博士論文
ユーザの多様な特性に適応する機械学習モデルのパーソナライズ化に関する研究

北海道大学 大学院情報科学院
情報科学専攻

上川 恭平

2026 年 3 月

目次

図目次

表目次

第1章

序論

1.1 本研究の背景

近年，機械学習に基づく基盤モデルは推薦，生成，認識および分類といった多様なタスクに対して高い性能を示し，日常的なサービスから産業応用に至るまで，その活用が急速に広がっている [1–3]．これらのシステムは，ユーザに対して推薦結果や生成物，あるいは物体認識や感情推定といった各種の推論結果を提示し，ユーザはこれらの出力を補助的な情報として意思決定を行う．したがって，システムが出力する情報がユーザのニーズ，価値観および判断基準にどの程度適合しているかは，モデルの信頼性および実用性を評価するうえで極めて重要である．しかしながら，多数のユーザから収集された大規模データに基づいて構築される汎用的なモデルは，総合的に高い性能を持ちつつも，その学習過程において多数派の傾向を主に捉え，それを反映した推論を行うため，ユーザ間に存在する細かな個人差を十分に認識することが困難である [4,5]．実際のユーザは，嗜好や興味だけでなく，情報の捉え方，注意の向け方，判断基準，さらには認知的バイアスといった多様な特性を有しており，これらの要因が意思決定に影響を与える．このような個人差が無視された場合，モデルが提示する出力がユーザの期待や価値基準と一致しないことによる不信感，不便さ，さらにはシステムの実用性の低下等の課題が生じる．

このような背景から，ユーザ自身に関する情報をモデルの学習および推論過程に取り入れることで，個々のユーザに適した出力を可能とするパーソナライズ化の研究が注目を集めている [6–8]．従来研究では，ユーザの行動履歴（購買履歴，視

聴履歴，評価履歴など），あるいは年齢・性別といった静的な属性情報を主な情報源として活用してきた [9,10]．これらのアプローチは，ユーザに依らずモデルが一定の出力を提示してしまう画一性の問題を部分的に解消してきたものの，行動履歴や静的な属性情報といった表層的な情報だけでは，ユーザがどのような観点に注意し，何を重視し，どのような判断過程を経て意思決定しているかを十分に説明することは困難である．また，ユーザの意思決定には，明示的な選択だけでなく，無意識的な注意配分や反応の差異などが含まれており，これらを表層的な行動履歴から推定することは限界が存在する [11]．そのため，近年では，ユーザがコンテンツを視聴している際の動作情報や生体反応など，より直接的にユーザの内部状態を反映する特徴が注目されている [12,13]．本論文では，このようなユーザの無意識的な反応や内部状態を反映する生体信号および動作を「ユーザセントリック特徴」と定義する．ユーザセントリック特徴は，ユーザの属性や，結果的な行動が同一の場合においても，個人差を有するものであるため，パーソナライズ化に有効であると考えられる．しかしながら，これらの特徴は，映像やテキストといったマルチメディアコンテンツに対して異質であり，ノイズを多く含むため，モデルの学習および推論にどのように利用すべきかについては，理論面・実装面の双方で未解決の課題が残されている [14]．さらに，ユーザに関する情報を用いたパーソナライズ化には，対象ユーザから十分な量のデータを収集する必要があるという実用上の大きな制約が存在する [15]．ユーザの個人差をより正確に把握するためには，対象ユーザ固有の情報が求められるが，その収集にはコストがかかり，規模の拡大に伴い非現実的となる．また，他のユーザと区別するための情報が乏しい場合には，学習自体が不安定になり，モデルが期待されるパーソナライズ化を実現できないという問題も生じる．以上の背景を踏まえると，ユーザの個人差をより明確に捉えつつ，実用的なコストでパーソナライズ化を実現する技術の確立が強く求められている．

1.2 従来研究とその限界

ユーザの特性を推論に反映するための研究は，主に以下の 3 つの観点から発展してきた．(1) ユーザの閲覧・購買などの行動履歴や静的属性を学習段階で利用し，それらをモデルに反映する手法，(2) ユーザセントリック特徴を学習段階で活用し，

それらをモデルに反映する手法，(3) すでに学習済みのモデルを，対象ユーザ向けに適応させる手法である．本節では，これら従来研究の代表的アプローチとその限界について整理する．

(1) 行動履歴・静的属性を活用したパーソナライズ化の限界

従来のパーソナライズ化手法は，主にユーザの閲覧履歴，購買履歴，あるいは年齢・性別などの静的属性に依存していた．これらは取得が容易かつ統一的であり，大規模サービスにおけるモデルパーソナライズ化の基盤として広く利用されている [16]．しかし，これらの情報はユーザの最終的な行動の結果に過ぎず，注意の向け方，価値判断の基準，認知的バイアスといった，意思決定を形成する過程そのものを反映していない [17]．そのため，表層的な履歴・属性情報だけでは，個人差の根源的な要因を把握することは困難である．

(2) ユーザセントリック特徴を活用する手法の限界

ユーザがコンテンツを視聴している際の視線，動作，身体反応などは，ユーザの内部状態を直接的に反映する重要な情報源である [18]．これらを学習に取り入れる研究が進められているが，ユーザセントリック特徴はマルチメディアコンテンツとは異質であり，モダリティ統合の方法論は依然確立途上にある [19]．また，ユーザごとに得られるデータの量が異なるため，モデルのスケール拡大時に情報の不足や偏りが顕在化し，学習の安定性が損なわれるという課題が残されている [20]．

(3) 学習済みモデルのユーザ適応手法の限界

複数人のユーザから取得された大規模データで事前学習済みのモデルは，多数のユーザに共通する特徴を反映した推論を行う [21]．対象ユーザに特化した出力を得るためには，再学習や追加学習を行う必要があるが，これには高い計算コストが発生し，ユーザごとにモデルを再構築することは現実的でない．また，対象ユーザに関するデータが少量である場合，追加学習によって過学習や破滅的忘却が生じやすく，むしろ性能が低下する可能性も存在する [22,23]．

以上より，ユーザセントリックな特徴を適切に取り入れつつ，少量データかつ低コストでパーソナライズ化を実現するという観点が重要な研究課題となっている．

1.3 本研究の目的

前節までに述べた背景および課題を踏まえ，本研究では，ユーザの多様な特性をより直接的かつ柔軟にモデルへ反映することを目的とし，**(i) 学習段階におけるユーザセントリック特徴の統合**，および**(ii) 学習済みモデルのユーザ適応**の二つの観点から新たな枠組みを構築する．以下ではそれぞれの目的と着想を述べる．

(1) 学習段階におけるユーザセントリック特徴の統合

従来のパーソナライズ化手法では，閲覧履歴や人口統計情報など，行動に基づく静的な特徴が主に利用されてきた．しかし，これらの情報はユーザの認知的特徴や意思決定過程を十分に表現できないという制約がある [24]．そこで本研究では，ユーザがコンテンツ閲覧時に示す動作情報などのユーザセントリック特徴と，映像やテキストといったマルチメディアコンテンツ特徴を統合した表現をモデルに導入する．これにより，ユーザ固有の注意傾向や評価傾向を学習段階から反映し，従来手法では扱いが困難であった認知的な個人差に基づく推論を実現する．

(2) 学習済みモデルのユーザ適応

大規模データで学習された汎用的なモデルは，多数ユーザの傾向を平均化したパラメータ構造を持つため，特定のユーザにとって不適合な知識が含まれている場合がある．このような不要な知識の忘却のため，本研究では，マシンアンラーニング [25] の技術をモデルパーソナライズ化の文脈に取り入れた．具体的に，対象ユーザと類似するユーザに由来する知識を保持しつつ，対象ユーザと大きく異なる傾向を持つユーザに起因する知識を選択的に忘却する新たなパーソナライズ化手法を検討する．これは追加学習とは異なり，大規模な再学習を必要とせず，モデルの部分的な再学習や，モデルパラメータの局所的な更新のみで個人化を実現できる点に特徴がある．データ量の少ないユーザに対しても適応可能であり，計算コスト・データ収集コストの両面において実用性が高いアプローチである．

本論文は，ユーザの多様な特性が反映されたユーザセントリック特徴を用いたモデル学習と，学習済みモデルから他ユーザ由来の知識を選択的に保持・忘却するという学習後のユーザ適応という，二つの方向性からパーソナライズ化手法を

提案する．これにより，個人差を反映した柔軟な推論と，実運用に耐えうる効率的なパーソナライズ化の両立を目指す．

1.4 本論文の構成

本論文は，全5章で構成され，各章の内容は以下のとおりである．

- 第1章では，本研究の背景，従来研究の限界および本研究の目的を述べた．
- 第2章では，モデルのパーソナライズ化に関連する既存研究を概観し，本研究が対象とする課題の整理および本研究の立ち位置の明確化を行う．
- 第3章では，ユーザセントリック特徴とマルチメディアコンテンツ特徴の特徴統合手法を提案し，映像コンテンツを対象とした関心度推定タスクにおいてその有効性を検証する．
- 第4章では，学習済みモデルから，他ユーザ由来の知識を選択的に保持・忘却することでユーザ適応を行う手法を提案し，ユーザの感情推定タスクにおいてその性能の評価を行う．
- 第5章では，本研究の成果を総括し，今後の課題について述べる．

以上より，本研究は，ユーザの個人差に基づく柔軟で実用的なパーソナライズ化手法を，学習段階と学習後の段階の双方から多面的に提案したものである．

第2章

関連研究

2.1 はじめに

本章では，本研究に関連するモデルパーソナライズ化に関する技術を，**(i) モデル学習時／モデル学習後**，および**(ii) 知識の追加／知識の忘却または削減**の二軸から体系的に整理する．また，これらの軸に基づき，図 2.1 に示すように，既存手法を以下の 4 つに分類する．(1) 学習時における知識追加：行動履歴・静的属性の利用，(2) 学習時における知識忘却または削減：知識の選択・圧縮による学習効率化，(3) 学習後における知識追加：追加データを用いたモデル適応，(4) 学習後における知識忘却または削減：マシンアンラーニング.

以下，各節で既存研究を整理し，それぞれの利点と限界を示したうえで，本研究がどの領域に立脚するかを明確にする．

2.2 学習時における知識追加：行動履歴・静的属性の利用

学習段階におけるパーソナライズ手法の多くは，ユーザの閲覧履歴，クリックログ，評価履歴，人口統計情報といった静的または行動ベースの特徴をモデルに直接取り込み，ユーザ固有の嗜好や長期的傾向を学習しようとするものである．これらは大規模ユーザを対象とした個人化の基盤として広く用いられてきたが，あくまで行動の結果であり，注意・判断基準・感情傾向といった内的要因を直接記述するものではない．

YouTube DNN [26]

Deep Neural Networks for YouTube Recommendations (YouTube DNN) は，深層学習

を用いた大規模推薦システムの草分け的研究であり，ユーザの視聴履歴と年齢・性別・居住地などの静的属性を結合して埋め込みベクトルを学習する手法である．数億規模のユーザに対して汎用的に適用可能である一方，すべての特徴を単一の密ベクトルに圧縮するため，ユーザの微細な文脈や動的な興味の変化を捉えきれないという課題がある．本研究における静的属性を利用したパーソナライズ化手法の代表的なベースラインとして位置づけられる．

BERT4Rec [27]

BERT4Rec は，自然言語処理で成功した Transformer アーキテクチャをユーザ行動系列のモデリングに応用した手法である．ユーザの過去の行動順序を文脈として捉え，双方向の Attention 機構を用いることで，従来の RNN ベースの手法よりも深く複雑な依存関係を学習可能とした．行動履歴を高度に抽象化して知識として取り込む手法であるが，入力情報は依然として行動履歴に限定されており，ユーザの内部状態には踏み込んでいない．

P5 [28]

Recommendation as Language Processing (P5) は，ユーザの ID や行動履歴，属性情報をすべて自然言語のテキストプロンプトとして記述し，大規模言語モデル（LLM）を用いて推薦を行う統一フレームワークである．履歴の言語化によって，モデルがユーザ知識を文脈として理解しやすくなる利点がある．しかし，これらの情報はあくまで過去の履歴や属性の記述に留まり，ユーザ固有の内部状態や反応を直接観測するものではない．本研究の第 3 章では，P5 が対象とするテキスト化された履歴情報ではなく，言語化が困難なユーザセントリック特徴に着目し，これをマルチメディア特徴と統合する新たなアプローチをとる．

2.3 学習時における知識忘却または削減：知識の選択・圧縮による学習効率化

データセットやモデル内部表現の一部を選択的に利用し，効率的な学習を実現する手法群である．本研究とは実行する段階は異なるものの，不要な知識を削減

するという観点で関連がある．これらは，モデルにとって真に必要な知識のみを抽出し，ノイズや冗長性を排除するプロセスと解釈可能である．

Dataset Distillation [29]

Dataset Distillationは，大規模な学習データセットを，その情報の要約となるごく少数の合成データへと圧縮する技術である．元の全データで学習した場合と同等の性能を，圧縮されたデータのみで再現することを目的としており，モデルにとって不可欠な知識成分のみを抽出する技術といえる．学習コストの削減だけでなく，冗長な知識を削減し，保持すべき知識の純化するという観点で，本研究の知識選択アプローチに関連する．

Pruning [30]

Pruningは，巨大なニューラルネットワークの中には，初期化状態で特定のスパースな部分構造が存在することを仮定し，それを学習することで元のモデルと同等の精度を達成する手法である．これは，学習に必要なパラメータは全体のごく一部であり，他は冗長であることを示唆している．知識の物理的な選択と削減によって学習効率化を図るアプローチの理論的支柱である．

Influence Function [31]

Influence Functionは，あるテストデータの予測結果に対して，学習データ中のどのサンプルがどの程度影響を与えたかを定量的に推定する技術である．モデルをブラックボックスとして扱わず，個々のデータが与える知識への貢献度を可視化できるため，モデルから特定のデータの影響を取り除くマシンアンラーニングや，第4章で扱う知識の選択的制御を実現するための理論的基盤として重要である．

2.4 学習後における知識追加：追加データを用いたモデル適応

モデル学習後のパーソナライズ手法として，追加データを用いてモデルを再調整するアプローチが広く研究されている．これらの手法は，学習済みモデルに新たな知識を追加することにより，対象ユーザに特化した表現を後から獲得できる点で有効であるが，既存知識への干渉や過学習，また，不要な知識がモデルに残存し続けるというリスクを伴う．

LoRA [32]

Low-Rank Adaptation (LoRA) は，大規模モデルの全パラメータを更新する代わりに，低ランク行列のみを追加して学習する効率的な適応手法である．計算コストを抑えつつ，少量の追加データでモデルを特定のドメインやユーザに適応させることができるため，現在最も普及している手法の一つである．しかし，これはあくまで知識の追加であり，モデル内に潜在する不要な知識を明示的に忘却する機能は持たない．

MAML [33]

Model-Agnostic Meta-Learning (MAML) は，少量のデータとわずかな勾配更新で新しいタスクに適応できる初期パラメータを学習するメタラーニングの基礎的枠組みである．パーソナライズの文脈では，個々のユーザを独立したタスクとみなすことで，未知のユーザに対しても即座に適応可能なモデルを構築する手法として応用される．高速な適応を実現するメタラーニングの代表的手法であり，個人化において少数データでの学習を効率化する標準的なアプローチとして位置づけられる．

Per-FedAvg [34]

Per-FedAvg は，MAML の概念を連合学習（Federated Learning）に拡張した手法である．従来の連合学習（FedAvg）が全ユーザに共通する平均的なモデルの構築を目指すのに対し，Per-FedAvg は，各ユーザがローカルデータで学習した後の性能を最大化するようにグローバルモデルを更新する．これにより，ユーザごとのデータ分布が大きく異なる場合でも，個別の適応能力が高いモデルを提供可能とする，分散環境におけるパーソナライズの代表的手法である．

kNN-LM [35]

kNN-LM は，モデルのパラメータを更新するのではなく，外部のデータストアから類似した事例（知識）を検索し，推論時に動的に知識を補完する手法である．Retrieval-Augmented Generation (RAG) [36] の基礎となる技術であり，モデル自体を変えずに知識を追加するアプローチである．しかし，これはあくまで知識の追加に特化した手法であり，パラメータ更新を伴わないため，汎

用的なモデルが元々保持している不要な知識や他者のバイアスを忘却することはできない．本研究は，知識を加えるだけでなく，不要な知識を削ぎ落とすこともパーソナライズに不可欠であるという立場をとる点で，本手法とは対照的な位置にある．

2.5 学習後における知識忘却または削減:マシンアンラーニング

マシンアンラーニングは，学習済みモデルから特定のデータに由来する知識を忘却させ，その影響をモデルの出力から削除することを目的とした技術群である．従来はプライバシー保護が主な用途であったが，近年では，誤ったデータやノイズの影響を後から取り除くことで，モデルの性能を改善する利用方法も注目されている [37].

SISA [25]

Sharded, Isolated, Sliced, Aggregated (SISA) は，学習データを複数のシャードに分割して独立に学習し，忘却要求があった場合に該当シャードのみを再学習するフレームワークである．完全な忘却を保証する重要なベースライン手法であるが，モデルの分割による性能低下や，推論時の計算コスト増大が課題となる．特定の知識を確実に消去する手法として，アンラーニングの性能評価における基準となる．

SCRUB [38]

SCRUB (Scalable Remembering and Unlearning Unbound) は，教師生徒モデルの枠組みを用い，忘却対象のデータに対しては生徒モデルの出力を抑制しつつ，保持対象データに対しては性能を維持するように学習する手法である．再学習を行わずに特定の知識を選択的に洗い流す（Scrub）ことが可能であり，高い実用性を持つ．本研究の第 4 章において，他者由来の知識を忘却するための具体的なアルゴリズムとして参照される重要な先行研究である．

Federated Unlearning [39]

Federated Unlearning は，連合学習によって構築されたグローバルモデルから，特定のクライアント（ユーザ）が寄与した知識のみを効率的に削除する技術である．従来は，システムから離脱するユーザの権利（忘れられる権利）を

保障するため，再学習を行うことなく特定ユーザの影響を無効化する手法として発展してきた．このような,「特定のユーザ由来の知識を選択的に消去可能」である機構は，本研究が目指す知識制御（不要な他者知識の除去）を実現するうえで，技術的な基盤となるものである．

2.6 まとめ

本章では，パーソナライズ技術を「学習時／学習後」，および「知識追加／知識忘却・削減」の二軸に基づき体系的に整理した．既存研究は個別の課題に対して効果的な手法を提供してきたが，以下の二点において未解決の課題が残されている．

第一に，学習段階において，行動履歴などの表層情報だけでなく，ユーザの内部状態を反映する**ユーザセントリック特徴をマルチメディアコンテンツとどのように統合すべきか**という点である．第二に，学習後のモデルに対して，**ユーザ間の類似度に基づいて保持すべき知識（他者の有用な知識）と忘却すべき知識（他者のノイズ）を選択的に制御する枠組み**が十分に確立されていない点である．

本研究はこのギャップを埋めるために，第3章においてユーザセントリック特徴の統合手法を，第4章において知識の選択的保持と忘却によるユーザ適応手法を提案する．

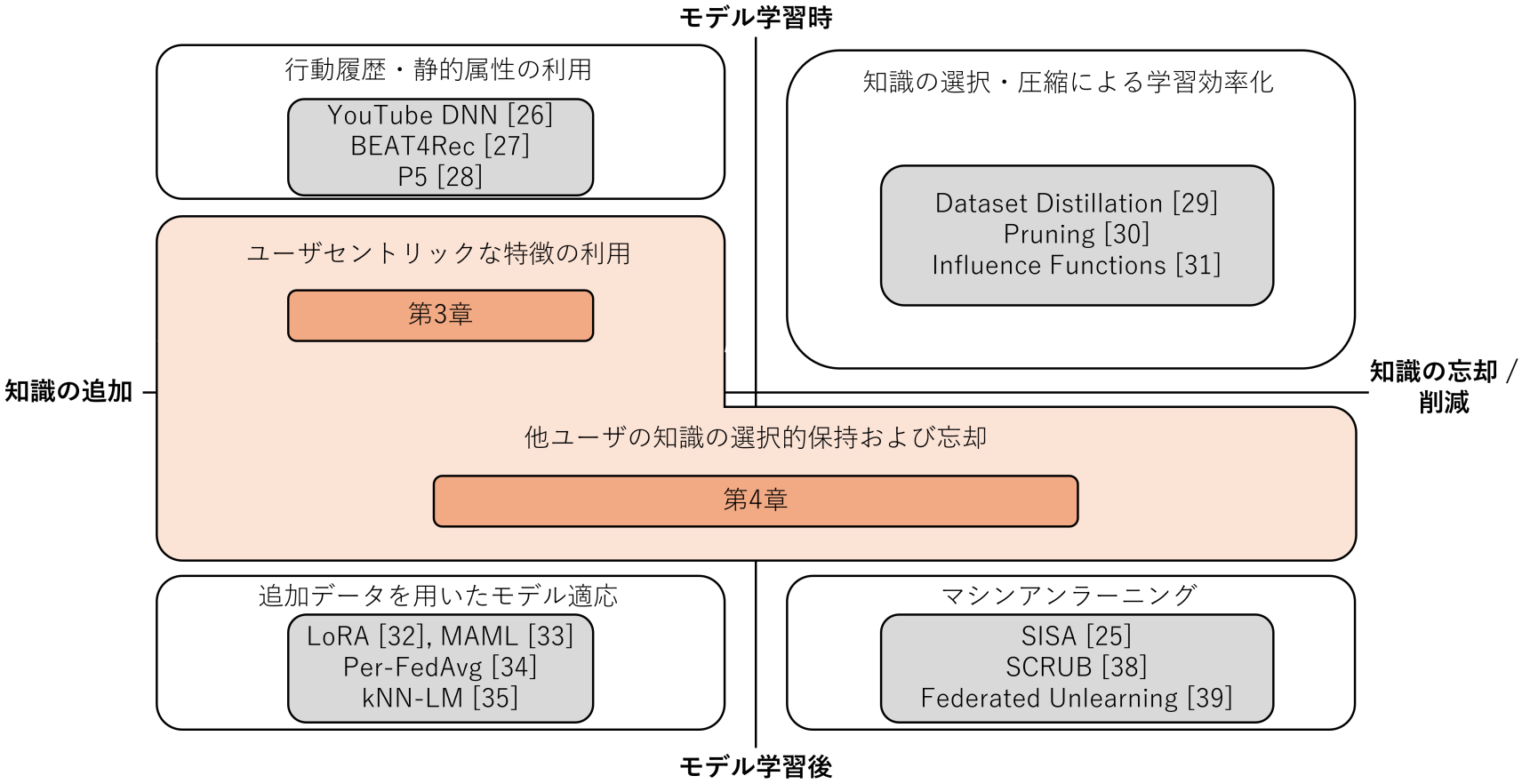

図 2.1: リサーチマップと本研究の位置付け.

第3章

ユーザセントリック特徴とコンテンツ特徴のマルチモーダル特徴統合

3.1 はじめに

近年のWebサービスにおけるマルチメディアコンテンツの爆発的な増加に伴い，ユーザごとの多様な特性や嗜好を正確に把握し，個々のユーザに適した出力を提供するモデルのパーソナライズ化の重要性が高まっている [40,41]. 従来のアプローチでは，主にコンテンツの映像やテキストといった特徴量や，閲覧履歴などのユーザの表層的な情報を利用してきたが，これらはユーザ個人の主観的な興味や反応（関心度）を直接的に反映するものではないため，その推定精度には限界があった．そこで，第1.1節で定義したユーザセントリック特徴，すなわちコンテンツ視聴中のユーザの視線や身体動作といった生体反応情報を利用し，ユーザの内部状態をより正確に捉えるアプローチが求められている [42,43].

ユーザセントリック特徴とコンテンツ特徴を統合して関心度を推定するためには，これら異種特徴間の関係性を適切にモデル化する必要がある．しかしながら，ユーザセントリック特徴は取得コストが高く大規模なデータ収集が困難であることに加え，ユーザの無意識的な振る舞いや環境要因によるノイズを多く含むという性質を持つ．そのため，大量の学習データを必要とする決定論的な手法や深層学習モデルでは，学習データへの過適合が生じやすく，汎化性能が低下するという課題があった [44].

本章では，これらの課題を解決するために，特徴のノイズに頑健な確率的生成

モデルである Multi-modal Gaussian Process Latent Variable Model (mGPLVM) [45] を特徴統合の基盤とする．mGPLVM は，少量のデータやノイズを含むデータであっても，観測値の背後にある共通の潜在空間を確率的に推定可能である．本章では，この mGPLVM を基盤とし，(i) 動作情報などの異種特徴の統合や，(ii) ラベル不足，順序ラベルおよび部分ラベルといった現実的な制約への対応という観点から段階的に拡張した手法について論じる．

本章の構成は以下の通りである．まず，第 3.2 節では，従来手法の枠組みにおいて連続値な特徴量を扱うための量子化手法を導入し，ユーザの動作特徴とコンテンツ特徴を統合する手法である **Multi-modal Gaussian Process Latent Variable Factorization (mGPLVF)** を提案する．次に，第 3.3 節では，ラベル無しサンプルおよび順序性を持つラベルを考慮した半教師あり学習モデル **Semi-supervised Ordinally Multi-modal Gaussian Process Latent Variable Model (Semi-OMGP)** により，ラベル不足環境下での特徴統合手法を提案する．第 3.4 節では，Semi-OMGP においてラベル無しデータに与える制約が推定精度に与える影響について検証する．第 3.5 節では，逆射影を導入した手法である **Back-projection Ordering Multi-modal Gaussian Process Latent Variable Model (BPomGP)** を提案し，動作特徴が取得不能である未視聴コンテンツに対する推論を可能にする枠組みについて述べる．第 3.6 節では，擬似ラベルを用いた半教師あり学習モデル **Multimodal Gaussian Process Latent Variable Model with Pseudo-Labels (Semi-MGPPL)** により，ラベル不足と未視聴コンテンツへの対応を同時に実現する拡張的な特徴統合手法を示す．最後に，3.7 節にて本章の結論を述べる．

3.2 特徴量の量子化およびmGPLVFに基づくユーザの関心度推定

3.2.1 はじめに

第3.1節で述べたように，ユーザセントリック特徴の一つであるユーザの動作特徴とコンテンツ特徴の統合は関心度推定において有効である．これらのマルチモーダルな特徴を統合し，推論を行う確率モデルとして，Gaussian Process Latent Variable Model Factorization (GPLVMF) [46] が挙げられる．GPLVMFは，ユーザの行動履歴や静的属性のようなカテゴリカルな特徴量（離散値）を連結させた，単一のモダリティを入力とすることを前提としており，この条件下で高精度な関心度推定が可能である．

しかしながら，本研究で扱うコンテンツ特徴や動作特徴は，連続値の実数ベクトルであり，かつこれらが複数存在するマルチモーダルなデータである．離散値を前提として設計されたGPLVMFの尤度関数やモデル構造に対し，連続値の特徴量を直接適用した場合，モデルの仮定との不整合により最適化が困難となる．また，単一モダリティを想定した構造のままでは，複数の異種特徴を適切に統合することが困難である．

そこで本節では，これらの課題を解決するために，**Multi-modal Gaussian Process Latent Variable Factorization (mGPLVF)** を提案し，これに基づくユーザの関心度推定法を構築する．提案手法では，連続値である各モダリティの特徴量に対し，それぞれ適切な量子化（離散化）を行う．これにより，連続値をGPLVMFが処理可能なカテゴリカル形式に変換し，モデルの最適化を可能にする．さらに，複数のモダリティを同時に扱えるようにモデルを拡張することで，コンテンツ特徴と動作特徴を統合した高精度なユーザの関心度推定を実現する．

3.2.2 提案手法

本章では，提案手法について説明する．提案手法の概要を図3.1に示す．提案手法では，まず，コンテンツ特徴およびユーザの動作特徴を算出し，それらの特徴量に対して量子化を行う．次に，得られた特徴量およびユーザの関心度を用いてモデルを学習し，潜在変数およびハイパーパラメータの最適化を行う．最終的に，

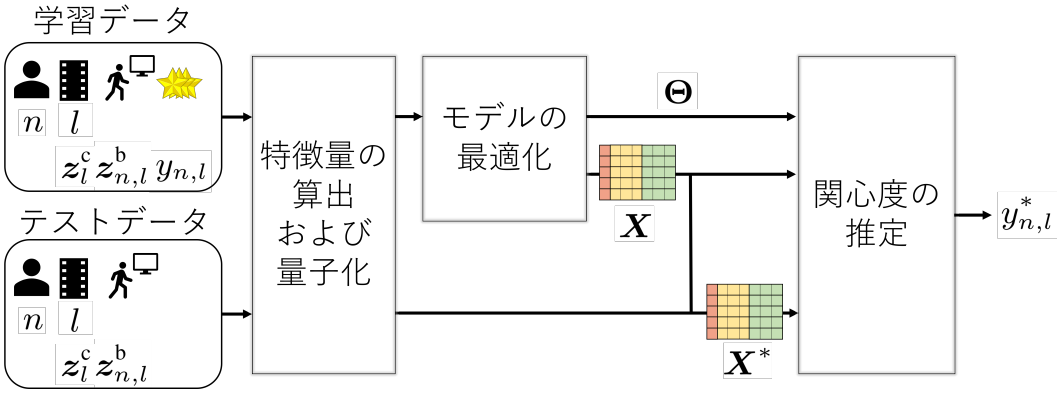

図 3.1: 提案手法の概要図

最適化された潜在変数およびハイパーパラメータを用いて，ユーザの関心度を推定する．

特徴量の算出および量子化

コンテンツ l $(= 1, \ldots, L; L$ はコンテンツの総数$)$ に対するユーザ n $(= 1, \ldots, N; N$ はユーザの総数$)$ の動作特徴 $\boldsymbol{z}_{n,l}^{\mathrm{b}} \in \mathbb{R}^{D^{\mathrm{b}}}$ を OpenPose [47] および Tobii Eye Tracker 4C[1]を用いて取得する．また，コンテンツ l に対するコンテンツ特徴 $\boldsymbol{z}_l^{\mathrm{c}} \in \mathbb{R}^{D^{\mathrm{c}}}$ を Inception-v3 [48] を用いて算出する．ユーザ n がコンテンツ l に与えた関心度を $y_{n,l}$ とし，$y_{n,l}$ をユーザ n について連結させたものを $\boldsymbol{y}_n \in \mathbb{R}^{L_n}$ とする．ただし，L_n はユーザ n が視聴したコンテンツの総数である．

次に，図 3.2 に示すように，特徴量 $\boldsymbol{z}_l^{\mathrm{c}}$ および $\boldsymbol{z}_{n,l}^{\mathrm{b}}$ の量子化を行う．特徴量のそれぞれの次元の値について，最小値から最大値まで $H^m (m \in \{\mathrm{c}, \mathrm{b}\})$ 個のビンに分割する．特徴量の各次元にビンの番号 $h^m \in \{1, 2, \ldots, H^m\}$ を代入したものを，量子化後の特徴量 $\tilde{\boldsymbol{z}}_l^{\mathrm{c}} = [\tilde{z}_l^{\mathrm{c}}(1), \tilde{z}_l^{\mathrm{c}}(2), \ldots, \tilde{z}_l^{\mathrm{c}}(D^{\mathrm{c}})]$ および $\tilde{\boldsymbol{z}}_{n,l}^{\mathrm{b}} = [\tilde{z}_{n,l}^{\mathrm{b}}(1), \tilde{z}_{n,l}^{\mathrm{b}}(2), \ldots, \tilde{z}_{n,l}^{\mathrm{b}}(D^{\mathrm{b}})]$ とする．

提案手法では，l, $\tilde{z}_l^{\mathrm{c}}(d)$ および $\tilde{z}_{n,l}^{\mathrm{b}}(d)$ に対して，それぞれ 2 種類の潜在変数を仮定する．これらを，カーネルに対する潜在変数および平均値に対する潜在変数と定義する．具体的に，l, $\tilde{z}_l^{\mathrm{c}}(d)$ および $\tilde{z}_{n,l}^{\mathrm{b}}(d)$ のカーネルに対する潜在変数を $\boldsymbol{v}_l \in \mathbb{R}^{Q_v}, \boldsymbol{v}_{d_l}^{\mathrm{c}} \in \mathbb{R}^{Q_v^{\mathrm{c}}}$ および $\boldsymbol{v}_{d_{n,l}}^{\mathrm{b}} \in \mathbb{R}^{Q_v^{\mathrm{b}}}$ と仮定する．l, $\tilde{z}_l^{\mathrm{c}}(d)$ および $\tilde{z}_{n,l}^{\mathrm{b}}(d)$ の平均値に対する潜在変数を $\boldsymbol{w}_l \in \mathbb{R}^{Q_w}, \boldsymbol{w}_{d_l}^{\mathrm{c}} \in \mathbb{R}^{Q_w^{\mathrm{c}}}$ および $\boldsymbol{w}_{d_{n,l}}^{\mathrm{b}} \in \mathbb{R}^{Q_w^{\mathrm{b}}}$ と仮定する．更に，$\boldsymbol{X}_n^v \in \mathbb{R}^{L_n \times Q_v^{\mathrm{sum}}}$ および $\boldsymbol{X}_n^w \in \mathbb{R}^{L_n \times Q_w^{\mathrm{sum}}}$ を 2 種類の潜在変数をユーザ n について連結させた行列とする．ただ

[1]https://tobiigaming.com/eye-tracker-4c

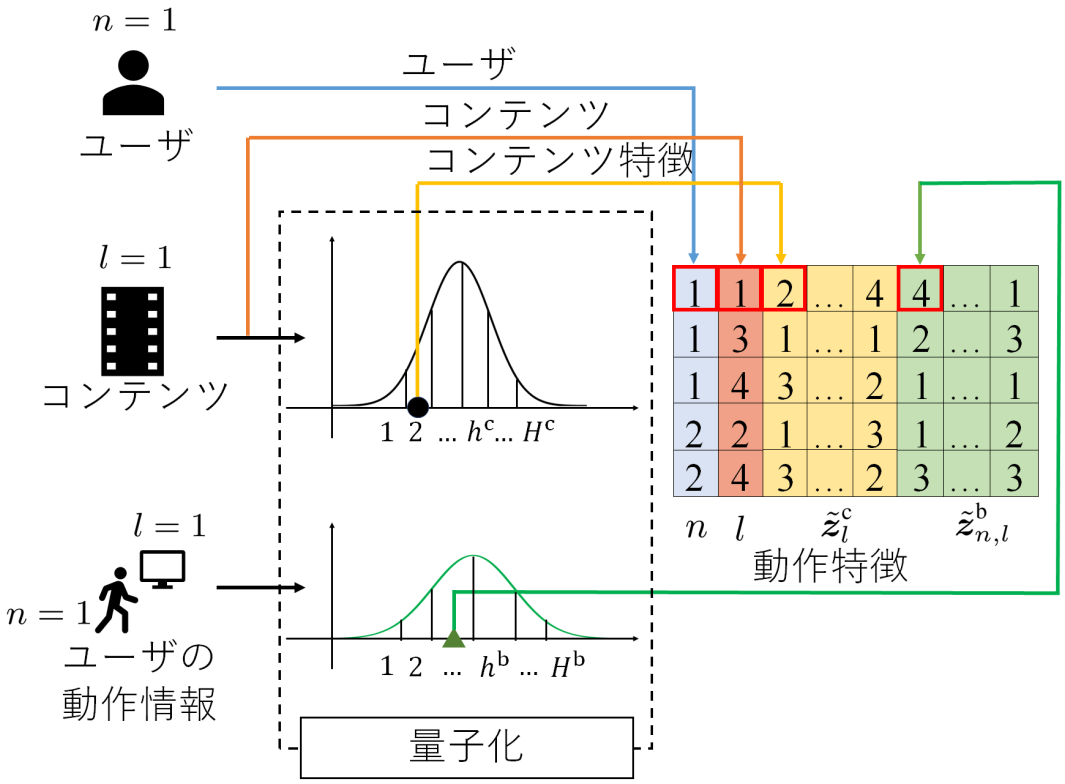

図 3.2: 特徴量の算出および量子化の概要図

し，$Q_v^{\mathrm{sum}} = Q_v + Q_v^{\mathrm{c}} D^{\mathrm{c}} + Q_v^{\mathrm{b}} D^{\mathrm{b}}$ および $Q_w^{\mathrm{sum}} = Q_w + Q_w^{\mathrm{c}} D^{\mathrm{c}} + Q_w^{\mathrm{b}} D^{\mathrm{b}}$ とする．また，ユーザ n に対するバイアス項を w_n^{u} とする．

モデルの最適化

$\boldsymbol{y}_n$ と潜在変数 $\boldsymbol{X}_n^v, \boldsymbol{X}_n^w$ がガウス過程に従うと仮定する．$\boldsymbol{y}_n$ の平均ベクトルを $\boldsymbol{m}_n$，共分散行列を $\boldsymbol{K}_n$ とすると，$\boldsymbol{y}_n$ の周辺尤度は次式で表される [49].

$$p(\boldsymbol{y}_n | \boldsymbol{X}_n^v, \boldsymbol{X}_n^w, \boldsymbol{\theta}_n) = \mathcal{N}(\boldsymbol{y}_n | \boldsymbol{m}_n, \boldsymbol{K}_n + \beta_n^{-1} \boldsymbol{I}_{L_n}) \tag{3.1}$$

ただし，$\boldsymbol{I}_{L_n}$ は L_n 次元の単位行列，$\boldsymbol{\theta}_n$ はノイズのハイパーパラメータ β_n およびカーネルのハイパーパラメータである．平均ベクトル $\boldsymbol{m}_n$ は次式により求める．

$$m_n(i) = w_n^{\mathrm{u}} + \sum_{q=1}^{Q_w} w_i(q) + \sum_{d=1}^{D^{\mathrm{c}}} \sum_{q=1}^{Q_w^{\mathrm{c}}} w_{d_i}^{\mathrm{c}}(q) + \sum_{d=1}^{D^{\mathrm{b}}} \sum_{q=1}^{Q_w^{\mathrm{b}}} w_{d_{n,i}}^{\mathrm{b}}(q) \tag{3.2}$$

また，共分散行列 $\boldsymbol{K}_n$ はカーネル関数 $k(\cdot,\cdot)$ を用いて次式により求める．

したがって，$\boldsymbol{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N]$ の周辺尤度は次式で表される．

$$p(\boldsymbol{Y}|\boldsymbol{X}^v, \boldsymbol{X}^w, \boldsymbol{\Theta}) = \prod_{n=1}^{N} p(\boldsymbol{y}_n|\boldsymbol{X}_n^v, \boldsymbol{X}_n^w, \boldsymbol{\theta}_n) \tag{3.3}$$

ただし，$\boldsymbol{\Theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_N]$ である．更に，$\boldsymbol{X}^v$ の事前分布を次式により定義する．

$$p(\boldsymbol{X}^v) = \prod_{l=1}^{L} \left\{ \mathcal{N}(\boldsymbol{v}_l|\boldsymbol{0}, \boldsymbol{I}_{Q_v}) \mathcal{N}(\boldsymbol{v}_{d_l}^{\mathrm{c}}|\boldsymbol{0}, \boldsymbol{I}_{Q_v^{\mathrm{c}}}) \prod_{n=1}^{N} \mathcal{N}(\boldsymbol{v}_{d_{n,l}}^{\mathrm{b}}|\boldsymbol{0}, \boldsymbol{I}_{Q_v^{\mathrm{b}}}) \right\}$$

同様にして，$\boldsymbol{X}^w$ の事前分布 $p(\boldsymbol{X}^w)$ を定義する．最大事後確率推定により，次式の最適化問題を解くことで，最適な潜在変数およびハイパーパラメータの算出を行う．

$$\max_{\boldsymbol{X}^v, \boldsymbol{X}^w, \boldsymbol{\Theta}} p(\boldsymbol{Y}|\boldsymbol{X}^v, \boldsymbol{X}^w, \boldsymbol{\Theta}) p(\boldsymbol{X}^v) p(\boldsymbol{X}^w) \tag{3.4}$$

式（3.4）は解析的に解くことが困難であるが，一般化変分推論 [46] により解くことが可能である．

関心度の推定

テストデータに対して，前節と同様に特徴量の算出および量子化を行う．また，前節で算出した潜在変数を用いて，テストデータの潜在変数 $\boldsymbol{X}_n^*$ を得る．関心度の推定値 $\boldsymbol{y}'_n \in \mathbb{R}^{L_n^*}$ を次式により算出する．

$$q(\boldsymbol{y}'_n) = \int \left\{ p(\boldsymbol{y}'_n|\boldsymbol{u}_n, \boldsymbol{X}_n^*) q(\boldsymbol{u}_n) d\boldsymbol{u}_n \right\} q(\boldsymbol{X}_n^*) d\boldsymbol{X}_n^* \tag{3.5}$$

ただし，$q(\cdot)$ は変分事後分布であり，$\boldsymbol{u}_n \in \mathbb{R}^M$ および $\boldsymbol{T}_n \in \mathbb{R}^{M \times Q_v^{\mathrm{sum}}}$ は誘導点および誘導変数 [50] である．$q(\boldsymbol{y}'_n)$ はガウス分布 $\mathcal{N}(\boldsymbol{y}'_n|\boldsymbol{\mu}_n^*, \boldsymbol{\Sigma}_n^*)$ に書き換えることが可能である．

ただし，

$$\boldsymbol{\mu}_n^* = \boldsymbol{\Psi}_{1_n}^* (\beta_n^{-1} \boldsymbol{K}_{M,M} + \boldsymbol{\Psi}_{2_n})^{-1} \boldsymbol{\Psi}_{1_n}^{\mathsf{T}} \hat{\boldsymbol{y}}_n + \boldsymbol{\phi}_{1_n}^*$$

$$\boldsymbol{\Sigma}_n^* = \boldsymbol{\Psi}_{1_n}^* \{\boldsymbol{K}_{M,M}^{-1} + (\boldsymbol{K}_{M,M} + \beta_n \boldsymbol{\Psi}_{2_n})^{-1}\} (\boldsymbol{\Psi}_{1_n}^*)^{\mathsf{T}}$$

$$\boldsymbol{\Psi}_{1_n} = \langle \boldsymbol{K}_{L_n,M} \rangle_{q(\boldsymbol{X}_n^v)}$$

$$\boldsymbol{\Psi}_{2_n} = \langle \boldsymbol{K}_{L_n,M}^{\mathsf{T}} \boldsymbol{K}_{L_n,M} \rangle_{q(\boldsymbol{X}_n^v)}$$

$$\boldsymbol{\phi}_{1_n} = \langle \boldsymbol{m}_n \rangle_{q(\boldsymbol{X}_n^w)}$$

$$\hat{\boldsymbol{y}}_n = \boldsymbol{y}_n - \boldsymbol{\phi}_{1_n}$$

である．動作特徴が類似したユーザは，類似した関心度を与えることを考慮して [42]，次式により最終的な関心度の推定値 $y_{n,l}^*$ を算出する．

$$y_{n,l}^* = \left(y'_{n,l} + \sum_{i=1}^{N-1} a_i s_{n,l}^i \right) / \left(1 + \sum_{i=1}^{N-1} a_i \right) \tag{3.6}$$

ただし，$a_i (a_1 \geqq a_2 \geqq \ldots \geqq a_{N-1})$ は重み係数である．$s_{n,l}^i$ は，$y_{:,l}$ および $y'_{:,l}$ の中で，対応する動作特徴が $y'_{n,l}$ と i 番目に類似したものである．

3.2.3 実験

本章では，mGPLVF を用いた特徴統合に基づくユーザの関心度推定の精度を検証するための実験を行う．

本実験では，データセットとして，YouTube から，映画の予告編（30 秒）を ‘Action’, ‘Comedy’, ‘Music’ および ‘Science’ のジャンルからそれぞれ 10 本，‘Sport’ のジャンルから 9 本，計 49 本取得した．本実験では，10 名の実験参加者に対して，各映像の視聴後に 4 段階の関心度 (1: 全く興味がない，2: あまり興味がない，3: 少し興味がある，4: とても興味がある) を付与させた．特徴量は，主成分分析 [51] を適用して次元を削減したものを用いた．その次元数は，$D^{\mathrm{c}} = 13$ および $D^{\mathrm{b}} = 13$ であった．本実験では，映像に対して付与された関心度を，$50, 60, \ldots, 90\%$ の割合でランダムに欠損させ，その欠損値の推定を行った．比較手法として，GPLVMF [46], mGPLVM [45], Multi-modal Similarity GPLVM (m-SimGP) [52], Multi-view Canonical Correlation Analysis (MVCCA) [53], Deep CCA [54] および Bayesian CCA (BCCA) [55] を用いた．ただし，提案手法と GPLVMF 以外の手法では，統合後の特徴を用いたテンソル補完 [56] に基

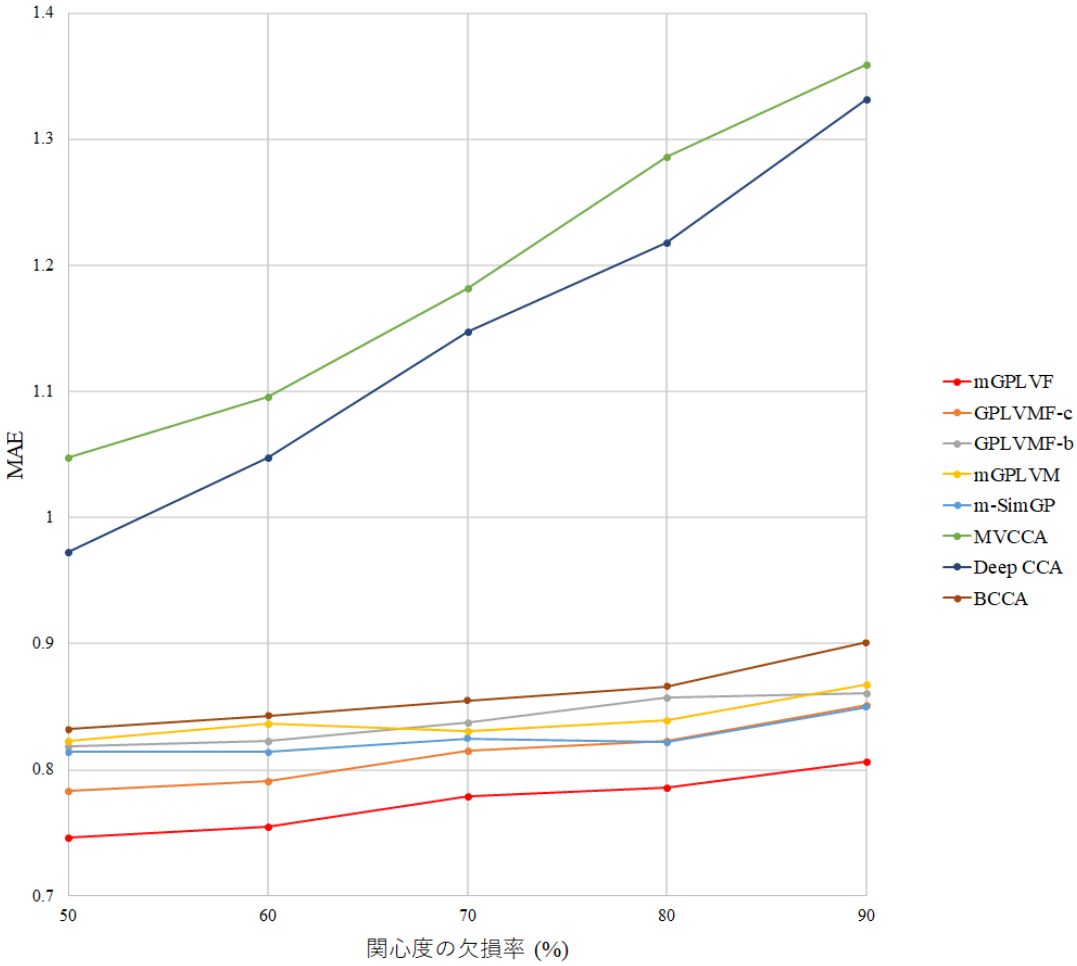

図 3.3: 関心度の各欠損率における MAE の平均

づき関心度の推定を行った．GPLVMFについては，コンテンツ特徴のみを用いた場合（GPLVMF-c）と，動作特徴のみを用いた場合（GPLVMF-b）について，それぞれ実験を行った．なお，Mean Absolute Error (MAE) を評価指標として推定精度の評価を行った．各実験参加者の各欠損率について，それぞれ10回の検証を行った．

図3.3に，実験結果を示す．図3.3より，提案手法は全ての欠損率において，比較手法と比較して最も低いMAEを達成しており，その推定精度の高さが確認できる．特に，コンテンツ特徴のみを用いたGPLVMF-cや，動作特徴のみを用いたGPLVMF-bと比較して，提案手法のMAEは大幅に低くなっている．これは，コンテンツ特徴と動作特徴という異なる性質を持つ情報を統合することで，単一のモダリティでは捉えきれないユーザの関心度に関わる潜在的な傾向を表現できたためと考えられる．また，MVCCAやDeep CCAといった決定論的な特徴統合手法と比較しても，提案手法は高い精度を示している．Deep CCAなどの深層学習に基づく手法は，高い表現能力を持つ一方で，本実験のようにサンプル数が限られている場合や，動作特徴にノイズが含まれる場合には過学習を起こしやすい傾向がある．これに対し，確率的生成モデルに基づく提案手法は，ノイズを考慮した柔軟なモデリングが可能であり，データの不確実性に対して頑健な推定ができたと推察される．さらに，mGPLVMやm-SimGPといった他のガウス過程に基づく手法と比較しても，提案手法は優れた性能を示している．これは，GPLVMFの枠組みに対し，連続値の特徴量を適切に量子化して組み込むという提案手法のアプローチが，関心度という順序性を持つラベルの推定に適していたことを示唆している．以上の結果より，提案手法は欠損率が高い（ラベル情報が少ない）状況下でも安定して高い推定精度を維持しており，その有効性が示された．

3.2.4 まとめ

本節では，mGPLVFを用いたユーザの関心度推定手法を提案した．提案手法では，連続値であるコンテンツ特徴および動作特徴を量子化し，GPLVMFの枠組みで統合することで，複数のモダリティを考慮した関心度推定を実現した．映画予告編動画を用いた評価実験の結果，提案手法は単一モダリティのみを用いる手法や，従来のマルチモーダル特徴統合手法と比較して，高い精度で関心度を推定できることを確認した．特に，ラベルの欠損率が高い状況においても提案手法の優

位性が確認され，実用的な観点からもその有効性が示された．

3.3 ラベルの順序性を考慮したSemi-OMGPに基づく特徴統合

3.3.1 はじめに

前節では，mGPLVFにより連続値の特徴量を量子化して統合する手法を示した．しかしながら，実環境におけるユーザの関心度推定には，さらに解決すべき二つの課題が存在する．第一に，コンテンツ特徴と動作特徴はその性質が大きく異なり，特徴空間における分布形状に乖離があるため，単純な共有空間への射影が困難である点である [52]．第二に，第3.1節でも触れた通り，ユーザによるアノテーションコストが高く，ラベル付きデータ（関心度が付与されたデータ）が不足しがちである点である [57]．加えて，ユーザの関心度は単純なクラス分類ではなく,「全く興味がない」から「とても興味がある」といった順序性 (Ordinality) を持つ値である [58] が，従来のmGPLVMに基づく手法ではこれらの性質を十分に考慮できていない．

そこで本節では，これらの課題を解決する **Semi-supervised Ordinally Multi-modal Gaussian Process Latent Variable Model (Semi-OMGP)** を提案する．本手法の主な貢献は以下の通りである．

- **分布の差異への対応**: 観測値そのものではなく，各モダリティの類似度行列が共通の潜在変数から生成されると仮定することで，モダリティ間の分布の差異に頑健な統合を実現する．
- **ラベル不足と順序性の考慮**: ラベルのあるデータだけでなくラベルのないデータも併せて学習する半教師あり学習の枠組みを導入する．さらに，ラベルの値の差が大きいほど潜在空間での距離が大きくなるような制約を導入することで，関心度の順序構造をモデルに反映させる．

3.3.2 従来手法

本章では，mGPLVMについて説明する．mGPLVMでは，M種類の観測値$\{\boldsymbol{Y}^m\}_{m=1}^M$を扱う．ただし，$\boldsymbol{Y}^m = [\boldsymbol{y}_1^m, \boldsymbol{y}_2^m, \ldots, \boldsymbol{y}_N^m]^\mathsf{T} \in \mathbb{R}^{N \times D_m}$である．mGPLVMの目的は，共通の潜在変数$\boldsymbol{X}$およびハイパーパラメータ$\boldsymbol{\Theta}$の最適化である．ただし，$\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_N]^\mathsf{T} \in \mathbb{R}^{N \times Q}$および$\boldsymbol{\Theta} = \{\boldsymbol{\theta}^1, \boldsymbol{\theta}^2, \ldots, \boldsymbol{\theta}^M\}$であり，$\boldsymbol{\theta}^m$はカーネル関数のハイパーパラメータお

よびノイズの大きさに関するハイパーパラメータである．mGPLVM は，以下の確率過程により，潜在変数から観測値が潜在関数 f_d^m $(d = 1, ..., D_m)$ を用いて生成されると仮定する．

$$y_{nd}^m = f_d^m(\boldsymbol{x}_n) + \epsilon_{nd}^m,\ \epsilon_{nd}^m \sim \mathcal{N}(0, \sigma_m^2) \tag{3.7}$$

ただし，y_{nd}^m は $\boldsymbol{Y}^m$ の (n, d) 要素である．さらに，全ての潜在関数 $\{f_d^m\}_{d=1}^{D_m}$ は以下の同時確率分布に従うと仮定する．

$$\boldsymbol{f}_d^m \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{K}_{\boldsymbol{XX}}^m) \tag{3.8}$$

ただし，$\boldsymbol{f}_d^m = [f_d^m(\boldsymbol{x}_1), f_d^m(\boldsymbol{x}_2), \ldots, f_d^m(\boldsymbol{x}_N)]^\mathsf{T} \in \mathbb{R}^N$ であり，$\boldsymbol{K}_{\boldsymbol{XX}}^m$ は潜在変数の共分散行列である．ここで，$\boldsymbol{K}_{\boldsymbol{XX}}^m$ の (i, j) 要素 k_{ij}^m はカーネル関数 k^m を用いて次式で算出される．

$$k_{ij}^m = k^m(\boldsymbol{x}_i, \boldsymbol{x}_j) \tag{3.9}$$

提案手法では．高次元のデータに有効である Radial Basis Function（RBF）カーネルを用いた [59].

Maximum a Posteriori (MAP) 推定 [60] では，潜在変数の事後確率分布 $p(\boldsymbol{X}|\boldsymbol{Y}^1, \boldsymbol{Y}^2, \ldots, \boldsymbol{Y}^M, \boldsymbol{\Theta})$ の最大化により，潜在変数およびハイパーパラメータを最適化する．ここで，ベイズの定理によると，

$$p(\boldsymbol{X}|\boldsymbol{Y}^1, \boldsymbol{Y}^2, \ldots, \boldsymbol{Y}^M, \boldsymbol{\Theta}) \propto p(\boldsymbol{Y}^1, \boldsymbol{Y}^2, \ldots, \boldsymbol{Y}^M|\boldsymbol{X}, \boldsymbol{\Theta})p(\boldsymbol{X}) \tag{3.10}$$

である．ただし，$p(\boldsymbol{Y}^1, \boldsymbol{Y}^2, \ldots, \boldsymbol{Y}^M|\boldsymbol{X}, \boldsymbol{\Theta})$ は $\{\boldsymbol{Y}^m\}_{m=1}^M$ の結合周辺尤度であり，$p(\boldsymbol{X})$ は $\boldsymbol{X}$ の事前確率分布である．したがって，以下の最適化問題を得る．

$$\arg\max_{\boldsymbol{X}, \boldsymbol{\Theta}}\ p(\boldsymbol{Y}^1, \boldsymbol{Y}^2, \ldots, \boldsymbol{Y}^M|\boldsymbol{X}, \boldsymbol{\Theta})p(\boldsymbol{X}) \tag{3.11}$$

ただし，

$$p(\boldsymbol{Y}^1, \boldsymbol{Y}^2, \ldots, \boldsymbol{Y}^M | \boldsymbol{X}, \boldsymbol{\Theta}) = \prod_{m=1}^{M} p(\boldsymbol{Y}^m | \boldsymbol{X}, \boldsymbol{\theta}^m) \tag{3.12}$$

$$\begin{aligned} p(\boldsymbol{Y}^m | \boldsymbol{X}, \boldsymbol{\theta}^m) &= \prod_{d=1}^{D_m} \frac{\exp(-\frac{1}{2}(\boldsymbol{y}^m_{:,d})^{\mathsf{T}}(\tilde{\boldsymbol{K}}^m_{\boldsymbol{XX}})^{-1}\boldsymbol{y}^m_{:,d})}{(2\pi)^{\frac{N}{2}}|\tilde{\boldsymbol{K}}^m_{\boldsymbol{XX}}|^{\frac{1}{2}}} \\ &= \frac{1}{(2\pi)^{\frac{ND_m}{2}}|\tilde{\boldsymbol{K}}^m_{\boldsymbol{XX}}|^{\frac{D_m}{2}}} \exp\left(-\frac{1}{2}\mathrm{tr}(\tilde{\boldsymbol{K}}^m_{\boldsymbol{XX}}{}^{-1}\boldsymbol{Y}\boldsymbol{Y}^{\mathsf{T}})\right) \end{aligned} \tag{3.13}$$

であり，$\tilde{\boldsymbol{K}}^m_{\boldsymbol{XX}} = \boldsymbol{K}^m_{\boldsymbol{XX}} + \sigma_m^2 \boldsymbol{I}_N$（$\boldsymbol{I}_N$ は，N 次元の単位行列）および $\boldsymbol{y}^m_{:,d}$ は $\boldsymbol{Y}^m$ の d 番目の列ベクトルである．

一般的には，$\boldsymbol{x}_n$ は多変量ガウス分布に従うと仮定し，$p(\boldsymbol{X})$ を次式で定義する．

$$p(\boldsymbol{X}) = \prod_{n=1}^{N} p(\boldsymbol{x}_n),\ p(\boldsymbol{x}_n) = \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_Q) \tag{3.14}$$

目的に合わせた $p(\boldsymbol{X})$ を選択可能であることが，mGPLVM の柔軟性を向上させている [61,62].

3.3.3 提案手法

本章では，提案手法である Semi-OMGP について説明する．まず，コンテンツから得られるコンテンツ特徴 $\boldsymbol{Y}^{\mathrm{c}}$ およびユーザの動作から得られる動作特徴 $\boldsymbol{Y}^{\mathrm{b}}$ を次式で定義する．

$$\boldsymbol{Y}^{\mathrm{c}} = [\boldsymbol{y}^{\mathrm{c}}_1, \boldsymbol{y}^{\mathrm{c}}_2, \ldots, \boldsymbol{y}^{\mathrm{c}}_N]^{\mathsf{T}} \in \mathbb{R}^{N \times D_{\mathrm{c}}} \tag{3.15}$$

$$\boldsymbol{Y}^{\mathrm{b}} = [\boldsymbol{y}^{\mathrm{b}}_1, \boldsymbol{y}^{\mathrm{b}}_2, \ldots, \boldsymbol{y}^{\mathrm{b}}_N]^{\mathsf{T}} \in \mathbb{R}^{N \times D_{\mathrm{b}}} \tag{3.16}$$

ただし，$N = U \times C$ であり，U はユーザの総数および C はコンテンツの総数である．コンテンツに付与されたラベル $\boldsymbol{l} = [l_1, l_2, \ldots, l_{\hat{N}}]^{\mathsf{T}} \in \mathbb{R}^{\hat{N}}$ はユーザのコンテンツに対する関心度であり，$\hat{N}\ (< N)$ はラベルを付与されたコンテンツの総数である．ただし，$l_n \in \{1, 2, \ldots, L_{\max} : L_{\max}$ は関心度の段階の数$\}$ であり，関心度には順序性がある．提案手法で用いる特徴量についての詳細は 3.3.4 節で述べる．

提案手法では，観測値の類似度行列が潜在変数から生成されると仮定する．各

モダリティの類似度行列を以下で定義する．

$$s_{pq}^{m} = \exp\left(-\frac{\|\boldsymbol{y}_p^m - \boldsymbol{y}_q^m\|_2^2}{2\gamma_m}\right) \tag{3.17}$$

ただし，$m \in \{\mathrm{c}, \mathrm{b}\}$ および s_{pq}^{m} は $\boldsymbol{S}^m$ の (p, q) 要素である．また，$\gamma_m > 0$ は帯域幅パラメータである．Semi-OMGP は，以下の確率過程により，潜在変数から観測値の類似度行列が潜在関数 f_q^m $(q = 1, ..., N)$ を用いて生成されると仮定する．

$$s_{pq}^{m} = f_q^m(\boldsymbol{x}_p) + \epsilon_{pq}^{m},\ \epsilon_{pq}^{m} \sim \mathcal{N}(0, \sigma_m^2) \tag{3.18}$$

Semi-OMGP では，全ての観測値を類似度行列に変換することで，マルチモーダルデータの分布の違いに頑健な潜在変数の算出が可能になる．これが，提案手法の **貢献 (i)** に対応する．

mGPLVM と同様にして，次の結合周辺尤度を得る．

$$\begin{aligned} p(\boldsymbol{S}^{\mathrm{c}}, \boldsymbol{S}^{\mathrm{b}} | \boldsymbol{X}, \boldsymbol{\Theta}) &= p(\boldsymbol{S}^{\mathrm{c}} | \boldsymbol{X}, \boldsymbol{\theta}^{\mathrm{c}}) p(\boldsymbol{S}^{\mathrm{b}} | \boldsymbol{X}, \boldsymbol{\theta}^{\mathrm{b}}) \\ &= \prod_m \frac{1}{(2\pi)^{\frac{ND_m}{2}} |\tilde{\boldsymbol{K}}_{\boldsymbol{XX}}^m|^{\frac{D_m}{2}}} \exp\left(-\frac{1}{2}\mathrm{tr}(\tilde{\boldsymbol{K}}_{\boldsymbol{XX}}^{m\,-1} \boldsymbol{S}^m (\boldsymbol{S}^m)^{\mathsf{T}})\right) \end{aligned} \tag{3.19}$$

また，MAP 推定に基づき，次の最適化問題を解く．

$$\begin{aligned} \arg\max_{\boldsymbol{X}, \boldsymbol{\Theta}}\ L &= \arg\max_{\boldsymbol{X}, \boldsymbol{\Theta}}\ \log p(\boldsymbol{X} | \boldsymbol{S}^{\mathrm{c}}, \boldsymbol{S}^{\mathrm{b}}, \boldsymbol{\Theta}) \\ &= \arg\max_{\boldsymbol{X}, \boldsymbol{\Theta}} \{\log p(\boldsymbol{S}^{\mathrm{c}}, \boldsymbol{S}^{\mathrm{b}} | \boldsymbol{X}, \boldsymbol{\Theta}) + \log p(\boldsymbol{X})\} \\ &= \arg\max_{\boldsymbol{X}, \boldsymbol{\Theta}} \{\log p(\boldsymbol{S}^{\mathrm{c}} | \boldsymbol{X}, \boldsymbol{\theta}^{\mathrm{c}}) + \log p(\boldsymbol{S}^{\mathrm{b}} | \boldsymbol{X}, \boldsymbol{\theta}^{\mathrm{b}}) + \log p(\boldsymbol{X})\} \end{aligned} \tag{3.20}$$

Semi-OMGP では，ラベルが一部のコンテンツのみに付与されている状況およびラベルの順序性を考慮し，潜在変数に制約を設けるため，事前確率分布 $p(\boldsymbol{X})$ を以下で定義する．

$$p(\boldsymbol{X}) = \frac{1}{Z} \exp\left(-\sum_{i,j=1}^{N} w_{ij} \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2\right) \tag{3.21}$$

$$w_{ij} = \begin{cases} \alpha(\Delta - |l_i - l_j|)\dfrac{e^t}{1 + e^t} & (l_i \text{ および } l_j \text{ が存在する場合}) \\ 0 & (\text{その他の場合}) \end{cases} \tag{3.22}$$

ただし，Z は定数，$t = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2$，$\alpha$ はパラメータおよび $\Delta\left(= \frac{(L_{\max} - 1)}{2}\right)$ は $|l_i - l_j|$ の平

均である．重み w_{ij} により，ラベルが一部のコンテンツのみに付与されている状況およびラベルの順序性を考慮を可能としている．式（3.22）では，半教師あり学習の手法である文献 [63] およびラベルの順序性を扱う手法である文献 [64] を組み合わせることにより，w_{ij} を導出している．l_i および l_j が得られている場合には，w_{ij} はラベルの順序性を反映する．具体的には，$|l_i - l_j|$ が小さくなるほど，w_{ij} は大きくなり，$\boldsymbol{x}_i$ および $\boldsymbol{x}_j$ は潜在空間上で互いに近づくように制約を受ける．それとは逆に，$|l_i - l_j|$ が大きくなるほど，w_{ij} は小さくなり，$\boldsymbol{x}_i$ および $\boldsymbol{x}_j$ は潜在空間上で互いに遠ざかるように制約を受ける．一方，l_i または l_j が得られていない場合には，w_{ij} は 0 になり，$\boldsymbol{x}_i$ および $\boldsymbol{x}_j$ は制約を受けない．式（3.21）および（3.22）において，一部のコンテンツに付与されたラベルおよびそのラベルの順序性を反映した事前確率分布 $p(\boldsymbol{X})$ を定義したことにより，ラベル情報を十分に利用した潜在変数の算出が可能になる．これが，提案手法の**貢献 (ii)** に対応する．さらに，式（3.21）は以下のように変形可能である．

$$\begin{aligned} p(\boldsymbol{X}) &= \frac{1}{Z}\exp\left(-\sum_{i,j=1}^{N} w_{ij}\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2\right) \\ &= \frac{1}{Z}\exp\left(-\mathrm{tr}(\boldsymbol{X}^\mathsf{T}\boldsymbol{W}\boldsymbol{X})\right) \\ &= \frac{1}{Z}\exp\left(-\mathrm{tr}(\boldsymbol{W}\boldsymbol{X}\boldsymbol{X}^\mathsf{T})\right) \end{aligned} \tag{3.23}$$

ただし，w_{ij} は $\boldsymbol{W}$ の各要素である．式（3.18）による仮定および式（3.23）による $p(\boldsymbol{X})$ の設定が，Semi-OMGP の具体的な新規点である．

式（3.20）の 1 番目，2 番目および 3 番目の項をそれぞれ $L_\mathrm{c}, L_\mathrm{b}$ および L_prior とすると，これらはそれぞれ以下のように変形可能である．

$$\begin{aligned} L_\mathrm{c} &= \log p(\boldsymbol{S}^\mathrm{c}|\boldsymbol{X}, \boldsymbol{\theta}^\mathrm{c}) \\ &= -\frac{D_\mathrm{c}N}{2}\log(2\pi) - \frac{D_\mathrm{c}}{2}\log|\tilde{\boldsymbol{K}}^\mathrm{c}_{\boldsymbol{X}\boldsymbol{X}}| - \frac{1}{2}\mathrm{tr}(\tilde{\boldsymbol{K}}^\mathrm{c}_{\boldsymbol{X}\boldsymbol{X}}{}^{-1}\boldsymbol{S}^\mathrm{c}(\boldsymbol{S}^\mathrm{c})^\mathsf{T}) \end{aligned} \tag{3.24}$$

$$\begin{aligned} L_\mathrm{b} &= \log p(\boldsymbol{S}^\mathrm{b}|\boldsymbol{X}, \boldsymbol{\theta}^\mathrm{b}) \\ &= -\frac{D_\mathrm{b}N}{2}\log(2\pi) - \frac{D_\mathrm{b}}{2}\log|\tilde{\boldsymbol{K}}^\mathrm{b}_{\boldsymbol{X}\boldsymbol{X}}| - \frac{1}{2}\mathrm{tr}(\tilde{\boldsymbol{K}}^\mathrm{b}_{\boldsymbol{X}\boldsymbol{X}}{}^{-1}\boldsymbol{S}^\mathrm{b}(\boldsymbol{S}^\mathrm{b})^\mathsf{T}) \end{aligned} \tag{3.25}$$

$$L_\mathrm{prior} = -\mathrm{tr}(\boldsymbol{W}\boldsymbol{X}\boldsymbol{X}^\mathsf{T}) - \log Z \tag{3.26}$$

以上により，潜在変数の対数事後確率分布 L は以下の式で示される．

$$
\begin{aligned}
L &= L_{\mathrm{c}} + L_{\mathrm{b}} + L_{\mathrm{prior}} \\
&= -\frac{D_{\mathrm{c}}}{2}\log|\tilde{\boldsymbol{K}}^{\mathrm{c}}_{\boldsymbol{XX}}| - \frac{D_{\mathrm{b}}}{2}\log|\tilde{\boldsymbol{K}}^{\mathrm{b}}_{\boldsymbol{XX}}| - \frac{1}{2}\mathrm{tr}(\tilde{\boldsymbol{K}}^{\mathrm{c}}_{\boldsymbol{XX}}{}^{-1}\boldsymbol{S}^{\mathrm{c}}(\boldsymbol{S}^{\mathrm{c}})^{\mathsf{T}} + \tilde{\boldsymbol{K}}^{\mathrm{b}}_{\boldsymbol{XX}}{}^{-1}\boldsymbol{S}^{\mathrm{b}}(\boldsymbol{S}^{\mathrm{b}})^{\mathsf{T}} + 2\boldsymbol{W}\boldsymbol{X}\boldsymbol{X}^{\mathsf{T}}) + \mathrm{constant}
\end{aligned}
\tag{3.27}
$$

式（3.27）を解くことにより，最適な潜在変数およびハイパーパラメータを算出可能である．提案手法では，従来の GPLVM に基づく手法 [62, 65] と同様に，Scaled Conjugate Gradient（SCG）法 [66] を用いて式（3.27）を解く．

3.3.4 実験

実験方法

本実験では，YouTube[2]から取得した 49 本の映画の予告編（コメディ，アクション，科学，音楽のジャンルから各 10 本，スポーツのジャンルから 9 本）をコンテンツとして使用した．これらの映像の各フレームを入力し，Inception-v3 [48] の中間層から 2,048 次元の出力を算出した．次に，出力の平均ベクトルをコンテンツ特徴量 $\boldsymbol{y}^{\mathrm{c}}_n$ として，[43] と同様に計算した．この実験では，実験参加者は１）１本の動画を 30 秒間視聴し，２）その動画を，１（全く面白くない），２（面白くない），３（少し面白い），４（とても面白い）の４つのクラスに分けて評価し，映像にラベルを付与した．そして，すべての映像を視聴するまで１）および２）を繰り返した．なお，被験者は女性 2 名および男性 8 名で，年齢は平均 22 歳であった．

本実験では，映像視聴時に OpenPose [47] と Tobii Eye Tracker 4C を用いて動作についての情報を取得した．OpenPose は，2 次元の体骨格の位置を推定する最も一般的で最新の手法の一つである．これらの位置は，身体の部位に基づいてディープニューラルネットワークによって推定することができる [47]．Tobii Eye Tracker 4C では，2 次元のユーザの視線位置を検出することが可能であり，ユーザの視線は，ユーザの興味関心と相関があることが知られている [67]．OpenPose で得られた身体の骨格の情報と Tobii Eye Tracker 4C で得られた視線情報からユーザの動作特徴を抽出した．具体的には，[43] と同様に，それらの位置の平均と分散を計算することで，ユーザの動作特徴 $\boldsymbol{y}^{\mathrm{b}}_n$ を得た．

[2]https://www.youtube.com

表 3.1: 欠損率が 50%の場合の各実験参加者の MAE の平均

	Semi-OMGP	mGPLVM [45]	m-SimGP [52]	MVCCA [53]	Deep CCA [54]	Bayesian CCA [55]
実験参加者 1	0.756	1.008	**0.668**	0.760	1.380	0.780
実験参加者 2	**0.952**	1.136	1.200	1.132	1.800	1.208
実験参加者 3	0.676	1.012	0.696	0.816	1.736	**0.576**
実験参加者 4	**0.740**	0.876	0.812	**0.740**	0.920	0.848
実験参加者 5	**0.944**	1.024	1.252	**0.944**	1.436	1.144
実験参加者 6	**0.560**	1.020	0.656	1.372	1.520	0.608
実験参加者 7	**0.988**	1.004	1.176	1.256	1.460	1.056
実験参加者 8	**0.756**	0.896	0.888	1.008	1.404	0.772
実験参加者 9	**0.796**	1.052	1.248	0.968	1.320	0.876
実験参加者 10	0.788	0.820	0.848	1.260	1.396	**0.744**
Mean ±SD	**0.796** ±0.1264	0.984 ±0.0888	0.944 ±0.2356	1.024 ±0.2112	1.436 ±0.2276	0.860 ±0.2028

本実験では，全映像の中から 50% の映像をランダムに選択し，未視聴の映像とした．提案した Semi-OMGP を以下の mGPLVM [45], m-SimGP [52], Multiview CCA (MVCCA) [53], Deep CCA [54] および Bayesian CCA [55] の 5 つの手法と比較した．mGPLVM は GPLVM 系列のベースラインとなる手法であり，m-SimGP は mGPLVM に基づいた最先端の手法である．さらに，CCA は基本的な特徴量統合の一つであるため，3 種類の CCA に基づく手法を採用した．特に，多層パーセプトロンと相関最大化を組み合わせて構築する Deep CCA は，近年いくつかのタスクに利用されている．Bayesian CCA は確率的なアプローチであり，ノイズに対して頑健な潜在空間を算出する．なお，提案手法では，Q, α および γ_m のパラメータは，それぞれ 4, 1.0×10^5 および 2 に設定されている．

Semi-OMGP と比較手法の性能を比較するために，従来手法で用いられたテンソル補完 [43] を，すべての手法で得られた統合後の特徴量と関心度に適用した．予測された関心度の評価には，以下で定義される MAE を用いた．

$$\mathrm{MAE} = \frac{1}{N - \hat{N}} \sum_{s=1}^{N-\hat{N}} |l_s^{\mathrm{PRE}} - l_s^{\mathrm{GT}}| \tag{3.28}$$

ただし，l_s^{PRE} および l_s^{GT} は，s 番目のサンプルの予測された関心度および真値である．

実験結果と考察

表 3.1 に，未視聴映像の割合を 50 % とした場合の各実験参加者の平均 MAE を示す．m-SimGP と mGPLVM を比較することで，共通潜在空間から各観測地の類似度行列が生成されていると仮定する **貢献 (i)** の有効性を確認することができる．Semi-OMGP と m-SimGP を比較することで，**貢献 (ii)** の有効性を確認することができる．さらに，Semi-OMGP は MVCCA や Deep CCA よりも高精度であるため，多視点アプローチや深層学習ベースのアプローチよりも効果的な潜在空間を構築することが可能である． また，確率的生成モデルの一つである Bayesian CCA の性能は，mGPLVM や m-SimGP よりも優れているが，Semi-OMGP は Bayesian CCA よりも優れている．このように，本章で提案したガウス過程潜在変数モデルに基づくアプローチの有効性を確認することができた．最後に，Semi-OMGP の推定性能は mGPLVM の推定性能よりも高いことから，**貢献 (i), (ii)** の共同利用が精度向上に貢献していることがわかる．以上の実験結果から，提案する Semi-OMGP は，コンテンツ特徴と動作

特徴の適切な特徴統合が可能であることが示された．

3.3.5 まとめ

本章では，新しい特徴統合手法である Semi-OMGP を提案した．本章では，共通の潜在変数から類似度行列を生成し，潜在変数の事前分布にラベルのないサンプルとラベルの順序性を考慮する．その結果，実験結果は，提案した Semi-OMGP が関心度推定に適した潜在変数を算出可能であることを示した．

3.4 潜在特徴に与える制約による特徴統合空間の変化に関する研究

3.4.1 はじめに

第 3.3 節では，Semi-OMGP により，ラベルの順序性を考慮した制約を潜在空間に与えることで，関心度推定の精度が向上することを示した．Semi-OMGP では，関心度が付与されたコンテンツ対（ペア）に対してのみ，そのラベル差に応じた距離制約を課している．一方で，関心度が付与されていないコンテンツ（ラベル無しデータ）に対しては，明確な制約を与えていない．しかし，半教師あり学習の分野では，ラベル無しデータに対しても一定の制約（正則化など）を与えることで，潜在空間の構造が安定し，タスクの精度が向上することが報告されている [63].

そこで本節では，Semi-OMGP の拡張として，ラベルが付与されていないコンテンツの潜在特徴に対しても一定の制約を与えた場合の影響を検証する．具体的には，ラベル無しデータ間に対しても潜在空間上での凝集性を促すような制約項を導入し，これが関心度推定精度にどのような変化をもたらすかを実験的に明らかにする．

3.4.2 提案手法

本節では，提案手法について説明する．コンテンツから得られるコンテンツ特徴 $\boldsymbol{Y}^{\mathrm{c}}$ およびユーザの動作から得られる動作特徴 $\boldsymbol{Y}^{\mathrm{b}}$ を次式で定義する．

$$\boldsymbol{Y}^{\mathrm{c}} = [\boldsymbol{y}_1^{\mathrm{c}}, \boldsymbol{y}_2^{\mathrm{c}}, \ldots, \boldsymbol{y}_N^{\mathrm{c}}]^{\mathsf{T}} \in \mathbb{R}^{N \times D_{\mathrm{c}}} \tag{3.29}$$

$$\boldsymbol{Y}^{\mathrm{b}} = [\boldsymbol{y}_1^{\mathrm{b}}, \boldsymbol{y}_2^{\mathrm{b}}, \ldots, \boldsymbol{y}_N^{\mathrm{b}}]^{\mathsf{T}} \in \mathbb{R}^{N \times D_{\mathrm{b}}} \tag{3.30}$$

ただし，$N = U \times C$ であり，U はユーザの総数および C はコンテンツの総数である．コンテンツに付与されたラベル $\boldsymbol{l} = [l_1, l_2, \ldots, l_{\hat{N}}]^{\mathsf{T}} \in \mathbb{R}^{\hat{N}}$ はユーザのコンテンツに対する関心度であり，$\hat{N}\,(< N)$ はラベルを付与されたコンテンツの総数である．ただし，$l_n \in \{1, 2, \ldots, L_{\max} : L_{\max}$ は関心度の段階の数$\}$ である．

提案手法では，各特徴の類似度行列が潜在特徴 $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_N]^{\mathsf{T}} \in \mathbb{R}^{N \times Q}$（$Q$ は潜在特徴の次元数）から生成されると仮定する [52]. 各特徴の類似度行列を以下で

定義する．

$$s_{pq}^{m} = \exp\left(-\frac{\|\boldsymbol{y}_p^m - \boldsymbol{y}_q^m\|_2^2}{2\gamma_m}\right) \tag{3.31}$$

ただし，$m \in \{\mathrm{c}, \mathrm{b}\}$ および s_{pq}^{m} は $\boldsymbol{S}^m$ の (p,q) 要素である．また，$\gamma_m > 0$ は帯域幅パラメータである．提案手法は，以下の確率過程により，潜在特徴から観測値の類似度行列が潜在関数 f_q^m $(q = 1, ..., N)$ を用いて生成されると仮定する．

$$s_{pq}^{m} = f_q^m(\boldsymbol{x}_p) + \epsilon_{pq}^{m},\ \epsilon_{pq}^{m} \sim \mathcal{N}(0, \sigma_m^2) \tag{3.32}$$

mGPLVM [45] と同様にして，次の結合周辺尤度を得る．

$$\begin{aligned} p(\boldsymbol{S}^{\mathrm{c}}, \boldsymbol{S}^{\mathrm{b}} | \boldsymbol{X}, \boldsymbol{\Theta}) &= p(\boldsymbol{S}^{\mathrm{c}} | \boldsymbol{X}, \boldsymbol{\theta}^{\mathrm{c}}) p(\boldsymbol{S}^{\mathrm{b}} | \boldsymbol{X}, \boldsymbol{\theta}^{\mathrm{b}}) \\ &= \frac{1}{(2\pi)^{\frac{ND_{\mathrm{c}}}{2}} |\tilde{\boldsymbol{K}}_{\boldsymbol{XX}}^{\mathrm{c}}|^{\frac{D_{\mathrm{c}}}{2}}} \exp\left(-\frac{1}{2}\mathrm{tr}(\tilde{\boldsymbol{K}}_{\boldsymbol{XX}}^{\mathrm{c}\,-1} \boldsymbol{S}^{\mathrm{c}} (\boldsymbol{S}^{\mathrm{c}})^{\mathsf{T}})\right) \\ &\quad \frac{1}{(2\pi)^{\frac{ND_{\mathrm{b}}}{2}} |\tilde{\boldsymbol{K}}_{\boldsymbol{XX}}^{\mathrm{b}}|^{\frac{D_{\mathrm{b}}}{2}}} \exp\left(-\frac{1}{2}\mathrm{tr}(\tilde{\boldsymbol{K}}_{\boldsymbol{XX}}^{\mathrm{b}\,-1} \boldsymbol{S}^{\mathrm{b}} (\boldsymbol{S}^{\mathrm{b}})^{\mathsf{T}})\right) \end{aligned} \tag{3.33}$$

ただし，$\tilde{\boldsymbol{K}}_{\boldsymbol{XX}}^{\mathrm{c}}$ は共分散行列であり，RBF kernel [59] 関数に基づき算出される．また，θ^m は，RBF kernel 関数に含まれるパラメータである．

提案手法では，両方に関心度が付与された潜在特徴のペアには Semi-OMGP と同様の制約を与え，それ以外の潜在特徴のペアには一定の制約を与えるため，事前確率分布 $p(\boldsymbol{X})$ を以下で定義する．

$$p(\boldsymbol{X}) = \frac{1}{Z} \exp\left(-\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij} \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2\right) \tag{3.34}$$

$$w_{ij} = \begin{cases} \alpha(\Delta - |l_i - l_j|)\dfrac{e^t}{1+e^t} & (l_i \text{ および } l_j \text{ が存在}) \\ \dfrac{e^t}{N^2(1+e^t)} & (\text{上記以外の場合}) \end{cases} \tag{3.35}$$

ただし，Z は定数，$t = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2$，$\alpha$ はパラメータおよび $\Delta\left(= \frac{(L_{\max}-1)}{2}\right)$ は $|l_i - l_j|$ の平均

表 3.2: 実験で取得した動作特徴の詳細

	特徴の種類	次元
Eye Tracker	2 次元空間での視線の動きの平均と分散	4
OpenPose	2 次元空間での首，鼻緒および腰の動きの平均と分散	12
	2 次元空間での両方の耳，目，肩，手首，肘および尻の動きの平均と分散	48
Total	**上記を連結した特徴量**	**64**

表 3.3: 関心度が付与された映像の割合が 50% のときの MAE の平均

特徴統合手法	MAE の平均
提案手法	**0.759**
Semi-OMGP	0.796

である．さらに，式（3.34）を以下のように変形する．

$$\begin{aligned} p(\boldsymbol{X}) &= \frac{1}{Z}\exp\left(-\sum_{i,j=1}^{N} w_{ij}\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2\right) \\ &= \frac{1}{Z}\exp\left(-\mathrm{tr}(\boldsymbol{X}^{\mathsf{T}}\boldsymbol{W}\boldsymbol{X})\right) \\ &= \frac{1}{Z}\exp\left(-\mathrm{tr}(\boldsymbol{W}\boldsymbol{X}\boldsymbol{X}^{\mathsf{T}})\right), \end{aligned} \tag{3.36}$$

ただし，w_{ij} は $\boldsymbol{W}$ の各要素である．次式により潜在特徴の事後確率を最大化することで，潜在特徴 $\boldsymbol{X}$ を導出する．

$$\arg\max_{\boldsymbol{X},\boldsymbol{\Theta}} \; p(\boldsymbol{S}^{\mathrm{c}},\boldsymbol{S}^{\mathrm{b}}|\boldsymbol{X},\boldsymbol{\Theta})p(\boldsymbol{X}) \tag{3.37}$$

3.4.3 実験

本節では，潜在特徴に与える制約による関心度推定の精度変化を検証するための実験を行う．具体的に，コンテンツ特徴および動作特徴を提案手法または比較手法により統合し，得られた潜在特徴に対してテンソル補完を適用し，未知の関心度を推定する実験を行う [43].

実験設定

本実験では，データセットとして，YouTube から，映画の予告編（30 秒）を ‘Action’, ‘Comedy’, ‘Music’, ‘Sport’ および ‘Science’ のジャンルから各 10 本の計 50 本取得した．

各映像のコンテンツ特徴を，文献 [43] と同様に算出した．本実験では，10 名の実験参加者は一つの映像を 30 秒間視聴した後に，その映像に 4 段階の関心度 (1: 全く興味がない，2: あまり興味がない，3: 少し興味がある，4: とても興味がある) を 5 秒間で付与した．実験参加者は，以上のタスクを全ての映像に対して行った．映像視聴中の実験参加者の身体の位置情報および視線情報を，それぞれ OpenPose [47] および Tobii Eye Tracker 4C を用いて取得し，各ユーザの各映像に対する動作特徴を，文献 [43] と同様に算出した．算出した動作特徴の詳細を，表 3.2 に示す．特徴量の次元数は，$D^{\mathrm{c}} = 2,048$ および $D^{\mathrm{b}} = 64$ であった．本実験では，映像に対して付与された関心度の 50% をランダムに欠損させ，未知の関心度とした．比較手法として，Semi-OMGP を用い，評価指標として，以下で定義される MAE を用いた．

$$\mathrm{MAE} = \frac{1}{\underline{N}} \sum_{s=1}^{\underline{N}} |l_s^{\mathrm{PRE}} - l_s^{\mathrm{GT}}|, \tag{3.38}$$

ただし，$\underline{N} = N - \hat{N}$，$l_s^{\mathrm{PRE}}$ および l_s^{GT} は，s 番目の映像の関心度の推定値および真値である．本実験は各手法について 10 回行い，全ての試行の MAE の平均を算出する．なお，提案手法および比較手法において，$\gamma_m = 2.0$ とした．

実験結果と考察

表 3.3 に，各特徴統合手法における MAE の平均を示す．提案手法および比較手法の MAE の平均を比較することにより，関心度が付与されていないコンテンツの潜在特徴に対して，式（3.35）に示される事前分布の定義に基づき，一定の制約を与えることが，関心度推定の精度向上に寄与することが確認された．この結果は，似たデータは近傍に存在しているべきという多様体仮説に裏付けられていると推察できる．

3.4.4 まとめ

本研究では，潜在特徴に与える制約による関心度推定の精度変化の検証を行った．実験により，関心度が付与されていないコンテンツの潜在特徴に一定の制約を与えることが関心度推定の精度向上に寄与することが確認された．

3.5 逆射影を仮定したmGPLVMに基づく特徴統合

3.5.1 はじめに

第3.4節までに述べた手法は，学習時および推論時の双方において，コンテンツ特徴と動作特徴の両方が揃っていることを前提とするものであった．しかしながら，実用的な推薦システムにおいては，ユーザがまだ視聴していない「未視聴コンテンツ」に対する関心度を予測する必要がある．未視聴コンテンツにおいては，当然ながらユーザの視聴動作およびそれから算出される動作特徴は取得できないため，コンテンツ特徴のみから，動作特徴を反映した潜在変数を推定する技術が不可欠となる．

そこで本節では，この課題を解決するために **Back-projection Ordering Multi-modal Gaussian Process Latent Variable Model (BPomGP)** を提案する．本手法では，特徴統合と同時に，観測空間（コンテンツ特徴）から潜在空間への写像である逆射影 [68] を学習する．これにより，動作特徴が得られない未視聴コンテンツに対しても，学習済みの逆射影を用いることで，動作特徴との相関を考慮した適切な潜在変数を算出することが可能となる．また，第3.3節と同様に，関心度の順序性を考慮した事前分布を導入することで，高精度な関心度推定を実現する．

3.5.2 提案手法

提案手法は，コンテンツ情報から得られるコンテンツ特徴，ユーザの動作情報から得られる動作特徴をモダリティとして用いる．最初に，コンテンツ特徴を $\boldsymbol{Y}^{\mathrm{c}} = [\boldsymbol{y}_1^{\mathrm{c}}, \boldsymbol{y}_2^{\mathrm{c}}, \ldots, \boldsymbol{y}_N^{\mathrm{c}}]^\top \in \mathbb{R}^{N \times D^{\mathrm{c}}}$，ユーザの動作特徴を $\boldsymbol{Y}^{\mathrm{b}} = [\boldsymbol{y}_1^{\mathrm{b}}, \boldsymbol{y}_2^{\mathrm{b}}, \ldots, \boldsymbol{y}_N^{\mathrm{b}}]^\top \in \mathbb{R}^{N \times D^{\mathrm{b}}}$ とする．ただし，N はコンテンツ数，D^{c} はコンテンツ特徴の次元数および D^{b} は動作情報の次元数である．次に，潜在特徴を $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N]^\top \in \mathbb{R}^{N \times D^{\mathrm{x}}} < \min(D^{\mathrm{b}}, D^{\mathrm{c}})$ と定義する．ただし，D^{x} は潜在特徴の次元数である．提案手法は，潜在特徴 $\boldsymbol{X}$ およびパラメータ $\boldsymbol{\Theta}, \boldsymbol{\Gamma}$ を算出する．ただし，$\boldsymbol{\Theta} = [\boldsymbol{\theta}^{\mathrm{c}}, \boldsymbol{\theta}^{\mathrm{b}}]$ は潜在特徴を各観測値に変換する射影を決定するパラメータ，$\boldsymbol{\Gamma} = [\boldsymbol{\gamma}^{\mathrm{c}}, \boldsymbol{\gamma}^{\mathrm{b}}]$ は各観測値を潜在特徴に変換する逆射影を決定するパラメータである．

提案手法は，mGPLVMと同様に，潜在特徴から各モダリティの観測値がガウス過程に従い生成されると仮定する．$\boldsymbol{Y}^m (m \in \{\mathrm{c}, \mathrm{b}\})$ の尤度 $p(\boldsymbol{Y}^m | \boldsymbol{X}, \boldsymbol{\theta}^m)$ は次式で算出さ

れる．

$$
\begin{aligned}
&p(\boldsymbol{Y}^m|\boldsymbol{X},\boldsymbol{\theta}^m) \\
&= \prod_{d=1}^{D^m} \frac{\exp(-\frac{1}{2}(\boldsymbol{y}^m_{:,d})^\top(\boldsymbol{K}^m_{\boldsymbol{XX}})^{-1}\boldsymbol{y}^m_{:,d})}{(2\pi)^{\frac{N}{2}}|\boldsymbol{K}^m_{\boldsymbol{XX}}|^{\frac{1}{2}}} \\
&= \frac{1}{(2\pi)^{\frac{ND^m}{2}}|\boldsymbol{K}^m_{\boldsymbol{XX}}|^{\frac{D^m}{2}}} \exp\left(-\frac{1}{2}\mathrm{tr}(\boldsymbol{K}^m_{\boldsymbol{XX}}{}^{-1}\boldsymbol{Y}^m\boldsymbol{Y}^{m\top})\right)
\end{aligned} \tag{3.39}
$$

ただし，$\boldsymbol{K}^m_{\boldsymbol{XX}} \in \mathbb{R}^{N\times N}$ は $\boldsymbol{X}$ の共分散行列であり，(i,j) 成分 k^m_{ij} は以下のカーネル関数に基づき算出される．

$$
k^m_{ij} = k^m(\boldsymbol{x}_i, \boldsymbol{x}_j, \theta^m) \tag{3.40}
$$

提案手法は，文献 [68] に従い，次式により逆射影を定義する．

$$
\begin{aligned}
p(\boldsymbol{X}|\boldsymbol{Y}^{\mathrm{c}},\boldsymbol{Y}^{\mathrm{b}},\boldsymbol{\Gamma}) &= \frac{1}{(2\pi)^{\frac{ND^{\mathrm{x}}}{2}}|\boldsymbol{K}_{\boldsymbol{Y}}|^{\frac{D^{\mathrm{x}}}{2}}} \exp\left(-\frac{1}{2}\mathrm{tr}((\boldsymbol{K}_{\boldsymbol{Y}})^{-1}\boldsymbol{X}\boldsymbol{X}^\top)\right) \\
\boldsymbol{K}_{\boldsymbol{Y}} &= \sum_m \boldsymbol{K}_{\boldsymbol{Y}^m}
\end{aligned} \tag{3.41}
$$

$\boldsymbol{K}_{\boldsymbol{Y}^m}$ の (i,j) 成分 $k^m_{\gamma,ij}$ は，以下のカーネル関数に基づき算出される．

$$
k^m_{\gamma,ij} = k^m_\gamma(\boldsymbol{y}^m_i, \boldsymbol{y}^m_j, \boldsymbol{\gamma}^m) \tag{3.42}
$$

提案手法は，未知の関心度の推定のための潜在特徴を算出する手法であり，既知の関心度の情報を潜在特徴に反映させることは，推定精度の向上に有効である [69]．そのため，ユーザが視聴済みのコンテンツに与えた関心度 $\boldsymbol{l} = [l_1, l_2, \ldots, l_N]$ を用いて，潜在変数の事前分布を以下で定義する．

$$
p(\boldsymbol{X}) = \frac{1}{\text{const.}} \exp\left(-\sum_{i=1}^N\sum_{j=1}^N w_{ij}\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2\right) \tag{3.43}
$$

$$
w_{ij} = \begin{cases} \alpha(\Delta - |l_i - l_j|)\dfrac{e^t}{1+e^t} & \\ \qquad\qquad (l_i \text{ and } l_j \text{ being obtained}) & \\ 0 \qquad\qquad (\text{otherwise}) & \end{cases} \tag{3.44}
$$

ただし，const. は定数，$t = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2$, α は事前に与えるパラメータ，$\Delta = \frac{L_{\max}-1}{2}$ である．

最後に，以下の目的関数を勾配降下法 [66] を用いて解き，$\boldsymbol{X}$，$\boldsymbol{\Theta}$ および $\boldsymbol{\Gamma}$ を算出

表 3.4: 各ユーザ毎の MAE の平均

実験参加者	**提案手法**	mGPLVM-MLP
A	**0.703**	0.757
B	**0.633**	0.744
C	0.651	**0.639**
D	**0.693**	0.746
E	**0.724**	0.787
F	**0.721**	0.760
G	**0.649**	0.748
H	**0.733**	0.748
I	**0.696**	0.736
J	**0.761**	0.792
平均 ± 標準偏差	**0.696** ±0.0390	0.746 ±0.0396

する．

$$\arg\min_{\boldsymbol{X},\boldsymbol{\Theta},\boldsymbol{\Gamma}} p(\boldsymbol{Y}^{\mathrm{c}}, \boldsymbol{Y}^{\mathrm{b}}|\boldsymbol{X}, \boldsymbol{\Theta}) p(\boldsymbol{X}|\boldsymbol{Y}^{\mathrm{c}}, \boldsymbol{Y}^{\mathrm{b}}, \boldsymbol{\Gamma}) p(\boldsymbol{X}) \tag{3.45}$$

$$p(\boldsymbol{Y}^{\mathrm{c}}, \boldsymbol{Y}^{\mathrm{b}}|\boldsymbol{X}, \boldsymbol{\Theta}) = \prod_{m} p(\boldsymbol{Y}^{m}|\boldsymbol{X}, \boldsymbol{\theta}^{m}) \tag{3.46}$$

式 (3.45) において，コンテンツ特徴と動作特徴の両方を用いて，逆射影のパラメータは最適化される．そのため，動作特徴が得られない未視聴コンテンツの特徴に対し，上記で算出した逆射影を用いることで，動作特徴を反映した統合特徴を算出することが可能である．

未視聴のコンテンツの特徴 $\boldsymbol{y}^{*c} \in \mathbb{R}^{D^c}$ に対する潜在変数 $\boldsymbol{x}^{*} \in \mathbb{R}^{D^{\mathrm{x}}}$ は，次式で算出される．

$$\boldsymbol{x}^{*} = \boldsymbol{K}_{\boldsymbol{y}^{*\mathrm{c}}\boldsymbol{Y}^{\mathrm{c}}} (\boldsymbol{K}_{\boldsymbol{Y}^{\mathrm{c}}})^{-1} \boldsymbol{X} \tag{3.47}$$

得られた潜在変数 $\boldsymbol{X}$ および $\boldsymbol{x}^{*}$ に分類器を適用することで，ユーザの関心度を推定することが可能である．

3.5.3 実験

本章では，提案手法の有効性を確認するための実験について説明する．本実験では，データセットとして，YouTube から，映画の予告編（30 秒）を ‘Action’, ‘Comedy’, ‘Music’ および ‘Science’ のジャンルからそれぞれ 10 本，‘Sport’ のジャンルから 9 本，

計 49 本取得し，各映像のコンテンツ特徴を，文献 [43] と同様に算出した．本実験では，10 名の実験参加者は一つの映像を 30 秒間視聴した後に，その映像に 4 段階の関心度 (1: 全く興味がない，2: あまり興味がない，3: 少し興味がある，4: とても興味がある) を 5 秒間で付与した．実験参加者は，以上のタスクを全ての映像に対して行った．映像視聴中の実験参加者の身体の位置情報および視線情報を，それぞれ OpenPose [47] および Tobii Eye Tracker 4C を用いて取得し，各ユーザの各映像に対する動作特徴を，文献 [43] と同様に算出した．特徴量の次元数は，$D^{\mathrm{c}} = 2,048$ および $D^{\mathrm{b}} = 64$ であった．本実験では，データセットのうち，20% を関心度が付与された学習データ，50% を関心度が付与されていない学習データおよび 30% をテストデータとした．比較手法として，Semi-OMGP に MLP に基づく逆射影 [61] を導入した手法である Semi-OMGP-MLP を用いた．提案手法と比較手法を比較するため，文献 [43] で用いられたテンソル補完を，それぞれの手法で得られた潜在変数と既知の関心度に適用し，未知の関心度を推定した．評価指標には，MAE を用いた．また，$Q = 20$，$\gamma_m = 2.0$ および $\lambda = 100,000$ とし，カーネル関数は線形カーネルを用いた．

各被験者に対する MAE を表 3.4 に示す．表 3.4 より，提案手法は，Semi-OMGP-MLP と比較して，低い MAE を達成していることが確認できる．Semi-OMGP-MLP は，逆射影に関数近似能力の高い MLP を用いているが，本実験のような小規模なデータセットにおいては過学習が生じやすく，テストデータに対する汎化性能が低下した可能性がある．一方，提案手法は確率モデルに基づき，コンテンツ特徴と動作特徴の両方の相関関係を考慮して逆射影を学習している．これにより，未知のコンテンツ特徴のみが入力された場合であっても，学習時に獲得した動作特徴との共起関係を利用することで，ユーザの反応傾向を反映した適切な潜在変数を推定できたと考えられる．以上の結果より，小規模かつノイズを含むマルチモーダルデータセットにおいて，提案する逆射影の枠組みが関心度推定に有効であることが示された．

3.5.4 まとめ

本稿では，逆射影を仮定した mGPLVM に基づくコンテンツおよび動作特徴の潜在空間の構築手法を提案した．提案手法は，コンテンツ特徴を統合特徴に変換す

る逆投影を推定可能であるため，新たに与えられたコンテンツ特徴から統合特徴を獲得することが可能である．また，提案手法は，ユーザの関心度の順序性を考慮して統合特徴の事前分布を定義し，関心度が潜在空間における分布を制約することを可能とした．その結果，提案手法はコンテンツ特徴を統合特徴に変換することができ，関心度推定に有効であることが実験的に示された．

3.6 擬似ラベルの付与によりサンプル間のペアワイズ制約を補完するSemi-MGPPLに基づく特徴統合

3.6.1 はじめに

前節までの検討により，動作特徴の統合や未視聴コンテンツへの対応といった課題に対する解決策を示した．しかしながら，ユーザセントリック特徴を用いた関心度推定において，最も深刻な課題の一つが「ラベル付きデータの極端な不足」である．ユーザに負担を強いるアノテーションは最小限に抑えられるべきであるが，ラベル付きデータが極端に少ない場合，Semi-OMGPのような半教師あり学習を用いても，潜在空間の構築に必要なペアワイズ制約が不足し，学習が不安定になる恐れがある．

そこで本節では，極少量のラベル付きデータからでも安定した学習を実現するために，**Multimodal Gaussian Process Latent Variable Model with Pseudo-Labels (Semi-MGPPL)** を提案する．本手法は，以下の二つの特徴により，データ不足と未視聴コンテンツの課題を同時に解決する．

1. **擬似ラベル (Pseudo-labels) の導入**: ラベル無しデータに対して，学習途中のモデルを用いて予測された「擬似ラベル」を付与する [70]．これにより，利用可能なラベル情報を仮想的に増幅させ，潜在空間に対する制約を強化することで，学習の安定化と精度向上を図る．

2. **逆射影の導入**: 第3.5節と同様に，観測空間から潜在空間への逆射影を導入することで，動作特徴が欠損している未視聴コンテンツへの対応を可能にする．

本手法により，ラベル付けされたサンプル数が少ない場合においても，擬似ラベルによる情報の補完と逆射影による推論の一般化により，高精度な関心度推定が可能となる．

3.6.2 提案手法

本章では，提案手法について説明する．提案手法は，複数のモダリティの特徴量を用いて共通の潜在変数を算出可能である．さらに，疑似ラベル [70,71] を仮定す

ることにより，少量のラベル情報を効率的に考慮可能である．また，逆射影 [61] の導入により，観測空間から潜在空間へのマッピング関数を算出可能である．

Semi-MGPPL の目的関数

本節では，Semi-MGPPL の目的関数について説明する．まず，ラベルのあるコンテンツ特徴 $\overline{\boldsymbol{Y}}^{\mathrm{c}} = [\overline{\boldsymbol{y}}_1^{\mathrm{c}}, \overline{\boldsymbol{y}}_2^{\mathrm{c}}, \ldots, \overline{\boldsymbol{y}}_{\overline{N}}^{\mathrm{c}}]^{\mathsf{T}} \in \mathbb{R}^{\overline{N} \times D^{\mathrm{c}}}$（$\overline{N}$ はラベルのあるサンプルの総数および D^{c} はコンテンツ特徴の次元数）およびラベルの無いコンテンツ特徴 $\underline{\boldsymbol{Y}}^{\mathrm{c}} = [\underline{\boldsymbol{y}}_1^{\mathrm{c}}, \underline{\boldsymbol{y}}_2^{\mathrm{c}}, \ldots, \underline{\boldsymbol{y}}_{\underline{N}}^{\mathrm{c}}]^{\mathsf{T}} \in \mathbb{R}^{\underline{N} \times D^{\mathrm{c}}}$（$\underline{N}$ はラベルの無いサンプルの総数）を定義する．次に，ラベルのある動作特徴 $\overline{\boldsymbol{Y}}^{\mathrm{b}} = [\overline{\boldsymbol{y}}_1^{\mathrm{b}}, \overline{\boldsymbol{y}}_2^{\mathrm{b}}, \ldots, \overline{\boldsymbol{y}}_{\overline{N}}^{\mathrm{b}}]^{\mathsf{T}} \in \mathbb{R}^{\overline{N} \times D^{\mathrm{b}}}$（$D^{\mathrm{b}}$ は動作特徴の次元数）およびラベルの無い動作特徴 $\underline{\boldsymbol{Y}}^{\mathrm{b}} = [\underline{\boldsymbol{y}}_1^{\mathrm{b}}, \underline{\boldsymbol{y}}_2^{\mathrm{b}}, \ldots, \underline{\boldsymbol{y}}_{\underline{N}}^{\mathrm{b}}]^{\mathsf{T}} \in \mathbb{R}^{\underline{N} \times D^{\mathrm{b}}}$ を定義する．さらに，それらを連結させた行列をそれぞれ $\boldsymbol{Y}^{\mathrm{c}} \in \mathbb{R}^{N \times D^{\mathrm{c}}}$（$N = \overline{N} + \underline{N}$）および $\boldsymbol{Y}^{\mathrm{b}} \in \mathbb{R}^{N \times D^{\mathrm{b}}}$ と定義する．ラベル $\bar{\boldsymbol{l}} = [\bar{l}_1, \bar{l}_2, \ldots, \bar{l}_{\overline{N}}]^{\mathsf{T}} \in \mathbb{R}^{\overline{N}}$ は，ユーザが視聴済みコンテンツに与える関心度である．ただし，$l_n \in \{1, 2, \ldots, L_{\max}\}$（$L_{\max}$ は関心度の段階の総数）であり，ラベルは順序性を持つ．$\boldsymbol{Y}^{\mathrm{c}}$ および $\boldsymbol{Y}^{\mathrm{b}}$ の共通の潜在変数を $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_N]^{\mathsf{T}} \in \mathbb{R}^{N \times Q}$ と定義する．ただし，Q は潜在変数の次元数である．また，$\boldsymbol{\Theta} = \{\boldsymbol{\theta}^1, \boldsymbol{\theta}^2, \ldots, \boldsymbol{\theta}^M\}$ はハイパーパラメータであり，$\boldsymbol{\theta}^m$ は $\boldsymbol{Y}^m$ におけるカーネル関数およびノイズの分散である．提案手法では，カーネル関数は全て RBF kernel [59] を用いる．

m-SimGP [52] と同様に，Semi-MGPPL は各モダリティの類似度行列が潜在変数から生成されると仮定する．類似度行列 $\boldsymbol{S}^m = [\boldsymbol{s}_1^m, \boldsymbol{s}_2^m, \ldots, \boldsymbol{s}_N^m]^{\mathsf{T}} \in \mathbb{R}^{N \times N}$ を次式により定義する．

$$s_{pq}^m = \exp\left(-\frac{\|\boldsymbol{y}_p^m - \boldsymbol{y}_q^m\|_2^2}{2\gamma_m}\right) \tag{3.48}$$

ただし，$m \in \{\mathrm{c}, \mathrm{b}\}$，$s_{pq}^m$ は $\boldsymbol{S}^m$ の (p, q) 要素および γ_m は正の定数である．類似度行列は，関数 f_q^m $(q = 1, \ldots, N)$ を用いて次式により生成されると仮定する．

$$s_{pq}^m = f_q^m(\boldsymbol{x}_p) + \epsilon_{pq}^m \tag{3.49}$$

$$\epsilon_{pq}^m \sim \mathcal{N}(0, \sigma_m^2) \tag{3.50}$$

類似度行列およびラベルを用いて，Semi-MGPPLの目的関数を次式により定義する．

$$\{\hat{\Psi}, \hat{\boldsymbol{\Theta}}\} = \arg\max_{\Psi,\boldsymbol{\Theta}} p(\boldsymbol{S}^{\mathrm{c}}, \boldsymbol{S}^{\mathrm{b}}|\Psi, \boldsymbol{\Theta}) p(\Psi)|\underline{\boldsymbol{\Sigma}}|^{-\lambda} \tag{3.51}$$

ただし，Ψ はマッピング関数 $g(\cdot)$ のパラメータであり，$\boldsymbol{x}_i = g(\boldsymbol{s}_i^{\mathrm{c}})$ と定義する．また，$p(\boldsymbol{S}^{\mathrm{c}}, \boldsymbol{S}^{\mathrm{b}}|\Psi, \boldsymbol{\Theta})$ は $\boldsymbol{S}^{\mathrm{c}}$ および $\boldsymbol{S}^{\mathrm{b}}$ の結合周辺尤度，$p(\Psi)$ は $\boldsymbol{X}$ の事前分布，$\underline{\boldsymbol{\Sigma}}$ は疑似ラベルの共分散行列および λ は事前に決定されるパラメータである．それぞれの変数の詳細な定義は以降で説明する．Semi-GPLVM に基づく手法 [63, 69] は取得済みのラベルを用いて $p(\Psi)$ を定義する．これに対し，Semi-MGPPL は，取得済みのラベルおよび疑似ラベルの両方を用いて用いて $p(\Psi)$ を定義する（**新規点 (i)**）．また，多くの mGPLVM に基づく手法は，潜在変数を直接最適化する．これに対し，Semi-MGPPL は，観測空間から潜在空間へのマッピング関数のパラメータを最適化する（**新規点 (ii)**）．式（3.51）は，以下の二つの手順を反復することで，解くことが可能である．

（a）Gaussian Process Regression（GPR）[72] に基づく疑似ラベルの更新

（b）MAP 推定に基づくマッピング関数のパラメータの更新

以降の節では，(a) および（b）のそれぞれについて説明する．

疑似ラベルの更新

本節では，GPR に基づく疑似ラベルの更新について説明する．提案手法は，$\boldsymbol{X}$ および $\bar{\boldsymbol{l}}$ を用いて，疑似ラベル $\underline{\boldsymbol{l}}$ を生成する．最初に，mGPLVM を用いて，潜在変数 $\boldsymbol{X}$ を初期化する．GPR で適用されるカーネル関数 $k'(\cdot)$ のパラメータ $\boldsymbol{\phi}$ を次式により最適化する．

$$\begin{aligned}\hat{\boldsymbol{\phi}} &= \arg\max_{\boldsymbol{\phi}} \log p(\bar{\boldsymbol{l}}|\overline{\boldsymbol{X}}, \boldsymbol{\phi}) \\ &= \arg\max_{\boldsymbol{\phi}} \log \mathcal{N}(\bar{\boldsymbol{l}}|\boldsymbol{0}, \tilde{\boldsymbol{K}}'_{\overline{\boldsymbol{X}}\,\overline{\boldsymbol{X}}})\end{aligned}$$

$\hat{\boldsymbol{\phi}}$ を用いて，Semi-MGPPL は疑似ラベルの予測分布 $\mathcal{N}(\underline{\boldsymbol{l}}|\underline{\boldsymbol{\mu}}, \underline{\boldsymbol{\Sigma}})$ を次式により算出する．

$$\underline{\boldsymbol{\mu}} = \boldsymbol{K}'_{\underline{\boldsymbol{X}}\,\overline{\boldsymbol{X}}}(\tilde{\boldsymbol{K}}'_{\overline{\boldsymbol{X}}\,\overline{\boldsymbol{X}}})^{-1}\bar{\boldsymbol{l}} \tag{3.52}$$

$$\underline{\boldsymbol{\Sigma}} = \boldsymbol{K}'_{\underline{\boldsymbol{X}}\,\underline{\boldsymbol{X}}} - \boldsymbol{K}'_{\underline{\boldsymbol{X}}\,\overline{\boldsymbol{X}}}(\tilde{\boldsymbol{K}}'_{\overline{\boldsymbol{X}}\,\overline{\boldsymbol{X}}})^{-1}\boldsymbol{K}'_{\overline{\boldsymbol{X}}\,\underline{\boldsymbol{X}}} \tag{3.53}$$

ただし，$\boldsymbol{K}'_{\boldsymbol{XX}}$ は共分散行列および $\tilde{\boldsymbol{K}}'_{\overline{\boldsymbol{X}}\,\overline{\boldsymbol{X}}} = \boldsymbol{K}'_{\overline{\boldsymbol{X}}\,\overline{\boldsymbol{X}}} + \sigma^2 \boldsymbol{I}_{\overline{N}}$（$\boldsymbol{I}_{\overline{N}}$ は $\overline{N}$ 次元の単位行列）である．Semi-MGPPL は，事前分布に制約を与えることで，取得済みのラベルおよび疑似ラベルの情報を潜在空間に反映させる．$\bar{\boldsymbol{l}}$ および $\underline{\boldsymbol{\mu}}$ の連結ベクトルを $\boldsymbol{l} \in \mathbb{R}^N$ と定義し，事前分布 $p(\Psi)$ を以下で定義する．

$$p(\Psi) = \frac{1}{Z} \exp\left(- \sum_{i,j=1}^{N} w_{ij} \| g(\boldsymbol{s}_i^{\mathrm{c}}) - g(\boldsymbol{s}_j^{\mathrm{c}}) \|_2 \right) \tag{3.54}$$

$$w_{ij} = \alpha(\Delta - |l_i - l_j|) \frac{e^t}{1 + e^t} \tag{3.55}$$

ただし，Z は正規化定数，$t = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2$，$\alpha$ は事前に決定されるパラメータおよび $\Delta = \frac{(L_{\max} - 1)}{2}$ である．さらに，式（3.54）は以下のように書き直される．

$$\begin{aligned} p(\Psi) &= \frac{1}{Z} \exp\left(- \sum_{i,j=1}^{N} w_{ij} \| g(\boldsymbol{s}_i^{\mathrm{c}}) - g(\boldsymbol{s}_j^{\mathrm{c}}) \|_2 \right) \\ &= \frac{1}{Z} \exp\left(-\mathrm{tr}(\boldsymbol{X}^{\mathsf{T}} \boldsymbol{W} \boldsymbol{X}) \right) \\ &= \frac{1}{Z} \exp\left(-\mathrm{tr}(\boldsymbol{W} \boldsymbol{X} \boldsymbol{X}^{\mathsf{T}}) \right) \end{aligned} \tag{3.56}$$

ただし，w_{ij} は $\boldsymbol{W}$ の (i, j) 要素である．しかしながら，疑似ラベルに大幅な誤推定が存在する場合，式（3.54）における疑似ラベルの利用のみでは，関心度推定の精度の向上に限界が存在する [73]．この問題を解決するため，式（3.53）で算出される $\underline{\boldsymbol{\Sigma}}$ を，式（3.51）の目的関数において最小化する．

疑似ラベルを生成することにより，サンプルのペアの全てに対して式（3.55）で示す $w_{i,j}$ の算出が可能になる．したがって，**新規点（i）**の導入により，**問題点（i）**が解決可能である．

マッピング関数のパラメータの更新

本節では，MAP 推定に基づくマッピング関数のパラメータの更新について説明する．結合周辺尤度 $p(\boldsymbol{S}^{\mathrm{c}}, \boldsymbol{S}^{\mathrm{b}}|\Psi, \boldsymbol{\Theta})$ は次式により算出される [74].

$$\begin{aligned} p(\boldsymbol{S}^{\mathrm{c}},\boldsymbol{S}^{\mathrm{b}}|\Psi, \boldsymbol{\Theta}) &= p(\boldsymbol{S}^{\mathrm{c}}|\Psi, \boldsymbol{\theta}^{\mathrm{c}}) p(\boldsymbol{S}^{\mathrm{b}}|\Psi, \boldsymbol{\theta}^{\mathrm{b}}) \\ &= \frac{1}{(2\pi)^{\frac{ND^{\mathrm{c}}}{2}} |\tilde{\boldsymbol{K}}^{\mathrm{c}}_{\boldsymbol{XX}}|^{\frac{D^{\mathrm{c}}}{2}}} \exp\left(-\frac{1}{2}\mathrm{tr}(\tilde{\boldsymbol{K}}^{\mathrm{c}}_{\boldsymbol{XX}}{}^{-1} \boldsymbol{S}^{\mathrm{c}} (\boldsymbol{S}^{\mathrm{c}})^{\mathsf{T}})\right) \\ &\quad \frac{1}{(2\pi)^{\frac{ND^{\mathrm{b}}}{2}} |\tilde{\boldsymbol{K}}^{\mathrm{b}}_{\boldsymbol{XX}}|^{\frac{D^{\mathrm{b}}}{2}}} \exp\left(-\frac{1}{2}\mathrm{tr}(\tilde{\boldsymbol{K}}^{\mathrm{b}}_{\boldsymbol{XX}}{}^{-1} \boldsymbol{S}^{\mathrm{b}} (\boldsymbol{S}^{\mathrm{b}})^{\mathsf{T}})\right) \end{aligned} \tag{3.57}$$

式（3.51）は次式で書き直される．

$$\begin{aligned} \{\hat{\Psi}, \hat{\boldsymbol{\Theta}}\} &= \arg\max_{\Psi, \boldsymbol{\Theta}} \log \mathcal{L} \\ \log \mathcal{L} &= \log p(\boldsymbol{S}^{\mathrm{c}}, \boldsymbol{S}^{\mathrm{b}}|\Psi, \boldsymbol{\Theta}) + \log p(\Psi) - \lambda \log |\underline{\boldsymbol{\Sigma}}| \end{aligned} \tag{3.58}$$

提案手法は，従来の GPLVM に基づく手法 [62,65,75] と同様に，scaled conjugate gradient [66] 法を用いて式（3.58）を解く．

式（3.52）および（3.58）を反復して解くことにより，マッピング関数のパラメータ Ψ およびハイパーパラメータ $\boldsymbol{\Theta}$ を算出可能である．マッピング関数を用いて，次式によりテストデータ $\boldsymbol{y}^{\mathrm{c}}_{(\mathrm{t})}$ の潜在変数 $\boldsymbol{x}_{(\mathrm{t})}$ を算出する．

$$\boldsymbol{x}_{(\mathrm{t})} = g(\boldsymbol{s}^{\mathrm{c}}_{(\mathrm{t})}) \tag{3.59}$$

$g(\cdot)$ をコンテンツ特徴から潜在空間への射影と定義することにより，提案手法は動作特徴の無いテストデータの潜在変数が算出可能である．したがって，**新規点（ii）**の導入により，**問題点（ii）**が解決可能である．

3.6.3 実験

本章では，提案手法の有効性および頑健性を確認するための実験を行う．

実験設定

本実験では，データセットとして，YouTube から，映画の予告編（30 秒）を「アクション」，「コメディ」，「ミュージック」および「サイエンス」のジャンルからそ

れぞれ10本,「スポーツ」のジャンルから9本，計49本取得し，各映像のコンテンツ特徴を，文献[43]と同様に算出した．本実験では，10名の実験参加者は一つの映像を30秒間視聴した後に，その映像に4段階の関心度(1: 全く興味がない，2: あまり興味がない，3: 少し興味がある，4: とても興味がある)を5秒間で付与した．実験参加者は，以上のタスクを全ての映像に対して行った．映像視聴中の実験参加者の身体の位置情報および視線情報を，それぞれOpenPose [47]およびTobii Eye Tracker 4Cを用いて取得し，各ユーザの各映像に対する動作特徴を，文献[43]と同様に算出した．特徴量の次元数は，$D^{\mathrm{c}} = 2,048$および$D^{\mathrm{b}} = 64$であった．本実験では，以下の3つの設定で実験を行った．

設定（i）: データセットのうち，10%をラベルあり学習データ，60%をラベルなし学習データおよび30%をテストデータとした．

設定（ii）: データセットのうち，20%をラベルあり学習データ，50%をラベルなし学習データおよび30%をテストデータとした．

設定（iii）: データセットのうち，30%をラベルあり学習データ，40%をラベルなし学習データおよび30%をテストデータとした．具体例として，設定（i）におけるデータセットの構成を図3.4に示す．ラベルありデータの数が異なる3種類の実験を行うことにより，提案手法の頑健性を確認する．

比較手法として，Back-constraint (BC) Semi-OMGP，BC-mGPLVM, BC-m-SimGP, Multi-view CCA (MVCCA) [53], Deep CCA [54]およびBayesian CCA (BCCA) [55]を用いた．BC-Semi-OMGP，BC-mGPLVMおよびBCm-SimGPは，それぞれSemi-OMGP [69]，mGPLVM [45]およびm-SimGP [52]にBCを導入した手法である．提案手法と比較手法を比較するために，文献[43]で用いられたテンソル補完を，それぞれの手法で得られた潜在変数と既知の関心度に適用し，未知の関心度を推定した．評価指標には，MAEを用いた．また，提案手法において，$Q = 20$，$\gamma_m = 2.0$および$\lambda = 100000$とし，マッピング関数はMultilayer perceptron，カーネル関数はRBF kernel [59]を用いた．

実験結果

表3.5，3.6および3.7に実験結果を示す．決定論的手法であるMVCCAおよびDeep learningに基づく手法であるDeep CCAと，その他の手法の精度を比較することに

表 3.5: 実験結果（設定（i））

実験参加者	**Semi-MGPPL**	BC-mGPLVM [45]	BC-m-SimGP [52]	BC-Semi-OMGP [69]	MVCCA [53]	BCCA [55]	Deep CCA [54]
A	**0.713**	0.785	1.013	0.759	1.421	0.724	1.553
B	0.756	**0.747**	1.065	0.789	1.476	0.769	1.507
C	**0.747**	0.761	0.982	0.833	1.387	0.770	1.507
D	0.683	**0.642**	1.171	0.783	1.426	0.793	1.652
E	**0.718**	0.781	1.093	0.768	1.369	0.755	1.480
F	0.791	**0.767**	1.132	0.779	1.372	0.795	1.453
G	0.791	**0.662**	1.067	0.725	1.433	0.794	1.527
H	**0.650**	0.773	1.038	0.730	1.409	0.773	1.372
I	0.729	0.737	1.178	**0.718**	1.421	0.815	1.419
J	**0.694**	0.729	1.030	0.746	1.409	0.716	1.553
平均 ± 標準偏差	**0.727** ±0.0432	0.738 ±0.0467	1.077 ±0.0629	0.763 ±0.0334	1.412 ±0.0302	0.770 ±0.0301	1.502 ±0.0742

表 3.6: 実験結果（設定（ii））

実験参加者	**Semi-MGPPL**	BC-mGPLVM [74]	BC-m-SimGP [52]	BC-Semi-OMGP [69]	MVCCA [53]	BCCA [55]	Deep CCA [54]
A	0.761	0.744	1.014	**0.683**	1.330	0.779	1.494
B	**0.709**	0.757	1.074	0.751	1.374	0.768	1.444
C	**0.728**	0.746	0.950	0.785	1.361	0.764	1.361
D	0.715	**0.639**	0.999	0.764	1.320	0.740	1.425
E	**0.671**	0.760	1.031	0.777	1.320	0.754	1.447
F	0.793	0.787	1.059	**0.715**	1.290	0.733	1.416
G	**0.687**	0.748	0.976	0.778	1.191	0.731	1.488
H	0.698	0.748	1.021	**0.677**	1.328	0.762	1.467
I	0.745	0.792	1.027	0.741	1.318	**0.731**	1.298
J	**0.705**	0.736	1.001	0.742	1.258	0.717	1.458
平均 ± 標準偏差	**0.721** ±0.0345	0.746 ±0.0396	1.015 ±0.0348	0.741 ±0.0366	1.309 ±0.0497	0.748 ±0.0191	1.430 ±0.0568

表 3.7: 実験結果（設定（iii））

実験参加者	**Semi-MGPPL**	BC-mGPLVM [45]	BC-m-SimGP [52]	BC-Semi-OMGP [69]	MVCCA [53]	BCCA [55]	Deep CCA [54]
A	0.713	**0.677**	0.985	0.704	1.336	0.710	1.339
B	**0.717**	0.748	0.936	0.762	1.330	0.726	1.265
C	**0.691**	0.760	1.005	0.726	1.323	0.717	1.302
D	0.766	**0.748**	0.997	0.808	1.354	0.784	1.419
E	**0.698**	0.729	0.971	0.754	1.260	0.765	1.387
F	0.720	**0.714**	0.992	0.737	1.303	0.737	1.379
G	**0.664**	0.732	0.932	0.711	1.371	0.693	1.419
H	**0.681**	0.746	0.907	0.724	1.351	0.696	1.406
I	0.717	0.754	0.949	0.820	1.285	**0.709**	1.389
J	**0.717**	0.788	0.901	0.730	1.438	0.730	1.376
平均	**0.708**	0.740	0.957	0.748	1.335	0.727	1.368
± 標準偏差	±0.0261	±0.0282	±0.0359	±0.0372	±0.0468	±0.0277	±0.0485

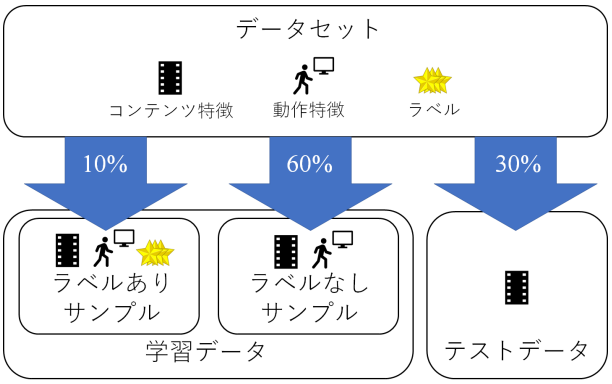

図 3.4: 設定（i）におけるデータセットの構成

より，確率論的手法を用いることの有効性が確認された．これは，ユーザの動作特徴が，小サンプルかつノイズの多い特徴量であるからだと考えられる．また，提案手法と BC-mGPLVM，BC-m-SimGP および BCCA を比較することにより，ユーザがコンテンツに与えた関心度を特徴統合に利用することの有効性が確認された．これは，関心度が類似している特徴量のペアは潜在空間で近くに，関心度が大きく離れている特徴量のペアは潜在空間で遠くに位置しているからであると考えられる．また，提案手法と，BC-Semi-OMGP を比較することにより，ラベル無し特徴量の擬似ラベルを考慮することの有効性が確認された．これは，擬似ラベルにより，ラベルあり特徴量のペアの数が増加し，各コンテンツにユーザが与えた関心度が潜在空間に効率的に反映されているからであると考えられる．最後に，ラベル付けされている特徴量の割合が小さい場合および大きい場合の両方で同様の結果が得られていることから，提案手法の頑健性が確認された．

3.6.4 まとめ

本文では，ユーザの関心度推定を目的としたコンテンツおよびユーザの動作の特徴統合手法を提案した．具体的に，ユーザの動作情報を用いた関心度推定に適した，Semi-MGPPL と呼ばれる mGPLVM ベースの新しい手法を提案した．提案する Semi-MGPPL は，ラベルのないサンプルに対して擬似ラベルを仮定し，ラベル付きサンプルのペアを増加させることができるため，提案手法は少量のラベル情報を潜在空間に効率的に反映させることが可能である．さらに，Semi-MGPPL では観測空間から潜在空間への射影である BC を導入しているため，提案手法は新たに

得られたテストデータの潜在変数を算出することができる．実験により，提案する Semi-MGPPL の有効性と非ラベル付きサンプル数の変化に対する頑健性が確認された．

3.7 まとめ

本章では，ユーザの多様な特性をモデル学習段階から反映させることを目的とし，ユーザセントリック特徴とコンテンツ特徴を統合する確率的生成モデルの枠組みを構築した．第 1 章で述べたように，ユーザの動作情報はノイズを含みやすく，かつ大規模な収集が困難であるという課題があったが，本章では mGPLVM を基盤とすることで，これらの不確実性を考慮した頑健な特徴統合を実現した．

具体的には，まず連続値特徴量を扱うための量子化手法と mGPLVF を提案し，異種特徴を共通の潜在空間上で統合することで，単一モダリティでは捉えきれないユーザの反応傾向を表現できることを示した．続いて，実環境におけるデータ収集の制約に対応するため，半教師あり学習や擬似ラベルの導入による拡張を行い，ラベル情報が極端に不足している状況や，未視聴コンテンツに対しても，高精度な関心度推定が可能であることを実証した．これらの成果は，表層的な行動履歴に依存していた従来手法とは異なり，ユーザの内部状態を反映した潜在表現を獲得することで，より本質的なパーソナライズ化が可能であることを示唆している．

本章で確立した「学習段階における個人差の考慮」は，次章で扱う「学習済みモデルに対する事後的なユーザ適応」と対をなすものである．次章では，本章のアプローチとは異なり，大規模な汎用モデルから必要な知識は保持しつつ，不要な他者の知識を忘却するという観点から，効率的なパーソナライズ化の枠組みについて論じる．

第4章

学習済みモデルからの他ユーザ由来の知識の選択的保持および忘却に基づくユーザ適応

4.1 はじめに

前章では，視線や動作情報といったユーザセントリック特徴とコンテンツ特徴を，確率的生成モデルを基盤として統合する手法について論じてきた．第3章で示したアプローチは，ユーザの内部状態を反映した潜在空間を学習段階から構築することで，関心度推定のような主観的なタスクにおいて高い有効性を示した．しかしながら，実社会における機械学習に基づくシステムの運用を考慮すると，常にゼロから学習を行うことは計算コストや運用コストの観点から現実的ではない．特に，近年急速に普及している大規模な基盤モデルや，すでに運用されている学習済みモデルに対しては,「モデル学習後」の段階で個々のユーザに適応させる技術が不可欠となる．

モデルのユーザ適応のための手法として，従来は，追加データを用いたモデル適応（知識の追加）が主流であった．これは，対象ユーザのデータを追加してファインチューニング等を行うアプローチであるが，対象ユーザから得られるデータが少量である場合，事前学習済みモデルが保持している膨大な他ユーザの知識（バイアス）に埋もれてしまい，十分な適応効果が得られないという課題がある．一方で，元来ユーザのプライバシー保護のため用いられるマシンアンラーニング手法は，不要な情報を削除することに特化しているが，パーソナライズの文脈で，ど

のような知識を保持し，何を忘却するかを統合的に制御する手法としては体系化されていない．

そこで本章では，図2.1の下半分全体をカバーするアプローチとして，学習済みモデルからの他ユーザ由来の知識を選択的保持および忘却するユーザ適応手法を提案する．本アプローチは，従来の「知識を追加する（適応）」か「知識を削除する（アンラーニング）」かという二元論的な対立ではなく，これらを統合し，対象ユーザにとって有用な知識（類似ユーザの傾向など）を選択的に保持しつつ，不要な知識（非類似ユーザの傾向）を忘却することで，相対的に対象ユーザへの適合度を最大化することを目指すものである．

本章の構成は以下の通りである．まず，第4.2節では，データ分割に基づく物理的な知識選別アプローチを提案する．ここでは，ユーザ間の感情反応の類似度に基づき学習データを事前にクラスタリングし，対象ユーザと非類似なグループを再学習プロセスから物理的に忘却することで，モデルのパーソナライズ化を実現する．これは，SISA [25]を応用し，再学習コストの削減と精度の向上を同時に目指すものである．次に，第4.3節では，より柔軟かつ連続的な制御を可能にする，勾配制御に基づくアプローチを提案する．本手法では，ユーザとアイテムのインタラクションに基づくグラフ構造から連続的なユーザ類似度を算出し，知識蒸留 [76]の考え方を用いて,「保持すべき知識」と「忘却すべき知識」のバランスを動的に調整する．これにより，データの物理的な分割が困難な場合や，より微細な類似度を反映させる必要がある場合においても，効果的なパーソナライズを実現することを目指す．最後に，第4.4節にて，本章の結論を述べる．

4.2 データ分割とサブモデル集約に基づくユーザ適応

4.2.1 はじめに

本節では，データの物理的な分割と選択的な再学習に基づくユーザ適応手法について詳述する．本アプローチの核心は，SISA [25] を応用し，対象ユーザと傾向が異なるユーザのデータを「不要な知識」として物理的にモデルから遮断することにある．

SISA は，ランダムにデータを分割することを前提としているが，パーソナライズを目的とする場合，この分割方法は最適ではない．なぜなら，ランダム分割では非類似ユーザのデータが全シャードに均等に分散してしまい，それらを除去するために多くのサブモデルを再学習する必要が生じるからである．そこで本節では，ユーザ間の感情反応の類似度（カッパ係数）に基づくクラスタリングを導入し，類似した傾向を持つユーザ同士を同一のシャードに凝集させる手法を提案する．これにより，非類似ユーザのデータが特定のシャードに偏在することになり，ごく一部のサブモデルを再学習するだけで，効率的かつ効果的なパーソナライズ（非類似知識の忘却）が可能となる．以下，提案手法の詳細および実験による評価結果について述べる．

4.2.2 提案手法

本節では，提案するマシンアンラーニングの考えに基づくのユーザ適応手法について説明する．提案手法の概要図を図 4.1 に示す．本手法は，学習フェーズ，適応フェーズおよび推論フェーズの 3 段階から構成される．なお，提案手法で用いる画像の感情ラベル予測モデルは特定の種類に限定されず，畳み込みニューラルネットワーク (CNN) [77] や Vision Transformer (ViT) [78] など様々な画像認識モデルに適用可能である．

学習フェーズ

学習フェーズでは，Cohen のカッパ係数 [79] に基づくユーザクラスタリングにより学習データを分割し，各シャードごとに感情ラベル予測サブモデルを学習する．まず，画像を $I_1, I_2, \ldots, I_{N^{\text{img}}}$，ユーザを $u_1, u_2, \ldots, u_{N^{\text{usr}}}$ とし，N^{img} を画像数，N^{usr} をユー

ザ数とする．各感情ラベルには種類に対応する番号を割り当て，ユーザ u_m が画像 I_n に付与した感情ラベルを $l_{m,n}$ と表す．ユーザ u_m が画像 I_n にラベルを付与していない場合は $l_{m,n} = 0$，ラベルが付与されている場合は $l_{m,n} = c$ ($\in \{1, 2, \ldots, N^{\text{label}}\}$) とし，$N^{\text{label}}$ を感情ラベルの総種類数とする．

次に，$l_{m,1}, l_{m,2}, \ldots, l_{m,N^{\text{img}}}$ を連結することで，ユーザベクトル $\boldsymbol{b}_m \in \mathbb{R}^{N^{\text{img}}}$ を得る．これらのユーザベクトル $\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_{N^{\text{usr}}}$ をクラスタリングし，ユーザを N^{shard} 個のクラスタに分割する．本論文では，画像に対する感情反応の類似度を表す指標として，ユーザベクトル間のカッパ係数 [79] に基づいてクラスタリングを行う．具体的には，k-means に類似した手順で任意のベクトルをクラスタ中心の初期値として選択し，各クラスタ内のデータ間のカッパ係数の和が最小となるようにクラスタ中心を更新する．ユーザベクトル $\boldsymbol{b}_m$ と $\boldsymbol{b}_{m'}$ のカッパ係数 $\kappa_{m,m'}$ は次式で計算される．

$$\kappa_{m,m'} = \frac{P_o - P_\epsilon}{1 - P_\epsilon},\ P_o = \frac{\sum_{c=1}^{N^{\text{label}}} \eta_{c,c}}{H} \tag{4.1}$$

$$P_\epsilon = \sum_{c=1}^{N^{\text{label}}} \frac{\left(\sum_{c'=1}^{N^{\text{label}}} \eta_{c,c'}\right)\left(\sum_{c'=1}^{N^{\text{label}}} \eta_{c',c}\right)}{H^2} \tag{4.2}$$

ここで $H = \sum_{c=1}^{N^{\text{label}}} \sum_{c'=1}^{N^{\text{label}}} \eta_{c,c'}$ は，ユーザ m が感情ラベル c を付与し，ユーザ m' が感情ラベル c' を付与した画像の総数を表す．ユーザベクトル間の類似度をカッパ係数に基づいて計算することで，ユーザ m と m' の両者がラベルを付与した画像数に依存せず，偶然一致を補正した信頼性の高い類似度が得られる [80]．この手順により，画像に対して類似した感情反応を示すユーザ同士を同一クラスタとしてまとめることができる．

次に，x 番目のクラスタに属するユーザが付与したデータ集合を $\mathcal{S}_x$ とおき，$\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_{N^{\text{shard}}}$ というシャードを得る．最後に，各シャード $\mathcal{S}_x$ に含まれるデータを用いて感情ラベル予測モデル $\mathcal{M}_x$ を学習する．実際のモデル学習においては，適応フェーズでの効率をさらに高めるため，SISA における slice 戦略 [25] に従い，シャード内のデータを細かく分割したスライスごとに学習パラメータを保存している．

適応フェーズ

適応フェーズでは，4.2.2 項で学習した感情ラベル予測モデルから，対象ユーザと非類似なユーザに由来するデータの影響を除去することでモデルのパーソナライ

ズ化を行う．

まず，パーソナライズの対象となる新規ユーザを u_{tgt} とする．このユーザは学習データには含まれていないと仮定する．対象ユーザは，$I_1, I_2, \ldots, I_{N^{\text{img}}}$ のうち N^{tgt} 枚に対して感情ラベルを付与する．対象ユーザが画像 I_n に付与した感情ラベルを $l_{\text{tgt},n}$ と表し，$N^{\text{tgt}} \ll N^{\text{img}}$ と仮定する．

次に，4.2.2 項と同様の手順でカッパ係数 $\kappa_{\text{tgt},1}, \kappa_{\text{tgt},2}, \ldots, \kappa_{\text{tgt},N^{\text{usr}}}$ を算出し，対象ユーザ u_{tgt} と既存ユーザ $u_1, u_2, \ldots, u_{N^{\text{usr}}}$ との類似度を求める．ここで，カッパ係数は偶然一致を補正できるため，N^{tgt} が小さい場合でも信頼性の高い類似度を算出できる．得られた類似度に基づき，カッパ係数が小さいユーザから順に d 人分の学習データを「不要データ」とみなし，これらを削除するためのアンラーニングを行う．具体的には，シャード $\mathcal{S}_x$ に不要データが含まれている場合，そのデータを取り除いた $\mathcal{S}_x$ を用いて $\mathcal{M}_x$ を再学習し，$\mathcal{S}_x$ から当該データの影響を除去する．再学習は不要データを含むシャードに対してのみ行い，他のシャードは再学習を行わない．

図 4.2 は，ユーザクラスタリングによるデータ分割と，対象ユーザにとって不要であると推定されるデータの選択の様子を模式的に示したものである．本手法の特長は以下の 2 点に集約される．

- **知識の選択的除去:** ユーザベクトル間の類似度（カッパ係数）に基づきモデルから除去すべきデータを選択することで，新たなデータを大量に追加することなく，非類似なデータの影響を取り除くだけでパーソナライズ化を実現する．

- **再学習コストの抑制:** 学習フェーズにおいてユーザ類似度に基づいてシャード分割を行っているため，削除対象となる非類似ユーザのデータは特定のシャードに偏って分布する．このシャードバイアスにより，再学習が必要となるシャード数を抑制でき，効率的な処理が可能となる．

推論フェーズ

推論フェーズでは，パーソナライズ済み感情ラベル予測モデルを用いて，対象ユーザに対する新規画像の感情ラベルを予測する．4.2.2 項で更新されたサブモデル群に新規画像 I_{test} を入力し，各サブモデルからの出力 $l^*_1, l^*_2, \ldots, l^*_{N^{\text{shard}}}$ を得る．これ

らの出力を次式のように統合し，最終出力 l^* を得る．

$$l^* = \sum_{x=1}^{N^{\text{shard}}} \frac{q_x}{q_{\text{sum}}} l_x^* \tag{4.3}$$

ここで，$q_{\text{sum}} = \sum_{x=1}^{N^{\text{shard}}} q_x$ である．q_x はサブモデル $\mathcal{M}_x$ の重要度に応じて設定できる．本論文では，シャード内のデータ量に比例する重みを用い，$q_x = |\mathcal{S}_x| / \sum_{x'=1}^{N^{\text{shard}}} |\mathcal{S}_{x'}|$ と定義する．ここで $|\mathcal{S}_x|$ はシャード $\mathcal{S}_x$ に含まれるデータ数を表す．

4.2.3 実験

本節では，提案手法（Proposed Framework: PF）の有効性を確認するための実験について述べる．4.2.3 項では，使用したデータセットおよび比較手法について説明する．4.2.3 項では，予測精度および再学習時間の観点から実験結果を示し，考察を行う．

実験設定

実験では，Affection データセット [81] と Personalized image aesthetics database with rich attributes（PARA データセット）[82] の 2 種類を使用した．Affection データセットは，5 つの代表的な画像データセット [83–87] から収集された 85,007 枚の画像で構成されており，各画像には “Amusement”，“Awe”，“Contentment”，“Excitement”，“Anger”，“Disgust”，“Fear”，“Sadness”，“Something else” のいずれかの感情ラベルが付与されている．本研究では，“Something else” のラベルが付与されたサンプルおよび 1,000 枚未満しかラベル付与を行っていないユーザを除外し，最終的に 72,869 枚の画像（= N^{img}）と 85 人のユーザから構成される 153,505 サンプルを使用した．PARA データセットは 31,200 枚の画像からなり，ユーザの主観評価に基づき “Amusement”，“Awe”，“Contentment”，“Excitement”，“Disgust”，“Fear”，“Sadness”，“Neutral” の感情ラベルが付与されている．本研究では “Neutral” とラベル付けされたサンプルを除外し，1,397 枚の画像（= N^{img}）と 272 人のユーザからなる 15,300 サンプルを使用した．なお，いずれのデータセットにおいても，全てのユーザが全ての画像にラベル付与しているわけではない．

両データセットにおいて，5 人のユーザをランダムに選択し対象ユーザとし，そ

表 4.1: 各シャード数における提案手法（PF）の精度．値は，非類似ユーザのデータを削除しなかった場合からの差分を示す．

Model	N^{shard}	Affection dataset [81]						PARA dataset [82]					
		User 1	User 2	User 3	User 4	User 5	Ave.	User 1	User 2	User 3	User 4	User 5	Ave.
ViT	1	-1.315	**+0.633**	**+0.646**	**+0.420**	**+0.347**	**+0.146**	0.000	-0.465	0.000	**+1.020**	0.000	**+0.111**
ViT	2	**+0.563**	**+0.286**	**+0.462**	**+0.420**	-0.289	**+0.288**	0.000	**+0.465**	**+5.882**	0.000	0.000	**+1.269**
ViT	3	**+1.127**	**+0.470**	**+0.831**	**+0.756**	**+1.100**	**+0.857**	0.000	0.000	**+11.765**	0.000	0.000	**+2.353**
ViT	4	-2.911	-0.143	-0.092	-0.084	-0.116	-0.669	**+2.500**	0.000	**+5.882**	0.000	0.000	**+1.676**
CNN	1	**+1.690**	**+7.843**	**+1.939**	-4.118	-3.414	**+0.788**	0.000	**+0.465**	0.000	0.000	0.000	**+0.093**
CNN	2	**+0.939**	**+4.902**	**+0.462**	-0.168	**+0.347**	**+1.296**	**+3.000**	**+0.930**	0.000	0.000	0.000	**+0.786**
CNN	3	**+2.254**	**+1.818**	**+0.646**	**+0.252**	-0.521	**+0.890**	**+7.000**	-3.721	**+17.647**	-3.061	**+5.926**	**+4.758**
CNN	4	**+3.005**	-1.797	-0.739	**+4.034**	**+3.935**	**+1.687**	**+4.839**	0.000	0.000	**+2.041**	0.000	**+1.376**

れ以外のユーザのデータを学習に用いた．各対象ユーザごとに感情ラベル予測モデルをパーソナライズ化した．対象ユーザから得られるデータがほとんど存在しない状況を想定し，パーソナライズに用いるサンプル数を 10（$N^{\mathrm{tgt}} = 10$）とし，残りのデータをテストデータとして用いた．評価指標として，予測結果と真の感情ラベルから算出した正解率を用いた．

実験では，感情ラベル予測モデルとして ViT [88] および CNN [89] を選択し，PF および全ての比較手法で同一のモデルを使用した．ViT ベースの予測モデルは，大規模データセットで事前学習されたモデルの分類層のみを微調整することで構築した．ViT は 12 層，12 ヘッド，隠れ次元 768，および 3,072 次元の全結合層から構成される．学習エポック数は 10，学習率は 0.0005 とした．CNN ベースの予測モデルはゼロから学習し，2 層の畳み込み層と 1 層の全結合層で構成した．学習エポック数は 20，学習率は 0.0005 とした．

PF に基づく感情ラベル予測精度を評価するため，まずアンラーニングを行わない場合の精度を算出し，その差分を求めた．提案手法の第一の貢献である「知識の選択的除去」の有効性を確認するため，PF においてシャード数 N^{shard} と削除ユーザ数 d を変化させ，各対象ユーザに対してモデルをパーソナライズし，テストデータに対する精度を比較した．さらに，第二の貢献である「パーソナライズ効率の向上」を評価するため，比較手法として，以下の 3 つを用いた．

- **CF1 (Retraining):** 非類似ユーザのデータを除外した後に，モデル全体の単純な全再学習を行う手法．
- **CF2 (Random-Sample):** サンプルレベルでランダムにデータを分割し，非類似ユーザのデータを削除する手法．
- **CF3 (Random-User):** ユーザ単位でシャードにランダムに割り当てるが，ユーザごとの類似度に基づくデータ分割は行わない手法．

全ての実験は，13th Gen Intel(R) Core(TM) i9-13900K（24 コア，32 スレッド，最大 5.8 GHz），NVIDIA RTX A5000（24 GB VRAM，CUDA 12.4，ドライバ 550.120），および 4000 MHz 動作の 32 GiB DIMM を 4 枚搭載した合計 128 GiB のメモリを備えた計算環境上で行った．

実験結果

表 4.1 に，シャード数 $N^{\text{shard}} \in \{1, 2, 3, 4\}$ を変化させた場合の PF による感情ラベル予測結果を示す．ここで，$N^{\text{shard}} = 1$ はシャーディングを行わない場合に相当する．Affection データセットを用いた結果から，シャード数が 4 でモデルが ViT の場合を除き，非類似ユーザの影響の除去が多くのユーザに対して精度を向上させていることが確認できる．一方，シャード数が 4 と小さい場合，すなわち各サブモデルに割り当てられる学習データ数が少ない場合には精度向上が確認できず，データ規模に応じた適切なシャード数と予測モデルの選択が必要であると考えられる．しかし，それ以外の条件では非類似ユーザの影響を除去することでモデルパーソナライズ化が可能であることが確認できる．PARA データセットを用いた実験結果からも，多くの場合で PF が予測精度を維持あるいは改善しており，その有効性が確認できる．

各画像に対し，対象ユーザならびに類似ユーザ・非類似ユーザが付与したラベルから，PF の予測結果が非類似ユーザよりも類似ユーザの影響を強く受けていることが定性的にも確認された．この結果から，PF は非類似ユーザの影響をモデルから除去することで，対象ユーザ固有の感情ラベルを適切に予測できているといえる．

さらに検証として，図 4.3 に Affection データセットを用い，シャード数 $N^{\text{shard}} = 3$ のもとで削除ユーザ数 d を変化させた場合の対象ユーザ平均精度の変化を示す．示した範囲において，使用モデルに関わらず PF は常に精度を改善している．これらの結果から，削除ユーザ数が全ユーザ数に対して過度に大きくない範囲であれば，PF に基づき非類似ユーザのデータを学習から除外することが有効であるといえる．表 4.1 および図 4.3 の実験結果は，マシンアンラーニング手法がモデルパーソナライズ化に有効であるという本論文の仮説を支持している．また，これらの精度向上が対象ユーザからのごく少量のデータのみで達成されていることから，知識の選択的除去の有効性が確認できる．

図 4.4 には，PF，CF1，CF2，CF3 の再学習時間削減率と予測精度を示す．この実験では，$N^{\text{shard}} = 3$，$d = 5$，モデルは ViT，データセットには Affection を使用した．予測精度は 5 人の対象ユーザに対する精度の範囲と平均値で示している．計算時間の削減率は，パーソナライズ前の元のモデルを学習するのに要した時間と比較し

た再学習時間の削減量を表す．結果から，CF2 および CF3 の削減率は CF1 とほぼ同程度となっていることがわかる．これは，異なるユーザのデータが多くのシャードに跨って分布しているため，多くのシャードで再学習が必要となっていることを示唆している．一方，PF は他の手法に比べて再学習時間を大幅に削減している．これは，提案手法の第二の貢献である「シャードバイアス」の効果により，再学習が必要なサブモデル数が減少していることを裏付ける結果である．その結果，PF は CF2，CF3 と同じ学習データを用いながら，計算コストを大きく削減した効率的なモデルパーソナライズ化を実現している．

以上を総合すると，提案手法はマシンアンラーニング手法とユーザ間類似度の算出を組み合わせることで，モデル再学習のコストを抑えつつモデルパーソナライズ化を実現できることが示された．

4.2.4 まとめ

本節では，事前学習済みモデルから学習データの影響を効率的に除去できるというマシンアンラーニングの物理的な特性に着目し，データ分割に基づくユーザ適応手法を提案した．提案手法は，ユーザ間の感情反応に基づく類似度により削除すべきデータを選択し，データ分割と部分的な再学習を組み合わせることで，対象ユーザから得られるごく少量のデータのみで事前学習済みモデルを対象ユーザに適応させる．実験により，特定のユーザに対するモデル精度を向上させつつ，完全な再学習と比較して再学習に要する時間を大幅に削減できることを示した．これは，削除対象ユーザのデータが特定のシャードに偏るようにクラスタリングを行うことで，再学習が必要なサブモデル数を最小限に抑えられたことに起因する．

しかしながら，本節で提案した物理的なデータ削除に基づくアプローチには，データの完全な分離が可能であるという前提が必要であり，また類似度の定義が離散的（削除するか否かの二値）にならざるを得ないという側面がある．より柔軟なパーソナライズを実現するためには，モデルのパラメータ空間において，知識の保持と忘却を連続的に制御する仕組みが求められる．次節では，この課題に対処するため，勾配制御に基づく新たなパーソナライズ手法について論じる．

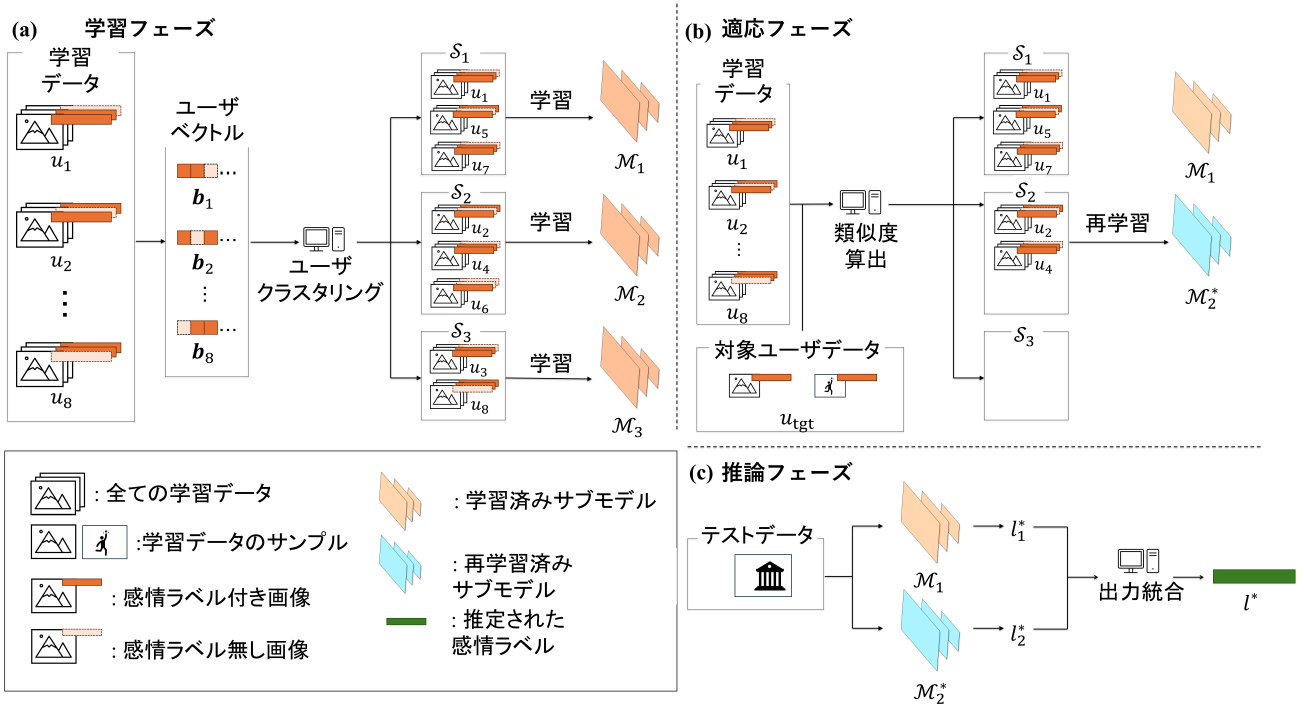

図 4.1: 提案手法の概要図．(a) は学習フェーズ，(b) は適応フェーズ，(c) は推論フェーズを示す．図の例では $N^{\mathrm{usr}} = 8$，$N^{\mathrm{shard}} = 3$，$d = 3$ の場合を表している．

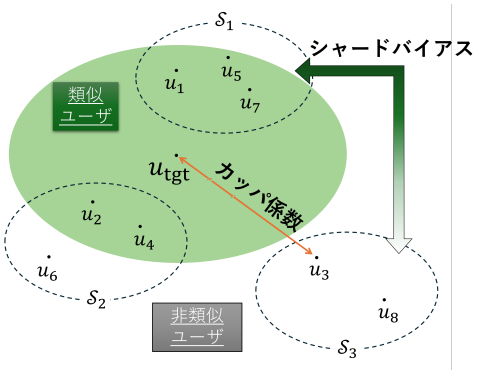

図 4.2: 対象ユーザに類似したユーザを選択する模式図．ユーザ間の類似・非類似の選択は，ユーザ間のカッパ係数に基づいて行われる．同じ類似度指標を用いてユーザクラスタリングに基づくシャーディングを行うことで，類似ユーザと非類似ユーザの間でシャードバイアスが生じる．この例は図 4.1 に示した例と対応している．

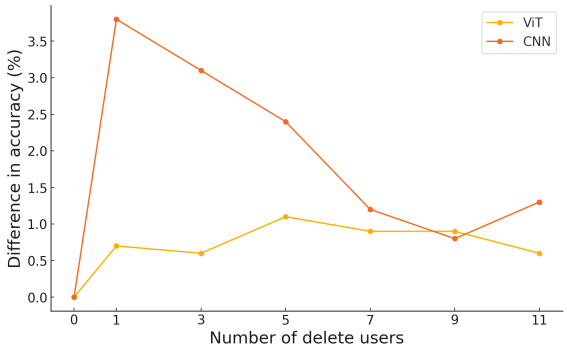

図 4.3: Affection データセットにおいて，$N^{\text{shard}} = 3$ としたときに，削除ユーザ数 d を変化させた場合の対象ユーザ平均精度．

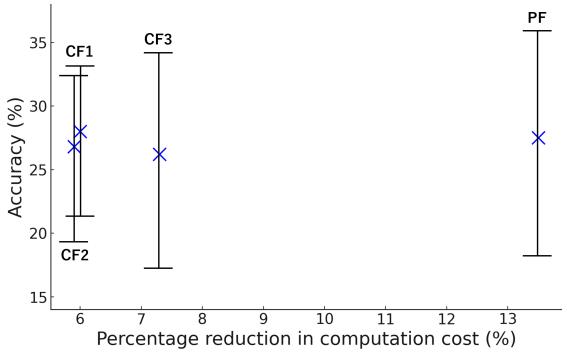

図 4.4: 各手法における精度と計算コスト削減率．精度は対象ユーザ 5 人分の範囲と平均値で示す．$N^{\text{shard}} = 3$，$d = 5$，モデルは ViT，データセットには Affection を使用した．

4.3 類似度に基づき知識を選択的保持および忘却するSimilarity-Aware Unlearningに基づくユーザ適応

4.3.1 はじめに

本節では，第4.1節で述べた第二のアプローチ，すなわちモデル学習時の勾配制御に基づくユーザ適応手法について詳述する．第4.2節で提案したデータ分割型のアプローチは，物理的にデータを遮断することで強力な忘却効果を発揮する一方で，データを離散的に分割する必要があり，類似度の微細なグラデーションを反映しにくいという課題があった．また，複雑な依存関係を持つユーザ間の類似性を，単純なクラスタリングのみで捉えることには限界がある．

そこで本節では，ユーザとアイテムのインタラクションをグラフ構造としてモデル化し，より高度なユーザ表現を獲得するとともに，その類似度に基づいて「知識の保持」と「忘却」を連続的に制御する手法「Similarity-Aware Unlearning」を提案する．本手法は，知識蒸留の枠組みを応用し，類似ユーザからの知識については教師モデル（事前学習済みモデル）の出力を模倣させ（保持），非類似ユーザからの知識については意図的に乖離させる（忘却）ことで，モデルの内部表現を対象ユーザに最適化する．これにより，データの物理的な分割に依存せず，各学習サンプルが持つ，対象ユーザにとっての有用性に応じた柔軟なユーザ適応を実現する．

4.3.2 関連研究

本手法は，モデルのパーソナライズ化とマシンアンラーニングという二つの研究領域の融合領域に位置する．

モデルのパーソナライズ化の課題

従来のパーソナライズ手法は，主に対象ユーザのデータを用いたファインチューニングなどの追加学習戦略に依存している [90]．LoRA [32] などのパラメータ効率の良いファインチューニング（PEFT）や，MAML [33] などのメタラーニングは，少量のデータでの適応能力を向上させてきた．しかし，これらの手法は「新しい知識をいかに効率的に追加するか」に焦点を当てており，事前学習済みモデル内に残存する「対象ユーザと相性の悪い他ユーザの知識（ノイズ）」の影響を明示的に

取り除く機構を持たない．本研究は，知識の注入だけでなく，不要な知識の除去を同時に行うことで，この限界を克服する．

マシンアンラーニングの応用

マシンアンラーニングは，本来プライバシー保護の観点から特定のデータを削除するために発展してきた [25]．近年では，有害なデータ（ポイズニングデータなど）を除去することでモデル性能を向上させる Corrective Machine Unlearning (CMU) [37] という概念も提唱されている．本手法は，この CMU の考え方をパーソナライズに応用し，「対象ユーザと傾向の異なるユーザデータ」を「除去すべき有害なデータ」とみなすことで，アンラーニングを精度向上のための手段として利用する．特に，従来の CMU がデータを「有害か無害か」の二値で扱っていたのに対し，本手法はユーザ類似度という連続値を用いて忘却の度合いを動的に制御する点が新規である．

4.3.3 提案手法

本項では，提案手法の詳細について述べる．全体フレームワークを図 4.5 に示す．本手法は主に二つの段階から構成される．第一に，ユーザとアイテムのインタラクション履歴から異種グラフを構築し，Graph Neural Network (GNN) [91] を用いて各ユーザの嗜好傾向を反映した特徴ベクトルを抽出する．第二に，これらの特徴ベクトルに基づいて算出されたユーザ間類似度を，連続的な制御信号として用い，知識の保持および忘却を安定的に制御する蒸留損失を用いてモデルを再学習する．

グラフベースのユーザ特徴抽出

ユーザ間の嗜好傾向の類似性を捉えるために，全ユーザと全アイテム間の関係をグラフ構造 $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ としてモデル化する．ノード集合 $\mathcal{V}$ はユーザノード $\mathcal{V}_u$ とアイテムノード $\mathcal{V}_x$ から構成され，エッジ $(u_i, x_j) \in \mathcal{E}$ はユーザ u_i がアイテム x_j に対して行ったインタラクション（例：感情ラベルの付与）を表す．

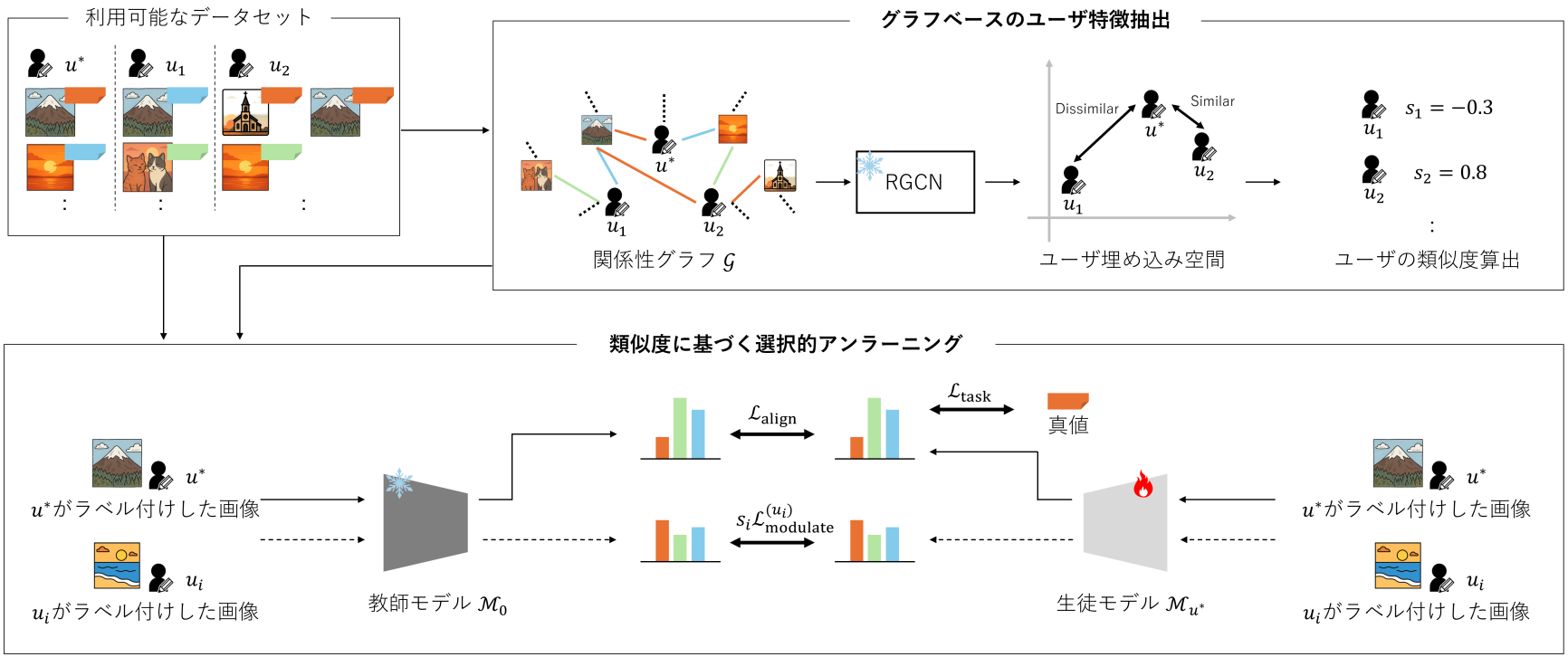

図 4.5: Similarity-Aware Unlearning の全体フレームワーク．本手法は，グラフニューラルネットワークによるユーザ特徴抽出と，ユーザ間類似度を連続的な重みとして用いる安定化蒸留型アンラーニングから構成される．

本研究では，カテゴリカルな関係を扱う分類タスクを対象とするため，関係タイプごとに異なる変換行列を利用する Relational Graph Convolutional Network (R-GCN) [92] を採用する．R-GNN をユーザ-アイテム間のリンク予測タスクとして学習させることで，各ユーザの嗜好傾向を反映した高次元特徴ベクトル（ユーザ埋め込み）$\boldsymbol{h}_u$ を獲得する．

類似度に基づく選択的アンラーニング

前項で得られたユーザ特徴ベクトルを活用し，対象ユーザ u^* に特化したパーソナライズモデル $\mathcal{M}_{u^*}$ を構築する．まず，全ユーザのデータで学習された汎用モデル $\mathcal{M}_0$ を用意し，そのパラメータを生徒モデル $\mathcal{M}_{u^*}$ の初期値として用いる．次に，対象ユーザ u^* と他ユーザ u_i との嗜好の近接性を，対応するユーザ埋め込み間のコサイン類似度として定義する：

$$s_i = \frac{\boldsymbol{h}_{u^*}^\top \boldsymbol{h}_{u_i}}{\|\boldsymbol{h}_{u^*}\|_2 \, \|\boldsymbol{h}_{u_i}\|_2}, \qquad s_i \in [-1, 1] \tag{4.4}$$

本研究では，この類似度スコアを学習過程で更新される変数とはみなさず，事前に推定されたユーザ嗜好の近接性を表す固定の重要度として扱う．

生徒モデルの学習においては，対象ユーザに対する性能維持，汎用モデルに含まれる有益な知識の保持，および非類似ユーザ由来の知識の抑制を同時に達成する必要がある．従来のアンラーニング手法では，類似ユーザと非類似ユーザを二値的に分離し，知識を一括で保持または削除する設計が多く用いられてきた．しかし，実際のユーザ間類似度は連続的に分布しており，このような離散的制御は，学習の不安定化や過剰な性能劣化を招く可能性がある．そこで本研究では，類似度スコアを連続的な重みとして用い，知識保持および忘却の強度を滑らかに制御する損失設計を採用する．

具体的には，類似度の正の成分のみを抽出し，ミニバッチ内で正規化することで，対象ユーザと嗜好が近いサンプルほど学習に強く寄与する重み

$$w_j^+ = \frac{\max(s_j, 0)}{\sum_{k \in \mathcal{B}} \max(s_k, 0)} \tag{4.5}$$

を定義する．ただし，$\mathcal{B}$ をミニバッチ内のサンプル集合とする．この重みを用いて，対象ユーザデータに対する教師あり損失と，汎用モデルとの出力分布の乖離を抑

制する知識整列損失を，それぞれ以下のように定義する：

$$\mathcal{L}_{\text{task}} = \sum_{j \in \mathcal{B}} w_j^+ \, \mathcal{L}_{\text{sup}}(\mathcal{M}_{u^*}(\boldsymbol{x}_j), y_j) \tag{4.6}$$

$$\mathcal{L}_{\text{align}} = \sum_{j \in \mathcal{B}} w_j^+ \, D_{\text{KL}}\Big(p(y|\boldsymbol{x}_j; \mathcal{M}_0) \,\|\, p(y|\boldsymbol{x}_j; \mathcal{M}_{u^*})\Big) \tag{4.7}$$

ただし，$\mathcal{L}_{\text{sup}}(\cdot)$ はタスクに応じた教師あり損失（分類タスクではクロスエントロピー損失）を表し，$p(y|\boldsymbol{x}; \mathcal{M})$ はモデル $\mathcal{M}$ が出力するクラス確率分布である．この設計により，対象ユーザと嗜好が近いユーザ由来の知識ほど強く保持され，一方で寄与の小さい知識は自然に減衰する．

一方，非類似ユーザ由来の知識については，単純に逆符号の損失を与えて強制的に忘却させるのではなく，必要な場合にのみ作用する制約付きの忘却損失を導入する．まず，類似度の負の成分に非線形変換を施すことで，忘却の強度を表す重みを定義する：

$$w_j^- = \begin{cases} (\max(-\tanh(s_j), 0))^2 & \text{if } -\tanh(s_j) > \tau \\ 0 & \text{otherwise} \end{cases} \tag{4.8}$$

ここで τ は忘却を開始するための閾値であり，弱い非類似性に対して不要な忘却が生じることを防ぐために導入される．さらに，生徒モデルと汎用モデルの出力分布が既に十分乖離している場合には，追加の忘却が不要であると考え，KL ダイバージェンスが所定のマージン m を下回る場合にのみ忘却が作用するよう，次の損失項を定義する：

$$\mathcal{L}_{\text{modulate}} = \frac{\sum_{j \in \mathcal{B}} w_j^- \, \max(0, \, m - D_{\text{KL}}(p_0 \,\|\, p_{u^*}))}{\sum_{j \in \mathcal{B}} w_j^-} \tag{4.9}$$

この制約により，不要な知識の過剰な削除を防ぎつつ，非類似ユーザ由来の影響のみを選択的に抑制することが可能となる．

以上を統合し，本研究では，対象ユーザ性能の維持，有益な知識の保持，および非類似知識の抑制を連続的かつ安定的に制御するため，次式で与えられる目的関数を最小化する：

$$\mathcal{L} = \gamma \mathcal{L}_{\text{task}} + \alpha \mathcal{L}_{\text{align}} + \beta \mathcal{L}_{\text{modulate}} \tag{4.10}$$

4.3.4 実験

本節では，提案手法の有効性を，画像に対するユーザの感情推定タスクを通じて検証する．特に，本研究では，汎用モデルや既存のパーソナライズ手法が十分に機能しない状況に着目し，そのような条件下において提案手法がどのような効果を発揮するかを分析する．

実験設定

データセット　データセットには，Affection Dataset [81] を独自にフィルタリングして使用した．具体的に，1,000 件以上の評価履歴を有するユーザかつ 8 人以上のユーザによって評価された画像によるサンプルのみを抽出し，合計 9,443 サンプルからなるデータセットを構築した．ランダムに 14 人の対象ユーザを選択し，それぞれについて，対象ユーザが付与したサンプルのうち 50 サンプルをテストデータとして用い，残りの対象ユーザのサンプルおよび対象ユーザ以外のサンプルを学習および検証に使用した．

バックボーンモデル，評価指標およびハイパーパラメータ　すべての実験において，ImageNet-21k [93] で事前学習された Vision Transformer (ViT) [78] をバックボーンとして用いた．評価指標には，Macro F1-Score を用いることで，クラス不均衡下における識別性能を総合的に評価する．本実験におけるハイパーパラメータは，$\gamma = 2, \alpha = 1, \beta = 2$ とした．

比較手法　提案手法の有効性を検証するため，本研究では，ユーザ適応および関連分野において広く用いられている代表的な学習手法を比較対象として採用した．まず，基礎的な比較対象として，ImageNet 事前学習済みモデルをそのまま評価する Pre-trained，全ユーザの学習データを用いてモデルを更新する Fine-tune (All)，それをさらに対象ユーザのデータのみに基づいて学習を行った Fine-tune (Target)，および初期化状態から対象データを用いて再学習を行う Retrain を用いた．これらの手法は，パーソナライズを行わない場合から，単純な追加学習および再学習までの性能を把握するための基準として位置付けられる．

次に，マシンアンラーニングに基づく手法として，非対象ユーザデータの影響を負の勾配として抑制する NegGrad および NegGrad+ [94]，削除対象サンプルに基づいてモデルの挙動を補正する Corrective Unlearning 系の手法である CF-k および EU-k [37]，ならびに教師モデルとの蒸留を通じて不要な知識の除去を行う SCRUB [38] を比較対象に含めた．これらの手法は，対象外データの影響を低減することを目的としており，知識の除去や抑制という観点において，本研究の問題設定と密接に関連している．

さらに，学習過程そのものを通じて高速な適応を実現する手法として，初期パラメータの更新を最適化する Reptile [95]，および少数サンプル下での分類を目的とする Prototypical Networks (ProtoNet) [96] を採用した．これらの手法は，明示的な知識の除去や抑制を行うことなく，学習戦略を通じてユーザ適応を実現する点に特徴がある．また，パラメータ効率の高い適応手法として，モデル本体を固定したまま低ランク行列を追加する LoRA [32] を比較対象とした．LoRA は計算コストを抑えつつ適応を行うことが可能である一方，既存モデルに内在する不要な知識の影響を直接的に制御することは想定していない．最後に，推論時にモデルを更新するテスト時適応手法として，追加の学習データを必要とせず，テストデータの分布に基づいて適応を行う ZERO [97] を採用した．ZERO は学習段階における知識制御を行わない点で，本研究が対象とする学習時の選択的知識抑制とは異なるアプローチに位置付けられる．

実験結果

Table 4.2 は，Fine-tune (All) における F1-score の昇順にユーザを並べ，各手法の推定精度をユーザ毎に比較した結果を示している．まず注目すべき点として，Fine-tune (All) の性能が低いユーザ群（User A-User D）では，多くの既存手法が十分な改善を示さない一方で，提案手法が各ユーザにおいて最大の F1-score を達成していることが確認できる．具体的に，User A および User B では，Fine-tune (All) の F1-score がそれぞれ 0.0306，0.0340 と極めて低いのに対し，提案手法は 0.1150，0.1302 と大幅な改善を示しており，汎用モデルでは性能が低下するユーザに対して有効に機能していることが分かる．

一方で，中程度の性能を示すユーザ群（User E-User I）においても，提案手法は

表 4.2: 各ユーザ毎のテストデータに対する感情推定結果（F1-score）．なお，一般的な汎用モデル（Fine-tune (All)）の結果の昇順でユーザは並べている．各ユーザにおける最大値を**太字**で示す.

Method	User A	User B	User C	User D	User E	User F	User G	User H	User I	User J	User K	User L	User M	User N
Pre-trained ViT	0.06725	0.1112	0.05518	0.1083	0.09935	0.1021	0.04521	0.04006	0.08182	0.06889	0.1330	0.08182	0.1199	0.2383
Fine-tune (All)	0.03064	0.03399	0.05574	0.07118	0.07250	0.07299	0.09055	0.1210	0.1273	0.1366	0.1413	0.1929	0.2110	0.3812
Retrain	0.04035	0.04021	0.07177	0.09961	0.1265	0.1211	0.05727	0.04298	0.05802	0.1411	0.1067	0.05802	0.04017	0.07292
Fine-tune (Target)	0.03064	0.03399	0.05574	0.07118	0.07250	0.07299	0.09055	0.1210	0.1273	0.1366	0.1413	0.1929	0.2110	0.3812
NegGrad [94]	0.01786	0.04715	0.04816	0.07373	0.06101	0.06974	0.03361	0.1411	0.1478	0.1564	0.1921	0.1153	0.2232	0.4046
NegGrad+ [94]	0.03038	0.03399	0.05574	0.07118	0.07250	0.07299	0.09055	0.1210	0.1334	0.1523	0.1413	0.1803	0.2090	0.3812
CF-k [37]	0.03609	0.03399	0.05574	0.06927	0.06134	0.07299	0.1003	0.1210	0.1400	0.1614	0.1374	**0.1968**	0.2090	0.3938
EU-k [37]	0.03819	0.08571	0.07392	0.04518	0.03084	0.1015	0.1042	0.1432	0.1183	0.1441	0.03971	0.1347	0.1046	0.07328
SCRUB [38]	0.07702	0.02000	0.09548	0.02632	0.04145	0.08408	0.1026	0.08741	0.1173	0.1596	**0.2172**	0.06070	0.2319	**0.4627**
ProtoNet [96]	0.07702	0.02000	0.09548	0.02632	0.04145	0.08408	0.1026	0.08741	0.1173	0.1596	**0.2172**	0.06070	0.2319	**0.4627**
Reptile [95]	0.03064	0.03399	0.05574	0.07118	0.07250	0.07299	0.09055	0.1210	0.1273	0.1366	0.1413	0.1929	0.2110	0.3812
LoRA [32]	0.09646	0.04923	0.09828	0.07623	0.07456	0.1126	0.1043	0.1345	0.1640	**0.1654**	0.1420	0.1706	0.2144	0.4247
ZERO [97]	0.09646	0.04923	0.09828	0.07623	0.07456	0.1126	0.1043	0.1345	0.1640	**0.1654**	0.1420	0.1706	0.2144	0.4247
提案手法	**0.1150**	**0.1302**	**0.1089**	**0.1163**	**0.1360**	**0.1347**	**0.1388**	**0.1444**	**0.1728**	0.1391	0.1921	0.1481	**0.2343**	0.3636

安定して高い F1-score を維持している．具体的に， User G では，既存の NegGrad, CF-k，SCRUB などの手法が 0.10–0.14 程度に留まるのに対し，提案手法は 0.1388 を達成している．この傾向は複数ユーザに共通しており，単に一部のユーザに特化した改善ではなく，ユーザ依存性が中程度の場合においても汎用的に有効であることを示唆している．

さらに，高性能な汎用モデルが既に成立しているユーザ（User M および User N）においては，提案手法が必ずしも最大値を達成しない場合を確認した．具体的に，User N では，SCRUB および ProtoNet が 0.4627 と最も高い F1-score を示しており，提案手法は 0.3636 に留まっている．これは，汎用モデルや既存の蒸留・メタ学習手法が十分に機能するユーザに対しては，提案手法による選択的抑制が必須ではないことを示しており，本手法が，全てのユーザで常に最良となることを前提としていない点は重要である．

以上の結果から，提案手法は，Fine-tune (All) による汎用的な学習では十分な性能を達成することが困難なユーザに対して特に有効であり，そのようなユーザを対象としたパーソナライズにおいて，既存の負勾配法，類似度ベース手法，蒸留・メタ学習手法と比較して一貫した性能向上を実現していることが確認された．このことは，本研究が対象としていた，少数派のユーザに対して汎用モデルが多数決的な推論に陥ってしまうという課題を解決可能であると考える．

4.3.5 まとめ

本節では，ユーザとアイテムのインタラクション履歴に基づく連続的な類似度を活用した，勾配制御型のパーソナライズ手法である Similarity-Aware Unlearning を提案した．本手法は，第 4.2 節の物理的なデータ削除アプローチとは異なり，損失関数を通じて知識の保持と忘却を動的に制御する．実験の結果，感情推定タスクにおいて，提案手法が既存の適応手法やアンラーニング手法を上回る性能を示し，特に，全てのユーザのデータで学習したモデルでは推論精度が低いユーザおける有効性が確認された．

4.4 まとめ

本章では，大規模な学習済みモデルを個々のユーザに適応させるための新たなパラダイムとして，従来の「ターゲットデータの追加（Fine-tuning）」ではなく,「非類似ユーザに由来する不要な知識の削除（Unlearning）」に基づくパーソナライズの枠組みを提案した．

第 4.2 節では，データの物理的な分割に基づく SISA フレームワークを応用した手法を提案した．ここでは，ユーザ間の感情反応の一致度（カッパ係数）に基づくクラスタリングを導入し，類似した傾向を持つユーザを同一のデータシャードに凝集させることで，効率的な忘却を実現した．実験の結果，対象ユーザと非類似なシャードを物理的に再学習プロセスから除外することで，モデルの精度を向上させつつ，再学習コストを大幅に削減できることを示した．このアプローチは，特に計算リソースに制約がある環境や，プライバシー保護の観点からデータの物理的な削除が求められるシナリオにおいて有効な解となる．第 4.3 節では，より柔軟かつ高精度な適応を実現するために，勾配制御に基づく Similarity-Aware Unlearning を提案した．ここでは，物理的な分割の代わりに，ユーザとアイテムのインタラクションに基づくグラフ構造から連続的なユーザ類似度を算出し，知識蒸留の損失関数を通じて知識の「保持」と「忘却」のバランスを動的に制御した．実験の結果，感情推定タスクにおいて，本手法は既存のアンラーニング手法や適応手法（LoRA 等）を上回る性能を達成した．これは，類似度に応じて連続的な重み付けを行うことで，中程度に類似した他ユーザからの有益な知識を保持しつつ，対象ユーザにとって有害なノイズのみを効果的に抑制できたことに起因する．

本論文全体の位置付けとして，第 3 章では「学習段階」においてユーザセントリック特徴を取り込む確率的な特徴統合手法を構築したのに対し，本章では「学習後」のモデルに対してユーザ類似度に基づき知識を選別するアプローチを確立した．前者がユーザの内部状態の表現獲得を，後者が汎用知識からの個人適応を担っており，これらを組み合わせることで，ユーザの多様性とデータの不確実性に一貫して対処可能なシステム構築が可能となる．

第5章

結論

5.1 本研究のまとめ

本論文では，機械学習モデルがユーザの多様な特性に適応し，個々のニーズに合致した出力を提供するための枠組みとして，学習段階における「ユーザセントリック特徴の統合」と，学習済みモデルに対する「他ユーザ由来の知識の選択的保持および忘却」という二つのアプローチに基づくパーソナライズ手法を構築した．従来のパーソナライズ技術は，主として行動履歴や静的属性といった表層的な情報に依存しており，ユーザの意思決定プロセスや認知的な個人差を十分に反映できていないという課題があった．また，近年主流となっている大規模な事前学習済みモデルに対しては，対象ユーザの少量のデータを追加するだけでは，モデル内部に強く残存する他者の知識の影響を解消しきれないという問題が存在した．本研究は，これらの課題に対し，ユーザの内部状態を反映した特徴量の活用と，モデルに含まれる不要な知識の削減という観点から解決を図ったものである．

第 3 章では，学習段階から個人差をモデル構造に取り込むアプローチとして，ユーザセントリック情報とコンテンツ特徴のマルチモーダル統合手法を提案した．具体的には，mGPLVM を核とする確率的生成モデルの枠組みにおいて，連続値特徴量の量子化や半教師あり学習，擬似ラベルの導入を行うことで，ノイズを含みやすく取得コストが高いユーザ動作情報を効果的に統合する手法 (mGPLVF, Semi-OMGP, Semi-MGPPL) を確立した．これにより，従来の履歴依存型手法では捉えきれなかったユーザの反応傾向を潜在空間上で表現し，ラベル不足や未視聴コンテンツといった実環境の制約下でも高精度な関心度推定が可能であることを実証した．

第4章では，学習済みモデルを出発点とし，事後的に個人適応を行うアプローチとして，マシンアンラーニングの概念を応用した「知識の選択的保持および忘却」の枠組みを提案した．ここでは，従来の「知識の追加」を中心とした適応戦略に対し,「対象ユーザにとって不要な知識（非類似ユーザの影響）を選択的に削除する」というアプローチを導入した．具体的には，データ分割と物理的なデータ削除に基づく手法と，勾配制御により連続的に忘却度合いを調整する手法 (Similarity-Aware Unlearning) の二つを提案し，これらが計算コストの削減と適応精度の向上を同時に達成できることを示した．特に，類似度に基づく知識蒸留制御は，主観性の強いタスクにおいて既存の適応手法よりも高い精度を示し，パーソナライズにおける「不要な知識の忘却」の有効性を決定づける成果となった．

以上の検討を通じて，本論文は，ユーザセントリック特徴を用いた学習手法と，他ユーザの知識を選別する適応手法を体系化し，機械学習モデルの学習から運用に至るプロセスにおいて，ユーザの多様性に対処するための基盤技術を確立した．

5.2 今後の課題と展望

本研究で構築した枠組みは，高精度かつ効率的なパーソナライズ技術に向けた重要な指針を与えるものであるが，実社会での広範な展開に向けては，以下の課題と展望が挙げられる．

第一に，**多様なモダリティへの拡張とスケーラビリティの検証**である．本論文では，動作情報や画像特徴を中心に扱ったが，大規模言語モデルやマルチモーダル生成モデルが普及する現在，テキスト，音声，生体信号など，より多様かつ非同期なモダリティを統合するニーズが高まっている．特に，第4章で提案したSimilarity-Aware Unlearningの概念を，大規模な基盤モデルに適用する際には，計算効率とメモリ効率をさらに最適化する技術が求められる．

第二に，**潜在表現の解釈性の向上**である．本研究では，ユーザの個人差を潜在空間上の距離や分布として表現したが，実用システムにおいては「なぜシステムがそのユーザを類似と判断したのか」「なぜそのアイテムを推薦したのか」をユーザに説明できることが信頼構築の鍵となる．構築された潜在空間の意味的な解釈や，アンラーニングによって削除された知識が具体的に何であったかを可視化する技術との統合は，今後の重要な研究テーマである．

第三に，**プライバシ保護や公平性の観点からのモデル制御**である．マシンアンラーニングを用いた本研究のアプローチは，パーソナライズだけでなく，プライバシ保護（忘れられる権利）や公平性の担保にも直接的に寄与する技術である．特定の属性やバイアスを「不要な知識」として定義し，それを意図的に忘却させることで，倫理的に公正なモデルを構築するための制御技術として発展させることが期待される．

最後に，**動的な環境下での継続的なパーソナライズ**である．ユーザの嗜好や状態は時間とともに変化するものであり，一度適応したモデルも，時間の経過とともに精度が低下する可能性がある．オンライン学習の枠組みにおいて，常に最新のユーザ状態を取り込みつつ，過去の重要な記憶は保持し，不要になった古い嗜好のみを忘却する継続的な適応の実現は，本研究の将来的な展望の一つである．

謝辞

本研究は，著者が北海道大学および北海道大学大学院に在学した2019年10月から2026年3月までの約6年間にわたって行ったものである．

本研究に関して，研究遂行のみならず，終始御指導および御鞭撻を頂きました長谷山美紀 教授および小川貴弘 教授に心より感謝申し上げます．加えて，多くの国内・国外学会への参加，論文執筆，および教育活動等，様々な有益な機会を頂けたこと，深く御礼申し上げます．本論文をまとめるにあたり，御助言いただき，また，副査をお引き受けいただいた北海道大学大学院情報科学研究院 情報メディア環境学研究室 土橋宜典 教授に感謝申し上げます．また，本研究の遂行において，多大なる御助力を賜りました北海道大学大学院情報科学研究院メディア創生学研究室 前田圭介 准教授に心より御礼申し上げます．研究活動のみならず公私に渡り様々な御助言いただけたこと，ご多忙の中においても真剣に対応していただきましたこと，感謝申し上げます．本研究室のゼミにおいて有益なフィードバックを提供してくださった北海道大学大学院情報科学研究院メディアダイナミクス研究室 藤後廉 准教授，北海道大学総合IR本部 斉藤直輝 助教，北海道大学 数理・データサイエンス教育研究センター 李広 特任助教に感謝申し上げます．

在学中，多くの御協力を賜りました北海道大学大学院情報科学院メディアダイナミクス研究室の先輩，同輩ならびに後輩学生の皆様に感謝申し上げます．皆様と共に高め合うことで，私は約6年間研究を成し遂げることができました．

最後に，日ごろから私の大学院生活を支えてくれた友人たち，そして遥か遠方，九州・熊本の地から私の選んだ道を見守り続けてくれた家族に，心より感謝の意を表し，本謝辞の結びといたします．

参考文献

[1] S. Wu, F. Sun, W. Zhang, X. Xie, and B. Cui, "Graph neural networks in recommender systems: A survey," *ACM Computing Surveys (CSUR)*, vol. 55, no. 5, pp. 1–37, 2022.

[2] P. V. Rouast, M. T. Adam, and R. Chiong, "Deep learning for human affect recognition: Insights and new developments," *IEEE Transactions on Affective Computing*, vol. 12, no. 2, pp. 524–543, 2019.

[3] P. P. Liang, A. Zadeh, and L.-P. Morency, "Foundations & trends in multimodal machine learning: Principles, challenges, and open questions," *ACM Computing Surveys*, vol. 56, no. 10, pp. 1–42, 2024.

[4] J. Li, A. Waleed, and H. Salam, "A survey on personalized affective computing in human-machine interaction," *arXiv preprint arXiv:2304.00377*, 2023.

[5] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A survey on bias and fairness in machine learning," *ACM Computing Surveys (CSUR)*, vol. 54, no. 6, pp. 1–35, 2021.

[6] C. Gao, Y. Zheng, W. Wang, F. Feng, X. He, and Y. Li, "Causal inference in recommender systems: A survey and future directions," *ACM Transactions on Information Systems*, vol. 42, no. 4, pp. 1–32, 2024.

[7] A. Z. Tan, H. Yu, L. Cui, and Q. Yang, "Towards personalized federated learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 12, pp. 9587–9603, 2022.

[8] N. S. Suhaimi, J. Mountstephens, and J. Teo, "Eeg-based emotion recognition: A state-of-the-art review of current trends and opportunities," *Computational Intelligence and Neuroscience*, vol. 2020, no. 1, pp. 8875426, 2020.

[9] S. Wang, L. Cao, Y. Wang, Q. Z. Sheng, M. A. Orgun, and D. Lian, "A survey on session-based recommender systems," *ACM Computing Surveys (CSUR)*, vol. 54, no. 7, pp. 1–38, 2021.

[10] P. Mateos and A. Bellogín, "A systematic literature review of recent advances on context-aware recommender systems," *Artificial Intelligence Review*, vol. 58, no. 1, pp. 20, 2024.

[11] T. Zhang, A. El Ali, A. Hanjalic, and P. Cesar, "Few-shot learning for fine-grained emotion recognition using physiological signals," *IEEE Transactions on Multimedia*, vol. 25, pp. 3773–3787, 2022.

[12] Y. Shou, T. Meng, W. Ai, F. Fu, N. Yin, and K. Li, "A comprehensive survey on multi-modal conversational emotion recognition with deep learning," *arXiv preprint arXiv:2312.05735*, 2023.

[13] I. B. Adhanom, P. MacNeilage, and E. Folmer, "Eye tracking in virtual reality: A broad review of applications and challenges," *Virtual Reality*, vol. 27, no. 2, pp. 1481–1505, 2023.

[14] M. Ma, J. Ren, L. Zhao, S. Tulyakov, C. Wu, and X. Peng, "Smil: Multimodal learning with severely missing modality," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, vol. 35, pp. 2302–2310.

[15] J. Wang, C. Lan, C. Liu, Y. Ouyang, T. Qin, W. Lu, Y. Chen, W. Zeng, and P. S. Yu, "Generalizing to unseen domains: A survey on domain generalization," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 8, pp. 8052–8072, 2022.

[16] K. Chen, D. Zhang, L. Yao, B. Guo, Z. Yu, and Y. Liu, "Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities," *ACM Computing Surveys (CSUR)*, vol. 54, no. 4, pp. 1–40, 2021.

[17] I. Ahmed, G. Jeon, and F. Piccialli, "From artificial intelligence to explainable artificial intelligence in industry 4.0: A survey on what, how, and where," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 8, pp. 5031–5042, 2022.

[18] V. Skaramagkas, G. Giannakakis, E. Ktistakis, D. Manousos, I. Karatzanis, N. S. Tachos, E. Tripoliti, K. Marias, D. I. Fotiadis, and M. Tsiknakis, "Review of eye tracking metrics

involved in emotional and cognitive processes," *IEEE Reviews in Biomedical Engineering*, vol. 16, pp. 260–277, 2021.

[19] R. Wu, H. Wang, H.-T. Chen, and G. Carneiro, "Deep multimodal learning with missing modality: A survey," *arXiv preprint arXiv:2409.07825*, 2024.

[20] J. Li, J. Li, X. Wang, X. Zhan, and Z. Zeng, "A domain generalization and residual network-based emotion recognition from physiological signals," *Cyborg and Bionic Systems*, vol. 5, no. 0074, 2024.

[21] R. Bommasani, "On the opportunities and risks of foundation models," *arXiv preprint arXiv:2108.07258*, 2021.

[22] Y. Luo, Z. Yang, F. Meng, Y. Li, J. Zhou, and Y. Zhang, "An empirical study of catastrophic forgetting in large language models during continual fine-tuning," *IEEE Transactions on Audio, Speech and Language Processing*, 2025.

[23] T. T. Nguyen, T. T. Huynh, Z. Ren, P. L. Nguyen, A. W.-C. Liew, H. Yin, and Q. V. H. Nguyen, "A survey of machine unlearning," *ACM Transactions on Intelligent Systems and Technology*, vol. 16, no. 5, pp. 1–46, 2025.

[24] Y. Zhang, X. Chen, et al., "Explainable recommendation: A survey and new perspectives," *Foundations and Trends® in Information Retrieval*, vol. 14, no. 1, pp. 1–101, 2020.

[25] L. Bourtoule, V. Chandrasekaran, C. A. Choquette-Choo, H. Jia, A. Travers, B. Zhang, D. Lie, and N. Papernot, "Machine unlearning," in *Proceedings of the IEEE Symposium on Security and Privacy*, 2021, pp. 141–159.

[26] P. Covington, J. Adams, and E. Sargin, "Deep neural networks for YouTube recommendations," in *Proceedings of the ACM Conference on Recommender Systems*, 2016, pp. 191–198.

[27] F. Sun, J. Liu, J. Wu, C. Pei, X. Lin, W. Ou, and P. Jiang, "BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer," in *Proceedings of the ACM International Conference on Information and Knowledge Management*, 2019, pp. 1441–1450.

[28] S. Geng, S. Liu, Z. Fu, Y. Ge, and Y. Zhang, "Recommendation as language processing (RLP): A unified pretrain, personalized prompt & predict paradigm (P5)," in *Proceedings of the ACM Conference on Recommender Systems*, 2022, pp. 299–315.

[29] T. Wang, J.-Y. Zhu, A. Torralba, and A. A. Efros, "Dataset distillation," *arXiv preprint arXiv:1811.10959*, 2018.

[30] J. Frankle and M. Carbin, "The lottery ticket hypothesis: Finding sparse, trainable neural networks," in *Proceedings of the International Conference on Learning Representations*, 2019.

[31] P. W. Koh and P. Liang, "Understanding black-box predictions via influence functions," in *Proceedings of the International Conference on Machine Learning*, 2017, pp. 1885–1894.

[32] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "LoRA: Low-rank adaptation of large language models," in *Proceedings of the International Conference on Learning Representations*, 2022.

[33] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proceedings of International Conference on Machine Learning*, 2017, pp. 1126–1135.

[34] A. Fallah, A. Mokhtari, and A. Ozdaglar, "Personalized federated learning with theoretical guarantees: A model-agnostic meta-learning approach," in *Proceedings of the Advances in Neural Information Processing Systems*, 2020, vol. 33, pp. 3557–3568.

[35] U. Khandelwal, O. Levy, D. Jurafsky, L. Zettlemoyer, and M. Lewis, "Generalization through memorization: Nearest neighbor language models," *arXiv preprint arXiv:1911.00172*, 2019.

[36] H. Yu, A. Gan, K. Zhang, S. Tong, Q. Liu, and Z. Liu, "Evaluation of retrieval-augmented generation: A survey," in *Proceedings of the CCF Conference on Big Data*, 2024, pp. 102–120.

[37] S. Goel, A. Prabhu, P. Torr, P. Kumaraguru, and A. Sanyal, "Corrective machine unlearning," *arXiv preprint arXiv:2402.14015*, 2024.

[38] M. Kurmanji, P. Triantafillou, J. Hayes, and E. Triantafillou, "Towards unbounded machine unlearning," in *Proceedings of the Advances in Neural Information Processing Systems*, 2023, vol. 36, pp. 1957–1987.

[39] Y. Liu, L. Xu, X. Yuan, C. Wang, and B. Li, "The right to be forgotten in federated learning: An efficient realization with rapid retraining," in *Proceedings of the IEEE International Conference on Computer Communications*. IEEE, 2022, pp. 1749–1758.

[40] J. Gantz and D. Reinsel, "The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east," *IDC IView: IDC Analyze The Future 2007*, pp. 1–16, 2012.

[41] M. Haseyama, T. Ogawa, and Y. Nobuyuki, "A review of video retrieval based on image and video semantic understanding," *ITE Transactions on Media Technology and Applications*, vol. 1, no. 1, pp. 2–9, 2013.

[42] Z. Ma, J. Wu, S. Zhong, J. Jiang, and S. J. Heinen, "Human eye movements reveal video frame importance," *Computer*, vol. 52, no. 5, pp. 48–57, 2019.

[43] T. Kushima, S. Takahashi, T. Ogawa, and M. Haseyama, "Interest level estimation based on tensor completion via feature integration for partially paired user ' s behavior and videos," *IEEE Access*, vol. 7, pp. 148576–148585, 2019.

[44] Y. Zhang, G. Zhou, J. Jin, M. Wang, X. Wang, and A. Cichocki, "L1-regularized multi-way canonical correlation analysis for SSVEP-based BCI," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 21, no. 6, pp. 887–896, 2013.

[45] A. Shon, K. Grochow, A. Hertzmann, and R. P. Rao, "Learning shared latent structure for image synthesis and robotic imitation," in *Proceedings of the Advances in Neural Information Processing Systems*, 2005, vol. 18.

[46] W. Huang and R. Y. Da Xu, "Gaussian process latent variable model factorization for context-aware recommender systems," *arXiv preprint arXiv:1912.09593*, 2019.

[47] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "Openpose: Realtime multi-person 2D pose estimation using part affinity fields," *arXiv preprint arXiv:1812.08008*, 2018.

[48] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2818–2826.

[49] N. D. Lawrence, "Gaussian process latent variable models for visualisation of high dimensional data," in *Proceedings of the Advances in Neural Information Processing Systems*, 2004, pp. 329–336.

[50] M. Titsias and N. D. Lawrence, "Bayesian Gaussian process latent variable model," in *Proceedings of the International Conference on Artificial Intelligence and Statistics*, 2010, pp. 844–851.

[51] K. Pearson, "LIII. on lines and planes of closest fit to systems of points in space," *Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 1901.

[52] G. Song, S. Wang, Q. Huang, and Q. Tian, "Multimodal similarity Gaussian process latent variable model," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4168–4181, 2017.

[53] J. Rupnik and J. Shawe-Taylor, "Multi-view canonical correlation analysis," in *Proceedings of the Conference on Data Mining and Data Warehouses*, 2010, pp. 1–4.

[54] G. Andrew, R. Arora, J. Bilmes, and K. Livescu, "Deep canonical correlation analysis," in *Proceedings of the International Conference on Machine Learning*, 2013, pp. 1247–1255.

[55] A. Klami, S. Virtanen, and S. Kaski, "Bayesian canonical correlation analysis," *Journal of Machine Learning Research*, vol. 14, no. Apr, pp. 965–1003, 2013.

[56] M. Nimishakavi, B. Mishra, M. Gupta, and P. Talukdar, "Inductive framework for multi-aspect streaming tensor completion with side information," in *Proceedings of the ACM International Conference on Information and Knowledge Management*, 2018, pp. 307–316.

[57] Z. Zhao and H. Liu, "Semi-supervised feature selection via spectral analysis," in *Proceedings of SIAM International Conference on Data Mining*, 2007, pp. 641–646.

[58] K. Maeda, T. Kushima, S. Takahashi, T. Ogawa, and M. Haseyama, "Estimation of interest levels from behavior features via tensor completion including adaptive similar user selection," *IEEE Access*, vol. 8, pp. 126109–126118, 2020.

[59] N. Lawrence, "Probabilistic non-linear principal component analysis with Gaussian process latent variable models," *Journal of Machine Learning Research*, vol. 6, pp. 1783–1816, 2005.

[60] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*, vol. 4, Springer, 2006.

[61] N. D. Lawrence and J. Quiñonero-Candela, "Local distance preservation in the GP-LVM through back constraints," in *Proceedings of International Conference on Machine Learning*, 2006, pp. 513–520.

[62] R. Urtasun and T. Darrell, "Discriminative Gaussian process latent variable model for classification," in *Proceedings of International Conference on Machine Learning*, 2007, pp. 927–934.

[63] X. Wang, X. Gao, Y. Yuan, D. Tao, and J. Li, "Semi-supervised Gaussian process latent variable model with pairwise constraints," *Neurocomputing*, vol. 73, no. 10-12, pp. 2186–2195, 2010.

[64] K. Maeda, S. Takahashi, T. Ogawa, and M. Haseyama, "Neural network maximizing ordinally supervised multi-view canonical correlation for deterioration level estimation," in *Proceedings of the IEEE International Conference on Image Processing*, 2019, pp. 919–923.

[65] G. Song, S. Wang, Q. Huang, and Q. Tian, "Harmonized multimodal learning with Gaussian process latent variable models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[66] M. F. Møller, "A scaled conjugate gradient algorithm for fast supervised learning," *Neural Networks*, vol. 6, no. 4, pp. 525–533, 1993.

[67] R. Vertegaal, R. Slagter, G. V. d. Veer, and A. Nijholt, "Eye gaze patterns in conversations: There is more to conversational agents than meets the eyes," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2001, pp. 301–308.

[68] J. Li, B. Zhang, and D. Zhang, "Shared autoencoder Gaussian process latent variable model for visual classification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 9, pp. 4272–4286, 2017.

[69] K. Kamikawa, K. Maeda, T. Ogawa, and M. Haseyama, "Feature integration via semi-supervised ordinally multi-modal Gaussian process latent variable model," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 2021, pp. 4130–4134.

[70] J. Wang, C. HQ Ding, S. Chen, C. He, and B. Luo, "Semi-supervised remote sensing image semantic segmentation via consistency regularization and average update of pseudo-label," *Remote Sensing*, vol. 12, no. 21, pp. 3603, 2020.

[71] D.-H. Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Proceedings of the Workshop on Challenges in Representation Learning*, 2013, vol. 3.

[72] M. Seeger, "Gaussian processes for machine learning," *International Journal of Neural Systems*, vol. 14, no. 02, pp. 69–106, 2004.

[73] V. A. Sindagi, R. Yasarla, D. S. Babu, R. V. Babu, and V. M. Patel, "Learning to count in the crowd from limited labeled data," in *Proceedings of the European Conference on Computer Vision*, 2020, pp. 212–229.

[74] C. H. Ek, P. H. Torr, and N. D. Lawrence, "Gaussian process latent variable models for human pose estimation," in *Proceedings of the International Workshop on Machine Learning for Multimodal Interaction*, 2007, pp. 132–143.

[75] M. Matsumoto, K. Maeda, N. Saito, T. Ogawa, and M. Haseyama, "Multi-modal label dequantized Gaussian process latent variable model for ordinal label estimation," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2021, pp. 3985–3989.

[76] J. Gou, B. Yu, S. J. Maybank, and D. Tao, "Knowledge distillation: A survey," *International Journal of Computer Vision*, vol. 129, no. 6, pp. 1789–1819, 2021.

[77] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 2002.

[78] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in *Proceedings of the International Conference on Learning Representations*, 2021.

[79] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 37–46, 1960.

[80] G. M. Foody, "Explaining the unsuitability of the kappa coefficient in the assessment and comparison of the accuracy of thematic maps obtained by image classification," *Remote Sensing of Environment*, vol. 239, pp. 111630, 2020.

[81] P. Achlioptas, M. Ovsjanikov, L. Guibas, and S. Tulyakov, "Affection: Learning affective explanations for real-world visual data," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 6641–6651.

[82] Y. Yang, L. Xu, L. Li, N. Qie, Y. Li, P. Zhang, and Y. Guo, "Personalized image aesthetics assessment with rich attributes," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 19861–19869.

[83] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *Proceedings of the European Conference on Computer Vision*, 2014, pp. 740–755.

[84] R. Krishna, Y. Zhu, O. Groth, J. Johnson, K. Hata, J. Kravitz, S. Chen, Y. Kalantidis, L.-J. Li, D. A. Shamma, et al., "Visual genome: Connecting language and vision using crowdsourced dense image annotations," *International Journal of Computer Vision*, vol. 123, pp. 32–73, 2017.

[85] B. A. Plummer, L. Wang, C. M. Cervantes, J. C. Caicedo, J. Hockenmaier, and S. Lazebnik, "Flickr30k entities: Collecting region-to-phrase correspondences for richer image-to-sentence models," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 2641–2649.

[86] Q. You, J. Luo, H. Jin, and J. Yang, "Building a large scale dataset for image emotion recognition: The fine print and the benchmark," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2016, vol. 30, pp. 308–314.

[87] H.-R. Kim, Y.-S. Kim, S. J. Kim, and I.-K. Lee, "Building emotional machines: Recognizing image emotions through deep neural networks," *IEEE Transactions on Multimedia*, vol. 20, no. 11, pp. 2980–2992, 2018.

[88] K. Han, Y. Wang, H. Chen, X. Chen, J. Guo, Z. Liu, Y. Tang, A. Xiao, C. Xu, Y. Xu, et al., "A survey on vision transformer," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 1, pp. 87–110, 2022.

[89] N. Ketkar and J. Moolayil, "Convolutional neural networks," *Deep Learning with Python: Learn Best Practices of Deep Learning Models with PyTorch*, pp. 197–242, 2021.

[90] Z. Zhang, R. A. Rossi, B. Kveton, Y. Shao, D. Yang, H. Zamani, F. Dernoncourt, J. Barrow, T. Yu, S. Kim, et al., "Personalization of large language models: A survey," *arXiv preprint arXiv:2411.00027*, 2024.

[91] G. Corso, H. Stark, S. Jegelka, T. Jaakkola, and R. Barzilay, "Graph neural networks," *Nature Reviews Methods Primers*, vol. 4, no. 1, pp. 17, 2024.

[92] M. Schlichtkrull, T. N. Kipf, P. Bloem, R. v. d. Berg, I. Titov, and M. Welling, "Modeling relational data with graph convolutional networks," in *Proceedings of the European Semantic Web Conference*. 2018, pp. 593–607, Springer.

[93] T. Ridnik, E. Ben-Baruch, A. Noy, and L. Zelnik-Manor, "Imagenet-21k pretraining for the masses," *arXiv preprint arXiv:2104.10972*, 2021.

[94] D. Choi and D. Na, "Towards machine unlearning benchmarks: Forgetting the personal identities in facial recognition systems," *arXiv preprint arXiv:2311.02240*, 2023.

[95] A. Nichol, J. Achiam, and J. Schulman, "On first-order meta-learning algorithms," *arXiv preprint arXiv:1803.02999*, 2018.

[96] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," in *Proceedings of the Advances in Neural Information Processing Systems*, 2017, vol. 30.

[97] M. Farina, G. Franchi, G. Iacca, M. Mancini, and E. Ricci, "Frustratingly easy test-time adaptation of vision-language models," in *in Proceedings of the Advances in Neural Information Processing Systems*.

著者の研究業績

(A) 学会誌

[**A-1**] K. Kamikawa, K. Maeda, T. Ogawa, and M. Haseyama, "Feature integration through semi-supervised multimodal Gaussian process latent variable model with pseudo-labels for interest level estimation," *IEEE Access* (2024 IF: 3.6), vol. 9, pp. 163843–163850, 2021.

[**A-2**] K. Kamikawa, K. Maeda, R. Togo, T. Ogawa, and M. Haseyama, "インフラ施設の変状の評価を支援する効率的な映像提示に向けた技術者の点検動作分類," 土木学会 *AI・データサイエンス論文集*, vol. 3, no. J2, pp. 811–818, 2022.

(B) 国際学会

[**B-1**] K. Kamikawa, K. Maeda, T. Ogawa, and M. Haseyama, "Interest level estimation based on feature integration considering distribution of partially paired user's behavior, videos and posters," in *Proc. IEEE Global Conference on Consumer Electronics (GCCE)*, pp. 558–559, Kobe, Japan, 2020. (TC: 2)

[**B-2**] K. Kamikawa, K. Maeda, T. Ogawa, and M. Haseyama, "Feature integration via semi-supervised ordinally multi-modal Gaussian process latent variable model," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4120–4124, Toronto, Canada (Online Participation), 2021. (TC: 4)

[**B-3**] K. Kamikawa, K. Maeda, T. Ogawa, and M. Haseyama, "Interest level estimation via multi-modal Gaussian process latent variable factorization," in *Proc. IEEE International Conference on Image Processing (ICIP)*, pp. 1209–1213, Anchorage, USA (Online Participation), 2021. (TC: 0)

[**B-4**] K. Kamikawa, K. Maeda, T. Ogawa, and M. Haseyama, "Interest level estimation using behavior information through multi-view feature integration considering partial and ordered

labels," in *Proc. International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)*, pp. 1035–1037, Phuket, Thailand (Online Participation), 2022. (TC: 0)

[**B-5**] K. Kamikawa, K. Maeda, T. Ogawa, and M. Haseyama, "Feature integration via back-projection ordering multi-modal Gaussian process latent variable model for rating prediction," in *Proc. IEEE International Conference on Image Processing (ICIP)*, pp. 3125–3129, Kuala Lumpur, Malaysia, 2023. (TC: 0)

[**B-6**] Y. Moroto, R. Yanagi, N. Ogawa, K. Kamikawa, K. Sakurai, R. Togo, K. Maeda, T. Ogawa, and M. Haseyama, "Personalized content recommender system via non-verbal interaction using face mesh and facial expression," in *Proc. ACM International Conference on Multimedia (MM)*, pp. 9399–9401, Ottawa, Canada, 2023. (TC: 1)

[**B-7**] S. Liu, K. Kamikawa, K. Maeda, T. Ogawa, and M. Haseyama, "Zero-shot controllable music generation from videos using facial expressions," in *Proc. IEEE Global Conference on Consumer Electronics (GCCE)*, pp. 1189–1190, Kitakyushu, Japan, 2024. (TC: 0)

[**B-8**] S. Liu, K. Kamikawa, K. Maeda, T. Ogawa, and M. Haseyama, "Stabilizing self-consuming training loop on music generation based on music embeddings," in *Proc. IEEE International Conference on Consumer Electronics – Taiwan (ICCE-Taiwan)*, pp. 11–12, Kaohsiung, Taiwan, 2025. (TC: 0)

[**B-9**] S. Liu, K. Kamikawa, K. Maeda, T. Ogawa, and M. Haseyama, "Context-aware image-to-music generation via bridging modalities through musical captions," in *Proc. ACM International Conference on Multimedia (MM)*, pp. 2235–2243, Dublin, Ireland, 2025. (TC: 0)

(C) 技術報告

[**C-1**] K. Kamikawa, K. Maeda, T. Ogawa, and M. Haseyama, "ユーザの動作情報を用いたコンテンツの関心度推定に関する検討-複数ユーザを導入した特徴統合の有効性検証-," 映像情報メディア学会技術報告, vol. 46, no. 6, pp. 103–107, Online, 2022.

[**C-2**] K. Kamikawa, K. Maeda, T. Ogawa, and M. Haseyama, “Feature integration introducing back-projection based on ordering in labels for rating prediction,” 第 26 回 画像の認識・理解シンポジウム (MIRU), Hamamatsu, Japan, 2023.

[**C-3**] K. Kamikawa, K. Maeda, T. Ogawa, and M. Haseyama, “Machine unlearning framework based on aggregation of Gaussian process-based submodels,” 第 27 回 画像の認識・理解シンポジウム (MIRU), Kumamoto, Japan, 2024.

[**C-4**] K. Kamikawa, R. Togo, K. Maeda, T. Ogawa, and M. Haseyama, “物体追跡モデルと Video-LLM の協調利用に基づくタイヤ点検の動作認識に関する検討,” 第 39 回人工知能学会全国大会 (JSAI2025), pp. 1–4, 2025.

(D) 講演

[**D-1**] K. Kamikawa, K. Maeda, T. Ogawa, and M. Haseyama, “m-SimGP を用いた特徴統合に基づくユーザの関心度推定に関する検討,” Proc. 電気・情報関係学会北海道支部連合大会, pp. 83–84, Online, 2020.

(E) 受賞

[**E-1**] IEEE GCCE 2020 Excellent Paper Award Bronze Prize (2020 年 10 月)

[**E-2**] The 2022 IEEE Sapporo Section Encouragement Award (2023 年 1 月)

[**E-3**] 電子情報通信学会北海道支部 学生奨励賞 (2023 年 3 月)

(F) 奨学金・助成金

[**F-1**] JEES・ソフトバンク AI 人材育成奨学金 奨学生 (2021 年 4 月〜2022 年 3 月)

[**F-2**] 日本学術振興会 特別研究員 DC1 (2023 年 4 月〜2026 年 3 月)

[**F-3**] 令和 5 年度 JEES・三菱商事科学技術学生奨学生 (2023 年 4 月〜2026 年 3 月)