of peaks, max and min of peaks, skew, kurtosis, max, min, and mean)

SDNN: Standard Deviation of Normal-to-Normal intervals, pNN50: The proportion of NN50 divided by the total number of NN intervals, RMSSD: Root Mean Square of the Successive Differences, HR: Heart Rate, HF/LF: High/Low Frequency, SCL: Skin Conductance Level, SCR: Skin Conductance Response

個，リファレンスおよびグラウンド-2 個． EEG の電極は，国際 10-20 法 [81] にしたがって，Fpz, C3, Cz, C4, Oz の頭部各所に配置した．すべての電極から 500 Hz のサンプリングレートで電位を測定した．

生体特徴量の抽出

生体データを収集した後，[136, 87] と同様に，各生体モダリティで広く用いられている特徴量抽出技術を用いて生体特徴量を抽出した．抽出した生体特徴量はウィンドウ特徴量と刺激特徴量の 2 種類である．概要を表 5.2 に示す．ウィンドウ特徴量については，1 つの刺激を提示している間に測定された生体データにスライディングウィンドウを適用し，各ウィンドウから特徴量を抽出した．ウィンドウサイズは 10 秒とし，スライドサイズは 3 秒と 5 秒の 2 種類を用いた．すなわち，各刺激（60 秒）にそれぞれ 17 個と 11 個のウィンドウを用意した．刺激特徴量は，1 つの刺激を提示している間に測定された生体データ全体から抽出した．生体特徴量の各次元について z 標準化を行った後，データセットに格納した．

各生体モダリティからの生体特徴量抽出の詳細を以下に示す．

EEG 各電極で測定した信号を用いて，デルタ（1-4Hz），シータ（4-8Hz），アルファ（8-12Hz），ベータ（12-30Hz）の各周波数帯のパワー値を計算した．また，Shestiyukらの報告 [129] を元にアルファのパワー減少とそれに伴うシータのパワーの増加も計算した (S)*¹．さらに，4 つの周波数帯それぞれについて，C3 と C4 のパワーの非対称性（C3-C4）を計算した．

ECG ECG 測定値から心拍数（HR）と心拍変動（HRV）に関連する特徴量を抽出した．時間ドメインでは，HR の平均値，最小値，最大値，NN 間隔の標準偏差（SDNN），

*¹ 人間の注意プロセスと関係があると報告されている

NN 間隔の連続差の二乗平均平方根 (RMSSD) を計算した。また、周波数ドメインでは、低周波帯域 (0.04-0.15Hz ; LF) と高周波帯域 (0.15-0.4Hz ; HF) のパワー、および LF と HF のパワーの比 (LF/HF) を用いた。各特徴量の詳細については、[127] を参照した。なお、周波数ドメインの特徴量については、[127] によれば、信頼できる値を計算するには少なくとも 30 秒間の測定が必要であるため、ウィンドウ特徴量については、周波数領域の特徴量は算出しなかった。

GSR GSR 信号は皮膚コンダクタンスレベル (SCL) と皮膚コンダクタンス反応 (SCR) で構成される。SCL はゆっくりと変化する信号で、注意の度合いと関連している。一方、SCR は散発的な刺激に対する反応であり、特定の時間窓内にのみ現れる高周波成分を含む。これらの成分から別々に特徴量を抽出した。SCL については、平均レベルと傾き、すなわち、ある期間 (ウィンドウまたは刺激全体) において減少/増加する程度を計算した。SCR については、ピークレベルと統計量 (歪度, 尖度, 最大値, 最小値, 平均値) を算出した。

5.4.2 ステップ 1-感情特徴量の抽出

提案方式

二種類の生体特徴量のうち、提案方式ではウィンドウ特徴量を入力として用いた。つまり、CML と MSL n へ入力されるシーケンス ($data_a^C$ と $data_a^n$) は、1 回の試行 (一人の被験者が一つの刺激を視聴) に対応する。 $data_a^C$ および $data_a^n$ の要素 ($i_{s,t}$, p_t^n) は刺激の t 番目のウィンドウと、当該ウィンドウから抽出された生体特徴量に対応する。入力シーケンスの総数は試行の総数と同じであり、そのうち 80% を訓練に、20% を検証 (validation) に、それぞれ用いた。なお、提案方式では刺激特徴量は入力として使用しなかった。仮に使用した場合、 $i_{s,t}$ は 1 つの刺激に対応し、入力シーケンスの数が被験者の数分しか確保できず、十分な量のデータで提案方式の RNN を訓練できないためである。

次のパラメタをハイパーパラメタとした。スライディングウィンドウのスライドサイズ = [3sec, 5sec], 学習率 = [5×10^{-4} , 1×10^{-3}], MSL の多層パーセプトロン (f_{MLP}^n) の隠れ層のサイズ = [(16, 8), (32, 16)] (入力に近い順に記載), バッチサイズ = [16, 32], UR, SR, CML と MSL の隠れ状態の次元数 = [8, 16]。ハイパーパラメタの全ての組合せについて、訓練と検証を 100 エポック実施し、検証データにおいてロスが最小となった時点の CML の隠れ状態を抽出した。このプロセスを訓練・検証データを入れ替えて実施し、全ての試行について CML の隠れ状態を取得した。認識対象は各試行後の感情であるため、感情特徴量としては CML 隠れ状態のうち、各試行の最後の隠れ状態を使用した。例えば、 $data_a^C$ の最後の要素が $i_{s,T}$ の場合、 h_T^C を感情特徴量として使用した。

ベースライン

Miranda ら [87] と Subramanian ら [136] の方式を参考に、まず刺激特徴量とウィンドウ特徴量の両方について、生体モダリティ間で生体特徴量を連結した。例えば、モダリティの組み合わせが A)EEG+ECG+GSR の場合、刺激特徴量から 137 ($116 + 10 + 11$) 次元の特徴量を、ウィンドウ特徴量から 748 ($(29 + 7 + 8) \times 17$) 次元の特徴量を各試行について構築した（刺激に 17 個のウィンドウがある場合）。次に PCA によって次元縮小を行い、累積寄与率が閾値以下で最大となるように寄与率上位 n 個の特徴量を抽出し、これらの特徴量を感情特徴量とした。閾値は 0.85, 0.90, 0.95 の 3 種類を設定した。なお、以下では、“ S ” は刺激特徴量から、“ W ” はウィンドウ特徴量から抽出した感情特徴量を表す。なお、Miranda らは、単一の生体モダリティを利用した感情認識が複数モダリティを利用した感情認識よりも精度が優れていたケースがある旨報告しているため、個別の生体モダリティからも S と W を抽出した。

5.4.3 ステップ 2 と 3 - 感情特徴量の選択と線形回帰

感情特徴量を抽出した後、特徴量選択と線形回帰を 6 つの感情次元それぞれに対して行った。

特徴量選択を行ったのは、ベースラインの感情特徴量の次元がサンプルサイズ（すなわち試行数）に対して大きかったためである。公平な比較のため、特徴量選択はベースラインと提案方式の両方で行った。最初に、LASSO のパラメタ λ の調整を行った。 λ は選択した特徴量の数に基づいて、正則化の強さを制御するパラメタである。予め用意した複数の λ の値のセットに対して、5-fold の交差検証を行い、最も正確な予測（感情スコアの実測値と予測値の平均二乗誤差が最小）を出力する値を求めた。次に、前ステップで決定した λ の値を用いて再度 LASSO 回帰を行い、回帰係数が 0 でない特徴量を選択した。

特徴量選択の後、2 種類の線形回帰を行った。一つは全サンプルを用いたモデルフィット評価であり、もう一方は、5-fold 交差検証による予測精度の評価である。

5.4.4 Ablation Study

提案方式の主要な特徴の効果を評価するために、それらを提案方式から除去した RNN を用意し、評価を行った。図 5.3 に評価した RNN を示す。一つは単一レイヤの RNN (AB1) であり、複数モダリティの生体特徴量を連結したもの（図中の $p_{s,t}$ ）を入力した。もう一つは、MSL UR を使用しない多層 RNN (AB2) である。双方から隠れ状態を抽出して（AB2 からは CML の隠れ状態を抽出）感情特徴量として使用し、提案方式と同じよ

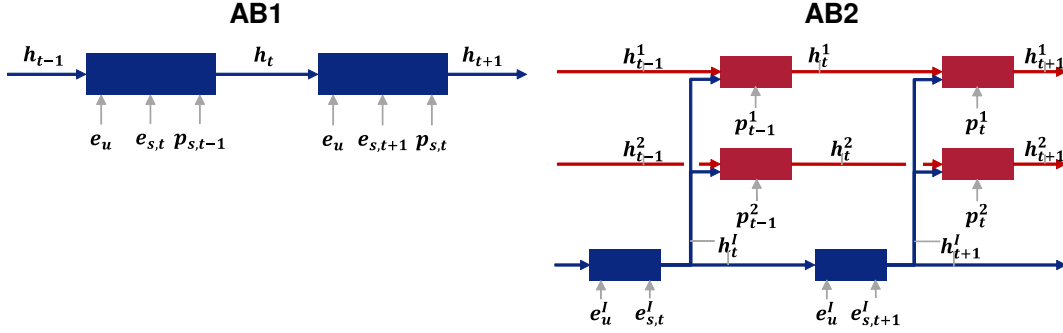


図 5.3 Ablation study で評価した RNN

うに評価した。

Li らの方式 [72] と同様に, AB1 では Short-term factor のみを反映した感情特徴量を抽出できるものの, Modality-specific factor と Cross-modal factor を区別して抽出できない. 一方, AB2 では Short-term かつ Cross-modal なファクターを抽出できるものの, AB2 の MSL は UR (e_u^n) を用いないため, 生体的な特徴の個人差による調整効果 (Moderating effect) がモデリングできない.

5.5 結果と考察

表 5.3 に結果を示す. 表に示すおとり, 刺激の種類 (音楽/動画), 感情次元, 生体モダリティの組み合わせによらず, ほとんどの条件で, 提案方式 (ours) の回帰モデルは, ベースライン (BL, U-BL), AB1, AB2 より良好な結果を示した.

生体データに影響を与える 4 種類のファクターを全く区別しないベースラインと比較して, 提案方式は 1 つの例外 (動画データセットの (b)-C の RMSE) を除いて, 両データセットのすべての条件において, 良好な性能を示した. 提案方式とベースラインの差は, 両者の回帰モデルの AIC から計算される相対尤度 (Relative likelihood; RL) から統計的に有意と言える. 相対尤度は, 次の式で計算され, 二つのモデルのうち, AIC が大きいモデルが, 小さいモデルよりも真のモデル (観測データに完全にフィットする仮想モデル) に近いことの尤度を表す.

$$RL = \exp\left(\frac{AIC(\text{ours}) - AIC(\text{BL or U-BL})}{2}\right). \quad (5.15)$$

ここで, $AIC(M)$ は回帰モデル M の AIC を表す. 表 5.4 に示すおとり, 全ての条件において提案方式と比較したベースラインの相対尤度は 0.05 を下回っていた. この結果は, 提案方式により抽出した感情特徴量が, ベースラインのそれよりも, 感情を反映する程度が有意に強いことを示している.

表 5.3 感情認識の結果． A-D は生体モダリティの組合せを表す． A) EEG+ECG+GSR, B) EEG+ECG, C) EEG+GSR, D) ECG+GSR. 各セルの色と斜線有無の意味は次のとおり． 灰色セル： 同じ条件（列）の提案方式（ours）の結果に劣ることを示す． 斜線有のセル： LASSO によって感情特徴量が一つも選択されなかったことを示す． 黒色セル： 該当する感情次元において最良の結果であることを示す． ベースライン（BL, U-BL）については， 3つの PCA 閾値のうち最良の結果を示した． また， U-BL は BL において単一の生体モダリティを使用し， 各列に対応するモダリティの組み合わせにおける最良の結果を示す（例えば， B 列の U-BL は EEG と ECG の良い方の結果を示す）．

		(a) sad-happy				(b) nervous-relaxed				(c) fear-relieved				(d) lethargic-excited				(e) depressed-delighted				(f) angry-serene			
		A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Results of Music dataset (n=2,336)																									
AIC ↓	BL(S)	11231	11235	11235	11234	11495	11501	11495	11503	11145	11145	11145	11162	11553	11553	11554	11544	12205	12205	12205	12205	10416	10426	10408	10430
	BL(W)	11235	11235	11235	11234	11503	11503	11503	11505	11150	11150	11151	11151	11521	11547	11533	11534	12191	12206	12181	12201	10414	10420	10420	10426
	U-BL(S)	11237	11237	11237	11237	11499	11499	11499	11503	11145	11145	11145	11160	11551	11551	11554	11551	12205	12205	12205	12211	10415	10415	10415	10428
	U-BL(W)	11234	11236	11234	11234	11503	11503	11503	11504	11151	11151	11151	11154	11543	11549	11543	11543	12198	12201	12198	12198	10416	10416	10416	10426
	AB1	11197	11128	11160	11206	11486	11439	11474	11486	11119	11113	11121	11142	11508	11493	11478	11499	12170	12044	12138	12182	10402	10380	10394	10429
	AB2	11130	11107	11166	11133	11431	11426	11454	11463	11097	11076	11088	11129	11421	11390	11495	11398	12086	12084	12116	12097	10350	10344	10350	10403
	Ours	11017	11059	11078	11077	11363	11398	11381	11411	11068	11060	11039	11094	11346	11343	11437	11378	11989	11961	12050	12057	10318	10376	10337	10346
RMSE ↓	BL(S)	2.232	2.232	2.235	2.233	2.422	2.430	2.423	2.434	2.159	2.159	2.159	2.149	2.392	2.391	2.392	2.401	2.925	2.926	2.926	2.926	1.856	1.865	1.851	1.860
	BL(W)	2.232	2.232	2.235	2.233	2.433	2.433	2.435	2.435	2.157	2.157	2.159	2.149	2.386	2.387	2.394	2.395	2.910	2.923	2.902	2.925	1.854	1.858	1.857	1.857
	U-BL(S)	2.231	2.233	2.231	2.231	2.428	2.428	2.428	2.434	2.154	2.154	2.159	2.154	2.392	2.392	2.392	2.397	2.923	2.923	2.923	2.928	1.859	1.860	1.859	1.859
	U-BL(W)	2.229	2.229	2.229	2.229	2.433	2.433	2.433	2.435	2.151	2.156	2.151	2.151	2.388	2.388	2.388	2.395	2.921	2.921	2.921	2.925	1.858	1.858	1.858	1.859
	AB1	2.218	2.158	2.199	2.217	2.410	2.371	2.396	2.414	2.159	2.150	2.155	2.160	2.389	2.357	2.364	2.376	2.889	2.778	2.850	2.895	1.847	1.841	1.838	1.858
	AB2	2.185	2.167	2.197	2.196	2.363	2.355	2.384	2.391	2.144	2.125	2.164	2.158	2.339	2.321	2.386	2.321	2.833	2.829	2.834	2.842	1.827	1.822	1.833	1.841
	Ours	2.135	2.136	2.161	2.158	2.318	2.342	2.330	2.367	2.135	2.131	2.121	2.146	2.271	2.286	2.331	2.312	2.752	2.743	2.797	2.815	1.797	1.808	1.806	1.828
Results of Movie dataset (n=2,119)																									
AIC ↓	BL(S)	8782	8782	8782	8782	9227	9227	9227	9227	9195	9195	9195	9197	9807	9825	9821	9821	9698	9699	9699	9699	9251	9282	9283	9274
	BL(W)	8781	8782	8781	8786	9228	9228	9229	9229	9195	9195	9194	9197	9812	9799	9814	9814	9694	9693	9699	9690	9290	9290	9287	9284
	U-BL(S)	8782	8782	8782	8790	9227	9227	9227	9235	9192	9192	9192	9192	9801	9801	9829	9801	9689	9689	9699	9689	9268	9268	9283	9268
	U-BL(W)	8780	8782	8780	8780	9214	9228	9214	9214	9171	9194	9171	9171	9795	9795	9823	9795	9686	9686	9695	9686	9264	9286	9264	9264
	AB1	8778	8777	8763	8776	9182	9223	9214	9211	9133	9186	9135	9164	9764	9774	9821	9781	9673	9668	9695	9670	9197	9210	9225	9161
	AB2	8776	8778	8740	8780	9224	9224	9206	9216	9191	9168	9117	9162	9761	9760	9722	9738	9656	9638	9625	9604	9205	9214	9225	9148
	Ours	8753	8774	8734	8748	9194	9197	9203	9125	9134	9129	9105	9136	9757	9646	9757	9715	9651	9505	9590	9558	9144	9119	9126	9162
RMSE ↓	BL(S)	1.438	1.438	1.438	1.438	1.604	1.604	1.604	1.604	1.607	1.607	1.607	1.607	1.800	1.804	1.812	1.812	1.736	1.736	1.736	1.736	1.759	1.772	1.773	1.760
	BL(W)	1.438	1.438	1.438	1.439	1.605	1.605	1.609	1.609	1.609	1.609	1.608	1.616	1.803	1.796	1.803	1.803	1.735	1.734	1.737	1.738	1.776	1.776	1.775	1.778
	U-BL(S)	1.438	1.438	1.438	1.444	1.604	1.604	1.604	1.611	1.607	1.607	1.607	1.607	1.797	1.797	1.813	1.797	1.736	1.736	1.736	1.736	1.759	1.759	1.773	1.759
	U-BL(W)	1.436	1.438	1.436	1.436	1.605	1.605	1.605	1.613	1.608	1.608	1.608	1.620	1.790	1.790	1.809	1.790	1.733	1.733	1.736	1.733	1.759	1.774	1.759	1.759
	AB1	1.432	1.432	1.424	1.435	1.601	1.604	1.607	1.617	1.617	1.617	1.615	1.626	1.778	1.794	1.808	1.781	1.739	1.741	1.740	1.732	1.722	1.744	1.743	1.727
	AB2	1.439	1.436	1.400	1.429	1.607	1.601	1.613	1.615	1.619	1.616	1.607	1.622	1.776	1.763	1.768	1.763	1.744	1.740	1.725	1.723	1.732	1.740	1.746	1.701
	Ours	1.426	1.427	1.400	1.421	1.603	1.594	1.609	1.586	1.606	1.602	1.599	1.614	1.754	1.709	1.764	1.777	1.726	1.678	1.723	1.711	1.704	1.687	1.691	1.718

同じことが，提案方式と AB1, AB2 の間にも当てはまる． 24 条件中，音楽データセットでは 23 条件，動画データセットでは 20 条件で，AIC, RMSE とともに提案方式が良好な結果を示した． また，表 5.5 に示すように，提案方式と比較した AB1, AB2 の相対尤度は，音楽データセットでは 23 条件すべてで，動画データセットでは 20 条件中 15 条件で 0.05 以下であった． これらの結果は，AB1 と AB2 には実装されていない提案方式の重要な特徴，すなわち，Cross-modal factor と Modality-specific factor を区別するための多層構造と，生体的特徴の個人差による調整効果をモデル化するための MSL UR が，特徴量に感情を反映することに大きく寄与したことを示している．

表 5.4 提案方式と比較したベースライン（最良のモデルと比較，太字で示す）の相対尤度。

		(a) sad-happy				(b) nervous-relaxed				(c) fear-relieved				(d) lethargic-excited				(e) depressed-delighted				(f) angry-serene			
		A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Results of Music dataset (n=2,336)																									
AIC ↓	BL(S)	11231	11235	11235	11234	11495	11501	11495	11503	11145	11145	11145	11162	11553	11553	11554	11544	12205	12205	12205	/	10416	10426	10408	10430
	BL(W)	11235	/	/	11234	/	11503	11503	11505	11150	11150	11151	/	11521	11547	11533	11534	12191	12206	12181	12201	10414	10420	/	10426
	U-BL(S)	11237	11237	11237	11237	11499	11499	11499	11503	11145	11145	11145	11160	11551	11551	11554	11551	12205	12205	12205	12211	10415	10415	10415	10428
	U-BL(W)	11234	11236	11234	11234	11503	11503	11503	11504	11151	11151	11151	11154	11543	11549	11543	11543	12198	12201	12198	12198	10416	10416	10416	10426
	Ours	11017	11059	11078	11077	11363	11398	11381	11411	11068	11060	11039	11094	11346	11343	11437	11378	11989	11961	12050	12057	10318	10376	10337	10346
RL		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Results of Movie dataset (n=2,119)																									
AIC ↓	BL(S)	8782	8782	8782	/	/	/	9227	/	9192	9193	/	/	9807	9825	9821	/	9698	9699	9699	/	9251	9282	9283	9274
	BL(W)	8781	8782	8781	8786	9228	/	9228	9229	9195	/	9194	9197	9812	9799	/	9814	9694	9693	9699	9690	9290	9290	9287	9284
	U-BL(S)	8782	8782	8782	8790	9227	9227	9227	9235	9192	9192	9192	/	9801	9801	9829	9801	9689	9689	9699	9689	9268	9268	9283	9268
	U-BL(W)	8780	8782	8780	8780	9214	9228	9214	9214	9171	9194	9171	9171	9795	9795	9823	9795	9686	9686	9695	9686	9264	9286	9264	9264
	Ours	8753	8774	8734	8748	9194	9197	9203	9125	9134	9129	9105	9136	9757	9646	9757	9715	9651	9505	9590	9558	9144	9119	9126	9162
RL		0.000	0.022	0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

表 5.5 提案方式と比較した AB1, AB2（最良のモデルと比較，太字で示す）の相対尤度。

		(a) sad-happy				(b) nervous-relaxed				(c) fear-relieved				(d) lethargic-excited				(e) depressed-delighted				(f) angry-serene			
		A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Results of Music dataset (n=2,336)																									
AIC ↓	AB1	11197	11128	11160	11206	11486	11439	11474	11486	11119	11113	11121	11142	11508	11493	11478	11499	12170	12044	12138	12182	10402	10380	10394	10429
	AB2	11130	11107	11166	11133	11431	11426	11454	11463	11097	11076	11088	11129	11421	11390	11495	11398	12086	12084	12116	12097	10350	10344	10350	10403
	Ours	11017	11059	11078	11077	11363	11398	11381	11411	11068	11060	11039	11094	11346	11343	11437	11378	11989	11961	12050	12057	10318	10376	10337	10346
	RL	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-	0.001	0.000
Results of Movie dataset (n=2,119)																									
AIC ↓	AB1	8778	8777	8763	8776	9182	9223	9214	9211	9133	9186	9135	9164	9764	9774	9821	9781	9673	9668	9695	9670	9197	9210	9225	9161
	AB2	8776	8778	8740	8780	9224	9224	9206	9216	9191	9168	9117	9162	9761	9760	9722	9738	9656	9638	9625	9604	9205	9214	9225	9148
	Ours	8753	8774	8734	8748	9194	9197	9203	9125	9134	9129	9105	9136	9757	9646	9757	9715	9651	9505	9590	9558	9144	9119	9126	9162
	RL	0.000	0.210	0.047	0.000	-	0.000	0.271	0.000	-	0.000	0.003	0.000	0.166	0.000	-	0.000	0.070	0.000	0.000	0.000	0.000	0.000	0.000	-

提案方式の結果で注目すべき点は，より多くの生体モダリティを使用しても，必ずしも感情特徴量（つまり CML の隠れ状態）が感情をより強く反映するわけではないということである．表 5.3 に示すように，3 つの生体モダリティを全て使って（つまり A），最良の結果が得られたのは 12 条件（6 つの感情次元 × 2 つのデータセット）のうち 3 条件のみであった．この結果は，既存の研究 [87, 136] と一致する．例えば，Arousal の認識では，ECG+GSR が EEG+ECG+GSR を上回ったことが報告されている．著者は，EEG が他の 2 つのモダリティほど Arousal を反映せず，認識のためのノイズとなったためだと報告している．

提案方式は既存方式と異なる方法で感情特徴量の抽出を行っているものの，上記は提案方式にも当てはまると考えられる．提案方式では CML は入力された全ての生体モダリティに影響する潜在的な共通ファクターを学習する．これにより，CML が Modality-specific factor を学習することを防がれるが，一方で，入力された生体モダリティのサブセットのみに共通し感情認識にも有用であるものの，他の生体モダリティには影響しないファクターを，CML が学習できないことも意味する．例えば，音楽データセットにおい

て、c) 怖い-落ち着く、について、C) EEG+GSR が A) EEG+ECG+GSR を上回る認識精度を示した。ECG を入力として使用したことで、EEG と GSR にのみ共通し、c) 怖い-落ち着く、の認識に有用なファクターを、CML で学習することが、妨げられたことが原因の一つとして考えられる。

以上を踏まえると、既存のアプローチと同様に、提案方式においても感情認識を実施する前に、生体モダリティの組み合わせを比較して、最適な組み合わせを特定しておく必要がある。最適な生体モダリティの組み合わせは、感情の次元や刺激の種類（音楽と動画）によって異なるため、生体モダリティの組み合わせの比較は、認識対象とする感情の次元や、刺激の種類ごとに行う必要がある。

5.6 結論

本章では、マルチモーダルな生体データから感情認識のための特徴量を抽出する多層 RNN を提案した。RNN を多層構造にすることで、提案方式は感情が複数のモダリティにまたがった生体活動に影響を与えるプロセスをモデリングする。これにより、提案方式では感情の特徴の 1 つであるものの、既存方式では見過ごされてきた Cross-modal な性質を持つ特徴量を生体データから抽出できる。EEG, ECG, GSR の測定データを用いた実験により、提案方式によって抽出された特徴量は、既存方式よりも有意に被験者の感情を反映していることが示された。

提案方式の課題として、感情と生体活動の間の一方向的な関係、すなわち感情が生体活動に影響を与える、という関係しかモデル化しない点が挙げられる。Roberts ら [118] によると、身体内部の生体活動の状態の知覚（内受容感覚）も感情に影響を与えることが知られている。こうした逆方向の関係を提案方式でモデリングすることにより、提案方式の特徴量はより感情を反映させられる可能性がある。今後の研究でこの可能性を探る必要がある。

本章で述べた研究のもう一つの課題は、被験者が静止している状態で収集された生体データのみを調査した点である。実世界のシナリオでは、生体データには体動によるノイズが含まれる。そのようなデータに対して提案方式がどのように機能するかを調査するために、さらなる研究が必要である。

第 6 章

結論

6.1 本研究の貢献

本研究では、ドメイン非依存の心理属性を用いたサービスのパーソナライズにおいて、従来研究では未解決だった 1. 推定対象とする心理属性の Ground truth の選択と、2. 行動ログ、生体ログからのドメイン非依存、Cross-modal な特徴量の抽出、の課題を解決した。また、本研究で得られた知見を活用することで、個々のユーザに対して所望のユーザ反応が得られる可能性がより高いサービスの提供が可能となること、すなわちより高いサービスパーソナライズ効果を得られることを確認した。

まず、3 章では、1. に焦点を当て、以下の疑問に答えるための検討を行った。

Q1 複数の質問紙が存在する「性格」について、測定に用いる質問紙によって、測定結果を用いたサービスパーソナライズの効果は異なるか。異なる場合、効果的なサービスパーソナライズを実現するには「性格」をどのような質問紙で測定すべきか

Q2 ともにドメイン非依存かつ長期安定の心理属性である「性格」と「価値観」は、サービスパーソナライズで利用した場合、独立した効果を持つか

これらの疑問に答えるため、心理属性を用いたサービスパーソナライズの研究が多数行われている広告デザインのパーソナライズに焦点を当て、約 8,000 名の被験者を対象とした広告デザインに対する選好度の質問紙調査を行い、以下を明らかにした。

A1 質問数、質問方法の異なる 3 つの Big Five 質問紙について調査を行い、異なる質問紙による測定結果は、個人の広告訴求力の知覚について、有意に異なる予測力を持つことを示した。また、視覚的、言語的なデザインを問わず、同一の質問紙が広告訴求力知覚に最も高い予測力を持つことを確認した。さらに、こうした質問紙が備える条件について検討を行い、各特性について十分な数の質問項目を備えているこ

と、また質問を形容詞単語で表現すること、の 2 条件が要件であるとの仮説を提起した。

A2 個人が広告デザインから知覚する広告訴求力には、視覚的・言語的なデザイン双方とも Big Five よりもシュワルツの価値観の方が大きな影響を与えることを示した。加えて、Big Five とシュワルツの価値観は個人の広告訴求力の知覚に、互いに独立した影響を与える成分を持つことを示した。広告デザインのパーソナライズを行う際には、Big Five とシュワルツの価値観の双方を用いてクリックやコンバージョンの予測を行うことで、単独で使用するよりも高いパーソナライズの効果を得られることを明らかにした。

3 章ではさらに、上記の知見を取り入れた広告予測モデル（提案モデル）の評価実験を行い、提案モデルが従来研究で検討されてきたベースラインよりも、広告反応の予測精度が高く、現在一般的に利用されている広告配信アプローチに比べ、クリック率やコンバージョン率を改善できる可能性があることを示した。

4 章では、2. のうち特に行動ログに焦点を当て、衣食住にまたがる様々な生活ドメインの商品・サービスを扱う EC ポータルにおいて、購買行動ログからドメイン非依存の心理属性を反映した特徴量を抽出する方式を提案し、そのサービスパーソナライズの効果を検証した。

具体的には、ドメイン非依存の心理属性が様々な生活ドメインにおける購買行動に影響を与えることに着目し、1) ドメイン毎の購買行動をモデリングする複数のレイヤ (Domain specific layer, DSL) と、2) 性格・価値観などのドメイン非依存なユーザ属性をモデリングする単一レイヤ (Domain independent layer, DIL) からなる多層 LSTM を導入し、性格・価値観などが各ドメインの購買行動に影響を与える過程をモデリングした。これにより、従来方式では分離できなかったドメイン特有/ドメイン非依存のユーザ属性を反映した特徴量を、分離して抽出することを可能にした。

Amazon 等の EC ポータルで収集した実際の購買行動のログを用いて実験を行い、提案方式で抽出した特徴量が従来方式により抽出した特徴量よりもドメイン非依存性が高いこと、すなわち特徴量の抽出元ドメインとは異なるドメインで、ユーザの購買行動をより精度よく予測できることを示した。さらに、従来方式と比べて、提案方式の特徴量は Big Five とより関連性が強いことを示した。

5 章では、2. のうち生体ログについて検討を行い、4 章の提案方式を元に、マルチモーダルな生体信号から感情認識のための特徴量を抽出する多層 LSTM を提案した。LSTM を多層構造にすることで、感情が複数のモダリティにまたがった生体活動に影響を与えるプロセスをモデリングした。これにより、感情の特徴の 1 つであるものの、従来方式では見過ごされてきた Cross-modal な性質、すなわち複数のモダリティにまたがった生体活

動に影響を与える性質を持つ特徴量を生体データから抽出することを可能にした。脳波 (EEG)、心電図 (ECG)、皮膚電位 (GSR) の測定データを用いた実験により、提案方式によって抽出された特徴量は、既存方式よりも有意に強く被験者の感情を反映していることを示した。

6.2 実用上の意義

本研究が対象としたドメイン非依存の心理属性は、一度推定しておけば、ユーザが利用したことのないサービスも含め、様々なサービスでパーソナライズに活用できる。

通常サービスのパーソナライズを行う場合、当該サービスで得られたログを用いてパーソナライズを行うのが一般的である。例えば、音楽配信サービスでレコメンドを行う場合、音楽配信サービスの利用履歴からユーザの音楽の好み等を把握して、これをレコメンドに用いる。この場合、当該サービスを利用したことのないユーザにはレコメンドを行うことができない。一方で、性格や感情などドメイン非依存の心理属性とユーザが好む音楽の関連性が把握できている場合（例えば Big Five の各因子を説明変数、楽曲の再生確率を目的変数とする予測モデルを構築済みの場合）、音楽配信サービスを利用したことのないユーザであっても、他のサービスの利用履歴（EC ポータルにおける商品の購買履歴や SNS の利用履歴、車の運転行動など）から当該ユーザの性格や感情が把握できれば、性格や感情に基づいてユーザの好む音楽を配信することが可能になる。

サービスの利用履歴がない（または浅い）ユーザに対するパーソナライズは、Cold start 問題と呼ばれ、解決のためのアプローチが提案されてきた [47]。基本的な解決アプローチは次のとおりである。複数のサービス（以下、A, B とする）を利用するユーザのデータを基に、A の利用履歴から B におけるサービス反応確率を、B の利用履歴から A におけるサービス反応確率を、それぞれ予測するモデルを構築しておくことで、A または B しか利用したことのないユーザに対するパーソナライズを実現する。しかし、この方法では、サービスの数が増えるにしたがって、構築すべき予測モデルの数は指数関数的に増えていく（図 6.1(1) 参照。前述の例の場合、サービス数 n では $C_2^n \times 2$ の予測モデルが必要）。また、予測モデルの精度は、複数サービスを利用するユーザのログの量に依存するため、当該ユーザの数に制約を受ける。さらに、複数サービスを同時に利用するユーザが稀な場合（サービス A, B を両方利用するユーザが A を利用するユーザの 1% 未満などの場合）、そうしたユーザの利用履歴が特殊で、それを基に構築した予測モデルがサービスを利用するユーザ全体のサービス反応をうまく予測できない可能性がある。

一方で、ドメイン非依存の心理属性を用いたパーソナライズでは、各サービスで心理属性からサービス反応確率を予測するモデルを構築するため、サービス数と同数の予測モデルの構築で Cold start 問題を解決できる（図 6.1(2) 参照）。また、ユーザの心理属性が推

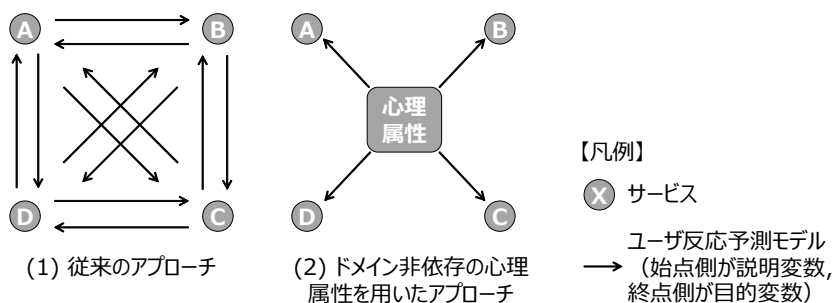


図 6.1 Cold start 問題解決に必要なユーザ反応予測モデルの比較

定/測定できていれば、当該サービスにおける全てのログを予測モデル構築に利用できる。

本研究で得られた知見を活用することで、こうしたドメイン非依存の心理属性を用いたサービスパーソナライズにおいて、高いパーソナライズ効果を得ることができる。すなわち、3章で得られた知見に基づき、ドメイン非依存の心理属性の Ground truth を選択することで、ユーザのサービス反応をより精度良く予測することができる。加えて、4章および5章で提案した方式で特徴量を抽出することでドメイン非依存の心理属性をより精度良く推定することができる。心理属性を精度良く推定することは、ドメイン非依存の心理属性を用いた Cold-start 問題の解決では特に重要である。ユーザ反応の予測モデルは、心理属性を説明変数として構築されているため、説明変数の推定精度が低い場合、予測モデルから期待した精度を得ることができない。

6.3 本研究の課題と今後の研究の方向性

本研究には今後の研究によって解決されるべき課題が複数存在する。1. 推定対象とする心理属性の Ground truth の選択と、2. 行動ログ、生体ログからのドメイン非依存、Cross-modal な特徴量の抽出に分けて課題を説明する。

まず、1. については3章でサービスのパーソナライズで使用するべき Big Five の質問紙の要件について仮説を提起したが、検証したサービスが広告デザインのパーソナライズに限られること、また、調査した質問紙の数が限られていたことから、この仮説の検証には至らなかった。広告デザインのパーソナライズについては広告対象の4つの商材すべてにおいて仮説が支持されたものの、他のサービスで同様に仮説が支持されるかは今後の検証が待たれる。また、調査した質問紙が3種類にとどまったことから、調査対象の質問紙を増やした上で仮説を検証するとともに、質問紙の他の要件を探索することが期待される。

また、2. については、多層 LSTM によって A) 複数ドメインの行動ログからドメイン非依存の特徴量を、B) マルチモーダルな生体信号から Cross-modal な特徴量を、それぞれ抽出する方式を提案し、A) については性格の、B) については感情の推定に有効である

ことを示したものの、A) の感情推定における有効性と B) の性格推定における有効性は、データセットの制約から評価できなかった。2 章で述べたとおり、感情も性格同様、様々な生活ドメインにおける行動に影響を及ぼすドメイン非依存性の心理属性であること、また性格も感情同様、様々なモダリティにおける生体活動に影響を及ぼす Cross-modal な心理属性であることから、A)、B) はともに性格、感情双方に有効であると考えられる。今後の研究でこの仮説の検証が期待される。

また、A) については 4 章で提案した方式の入力データの構築が課題として挙げられる。EC ポータルの購買履歴を入力データとして用いたが、データに適切なドメインのラベル付けがなされていない場合（すなわち 同一の DSL に無関係のドメインが混在した状態で購買行動履歴を入力した場合）、提案方式は期待通りに機能しないことが、実験により明らかとなった。人間のアノテーターがデータにラベルを付けるのはコストがかかるし、必ずしも実行可能ではない。データを自動的にラベル付けする機能は将来研究されるべきである。

入力データのドメイン数が増えるにつれて、提案方式で学習した特徴量のパフォーマンス向上の度合いが減少することも課題である。これは、複数ドメインから学習したドメイン非依存の属性と、そこに新しくドメインを追加することで学習できるドメイン非依存の属性との重複が原因であると考えられる。すでに多くのドメイン非依存属性を学習している場合、新しくドメインを追加しても、それ以上学習できることは少ない。一方で、ドメインを追加すると、DSL の追加により学習すべきパラメタ数が増加する。このように、ドメインの数が増えるにつれて、利得とコストのバランスが悪化する。この解決策の 1 つは、複数の「関連する」ドメインを 1 つの DSL にまとめることである。課題はドメイン間の関係をどのように自動的に決定するかである。これは将来的に研究されるべきである。

B) の課題としては、5 章で提案した方式が、感情と生体活動の間の関係性のうち一部しかモデリングできていない点が挙げられる。提案方式では、感情が生体活動に影響を与える、という一方向的な関係しかモデル化できていない。既存研究において、身体内部の生体活動の状態の知覚（内受容感覚）が感情に影響を与えることが知られている。こうした生体活動 → 感情の関係も提案方式でモデリングすることにより、提案方式の特徴量により強く感情を反映させられる可能性がある。今後の研究でこの可能性を探る必要がある。

他の課題は、被験者が静止している状態で収集された生体データのみを調査した点である。実世界のシナリオでは、生体データには体動によるノイズが含まれる。そのようなデータに対して提案方式がどのように機能するかを調査するために、さらなる研究が必要である。

最後に A)、B) に共通する課題として、提案方式で学習した特徴量の解釈性が挙げられる。A) については 4 章の実験により、提案方式の特徴量にはドメイン非依存のユーザ属性が反映されていること、さらに Big Five との関連性が強いことを示したが、特徴量に

は価値観や、年齢、性別、居住地といったデモグラフィックな属性など、他のドメイン非依存の属性も反映されていると考えられる。このように複数種類のユーザ属性が反映されていると、仮にこれらのうち一部の属性がユーザによって提供された場合、提案方式の特徴量とユーザ提供の属性に重複が生じ、共に利用した場合にサービスパーソナライズの効果を下げる可能性がある。他の属性を反映している可能性は、B) にも当てはまる。複数のモダリティにおける生体活動に影響を与える Cross-modal なユーザ属性としては、体調やストレス、空腹度などが考えられ、B) の特徴量にも反映されている可能性があるが、これらの属性は必ずしも様々なサービスのパーソナライズに有効とは言えない。以上を踏まえると、今後の研究では、対応する属性毎に特徴量を分離するなど提案方式の特徴量の解釈性を高めるための拡張が望まれる。

謝辞

本論文の執筆にあたり，博士課程の三年間にわたり多大なるご指導を頂きました九州大学大学院システム情報科学研究院の荒川豊教授，峯恒憲准教授，中村優吾助教，ならびに，公立はこだて未来大学の石田繁巳准教授に深く御礼申し上げます．また，本論文をまとめる過程においてご助言ご指導を頂きました九州大学大学院システム情報科学研究院の福嶋政期准教授に深く感謝申し上げます．

参考文献

- [1] Covid-19 boost to e-commerce sustained into 2021, new unctad figures show. <https://unctad.org/news/covid-19-boost-e-commerce-sustained-2021-new-unctad-figures-show>, Apr 2022.
- [2] Hirotugu Akaike. Information theory and an extension of the maximum likelihood principle. In *Selected papers of hirotugu akaike*, pp. 199–213. Springer, 1998.
- [3] Renan Vinicius Aranha, Cléber Gimenez Corrêa, and Fátima LS Nunes. Adapting software with affective computing: a systematic review. *IEEE Transactions on Affective Computing*, Vol. 12, No. 4, pp. 883–899, 2019.
- [4] Michael C Ashton, Kibeom Lee, Marco Perugini, Piotr Szarota, Reinout E De Vries, Lisa Di Blas, Kathleen Boies, and Boele De Raad. A six-factor structure of personality-descriptive adjectives: solutions from psycholexical studies in seven languages. *Journal of personality and social psychology*, Vol. 86, No. 2, p. 356, 2004.
- [5] Danny Azucar, Davide Marengo, and Michele Settanni. Predicting the big 5 personality traits from digital footprints on social media: A meta-analysis. *Personality and individual differences*, Vol. 124, pp. 150–159, 2018.
- [6] Murray R Barrick and Michael K Mount. The big five personality dimensions and job performance: a meta-analysis. *Personnel psychology*, Vol. 44, No. 1, pp. 1–26, 1991.
- [7] Verónica Benet-Martínez and Oliver P John. Los cinco grandes across cultures and ethnic groups: Multitrait-multimethod analyses of the big five in spanish and english. *Journal of personality and social psychology*, Vol. 75, No. 3, p. 729, 1998.
- [8] Erdal Binboga, Senol Guven, Fatih Çatıkkaş, Onur Bayazıt, and Serdar Tok.

- Psychophysiological responses to competition and the big five personality traits. *Journal of human kinetics*, Vol. 33, p. 187, 2012.
- [9] Tim Bogg and Brent W Roberts. Conscientiousness and health-related behaviors: a meta-analysis of the leading behavioral contributors to mortality. *Psychological bulletin*, Vol. 130, No. 6, p. 887, 2004.
- [10] Ciro Bologna, Anna Chiara De Rosa, Alfonso De Vivo, Matteo Gaeta, Giuseppe Sansonetti, and Valeria Viserta. Personality-based recommendation in e-commerce. In Shlomo Berkovsky, Eelco Herder, Pasquale Lops, and Olga C. Santos, editors, *Late-Breaking Results, Project Papers and Workshop Proceedings of the 21st Conference on User Modeling, Adaptation, and Personalization.*, Rome, Italy, June 10-14, 2013, Vol. 997 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2013.
- [11] Turhan Canli, Heidi Sivers, Susan L Whitfield, Ian H Gotlib, and John DE Gabrieli. Amygdala response to happy faces as a function of extraversion. *Science*, Vol. 296, No. 5576, pp. 2191–2191, 2002.
- [12] Turhan Canli, Zuo Zhao, John E Desmond, Eunjoo Kang, James Gross, and John DE Gabrieli. An fmri study of personality influences on brain reactivity to emotional stimuli. *Behavioral neuroscience*, Vol. 115, No. 1, p. 33, 2001.
- [13] Tomas Chamorro-Premuzic and Adrian Furnham. Personality predicts academic performance: Evidence from two longitudinal university samples. *Journal of research in personality*, Vol. 37, No. 4, pp. 319–338, 2003.
- [14] Li Chen, Wen Wu, and Liang He. How personality influences users’ needs for recommendation diversity? In *CHI’13 extended abstracts on human factors in computing systems*, pp. 829–834. 2013.
- [15] Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishi Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ispir, et al. Wide & deep learning for recommender systems. In *Proceedings of the 1st workshop on deep learning for recommender systems*, pp. 7–10, 2016.
- [16] Gokul Chittaranjan, Jan Blom, and Daniel Gatica-Perez. Mining large-scale smartphone data for personality studies. *Personal and Ubiquitous Computing*, Vol. 17, No. 3, pp. 433–450, 2013.
- [17] Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*, 2014.
- [18] RB Cialdini. *Influence: The psychology of persuasion*: Harpercollins. 2006.

-
- [19] Ana Ciocarlan, Judith Masthoff, and Nir Oren. Actual persuasiveness: impact of personality, age and gender on message type susceptibility. In *Persuasive Technology: Development of Persuasive and Behavior Change Support Systems: 14th International Conference, PERSUASIVE 2019, Limassol, Cyprus, April 9–11, 2019, Proceedings 14*, pp. 283–294. Springer, 2019.
 - [20] Michael X Cohen, Jennifer Young, Jong-Min Baek, Christopher Kessler, and Charan Ranganath. Individual differences in extraversion and dopamine genetics predict neural reward responses. *Cognitive brain research*, Vol. 25, No. 3, pp. 851–861, 2005.
 - [21] Paul T Costa and Robert R McCrae. The revised neo personality inventory (neo-pi-r). *The SAGE handbook of personality theory and assessment*, Vol. 2, No. 2, pp. 179–198, 2008.
 - [22] Paul Costa and R. McCrae. Neo pi-r professional manual. *Psychological Assessment Resources*, Vol. 396, , 01 1992.
 - [23] Kerry S Courneya and Laurie-Ann M Hellsten. Personality correlates of exercise behavior, motives, barriers and preferences: An application of the five-factor model. *Personality and Individual differences*, Vol. 24, No. 5, pp. 625–633, 1998.
 - [24] Yves-Alexandre de Montjoye, Jordi Quoidbach, Florent Robic, and Alex Pentland. Predicting personality using novel mobile phone-based metrics. In *Social Computing, Behavioral-Cultural Modeling and Prediction: 6th International Conference, SBP 2013, Washington, DC, USA, April 2-5, 2013. Proceedings 6*, pp. 48–55. Springer, 2013.
 - [25] Reinout E de Vries. No evidence for a general factor of personality in the hexaco personality inventory. *Journal of Research in Personality*, Vol. 45, No. 2, pp. 229–232, 2011.
 - [26] Thilo Deckersbach, Karen K Miller, Anne Klibanski, Alan Fischman, Darin D Dougherty, Mark A Blais, David B Herzog, and Scott L Rauch. Regional cerebral brain metabolism correlates of neuroticism and extraversion. *Depression and anxiety*, Vol. 23, No. 3, pp. 133–138, 2006.
 - [27] James J Deng, Clement HC Leung, Alfredo Milani, and Li Chen. Emotional states associated with music: Classification, prediction of changes, and consideration in recommendation. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, Vol. 5, No. 1, pp. 1–36, 2015.
 - [28] Colin G. DeYoung and Jeremy R. Gray. *Personality neuroscience: explaining*

- individual differences in affect, behaviour and cognition*, pp. 323–346. Cambridge Handbooks in Psychology. Cambridge University Press, 2009.
- [29] Colin G DeYoung, Jordan B Peterson, and Daniel M Higgins. Higher-order factors of the big five predict conformity: Are there neuroses of health? *Personality and Individual differences*, Vol. 33, No. 4, pp. 533–552, 2002.
- [30] Colin G DeYoung, Lena C Quilty, and Jordan B Peterson. Between facets and domains: 10 aspects of the big five. *Journal of personality and social psychology*, Vol. 93, No. 5, p. 880, 2007.
- [31] Sahraoui Dhelim, Nyothiri Aung, Mohammed Amine Bouras, Huansheng Ning, and Erik Cambria. A survey on personality-aware recommendation systems. *Artificial Intelligence Review*, pp. 1–46, 2022.
- [32] Sahraoui Dhelim, Huansheng Ning, Nyothiri Aung, Runhe Huang, and Jianhua Ma. Personality-aware product recommendation system based on user interests mining and metapath discovery. *IEEE Transactions on Computational Social Systems*, Vol. 8, No. 1, pp. 86–98, 2021.
- [33] Tao Ding and Shimei Pan. Personalized emphasis framing for persuasive message generation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 1432–1441, 2016.
- [34] Tim Donkers, Benedikt Loepp, and Jürgen Ziegler. Sequential user-based recurrent neural network recommendations. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*, RecSys '17, pp. 152–160, New York, NY, USA, 2017. Association for Computing Machinery.
- [35] Paul Ekman. Facial expressions of emotion: New findings, new questions, 1992.
- [36] Bruce Ferwerda, Markus Schedl, and Marko Tkalcić. Using instagram picture features to predict users' personality. In *MultiMedia Modeling: 22nd International Conference, MMM 2016, Miami, FL, USA, January 4-6, 2016, Proceedings, Part I 22*, pp. 850–861. Springer, 2016.
- [37] Bruce Ferwerda and Marko Tkalcić. Predicting users' personality from instagram pictures: using visual and/or content features? In *Proceedings of the 26th conference on user modeling, adaptation and personalization*, pp. 157–161, 2018.
- [38] FFPQ 研究会. FFPQ(5 因子性格検査). 北大路書房, 1998.
- [39] Elena Fraj and Eva Martinez. Environmental values and lifestyles as determining factors of ecological consumer behaviour: an empirical analysis. *Journal of consumer marketing*, Vol. 23, No. 3, pp. 133–144, 2006.

-
- [40] Elena Fraj and Eva Martinez. Ecological consumer behaviour: an empirical analysis. *International journal of consumer studies*, Vol. 31, No. 1, pp. 26–33, 2007.
 - [41] Howard S Friedman, Joan S Tucker, Carol Tomlinson-Keasey, Joseph E Schwartz, Deborah L Wingard, and Michael H Criqui. Does childhood personality predict longevity? *Journal of personality and social psychology*, Vol. 65, No. 1, p. 176, 1993.
 - [42] Adrian Furnham and Sumedh Rao. Personality and the aesthetics of composition: A study of mondrian and hirst. *North American Journal of Psychology*, Vol. 4, No. 2, pp. 233–242, 2002.
 - [43] Jennifer Golbeck and Eric Norris. Personality, movie preferences, and recommendations. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 1414–1415, 2013.
 - [44] Jennifer Golbeck and Eric Norris. Personality, movie preferences, and recommendations. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, ASONAM '13, pp. 1414–1415, New York, NY, USA, 2013. Association for Computing Machinery.
 - [45] Jennifer Golbeck, Cristina Robles, Michon Edmondson, and Karen Turner. Predicting personality from twitter. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, pp. 149–156, 2011.
 - [46] Lewis R Goldberg. An alternative” description of personality”: the big-five factor structure. *Journal of personality and social psychology*, Vol. 59, No. 6, p. 1216, 1990.
 - [47] Jyotirmoy Gope and Sanjay Kumar Jain. A survey on solving cold start problem in recommender systems. In *2017 International Conference on Computing, Communication and Automation (ICCCA)*, pp. 133–138. IEEE, 2017.
 - [48] Samuel D Gosling, Peter J Rentfrow, and William B Swann Jr. A very brief measure of the big-five personality domains. *Journal of Research in personality*, Vol. 37, No. 6, pp. 504–528, 2003.
 - [49] Jie Gu, Feng Wang, Qinghui Sun, Zhiquan Ye, Xiaoxiao Xu, Jingmin Chen, and Jun Zhang. Exploiting behavioral consistence for universal user representation. *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35, No. 5, pp. 4063–4071, May 2021.
 - [50] Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, and Xiuqiang He.

- Deepfm: a factorization-machine based neural network for ctr prediction. *arXiv preprint arXiv:1703.04247*, 2017.
- [51] Terry Halfhill, Eric Sundstrom, Jessica Lahner, Wilma Calderone, and Tjai M Nielsen. Group personality composition and group effectiveness: An integrative review of empirical research. *Small group research*, Vol. 36, No. 1, pp. 83–105, 2005.
- [52] Frank Harvey and Richard Hirschmann. The influence of extraversion and neuroticism on heart rate responses to aversive visual stimuli. *Personality and Individual Differences*, Vol. 1, No. 1, pp. 97–100, 1980.
- [53] Jonathan Herzig, Guy Feigenblat, Michal Shmueli-Scheuer, David Konopnicki, and Anat Rafaeli. Predicting customer satisfaction in customer support conversations in social media using affective features. In *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization*, pp. 115–119, 2016.
- [54] Jacob B Hirsh, Sonia K Kang, and Galen V Bodenhausen. Personalized persuasion: Tailoring persuasive appeals to recipients’ personality traits. *Psychological science*, Vol. 23, No. 6, pp. 578–581, 2012.
- [55] Jacob B. Hirsh and Jordan B. Peterson. Personality and language use in self-narratives. *Journal of Research in Personality*, Vol. 43, No. 3, pp. 524–527, 2009.
- [56] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, Vol. 9, No. 8, pp. 1735–1780, 1997.
- [57] Christopher J Hopwood, Aidan GC Wright, and M Brent Donnellan. Evaluating the evidence for the general factor of personality across multiple inventories. *Journal of Research in Personality*, Vol. 45, No. 5, pp. 468–478, 2011.
- [58] Rong Hu and Pearl Pu. Enhancing collaborative filtering systems with personality information. In *Proceedings of the Fifth ACM Conference on Recommender Systems*, RecSys ’11, pp. 197–204, New York, NY, USA, 2011. Association for Computing Machinery.
- [59] Yuichi Ishikawa, Akihiro Kobayashi, Atsunori Minamikawa, and Chihiro Ono. Predicting a driver’s personality from daily driving behavior. In *Driving Assessment Conference*, Vol. 10. University of Iowa, 2019.
- [60] Donahue E. M. John, O. P. and R. L. Kentle. The big five inventory—versions 4a and 54. 1991.
- [61] Oliver P John, Laura P Naumann, and Christopher J Soto. Paradigm shift to

-
- the integrative big five trait taxonomy. *Handbook of personality: Theory and research*, Vol. 3, No. 2, pp. 114–158, 2008.
- [62] Christopher C Johnson. Logistic matrix factorization for implicit feedback data. *Advances in Neural Information Processing Systems*, Vol. 27, No. 78, pp. 1–9, 2014.
 - [63] Donghyun Kim, Sungchul Kim, Handong Zhao, Sheng Li, Ryan A. Rossi, and Eunye Koh. Domain switch-aware holistic recurrent neural network for modeling multi-domain user behavior. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, WSDM '19, pp. 663–671, New York, NY, USA, 2019. Association for Computing Machinery.
 - [64] Jonghwa Kim and Elisabeth André. Emotion recognition based on physiological changes in music listening. *IEEE transactions on pattern analysis and machine intelligence*, Vol. 30, No. 12, pp. 2067–2083, 2008.
 - [65] Akihiro Kobayashi, Yuichi Ishikawa, Kazushi Ikeda, Daisuke Kamisaka, and Roberto Legaspi. Composing groups in collaborative learning by pair personality differences. In *Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization*, pp. 116–123, 2023.
 - [66] Akihiro Kobayashi, Yuichi Ishikawa, and Atsunori Minamikawa. A study on effect of big five personality traits on ad targeting and creative design. In *International Conference on Persuasive Technology*, pp. 257–269. Springer, 2019.
 - [67] Akihiro Kobayashi, Yuichi Ishikawa, and Roberto Sebastian Legaspi. Psychographic matching between a call center agent and a customer. In *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*, pp. 229–234, 2021.
 - [68] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. Deap: A database for emotion analysis; using physiological signals. *IEEE Transactions on Affective Computing*, Vol. 3, No. 1, pp. 18–31, 2011.
 - [69] Meera Komarraju and Steven J Karau. The relationship between the big five personality traits and academic motivation. *Personality and individual differences*, Vol. 39, No. 3, pp. 557–567, 2005.
 - [70] Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, Vol. 42, No. 8, pp. 30–37, 2009.
 - [71] Marco Lauriola and Irwin P Levin. Personality traits and risky decision-making in a controlled experimental task: An exploratory study. *Personality and indi-*

- vidual differences*, Vol. 31, No. 2, pp. 215–226, 2001.
- [72] Chao Li, Zhongtian Bao, Linhao Li, and Ziping Zhao. Exploring temporal representations by leveraging attention-based bidirectional lstm-rnns for multi-modal emotion recognition. *Information Processing & Management*, Vol. 57, No. 3, p. 102185, 2020.
 - [73] Seth Siyuan Li and Elena Karahanna. Online recommendation systems in a b2c e-commerce context: a review and future directions. *Journal of the association for information systems*, Vol. 16, No. 2, p. 2, 2015.
 - [74] Sheng Li and Handong Zhao. A survey on representation learning for user modeling. In Christian Bessiere, editor, *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, pp. 4997–5003. International Joint Conferences on Artificial Intelligence Organization, 7 2020. Survey track.
 - [75] Jianxun Lian, Xiaohuan Zhou, Fuzheng Zhang, Zhongxia Chen, Xing Xie, and Guangzhong Sun. Xdeepfm: Combining explicit and implicit feature interactions for recommender systems. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '18*, pp. 1754–1763, New York, NY, USA, 2018. Association for Computing Machinery.
 - [76] Beng-Chong Lim and Robert E Ployhart. Transformational leadership: relations to the five-factor model and team performance in typical and maximum contexts. *Journal of applied psychology*, Vol. 89, No. 4, p. 610, 2004.
 - [77] Wei Liu, Wei-Long Zheng, and Bao-Liang Lu. Emotion recognition using multimodal deep learning. In *ICONIP 2016, Proceedings, Part II 23*, pp. 521–529. Springer, 2016.
 - [78] Zhaoyang Liu, Haokun Chen, Fei Sun, Xu Xie, Jinyang Gao, Bolin Ding, and Yanyan Shen. Intent preference decoupling for user representation on online recommender system. In Christian Bessiere, editor, *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, pp. 2575–2582. International Joint Conferences on Artificial Intelligence Organization, 7 2020. Main track.
 - [79] Wei Lü, Wanying Xing, Brian M Hughes, and Zhenhong Wang. Extraversion and cardiovascular responses to recurrent social stress: Effect of stress intensity. *International Journal of Psychophysiology*, Vol. 131, pp. 144–151, 2018.
 - [80] Yukun Ma, Khanh Linh Nguyen, Frank Z Xing, and Erik Cambria. A survey on empathetic dialogue systems. *Information Fusion*, Vol. 64, pp. 50–70, 2020.

-
- [81] Jaakko Malmivuo and Robert Plonsey. *Bioelectromagnetism. 13. Electroencephalography*, pp. 247–264. 01 1995.
- [82] Leslie R Martin and Howard S Friedman. Comparing personality scales across time: An illustrative study of validity and consistency in life-span archival data. *Journal of personality*, Vol. 68, No. 1, pp. 85–110, 2000.
- [83] Sandra C Matz, Michal Kosinski, Gideon Nave, and David J Stillwell. Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the national academy of sciences*, Vol. 114, No. 48, pp. 12714–12719, 2017.
- [84] Sandra C Matz, Cristina Segalin, David Stillwell, Sandrine R Müller, and Maarten W Bos. Predicting the personal appeal of marketing images using computational methods. *Journal of Consumer Psychology*, Vol. 29, No. 3, pp. 370–390, 2019.
- [85] Albert Mehrabian and James A Russell. *An approach to environmental psychology*. the MIT Press, 1974.
- [86] Yash Mehta, Navonil Majumder, Alexander Gelbukh, and Erik Cambria. Recent trends in deep learning based personality detection. *Artificial Intelligence Review*, Vol. 53, No. 4, pp. 2313–2339, 2020.
- [87] Juan Abdon Miranda-Correa, Mojtaba Khomami Abadi, Nicu Sebe, and Ioannis Patras. Amigos: A dataset for affect, personality and mood research on individuals and groups. *IEEE Transactions on Affective Computing*, Vol. 12, No. 2, pp. 479–493, 2018.
- [88] Dean Mobbs, Cindy C Hagan, Eiman Azim, Vinod Menon, and Allan L Reiss. Personality predicts activity in reward and emotional regions associated with humor. *Proceedings of the National Academy of Sciences*, Vol. 102, No. 45, p. 1685, 2005.
- [89] Ross Moen and Kenneth O Doyle Jr. Construction and development of the academic motivations inventory (ami. *Educational and Psychological Measurement*, Vol. 37, No. 2, pp. 509–512, 1977.
- [90] R Matthew Montoya, Robert S Horton, and Jeffrey Kirchner. Is actual similarity necessary for attraction? a meta-analysis of actual and perceived similarity. *Journal of Social and Personal Relationships*, Vol. 25, No. 6, pp. 889–922, 2008.
- [91] Kelly Moore and James C McElroy. The influence of personality on facebook usage, wall postings, and regret. *Computers in human behavior*, Vol. 28, No. 1, pp. 267–274, 2012.
- [92] Jon D Morris. Observations: Sam: the self-assessment manikin; an efficient

- cross-cultural measurement of emotional response. *Journal of advertising research*, Vol. 35, No. 6, pp. 63–68, 1995.
- [93] Riza Casidy Mulyanegara, Yelena Tsarenko, and Alastair Anderson. The big five and brand personality: Investigating the impact of consumer personality on preferences towards particular brand personality. *Journal of brand management*, Vol. 16, No. 4, pp. 234–247, 2009.
- [94] Steven J Muncer. The general factor of personality: Evaluating the evidence from meta-analysis, confirmatory factor analysis and evolutionary theory. *Personality and Individual Differences*, Vol. 51, No. 6, pp. 775–778, 2011.
- [95] Janek Musek. A general factor of personality: Evidence for the big one in the five-factor model. *Journal of research in personality*, Vol. 41, No. 6, pp. 1213–1233, 2007.
- [96] Janek Musek. *The general factor of personality*. Academic Press, 2017.
- [97] Susan D Myers, Sandipan Sen, and Aliosha Alexandrov. The moderating effect of personality traits on attitudes toward advertisements: A contingency framework. *Management & Marketing*, Vol. 5, No. 3, 2010.
- [98] Nils Myszkowski and Martin Storme. How personality traits predict design-driven consumer choices. *Europe’s Journal of Psychology*, Vol. 8, No. 4, pp. 641–650, 2012.
- [99] Daniel Nettle. *Personality: What makes you the way you are*. OUP Oxford, 2007.
- [100] Jianmo Ni, Jiacheng Li, and Julian McAuley. Justifying recommendations using distantly-labeled reviews and fine-grained aspects. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 188–197, 2019.
- [101] Yabo Ni, Dan Ou, Shichen Liu, Xiang Li, Wenwu Ou, Anxiang Zeng, and Luo Si. Perceive your users in depth: Learning universal user representations from multiple e-commerce tasks. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD ’18*, pp. 596–605, New York, NY, USA, 2018. Association for Computing Machinery.
- [102] Maria Augusta S.N. Nunes and Rong Hu. Personality-based recommender systems: An overview. In *Proceedings of the Sixth ACM Conference on Recommender Systems, RecSys ’12*, pp. 5–6, New York, NY, USA, 2012. Association for Computing Machinery.

-
- [103] Adam O’Riordan, Danielle A Young, Alexandra T Tyra, and Annie T Ginty. Extraversion is associated with lower cardiovascular reactivity to acute psychological stress. *International Journal of Psychophysiology*, Vol. 189, pp. 20–29, 2023.
 - [104] Atsushi Oshio, ABE Shingo, and Pino Cutrone. Development, reliability, and validity of the japanese version of ten item personality inventory (tipi-j). *Japanese Journal of Personality*, Vol. 21, No. 1, 2012.
 - [105] Kiemute Oyibo, Rita Orji, and Julita Vassileva. Investigation of the Influence of Personality Traits on Cialdini’s Persuasive Strategies. *PPT@PERSUASIVE*, Vol. 2017, pp. 8–20, 2017.
 - [106] Delroy L Paulhus and Kevin M Williams. The dark triad of personality: Narcissism, machiavellianism, and psychopathy. *Journal of research in personality*, Vol. 36, No. 6, pp. 556–563, 2002.
 - [107] Miranda AG Peeters, Harrie FJM Van Tuijl, Christel G Rutte, and Isabelle MMJ Reymen. Personality and team performance: a meta-analysis. *European journal of personality*, Vol. 20, No. 5, pp. 377–396, 2006.
 - [108] Wei Peng, Yue Hu, Luxi Xing, Yuqiang Xie, and Yajing Sun. Do you know my emotion? emotion-aware strategy recognition towards a persuasive dialogue system. In *ECML PKDD 2022, Proceedings, Part II*, pp. 724–739. Springer, 2023.
 - [109] James Pennebaker, Ryan Boyd, Kayla Jordan, and Kate Blackburn. The development and psychometric properties of liwc2015, 09 2015.
 - [110] Stephen F Pirog and James A Roberts. Personality and credit card misuse among college students: The mediating role of impulsiveness. *Journal of Marketing Theory and Practice*, Vol. 15, No. 1, pp. 65–77, 2007.
 - [111] Lin Qiu, Han Lin, Jonathan Ramsay, and Fang Yang. You are what you tweet: Personality expression and perception on twitter. *Journal of research in personality*, Vol. 46, No. 6, pp. 710–718, 2012.
 - [112] Zhaopeng Qiu, Xian Wu, Jingyue Gao, and Wei Fan. U-bert: Pre-training user representations for improved recommendation. *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35, No. 5, pp. 4320–4327, May 2021.
 - [113] Yanru Qu, Han Cai, Kan Ren, Weinan Zhang, Yong Yu, Ying Wen, and Jun Wang. Product-based neural networks for user response prediction. In *2016 IEEE 16th International Conference on Data Mining (ICDM)*, pp. 1149–1154. IEEE, 2016.

- [114] Massimo Quadrana, Alexandros Karatzoglou, Balázs Hidasi, and Paolo Cremonesi. Personalizing session-based recommendations with hierarchical recurrent neural networks. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*, RecSys '17, pp. 130–137, New York, NY, USA, 2017. Association for Computing Machinery.
- [115] Steffen Rendle. Factorization machines with libfm. *ACM Trans. Intell. Syst. Technol.*, Vol. 3, No. 3, May 2012.
- [116] Retailrocket. Retailrocket recommender system dataset — kaggle, 09 2017.
- [117] Brent W Roberts, Nathan R Kuncel, Rebecca Shiner, Avshalom Caspi, and Lewis R Goldberg. The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological science*, Vol. 2, No. 4, pp. 313–345, 2007.
- [118] Tomi-Ann Roberts and James W Pennebaker. Gender differences in perceiving internal state: Toward a his-and-hers model of perceptual cue use. In *Advances in experimental social psychology*, Vol. 27, pp. 143–175. Elsevier, 1995.
- [119] Milton Rokeach. *The nature of human values*. Free press, 1973.
- [120] James A Russell. A circumplex model of affect. *Journal of personality and social psychology*, Vol. 39, No. 6, p. 1161, 1980.
- [121] Gerard Saucier. Mini-markers: A brief version of goldberg’s unipolar big-five markers. *Journal of personality assessment*, Vol. 63, No. 3, pp. 506–516, 1994.
- [122] Shalom H Schwartz. Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In *Advances in experimental social psychology*, Vol. 25, pp. 1–65. Elsevier, 1992.
- [123] Shalom H Schwartz. A proposal for measuring value orientations across nations. *Questionnaire package of the european social survey*, Vol. 259, No. 290, p. 261, 2003.
- [124] Shalom H Schwartz. An overview of the schwartz theory of basic values. *Online readings in Psychology and Culture*, Vol. 2, No. 1, p. 11, 2012.
- [125] Gideon Schwarz. Estimating the dimension of a model. *The annals of statistics*, pp. 461–464, 1978.
- [126] Crisitina Segalin, Alessandro Perina, Marco Cristani, and Alessandro Vinciarelli. The pictures we like are our image: continuous mapping of favorite pictures into self-assessed and attributed personality traits. *IEEE Transactions on Affective Computing*, Vol. 8, No. 2, pp. 268–285, 2016.

-
- [127] Fred Shaffer and Jay P Ginsberg. An overview of heart rate variability metrics and norms. *Frontiers in public health*, p. 258, 2017.
 - [128] R Shepherd and P Sparks. Modelling food choice. In *Measurement of food preferences*, pp. 202–226. Springer, 1994.
 - [129] Avgusta Y Shestyuk, Karthik Kasinathan, Viswajith Karapoondinott, Robert T Knight, and Ram Gurumoorthy. Individual eeg measures of attention, memory, and motivation predict population level tv viewership and twitter engagement. *PloS one*, Vol. 14, No. 3, p. e0214507, 2019.
 - [130] Kyuyong Shin, Hanock Kwak, Kyung-Min Kim, Minkyu Kim, Young-Jin Park, Jisu Jeong, and Seungjae Jung. One4all user representation for recommender systems in e-commerce. *arXiv preprint arXiv:2106.00573*, 2021.
 - [131] Lin Shu, Jinyan Xie, Mingyue Yang, Ziyi Li, Zhenqi Li, Dan Liao, Xiangmin Xu, and Xinyi Yang. A review of emotion recognition using physiological signals. *Sensors*, Vol. 18, No. 7, p. 2074, 2018.
 - [132] Gkika Sofia, Skiada Marianna, Lekakos George, and Kourouthanasis Panos. Investigating the role of personality traits and influence strategies on the persuasive effect of personalized recommendations. In *4th Workshop on emotions and personality in personalized systems (EMPIRE)*, Vol. 9, 2016.
 - [133] Christopher J Soto and Oliver P John. The next big five inventory (bfi-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of personality and social psychology*, Vol. 113, No. 1, p. 117, 2017.
 - [134] Clemens Stachl, Quay Au, Ramona Schoedel, Daniel Buschek, Sarah Völkel, Tobias Schuwerk, Michelle Oldemeier, Theresa Ullmann, Heinrich Hussmann, Bernd Bischl, et al. Behavioral patterns in smartphone usage predict big five personality traits. 2019.
 - [135] Christopher Steiner and Walter Dixon. *Automate This: How Algorithms Came to Rule Our World (Your Coach in a*. 2012.
 - [136] Ramanathan Subramanian, Julia Wache, Mojtaba Khomami Abadi, Radu L Vieriu, Stefan Winkler, and Nicu Sebe. Ascertain: Emotion and personality recognition using commercial sensors. *IEEE Transactions on Affective Computing*, Vol. 9, No. 2, pp. 147–160, 2016.
 - [137] Chris Sumner, Alison Byers, Rachel Boochever, and Gregory J. Park. Predicting dark triad personality traits from twitter usage and a linguistic analysis of tweets. In *2012 11th International Conference on Machine Learning and*

- Applications*, Vol. 2, pp. 386–393, 2012.
- [138] Qinghui Sun, Jie Gu, Bei Yang, XiaoXiao Xu, Renjun Xu, Shangde Gao, Hong Liu, and Huan Xu. Interest-oriented universal user representation via contrastive learning. *arXiv preprint arXiv:2109.08865*, 2021.
- [139] Viren Swami and Adrian Furnham. The effects of symmetry and personality on aesthetic preferences. *Imagination, Cognition and Personality*, Vol. 32, No. 1, pp. 41–57, 2012.
- [140] Zhiqiang Tao, Sheng Li, Zhaowen Wang, Chen Fang, Longqi Yang, Handong Zhao, and Yun Fu. Log2intent: Towards interpretable user modeling via recurrent semantics memory unit. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '19, pp. 1055–1063, New York, NY, USA, 2019. Association for Computing Machinery.
- [141] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 58, No. 1, pp. 267–288, 1996.
- [142] Marko Tkalcić and Li Chen. Personality and recommender systems. In *Recommender systems handbook*, pp. 715–739. Springer, 2015.
- [143] Serdar Tok. The big five personality traits and risky sport participation. *Social Behavior and Personality: an international journal*, Vol. 39, No. 8, pp. 1105–1111, 2011.
- [144] Alessandro Vinciarelli and Gelareh Mohammadi. A survey of personality computing. *IEEE Transactions on Affective Computing*, Vol. 5, No. 3, pp. 273–291, 2014.
- [145] Alessandro Vinciarelli and Gelareh Mohammadi. A survey of personality computing. *IEEE Transactions on Affective Computing*, Vol. 5, No. 3, pp. 273–291, 2014.
- [146] Scott I Vrieze. Model selection and psychological theory: a discussion of the differences between the akaike information criterion (aic) and the bayesian information criterion (bic). *Psychological methods*, Vol. 17, No. 2, p. 228, 2012.
- [147] Ruoxi Wang, Bin Fu, Gang Fu, and Mingliang Wang. Deep & cross network for ad click predictions. In *Proceedings of the ADKDD'17*, pp. 1–7. 2017.
- [148] David Watson, Lee Anna Clark, and Auke Tellegen. Development and validation of brief measures of positive and negative affect: the panas scales. *Journal of personality and social psychology*, Vol. 54, No. 6, p. 1063, 1988.
- [149] James B Weaver III. Exploring the links between personality and media pref-

- erences. *Personality and individual differences*, Vol. 12, No. 12, pp. 1293–1299, 1991.
- [150] Ralph Westfall. Psychological factors in predicting product choice. *Journal of Marketing*, Vol. 26, No. 2, pp. 34–40, 1962.
- [151] Jason Weston, Samy Bengio, and Nicolas Usunier. Wsabie: Scaling up to large vocabulary image annotation. In *Twenty-Second International Joint Conference on Artificial Intelligence*, 2011.
- [152] Chao Yang, Shimei Pan, Jalal Mahmud, Huahai Yang, and Padmini Srinivasan. Using personal traits for brand preference prediction. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 86–96, 2015.
- [153] Zhong Yin, Mengyuan Zhao, Yongxiong Wang, Jingdong Yang, and Jianhua Zhang. Recognition of emotions using multimodal physiological signals and an ensemble deep learning model. *Computer methods and programs in biomedicine*, Vol. 140, pp. 93–110, 2017.
- [154] Yusuke Yokota and Yasushi Naruse. Temporal fluctuation of mood in gaming task modulates feedback negativity: Eeg study with virtual reality. *Frontiers in human neuroscience*, Vol. 15, p. 246, 2021.
- [155] Junqi Zhang, Bing Bai, Ye Lin, Jian Liang, Kun Bai, and Fei Wang. General-purpose user embeddings based on mobile app usage. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '20, pp. 2831–2840, New York, NY, USA, 2020. Association for Computing Machinery.
- [156] 下仲順子, 中里克治, 榎藤恭之, 高山緑. 日本版 neo-pi-r の作成とその因子的妥当性の検討. 性格心理学研究, Vol. 6, No. 2, pp. 138–147, 1998.
- [157] 広岡博之, 大石風人, 長命洋佑, 園田裕太. 人間の価値観と消費者の動物福祉や地球環境に配慮した牛肉に対する購買行動との関連性: 牛肉の新しい評価軸の検討. 畜産の情報 = Livestock industries information/農畜産業振興機構調査情報部情報課編, No. 339, pp. 52–61, 2018.
- [158] 安藤寿康. 日本人の 9 割が知らない遺伝の真実. SB クリエイティブ株式会社, 2016.
- [159] 小塩真司, 阿部晋吾. 日本語版 ten item personality inventory (tipi-j) 作成の試み. パーソナリティ研究, Vol. 21, No. 1, pp. 40–52, 2012.
- [160] 小塩真司. パーソナリティ心理学: progress & application. Progress & application. サイエンス社, 2014.
- [161] 村上宣寛, 村上千恵子. 主要 5 因子性格検査の尺度構成. 性格心理学研究, Vol. 6,

- No. 1, pp. 29–39, 1997.
- [162] 並川努, 谷伊織, 脇田貴文, 熊谷龍一, 中根愛, 野口裕之. Big five 尺度短縮版の開発と信頼性と妥当性の検討. 心理学研究, Vol. 83, No. 2, pp. 91–99, 2012.
- [163] 和田さゆり. 性格特性用語を用いた big five 尺度の作成. 心理学研究, Vol. 67, No. 1, pp. 61–67, 1996.