数理最適化モデルを利用した
形状制約ノンパラメトリック推定

２０２１年　３月

岩永二郎

数理最適化モデルを利用した
形状制約ノンパラメトリック推定

岩永 二郎

システム情報工学研究群
筑波大学

2021 年 3 月

概要

本論文は，インターネット上のサービスで収集される EC（Electronic Commerce：電子商取引）サイトの閲覧履歴やスマホアプリの検索データを対象に，商品の選択確率や時系列検索確率分布をノンパラメトリックに推定する手法を提案する．特にノイズが多く含まれる大規模データが作る経験分布を対象に数理最適化モデルを利用して形状を制約した推定を行う．ここで，ノンパラメトリックとは特定の関数形を仮定しないことを指す．単調性などの順序構造や凸性凹性，極大値，極小値を持つなどの形状を制限する弱い仮定をすることで経験分布の特徴を活かした柔軟で精度の高い推定を行うことができる．

商品の選択確率を推定する際のノンパラメトリックな手法としてカーネルサポートベクトルマシンがあるが，複雑な仮定をするため過剰適合しやすくハイパーパラメータの調整が必要である．また，時系列確率分布のノンパラメトリックな手法として移動平均やスプライン平滑化，カーネル回帰があるが，分布のノイズに過剰適合しやすい点や滑らかさの仮定のために尖った分布を表現しにくいという課題をもつ．一方，滑らかさの仮定をもたないノンパラメトリックな手法として単調性や単峰性などの形状を制約する単調回帰や単峰回帰の研究がある．本研究はこれらの研究で扱う形状制約を統合的に利用可能な数理最適化モデルを提案し，応用事例を通して有効性を示した．特に経験分布からノイズを除去し，データの特徴を活かした柔軟な推定を可能にする．

本研究は 2 つの数理最適化モデルを利用した形状制約付きのノンパラメトリックな推定手法を提案している．1 つ目は，EC サイトの閲覧履歴を対象に数理最適化モデルを利用して商品の選択確率を推定する方法である．利用者の商品に対する閲覧の最新度と頻度に基づく選択確率を経験分布として求め，単調性，凸性凹性を仮定した最尤推定の問題を非線形計画問題として定式化することで商品の選択確率を推定する．数値実験から，パラメトリックな手法のロジスティック回帰やノンパラメトリックな手法のカーネルサポートベクトルマシンと比較して提案手法の予測性能が良いことが示された．また，利用者の商品選択確率は商品に対する選好度と考えることができるため協調フィルタリングの評価値行列の作成に利用できる．数値実験から，提案手法で推定した選好度を利用することで協調フィルタリングを用いた商品推薦タスクにおいて予測性能が良くなることがわかった．2 つ目は，スマホアプリの検索履歴を対象に数理最適化モデルを利用して時系列確率分布を推定する方法である．単峰性または二峰性の分布で，単調性や裾で検索が無くなるといった制約をもつ混合整数凸二次計画問題に定式化することで時系列確率分布を推定する．実データにおける数値実験から，ノンパラメトリックな手法である経験分布，移動平均，カーネル回帰と比較して提案手法の予測性能が良いことが示された．また，人工データにおける数値実験から提案手法が尖った分布やノイズが含まれる分布に対して頑健な予測ができることが示された．

目次

図目次

表目次

第 1 章

序論

1.1 研究背景

インターネット上の様々なサービスにおいてデータサイエンスの有効性が実証されるようになったのは，データの収集，データの蓄積，データへのアクセス，データの処理などの技術的な発展が背景にある．インターネットの発達により，多くの利用者がネット上で活動するようになった．ホームページやブログなどの静的な情報だけでなく，利用者の閲覧行動や検索行動などの動的な情報もネットを通して収集され，データとして蓄積されている．また，ストレージの発展により大規模データの蓄積が可能となり，クラウドの進歩によりネットに接続できる様々な場所から簡単にデータにアクセスできるようになった．一方でデータの処理技術も発展した．メモリや CPU などのハードだけでなく高速で効率的なアルゴリズムが開発されたことで，大規模データでも実時間内に処理することが可能になった．上記の背景により，データの収集，蓄積，アクセス，処理までが可能となり，インターネット上のサービスにデータサイエンスを応用する準備が整ったのである．しかし，インターネット上のサービスで収集されたデータを分析する際には注意が必要である．インターネット上ではサービス利用者の統制が難しいだけでなく運営側の施策の影響を受けるため多くのノイズが含まれており，その除去は困難である．ノイズが含まれるデータに対して適切な仮定をせずに作成した数理モデルは，実際のサービスに適用するとノイズに過剰適合した分だけ性能の劣化が起きる．

本論文では，データサイエンスの中でも特に数理モデルに注目し，インターネット上のサービスで収集されるようなノイズが含まれるデータへの数理モデルの適用を研究対象とする．

さて，以下では実データに数理モデルを適用する場面でパラメトリックなモデルとノンパラメトリックなモデルの利用を考える．まず，実データにパラメトリックな数理モデルを適用する点について触れたい．インターネット上のサービスで扱うコンテンツや利用者の行動は多種多様であるため，強い仮定をする単純なパラメトリックモデルでは表現力に乏しい．一方，複雑なパラメトリックモデルは，多くの仮定を必要とする場合が多く，実際にデータが仮定を満たすかどうかを確認することは現実的に困難であり，モデル適用の妥当性を検証できない．また，すでに述べたように，複雑なモデルの中にはノイズに過剰適合する場合がある．一般的には，正則化項を入れることで過剰適合を回避するが，過度な正則化はパラメータの増加を招く

だけでなく，数理モデルを実際のサービス上のタスクに適用した際に性能劣化がおきる．また，複雑なモデルは解釈性が乏しい点が問題視されることがある．例えば，意思決定の場面において，数理モデルが算出した結果に明確な説明がなければその結果を採用する心理的な障壁は高い．商品推薦の場面では数理モデルが推薦する商品について「なぜ，この商品が推薦されたのか」がわかると利用者は安心してその商品を購入することができる．このように解釈性が高く，明確に説明できる数理モデルは実務において利点となる．

次に実データにノンパラメトリックな数理モデルを適用する点について触れたい．ノンパラメトリックなモデルの欠点は，データが少ない場合には適切なモデルの構築が困難なことである．一方で，データが多い場合には複雑な仮定をせずとも適切なモデルの構築ができることが利点である．本研究ではノンパラメトリックなモデルの中でもデータの集計処理のみで作成した実績値による分布に注目し，総称して経験分布と呼ぶことにする．ノンパラメトリックなモデルである経験分布はデータが増えるほど真の分布に近づき，パラメトリックなモデルよりも相対的に性能が上昇する．しかし，経験分布は稀にしか起きないケースやノイズに対して過剰適合する問題を回避できない．本研究では数理最適化モデルを利用することで，パラメトリックなモデルが仮定する関数や分布ほど強くない仮定をする．すなわち，単調性や凸性凹性，極大値や極小値の存在に代表される形状を制限する構造を入れることで経験分布を活かしつつ，過剰適合を回避するアプローチをとる．また，領域知識に基づく適切な構造を仮定することで数理モデルのブラックボックス化も解消し，説明力のある数理モデルの構築が可能となる．

研究背景をまとめる．データサイエンスの中でも特に数理モデルに注目し，インターネット上のサービスで収集されるようなノイズが多く含まれるデータへの数理モデルの適用を研究対象とする．パラメトリックな数理モデルは，妥当性，解釈性の問題がある．そこでノンパラメトリックな数理モデルとして経験分布に注目する．データが大規模になるほど経験分布は真の分布に近づくが稀にしか起きない事例やノイズへの過剰適合を回避できない．本研究では，数理最適化モデルを利用して経験分布の形状に制約を課すことで経験分布の特性を活かしつつ過剰適合を回避する．領域知識に基づく適切な仮定をすることで，モデルの妥当性を担保するとともに，解釈性も高めることも可能となる．

さて，本研究は上記の背景のもとで形状制約付きのノンパラメトリックな推定手法を提案する．特に形状制約を統合的に利用可能な数理最適化モデルを提案し，応用事例を通して有効性を示す．具体的には次の 2 つの数理最適化モデルを提案する．1 つ目は，EC サイトの閲覧履歴を対象に数理最適化モデルを利用して商品の選択確率を推定する方法である．利用者の商品に対する閲覧の最新度と頻度に基づく選択確率を経験分布として求め，単調性，凸性凹性を仮定した最尤推定の問題を非線形計画問題として定式化することで選択確率を推定する．2 つ目は，スマホアプリの検索履歴を対象に数理最適化モデルを利用して時系列確率分布を推定する方法である．単峰性または二峰性の分布であり，単調性や裾で検索が無くなるといった制約をもつ混合整数凸二次計画問題に定式化することで時系列確率分布を推定する．

1.2 既存研究

本節では本研究と関わりが深いノンパラメトリック推定の手法について述べる．また，関連する研究分野であるマーケティングと認知心理学との関連についても述べる．関連研究の詳細については各章に委ねる．

1.2.1 ノンパラメトリック推定

統計学において，少ないパラメータや適当な確率分布を仮定する手法をパラメトリックな手法と呼ぶのに対し，パラメトリックではない手法をノンパラメトリックな手法と呼ぶ．そのため，ノンパラメトリックの意味は対象によって異なる．例えば，検定を対象とする場合は，特定の確率分布に依存しない仮説検定のことをノンパラメトリック検定と呼ぶ．また，パラメトリック回帰はなるべく少ない回帰係数で回帰式をデータに適合させることを目的とした回帰である一方で，ノンパラメトリック回帰は回帰係数の数が多いままデータの性質をより適切に反映させた回帰をすることを目的とした回帰である [141]．本研究で提案する手法は特定の関数を仮定せずに予測を行うという点でノンパラメトリックな手法である．関数形を仮定しないノンパラメトリックな手法として単調回帰に代表される形状制約回帰とカーネル法を用いたカーネル回帰がある．

単調回帰（isotonic regression）の研究の歴史は古く，1900 年代中盤に盛んであった．古典的な結果については Barlow の本 [9] にまとめられている．単調回帰は，推定するパラメータ間に大小関係が成り立つ回帰であり，数理モデルの拡張とアルゴリズム開発の研究が中心である．単調回帰の古典的な定式化とは，半順序構造 $(I, \leq)$ と各種パラメータ w_i $(i \in I)$，a_i $(i \in I)$ が与えられたときに次の凸二次計画問題によって x_i $(i \in I)$ を求める問題を指す．

$$\left| \begin{array}{ll} \text{minimize} & \displaystyle\sum_{i \in I} w_i \left(x_i - a_i\right)^2 \\ \text{subject to} & x_i \leq x_{i'} \quad (i, i' \in I, i \leq i') \end{array} \right.$$

ここで，a_i $(i \in I)$ は単調性を示す分布から得られる観測値を表し，w_i $(i \in I)$ は各観測点の重みを表している．具体的な数理モデルの拡張として，パラメータ間の順序関係により定義される単調性に加え，凸性や凹性への拡張や 1 つの峰（極大値）を与える単峰回帰（unimodal regression）への拡張がある．これらは一般に形状を制約したモデル推定として形状制約回帰（shape-restricted regression） [40] にまとめられる．そのため単調回帰は形状制約回帰の特殊な場合に分類される．また，数理モデルの拡張研究として Geyer [37] は目的関数として最小二乗法ではなく最尤法を用いた．単調回帰の研究は 1 次元の例が多いが，Bril et al. [15] によって 2 次元に拡張する研究もあり，さらに多次元に拡張した研究もある [43]．一方，単峰回帰は，Frisén [33] によって研究されており，2 次元に拡張した研究は Geng and Shi による傘型回帰（umbrella orderings） [35] が知られている．傘型回帰でピークが与えられている問題は単調回帰 [15] をサブルーチンとして解くことができるが，Geng and Shi [36] はさらにピー

クの位置を同時に推定するアルゴリズムも開発した．しかし，より一般的に多峰性の分布に対して極大値（ピーク）や極小値を同時に推定する研究はない．また，これらの研究は新しく数理モデルを拡張し，そのアルゴリズムを開発する研究が多い．近年では形状制約回帰を実務に応用する研究が増え，遺伝子の解析や医療，金融などを含む様々な分野で応用研究がされているが [3, 17, 80]，本研究が対象とする EC（Electronic Commerce：電子商取引）サイトの閲覧履歴やスマホアプリの検索データに応用した事例はない．一方，カーネル回帰（kernel regression）は確率変数の条件付き期待値を推定するためのノンパラメトリック回帰の代表的な手法であり，カーネル関数を用いて滑らかな関数を出力する．本研究では比較手法として利用する．

1.2.2　マーケティングにおける RFM 分析

本研究はマーケティングの分野とも関連が深い．マーケティングには RFM 分析 [48, 112] と呼ばれる手法があり，R は Recency（最新購買日）を，F は Frequency（購買頻度）を，M は Monetary（購買金額）の指標を表す．RFM 分析は，各指標の組合せで顧客をグループ化することでそれぞれのグループの性質を分析し，マーケティング施策を講じる手法である．顧客の分類から顧客生涯価値の算出など様々な研究がされている [30, 134–136]．なお，顧客は次の Recency（最新購買日），Frequency（購買頻度），Monetary（購買金額）の指標に従ってグループ化される．

- Recency（最新購買日）: 最終購入日が新しい顧客ほどよい顧客である．
- Frequency（購買頻度）: 購買頻度が多い顧客ほどよい顧客である．
- Monetary（購買金額）: 購買金額が大きいほどよい顧客である．

本研究で用いる商品に対する最新度と頻度は，RFM 分析の Recency と Frequency に相当する．

RFM 分析は顧客のグループ化をするのに対して，本研究は特定の顧客が過去に閲覧した商品をグループ化する．最新度や頻度が高いほど良い顧客／商品であるという点で類似している．しかし，RFM 分析は各グループの性質を分析した上で施策を講じることを目的としているが，本研究はどの商品に興味があるか，すなわち過去に閲覧した商品に対する選好度を定量化し，選好度順に並び替えることを目的としている．

1.2.3　認知心理学における忘却曲線と単純接触効果

本研究は認知心理学とも関連が深い．認知心理学には，忘却曲線と単純接触効果と呼ばれる研究があり，本研究で扱う EC サイトの商品に対する最新度が忘却曲線と，頻度が単純接触効果と類似した概念である．

忘却曲線

忘却曲線（forgetting curve）は，Ebbinghaus [28] により発見されたもので，学習した情報をどれだけ早く忘れるかを示す指数曲線を表す．

本研究で扱う EC サイトの商品に対する最新度は，興味がある商品ほど最近閲覧していることが多いという行動を表しており，因果関係は異なるが忘却曲線と非常に似た性質をもっている．本研究が対象とする EC サイトにおいて商品を最近閲覧するほど購入確率が高くなるが，忘却曲線の考え方は昔に閲覧した商品ほど記憶に残らないことを表す．また，EC サイトでは昔に閲覧した商品ほど購買確率の下降率が低減すること（凸性）と同様に，忘却曲線も指数関数的性質により記憶に残る確率の下降率が低減すること（凸性）が知られている．

単純接触効果

単純接触効果（mere exposure effect）は，認知心理学の分野で 100 年以上も研究されており，特に 1968 年の Zajonc による研究 [128] から活発になり，1980 年の Kunst-Wilson と Zajonc による閾下単純接触効果（subliminal mere exposure effect）の発見 [64] は代表的な研究である．

単純接触効果とは対象への単純な繰り返し接触がその対象に対する好意度を高める現象のことをいう．例えば，楽曲を繰り返し聴いているとその曲をはじめて聴いたときに比べて楽曲への印象が好意的になったり，電車のつり広告で繰り返し見ていたツアー旅行に行きたくなる現象である [139]．一方，閾下単純接触効果は主観的に見えておらず，思い出すこともできない刺激に対してもその対象に対する好意度を高める現象のことをいう．

本研究の EC サイトの商品に対する頻度は，興味がある商品を何度も閲覧するという行動を表しており，因果関係は異なるが単純接触効果と非常に似た性質をもっている．本研究が対象とする EC サイトにおいて商品を閲覧する頻度が多くなるほど購買確率が高くなるが単純接触効果は商品への接触が多くなるほど好意度が高くなることを表す．また，EC サイトでは商品を閲覧する頻度が多くなるほど購買確率の上昇率が低減すること（凹性）と同様に，単純接触効果においても商品への接触が多くなるほど好意度の上昇率が低減すること（凹性）が報告されている [23, 116, 129]．ここで，提案手法で扱う頻度の概念は意識化における刺激のため閾下単純接触効果よりも単純接触効果に近いと言える．

認知心理学における単純接触効果の研究では，単純接触効果研究の方法論，および単純接触効果の説明モデルについて議論されている．本研究はこの文脈で後者の説明モデルについて議論する．

1.3　本論文の構成と概要

本論文は，数理最適化モデルを利用して形状制約を課したノンパラメトリックな推定手法について論じており，実務で得られるデータとタスクによって 2 つのテーマに分けられる（図 1.1）．1 つ目のテーマは EC サイトの閲覧履歴を用いた商品推薦であり，第 2 章と第 3 章

で論じる．第 2 章では EC サイト利用者に対して，過去に閲覧した商品の選択確率を推定し，利用者にとって既知の商品を推薦するタスクである．第 3 章では第 2 章で求めた商品の選択確率を利用した協調フィルタリングアルゴリズムを適用することで，利用者にとって新しい商品を推薦するタスクである．2 つ目のテーマはイベント付き検索データの解析であり，具体的にはスマホアプリの検索履歴を用いて時系列検索確率分布を推定するタスクである．

以下では，本章を除く第 2 章から第 5 章までの概要を説明する．なお，第 2 章「最新度と頻度に基づく商品選択確率の推定」は学術論文 [52, 138] として掲載済み，第 3 章「協調フィルタリングにおける評価値行列の推定」は学術論文 [53] として掲載済みである．

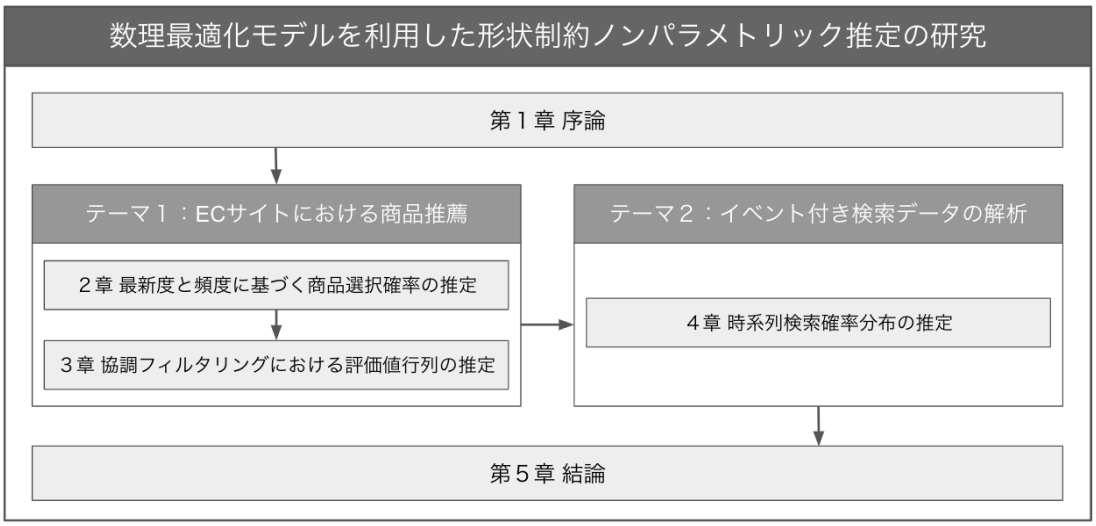


図 1.1. 本論文の構成

1.3.1　第 2 章：最新度と頻度に基づく商品選択確率の推定

第 2 章では，EC サイトにおける商品推薦タスクを扱い，特に利用者にとって既知の商品を推薦するタスクを研究対象とする．本研究の目的は，利用者の商品閲覧行動と商品の選択確率との関係を明らかにすることであり，具体的には EC サイト利用者が過去に閲覧した商品に対する興味を定量化することである．

提案手法は，閲覧商品に対する特徴量として最新度と頻度を算出し，当該商品を購入したかどうかを紐付ける．閲覧履歴に現れる全利用者のデータを集計することで，最新度と頻度の組に対する商品の選択確率を計算できる．すなわち，商品の最新度と頻度に対して選択確率を対応づける 2 次元確率表が作成される．集計によって作成した 2 次元確率表は経験分布であり，データが十分にある場合に真の分布に近づく．しかし，データが少量の場合，ノイズにより過剰適合が起きる．そこで，最新度に対する商品選択確率と頻度に対する商品選択確率に単調性の制約を導入することを検討する．このような回帰は単調回帰 [9] と呼ばれ，特に 2 次元に拡張したモデル [15, 115] や最尤推定モデル [37] が知られている．提案手法は，数理最適化モデルを利用することで，最新度と頻度の 2 次元の特徴量に対して商品選択確率との関係に単調性

と凸性凹性の制約を入れた上で最尤推定を行なう．非線形計画問題として定式化されており，特定の関数を仮定しないノンパラメトリックな推定手法である．また，最新度と頻度による交互作用を完全に表現できる点で柔軟な予測モデルである．単調回帰の応用は様々知られているが，EC のサイトの閲覧履歴に応用した事例は知られていない．数値実験からパラメトリックな手法であるロジスティック回帰とノンパラメトリックな手法であるカーネルサポートベクトルマシンと比較して提案手法の予測性能が良いことを示す．また，データが少量である場合にも単調性と凸性凹性による形状制約が正則化として効果を発揮し，提案手法の予測性能が良いことを示す．

1.3.2 第 3 章：協調フィルタリングにおける評価値行列の推定

第 3 章では，EC サイトにおける商品推薦タスクを扱い，特に利用者にとって新しい商品を推薦するタスクを研究対象とする．本研究の目的は，利用者の商品に対する評価（特に未評価）を高い精度で予測することを通して商品推薦タスクの精度改善に繋げることである．

EC サイト利用者に新しい商品を推薦する代表的なアルゴリズムに協調フィルタリングがある [1, 2, 12, 29]．協調フィルタリングの実装では，利用者の商品への選好を表す利用者-商品間の評価値行列を必要とする．選好を表すデータは，利用者から明示的に取得する方法と暗黙的に取得する方法があり [55, 72]，後者はノイズが多く [5, 96]，商品推薦の精度の低下に繋がることが実証されている [6]．利用者の選好データを暗黙的に取得する場合，商品の閲覧数（本研究では頻度に対応），または経過時間（本研究では最新度に対応）が採用されており，どちらか一方を利用して評価値行列を作成する方法が一般的である．本研究では 2 章で提案する商品の最新度と頻度に関する 2 次元確率表を用いて評価値行列を作成する方法を提案する．提案手法は，最新度と頻度を同時に利用するだけでなく，最新度と頻度の交互作用を完全に表現する評価値行列を作成する点で新しい手法である．数値実験では，次の 4 つの方法で評価値行列を作成し，協調フィルタリングによる商品推薦タスクの予測性能を比較する．4 つの評価値行列作成方法とは，最新度のみで作成する方法，頻度のみで作成する方法，2 次元確率表の経験分布から作成する方法，数理最適化モデルを利用してノイズを除去した 2 次元確率表から作成する提案方法である．利用者間型，および非負値行列分解を用いた協調フィルタリングによる商品推薦タスクで予測性能を比較し，提案手法の予測性能が最も良いことを示す．また，データ量が少量の場合にも提案手法の予測性能が良いことを示す．

1.3.3 第 4 章：時系列検索確率分布の推定

第 4 章では，イベント付き検索データの解析をテーマとし，特に出産イベントの前後においてスマホアプリで検索したデータを用いた時系列確率分布の推定を行う．本研究の目的は，精度の高い時系列確率分布を推定するとともに，推定した確率分布を用いて確率分布間の解析をすることである．

研究対象とする時系列検索確率分布は，単峰性か二峰性の分布となっており，部分的な単調

性を持ち，分布の裾で検索がなくなるという特徴がある．単峰性の分布を推定する手法として単峰回帰がある．特に単峰性のピークも同時に推定する研究は Geng and Shi [36] による研究があるが，多峰性の分布に対して極大値や極小値を同時に推定する研究はない．本研究では数理最適化モデルを利用することで，単峰性，および二峰性の極大値（ピーク）や極小値を自動で特定し，単峰性や二峰性を満たすように時系列確率分布を推定する方法を提案する．提案手法は，従来の研究と異なり数理最適化モデルを利用することで多峰性の分布に対して極大値，および極小値となる時点も同時に推定する新しい手法である．数値実験では実データと人工データを用いて提案手法を評価する．実データを用いた実験では，経験分布としての時系列確率分布，移動平均，カーネル回帰を用いて推定した時系列確率分布と比較して提案手法は推定誤差が小さいことを示す．一方，人工データを用いた実験では提案手法が尖った分布やノイズが混ざっている分布の予測に対して頑健であることを示す．

1.3.4　第 5 章：結論

第 5 章では，本論文の主要な結果をまとめるとともに，数理最適化モデルを利用した形状制約ノンパラメトリック推定の注意点について触れる．データ量やノイズを考慮しつつ，適切で過不足がない仮定を数理モデルに課すことが重要であることについて述べる．また，第 3 章，第 4 章の研究の社会実装の方法についても解説をする．具体的には EC サイトにおける商品推薦の仕組みの実装方法とその実用上の有効性について述べる．最後に今後の展望として，他の分野への応用，形状制約の表現可能性，形状制約をもつ関数を用いた解析について触れる．

第 2 章

最新度と頻度に基づく商品選択確率の推定

2.1 はじめに

インターネットを利用して製品やサービスを提供する EC サイトを運営する企業が増えている [121]．顧客は実店舗に足を運ぶことなく，EC サイトで様々な商品を比較検討して購買をする．一方で企業は EC サイトに蓄積された詳細なデータを利用して顧客との関係構築を行っている．特に，利用者のページ閲覧（PV）を含むクリックストリームデータは，利用者の行動理解をするために有益であることが実証されている [16, 49, 54, 85, 89]．

本研究の目的は，利用者の閲覧行動と商品選択確率との関係を明らかにすることである．具体的には，クリックストリームデータから得られる商品閲覧履歴を用いて，利用者が過去に閲覧した商品に対する興味を定量化する．利用者の商品に対する選好度合いを理解することで，利用者の嗜好に合わせた商品の推薦をすることが可能となる．また，利用者が過去に閲覧した商品に対する選択確率がわかるため在庫管理における需要予測にも有効である [47]．

統計的推定方法は，主にパラメトリックな手法とノンパラメトリックな手法に分類される．パラメトリックな手法は推定モデルを強制的にパラメトリックな関数に当てはめるため，利用者の閲覧と商品選択確率の関係を表現する際の自由度は小さい．一方，ノンパラメトリックな手法は特定のパラメトリックな関数を仮定しないため，利用者の閲覧と商品選択確率の関係を高い自由度で表現できる．

本研究は，商品の選択確率を推定するノンパラメトリックな新しい手法を提案する．利用者の過去の購買行動における商品に対する最新度と頻度は，反復購買を予測するために重要な指標であることが実証されている [31, 32, 56, 101, 102]．これらの事実を考慮して，本研究では利用者の商品閲覧履歴における商品の最新度と頻度を商品選択確率推定のための特徴量として採用する．

具体的には利用者の商品に対する最新度と頻度の組に対して商品選択確率を計算し，「2 次元確率表」を作成する．このアプローチは商品閲覧の最新度と頻度の間にある交互作用を完全に表現できるが，一方で少数の学習データから推定された確率には大きな推定誤差が含まれ

る．この問題を解消するために商品閲覧の最新度と頻度の性質を利用して商品選択確率を最尤推定する．具体的には，最新度と頻度がもつ単調性，および凸性凹性の性質を満たすように商品選択確率を推定する最適化モデルを利用する．

提案手法の予測性能を評価するため，2 値分類器の一般的なモデルであるロジスティック回帰とカーネルベースのサポートベクトルマシン（カーネル SVM）を用いて提案手法と比較した．数値実験では，最新度（3 種類）と頻度（3 種類）の具体的な特徴量を構築し，これらの特徴量の組合せの有効性を検証した．

提案する予測モデルの利点は次のように要約される:

安定性　商品閲覧の最新度と頻度の性質を利用することで小規模な学習データでも提案手法は高い予測性能を持つ．数値実験から学習データが少ない場合に極めて有効であることが示された．

柔軟性　提案するノンパラメトリックな予測モデルは，ロジスティック回帰のような多くのパラメトリックな予測モデルと対照的に，利用者の商品閲覧と商品選択確率との間の関係を交互作用により柔軟に表現する．実際，数値実験でパラメトリックな予測モデルよりも高い性能を示した．

拡張性　2 次元確率表のサイズが学習データのサイズに依存せず一定であるため，学習コストがデータサイズに依存しないという点で拡張性が高い．そのため数値実験では大規模データを利用して学習することができた．一方，カーネル SVM は学習データのサイズに計算負荷が依存するため小規模なデータセットにしか適用できなかった．

本章の構成を説明する．本節では，本研究の背景と要約を述べた．2.2 節では関連研究について説明する．2.3 節では商品選択確率を推定するための最適化モデル，すなわち提案手法について述べる．2.4 節では数値実験を通して提案手法の有効性を評価する．2.5 節で本章のまとめと今後の研究課題について述べる．

2.2　関連研究

クリックストリームデータに関する研究で最も活発な分野の 1 つに，EC サイトにおけるオンライン購買行動の分析がある．この目的のために Moe and Fader [84] は各利用者の訪問と購買の履歴を観察することで購買確率を予測する確率モデルを提案した．他の多くの研究ではロジットモデルやプロビットモデルを用いて，様々なタイプの変数を入力としてオンライン購買行動を予測している [86, 95, 111, 114, 122]．一方，Boroujerdi et al. [13] は決定木，サポートベクトルマシンなどの異なる分類アルゴリズムを適用しており，利用者の購買意向を予測している．しかしながら，これらの研究は購買に繋がる利用者の訪問の予測に焦点を当てており，各商品の購買確率まではわからない．

オンライン商品選択行動を分析する研究には，クリックストリームデータではなく，他の詳細なデータに重点を置いているものが多い．例えば，Chen and Fan [21] は，静的なデータ，時系列データ，記号列データ，文脈データなどの複数のデータを扱い，マルチカーネルサポー

トベクトルマシンの改善を行った．Zhang and Pennacchiotti [132] は，ソーシャルメディアのプロファイルを利用して機械学習モデルを構築し，利用者の選択する商品のカテゴリの予測を行った．Qiu [100] は，商品のレビューと評価の予測にサポートベクトル回帰モデルを利用した．本研究は利用者のクリックストリームデータから得られる利用者の商品閲覧と商品選択確率の関係に注目する．

推薦システムは利用者にとって有益な商品を推薦するソフトウェアや技術を指す [105]．一般的な推薦システムのアルゴリズムである協調フィルタリング [29, 103] は，他の利用者の嗜好に基づいて商品を推薦する．推薦システムの目的の 1 つは，利用者にとって未知の商品を推薦することで価値ある商品を発見することにあるが，本研究では利用者が過去に閲覧した商品，すなわち既知の商品を推薦することに注意されたい．利用者にとって既知の商品が推薦されたとしても興味のある商品であれば購買される可能性は高く，EC サイトの様々な場面で利用することができる．例えば，利用者が EC サイトに再訪問した際に，過去に閲覧したことのある商品ページに誘導して購買に繋げたり，利用者が購買したタイミングで「ついで買い」を誘発させることもできる．本研究では，各利用者が過去に閲覧した商品を対象に，それぞれの商品を選択（購買）する確率を予測する．推定する商品選択確率は利用者の商品に対する選好を表現するため，協調フィルタリングにおける利用者の商品に対する評価値行列に利用できる．すなわち，提案手法を協調フィルタリングに応用することで，利用者が過去に閲覧したことがない，未知の商品を推薦することに繋げることができる．協調フィルタリングへの応用は 3 章で論じる．

商品選択確率を推定する最適化モデルは，クリックストリームデータの分析に単調回帰と凸性凹性制約を応用した形状制約回帰の新しい手法である．従来の単調回帰は変数列に単調性のみを課すものであり，強多項式時間のアルゴリズム [82, 108] を含む様々なアルゴリズムがあることが知られている [10, 27]．現在，最良の多次元アルゴリズムは Stout [118] によるもので，L_1，L_2，または L_∞ のいずれかの距離尺度で利用できる．一方，与えられたデータセットを柔軟に扱うために様々な代替案が議論されてきた．単調性の代替案としては凸性凹性による制約の研究がある [26, 45]．また，統計的性質を考慮して Geyer [37] は単調回帰に最尤推定を用いた．形状制約回帰の研究は，Robertson et al. [107] の本にアルゴリズムの概要が書かれている．しかしながら形状制約回帰の研究において単調性，凸性凹性の制約を課した上で最尤推定法を実装した研究はない．さらに，数理最適化モデルは従来の形状制約回帰の拡張としての価値をもつ．

2.3 提案手法

本節では，クリックストリームデータから個々の利用者の商品選択確率を予測する方法について解説する．本手法は利用者の商品閲覧履歴から計算される最新度と頻度の 2 つの特徴量を用いる．本研究で解くべき問題は，利用者と商品の間に計算される最新度と頻度から利用者の商品選択確率を推定することである．

2.3.1　最新度と頻度の算出方法

各利用者の過去の商品閲覧から最新度と頻度を特徴量として計算し，商品選択確率を推定する．最新度と頻度の概念を簡単に説明する．利用者 u に対する商品 v の最新度とは，利用者 u が最後に商品 v を閲覧したのがどのくらい最近であるか，すなわち予測時点から最終閲覧までの時間距離に反比例する「新しさ」を表現している．一方，利用者 u に対する商品 v の頻度とは，利用者 u が商品 v にどのくらい接触したか，すなわち，予測時点までの「接触度合い」を表現している．

最新度と頻度の具体的な算出方法について説明する．表 2.1 はある利用者の閲覧履歴の要約である．閲覧履歴は 3 月 1 日〜3 月 3 日までの期間の商品 1〜商品 5 の閲覧数が記録されており，閲覧履歴の要約は基準日を 3 月 4 日とした場合の各商品の最新度と頻度に対応する．

表 2.1. ある利用者の閲覧履歴の要約

	閲覧数				
商品	3 月 1 日	3 月 2 日	3 月 3 日	最新度	頻度
商品 1	2	1	1	3	4
商品 2	0	2	0	2	2
商品 3	1	0	0	1	1
商品 4	0	1	0	2	1
商品 5	1	1	0	2	2

頻度は集計により算出することができる．商品 1 を 3 月 1 日に 2 回，3 月 2 日に 1 回，3 月 1 日に 1 回閲覧しており，期間内で 4 回閲覧していることから頻度は 4 と算出される．一方，最新度は基準日と閲覧履歴が対象とする期間を考慮して算出される．3 月 4 日を基準日とすると 3 月 1 日から 3 月 3 日までの 3 期間を対象とするため，最終閲覧が 3 月 1 日の場合は最新度が 1，最終閲覧が 3 月 2 日の場合は最新度が 2，最終閲覧が 3 月 3 日の場合は最新度が 3 となる．そのため，3 月 3 日に最終閲覧した商品 1 は最新度が 3 で，3 月 2 日に最終閲覧した商品 2 は最新度が 2 と計算される．

2.3.2　2 次元確率表

連続する自然数の有限部分集合として最新度の集合を $R = \{1, 2, 3, \ldots\}$，頻度の集合を $F = \{1, 2, 3, \ldots\}$ とおく．最新度と頻度の間の交互作用をノンパラメトリックに表現するために次の 2 次元確率表，

$$\boldsymbol{P} = \left[\, p_{ij} \,\right]_{(i,j)\in R\times F}$$

を定める．各成分 p_{ij} は最新度 i と頻度 j のときの商品選択確率を示す．

閲覧履歴の要約を集計することで商品選択確率を計算することができる．まず，データセッ

トにおいて適当な基準日と期間を定義する．それぞれの $(i,j) \in R \times F$ について，基準日と期間内の閲覧履歴を要約し，最新度が i で頻度が j となる利用者と商品の組の件数を数え上げて n_{ij} とする．同様に最新度が i で頻度が j となる利用者と商品の組のうち，基準日にその商品を購買した件数を数え上げて q_{ij} とする．このとき，各 $(i,j) \in R \times F$ について $n_{ij} > 0$ の場合，商品選択確率を直接計算でき，経験分布としての 2 次元確率表を作成することができる．

$$\bar{\boldsymbol{P}} = \left[\bar{p}_{ij} := \frac{q_{ij}}{n_{ij}} \right]_{(i,j) \in R \times F} \tag{2.1}$$

ここで，便宜上 $n_{ij} = 0$ の場合は $\bar{p}_{ij} := 0$ と定める．また，n_{ij} は $(i,j) \in R \times F$ において確率を計算するためのデータ数を表している．

商品選択確率の具体的な算出方法について説明する．表 2.2 は利用者 1〜利用者 3 の閲覧履歴の要約と各商品を基準日に購買したかどうかのフラグを表している．

表 2.2. 利用者 1〜利用者 3 の閲覧履歴の要約と購買フラグ

利用者	商品	最新度	頻度	購買フラグ
利用者 1	商品 1	3	4	0
	商品 2	2	2	1
	商品 3	1	1	1
	商品 4	2	1	0
	商品 5	2	2	0
利用者 2	商品 6	1	3	0
	商品 7	3	3	1
	商品 8	1	1	0
	商品 9	3	2	1
利用者 3	商品 10	1	1	0
	商品 11	3	4	1
	商品 12	1	3	0

このとき，経験分布による 2 次元確率表は表 2.3 のように与えられる．経験分布による 2 次元確率表はデータ数 n_{ij} が十分にある $(i,j) \in R \times F$ に対しては $\bar{p}_{ij}$ の信頼性は高いが，一方でデータ数 n_{ij} が極端に少ない $(i,j) \in R \times F$ に対しては $\bar{p}_{ij}$ の信頼性は低く，計算される商品選択確率は大きな誤差を含む．それゆえ最新度と頻度の性質を利用することで，より信頼性の高い 2 次元確率表を作成する必要がある．

表 2.3. 経験分布による 2 次元確率表

	頻度			
最新度	1	2	3	4
1	$1/3 = 0.33$	0	$0/2 = 0$	0
2	$0/1 = 0$	$1/2 = 0.50$	0	0
3	0	$1/1 = 1.00$	$1/1 = 1.00$	$1/2 = 0.50$

2.3.3　最新度と頻度の性質

2 次元確率表に対応する決定変数を

$$\boldsymbol{X} = \left[\ x_{ij}\ \right]_{(i,j)\in R\times F}$$

と定める．

まず，商品選択確率に単調性制約を課すことを検討する．Fader et al. [31] は，過去の購買の最新度と頻度が未来の購買と正の相関があることを示している．言い換えると，頻度の値を固定した場合，最新度が大きくなると商品選択確率も高くなり，同様に，最新度の値を固定した場合，頻度が大きくなると商品選択確率も高くなる．すなわち，商品選択確率に次の単調性制約を課すことは妥当である．

$$x_{ij} \le x_{i+1,j} \quad ((i,j) \in R \times F,\ i \le |R| - 1) \tag{2.2}$$
$$x_{ij} \le x_{i,j+1} \quad ((i,j) \in R \times F,\ j \le |F| - 1) \tag{2.3}$$

次に商品選択確率に凸性凹性制約を課すことを検討する．利用者の商品閲覧を遡ることを考えると，最新度が小さくなればなるほど商品選択確率への影響は小さくなる．言い換えると頻度の値を固定した場合，最新度の値が大きくなるほど商品選択確率の増分が大きくなる．また，利用者の商品閲覧が増えていくことを考えると，頻度が大きくなればなるほど商品選択確率への影響は小さくなる．つまり，最新度の値を固定した場合，頻度の値が大きくなるほど商品選択確率の増分は小さくなる．よって最新度と商品選択確率の関係，および頻度と商品選択確率の関係に次の制約を課すことができる．

$$x_{i+1,j} - x_{i,j} \le x_{i+2,j} - x_{i+1,j} \quad ((i,j) \in R \times F,\ i \le |R| - 2) \tag{2.4}$$
$$x_{i,j+1} - x_{ij} \ge x_{i,j+2} - x_{i,j+1} \quad ((i,j) \in R \times F,\ j \le |F| - 2) \tag{2.5}$$

上記の制約は凸関数と凹関数の 2 次の条件に相当するため [14]，凸性凹性制約と呼ぶ．

2.3.4　最適化モデル

ここでは，最新度と頻度の性質を満たす 2 次元確率表 $\boldsymbol{P}$ を作成するための最適化モデルを定義する．

各 $(i,j) \in R \times F$ について商品選択確率 x_{ij} を与え，n_{ij} 回の商品閲覧のもと商品選択した回数を q_{ij} とすると，二項分布から独立にデータが発生する確率は

$$\binom{n_{ij}}{q_{ij}} (x_{ij})^{q_{ij}} (1 - x_{ij})^{n_{ij} - q_{ij}}$$

と表せる．よって，対数尤度関数は次のようになる．

$$\begin{aligned} &\log \left(\prod_{(i,j) \in R \times F} \binom{n_{ij}}{q_{ij}} (x_{ij})^{q_{ij}} (1 - x_{ij})^{n_{ij} - q_{ij}} \right) \\ &= \sum_{(i,j) \in R \times F} \left(\log \binom{n_{ij}}{q_{ij}} + q_{ij} \log x_{ij} + (n_{ij} - q_{ij}) \log(1 - x_{ij}) \right) \end{aligned} \tag{2.6}$$

上記より単調性制約 (2.2) と (2.3) のもと，定数項を削除した対数尤度関数 (2.6) を最大化する最適化モデルは次のようになる．

$$\left| \begin{array}{ll} \text{maximize} & \displaystyle\sum_{(i,j) \in R \times F} (q_{ij} \log x_{ij} + (n_{ij} - q_{ij}) \log(1 - x_{ij})) \\ \text{subject to} & x_{ij} \leq x_{i+1,j} \quad ((i,j) \in R \times F,\ i \leq |R| - 1) \\ & x_{ij} \leq x_{i,j+1} \quad ((i,j) \in R \times F,\ j \leq |F| - 1) \\ & 0 < x_{ij} < 1 \quad ((i,j) \in R \times F) \end{array} \right. \tag{2.7}$$

単調性制約に加えて，さらに凸性凹性制約 (2.4)，(2.5) を加えた最適化モデルは次のようになる．

$$\left| \begin{array}{ll} \text{maximize} & \displaystyle\sum_{(i,j) \in R \times F} (q_{ij} \log x_{ij} + (n_{ij} - q_{ij}) \log(1 - x_{ij})) \\ \text{subject to} & x_{ij} \leq x_{i+1,j} \quad ((i,j) \in R \times F,\ i \leq |R| - 1) \\ & x_{ij} \leq x_{i,j+1} \quad ((i,j) \in R \times F,\ j \leq |F| - 1) \\ & x_{i+1,j} - x_{i,j} \leq x_{i+2,j} - x_{i+1,j} \quad ((i,j) \in R \times F,\ i \leq |R| - 2) \\ & x_{i,j+1} - x_{ij} \geq x_{i,j+2} - x_{i,j+1} \quad ((i,j) \in R \times F,\ j \leq |F| - 2) \\ & 0 < x_{ij} < 1 \quad ((i,j) \in R \times F) \end{array} \right. \tag{2.8}$$

上記 2 つの最適化モデルは線形制約のもとで凹関数を最大化する数理最適化問題となる．そのため，標準的な非線形最適化ソルバーで効率的に最適解を求めることができる．

また，データ数 n_{ij} が大きくなるほど経験分布として得られる商品選択確率 $\bar{p}_{ij}$ の信頼性が高くなることに注目すれば，加重残差平方和

$$\sum_{(i,j) \in R \times F} n_{ij} (x_{ij} - \bar{p}_{ij})^2$$

の最小化に目的関数を置き替えることも考えられる．さらに，よりデータ数への加重を与えたい場合には n_{ij} を n_{ij}^2 に書き換えることもできる．加重残差平方和の最小化は，凸二次計画問題となるためより効率的に解くことができる．そのため非線形最適化ソルバーよりも多く存在する凸二次計画ソルバーで解くことができ，実務上有効である．

ここで $n_{ij} = 0$ となる $(i,j) \in R \times F$ が存在する場合に触れておく．このとき，$n_{ij} = q_{ij} = 0$ であるため目的関数から x_{ij} の項が除外される．その結果，x_{ij} の周辺の $x_{i-1,j}, x_{i+1,j}, x_{i,j-1}, x_{i,j+1}$ の値と制約 (2.2)–(2.5) によって x_{ij} の値が補完される．

以上の最適化モデルの最適解を用いて 2 次元確率表 $\boldsymbol{P}$ は再構成される．すなわち，最適化問題 (2.7)，または (2.8) の最適解を x^*_{ij} とすると，各 $(i,j) \in R \times F$ について $p_{ij} := x^*_{ij}$ と定義することで 2 次元確率表 $\boldsymbol{P}$ を構築できる．

表 2.4 は表 2.3 と同じデータから単調性モデル (2.7) を用いて生成した 2 次元確率表である．表 2.1， 2.4 を参照すると，利用者 1 について商品 1 は最新度 3，頻度 4 であり商品選択確率は 0.75，商品 2 は最新度 2，頻度 2 であり商品選択確率は 0.5，商品 3 は最新度 1，頻度 1 であり商品選択確率は 0.17 となっている．

表 2.4. 単調性制約を課した最適化モデルから作成した 2 次元確率表

	頻度			
最新度	1	2	3	4
1	0.17	0.17	0.17	0.44
2	0.17	0.50	0.57	0.67
3	0.47	0.75	0.75	0.75

2.4　数値実験

本節では提案手法の予測性能をロジスティック回帰モデルとカーネル SVM モデルと比較する．

2.4.1　データセット

本実験では EC サイトにおけるクリックストリームデータを利用する．本データセットは，経営科学系研究部会連合協議会（JASMAC）が主催するデータ解析コンペティションで提供された．対象となる EC サイトではシャツ，パンツ，帽子，時計などのアパレル商品を扱っている．データセットには閲覧履歴や購買履歴が記録されており，具体的には，時刻，利用者 ID，商品 ID，商品閲覧（商品購買）のフラグからなる．いつ（時刻），どの利用者（利用者 ID）がどの商品（商品 ID）を閲覧したか，購買したかの情報と解釈することができる．なお，本データセットには 44,080 人の利用者が含まれている．

表 2.5 に示すように，ページ閲覧（PV : View），セッション（Session : Ses），日付（Day）に基づいて最新度と頻度を表す 6 個の特徴量を作成した．各特徴量の最大値（閾値）は本表に記載されている．最終閲覧日からの経過日数を基にした DayR と過去の閲覧数を基にした ViewF は表 2.1 で例示した．実験では，DayR の閾値が 12 であるので最後の商品閲覧（PV）が 5 日前の場合は DayR を 8 に，最後の商品閲覧（PV）が 13 日以上前の場合は DayR を 0 に設定する．また，ページ閲覧（PV）の回数が 15 を超える場合は ViewF を 15 に丸めるという処理を施している．本研究の中間目標は商品選択確率を予測するために最も有効な最新度と頻度の特徴量の組合せを決定することである．

表 2.5. 最新度と頻度を表す特徴量

略称	詳細
ViewR	最終閲覧以降のページ閲覧数に基づく最新度（最大値 60）
SesR	最終閲覧以降のセッション数に基づく最新度（最大値 12）
DayR	最終閲覧以降の経過日数に基づく最新度（最大値 12）
ViewF	閲覧数に基づく頻度（最大値 15）
SesF	セッション数に基づく頻度（最大値 15）
DayF	閲覧日数に基づく頻度（最大値 15）

表 2.5 で与えた閾値はデータセットを分析して決めた．図 2.1 は，各頻度における購買数を表しており，閾値である 15 を超えると購買が極端に少なくなることが確認できる．そこで，頻度に対する閾値を一律で 15 と設定した．一方，図 2.2 は各最新度の商品選択確率を表しており，最新度の最大値を ViewR は 120，SesR は 25，DayR は 25 とした．商品選択確率は表 2.5 に定めた閾値を超えると，最新度の増減による商品選択確率の変化が小さくなり，定数のように振る舞うことがわかる．そのため，商品選択確率の推定の観点から大きな最新度の値は閾値によって切り捨てても影響は小さいといえる．

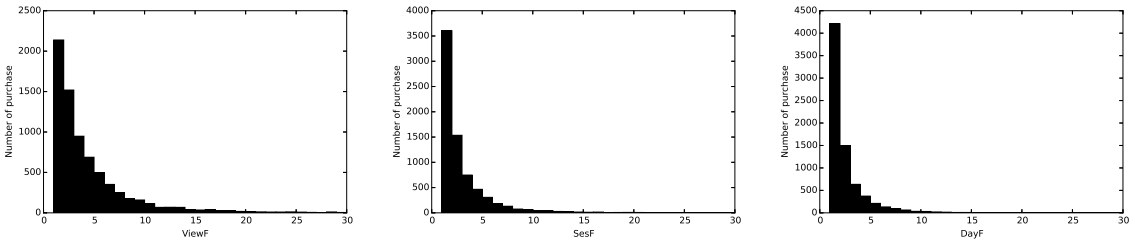


図 2.1. 各頻度における購買数

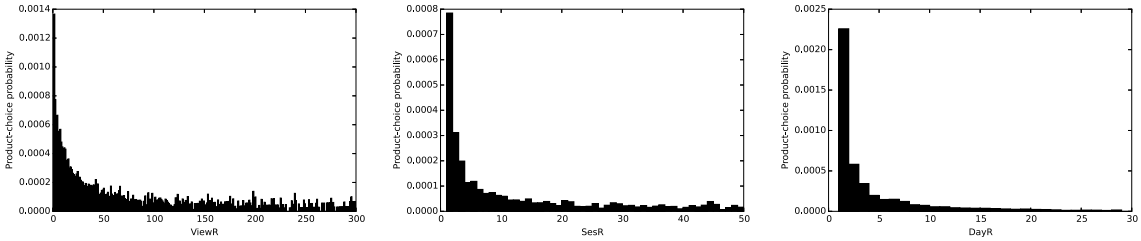


図 2.2. 各最新度における商品選択確率

2.4.2　評価方法

表 2.6 にある 6 個の手法（OneF，EPT，Mono，MCC，LR，SVM）を推薦システムとしてトップ N 推薦による予測性能を評価する．OneF は表 2.5 にある特徴量の 1 つを選択して特徴量の値で降順に並び替えた上位 N 個の商品を推薦する．商品選択確率を推定する際，EPT は経験的に得られた 2 次元確率表 (2.1) を使い，Mono は単調性制約をもつ最適化モデル (2.7) を，MCC は単調性制約と凸性凹性制約をもつ最適化モデル (2.8) を利用する．ここで，最適化モデル (2.7)，(2.8) における不等式制約 $0 < x_{ij} < 1$ は，$\epsilon = 10^{-5}$ として，$\epsilon \leq x_{ij} \leq 1 - \epsilon$ に置き換える．これら EPT，Mono，MCC の手法は各利用者に対して推定された商品選択確率を降順に並び替えて上位 N 個の商品を推薦する．他の 2 つの手法，ロジスティック回帰モデル（LR）とカーネル SVM モデル（SVM）は予測値を降順に並び替えて上位 N 個の商品を推薦する．以下，OneF の特徴量の値，EPT・Mono・MCC の商品選択確率，LR・SVM の予測値をスコアと呼ぶことにする．ここで，降順による並び替えでスコアが等しい商品があった場合，頻度が高い商品を優先的に推薦した．

表 2.6. 比較手法

略称	詳細
OneF	1 つの特徴量を利用した並び替え
EPT	経験的な 2 次元確率表 (2.1)
Mono	単調性モデル (2.7)
MCC	単調性&凸性凹性モデル (2.8)
LR	ロジスティック回帰モデル
SVM	カーネル SVM モデル

次に，学習データセットとテストデータセットの作成方法について説明する．テストデータセットは 2013 年 4 月 1 日から 4 月 28 日の各日を予測日とする 28 個のテストデータから構成される．図 2.3 を用いて 4 月 1 日を予測日とする場合のテストデータと学習データセットの作成方法について説明する．本実験では 4 月 1 日の商品選択確率を予測する場合には前日の 3 月 31 日までのデータが利用できると仮定する．テストデータは，4 月 1 日の商品購買フラグを正解データとして，3 月 4 日から 3 月 31 日までの 28 日間の閲覧履歴から特徴量を作成する．一方，学習データセットは基準日を 3 月 31 日から 28 日前の 3 月 4 日まで遡って作成する．1 つの基準日に対して 1 つの学習データを作成できるため 28 個の学習データから 1 つの学習データセットが作成される．具体的に 3 月 31 日を基準日とした学習データは，3 月 3 日から 3 月 30 日までの 28 日間の閲覧履歴を対象として特徴量を作成し，3 月 31 日の商品購買フラグを付与して作成される．上記の方法で 4 月 1 日を予測日とするテストデータと学習データセットが作成できる．以上のようにして作成した 1 つの学習データセットは，7,365 件の購買データと 42,421,814 件の未購買データで構成されていた．

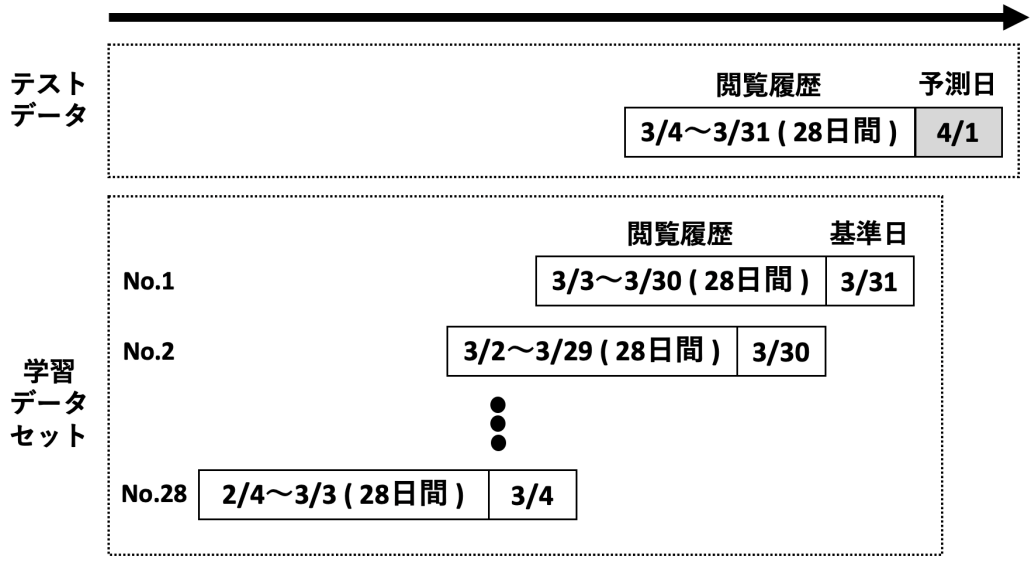


図 2.3. 4 月 1 日を予測日とする場合のテストデータと学習データセットの概念図

最後に評価方法を説明する．例えば 4 月 1 日の購買を予測する場合を考える．各利用者の 3 月 4 日から 3 月 31 日までの 28 日間の閲覧履歴から抽出される商品に対して，表 2.6 にある 6 個の手法を用いてスコアを付与する．商品をスコアにより降順に並び替えて推薦する N 個の商品を決める．本実験では次の評価尺度を用いる．

- 再現率 $= \dfrac{\#(\text{推薦商品}\&\text{購買商品})}{\#(\text{購買商品})}$
- 適合率 $= \dfrac{\#(\text{推薦商品}\&\text{購買商品})}{\#(\text{推薦商品})}$
- F1 値 $= \dfrac{2 \cdot \text{再現率} \cdot \text{適合率}}{\text{再現率} + \text{適合率}}$

ここで $\#(\cdot)$ は商品数を表す．なお，28 個のテストデータセットは，平均で 147.8 個の購買データと 827,399.1 個の非購買データが含まれていた．説明を簡単にするため以下では F1 値を用いて説明する．

2.4.3 実験環境

最適化問題 (2.7)， (2.8) は，株式会社 NTT データ数理システムの数理最適化ソルバー RNUOPT （ver. 1.15.5）を利用して解いた．ロジスティック回帰モデルは統計解析言語 R[*1]（ver. 3.1.1）に実装されている glm 関数を，カーネル SVM モデルは scikit-learn[*2]（ver. 0.15.2）に実装されている SVC 関数を利用した．ここで，カーネル SVM の rbf カーネルは，C $\in \{10^{-4}, 10^{-3}, \ldots, 10^{6}\}$ と gamma $\in \{10^{-6}, 10^{-5}, \ldots, 10^{4}\}$ の組合せを GridSearchCV 関数を用いて 3-分割交差検証により選択した．また，ロジスティック回帰モデルとカーネル

[*1] http://www.R-project.org
[*2] http://scikit-learn.org

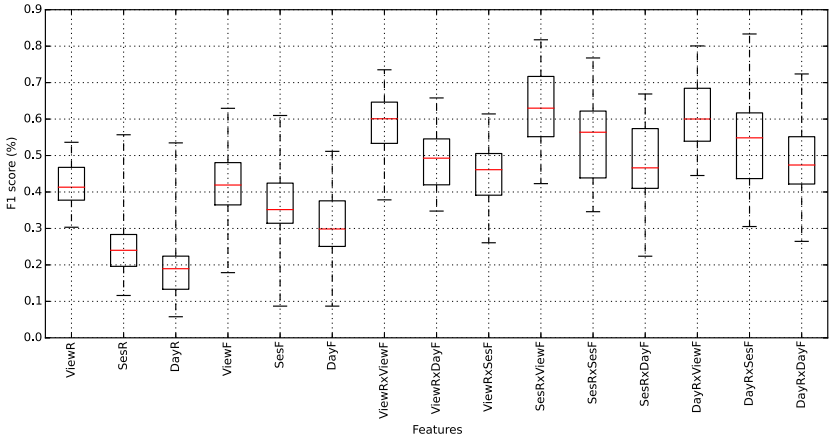
SVM モデルの前処理として表 2.5 にある各特徴量の値は最大値を 1，最小値が 0 となるように正規化した．

カーネル SVM モデルは学習データの数の影響を受けやすく膨大な計算時間が必要なため，直接大規模なデータセットを学習することはできない．そのため，カーネル SVM モデルの学習では次の方法で学習データのサイズを縮小した．具体的には，交差検証の際に元の学習データから無作為に 1,000 購買データと 1,000 非購買データを抽出し，10 回の 3 分割交差検証を繰り返し行った．ここで，`C` と `gamma` は最も多く選択された値を採用した．そして，元のデータセットから無作為に抽出した 7,365 の購買データと 7,365 の非購買データから学習を行い，予測モデルを構築した．

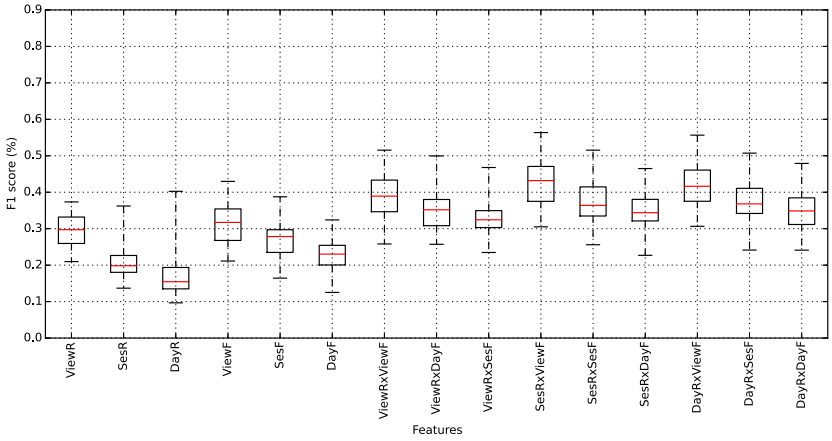
`Mono` と `MCC` の学習時間は通常の PC にて 1 秒未満であったが，ロジスティック回帰モデル（`LR`）は数分の学習時間がかかった．また，カーネル SVM モデルは学習コストが大きいため数十時間を要した．ここで，データから予測モデルを学習する観点で提案手法が最も高速である．また，提案手法は最新度と頻度の最大値を決めれば常に同じサイズの 2 次元データを利用して学習するため学習データの量に依存せずに学習が行うことができ，かつ 2 次元データの作成は集計のみの処理で可能である．そのため使用メモリ量の効率が最もよく，かつ前処理に必要な計算時間も最も高速であった．以上より，提案手法は拡張性がある手法である．実務では安定して解を求めることができるという利点がある．

2.4.4　2 次元確率表の性能

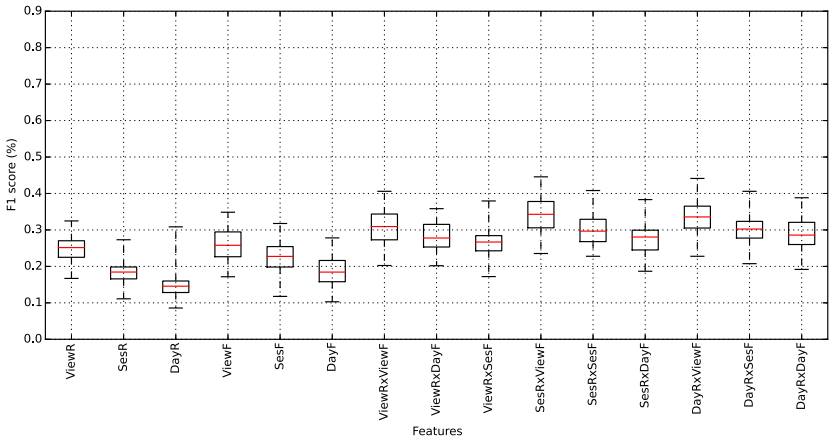
はじめに，2 次元の特徴量によって作成された 2 次元確率表の有効性を確認するため，2 次元確率表の最も単純な手法である `EPT` と 1 次元の特徴量による手法である `OneF` を比較する．図 2.4 の箱ひげ図は 28 個のテストデータセットにおける F1 値を示しており，推薦する商品の数は $N = 1, 3, 5$ としている．ここで，左の 6 個の箱ひげ図は表 2.5 にある 6 個の特徴量の 1 つを選んだ `OneF` の結果である．また，右の 9 個の箱ひげ図は最新度と頻度の特徴量の組合せを用いた `EPT` の結果である．`OneF` の結果からページ閲覧に基づく特徴量である最新度 ViewR と頻度 ViewF が購買を予測するのに有効であることがわかる．セッションに基づく頻度 SesF も次点で有効であることが確認できる．図 2.4 の `OneF` と `EPT` の結果を比較すると，2 次元の特徴量を用いた手法において 1 次元の特徴量を用いた手法よりも F1 値に大幅な上昇が確認できる．例えば，$N = 3$ の場合，`OneF` の F1 値の中央値は 0.32% 未満だが，最新度と頻度の組合せが SesR×ViewF である `EPT` の F1 値の中央値は，約 0.43% であった．`OneF` において最新度 ViewR と 頻度 ViewF が有効であるため，2 次元の特徴量では ViewR×ViewF が `EPT` において有効であることが期待されるが，これは正しくない．実際，最良の組合せは全ての N において SesR×ViewF であった．さらに，F1 値上位 3 つの特徴量の組合せは，すべてページ閲覧に基づく頻度 ViewF が選ばれていた．上記の観察を考慮すると，ViewF は購買を予測するために最も有効な特徴量であると言える．



(a) $N = 1$



(b) $N = 3$



(c) $N = 5$

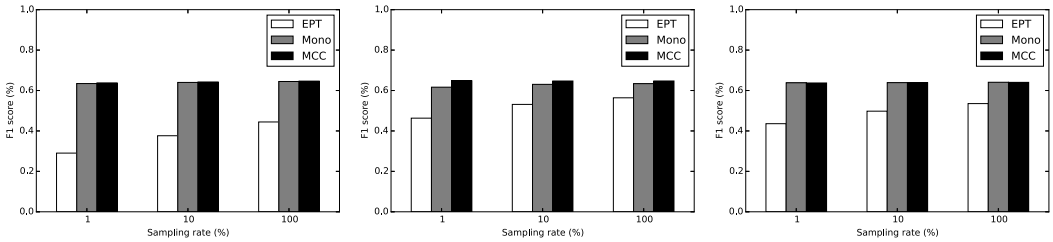
図 2.4. 1 次元の特徴量（`OneF`） vs. 2 次元の特徴量（`EPT` ）

2.4.5　最適化モデルの分析

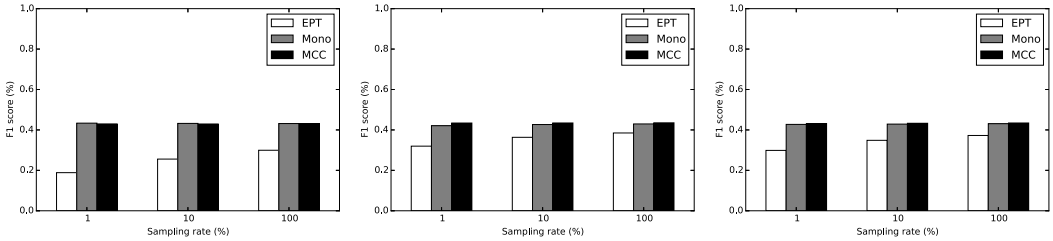
本項では提案する単調性制約 (2.2)，(2.3) と凸性凹性制約 (2.4)，(2.5) の有効性を評価する．学習データの数と予測性能の関係を調べるために，元の学習データセットから 1%，および 10% のデータを抽出した学習データセットを用意する．

予測性能

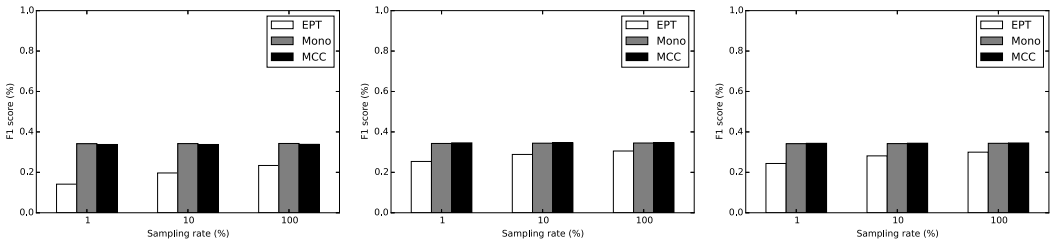
図 2.5 は 28 個のテストデータにおける EPT, Mono, MCC の F1 値の平均を表している．図 2.4 では頻度 ViwF が購買予測に最も有効な特徴量であったため，図 2.5 には ViewR×ViewF, SesR×ViewF，DayR×ViewF の実験結果を示す．図 2.5 から経験的な 2 次元確率表である EPT は最も性能が悪いこと，さらにデータの抽出率が小さくなるほど EPT の F1 値が悪化している．これは学習データが少ない場合には経験的な 2 次元確率表の信頼性が低下することを意味する．対照的に Mono と MCC の F1 値はデータの抽出率が小さくなってもほとんど変化しなかった．これは学習データの数に関して提案手法は安定していることを示している．特に Mono と MCC は ViewR×ViewF の特徴量の組合せを利用した場合に F1 値で EPT を大幅に上回る．さらに，Mono と MCC の F1 値については大きな差は見られなかった．



(a) $N = 1$（左：ViewR×ViewF ／ 中央： SesR×ViewF ／ 右： DayR×ViewF）



(b) $N = 3$（左：ViewR×ViewF ／ 中央： SesR×ViewF ／ 右： DayR×ViewF）



(c) $N = 5$（左：ViewR×ViewF ／ 中央： SesR×ViewF ／ 右： DayR×ViewF）

図 2.5. 単調性制約と凸性凹性制約の有効性

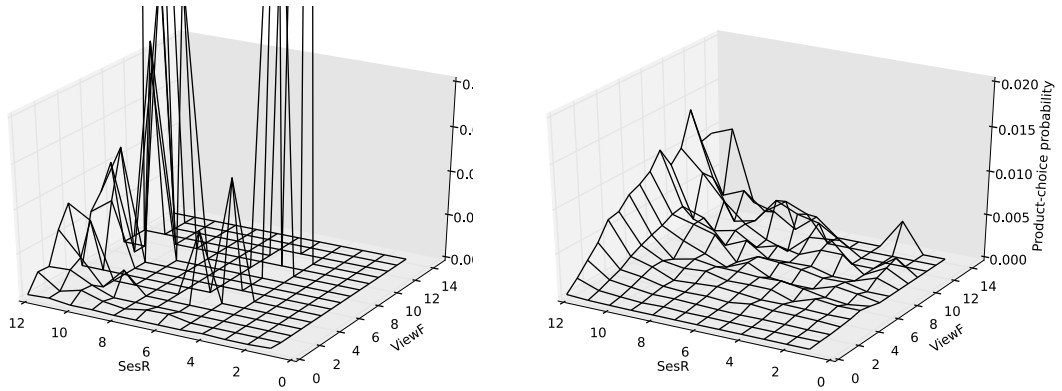
商品選択確率の予測

図 2.6 は，SesR×ViewF を用いて推定された商品選択確率を示している．2 次元確率表の作成において，右側の図は元の学習データセット（100% の学習データセット）を利用し，左側の図は元の学習データセットから 1% のデータを抽出した学習データセットを利用している．

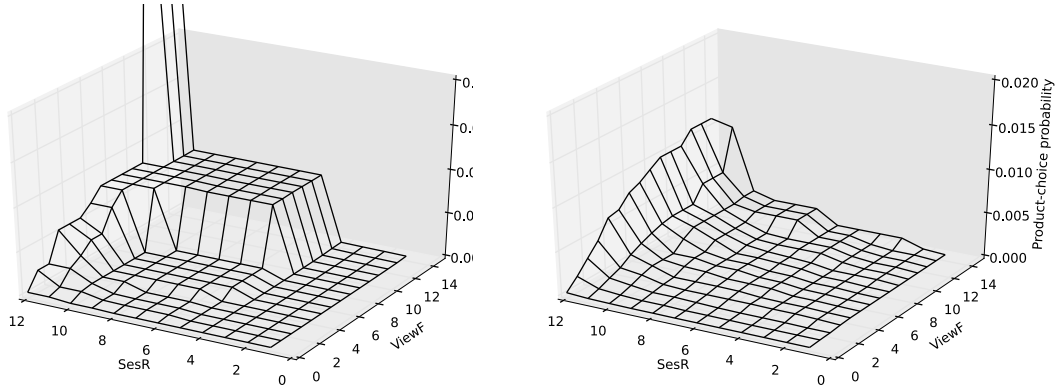
図 2.6(a) の右図から元の学習データが十分にある場合，経験的な 2 次元確率表であっても最新度と頻度に関する単調性の傾向は見られる．しかし，ところどころ単調性が成り立たない箇所がある．特に最新度が小さく，頻度が大きくなる組合せではデータ量が少なくなるため単調性を違反する箇所が目立つ．また，図 2.6(a) の左図から学習データが少ない場合，最新度と頻度に関する単調性は明らかに満たされていないことがわかる．この非単調性は少ないデータ数から $\bar{p}_{ij}$ を計算したことに起因する．特に抽出率 1% の学習データセットにおいては多峰型となり，一部の最新度と頻度では推定値に異常値が観測された．例えば，最新度と頻度の組に対して 1 件しかデータがない場合に 1 件の購買があると商品選択確率が $\bar{p}_{ij} = 1$ となる．

一方，`Mono` では，図 2.6(b) に示すように単調性を満たす 2 次元確率表が得られている．抽出率 1% の学習データセットから作成した 2 次元確率表は元の学習データセットから作成した 2 次元確率表と大きく形状が異なっていたものの，単調性に基づく商品選択確率の補正をすることで図 2.5 に示すように予測性能を改善したと考えられる．

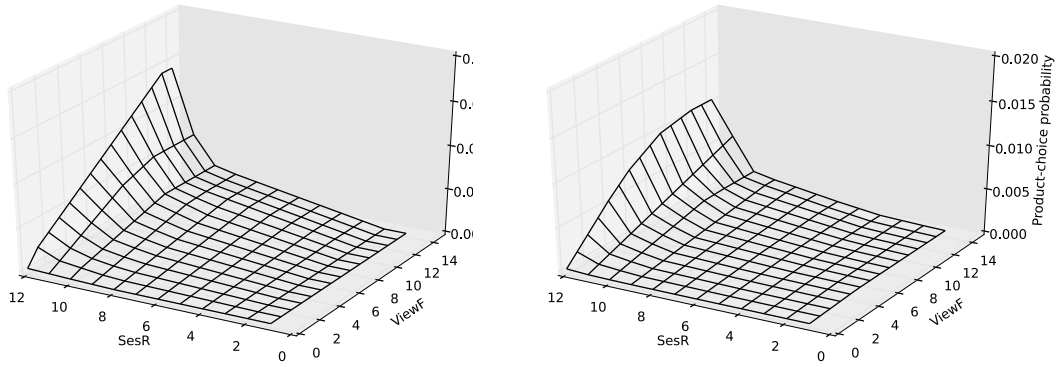
図 2.6(c) では，`MCC` によって推定された商品選択確率は，他の 2 次元確率表と比べて滑らかになることが確認できる．さらに抽出率 1% の学習データセットから作成された 2 次元確率表は元の学習データセットから作成した 2 次元確率表と似た形状となっていた．凸性凹性制約 (2.4)，(2.5) は非常に厳しい制約であるため，`MCC` は学習データが少ない場合にでも商品選択確率をより正確に推定することができる．



(a) 経験的な 2 次元確率表（抽出率：左：1% ／ 右：100%）



(b) 単調性制約モデル（抽出率：左：1% ／ 右：100%）



(c) 凸性凹性制約モデル（抽出率：左：1% ／ 右：100%）

図 2.6. SesR×ViewF による推定された商品選択確率

2.4.6 ロジスティック回帰とカーネル SVM との比較

本項では提案手法の Mono と MCC の予測性能を一般的な 2 値分類器であるロジスティック回帰モデル（LR）とカーネル SVM モデル（SVM）と比較する．

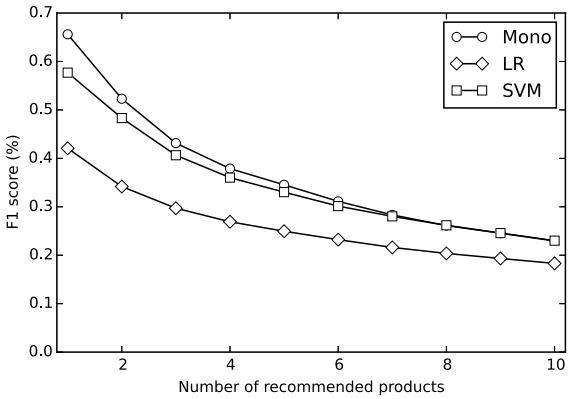
はじめにトップ 3 推薦 ($N = 3$) の結果全体を示す．表 2.7 は，28 個のテストデータセットで実験した再現率，適合率，F1 値の平均を示している．それぞれの特徴量の組で最も性能のよい推薦手法を太字とし，再現率，適合率，F1 値のそれぞれの評価尺度で最も性能の良い特徴量の組は下線で強調した．対象とする EC サイトは，一般的な EC サイトと同様に購買確率が非常に低い値となっている点に注意する．トップ N 推薦の場合，適合率と F1 値は非常に低い値となるが，一方で再現率は比較的高い値となる．

表 2.7 から全ての評価尺度において，特徴量 SesR×ViewF の組合せの MCC が最も良い予測性能であった．また，最も予測性能が低い特徴量の組合せである SesR×DayF を除く，8 個全ての特徴量の組合せにおいて提案手法である Mono か MCC のどちらかが最も良い予測性能であった．これは提案手法が比較手法と比べて特徴量の組合せに対して安定した予測性能を発揮することを示している．また，予測性能は頻度の特徴量に強く依存していることもわかる．例えば，Mono の F1 値は，ViewF のとき 0.4297〜0.4357，SesF のとき 0.3841〜0.3867，DayF のとき 0.3528〜0.3550 であり，予測性能が頻度の特徴量に強く依存していることが確認できる．一方で LR と SVM の予測性能は特徴量の組合せによって提案手法と大きな差があり，提案手法と比較して不安定であるこがわかる．

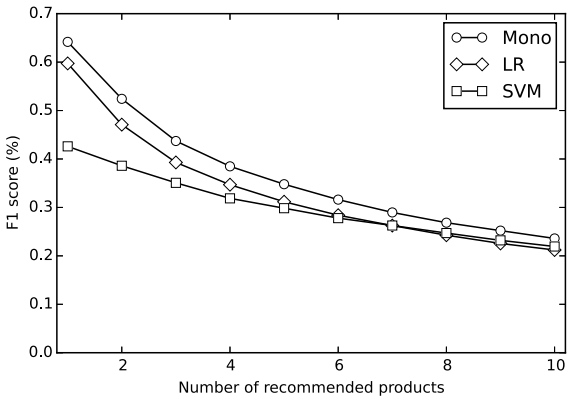
次に，各手法の推薦商品数 N と予測性能の関係を示す．図 2.7 は ViewR×ViewF，SesR×ViewF，DayR×ViewF の特徴量の組合せについて 28 個のテストデータセットで実験した F1 値の平均を示している．ただし，MCC の結果は Mono の結果と類似しているため省略する．また，半数を超える利用者が高々 10 商品しか購買をしていないため，大きな N で比較する必要はないことに注意する．実際，推薦商品数が多くなるほど 3 つの手法 Mono，LR，SVM の F1 値は近づいており，$N = 10$ において予測性能の差は小さくなっている．まず，図 2.7 の結果として全ての特徴量の組合せにおいて提案手法である Mono が比較手法である LR，SVM よりも良い予測性能を示した．また，図 2.7(a) では LR よりも SVM が優れており，図 2.7(b) と図 2.7(c) は SVM よりも LR が優れていることが確認できる．これは，LR と SVM は提案手法と異なり，様々な特徴量に対して頑健な予測モデルを構築することが難しいことを意味している．以上より，提案手法が比較手法と比べて商品の推薦数に対しても安定した予測性能を発揮することを示している．

表 2.7. $N = 3$ における各評価尺度の結果

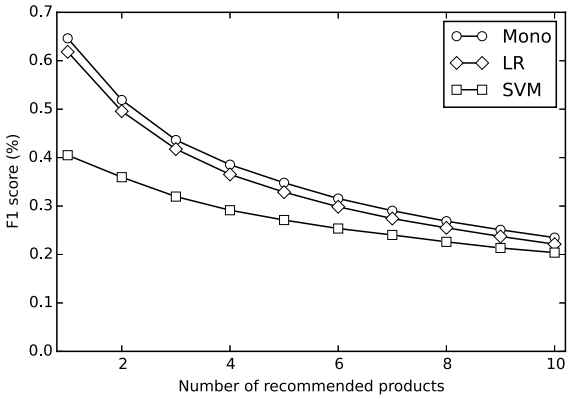
評価尺度	特徴量			推薦手法 Mono	MCC	SVM	LR
再現率（%）	ViewR	×	ViewF	40.50	**40.82**	33.88	28.01
		×	SesF	36.28	**36.62**	23.84	35.26
		×	DayF	33.50	**33.56**	22.98	32.34
	SesR	×	ViewF	41.08	**41.19**	38.38	37.03
		×	SesF	**36.33**	36.29	21.26	34.33
		×	DayF	33.25	33.03	22.04	**33.69**
	DayR	×	ViewF	40.92	**40.93**	29.85	39.23
		×	SesF	**36.48**	36.41	28.06	35.45
		×	DayF	**33.43**	33.29	23.28	32.62
適合率（%）	ViewR	×	ViewF	0.2160	**0.2176**	0.1807	0.1488
		×	SesF	0.1931	**0.1945**	0.1266	0.1877
		×	DayF	0.1785	**0.1787**	0.1220	0.1722
	SesR	×	ViewF	0.2190	**0.2193**	0.2045	0.1969
		×	SesF	**0.1936**	0.1933	0.1125	0.1825
		×	DayF	0.1774	0.1761	0.1174	**0.1795**
	DayR	×	ViewF	0.2186	**0.2186**	0.1589	0.2092
		×	SesF	**0.1944**	0.1940	0.1491	0.1885
		×	DayF	**0.1779**	0.1773	0.1239	0.1734
F1 値（%）	ViewR	×	ViewF	0.4297	**0.4329**	0.3594	0.2961
		×	SesF	0.3841	**0.3869**	0.2518	0.3735
		×	DayF	0.3550	**0.3556**	0.2428	0.3425
	SesR	×	ViewF	0.4357	**0.4362**	0.4069	0.3917
		×	SesF	**0.3851**	0.3845	0.2238	0.3631
		×	DayF	0.3528	0.3504	0.2335	**0.3570**
	DayR	×	ViewF	0.4348	**0.4349**	0.3160	0.4162
		×	SesF	**0.3867**	0.3859	0.2966	0.3751
		×	DayF	**0.3540**	0.3528	0.2465	0.3450



(a) ViewR×ViewF



(b) SesR×ViewF



(c) DayR×ViewF

図 2.7. $N = 1, 2, \ldots, 10$ における提案手法と比較手法の F1 値

次に，表 2.8 は，商品推薦数 $N = 5, 10, 20, 100$ における 28 個のテストデータセットにおける適合率 (%) の平均値である．それぞれの特徴量の組で最も性能のよい推薦手法を太字とし，再現率，適合率，F1 値のそれぞれの評価尺度で最も性能の良い特徴量の組は下線で強調した．表 2.8 から全ての N において，特徴量 SesR×ViewF か DayR×ViewF の組合せの `Mono` か `MCC` が最も良い予測性能であった．また，最も予測性能が低い特徴量の組合せである SesR×DayF を除く，8 個全ての特徴量の組合せにおいて提案手法である `Mono` か `MCC` のどちらかが最も良い予測性能であった．これは提案手法が比較手法と比べて特徴量の組合せに対して安定した予測性能を発揮することを示している．

以上の議論から図 2.4，表 2.7，表 2.8 の結果を総合すると，最も効果的な特徴量の組合せは，$N \leq 10$ で SesR×ViewF であり，$N \geq 20$ で DayR×ViewF であることが示された．

最後に類似した結果として MAP（mean average precision）による評価を表 2.9 にまとめる．MAP は大まかに述べると適合率-再現率曲線の平均的な面積を表している（ Manning et al [113]）．それぞれの特徴量の組で最も性能のよい推薦手法を太字とし，最も性能の良い特徴量の組は下線で強調した．提案手法は全体的に高い MAP 値であり，特徴量 DayR×ViewF の組合せの `MCC` が最も良い予測性能であった．

以上の実験を通して提案手法は比較手法に比べて予測性能が優れており，安定していることが示された．これは，提案手法は最新度と頻度の間にある交互作用をノンパラメトリックに表現できる柔軟性があったことに起因すると考察できる．

表 2.8. $N = 5, 10, 20, 100$ における適合率 (%)

	特徴量			推薦手法			
				Mono	MCC	SVM	LR
$N = 5$	ViewR	×	ViewF	**0.17261**	0.17087	0.14685	0.12473
		×	SesF	**0.15114**	0.15079	0.11368	0.14800
		×	DayF	**0.14287**	0.14087	0.11178	0.13900
	SesR	×	ViewF	0.17403	**0.17452**	0.16678	0.15580
		×	SesF	0.15403	**0.15426**	0.10025	0.14673
		×	DayF	0.14218	0.14187	0.10852	**0.14254**
	DayR	×	ViewF	0.17401	**0.17443**	0.13312	0.16440
		×	SesF	**0.15597**	0.15507	0.12414	0.15032
		×	DayF	**0.14380**	0.14361	0.10605	0.14131
$N = 10$	ViewR	×	ViewF	**0.11504**	0.11354	0.10675	0.09145
		×	SesF	**0.10734**	0.10474	0.09662	0.10085
		×	DayF	0.10345	0.10019	0.09408	**0.10379**
	SesR	×	ViewF	**0.11808**	0.11731	0.11595	0.10586
		×	SesF	**0.10770**	0.10750	0.08356	0.10227
		×	DayF	**0.10410**	0.10329	0.09089	0.10298
	DayR	×	ViewF	**0.11745**	0.11728	0.09946	0.11051
		×	SesF	**0.10839**	0.10802	0.09527	0.10463
		×	DayF	**0.10392**	0.10330	0.08466	0.10174
$N = 20$	ViewR	×	ViewF	**0.07589**	0.07478	0.07410	0.06633
		×	SesF	**0.07222**	0.07101	0.07167	0.06910
		×	DayF	0.07009	0.06875	0.07007	**0.07126**
	SesR	×	ViewF	**0.07659**	0.07654	0.07655	0.07127
		×	SesF	**0.07322**	0.07313	0.06530	0.06995
		×	DayF	**0.07168**	0.07162	0.06800	0.07129
	DayR	×	ViewF	0.7722	**0.07728**	0.06988	0.07307
		×	SesF	0.07321	**0.07328**	0.06935	0.07094
		×	DayF	0.07156	**0.07159**	0.06274	0.07061
$N = 100$	ViewR	×	ViewF	0.03170	0.03162	**0.03182**	0.03093
		×	SesF	0.03143	0.03132	**0.03165**	0.03102
		×	DayF	0.03113	0.03110	**0.03155**	0.03155
	SesR	×	ViewF	**0.03186**	**0.03186**	0.03175	0.03115
		×	SesF	**0.03165**	0.03164	0.03102	0.03109
		×	DayF	**0.03152**	0.03148	0.03108	0.03122
	DayR	×	ViewF	0.03195	**0.03197**	0.03159	0.03124
		×	SesF	0.03173	**0.03174**	0.03164	0.03115
		×	DayF	**0.03161**	0.03160	0.03128	0.03120

表 2.9. MAP(%) の結果

特徴量			推薦手法			
			Mono	MCC	SVM	LR
ViewR	×	ViewF	**0.3741**	0.3735	0.3105	0.2616
	×	SesF	**0.3299**	0.3266	0.2233	0.3252
	×	DayF	**0.2979**	0.2937	0.2104	0.2839
SesR	×	ViewF	0.3749	**0.3768**	0.3516	0.3505
	×	SesF	0.3417	**0.3422**	0.2064	0.3320
	×	DayF	0.3082	0.3096	0.2003	**0.3104**
DayR	×	ViewF	0.3793	**0.3825**	0.2772	0.3622
	×	SesF	**0.3490**	0.3478	0.2821	0.3352
	×	DayF	**0.3199**	0.3174	0.2495	0.3103

2.5　まとめ

本節では各利用者の過去の閲覧履歴から最新度と頻度の特徴量を作成し，商品選択確率を推定する最適化モデルを提案した．提案する最適化モデルは，最新度と頻度について単調性制約と凸性凹性制約を満たす 2 次元確率表を作成する．制約を与えることで学習データが少ない場合でも高い予測性能を発揮する．

本研究は，多次元のノンパラメトリックな関数を推定する新しい計算手法を確立し，クリックストリームデータの分析において従来の予測モデルを超える予測性能を持つ手法を提案している点で大きな貢献がある．

数値実験の結果は，提案手法はロジスティック回帰モデルとカーネル SVM モデルの予測性能と比べて明らかに優れていることを示している．すなわち，適切な制約をもつノンパラメトリックな予測モデルは高い予測性能を発揮する潜在能力をもつと言える．

2.1 節で述べたように提案手法は予測タスクにおいて，安定性，柔軟性，拡張性の点で利点を持つ．特に提案する最適化モデルは学習データのデータ数が少ない場合でも学習でき，逆に大規模データでも安定してモデル構築ができる．加えて，提案手法はカーネル SVM とは対照的に複雑なハイパーパラメータの調整の必要がない．また，一度最適化モデルを解いて 2 次元確率表を作成しておけば，EC サイトに簡単に接続することができる．

本研究には，いくつかの課題と発展性がある．まず，学習データの量と制約の強さに関するトレードオフの課題がある．学習データが少量の場合，正規化の観点で MCC や Mono により過剰適合を防ぐことができる．一方で学習データが十分にある場合，MCC による制約は強すぎる可能性があり，あらゆるノイズが排除されているならば理想的には EMP で十分である．また，本研究において MCC と Mono は 2 次元確率表全体に制約を課しているが，2 次元確率表において部分的に制約を課す方法も考えられる．本研究で用いた EC サイトのデータではなく，購買

に周期性があるようなデータ，すなわち最新度に単調性が成り立たないようなデータには本手法を適用しても高い予測性能は見込めないことに注意されたい．

次に発展性として，本研究では 2 次元の特徴量を扱ったが，3 次元以上のより大きな次元に拡張することが考えられる．次元が増えることで特徴量間に与えられる制約もより複雑になる．また，本提案手法は，各利用者が過去に閲覧した商品に対して選好順序を与えることで商品推薦をしたが，選好スコアを計算していることにも注目したい．すなわち，利用者がどの商品にどの程度の興味があるのかを定量化しているため，協調フィルタリングに応用することが可能である．本発展については 3 章で議論する．

最後に実問題への応用について触れたい．提案手法は，実験結果から明らかなように学習データの数が少ないほど他の予測モデルよりも高い予測性能を示す．また，支配的な特徴量がある場合や得られる特徴量が少ない場合に特に重宝される．そのため，病気などの症例が少なく，検査により取得できる特徴量が少ないケースにも応用することが可能である．他の研究分野への応用の可能性も検討できる．

第 3 章

協調フィルタリングにおける評価値行列の推定

3.1 はじめに

EC サイトの利用者に対して，興味がある商品を推測して推薦する仕組みが一般的になりつつある．利用者は様々な商品を閲覧することを楽しむとともに，数ある商品の中から選択することにストレスを感じる．このような情報過多に対処するために推薦システムはより重要になっている [106]．協調フィルタリングは推薦システムの代表的なアルゴリズムの 1 つであり，対象とする利用者に対して，他の利用者の嗜好を学ぶことで商品推薦をすることができる [1, 2, 12, 29]．

協調フィルタリングのアルゴリズム実装では，利用者の商品への選好を表す利用者-商品間の評価値行列を必要とする．利用者の選好を表すデータは，利用者から明示的に取得する方法と暗黙的に取得する方法がある [55, 72]．明示的に取得する方法は利用者に直接商品を評価してもらうことで商品の選好データを得る（例えば，アンケートに回答してもらう）．一方，暗黙的に取得する方法は利用者の行動を観察することで利用者に気づかれることなく商品の選好データを得る（例えば，購買履歴から得る）．しかし，利用者の選好を推定することは，多くのノイズの影響を受けることが知られており [5, 96]，このようなノイズが推薦精度の大幅な低下に繋がることが実証されている [6]．

EC サイトはクリックストリームデータを自動的に収集する仕組みを提供し，利用者の閲覧履歴を得ることができる．これらのデータはオンラインでの購買行動を予測する上で有益であることが実証されている [7, 76, 97]．また，推薦システムのアルゴリズムにクリックストリームデータを利用する研究は過去に多く存在する [22, 60, 83, 119]．しかし，これらの既存研究はいずれもクリックストリームデータの効果的な利用方法には焦点を当てておらず，質の高い評価値行列を作成するという目的の研究ではない．

2 章では，形状制約付きの最適化モデルである MCC モデル (2.8) を利用して，各利用者の過去の閲覧履歴から商品に対する最新度と頻度に基づく商品選択確率を推定する方法を提案し，購買商品の予測精度が向上することを示した．本手法は最新度と頻度に対して単調性と凸

性凹性の制約を満たす予測モデルを用いて商品選択確率の予測を行っている．また，MCC モデルを利用することで 2 値分類の標準的な手法であるロジスティック回帰モデルとカーネル SVM モデルよりも良い予測性能であることを示した．さらに，商品の異質性を MCC モデルに取り込むため，Nishimura et al. [94] は商品の潜在クラスごとに商品選択確率を予測する潜在クラスモデルを提案した．これらのモデルは，各利用者が過去に閲覧した商品を対象に，それぞれの商品を選択する確率を推定している．一方で，本章で提案する手法は利用者にとって未知の商品を推薦するための協調フィルタリングに焦点を当てている．

本章で提案する手法は，暗黙的に得られるクリックストリームデータ，特に閲覧履歴から高い品質の評価値行列を作成することを目的とする．このため，MCC モデルを利用して閲覧履歴から利用者の選好をより正確に推定する．提案手法は商品閲覧の最新度と頻度の間にある交互作用を完全に表現できるため，利用者の予測時点における選好をより正確に捉えることが可能である．本章では，実際に EC サイトから得られた閲覧履歴を用いて数値実験を行い提案手法の有効性を評価する．協調フィルタリングアルゴリズムとしては，利用者間型協調フィルタリングと非負値行列分解を用いる．

本研究の貢献は以下にまとめられる．

- 閲覧履歴から得られる暗黙的なフィードバック評価を用いて，高い品質の評価値行列を作成する方法を提案する．本手法は，利用者-商品間の評価値行列を利用する様々な種類の推薦システムアルゴリズムの性能を向上させる可能性をもつ．
- 提案手法で作成した評価値行列と比較手法で作成した評価値行列を用いて，協調フィルタリングの数値実験を行った．その結果，利用者間型の協調フィルタリングと非負値行列分解ともに提案手法を用いた場合に予測性能が高くなることが示された．また，学習データが少ない場合にも，形状制約ノンパラメトリック推定による評価値予測は有効であることを示した．
- EC サイトにおける推薦システムへの実装を想定し，最も有効な最新度と頻度の特徴量の組合せを決定した．この特徴量は閲覧履歴から簡単に作成でき，EC サイトに推薦システムを実装する際に有益な知見である．

3.2 関連研究

本節では協調フィルタリングとその改善に関する研究について体系的に総説する．

3.2.1 協調フィルタリング

協調フィルタリングにおいて様々なアルゴリズムが提案されてきたが，主に 2 つのタイプに分類される [2]．すなわち近傍ベースの手法とモデルベースの手法である．

近傍ベースの方法は，利用者間または商品間の類似性に焦点を当てて推薦システムを構築する．利用者間型の協調フィルタリングは，推薦システムにおいて協調フィルタリングが実装さ

れた初期の手法である [104]．推薦対象の利用者の嗜好に似ている利用者グループを抽出して，当該グループで評価の高い商品を推薦する方法である．推薦対象の利用者に特化されている点に注目されたい．

一方，商品間型の協調フィルタリングは，対象となる商品と関係の強い商品を推薦する方法で，Amazon.com で実際に利用された方法である [75]．このアルゴリズムは商品間の類似度を利用者の評価を用いて計算する．商品間の類似度を利用することで，利用者が過去に高く評価した商品と類似した商品を推薦することができる [110]．利用者間型，および商品間型の協調フィルタリングはそれぞれ異なる利点がある [29]．本研究では，推薦システムの研究で一般的に利用される利用者間型の協調フィルタリングを用いて実験を行う．

モデルベースの協調フィルタリングは，決定木，相関ルール分析，ナイーブベイズ分類器，行列分解などの機械学習やデータマイニングの手法を利用して，推薦対象の利用者に商品を推薦する [2]．特に，非負値行列分解 [68, 69] の協調フィルタリングは，非負値の評価値行列から商品推薦を行う方法である [20, 39, 44, 79, 131]．非負値行列分解のアルゴリズムは，評価値行列を潜在因子を用いて低ランクの行列に分解することで評価値行列の評価済みの値から未評価の評価値を予測する．潜在因子を通して特定の商品が推薦される理由が明確であるという利点がある．このような透過性は EC サイトに推薦システムを実装する際の信頼性を担保する要素になる [92]．

3.2.2　協調フィルタリング推薦システムの改善

協調フィルタリングの性能を改善することを目的としたいくつかの研究がある [12]．これらは表 3.1 のように分類できる．まず，利用者・商品クラスタリングの研究において，予め利用者や商品をクラスタリングすることでコールドスタート問題に対処する方法がある．これは，協調フィルタリングの評価値行列が疎である場合に，協調フィルタリングアルゴリズムが適切に機能しない問題を解決する．また，利用者間型や商品間型の協調フィルタリングにおいて類似尺度の定義が評価値行列を補完する上で重要なため，類似尺度の改善の研究もある．現在の推薦システムでは，利用者属性，商品属性，ソーシャルネットワークから取得した利用者間の情報など豊富な追加情報を利用できるように拡張されている．また，文脈依存の推薦システムは，利用者の状況（場所，時間，デバイス，感情など）を考慮して，利用者が求めている商品を推薦する [123]．

一方，本研究は協調フィルタリングにおいて，暗黙的に得られた評価（閲覧履歴や購入履歴）から品質の高い評価値行列を計算し推薦システムの性能を改善することを目的とする．多くの協調フィルタリングアルゴリズムは利用者-商品間の評価値行列を作成する必要があるため，利用者の評価にノイズが多いと推薦システムの性能は著しく低下する．そのため評価値行列の品質を改善する様々な方法が提案されてきた．例えば，利用者の明示的な評価に加え，暗黙的な評価のデータを追加して評価値行列を作成する方法や，ノイズのある評価を抽出して削除，または修正をする方法がある．本研究では，明示的な評価を利用せず，多くの EC サイトから簡単に取得できるクリックストリームデータから得られる暗黙的な評価データのみを利用

する．また，提案手法は，ノイズのある評価の検出や修正をする労力を必要としない．2 章で解説した商品選択確率を推定するための形状制約付きの最適化モデル [52] を利用して評価値行列の品質を改善する．具体的には，閲覧履歴から商品の最新度と頻度の特徴量を算出し，最適化モデルによって計算された商品選択確率を評価値として採用する方法を提案する．

表 3.1. 協調フィルタリングによる推薦システムの改善手法

研究の種類	文献	研究内容	主な主張
利用者・商品クラスタリング	[93]	利用者自動クラスタリング	10% 以上の予測精度の改善
	[57]	評価パターンのクラスタリング	行列近似モデルによる予測精度の改善
	[74]	自己構成クラスタリング	推薦処理時間の短縮
	[109]	時刻重み付きクラスタリング	増分データに対する効率的なデータ処理
	[130]	利用者の局所的，大域的な選好度	類似尺度に基づく効果的な利用者の選好度
	[61]	部分空間クラスタリング	事前のパラメータチューニングなしで高い予測精度を達成
	[81]	信頼性によるクラスタリング	ベースラインの予測精度を更新，コールドスタート問題の緩和
	[88]	相関ルールによるクラスタリング	疎なデータに対する相関ルールの有効性
	[41]	利用者の性格特性	新規利用者とコールドスタート問題に効果的
	[91]	商品のオントロジー	規模拡張性とスパース性の課題に対して効果的にオントロジーを適用
類似尺度の改善			
	[8]	遺伝的アルゴリズムによる改良	遺伝的アルゴリズムを利用して効果的に重み付けを取得するためのフレームワーク
	[99]	多段階のピアソン相関係数	新規利用者とコールドスタート問題に効果的
	[59]	適応型ニューロファジィ推論	ベースラインよりも適合率を 21% 改善
	[77]	混合類似性モデル	類似性を混合する際の有効な方法
	[50]	ピアソンの相関係数に基づくモデル	2 値購買データに特化した効果的なモデル
ソーシャルネットワーク推薦			
	[90]	ソフト評価とソーシャルネットワーク	ハードなモデルよりも柔軟で高い予測精度を達成
	[4]	重み付き多重属性モデル	拡張した利用者行動の解析からの効果的な重み付け
	[42]	総説論文	2006 年から 2017 年の間の 407 本の論文を整理したソーシャルコマース研究の系統的な調査
文脈依存推薦システム			
	[123]	総説論文	1997 年から 2017 年の間の 114 本の論文を整理した文脈依存情報推薦研究の系統的な調査

表 3.1 続き：協調フィルタリングによる推薦システムの改善手法

研究の種類	文献	研究内容	主な主張
評価値行列の改善			
	明示的な評価と暗黙的な評価		
	[34]	暗黙的な選好	学習データに対する効果的な重み付けと選択戦略
	[63]	明示的／暗黙的なフィードバック	因子分解は精度を維持しつつ規模拡張性を強化する
	[78]	明示的／暗黙的なフィードバック	明示的／暗黙的なフィードバックを統合する有益な方法
	[97]	明示的／暗黙的なフィードバック	楽曲推薦における効果的な対数変換
	ノイズを含む評価の検出・修正		
	[127]	データ縮約によるノイズ除去	インスタンスの効果的な選択によって予測精度を向上させる
	[96]	悪意のあるノイズの修正	悪意のあるノイズを検知して修正することで予測精度が向上する
	[65]	評価の分散	予測精度を高めるための分散の有効な利用方法
	[6]	評価時のノイズ低減	予測精度を高めるための再評価データの効果的な利用
	[11]	商品の重要性	商品の重要性を考慮することで予測精度が高まる
	[71]	ノイズの多い利用者の検出	自己矛盾に関するノイズが多い利用者の検知の方法
	[98]	誤った評価の修正	利用者の選好に基づいた効果的な修正ルール
	[67]	評価における人気スコア	人気商品の選好は有効な指標である
	[120]	一貫性のない評価の修正	補足情報がない上での効果的な修正ルール
	本研究	商品選択確率の予測	利用者間型，非負値行列分解による協調フィルタリングの予測精度の改善

3.2.3　最新度に基づく協調フィルタリング

品質の高い評価値行列を計算するために提案手法では商品閲覧の最新度を用いる．その点で時間の影響を考慮した推薦システムと関係が強い [19]．表 3.2 は，最新度の影響を考慮した協調フィルタリングの方法の一覧である．最も採用される手法は，時間加重関数を用いて評価値行列を作成する方法で，経過時間に対して指数的に減衰する加重関数を採用した方法がある [24]．提案手法は，最新度に加え，頻度の影響を考慮しているため時間加重関数の拡張としても特徴づけられる．具体的な提案手法は，商品閲覧に関する最新度と頻度の全ての組合せに対して商品選択確率を計算することでノンパラメトリックに 2 次元確率表を作成する手法である．本手法は商品閲覧についての最新度と頻度に関する交互作用を商品推薦に効果的に利用できる．さらに，時間に依存するハイパーパラメータの調整が必要ない点にも注目したい．商品閲覧に関する最新度と頻度を用いてノンパラメトリックに品質の高い評価値行列を作成する方法は著者が知る得る限り初めての手法である．

表 3.2. 最新度に関する協調フィルタリングの研究

手法	文献
時系列予測	[46, 133]
時間加重関数	[24, 25, 51, 66, 125]
時間距離の閾値設定	[18, 38]
最新の評価ほど重みをつける	[70]
時間に基づく行列分解	[58, 62, 63]
半順序に基づく選好の定式化	[73]
利用者の興味の上昇と下降の定式化	[126]
最新度と頻度に基づく 2 次元確率表	**本研究**

3.3 協調フィルタリングのアルゴリズム

本節では，利用者-商品間の評価値行列を説明した後，協調フィルタリングアルゴリズムである利用者間型協調フィルタリングと非負値行列分解の解説をする．

3.3.1 利用者-商品間の評価値行列

U を利用者の集合，I を商品の集合とする．利用者の商品への選好度を表す利用者-商品評価値行列を

$$\boldsymbol{R} := (r_{ui})_{(u,i)\in U\times I} \in \mathbb{R}^{U\times I}$$

と定める．ただし，r_{ui} は利用者 u の商品 i に対する評価値である．また，利用者 u の評価値ベクトルを $\boldsymbol{r}_u := (r_{ui})_{i\in I}$ と定義する．次項ではクリックストリームデータから評価値行列を計算するためのフレームワークを紹介する．

3.3.2 利用者間型協調フィルタリング

利用者 u へ推薦する商品を決めるために，利用者間型の協調フィルタリングアルゴリズムは，当該利用者と好みが似ている他の利用者の情報を利用する．このため，利用者 u と u' の間の好みの類似性を定量化する類似度尺度 $s(u,u')$ に依存する．最も一般的な 2 つの類似尺度はピアソンの相関係数とコサイン類似度であり [1]，それぞれ次のように定義される．

$$s(u,u') = \frac{\displaystyle\sum_{i\in I(u,u')}(r_{ui}-\bar{r}_u)(r_{u'i}-\bar{r}_{u'})}{\sqrt{\displaystyle\sum_{i\in I(u,u')}(r_{ui}-\bar{r}_u)^2}\sqrt{\displaystyle\sum_{i\in I(u,u')}(r_{u'i}-\bar{r}_{u'})^2}} \tag{3.1}$$

$$s(u,u') = \frac{\boldsymbol{r}_u\cdot\boldsymbol{r}_{u'}}{\|\boldsymbol{r}_u\|_2\|\boldsymbol{r}_{u'}\|_2} = \frac{\displaystyle\sum_{i\in I(u,u')} r_{ui}r_{u'i}}{\sqrt{\displaystyle\sum_{i\in I(u,u')} r_{ui}^2}\sqrt{\displaystyle\sum_{i\in I(u,u')} r_{u'i}^2}} \tag{3.2}$$

ここで，$I(u)$ は利用者 u に評価された商品の集合で，$I(u,u')$ は利用者 u と u' の両方に評価された商品の集合である．また，$\bar{r}_u := (\sum_{i\in I(u)} r_{ui})/|I(u)|$ は利用者 u の評価値平均である．

利用者 u が未評価である商品 i の評価値を予測／補完するために，利用者間型の協調フィルタリングアルゴリズムは，利用者 $u' \in N(u,i)$ の評価値を集計する．ここで，$N(u,i)$ は利用者 u と好みが似ていて，かつ商品 i を評価した利用者の集合である．以下に集計関数の例 [1] を示す．

$$p_{ui} = \frac{1}{|N(u,i)|} \sum_{u'\in N(u,i)} r_{u'i} \tag{3.3}$$

$$p_{ui} = \alpha \sum_{u'\in N(u,i)} s(u,u') r_{u'i} \tag{3.4}$$

$$p_{ui} = \bar{r}_u + \alpha \sum_{u'\in N(u,i)} s(u,u')(r_{u'i} - \bar{r}_{u'}) \tag{3.5}$$

ここで α は正規化項であり，次のように定義される．

$$\alpha = \frac{1}{\displaystyle\sum_{u'\in N(u,i)} |s(u,u')|}$$

3.3.3　非負値行列分解

協調フィルタリングは，評価値行列で欠損している要素，すなわち未知の評価を予測する．そのため協調フィルタリングは行列補完問題と捉えることができ，評価値行列を行列積で近似する行列分解によって解決される．このとき評価値行列のすべての成分が非負である仮定が自然に成り立つ問題設定が多く，非負値行列分解 (NMF) [68, 69] により利用者と商品の関係を表現することができる．評価値行列を行列積に分解する際の因子の集合を F とし，そのサイズは利用者の集合や商品の集合よりも十分に小さなサイズとする（すなわち，$|F| \ll \min\{|U|,|I|\}$）．非負値行列分解は次のように定式化される．

$$\left| \begin{array}{ll} \text{minimize} & \|\boldsymbol{R} - \boldsymbol{W}\boldsymbol{H}\|_{\mathrm{F}}^2 \\ \text{subject to} & \boldsymbol{W},\ \boldsymbol{H} \geq \boldsymbol{O} \end{array} \right. \tag{3.6}$$

ここで $\boldsymbol{W} \in \mathbb{R}^{U\times F}$ と $\boldsymbol{H} \in \mathbb{R}^{F\times I}$ は分解のための行列であり，$\boldsymbol{O}$ は零行列である．また，$\|\cdot\|_{\mathrm{F}}$ はフロベニウスノルムであり，$\boldsymbol{R}$ の欠損している箇所は計算では除外されることに注意する．最後に $\boldsymbol{R} \approx \boldsymbol{W}\boldsymbol{H}$ から導かれる $p_{ui} = (\boldsymbol{W}\boldsymbol{H})_{ui}$ を用いて利用者 u が未評価である商品 i への評価を予測する．

3.4　評価値行列構成に関する提案手法

本節では閲覧履歴から得られる最新度と頻度の組合せに対する商品選択確率から構成される 2 次元確率表を用い評価値行列を構成するアルゴリズムについて述べる．2 次元確率表は 2 章で提案した最適化問題 (2.8) を利用することに注意する．

3.4.1　評価値行列構成アルゴリズム

図 3.1 は，本研究で提案する閲覧履歴から評価値行列を構成するアルゴリズムの手順を示している．

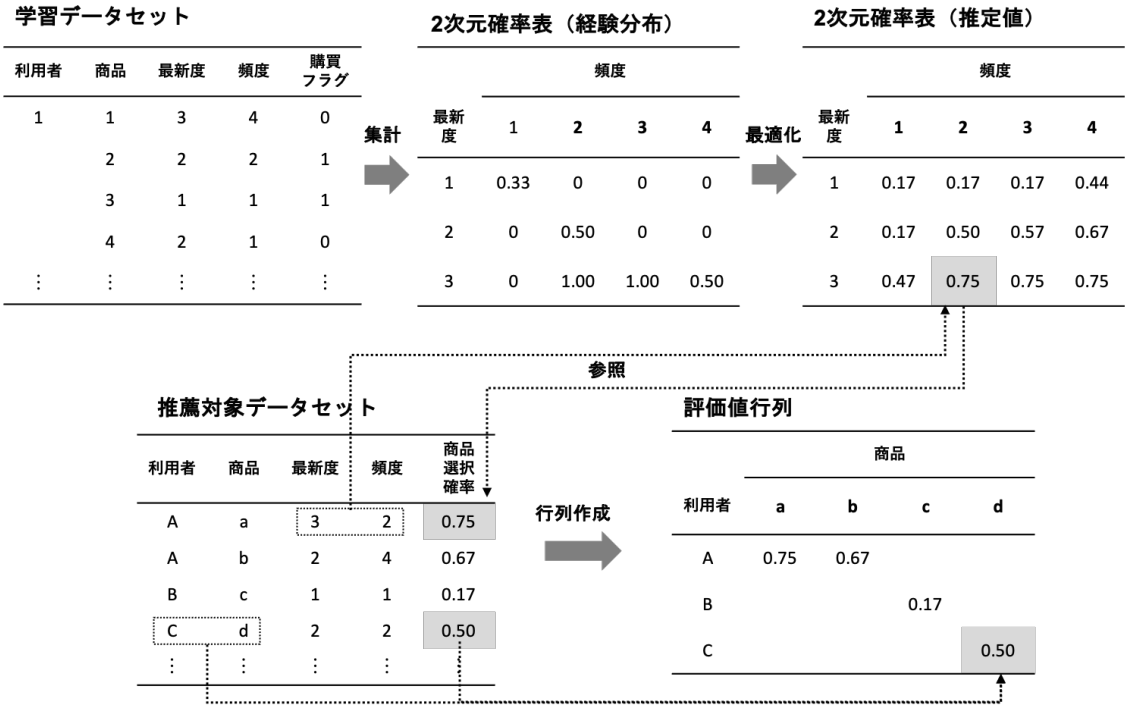


図 3.1. 評価値行列の構成の手順

はじめに学習データセットを準備する．クリックストリームデータに含まれる閲覧履歴から各利用者が閲覧した商品に対して最新度と頻度の特徴量を算出し，購買フラグを付与する．次に学習データセットを集計し，最新度と頻度に対する商品選択確率を計算し，経験分布としての 2 次元確率表を得る．この 2 次元確率表を最適化問題 (2.8) の入力として最適化し，推定値としての 2 次元確率表が得られる．推定された 2 次元確率表の商品選択確率は最新度と頻度に対して単調性が満たされているだけでなく，最新度に対して凸性をもち，頻度に対して凹性を持つ．一方で推薦対象データセットを準備し，学習データセットと同様にクリックストリームデータに含まれる閲覧履歴から各利用者が閲覧した商品に対して最新度と頻度の特徴量を算出する．ここで，2 次元確率表（推定値）を参照することで利用者と商品の組合せに対して商品選択確率を紐付けることができる．最後に推薦対象データセットから利用者と商品，商品選択確率の列を抽出し，協調フィルタリングの入力となる評価値行列を構成する．

図 3.1 を用いて具体的に説明する．例えば，推薦対象データセットにおける利用者-商品の組 (A, a) は特徴量として最新度 3 と頻度 2 を持っており，2 次元確率表（推定値）の $(3, 2)$ 成分を参照することで，利用者-商品の組　$(3, 2)$ の商品選択確率は 0.75 であることがわかる．

同様にして推薦対象データセットの全ての利用者-商品の組みに対して商品選択確率を紐付

けることができる．最後に推薦対象データセットから評価値行列を構成する．例えば，利用者-商品の組 (C, d) の商品選択確率 0.50 を評価値行列の (C, d) 成分に対応させる．

以上より，提案する評価値行列の構成アルゴリズムは次のように要約される．

ステップ 1（学習データセットの準備）： 閲覧履歴から利用者-商品の組合せに対して，最新度と頻度の特徴量を算出し，購買フラグを付与する．

ステップ 2（経験分布としての 2 次元確率表の作成）： 学習データセットを集計することで最新度と頻度に対して商品選択確率を対応付ける 2 次元確率表を作成する．

ステップ 3（推定値としての 2 次元確率表の作成）： 経験分布としての 2 次元確率表を入力として最適化問題を解き，2 次元確率表を推定する．

ステップ 4（商品選択確率の参照）： 推薦対象データセットを閲覧履歴から作成し，最新度と頻度の特徴量を算出する．推定された 2 次元確率表を参照することで商品選択確率を紐付ける．

ステップ 5（評価値行列の構成）： 利用者と商品の組に対する商品選択確率を利用して評価値行列を構築する．

一般的な協調フィルタリングは，最新度または頻度のどちらかの特徴量を利用して評価値行列を構成する．提案手法は，最新度と頻度の間にある交互作用を完全に表現できる点で注目すべきである．そのため，協調フィルタリング適用時点での利用者の商品への興味を最新度と頻度の両方の観点で見積もることができる．

3.5　数値実験

本節では 3.4.1 節で説明した提案手法の有効性を評価するために数値実験を行い，推薦システムとして協調フィルタリングの性能評価を行った．

3.5.1　実験方法

データセット　数値実験には，経営科学系研究部会連合協議会主催平成 25 年度データ解析コンペティションで提供されたデータセットを利用した．アパレル系（衣類，装飾品，靴，鞄など）の EC サイトにおいて 2015 年 8 月 1 日～2015 年 10 月 30 日の期間に取得した閲覧履歴である．また，2 次元確率表を推定するための学習データは，2015 年 9 月に取得された購買データを利用した．一方，推薦システムの評価をするためのテストデータは，2015 年 10 月に取得された購買データを利用した．表 3.3 は，各データセットの利用者数，商品数，データ数，購買数を示している．なお，1 件のデータは利用者と商品の組を表す．

表 3.3. 学習データとテストデータの統計値

	学習データ	テストデータ
利用者数	323,798	331,661
商品数	426,865	427,776
データ数	26,683,669	34,148,017
購買数	7,545	7,381

特徴量設計 最新度と頻度の特徴量については，閲覧（View），セッション（Ses），日（Day）の 3 つの観点に基づいて定義されている．例えば，利用者と商品の組 (u, i) のセッション観点の頻度（SesF）は，利用者 u が商品 i を閲覧したセッション数として定義される．他の特徴量も同様に定義でき，合計で 6 つの特徴量を定義した．表 3.4 は，6 個の特徴量の略語と閾値のリストである．

表 3.4. 最新度と頻度の特徴量の略語と閾値

	閲覧（View）	セッション（Ses）	日（Day）
最新度	ViewR（$|J| = 42$）	SesR（$|J| = 8$）	DayR（$|J| = 22$）
頻度	ViewF（$|K| = 15$）	SesF（$|K| = 7$）	DayF（$|K| = 5$）

これらの閾値は次のように利用されている．例えば，ある利用者の閲覧した適当な商品について m 日前に閲覧したとすると，利用者と商品の組に対して DayR は $\max\{23 - m, 1\}$ と計算される．同様に，ViewF の場合はある利用者の適当な商品の閲覧数を n とすると，利用者と商品の組に対して ViewF は $\min\{n, 15\}$ で計算される．ここで，利用者と商品の購買の組が 97.5% 以上となるように最新度と頻度の閾値を決定した．

比較手法 提案する評価値行列の構成方法を評価するために，次の評価値行列の作成方法を比較した．

- Rece: 利用者 u の商品 i に対する最新度
- Freq: 利用者 u の商品 i に対する頻度
- EMP: 経験的に得られた 2 次元確率表から得られる商品選択確率（$x_{jk} = q_{jk}/n_{jk}$, $(j, k) \in J \times K$）
- MCC: 2 次元確率表から得られる商品選択確率（最適化問題 (2.8) の解）

例えば，図 3.1 における利用者と商品の組 (A, b) に注目すると，Rece なら評価値は 2，Freq なら 4，MCC であれば 0.67 となる．推薦システムの研究 [22, 60, 119] では，利用者の評価値として閲覧数，すなわち ViewF が頻繁に使われるため，ベースラインとみなすことができる．

評価尺度　4 つの手法によって求めた評価値行列を用いて 3.3 節で説明した協調フィルタリングアルゴリズムを適用し，トップ N 推薦により性能を評価した．N 個の推薦商品には，利用者が過去に閲覧した商品は除かれている．特定の利用者に対する推薦システムの性能は，アルゴリズムが推薦した商品のうち実際に利用者が購買した商品の数により評価した．一般的な評価尺度として再現率と適合率が利用される．再現率は利用者が購買した商品数のうちアルゴリズムが推薦した商品数の割合を表す．一方，適合率はアルゴリズムが推薦した商品数のうち利用者が購買した商品数の割合を表す．また，F1 値は再現率と適合率の調和平均で計算される．

$$\text{F1 値} := \frac{2 \cdot \text{適合率} \cdot \text{再現率}}{\text{適合率} + \text{再現率}}$$

F1 値は，日別に全利用者の平均値をとり，さらにテストセットにおいて 28 日間で平均をとった．

アルゴリズムの実装　評価値行列の品質を評価するために，協調フィルタリングのアルゴリズムである利用者間型協調フィルタリングと非負値行列分解でテストを実施した．利用者間型協調フィルタリングは Python 言語で実装し，非負値行列分解は Python 言語の機械学習パッケージである scikit-learn 3（ver. 0.19.1）の `sklearn.decomposition.NMF` 関数を利用した．利用者間型協調フィルタリングの実装はコサイン類似度 (3.2) を利用し，集計関数 (3.4) と $N(u,i) = U$ ，$\alpha = 1$ を用いた．非負値行列分解の計算では分解する行列のサイズ $|F|$ を 4 とした．最適化問題 (2.8) は，株式会社 NTT データ数理システム[*1]の数理最適化ソルバー Numerical Optimizer V17 を用いて解いた．ここで，問題規模（すなわち，$|J| \times |K|$）は非常に小さく，最適化問題 (2.8) は 1 秒未満で最適解を得ることができる．

3.5.2　利用者間型協調フィルタリングの結果

本項では，協調フィルタリングアルゴリズムとして利用者間型協調フィルタリングを利用した結果の考察を与える．はじめに利用者の好みを推定する 2 次元確率表の有効性について確認する．図 3.2 は，最新度と頻度の特徴量の 9 種の組合せについて Rece，Freq，MCC を評価した結果である．各グラフは，推薦商品数 $N = 1, 2, \ldots, 100$ までの F1 値の折れ線グラフである．

次のことが図 3.2 から読み取れる．

- MCC はどの特徴量の組合せでもほとんど全ての N に対して最も高い F1 値を示している．特に，推薦商品数 N が $N \leq 10$ の周辺で最も F1 値の差が大きく，それよりも N が大きくなるとその差は小さくなる．
- Rece の予測性能は，最新度の特徴量によって異なり，ViweR を利用したときに非常に小さい N で高い F1 値を示した（図 3.2(a)–(c)）．

[*1] `http://www.msi.co.jp/english/`

- Freq の予測性能は，頻度の特徴量によってあまり差が見られない．さらに，Freq は手法の中で最も悪い F1 値を示していた．
- MCC の予測性能は最新度の性能に大きく依存しており，最新度として DayR を利用したときに他の手法を大幅に上回った．

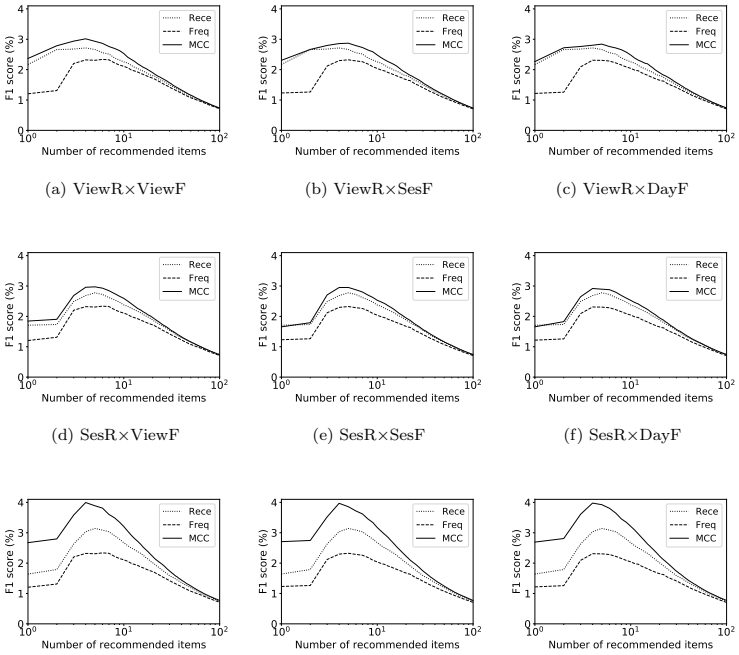
上記の観察から利用者間型協調フィルタリングにおいて MCC（2 次元確率表から評価値行列を作成する方法）は推薦システムの性能を向上させるために有効である．また，最新度の特徴量 DayR は商品推薦における利用者の嗜好をよく表現していると言える．一方で，これらの知見は実験で対象としているアパレル系の EC サイト，および取扱商品と密接に関係している．そのため，他の特徴量も異なる EC サイト，異なる商品の場合にも有効である可能性を持つ．以下では，最も効果的な特徴量の組合せである DayR×ViewF，DayR×SesF，DayR×DayF について論じる．

次に 2 次元確率表に対して形状制約を課す利点を示すため，学習データのデータ数に対する安定性を実験で確かめた．ここでは，元の学習データセットから抽出率 1%，および抽出率 10% の無作為抽出したデータを学習データセットとして追加した．図 3.3 は，最新度と頻度の 3 つの組合せ（DayR×ViewF，DayR×SesF，DayR×DayF）と 3 つの商品推薦数（$N = 5, 10, 15$）についてのグラフである．抽出率 1% の学習データセット， 抽出率 10% の学習データセット, 全学習データセット（元の学習データセット）について，Rece，Freq，EMP，MCC を F1 値で比較する．Rece と Freq は 2 次元確率表の推定をする必要がないので，学習データの数に無関係であることに注意する．

次のことが図 3.3 から読み取れる．

- EMP は，全学習データを利用した場合に MCC と同等の高い F1 値を示した．しかし，抽出率 1% 学習データセットを利用したときには，Rece や Freq と同等の性能にまで劣化した．
- MCC は，EMP と比較して学習データの数に関係なく高い F1 値を維持している．すなわち，形状制約を課した 2 次元確率表を利用することで高い予測性能を発揮する．
- MCC は抽出率 1% の学習データセットを利用したときでも Rece や Freq よりも常に良い性能を示している．これは，抽出率 1% の学習データセットに 75 個の購買データしか含まれていないにも関わらず，高い性能を示したという点で興味深い結果である．

上記の観察から，形状制約を課した 2 次元確率表を用いて評価値行列を構成することで，学習データが少量しか得られない状況でも，利用者間型協調フィルタリングを用いた推薦システムの高い性能を引き出すことができると言える．これは，コールドスタート問題等が起きる実用上で大きな利点となる．

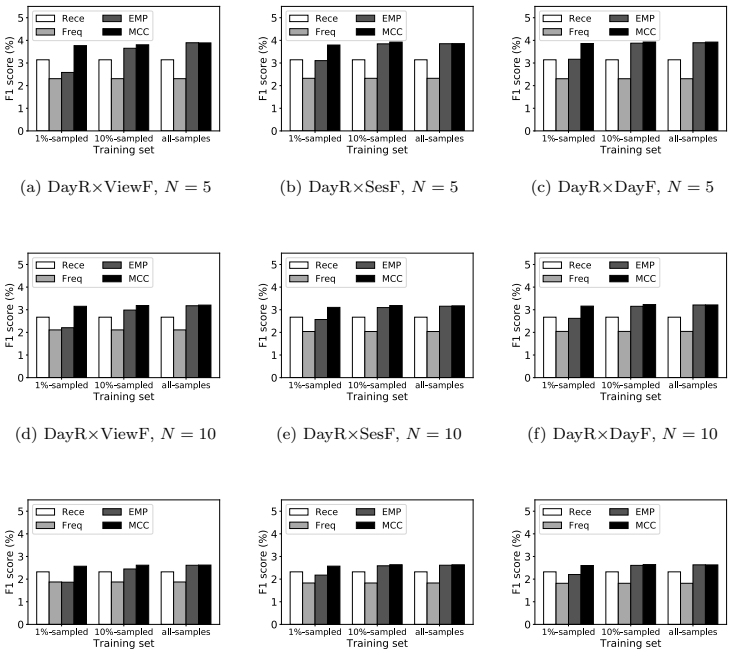


(a) ViewR×ViewF　(b) ViewR×SesF　(c) ViewR×DayF

(d) SesR×ViewF　(e) SesR×SesF　(f) SesR×DayF

(g) DayR×ViewF　(h) DayR×SesF　(i) DayR×DayF

図 3.2. 推薦商品数に対する推薦精度（利用者間型協調フィルタリング）



(a) DayR×ViewF, $N = 5$　(b) DayR×SesF, $N = 5$　(c) DayR×DayF, $N = 5$

(d) DayR×ViewF, $N = 10$　(e) DayR×SesF, $N = 10$　(f) DayR×DayF, $N = 10$

(g) DayR×ViewF, $N = 15$　(h) DayR×SesF, $N = 15$　(i) DayR×DayF, $N = 15$

図 3.3. データ抽出率に対する推薦精度（利用者間型協調フィルタリング）

3.5.3 非負値行列分解の結果

本項では，協調フィルタリングアルゴリズムとして非負値行列分解を利用した結果の考察を与える．利用者間型の協調フィルタリングの場合と同様に利用者の好みを推定する 2 次元確率表の有効性について確認する．図 3.4 に示される結果から次のことが読み取れる．

- MCC は，どの特徴量の組合せでもほとんど全ての N に対して最も高い F1 値を示しているが，利用者間型協調フィルタリングと異なり，最新度の特徴量 SesR を利用した際に Rece と競合している（図 3.4(d)–(f)）．しかしながら，MCC は最新度の特徴量 DayR を利用した際に実験上で最も高い性能を発揮した（図 3.4(g)–(i)）．
- Rece の性能は最新度の特徴量の選び方にわずかに依存していた．Rece が最新度 ViewR を選んだときに最も低い F1 値であり（図 3.4(a)–(c)），最新度 DayR を選んだときに最も高い F1 値であった（図 3.4(g)–(i)）．
- Freq の性能は特徴量の組合せに関係なく最も性能が低かった（図 3.2）．
- 図 3.2 とは異なり，MCC の性能は頻度の特徴量にわずかに影響を受けている．これは，小さな N における最新度 DayR の F1 値を比較すれば明らかである（図 3.4(g)–(i)）．

上記の観察から利用者間型の協調フィルタリングと同様に非負値行列分解においても MCC（2 次元確率表から評価値行列を作成する方法）を用いて推薦性能を改善できる．また，非負値行列分解の性能は利用者間型協調フィルタリングと比べて利用する特徴量の選択に対して敏感である．すなわち，非負値行列分解を利用して推薦システムを実装する場合には，最新度と頻度の特徴量を慎重に選択する必要があることを示唆している．加えて利用者間型協調フィルタリングと非負値行列分解は，それぞれの特徴量の組合せにおいて類似した予測性能の傾向があったが，推薦商品数 N によって大きく予測性能が異る．例えば，両方の協調フィルタリングアルゴリズムで最も性能が良かった特徴量の組合せ DayR$\times\times$ DayF の結果に注目すると，$N = 1$ のときに，非負値行列分解は利用者間型協調フィルタリングよりも性能が良かった．一方，この関係は $N = 4$ 前後で逆転した（図 3.2(i) ，3.4(i)）．

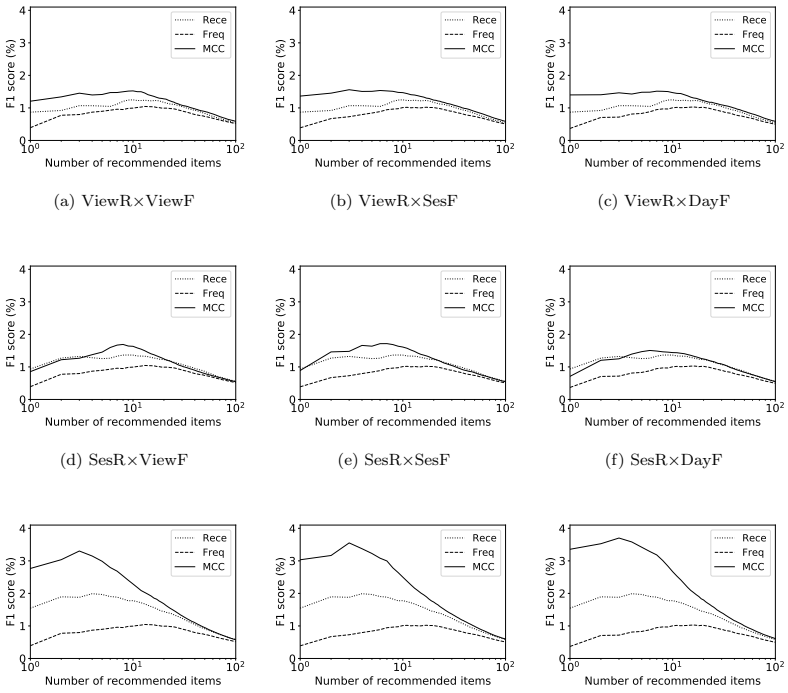
次に 2 次元確率表に対して形状制約を課す利点を示すため，学習データのデータ数に対する安定性を実験で確かめた．無作為抽出する学習データは，利用者間型協調フィルタリングの実験と同じデータセットを利用している．図 3.5 が示す結果から以下のことが読み取れる．

- 利用者間型協調フィルタリングの結果と同様に EMP は全学習データセットを利用した場合に非常に高い F1 値を示したが，学習データ数が減少すると性能が低下した．
- 抽出率 1% の学習データセットを利用した場合に MCC の F1 値がわずかに低下したが，ほぼ全ての場合で MCC が他の手法と比べて最も高い F1 値を示した．

上記の観察から，形状制約を課した 2 次元確率表を用いて評価値行列を構成することで，学習データが少量しか得られない状況でも，非負値行列分解を用いて推薦システムの高い性能を引き出すことができる．一方，MCC の推薦システムの性能は，非負値行列分解の場合には，抽

出率 1% のデータセットの場合にわずかに性能が劣化した．これは，推薦システムに非負値行列分解を利用する場合には，利用者間型協調フィルタリングを利用する場合と比べて学習データ数が多いことが重要であると示唆している．

本研究の全ての実験結果は表 3.5 に要約される．ここでは利用者間型協調フィルタリングと非負値行列分解に関する F1 値を推薦商品数 $N = 1, 2, \ldots, 100$ について平均している．4 つの手法（Rece，Freq，EMP，MCC）について最も高い F1 値を太字で強調した．この結果は MCC がすべての状況で安定して高い性能を発揮することを示している．

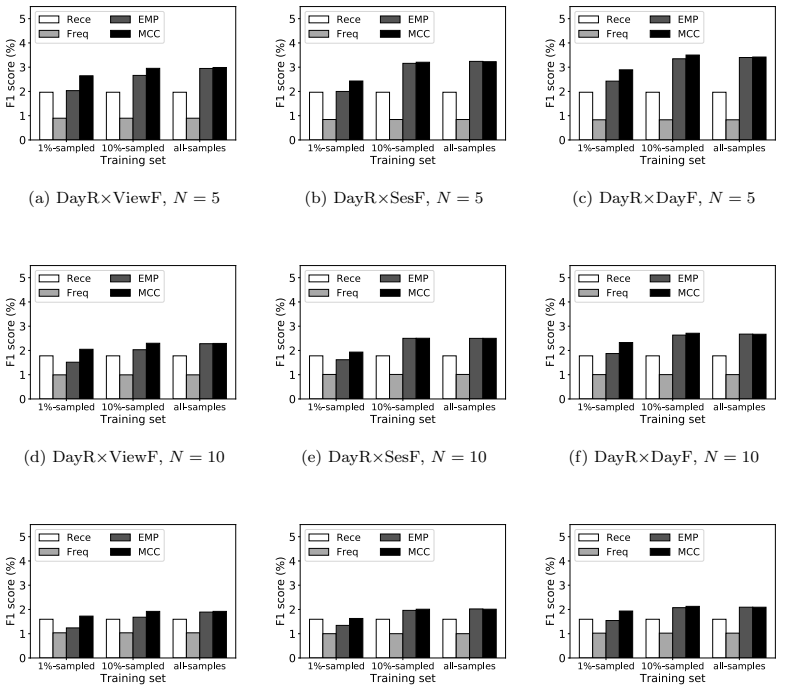


(a) ViewR×ViewF (b) ViewR×SesF (c) ViewR×DayF

(d) SesR×ViewF (e) SesR×SesF (f) SesR×DayF

(g) DayR×ViewF (h) DayR×SesF (i) DayR×DayF

図 3.4. 推薦商品数に対する推薦精度（非負値行列分解）



(a) DayR×ViewF, $N = 5$ (b) DayR×SesF, $N = 5$ (c) DayR×DayF, $N = 5$

(d) DayR×ViewF, $N = 10$ (e) DayR×SesF, $N = 10$ (f) DayR×DayF, $N = 10$

(g) DayR×ViewF, $N = 15$ (h) DayR×SesF, $N = 15$ (i) DayR×DayF, $N = 15$

図 3.5. データ抽出率に対する推薦精度（非負値行列分解）

表 3.5. 利用者間型協調フィルタリングと非負値行列分解の $N = 1, 2, \ldots, 100$ の平均 F1 値

		利用者間型協調フィルタリング				非負値行列分解			
学習データ	特徴量	Rece	Freq	EMP	MCC	Rece	Freq	EMP	MCC
抽出率 1%	ViewR × ViewF	1.31	1.22	1.17	**1.39**	0.85	0.75	0.77	**1.03**
	× SesF	1.31	1.21	1.24	**1.35**	0.85	0.74	0.62	**0.96**
	× DayF	1.31	1.21	1.27	**1.37**	0.85	0.74	0.79	**1.00**
	SesR × ViewF	1.33	1.22	1.22	**1.36**	**0.90**	0.75	0.78	**0.90**
	× SesF	1.33	1.21	1.27	**1.35**	**0.90**	0.74	0.78	0.89
	× DayF	1.33	1.21	1.26	**1.36**	**0.90**	0.74	0.77	0.88
	DayR × ViewF	1.41	1.22	1.13	**1.54**	1.04	0.75	0.86	**1.10**
	× SesF	1.41	1.21	1.31	**1.54**	**1.04**	0.74	0.88	1.03
	× DayF	1.41	1.21	1.33	**1.55**	1.04	0.74	1.00	**1.18**
抽出率 10%	ViewR × ViewF	1.31	1.22	1.34	**1.38**	0.85	0.75	**1.00**	0.96
	× SesF	1.31	1.21	1.35	**1.37**	0.85	0.74	0.88	**0.96**
	× DayF	1.31	1.21	1.36	**1.38**	0.85	0.74	0.93	**0.96**
	SesR × ViewF	1.33	1.22	1.37	**1.38**	**0.90**	0.75	0.83	**0.90**
	× SesF	1.33	1.21	**1.38**	**1.38**	0.90	0.74	0.92	**0.93**
	× DayF	1.33	1.21	1.37	**1.38**	0.90	0.74	**0.91**	0.90
	DayR × ViewF	1.41	1.22	1.48	**1.56**	1.04	0.75	1.10	**1.19**
	× SesF	1.41	1.21	1.54	**1.57**	1.04	0.74	1.23	**1.26**
	× DayF	1.41	1.21	1.55	**1.57**	1.04	0.74	1.30	**1.32**
全データ	ViewR × ViewF	1.31	1.22	1.37	**1.38**	0.85	0.75	0.93	**0.95**
	× SesF	1.31	1.21	1.36	**1.37**	0.85	0.74	**0.95**	**0.95**
	× DayF	1.31	1.21	1.36	**1.37**	0.85	0.74	0.94	**0.95**
	SesR × ViewF	1.33	1.22	1.37	**1.38**	**0.90**	0.75	0.89	**0.90**
	× SesF	1.33	1.21	**1.37**	**1.37**	0.90	0.74	0.91	**0.92**
	× DayF	1.33	1.21	**1.37**	**1.37**	0.90	0.74	0.89	**0.91**
	DayR × ViewF	1.41	1.22	1.56	**1.57**	1.04	0.75	1.19	**1.20**
	× SesF	1.41	1.21	**1.56**	**1.56**	1.04	0.74	**1.26**	**1.26**
	× DayF	1.41	1.21	**1.57**	**1.57**	1.04	0.74	**1.31**	**1.31**

3.6　まとめ

本研究では，クリックストリームデータから高品質の評価値行列を計算することを目的とした．高品質の評価値行列を計算することができれば，協調フィルタリングによる推薦システムの性能を改善することができる．提案する手法は，利用者の閲覧履歴から商品の最新度と頻度を計算し，形状制約を課した最適化モデルを用いて商品選択確率を推定する．利用者の商品選択確率に関係する最新度と頻度の交互作用を完全に再現する適切な推定値を得ることができる．さらに，提案手法は利用者と商品の間に定義される評価値行列に基づく様々な推薦アルゴリズムの性能を向上させる可能性を持つ．

本研究では，一般的な協調フィルタリングアルゴリズムである利用者間型協調フィルタリングと非負値行列分解を利用した数値実験を行い，提案手法の有効性を検証した．数値実験の結果，両方の協調フィルタリングアルゴリズムで性能が向上するとともに，学習データ数が少ない場合にも頑健な予測性能を発揮することが示された．これは協調フィルタリングにおけるコールドスタート問題を緩和する．さらに推薦システムに有効な最新度と頻度の効果的な特徴量についても考察した．これらの特徴量は推薦システムを社会実装する上で重要な知見である．

一般的な評価値行列の構成には最新度よりも頻度が使われることが多い．本研究は，実務の観点で利用者の暗黙的評価（評価，閲覧，購買）における最新度が推薦システムの性能を向上させることを示唆している．また，利用者の暗黙的評価における最新度と頻度の交互作用の効果も確認することができた．実際，ページ閲覧の最新度と頻度に基づいて利用者の好みを推定する提案手法は，推薦システムの性能評価において，最新度に基づく手法と頻度に基づく手法を大幅に上回っている．

提案手法は実務上の施策を実施する際にも利点がある．例えば，キャンペーンなどの施策を立案する際に，提案手法を用いることで利用者に対して適切なタイミングで商品を提案することができる．また，実装上の利点も多く持つ．評価値行列を構成するアルゴリズムを簡単に実装できる点や，EC サイトから自動的に収集される閲覧履歴のみを利用している点は実装上の利点である．また，閲覧履歴から 2 次元確率表を作成しておけば，EC サイトに実装されている協調フィルタリングアルゴリズムに簡単に接続できる．提案手法は，閲覧履歴のみを利用しており，利用者の明示的評価を利用していない点も重要である．EC サイトなどのサービスを運営する上で利用者から明示的な評価を得るのは非常にコストが高い．一方で，閲覧履歴による暗黙的な評価を得るのはコストが低く，実用上の利点といえる．

第 4 章

時系列検索確率分布の推定

4.1　はじめに

本章では，時系列確率分布のノンパラメトリック推定について論じる．研究対象はコネヒト株式会社[*1]が運営するサービス「ママリ」の検索データである．「ママリ」は，「ママの一歩を支える，女性向け Q&A アプリサービス」として月間 100 万件の Q&A が投稿される大規模サービスである．

サービスの利用者は子どもの誕生日を登録することができるため，出産前であれば妊娠何周目にどのような検索をしたか，出産後であれば子どもの月齢が何ヶ月の時にどのような検索をしたかを検索データから解析することができる[*2]．また，検索には母親の体の変化や体調だけでなく，子どもに対する興味も現れる．すなわち，本検索データを解析することにより出産日を起点として母親のニーズの推移を理解することができるため，ニーズに合わせた様々な情報推薦に応用することができる．

本研究では検索数の推移を確率分布として正規化して解析をする．確率分布を正規化することでいくつかの利点がある．1 つ目の利点は，正規化することにより検索語間で検索量に依存しない比較ができる点である．一般に，検索行動は顕在化されたニーズが現れる．顕在化されたニーズほど検索量が多くなるが，正規化することで潜在的なニーズも同等に扱うことができる．情報推薦の文脈では顕在化されたニーズと同等に潜在的なニーズも重要である．もちろん，確率分布として解析しておけば，期間内検索量から実際の検索数を算出できるため，検索量に応じた施策を検討することもできる．2 つ目の利点は，確率分布間の解析がしやすくなる点である．例えば，確率分布間に順序を定義することで検索語間のニーズが生じる順番を解析することができる．検索語間に順序が定義できれば，特定のニーズが生じたタイミングで次にどのようなニーズが生じるか情報推薦をすることが可能となる．

上記のように検索数の推移を確率分布として解析することで様々な応用がある．出産日を起点として母親のニーズを時系列確率分布として表現すると様々な情報推薦の仕組みを構築できる．まず，子どもの誕生日が既知な母親に対して，ニーズが生じるタイミングで適切な情報を

[*1] https://connehito.com/

[*2] 本研究で用いた検索データは統計処理されており，個人の特定が不能なデータである．

推薦することを検討できる．例えば，妊娠 N 周，または月齢 N ヶ月のタイミングで適切な記事，Q&A，商品の推薦が行える．また，ニーズを時系列確率分布として保持しておくことで興味を持ち始めるタイミング（例えば累積分布関数が 0.25 の時点）で記事情報を推薦し，興味がピークになったタイミング（例えば時系列確率分布で最大値をとる時点）で商品を推薦するなどの施策が可能となる．なお，情報推薦はアプリなどのサービス上だけでなく，メルマガや EC サイトなど様々な媒体を通して可能である．また，子育てに関係する支援者にとっても有用である．例えば，父親の視点で母親とこどものニーズ知ることは育児参加の敷居を下げるだけでなく，適切な行動を促すことに繋がる．そのため，母親の周辺を巻き込んだ新しい育児の形を検討することができるようになる．

さて，実務の便益を考慮すると本研究の目的は次のように整理される．すなわち，精度の高い時系列確率分布を推定するとともに，推定した確率分布を用いて確率分布間の解析をすることである．

本研究では出産日を起点とする検索数の推移から数理最適化モデルを利用することで，時系列確率分布をノンパラメトリックに推定する方法を提案する．本研究で扱う検索データの時系列確率分布は単峰型か二峰型の分布であり，出産日を起点（0 時点）として分布の裾で検索がなくなる．単峰型の場合，ピーク前で単調増加し，ピーク後で単調減少する傾向がある．二峰型の場合は，1 つ目のピーク前で単調増加し，1 つ目のピーク後と 2 つ目のピーク前で単調減少と単調増加し，2 つ目のピーク後で単調減少する傾向がある．しかし，ノイズが含まれる実データで単調性を完全に満たすことはない．本研究では数理最適化モデルとして混合整数凸二次計画問題を利用することで，単峰型，および二峰型の分布のピークを自動で特定し，ピークの前後で単調性を満たすように時系列確率分布を推定する単調単峰性モデルと単調二峰性モデルを提案する．また，提案手法を利用して推定した時系列確率分布には順序が定義しやすくなることを示す．

本研究では，提案する数理最適化モデルの予測性能を評価するため実データと人工データを用いて数値実験を行う．実データを用いた実験では，提案手法である単調二峰性モデルと単調単峰性モデルを経験分布，移動平均，カーネル回帰と比較する．具体的にはデータ量を調整した 3 つの学習データを用いて推定した時系列確率分布を，テストデータから作成した時系列確率分布を正解データとして，平均平方二乗誤差（RMSE）で評価する．一方，人工データを用いた実験では単峰分布である離散ラプラス分布とポアソン分布の推定に関して単調単峰性モデルとカーネル回帰を比較する．具体的には離散ラプラス分布とポアソン分布を時系列確率分布の正解データとし，それらの分布から生成した学習データを用いて推定した時系列確率分布を平均平方二乗誤差（RMSE）で評価する．

本章の構成を説明する．本節では本研究の概要を述べた．4.2 節では本研究に関連する知識について説明する．4.3 節では時系列確率分布を推定するための最適化モデルについて述べる．4.4 節では実データを用いた数値実験を通して提案手法の有効性を評価し，4.5 節では人工データを用いた数値実験を通して提案手法の有効性を評価する．4.6 節で本章のまとめと今後の研究課題について述べる．

4.2　関連研究

本節では，本研究と関連が深い単峰回帰に関する先行研究を紹介し，数値実験で比較手法として用いる移動平均とカーネル回帰についてふれる．

単峰回帰（unimodal regression）は，Frisén [33] によって研究が開始され，単峰回帰のアルゴリズム [117] も開発されている．2 次元に拡張した研究は Geng and Shi による傘型回帰（umbrella orderings） [35] が知られている．傘型回帰でピークが与えられている問題は単調回帰をサブルーチンとして解くことができるが，Geng and Shi [36] はピークの位置も同時に推定するアルゴリズムを開発した．しかし，より一般的に多峰性の分布に対して極大値（ピーク）や極小値を同時に推定する研究はない．また，相関ルール分析と関連した 2 次元の単峰回帰として最適ピラミッド問題 [137] と呼ばれる問題がある．

単峰回帰とは，観測点の集合 I に対して単峰型の分布から観測値 a_i $(i \in I)$ が与えられている際に，単峰性を満たす推定値 x_i $(i \in I)$ を求める問題である．具体的には，重み w_i $(i \in I)$ とピークの位置である i_{peak} を与え，次の凸二次計画問題として定式化できる．

$$\left|\begin{array}{lll} \text{minimize} & \displaystyle\sum_{i \in I} w_i \left(x_i - a_i\right)^2 & \\ \text{subject to} & x_i \leq x_{i'} & (i, i' \in I, i \leq i' \leq i_{peak}) \\ & x_i \geq x_{i'} & (i, i' \in I, i_{peak} \leq i \leq i') \end{array}\right.$$

目的関数は観測値と推定値の誤差を重み付き二乗誤差として設定し，最小化する．一方，制約式はピークの位置 i_{peak} 前では単調増加し，i_{peak} 後で単調減少をすることを表している．

最適ピラミッド問題 [137] とは，相関ルール分析と関連した単峰型の分布を推定する問題であり，一般に多次元上のデータを単峰関数で近似する手法である．最適ピラミッド問題は画像解析を含む様々な分野に応用されている [140].

本研究は，従来の研究と異なり単峰型の分布だけでなく，二峰型の分布に対して極大値（および極小値）となる時点を同時に推定することができる．特に従来の形状制約は単調性や凸性凹性により形状を制限するものであったが，提案手法は極大値の存在や極小値の存在まで柔軟に定義できる点で新しい拡張である．また，提案手法は，単峰型だけでなく二峰型の確率分布の推定も行うことができ，多峰性の確率分布の推定にも自然に拡張できる汎用的な数理最適化モデルである．

本研究では比較手法の 1 つとして移動平均（moving average）を用いる．時点の集合を T，時系列データを v_t $(t \in T)$ とすると，単純移動平均（simple moving average）は予測時点から直近の n 時点のデータの平均を用いて計算される．時点 t $(\in T)$ の単純移動平均の値を SMA_t とすると次の式で表現される．

$$SMA_t = \frac{\sum_{s=t-n+1}^{t} v_s}{n}$$

移動平均は時系列データを平滑化することを目的とした手法である．

また，比較手法の 1 つとしてカーネル回帰も用いる．カーネル回帰（kernel regression）は確率変数の条件付き期待値を推定するためのノンパラメトリック回帰の代表的な手法である．ノンパラメトリック回帰は，データ点の集合 $\{(X_i, Y_i) \mid i = 1, 2, \cdots, n\}$ を与えたときに

$$Y_i = m(X_i) + \epsilon_i$$

となる m を推定する問題である．ただし，ϵ_i $(i = 1, 2, \cdots, n)$ はノイズで $\mathrm{E}[\epsilon] = 0$ かつ $\mathrm{E}[\epsilon^2] = \sigma^2$ であり，独立同分布を仮定すると $m(x) = \mathrm{E}[Y \mid x]$ と書ける．このとき適当なカーネル関数 K を用いて Nadaraya-Watoson の推定量 [87, 124] は次の式で与えられる．

$$\hat{m}_h(x) = \frac{\sum_{i=1}^{n} K(\frac{X_i - x}{h}) Y_i}{\sum_{i=1}^{n} K(\frac{X_i - x}{h})}$$

h (> 0) はバンド幅と呼ばれ，最小二乗交差確認によって決定することで精度の良いノンパラメトリック回帰モデルを構築することができる．カーネル回帰も移動平均と同様に時系列データを平滑化することを目的とした手法である．

4.3 提案手法

本節では時系列確率分布を数理最適化モデルを利用してノンパラメトリック推定する方法を提案する．はじめに研究対象である時系列検索データの特徴について整理をした後，問題設定を行う．最後に提案する最適化モデルについて説明を行う．

4.3.1 時系列検索データの特徴

本項では出産日を起点とする時系列検索データの特徴を具体例を通して説明する．以下では時系列の粒度を 1 週間とし，出産日を 0 として −55 週（出産 1 年前）〜110 週（出産 2 年後）までの 3 年間の検索数の推移を表した時系列検索分布を用いる．なお，対象とする検索語は期間内にニーズが発生し，期間内にニーズが終了するものとする．すなわち，分布の裾で検索数がなくなることを前提とする．

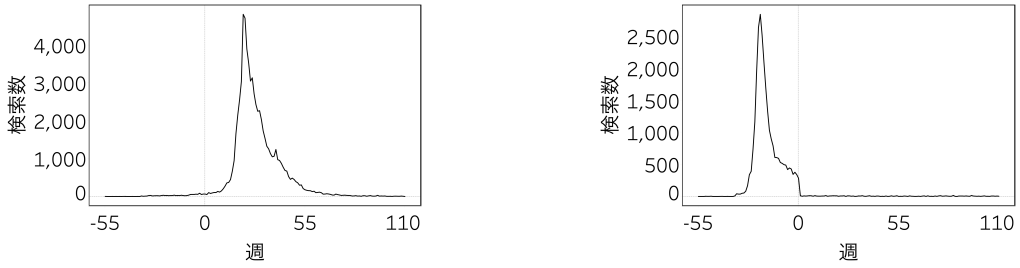
単調・非対称性

時系列検索データはピークの前後で単調に増加・減少する傾向があり，その確率分布の形状は非対称となる場合が多い．図 4.1 は検索語「離乳食」「胎動」の時系列検索分布のグラフである．

検索語「離乳食」は本研究対象とする分布の代表的な形状をもつ．一般的に離乳食に切り替わるのは産後 5 ヶ月〜6 ヶ月であり，産後 5 ヶ月頃である 21 週のピークに向けて単調に検索数が増加し，ピークの後に単調に検索数が減少する傾向が見られる．また，ピーク後はピーク前と比較して緩やかに検索数が減っており，ピークに対して非対称である．

一方，検索語「胎動」は出産前 21 週，すなわち，妊娠 19 週（5 ヶ月頃）で検索数のピークとなり，ピークに向けて単調に検索数が増加し，ピークの後は単調に検索数が減少する傾向が

見られる．また，ピーク後の検索数の減少は特徴的であり，出産前 13 週まで下降した後，出産（0 週）に向けて緩やかに検索数が減少する．最後，出産（0 週）をきっかけに検索されなくなる傾向があり，ピークに対して非対称である．



(a) 検索語「離乳食」　　(b) 検索語「胎動」

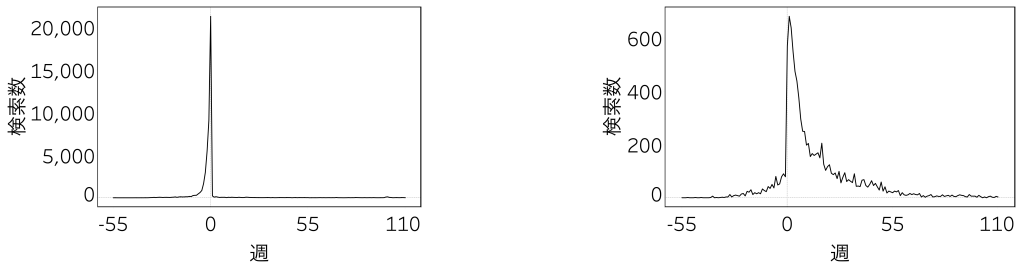
図 4.1. 単調・非対称な時系列検索分布の例

急上昇・急下降性

時系列検索分布は特定のイベントに対して急上昇，急下降する性質を持つ．図 4.2 は検索語「陣痛」「ミルク」の時系列検索分布のグラフである．

検索語「陣痛」は出産前 8 週から徐々に検索数が増えピークとなる出産前 1 週まで上昇し，出産週（0 週）で急に検索がなくなる．これは母親が出産により陣痛がなくなるという身体の変化を表している．

一方，検索語「ミルク」は出産前から検索数は徐々に増えるものの，出産週（0 週）に急上昇し，1 週目でピークに達している．これは母親が出産により子どもへの興味が生まれることを表している．「陣痛」も「ミルク」も出産前後で急激に検索数が変化する．急上昇や急下降する分布は滑らかな関数で近似することが難しく，本研究で解決したい課題の 1 つである．



(a) 検索語「陣痛」　　(b) 検索語「ミルク」

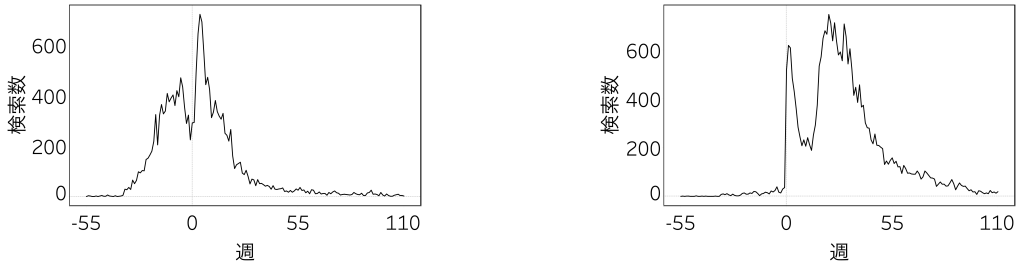
図 4.2. 急上昇・急下降する時系列検索分布の例

単峰型と二峰型

上記の例のように多くの時系列検索分布は単峰型であるが，一部で二峰型の分布もある．図 4.3 は検索語「抱っこ紐」「夜泣き」の時系列検索分布のグラフである．

検索語「抱っこ紐」は出産前 5 週と出産後 4 週にピークをもつ．ただし，出産後 4 週のピークのほうが大きなピークであり，全体の形状を考察すると出産（0 週）の前後で検索が一時的に控えられているように見える．すなわち「抱っこ紐」の興味が出産後 4 週に向けて上昇するものの，出産イベントの周辺で一時的に他の話題に興味が移ることで二峰型の分布になったと考えられる．

一方，検索語「夜泣き」は出産 1 週と出産 22 週にピークをもつ．これは，新生児が体の不快感を感じて行う「夜泣き」と産後 6 ヶ月頃に睡眠サイクルが短いことや脳が未発達のために行う「夜泣き」とで根本的に原因が異なる．



(a) 検索語「抱っこ紐」　(b) 検索語「夜泣き」

図 4.3. 二峰型の時系列検索分布の例

以上を考察すると研究対象とする時系列検索分布の特徴は次の性質にまとめられる．

- 分布の裾で検索がなくなる
- 単峰型と二峰型の確率分布が存在する
- 極大値と極小値の間で単調性が成り立つ
- ピークに対して左右非対称であり，急上昇・急下降など様々な特徴をもつ

ここでピークについては極大値と言い換え，二峰型の分布についての 1 つ目のピークと 2 つ目のピークの間の最小値を極小値と言い換えていることに注意する．上記の特徴を包含する数理最適化モデルを提案することが本研究の目標の 1 つである．

4.3.2 問題設定

本研究では，出産日を起点とする時系列検索データを扱う．まず，出産日を含む時系列の集合 T を与える．出産日を 0 として，T は 0 を含む連続する整数の有限部分集合とする．例えば，出産日を 0 として 2 時点前から 4 時点後までの時系列集合を扱う場合，$T = \{-2, -1, 0, 1, 2, 3, 4\}$

と表せる．次に，ある検索語の時点 $t \in T$ における検索数を n_t として時系列検索数列

$$(n_t \mid t \in T)$$

を定める．最後に経験分布としての時系列確率分布を定める．ある検索語が時点 $t \in T$ において検索される割合を $r_t := n_t / \sum_{t' \in T} n_{t'}$ とすれば時系列確率分布

$$(r_t \mid t \in T)$$

が得られる．$(r_t \mid t \in T)$ が確率分布の定義を満たすことは明らかである．

さて，本研究では最適化モデルを用いて確率分布の推定を行う．ある検索語が時点 $t \in T$ において検索される確率を決定変数 x_t とおけば，本問題は次の時系列確率分布

$$(x_t \mid t \in T)$$

を求める問題である．ここで，x_t は確率であるため，

$$0 \leq x_t \leq 1 \quad (t \in T) \tag{4.1}$$

が成り立ち，また，$(x_t \mid t \in T)$ は確率分布であるため

$$\sum_{t \in T} x_t = 1 \tag{4.2}$$

を満たす必要がある．また，確率分布の裾は 0 であることから

$$x_{\min(T)} = x_{\max(T)} = 0 \tag{4.3}$$

であると仮定する．また，推定する確率分布 $(x_t \mid t \in T)$ は，経験分布 $(r_t \mid t \in T)$ との誤差を最小化するため，次の目的関数を設定する．

$$\sum_{t \in T} (x_t - r_t)^2 \tag{4.4}$$

上記で定義される制約式 (4.1)，(4.2)，(4.3) に目的関数 (4.4) を最小化する最適化問題は凸二次計画問題のクラスに分類される．

4.3.3　最適化モデル

4.3.2 節で定めた問題設定に制約を加え 2 つの最適化モデルを提案する．単調単峰性モデル（MONO-UNI）と単調二峰性モデル（MONO-BI）である．最適化モデルに極大値，および極小値となる時点を定数として与えられるが，次の 2 つの点で課題がある．

- 極大値と極小値となる時点を求める方法を設計する必要がある
- 誤った極大値，または極小値を与えてしまった場合，得られる解は局所最適解となる

そのため，極大値と極小値となる時点は数理最適化モデルの中で自動で推定できることが好ましい．すなわち，極大値の存在や極小値の存在を形状として制約する．提案する数理モデルは，0-1 整数変数を導入することで極大値，および極小値となる時点を自動で推定することができる． 4.3.2 節で定めた問題は凸二次計画問題であるが，以下では混合整数凸二次計画問題に拡張される．

単調単峰性モデル：MONO-UNI

単調単峰性モデルを表現するためには極大値が存在するという形状を課す必要がある．極大値の存在を表現するための 0-1 整数変数を導入すればよい．具体的には時点 $t \in T$ において極大値前であれば 0，極大値および極大値後であれば 1 となる 0-1 整数変数 s_t の列

$$(s_t \mid t \in T)$$

を定める．このとき，

$$s_t \leq s_{t+1} \quad (t \in T, t \neq \max(T)) \tag{4.5}$$

とする．以下，単調性の制約を定義するための十分大きな定数を M とする．

まず，極大値前で単調増加する制約は次のように表現できる．

$$x_t \leq x_{t+1} + M \cdot s_t \quad (t \in T, t \neq \max(T)) \tag{4.6}$$

$s_t = 0$，すなわち極大値前では

$$x_t \leq x_{t+1}$$

となり，単調増加を表す．一方，$s_t = 1$，すなわち極大値後では

$$x_t \leq x_{t+1} + M$$

となり，M が十分大きな数であるため，常に充足される制約となり，単調増加の制約が課されなくなる．

次に極大値後で単調減少する制約は次のように表現できる．

$$x_t \geq x_{t+1} - M \cdot (1 - s_t) \quad (t \in T, t \neq \max(T)) \tag{4.7}$$

$s_t = 0$，すなわち極大値前では

$$x_t \geq x_{t+1} - M$$

となり，M が十分大きな数であるため，常に充足される制約となり，単調減少の制約が課されなくなる．一方，$s_t = 1$，すなわち極大値後では

$$x_t \geq x_{t+1}$$

となり，単調減少を表す．

以上より問題設定の目的関数 (4.4)，制約式 (4.1)，(4.2)，(4.3) に加え，制約式 (4.5)，(4.6)，(4.7) より単調単峰性モデル MONO-UNI は，次のように定められる．

$$\begin{array}{ll} \text{minimize} & \displaystyle\sum_{t \in T} (x_t - r_t)^2 \\ \text{subject to} & 0 \leq x_t \leq 1 \quad (t \in T) \\ & \displaystyle\sum_{t \in T} x_t = 1 \quad (t \in T) \\ & x_{\min(T)} = x_{\max(T)} = 0 \\ & s_t \leq s_{t+1} \quad (t \in T, t \neq \max(T)) \\ & x_t \leq x_{t+1} + M \cdot s_t \quad (t \in T, t \neq \max(T)) \\ & x_t \geq x_{t+1} - M \cdot (1 - s_t) \quad (t \in T, t \neq \max(T)) \end{array} \tag{4.8}$$

単調単峰性モデルは単峰回帰の拡張であり，極大値をとる時点 t を自動で推定できる．また，標準的な数理最適化ソルバーで厳密解を簡単に得られるという実用上の利点がある．

単調二峰性モデル：MONO-BI

単調二峰性モデルを表現するためには 1 つ目の極大値，1 つ目の極小値，2 つ目の極大値が存在するという形状を課す必要がある．1 つ目の極大値，1 つ目の極小値，2 つ目の極大値を表現するためには 0-1 整数変数を 3 種類導入すればよい．具体的には時点 $t \in T$ において 1 つ目の極大値前であれば 0，極大値，および極大値後であれば 1 となる 0-1 整数変数 s_t の列

$$(s_t \mid t \in T)$$

時点 $t \in T$ において 1 つ目の極小値前であれば 0，極小値，および極小値後であれば 1 となる 0-1 整数変数 u_t の列

$$(u_t \mid t \in T)$$

時点 $t \in T$ において 2 つ目の極大値前であれば 0，極大値，および極大値後であれば 1 となる 0-1 整数変数 v_t の列

$$(v_t \mid t \in T)$$

を定める．このとき，単調単峰性モデルでも定義したように以下の制約を加える．

$$s_t \leq s_{t+1} \quad (t \in T, t \neq \max(T)) \tag{4.9}$$
$$u_t \leq u_{t+1} \quad (t \in T, t \neq \max(T)) \tag{4.10}$$
$$v_t \leq v_{t+1} \quad (t \in T, t \neq \max(T)) \tag{4.11}$$

以下，単調性の制約を定義するための十分大きな定数を M とする．単調単峰性モデルの定義と同様に各単調性の制約は次のように定める．

$$x_t \leq x_{t+1} + M \cdot s_t \quad (t \in T, t \neq \max(T)) \tag{4.12}$$
$$x_t \geq x_{t+1} - M \cdot (1 - s_t) - M \cdot u_t \quad (t \in T, t \neq \max(T)) \tag{4.13}$$
$$x_t \leq x_{t+1} + M \cdot (1 - u_t) + M \cdot v_t \quad (t \in T, t \neq \max(T)) \tag{4.14}$$
$$x_t \geq x_{t+1} - M \cdot (1 - v_t) \quad (t \in T, t \neq \max(T)) \tag{4.15}$$

さらに 1 つ目の極大値の後に 1 つ目の極小値をとることを課する制約と 1 つ目の極小値の後に 2 つ目の極大値をとることを課する制約として次の 2 つの制約を定める．

$$s_t \geq u_t \quad (t \in T) \tag{4.16}$$
$$u_t \geq v_t \quad (t \in T) \tag{4.17}$$

以上より問題設定の目的関数 (4.4)，制約式 (4.1)，(4.2)，(4.3) に加え，制約式 (4.9)，(4.10)，(4.11)，(4.12)，(4.13)，(4.14)，(4.15)，(4.16)，(4.17) より単調二峰性モデル MONO-BI は，

次のように定められる．

$$
\begin{aligned}
\text{minimize} \quad & \sum_{t\in T}(x_t - r_t)^2 \\
\text{subject to} \quad & 0 \le x_t \le 1 \quad (t \in T) \\
& \sum_{t\in T} x_t = 1 \quad (t \in T) \\
& x_{\min(T)} = x_{\max(T)} = 0 \\
& s_t \ge u_t \quad (t \in T) \\
& u_t \ge v_t \quad (t \in T) \\
& s_t \le s_{t+1} \quad (t \in T, t \ne \max(T)) \\
& u_t \le u_{t+1} \quad (t \in T, t \ne \max(T)) \\
& v_t \le v_{t+1} \quad (t \in T, t \ne \max(T)) \\
& x_t \le x_{t+1} + M \cdot s_t \quad (t \in T, t \ne \max(T)) \\
& x_t \ge x_{t+1} - M \cdot (1 - s_t) - M \cdot u_t \quad (t \in T, t \ne \max(T)) \\
& x_t \le x_{t+1} + M \cdot (1 - u_t) + M \cdot v_t \quad (t \in T, t \ne \max(T)) \\
& x_t \ge x_{t+1} - M \cdot (1 - v_t) \quad (t \in T, t \ne \max(T))
\end{aligned}
\tag{4.18}
$$

単調二峰性モデルは単峰回帰の拡張であり，二峰型の分布を表現でき，極大値，極小値を自動で推定できる．また，標準的な数理最適化ソルバーで厳密解を簡単に得られると言う実用上の利点がある．

十分大きな定数 M の定義

単調単峰性モデル，単調二峰性モデルで十分大きな定数 M を与えたが，ここでは具体的な M の値を定める．一般に十分大きな M としてなるべく小さい値をとることで実行可能領域が狭まり，数理最適化ソルバーが効率よく解を探索することができる．ここでは，簡単に求められる M の値を定める．式 (4.6) において $s_t = 1$ となる任意の $t \in T$ について $x_t \le x_{t+1} + M$ を必ず充足される制約とするためには $M = max\{x_t | t \in T\}$ とおけばよい．具体的には式 (4.1) から x_t の最大値は 1 なので自明な値として $M = 1$ とおける．

4.4 実データによる数値実験

本節では実データを利用して提案手法の有効性を検証する．提案手法である単調単峰性モデルと単調二峰性モデルの予測性能を経験分布，移動平均，カーネル回帰と比較する．いずれの手法もノンパラメトリックな手法であり，予測精度がデータ量に依存するモデルであるため，学習データの量を調整した 3 つの学習データセットで数値実験を行った．データセット，評価方法，実験環境について説明し，最後に評価と考察を行う．

4.4.1 データセット

本実験ではコネヒト株式会社が運営するサービス「ママリ」における検索履歴を用いる．2018 年 9 月 1 日から 2020 年 8 月 31 日までの 2 年間の検索履歴を対象とした．本検索履歴は検索日時だけでなく，出産前後何日の時点で検索したか特定できるデータである．本データに

次の前処理を施して下記の 3 種類の条件をすべて満たす 76 語を最終的な分析対象とした．なお，検索語 76 語の期間内検索数の総数は 1,252,286 件であった．

- 出産前 1 年から出産後 2 年までの期間に全体の検索数の 95% が含まれる．
- コネヒト株式会社指定のマーケティングに関わる検索語リストに含まれる．
- 期間内検索数が 8,000 件以上である．

本データセットから数値実験用のデータセットを作成する．数値実験用のデータセットは学習用データとテスト用データが必要である．

まず，本データセットから各検索語に対して 5,000 件の検索履歴を抽出してテストデータとした．一方，学習データはテストデータと重複しないように 3 種類作成し，学習データ 1 は 3,000 件，学習データ 2 は 1,500 件，学習データ 3 は 100 件の検索履歴を無作為に抽出した．学習データ 1，2 は学習データが十分にある場合の数理モデルの性能評価を目的としており，学習データ 3 は学習データが少量の場合の数理モデルの性能評価を目的としている．

4.4.2　評価方法

提案手法である単調単峰性モデル MONO-UNI，単調二峰性モデル MONO-BI に加え，比較手法として経験分布 EMP，移動平均 MA，カーネル回帰 KR を用いて予測性能を評価した．なお，経験分布は学習データと同じ分布であり，移動平均の期間は 2 期間を採用した．カーネル回帰は，ガウシアンカーネルを用いた Nadaraya-Watoson の推定量を用いており，最小二乗交差確認によりバンド幅を決定した．評価尺度は平方平均二乗誤差（RMSE）とする．

4.4.3　実験環境

単調単峰性モデル MONO-UNI，単調二峰性モデル MONO-BI は混合整数凸二次計画問題を解く必要があるため数理最適化問題を解くソフトウェアである Gurobi Optimizer （ver. 9.0.3)[*3]を用いた．移動平均 MA は，Python 言語のライブラリ pandas（ver. 1.1.1）に実装されている pandas.DataFrame.rolling 関数[*4]を利用した．また，カーネル回帰 KR は，Python 言語のライブラリ scipy（ver. 1.5.2）に実装されている scipy.statsmodels.nonparametric.kernel_regression.KernelReg 関数[*5]を利用した．

4.4.4　評価

本項では提案手法の有効性について評価する．表 4.1 は各データ量に対する経験分布 EMP，移動平均 MA，カーネル回帰 KR，単調単峰性モデル MONO-UNI，単調二峰性モデル MONO-BI の

[*3] https://www.gurobi.com/products/gurobi-optimizer/
[*4] https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.rolling.html
[*5] https://www.statsmodels.org/stable/generated/statsmodels.nonparametric.kernel_regression.KernelReg.html

5 つの手法それぞれについての 76 種類の検索語に対する RMSE の平均である．最も予測性能が良かった（予測誤差が小さかった）値を太字としている．なお，括弧内は標準誤差である．

はじめに，学習データ量に対する全体の予測性能について確認する．いずれの手法も学習データ量が最も多い学習データ 1 で予測性能が最も高く，学習データ量が少なくなるほど予測性能が劣化する．特に学習データ 3 における予測性能は著しく悪化している．

次に各手法を学習データ量の観点で評価をする．まず，いずれの学習データでも MA の予想性能が最も悪かった．学習データが十分にある場合，すなわち学習データ 1 では提案手法 MONO-BI が最も良い予測性能を示し，学習データ 2 では提案手法 MONO-UNI と MONO-BI の 2 つが同等に良い予測性能を示した．また，学習データ 1，学習データ 2 ともに MONO-BI と MONO-UNI に続き EMP，KR の順で予測性能が良かった．

表 4.1. 実データに対する RMSE 平均と標準誤差（$\times 10^3$）

	学習データ 1 3,000 件	学習データ 2 1,500 件	学習データ 3 100 件
EMP	1.70 (± 0.03)	2.19 (± 0.04)	7.68 (± 0.12)
MA	3.72 (± 0.38)	3.90 (± 0.36)	6.60 (± 0.30)
KR	2.01 (± 0.28)	2.32 (± 0.27)	**4.82** (± 0.38)
MONO-UNI	1.65 (± 0.04)	**1.94** (± 0.04)	5.07 (± 0.17)
MONO-BI	**1.57** (± 0.03)	1.95 (± 0.04)	6.01 (± 0.16)

具体的に学習データ 1 における検索語「ファーストシューズ」に対する考察を与える．表 4.2 は検索語「ファーストシューズ」の RMSE である．比較手法と比べて提案手法である MONO-UNI，MONO-BI の順で性能が良いことがわかる．

表 4.2. 学習データ 1 における検索語「ファーストシューズ」に対する各手法の RMSE（$\times 10^3$）

	RMSE
EMP	1.71
MA	1.78
KR	1.72
MONO-UNI	1.33
MONO-BI	1.63

実際の時系列確率分布を図 4.4 に示す．学習データと同じ分布である EMP と各手法を比較することで学習の傾向を確認できる．MA と KR は EMP に比べて滑らかであり，特に KR は EMP の尖りが滑らかに補正されている．一方，MONO-UNI と MONO-BI は，EMP の尖りを残したまま単調性が成り立つ分布に補正されている．また，EMP，MA，KR は分布の裾でノイズが残っているが，一方で MONO-UNI と MONO-BI は裾ではノイズが排除されている．以上を考察すると，

MA と KR が MONO-UNI と MONO-BI に比べて予測性能で劣るのは，極大値周辺の急上昇，急下降を表現できないことと分布の裾のノイズを排除できないためであると考えられる．

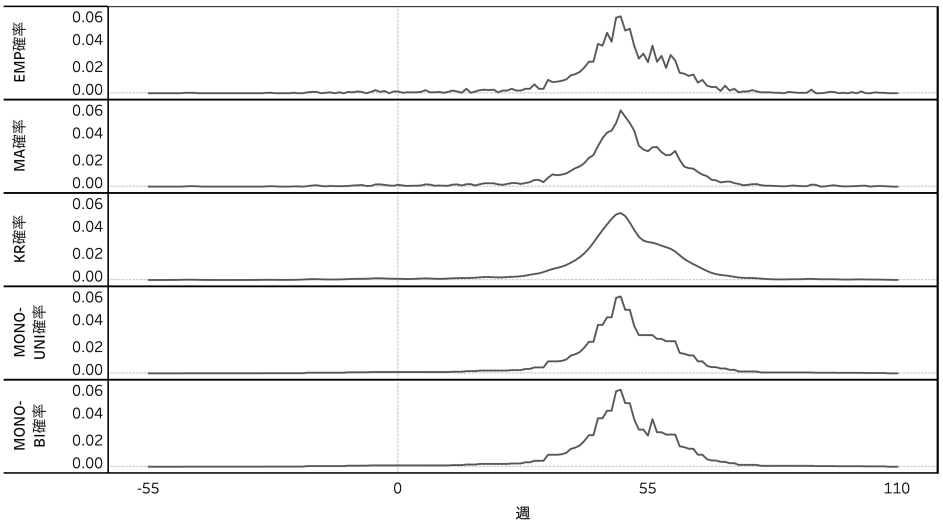


図 4.4. 検索語「ファーストシューズ」

また，学習データ 1 において MONO-BI が MONO-UNI よりも予測性能が良いのは，二峰型の確率分布も表現できるためである．図 4.5 は学習データ 1 における検索語「夜泣き」に対する EMP，MONO-UNI，MONO-BI の時系列確率分布である．MONO-BI は MONO-UNI と異なり，二峰性の特徴を学習できることが確認できる．実際，MONO-UNI の RMSE（$\times 10^3$）は 2.26，MONO-BI の RMSE（$\times 10^3$）は 1.55 と大幅に MONO-BI の予想性能が MONO-UNI の予測性能を上回っている．

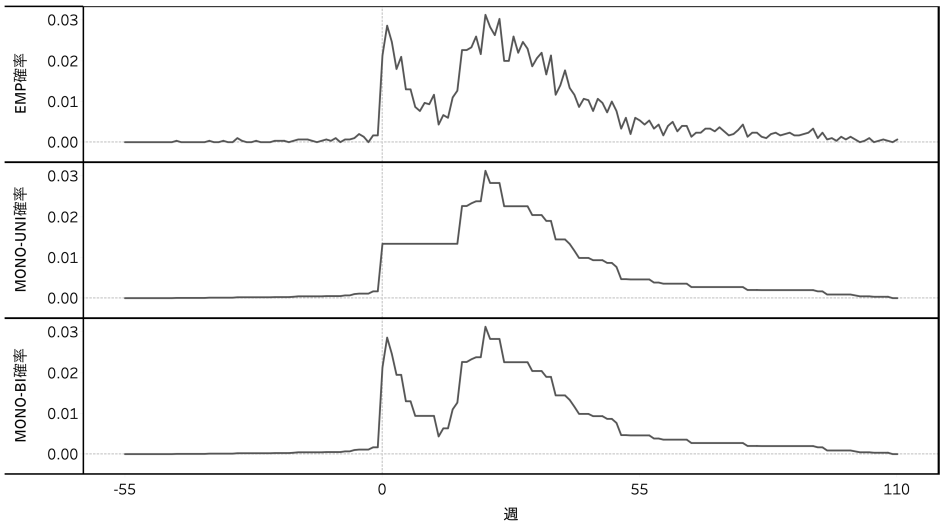


図 4.5. 検索語「夜泣き」の時系列確率分布（学習データ 1）

続けて学習データが少量である場合を評価する．学習データ 3 では提案手法である MONO-BI

と MONO-UNI の予測性能が劣化し，KR の予測性能が最も良い．具体的に学習データ 3 における検索語「鼻水吸引器」に対する考察を与える．表 4.3 は検索語「鼻水吸引器」の RMSE である．

表 4.3. 学習データ 3 における検索語「鼻水吸引器」に対する各手法の RMSE（$\times 10^3$）

	RMSE
EMP	8.74
MA	6.47
KR	2.74
MONO-UNI	4.86
MONO-BI	5.67

MONO-UNI と MONO-BI の予測性能よりも KR の予測性能が良い．実際の時系列確率分布を図 4.6 に示す．データ量が少ないため EMP の形状が多峰性の分布となっており，MONO-BI が EMP の分布に過剰に適合している．特に分布の裾が広い検索語でこの傾向がある．また，KR の予測性能の劣化が他の手法と比べて小さいのは平滑化によって頑健性が保たれるためと考えられる．

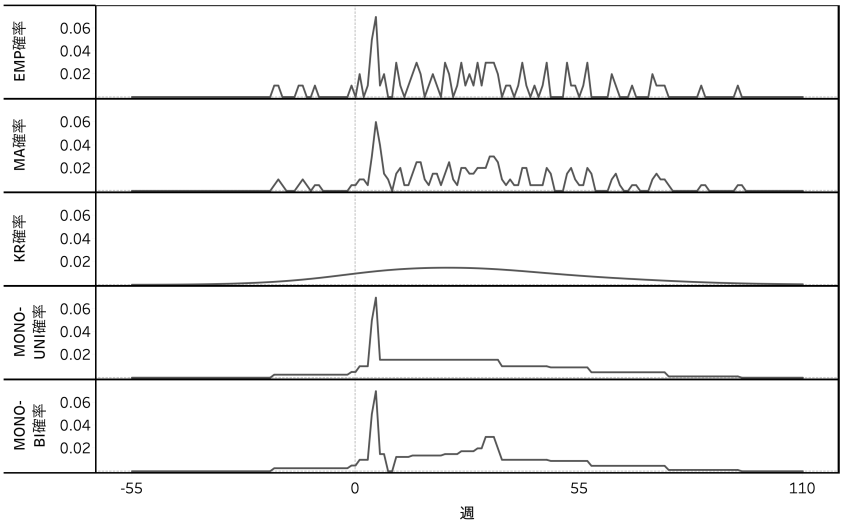


図 4.6. 検索語「鼻水吸引器」の時系列確率分布（学習データ 3）

最後に各手法の安定性について評価する．表 4.1 から明らかなように提案手法である MONO-UNI，MONO-BI は EMP と同等の標準誤差となっており，予測モデルとして安定している．一方，MA，KR は標準誤差の値が非常に大きく検索語によって性能が不安定である．これは，MA や KR が滑らかさを仮定するモデルであるため，急上昇や急降下する分布に対応できず，大きな予測誤差が生じることによる．この点で MONO-UNI，MONO-BI は非常に有効なモデルであると言える．

4.5　人工データによる数値実験

4.4 節では，実データにおける提案手法の有効性の検証をした．提案手法の単調単峰性モデルと単調二峰性モデルは，尖った分布やノイズが含まれる分布に有効であると考察を与えたが，本節では人工データを利用してその有効性の検証をする．具体的には，単峰型の確率分布である離散ラプラス分布とポアソン分布の推定についてノイズを含まない場合とノイズを含む場合，学習データが十分にある場合と少量の場合の組合せで学習データセットを用意して単調単峰性モデルとカーネル回帰を比較する．離散ラプラス分布は尖った分布における精度評価を目的としており，一方でポアソン分布は尖ってない分布における精度評価を目的としている．評価方法，および実験環境は実データによる数値実験と同じである．以下ではデータセットの作成方法について説明した後，評価と考察を行う．

4.5.1　データセット

本実験では単峰型の確率分布である離散ラプラス分布とポアソン分布を用いる．以下，離散ラプラス分布とポアソン分布で用いられる a と μ は定数であることに注意する．離散ラプラス分布は整数 k に対して

$$f(k) = \tanh(1/2)\exp(-a|k|)$$

で表される．本実験では尖った分布の代表として用いる．一方，ポアソン分布は非負整数 k に対して

$$f(k) = \exp(-\mu)\frac{\mu^k}{k!}$$

で表され，大きな μ の値では正規分布と似た分布になることが知られている．本実験では尖ってない分布の代表として用いる．

実データにおける実験と同様にデータセットは学習データとテストデータを用意する．まず，それぞれの分布の確率質量関数を用いてテストデータを作成した．続けて，学習データは離散ラプラス分布とポアソン分布から無作為抽出によって 2 種類を作成した．データ量は 10,000 件と 100 件である．また，離散ラプラス分布のパラメータは $a \in \{1, 2\}$，ポアソン分布のパラメータは $\mu \in \{1, 10\}$ と設定した．図 4.7 は，各種パラメータにおける離散ラプラス分布とポアソン分布である．ポアソン分布について $\mu = 1$ と $\mu = 10$ を比べると $\mu = 10$ のほうが正規分布に近くなり，尖ってない分布となる．一方，離散ラプラス分布について $a = 1$ と $a = 2$ を比べると $a = 2$ のほうが尖った分布となる．上記の条件から作成する 2 通りの分布，2 通りの学習データ，2 通りのパラメータから 8 種類のデータについて提案手法の有効性を検証する．

また，分布にノイズが含まれる場合にも提案手法が有効であることを検証するため，上記で作成した 8 種類の学習データに，対象期間内から離散一様分布を用いて 10 個の時点を無作為に復元抽出し，各時点に学習データ量の 2% にあたるノイズを加え学習データを作成した．そ

のためノイズを含む学習データは元の学習データの量に対して 20% のノイズ（検索数）を含むことになる．

以上より 16 種類の分布が得られるが，それぞれ無作為抽出を行っているため乱数の種に予測精度が依存する．そのため，乱数の種を 100 種類用意し，16 種類の分布に対してそれぞれ 100 個の学習データ，全部で 1600 個の学習データを用意した．

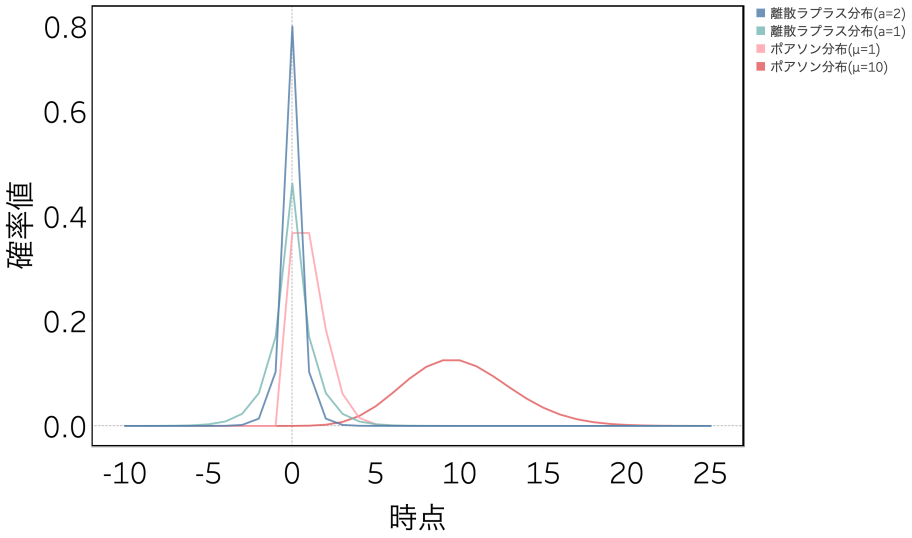


図 4.7. 離散ラプラス分布とポアソン分布

4.5.2　評価

本項では提案手法が尖った分布に有効であること，およびノイズが含まれる分布に有効であることを評価する．

まず，与えられた正解データと学習データに対して，どのような推定がされているか確認する．図 4.8 は，ノイズありのポアソン分布（$\mu = 10$）のデータに対してカーネル回帰 KR と単調単峰性モデル MONO-UNI で予測した結果である．カーネル回帰 KR は全てのノイズの影響を受けているが，一方で単調単峰性モデル MONO-UNI はピークに近い時点のみノイズの影響を受けている．

表 4.4 は，各種条件における各手法の RMSE 平均と標準誤差である．RMSE 平均は 100 個の学習データセットに各手法を適用した RMSE から計算されている．条件ごとに手法 EMP，KR，MONO-UNI の中で最も良かった手法の精度を太文字とした．学習データ量が十分にある場合（10,000 件）と学習データ量が少ない場合（100 件）について，ノイズがない場合とノイズがある場合，および尖った分布と尖ってない分布を比較することができる．なお，離散ラプラス分布（$a = 2$），離散ラプラス分布（$a = 1$），ポアソン分布（$\mu = 1$），ポアソン分布（$\mu = 10$）の順番で分布が尖っていることに注意する．

はじめに全体の傾向について確認する．学習データ量が十分にある場合（10,000 件）と学習

データ量が少ない場合（100 件）を比較すると，いずれの場合も学習データ量が十分にある場合のほうが予測精度が良い．また，ノイズがない場合とノイズがある場合を比較すると，いずれの場合もノイズがない場合のほうが予測精度が良い．

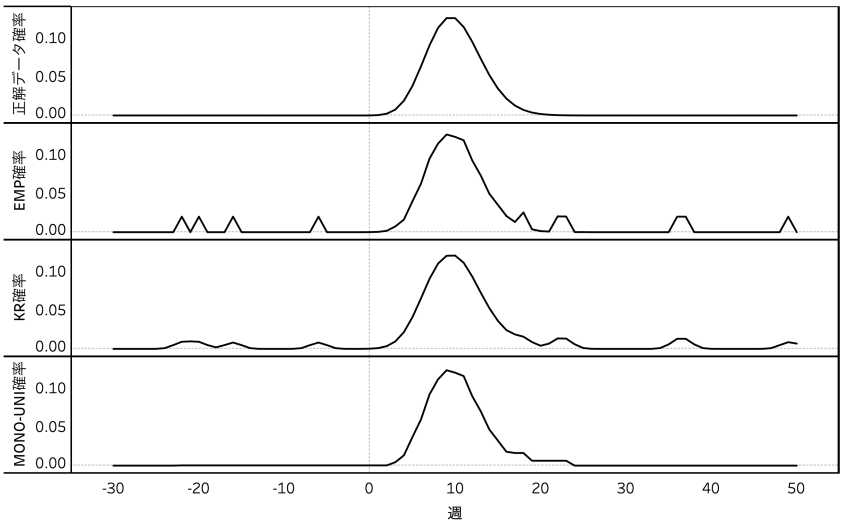


図 4.8. ノイズありポアソン分布（$\mu = 10$）の推定結果の例

表 4.4. 人工データにおける各手法の RMSE 平均と標準誤差（$\times 10^3$）

			学習データ量					
			10,000 件			100 件		
ノイズ	分布	パラメータ	EMP	KR	MONO-UNI	EMP	KR	MONO-UNI
なし	離散ラプラス	$a = 2$	**0.65** (± 0.03)	58.78 (± 0.11)	**0.65** (± 0.03)	6.49 (± 0.29)	59.87 (± 1.36)	**6.44** (± 0.29)
		$a = 1$	**0.92** (± 0.03)	**0.92** (± 0.03)	**0.92** (± 0.03)	9.22 (± 0.34)	13.11 (± 0.74)	**9.02** (± 0.34)
	ポアソン	$\mu = 1$	**0.80** (± 0.04)	**0.80** (± 0.04)	**0.80** (± 0.04)	8.64 (± 0.38)	9.07 (± 0.46)	**8.62** (± 0.38)
		$\mu = 10$	**1.02** (± 0.02)	**1.02** (± 0.02)	**1.02** (± 0.02)	10.55 (± 0.22)	**5.89** (± 0.29)	8.37 (± 0.21)
あり	離散ラプラス	$a = 2$	7.40 (± 0.05)	59.24 (± 0.29)	**2.02** (± 0.07)	16.20 (± 0.31)	63.35 (± 1.33)	**15.20** (± 0.33)
		$a = 1$	7.41 (± 0.05)	7.65 (± 0.17)	**2.53** (± 0.07)	13.35 (± 0.26)	18.42 (± 0.68)	**12.02** (± 0.28)
	ポアソン	$\mu = 1$	7.43 (± 0.05)	7.43 (± 0.05)	**2.17** (± 0.08)	13.58 (± 0.20)	14.29 (± 0.36)	**12.40** (± 0.22)
		$\mu = 10$	7.46 (± 0.04)	4.79 (± 0.08)	**3.12** (± 0.07)	11.90 (± 0.15)	**8.70** (± 0.19)	8.86 (± 0.15)

尖った分布への有効性

提案手法である単調単峰性モデル MONO-UNI の尖った分布への有効性を評価するため，ノイズが含まれていない場合の実験結果を確認する．

まず，学習データ量が十分にある場合（10,000 件）に注目すると，最も尖った分布である離散ラプラス分布（$a = 2$）以外の 3 つの分布で経験分布 EMP，カーネル回帰 KR，単調単峰性モデル MONO-UNI の全ての手法で同じ予測精度となっている．これはデータが十分にある場合にはいずれのモデルを利用しても同程度の予測性能であることを表している．しかし，最も尖った分布である離散ラプラス分布（$a = 2$）の場合でカーネル回帰の RMSE 平均が 58.78 となっており，著しく予測精度が悪い．次に，学習データ量が十分にある場合（100 件）に注目すると，同様に最も尖った分布である離散ラプラス分布（$a = 2$）の場合でカーネル回帰の RMSE

平均が 59.87 となっており，著しく予測精度が悪い．しかし，最も尖ってない分布であるポアソン分布（$\mu = 10$）の場合では，RMSE 平均が 5.89 となっており，最も予測精度が良い．これは，カーネル回帰 KR が平滑化を目的とする予測モデルであるため，尖ってない分布の予測は得意であるが，尖った分布の予測が苦手であることを表している．対照的に，提案手法である単調単峰性モデル MONO-UNI は，学習データ量が少ない場合（100 件）でも，最も尖ってない分布であるポアソン分布（$\mu = 10$）以外の全ての分布，すなわちポアソン分布（$\mu = 1$），離散ラプラス分布（$a = 1$），離散ラプラス分布（$a = 2$）において予測精度が最も良い．これは実データにおける数値実験で考察を与えたように尖ってない分布でかつ学習データ量が少ないと多峰型の分布となり，単調単峰性モデルの適用が不適切となるためである．以上の評価から，提案手法である単調単峰性モデル MONO-UNI は学習データが十分にある，または尖った分布に対して有効であることがわかる．

ノイズが含まれる分布への有効性

提案手法である単調単峰性モデル MONO-UNI のノイズが含まれる分布への有効性を評価する．ノイズがある 8 個の結果のうち 7 個の結果で提案手法である単調単峰性モデル MONO-UNI の予測精度が最も良い．唯一，提案手法が不得意とする尖っておらず，データ量が少ない場合であるポアソン分布（$\mu = 10$）の場合でもカーネル回帰 KR と近い RMSE 値を示している．

また，ノイズがない場合と比較すると，経験分布 EMP とカーネル回帰 KR の予測精度が大幅に悪くなっているが，一方で提案手法である単調単峰性モデル MONO-UNI の予測精度は悪化の割合が小さい．これは，経験分布 EMP とカーネル回帰 KR がノイズに過剰適合しやすく，提案手法である単調単峰性モデル MONO-UNI はノイズに強いことを表している．

4.6 まとめ

本章では，コネヒト株式会社が運営するサービス「ママリ」から提供された検索データを研究対象に時系列確率分布のノンパラメトリック推定について論じた．本検索データの特徴は，出産前後の時点で何を検索したかがわかる点である．そのため出産日を起点とした時系列検索データとして表現でき，特に確率分布として解析を行った．

経験分布としての時系列確率分布は次の特徴をもつ．

- 分布の裾で検索がなくなる
- 単峰型と二峰型の確率分布が存在する
- 極大値と極小値の間で単調性が成り立つ
- ピークに対して左右非対称，急上昇・急下降など様々な特徴をもつ

そこで，本研究では数理最適化モデルとして単調単峰性モデルと単調二峰性モデルを提案して時系列確率分布の推定を行った．これらのモデルは極大値，および極小値を自動で推定することができ，混合整数凸二次計画問題として定式化されているため，標準的な数理最適化ソルバーで厳密解を得られる．

提案手法の有効性を示すために実データと人工データを用いて数値実験を行った．実データを利用した実験では提案手法である単調単峰性モデルと単調二峰性モデルを経験分布，移動平均，カーネル回帰と比較した．予測精度を RMSE で評価した結果，提案する単調二峰性モデルの精度が最も良く，標準誤差も小さいために安定して予測できることが示された．ただし，学習データが極端に少なく，単峰型，二峰型の分布の外形が得られないような多峰型の分布となっている場合には，単調二峰性モデルの性能が劣化し，カーネル回帰が最もよい精度であった．一方，人工データを利用した実験では離散ラプラス分布とポアソン分布を用い，ノイズがない場合とノイズがある場合について実験を行った．ノイズがない場合の実験において提案手法である単調単峰性モデルとカーネル回帰の予測精度を RMSE で評価した結果，データ量が十分にある場合，または尖った分布に対して単調単峰性モデルが有効であることがわかった．しかし，尖ってない分布でかつデータ量が少ない場合には単調単峰性モデルよりもカーネル回帰のほうが精度がよかった．また，ノイズがある場合の実験については単調単峰性モデルがカーネル回帰や経験分布と比べてノイズに過剰適合しにくいことが確認できた．

次に本研究の課題について述べ今後の課題について言及する．提案手法は混合整数凸二次計画問題に定式化をしているため計算コストが高い．本研究が対象とする出産前後の検索語に関する時系列検索データは，時系列の粒度を 1 週間に設定し，出産前 1 年から出産後 2 年ほどの期間を対象としているため提案手法を用いて現実的な時間で時系列確率分布の推定が可能である．しかし，1 日単位のようなより細かい時系列の粒度で分析する必要がある場合や出産後数十年後までの期間に対象が拡大した場合に，計算コストが高くなるため提案手法の適用が困難になることが想定される．その場合，極大値や極小値が存在する期間を予め推定し，その期間にのみ 0-1 整数変数を定義することで問題規模を縮小するなどの実装上の工夫が必要である．なお，極大値や極小値が存在する期間を理論的に保証できるかどうかについては今後の研究課題である．また，本研究の興味の対象である時系列確率分布は，1 次元上の単峰型と二峰型の分布であったが，三峰以上の多峰型の分布や多次元上の分布に提案手法を適用する場合には同様に問題規模を縮小するための実装上の工夫が必要である．多峰性の分布の推定，および多次元上の分布の推定についても今後の研究課題である．また，本提案手法により各検索語の時系列確率分布を推定することができるが，検索語間の詳細な解析も課題である．例えば，検索語間の分布を比較することで類似性や順序構造を解析することである．

最後に本提案手法の実用上の有効性について触れる．提案手法は計算コストは高いが，一方で様々な特徴を持つ分布に対して柔軟で汎用的に推定できる点で優れている．実際，単調二峰性モデルにより今回の研究対象である検索データに対して高い推定精度が得られている点は注目すべきである．パラメトリックな推定手法では様々な特徴を持つ分布を汎用的に推定することは困難であり，検索語ごとに適切な分布を選ぶ必要やパラメータの調整が必要である．一方でノンパラメトリックな推定手法であるカーネル回帰は，平滑化を目的とするため急上昇や急下降する時系列データの推定を不得意とし，ノイズに敏感である点で予測性能の劣化が起きる．以上より，提案手法はパラメトリックな推定手法よりも柔軟で汎用的であり，カーネル回帰と比べても平滑化の仮定がないため尖った分布にも対応できる．また，他の手法と比較して，形状制約によりノイズへの過剰適合を軽減できる点も注目に値する．

第 5 章

結論

本論文は，インターネット上のサービスで収集される EC サイトの閲覧履歴やスマホアプリの検索データを対象に，商品の購買確率や時系列確率分布をノンパラメトリックに推定する手法を提案した．特にノイズが多く含まれる大規模データが作る経験分布に対して数理最適化モデルを利用して形状を制約した推定を行った．ここで，ノンパラメトリックとは特定の関数形を仮定しないことを指す．単調性に加えて，凸性凹性，極大値や極小値を持つなどの形状を制限する弱い仮定をすることで，経験分布の性質を活かした柔軟で精度の高い推定を行うことができる．

5.1　主要な結果

2〜4 章において，形状制約を統合的に利用可能な数理最適化モデルを提案し，応用事例を通して有効性を示した．以下では各章の主要な結果をまとめる．

2 章では EC サイトの閲覧履歴を対象に数理最適化モデルを利用して商品の選択確率をノンパラメトリックに推定する方法について論じた．閲覧履歴から利用者の商品に対する最新度と頻度に基づく選択確率を経験分布として求め，単調性，凸性凹性を仮定した最尤推定の問題を非線形計画問題として定式化し，商品の選択確率を 2 次元確率表として推定した．数値実験の結果，パラメトリックな手法であるロジスティック回帰やノンパラメトリックな手法であるカーネルサポートベクトルマシンと比較して提案手法の予測性能が良いことが示された．また，単調性や凸性凹性の制約を与えることで学習データが少ない場合でも高い予測性能を発揮することがわかった．提案手法は予測タスクにおいて安定性，柔軟性，拡張性の利点を持つ．

3 章では 2 章で提案した最新度と頻度に対して商品選択確率を紐付ける 2 次元確率表を利用して，協調フィルタリングの性能改善ができることについて論じた．2 次元確率表は利用者の商品に対する選好を数値化できるため，協調フィルタリングを実行するときに必要な評価値行列の作成に利用する方法を提案した．数値実験の結果，利用者間型，および非負値行列分解による商品推薦タスクにおいて予測性能が良くなることが示された．また，学習データが少ない場合でも高い予測性能を発揮することがわかった．これは協調フィルタリングにおけるコールドスタート問題を緩和する．

4 章では，スマホアプリの検索履歴を対象に数理最適化モデルを利用して時系列確率分布をノンパラメトリックに推定する方法について論じた．出産前後の検索データには次の特徴がある．すなわち，単峰性または二峰性の分布であり，単調性や裾で検索が無くなるといった特徴をもつ．そこで，極大値や極小値の存在に加え，単調性や分布の裾で 0 になるといった形状を制約した単調単峰性モデルと単調二峰性モデルを提案した．実データを用いた数値実験の結果，学習データが十分にある場合にはノンパラメトリックな手法である経験分布，移動平均，カーネル回帰と比較して提案手法の予測性能が良いことが示された．一方で学習データが極端に少なく，単峰性，二峰性の外形すら得られない多峰性の分布となっている場合には，単調二峰性モデルの性能が劣化し，カーネル回帰が最も良い性能であることが確認できた．また，推定した時系列確率分布間に通常の確率順序を導入することで時系列確率分布の集合に順序構造を定義した．提案手法は経験分布やカーネル回帰よりも多くの順序を定義することができるため，時系列確率分布間の順序に関する解析がしやすいことがわかった．また，人工データを用いた数値実験により提案手法が尖っている分布やノイズを含む分布の予測にも頑健であることが示された．

5.2　形状制約ノンパラメトリック推定の注意点

本研究で提案する数理最適化モデルを利用したノンパラメトリックな推定手法は，ノイズの有無，学習データ量，およびデータの特徴に注意する必要がある．

一般に学習データ量が多いほど経験分布は有効であり，さらにノイズが一切ない状況であればデータが十分にある場合，経験分布が最適なモデルとなる．しかし，実際に得られるデータは測定誤差などのノイズが含まれる．特にインターネット上で得られるデータはサービス利用者の独自性だけでなく，運営企業の施策の影響も受けるためノイズが多く含まれる．ノイズの除去には，数理最適化モデルを利用して形状制約を入れる手法が有効である．実際，2 章で閲覧履歴から商品の選択確率を推定する場合と 4 章で検索履歴から時系列確率分布を推定する場合にも提案手法と経験分布の予測性能を比較することで有効性を確認できる．また，数理最適化モデルを利用する場合，関数形を仮定するパラメトリックな推定手法と同様にデータの特徴を把握することが重要である．実際，閲覧履歴から商品の選択確率を推定する場合には最新度と頻度に単調性や凸性凹性が成り立つ特徴が重要である．また，検索履歴から時系列確率分布を推定する場合には単峰性や二峰性，分布の裾で検索がなくなる特徴が重要である．これらの特徴を形状制約として課すことは，関数を仮定するほど強くない過不足がない仮定といえる．適切な仮定をすれば，閲覧履歴から商品の選択確率を推定する場合に確認できたように学習データが少量の場合でも頑健な予測モデルを構築できる．一方，検索履歴から時系列確率分布を推定する場合には学習データが少量のときに経験分布が多峰性の分布となり，単調二峰性モデルでは過剰適合してしまうことが確認できた．これは学習データ量が少量の場合には単調二峰性モデルの仮定が弱すぎることを表している．

以上をまとめると，ノイズを含むような大規模データに対して数理最適化モデルを利用して形状制約を課した推定手法は，経験分布の柔軟な性質を保つと同時にノイズを除去できるため

高い予測性能を発揮する．また，学習データ量とデータの特徴に応じてどの程度の強さの仮定をするかも重要である．学習データ量に応じた適切な形状を仮定することでより高い予測性能を発揮することができる．

5.3 EC サイトへの社会実装

本研究で取り上げたテーマはいずれも社会実装を前提とした研究である．本節では第 2 章と第 3 章の社会実装の方法について触れる．第 2 章と第 3 章のテーマは EC サイトの閲覧履歴を用いた商品推薦である．第 2 章では EC サイト利用者に対して，過去に閲覧した商品の選択確率を推定し，利用者にとって既知の商品を推薦するタスクである．第 3 章では第 2 章で求めた商品の選択確率を利用した協調フィルタリングアルゴリズムを適用することで，利用者にとって新しい商品を推薦するタスクである．以下では，EC サイトにおける商品推薦タスクへの社会実装について説明する．本研究を社会実装するために，2 次元確率表作成と商品推薦の 2 段階に分けて説明する．

5.3.1 2 次元確率表作成

提案手法は，コールドスタート問題にも頑健であるため，2 次元確率表は 1 度作成すれば以降も実用上十分な性能を発揮する．次のステップを踏んで 2 次元確率表は作成される．

ステップ 1（閲覧履歴の取得） 利用者が EC サイトを利用して商品閲覧や商品購買を行なう．閲覧，および購買した商品の ID とタイムスタンプをデータベースに保存する．

ステップ 2（2 次元確率表の作成） データベースに保存されている閲覧履歴と購買履歴から商品の最新度と頻度を計算し，提案手法を用いて 2 次元確率表を作成する．作成した 2 次元確率表はデータベースに保存する．

本研究では，数理最適化モデルを利用して 2 次元確率表の推定を行ったが，実用上では数理最適化ソルバーが入手できない場合や，数理最適化の技術に通ずる技術者が現場にいない場合がある．その際には数理最適化ソルバーを利用せず，経験的に得られた 2 次元確率表を利用すればよい．データ量が十分にある場合には経験的に得られた 2 次元確率表を用いても推定した 2 次元確率表と同程度の予測精度となることが本研究で確認されており，利用者の体験には大きな差はない．

5.3.2 商品推薦

商品推薦では，過去に閲覧した商品からの推薦リストと過去に閲覧したことがない商品から推薦リストの作成を行なう．

ステップ 1（閲覧履歴の取得） 利用者が EC サイトを利用して商品閲覧を行なった際に，閲覧した商品の ID とタイムスタンプを閲覧履歴としてデータベースに保存する．

ステップ 2（商品の選好度の算出）　商品推薦の対象となる利用者を選定して，データベースから閲覧履歴を取得する．閲覧履歴から各商品の頻度と最新度を計算し，予めデータベースに保存しておいた 2 次元確率表を参照して商品の選好度（商品選択確率）を紐付ける．各利用者の商品選好リストはデータベースに保存する．

ステップ 3（過去に閲覧した商品からの推薦リストの作成）　データベースから各利用者の商品選好リストを取得して，選好度順で商品推薦リストを作成する．過去に閲覧した商品からの推薦リストはデータベースに保存する．

ステップ 4（過去に閲覧したことがない商品からの推薦リストの作成）　データベースから各利用者の商品選好リストを取得して，提案手法の通りに評価値行列を作成し，協調フィルタリングアルゴリズムを実行する．各利用者と商品との間に推薦スコアが計算できるので，スコア順で商品推薦リストを作成する．このとき，過去に閲覧した商品を除外することに注意する．過去に閲覧したことがない商品からの推薦リストはデータベースに保存する．

ステップ 5（EC サイトにおける商品推薦）　利用者が EC サイトへ再来訪した際に，データベースを参照して過去に閲覧した商品からの推薦リストと過去に閲覧したことがない商品からの推薦リストを取得する．適当なページで推薦リストから商品を選択して商品推薦を配信する．

上記の実装はリアルタイムの配信を前提とせず，ステップ 2，ステップ 3，ステップ 4 をバッチ処理することを想定している．しかし，実装の設計を工夫することでリアルタイムで過去に閲覧した商品を推薦することは可能である．ステップ 1 において閲覧した商品の ID とタイムスタンプをデータベースに保存する方法が一般的であるが，閲覧履歴を利用者のブラウザ（例えば，ローカルストレージ）に保存する手段がある．閲覧履歴を利用者のブラウザに保存しておけば，利用者がサイトに再来訪した際にブラウザに保存されている閲覧履歴をサーバーに提供することで，各商品に選好度を付与し，商品推薦を行なうことが可能である．このとき，提案手法は商品の最新度と頻度の計算，データベースへのアクセス，およびソート処理のみで商品推薦ができるため，非常に計算コストが低い．特にデータベースへのアクセスは最新度と頻度による主キーで参照し，選好度を紐付けることができるため複雑な計算をしてスコアを算出する必要がない点は実用上，非常に有効である．

一方，協調フィルタリングによる商品推薦はコストが高いためリアルタイムで商品推薦する仕組みを構築するためには工夫が必要になるが，他の商品推薦の方法と組合せてリアルタイムで商品推薦を実装することは可能である．例えば，予め類似商品を計算しておき，データベースに保存しておけば，選好度が高い商品と関連付けて，類似商品を推薦することは簡単に実装できる．2 次元確率表を用いた手法は計算コストが低いため，様々な推薦手法への応用がしやすい点にも注目されたい．

提案手法は取得が難しい利用者個人の属性をできるだけ利用せず，取得が容易である閲覧履歴のみを利用している点も注目すべきである．閲覧履歴のみを利用しているため多くの EC サイトで提案手法を利用することができ，実用上の汎用性が高いと言える．また，提案手法は個

人の属性に関する情報を利用していないという点で，セキュリティや情報管理の点でもリスクが低い．

本研究では過去に閲覧した商品からの推薦と過去に閲覧したことがない商品からの推薦の 2 つのタスクに注目したが，学術的な商品推薦の文脈では後者の推薦手法が重要視されることが多い．しかし，前者の推薦手法も実用上は重要である．実際，過去に閲覧した商品からの推薦をサービスに実装することで商品推薦が動線として機能することが期待できる．例えば，利用者が EC サイトのトップページに流入した場合に，商品推薦機能を利用することで過去に閲覧した商品のページにスムーズに移動することができる．EC サイトにおける購買，特に単価が高い商品の購買をする際，利用者は当該サイトに何度も訪問して商品検討を行なう．そのため，過去に閲覧した商品の推薦が適切なページで行われることで利用者の体験の向上が期待できる．

5.4　今後の展望

本節では，今後の展望について述べる．今後の研究課題として，他の分野への応用，形状制約の拡張，形状制約をもつ関数の応用が考えられる．

5.4.1　他の分野への応用

本研究では EC サイトの閲覧履歴やスマホアプリの検索履歴といったノイズを含む大規模データを対象とした．本研究は順序構造を持つ様々な現象に応用することができる．

EC サイトの閲覧履歴から商品の購買確率を推定する手法において，最新度と頻度に基づき半順序構造を定義して 2 次元確率表を作成する方法は，稀にしか起きない事象やデータ取得が高価な場合にも有効である．例えば，医療などで症例数が少ない病気の陽性予測において，検査のコストが高く，少数の特徴量で陽性予測をしなければならない場合が考えられる．検査により取得した数値は陽性と相関関係があるので単調性の制約を自然に入れることができる．そのため，医療の分野などで少数のデータから陽性予測をしなければならないタスクに応用できる可能性がある．

また，スマホアプリの検索履歴から時系列確率分布を推定する手法においては，一次元の単峰性や二峰性ではなく，二次元の単峰性や二峰性に拡張すれば地理的な事例への応用も考えられる．すなわち，ある二次元の空間で何かしらのイベントが複数発生しており，イベントの発生確率がその頻度と距離に相関関係がある場合である．例えば，犯罪の発生はイベントが発生した地点から近い地点で発生する確率が高いと言われており，犯罪の発生予測に応用できる可能性がある．

本研究における最新度と頻度による 2 次元確率表は 1.2.3 項で述べたように認知心理学の分野における忘却曲線，および単純接触効果と強い関連がある．認知心理学では忘却曲線と単純接触効果をそれぞれ単独で研究することが多いが，忘却曲線と単純接触効果の交互作用を考えた研究は筆者が知り得る限りない．また，心理学の分野では特定の現象に対する関数形を見つ

ける研究は多いが，本研究のように数理最適化モデルを用いてノンパラメトリックに推定する方法はない．そのため心理学への分野への応用も期待できる．

5.4.2　形状制約の表現可能性

本研究では数理最適化モデルを用いて形状を制約する推定手法を提案した．特に二項関係である順序構造を用いた．EC サイトの閲覧履歴から商品の購買確率を推定する手法において，最新度と頻度を用いて半順序構造を定義しているが，扱う現象によってより複雑な順序を定義できる可能性がある．複雑な順序構造を定義する場合，理論的な解析も必要になるであろう．

また，スマホアプリの検索履歴から時系列確率分布を推定する手法において，単調単峰性モデルと単調二峰性モデルは極大値や極小値を自動で推定した．2 次元確率表の場合と異なり，単調性などの順序構造だけでなく極大値や極小値の存在を形状制約として課していることに注意されたい．2 次元確率表における頻度に関する単調性を論理式で表現すると

$$\forall r \in R\ \forall f, f' \in F\ (f \leq f' \rightarrow x_{r,f} \leq x_{r,f'})$$

のように全称量化子 $(\forall)$ のみを用いて表されるが，単調単峰性モデルを論理式で表現すると

$$\exists s \in T\ \forall t, t' \in T\ ((t \leq t' \leq s \rightarrow x_t \leq x_{t'})\ \&\ (s \leq t \leq t' \rightarrow x_t \geq x_{t'}))$$

のように存在量化子 $(\exists)$ が必要になる．すなわち，本研究は順序構造に代表される二項関係だけでなく，存在量化子を用いて表すような構造へ部分的に拡張できることを示唆している．数理論理学の分野では全称量化子のみで表現できる論理式と存在量化子が含まれる論理式とで表現力が大きく異ることに注意されたい．これは，数理最適化問題において 0-1 整数変数を用いることで一部の論理的な制約を表現できることと関係しており，従来の形状制約回帰の問題クラスよりも真に広い問題クラスを汎用的に扱うことができる．すなわち，本研究の延長として数理最適化問題の表現可能性について議論することができる．

5.4.3　形状制約をもつ関数を用いた解析

本研究では主に数理最適化モデルを利用して形状制約を課した推定手法について予測性能の観点で評価をした．しかし，推定した後のモデルの応用についても様々な議論ができる．例えば，2 次元確率表は $R \times F$ から $[0, 1]$ への関数 x で次の性質を持つ．

- $\forall r \in R\ \forall f, f' \in F\ (f \leq f' \rightarrow x_{r,f} \leq x_{r,f'})$
- $\forall f \in F\ \forall r, r' \in R\ (r \leq r' \rightarrow x_{r,f} \leq x_{r',f})$

一方，単調単峰性モデルは T から $[0, 1]$ への関数 x で次の性質を持つ．

- $\exists s \in T\ \forall t, t' \in T\ ((t \leq t' \leq s \rightarrow x_t \leq x_{t'})\ \&\ (s \leq t \leq t' \rightarrow x_t \geq x_{t'}))$
- $\sum_{t \in T} x_t = 1$

上記のように推定した関数 x を用いて解析を行う場合，形状の制約が課されているため，その

後の解析がしやすい．例えば，単調単峰性モデルを利用して推定した関数 x に確率順序を定義する場合，関数 x が形状に関する性質を持つことで解析がしやすくなることが期待できる．これは経験分布ではノイズの影響で確率順序が定義しにくいが，単調単峰性モデルによって得られた分布はノイズが減少するため確率順序が定義しやすくなるためである．このように形状制約を課して得られた出力は様々な応用が考えられる．

謝辞

本研究を進めることができたのは多くの先生方のご指導と同僚の皆様からのご支援のおかげです．

筑波大学大学院理工情報生命学術院システム情報工学研究群社会工学学位プログラムにて本論文を書き上げるにあたり，高野祐一准教授には指導教員として，吉瀬章子教授，馬場雪乃准教授，八森正泰准教授，金澤輝代士助教には副指導教員としてご指導いただきました．私の気付かぬ視点からのご助言のおかげで多くの示唆を得ることができ，博士論文として形にすることができました．特に金澤先生からは，第 4 章の研究で人工データを用いた数値実験をするご提案をいただきました．追加実験を通して研究アプローチの幅を広げることができただけでなく，自身の研究内容をより深く理解することができました．深く感謝申し上げます．

本研究は，株式会社リクルートライフスタイルの西村直樹博士，東京理科大学の鮭川矩義助教，そして指導教員でもある高野祐一准教授と長年続けた共同研究であり，経営科学系研究部会連合協議会主催のデータ解析コンペティションにおける議論からこの研究がはじまりました．その際に，高野准教授（当時，東京工業大学助教）からのお声がけがなければ本研究が進むことはありませんでしたし，鮭川助教と西村博士のご支援がなければ本研究を深めることはできませんでした．鮭川助教と西村博士に感謝の意を，また高野准教授には改めて感謝申し上げます．

経営科学系研究部会連合協議会にも感謝致します．経営科学系研究部会連合協議が主催するデータ解析コンペティションにて提供されたデータがなければ本研究は生まれませんでした．本コンペティションの運営をされている中央大学の生田目崇教授，慶應義塾大学の鈴木秀男教授にも大変お世話になりました．当時株式会社リクルート住まいカンパニーに所属しておりました吉永恵一氏，および野村晋平氏にも様々なご助言をいただきました．この場を借りて感謝申し上げます．また，コネヒト株式会社様からもデータを提供していただきました．国内では非常に貴重なデータである出産日付きの時系列検索データを提供いただいたおかげで新たな研究に着手することができました．代表取締役北吉竜也様をはじめとし，CTO 伊藤翔様，執行役員高橋恭文様，そして小椋友季様には多大なるご厚意をいただきました．深く感謝申し上げます．

本研究を進めることができたのは同僚の支援のおかげであることは言うまでもありません．研究を始めた頃に所属していた株式会社 NTT データ数理システムの同僚の皆様にも感謝を申し上げます．論文執筆を渋っていた私に「論文が書けるなら時間をかけても書いほうがい

い.」という一言と皆様の後押しがなければ，本研究を始めることはありませんでした．その後，Retty 株式会社では，様々な産学連携プロジェクトを手掛けつつ，研究活動を進めさせていただきました．感謝申し上げます．

最後に，仕事を辞めて博士課程進学をすることに快諾してくれただけでなく，本論文の校正までしてくれた妻めぐみに心から感謝します．

参考文献

[1] Gediminas Adomavicius and Alexander Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 17, No. 6, pp. 734–749, 2005.

[2] Charu C. Aggarwal. *Recommender Systems*. Springer, 2016.

[3] Yacine Aıt-Sahalia and Jefferson Duarte. Nonparametric option pricing under shape restrictions. *Journal of Econometrics*, Vol. 116, No. 1-2, pp. 9–47, 2003.

[4] M. Ali Akcayol, Anıl Utku, Ebru Aydoğan, and Begüm Mutlu. A weighted multi-attribute-based recommender system using extended user behavior analysis. *Electronic Commerce Research and Applications*, Vol. 28, pp. 86–93, 2018.

[5] Xavier Amatriain, Josep M. Pujol, and Nuria Oliver. I like it... i like it not: Evaluating user ratings noise in recommender systems. In *International Conference on User Modeling, Adaptation, and Personalization*, pp. 247–258. Springer, 2009.

[6] Xavier Amatriain, Josep M. Pujol, Nava Tintarev, and Nuria Oliver. Rate it again: Increasing recommendation accuracy by user re-rating. In *Proceedings of the 2009 ACM Conference on Recommender Systems*, pp. 173–180. ACM, 2009.

[7] Suhail Ansari, Ron Kohavi, Llew Mason, and Zijian Zheng. Integrating e-commerce and data mining: Architecture and challenges. In *Proceedings of the 2001 IEEE International Conference on Data Mining*, pp. 27–34. IEEE, 2001.

[8] Yilmaz Ar and Erkan Bostanci. A genetic algorithm solution to the collaborative filtering problem. *Expert Systems with Applications*, Vol. 61, pp. 122–128, 2016.

[9] R.E. Barlow, D.J. Bartholomew, J.M. Bremner, and H.D. Brunk. *Statistical Inference Under Order Restrictions: The Theory and Application of Isotonic Regression*. Out-of-print Books on demand. J. Wiley, 1972.

[10] Michael J. Best and Nilotpal Chakravarti. Active set algorithms for isotonic regression; a unifying framework. *Mathematical Programming*, Vol. 47, No. 1-3, pp. 425–439, 1990.

[11] Jesús Bobadilla, Antonio Hernando, Fernando Ortega, and Abraham Gutiérrez. Collaborative filtering based on significances. *Information Sciences*, Vol. 185, No. 1,

pp. 1–17, 2012.

[12] Jesús Bobadilla, Fernando Ortega, Antonio Hernando, and Abraham Gutiérrez. Recommender systems survey. *Knowledge-based systems*, Vol. 46, pp. 109–132, 2013.

[13] Emad Gohari Boroujerdi, Soroush Mehri, Saeed Sadeghi Garmaroudi, Mohammad Pezeshki, Farid Rashidi Mehrabadi, SeyyedSalim Malakouti, and Shahram Khadivi. A study on prediction of user's tendency toward purchases in websites based on behavior models. In *2014 6th Conference on Information and Knowledge Technology (IKT)*, pp. 61–66. IEEE, 2014.

[14] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[15] Gordon Bril, Richard Dykstra, Carolyn Pillers, and Tim Robertson. Algorithm as 206: isotonic regression in two independent variables. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 33, No. 3, pp. 352–357, 1984.

[16] Randolph E. Bucklin and Catarina Sismeiro. Click here for internet insight: Advances in clickstream data analysis in marketing. *Journal of Interactive marketing*, Vol. 23, No. 1, pp. 35–48, 2009.

[17] Bo Cai and David B. Dunson. Bayesian multivariate isotonic regression splines: Applications to carcinogenicity studies. *Journal of the American Statistical Association*, Vol. 102, No. 480, pp. 1158–1171, 2007.

[18] Pedro G. Campos, Alejandro Bellogín, Fernando Díez, and J. Enrique Chavarriaga. Simple time-biased knn-based recommendations. In *Proceedings of the Workshop on Context-Aware Movie Recommendation*, pp. 20–23. ACM, 2010.

[19] Pedro G. Campos, Fernando Díez, and Iván Cantador. Time-aware recommender systems: a comprehensive survey and analysis of existing evaluation protocols. *User Modeling and User-Adapted Interaction*, Vol. 24, No. 1-2, pp. 67–119, 2014.

[20] Gang Chen, Fei Wang, and Changshui Zhang. Collaborative filtering using orthogonal nonnegative matrix tri-factorization. *Information Processing & Management*, Vol. 45, No. 3, pp. 368–379, 2009.

[21] Zhen-Yu Chen and Zhi-Ping Fan. Distributed customer behavior prediction using multiplex data: a collaborative mk-svm approach. *Knowledge-Based Systems*, Vol. 35, pp. 111–119, 2012.

[22] Yoon Ho Cho and Jae Kyeong Kim. Application of web usage mining and product taxonomy to collaborative recommendations in e-commerce. *Expert Systems with Applications*, Vol. 26, No. 2, pp. 233–246, 2004.

[23] James E. Crandall, Victor E. Montgomery, and Willis W. Rees. "mere" exposure versus familiarity, with implications for response competition and expectancy arousal hypotheses. *The Journal of general psychology*, Vol. 88, No. 1, pp. 105–120, 1973.

[24] Yi Ding and Xue Li. Time weight collaborative filtering. In *Proceedings of the 14th ACM international conference on Information and knowledge management*, pp. 485–492. ACM, 2005.

[25] Yi Ding, Xue Li, and Maria E. Orlowska. Recency-based collaborative filtering. In *Database Technologies 2006, Proceedings of the 17th Australasian Database Conference, ADC 2006, Hobart, Tasmania, Australia, January 16-19 2006*, pp. 99–107, 2006.

[26] Richard L. Dykstra. An algorithm for restricted least squares regression. *Journal of the American Statistical Association*, Vol. 78, No. 384, pp. 837–842, 1983.

[27] Richard L. Dykstra and Tim Robertson. An algorithm for isotonic regression for two or more independent variables. *The Annals of Statistics*, pp. 708–716, 1982.

[28] Hermann Ebbinghaus. Memory: A contribution to experimental psychology. *Annals of neurosciences*, Vol. 20, No. 4, p. 155, 2013.

[29] Michael D. Ekstrand, John T. Riedl, and Joseph A. Konstan. *Collaborative filtering recommender systems*. Now Publishers Inc, 2011.

[30] Peter S. Fader, Bruce GS Hardie, and Ka Lok Lee. Rfm and clv: Using iso-value curves for customer base analysis. *Journal of marketing research*, Vol. 42, No. 4, pp. 415–430, 2005.

[31] Peter S. Fader, Bruce G.S. Hardie, and Ka Lok Lee. Rfm and clv: Using iso-value curves for customer base analysis. *Journal of marketing research*, Vol. 42, No. 4, pp. 415–430, 2005.

[32] Peter S. Fader, Bruce G.S. Hardie, and Ka Lok Lee. "counting your customers" the easy way: An alternative to the pareto/nbd model. *Marketing science*, Vol. 24, No. 2, pp. 275–284, 2005.

[33] M. Frisén. Unimodal regression. *Journal of the Royal Statistical Society: Series D (The Statistician)*, Vol. 35, No. 4, pp. 479–485, 1986.

[34] Sandra Clara Gadanho and Nicolas Lhuillier. Addressing uncertainty in implicit preferences. In *Proceedings of the 2007 ACM Conference on Recommender Systems*, pp. 97–104. ACM, 2007.

[35] Zhi Geng and Ning-Zhong Shi. Algorithm as 257: isotonic regression for umbrella orderings. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 39, No. 3, pp. 397–402, 1990.

[36] Zhi Geng and Ning-Zhong Shi. Isotonic regression in two independent variables under umbrella orderings. *Journal of the Japanese Society of Computational Statistics*, Vol. 4, No. 1, pp. 49–61, 1991.

[37] Charles J. Geyer. Constrained maximum likelihood exemplified by isotonic convex logistic regression. *Journal of the American Statistical Association*, Vol. 86, No. 415, pp. 717–724, 1991.

[38] Sergiu Gordea and Markus Zanker. Time filtering for better recommendations with small and sparse rating matrices. In *International Conference on Web Information Systems Engineering*, pp. 171–183. Springer, 2007.

[39] Quanquan Gu, Jie Zhou, and Chris Ding. Collaborative filtering: Weighted nonnegative matrix factorization incorporating user and item graphs. In *Proceedings of the 2010 SIAM International Conference on Data Mining*, pp. 199–210. SIAM, 2010.

[40] Adityanand Guntuboyina, Bodhisattva Sen, et al. Nonparametric shape-restricted regression. *Statistical Science*, Vol. 33, No. 4, pp. 568–594, 2018.

[41] Zahra Yusefi Hafshejani, Marjan Kaedi, and Afsaneh Fatemi. Improving sparsity and new user problems in collaborative filtering by clustering the personality factors. *Electronic Commerce Research*, Vol. 18, No. 4, pp. 813–836, 2018.

[42] Hui Han, Hongyi Xu, and Hongquan Chen. Social commerce: A systematic review and data synthesis. *Electronic Commerce Research and Applications*, Vol. 30, pp. 38–50, 2018.

[43] Qiyang Han, Tengyao Wang, Sabyasachi Chatterjee, Richard J. Samworth, et al. Isotonic regression in general dimensions. *The Annals of Statistics*, Vol. 47, No. 5, pp. 2440–2471, 2019.

[44] Antonio Hernando, Jesús Bobadilla, and Fernando Ortega. A non negative matrix factorization for collaborative filtering recommender systems based on a bayesian probabilistic model. *Knowledge-Based Systems*, Vol. 97, pp. 188–202, 2016.

[45] Clifford Hildreth. Point estimates of ordinates of concave functions. *Journal of the American Statistical Association*, Vol. 49, No. 267, pp. 598–619, 1954.

[46] Wenxing Hong, Lei Li, and Tao Li. Product recommendation with temporal dynamics. *Expert Systems with Applications*, Vol. 39, No. 16, pp. 12398–12406, 2012.

[47] Tingliang Huang and Jan A. Van Mieghem. Clickstream data and inventory management: Model and empirical analysis. *Production and Operations Management*, Vol. 23, No. 3, pp. 333–347, 2014.

[48] Arthur Middleton Hughes. *Strategic database marketing: the masterplan for starting and managing a profitable, customer-based marketing program*, Vol. 12. McGraw-Hill New York, NY, 2000.

[49] Sam K. Hui, Peter S. Fader, and Eric T. Bradlow. Path data in marketing: An integrative framework and prospectus for model building. *Marketing Science*, Vol. 28, No. 2, pp. 320–335, 2009.

[50] Wook-Yeon Hwang. Assessing new correlation-based collaborative filtering approaches for binary market basket data. *Electronic Commerce Research and Applications*, Vol. 29, pp. 12–18, 2018.

[51] Hyunwoo Hwangbo, Yang Sok Kim, and Kyung Jin Cha. Recommendation system development for fashion retail e-commerce. *Electronic Commerce Research and*

Applications, Vol. 28, pp. 94–101, 2018.

[52] Jiro Iwanaga, Naoki Nishimura, Noriyoshi Sukegawa, and Yuichi Takano. Estimating product-choice probabilities from recency and frequency of page views. *Knowledge-Based Systems*, Vol. 99, pp. 157–167, 2016.

[53] Jiro Iwanaga, Naoki Nishimura, Noriyoshi Sukegawa, and Yuichi Takano. Improving collaborative filtering recommendations by estimating user preferences from clickstream data. *Electronic Commerce Research and Applications*, Vol. 37, p. 100877, 2019.

[54] M. Jamalzadeh. *Analysis of clickstream data*. PhD thesis, Durham University, 2011.

[55] Gawesh Jawaheer, Peter Weller, and Patty Kostkova. Modeling user preferences in recommender systems: A classification framework for explicit and implicit user feedback. *ACM Transactions on Interactive Intelligent Systems*, Vol. 4, No. 2, pp. 8:1–8:26, 2014.

[56] Kinshuk Jerath, Peter S. Fader, and Bruce G.S. Hardie. New perspectives on customer "death" using a generalization of the pareto/nbd model. *Marketing Science*, Vol. 30, No. 5, pp. 866–880, 2011.

[57] Ke Ji, Runyuan Sun, Xiang Li, and Wenhao Shu. Improving matrix approximation for recommendation via a clustering-based reconstructive method. *Neurocomputing*, Vol. 173, pp. 912–920, 2016.

[58] Alexandros Karatzoglou. Collaborative temporal order modeling. In *Proceedings of the fifth ACM conference on Recommender systems*, pp. 313–316. ACM, 2011.

[59] Naime Ranjbar Kermany and Sasan H. Alizadeh. A hybrid multi-criteria recommender system using ontology and neuro-fuzzy techniques. *Electronic Commerce Research and Applications*, Vol. 21, pp. 50–64, 2017.

[60] Yong Soo Kim and Bong-Jin Yum. Recommender system based on click stream data using association rule mining. *Expert Systems with Applications*, Vol. 38, No. 10, pp. 13320–13327, 2011.

[61] Hamidreza Koohi and Kourosh Kiani. A new method to find neighbor users that improves the performance of collaborative filtering. *Expert Systems with Applications*, Vol. 83, pp. 30–39, 2017.

[62] Yehuda Koren. Collaborative filtering with temporal dynamics. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, June 28 - July 1, 2009*, pp. 447–456, 2009.

[63] Yehuda Koren. Factor in the neighbors: Scalable and accurate collaborative filtering. *ACM Transactions on Knowledge Discovery from Data*, Vol. 4, No. 1, pp. 1:1–1:24, 2010.

[64] William R. Kunst-Wilson and Robert B. Zajonc. Affective discrimination of stimuli that cannot be recognized. *Science*, Vol. 207, No. 4430, pp. 557–558, 1980.

[65] YoungOk Kwon. Improving top-n recommendation techniques using rating variance. In *Proceedings of the 2008 ACM conference on Recommender systems*, pp. 307–310. ACM, 2008.

[66] Santiago Larrain, Christoph Trattner, Denis Parra, Eduardo Graells-Garrido, and Kjetil Nørvåg. Good times bad times: A study on recency effects in collaborative filtering for social tagging. In *Proceedings of the 9th ACM Conference on Recommender Systems, RecSys 2015, Vienna, Austria, September 16-20, 2015*, pp. 269–272, 2015.

[67] R. Latha and R. Nadarajan. Ranking based approach for noise handling in recommender systems. In *International Conference on Multimedia Communications, Services and Security*, pp. 46–58. Springer, 2015.

[68] Daniel D. Lee and H. Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, Vol. 401, No. 6755, p. 788, 1999.

[69] Daniel D. Lee and H. Sebastian Seung. Algorithms for non-negative matrix factorization. In *Advances in neural information processing systems*, pp. 556–562, 2001.

[70] Tong Queue Lee, Young Park, and Yong-Tae Park. A time-based approach to effective recommender systems using implicit feedback. *Expert systems with applications*, Vol. 34, No. 4, pp. 3055–3062, 2008.

[71] Bin Li, Ling Chen, Xingquan Zhu, and Chengqi Zhang. Noisy but non-malicious user detection in social recommender systems. *World Wide Web*, Vol. 16, No. 5-6, pp. 677–699, 2013.

[72] Seth Siyuan Li and Elena Karahanna. Online recommendation systems in a b2c e-commerce context: A review and future directions. *Journal of the Association for Information Systems*, Vol. 16, No. 2, pp. 72–107, 2015.

[73] Xin Li, Guandong Xu, Enhong Chen, and Yu Zong. Learning recency based comparative choice towards point-of-interest recommendation. *Expert Systems with Applications*, Vol. 42, No. 9, pp. 4274–4283, 2015.

[74] Chih-Lun Liao and Shie-Jue Lee. A clustering based approach to improving the efficiency of collaborative filtering recommendation. *Electronic Commerce Research and Applications*, Vol. 18, pp. 1–9, 2016.

[75] Greg Linden, Brent Smith, and Jeremy York. Amazon. com recommendations: Item-to-item collaborative filtering. *IEEE Internet computing*, Vol. 7, No. 1, pp. 76–80, 2003.

[76] Guimei Liu, Tam T Nguyen, Gang Zhao, Wei Zha, Jianbo Yang, Jianneng Cao, Min Wu, Peilin Zhao, and Wei Chen. Repeat buyer prediction for e-commerce. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 155–164. ACM, 2016.

[77] Mengsi Liu, Weike Pan, Miao Liu, Yaofeng Chen, Xiaogang Peng, and Zhong Ming. Mixed similarity learning for recommendation with implicit feedback. *Knowledge-*

Based Systems, Vol. 119, pp. 178–185, 2017.

[78] Nathan N. Liu, Evan W. Xiang, Min Zhao, and Qiang Yang. Unifying explicit and implicit feedback for collaborative filtering. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, pp. 1445–1448. ACM, 2010.

[79] Xin Luo, Mengchu Zhou, Yunni Xia, and Qingsheng Zhu. An efficient non-negative matrix-factorization-based approach to collaborative filtering for recommender systems. *IEEE Transactions on Industrial Informatics*, Vol. 10, No. 2, pp. 1273–1284, 2014.

[80] Ronny Luss, Saharon Rosset, Moni Shahar, et al. Efficient regularized isotonic regression with application to gene–gene interaction search. *The Annals of Applied Statistics*, Vol. 6, No. 1, pp. 253–283, 2012.

[81] Xiao Ma, Hongwei Lu, Zaobin Gan, and Jiangfeng Zeng. An explicit trust and distrust clustering based collaborative filtering recommendation approach. *Electronic Commerce Research and Applications*, Vol. 25, pp. 29–39, 2017.

[82] William L. Maxwell and John A. Muckstadt. Establishing consistent and realistic reorder intervals in production-distribution systems. *Operations Research*, Vol. 33, No. 6, pp. 1316–1341, 1985.

[83] Bamshad Mobasher, Honghua Dai, Tao Luo, and Miki Nakagawa. Improving the effectiveness of collaborative filtering on anonymous web usage data. In *Proceedings of the IJCAI 2001 Workshop on Intelligent Techniques for Web Personalization*, pp. 53–61, 2001.

[84] Wendy W. Moe and Peter S. Fader. Dynamic conversion behavior at e-commerce sites. *Management Science*, Vol. 50, No. 3, pp. 326–335, 2004.

[85] Alan L. Montgomery. Applying quantitative marketing techniques to the internet. *Interfaces*, Vol. 31, No. 2, pp. 90–108, 2001.

[86] Alan L. Montgomery, Shibo Li, Kannan Srinivasan, and John C. Liechty. Modeling online browsing and path analysis using clickstream data. *Marketing science*, Vol. 23, No. 4, pp. 579–595, 2004.

[87] Elizbar A. Nadaraya. On estimating regression. *Theory of Probability & Its Applications*, Vol. 9, No. 1, pp. 141–142, 1964.

[88] Maryam Khanian Najafabadi, Mohd Naz'ri Mahrin, Suriayati Chuprat, and Haslina Md Sarkan. Improving the accuracy of collaborative filtering recommendations using clustering and association rules mining on implicit data. *Computers in Human Behavior*, Vol. 67, pp. 113–128, 2017.

[89] Eric W.T. Ngai, Li Xiu, and Dorothy C.K. Chau. Application of data mining techniques in customer relationship management: A literature review and classification. *Expert systems with applications*, Vol. 36, No. 2, pp. 2592–2602, 2009.

[90] Van-Doan Nguyen, Songsak Sriboonchitta, and Van-Nam Huynh. Using community preference for overcoming sparsity and cold-start problems in collaborative filtering system offering soft ratings. *Electronic Commerce Research and Applications*, Vol. 26, pp. 101–108, 2017.

[91] Mehrbakhsh Nilashi, Othman Ibrahim, and Karamollah Bagherifard. A recommender system based on collaborative filtering using ontology and dimensionality reduction techniques. *Expert Systems with Applications*, Vol. 92, pp. 507–520, 2018.

[92] Mehrbakhsh Nilashi, Dietmar Jannach, Othman bin Ibrahim, Mohammad Dalvi Esfahani, and Hossein Ahmadi. Recommendation quality, transparency, and website quality for trust-building in recommendation agents. *Electronic Commerce Research and Applications*, Vol. 19, pp. 70–84, 2016.

[93] Mehrbakhsh Nilashi, Dietmar Jannach, Othman bin Ibrahim, and Norafida Ithnin. Clustering-and regression-based multi-criteria collaborative filtering with incremental updates. *Information Sciences*, Vol. 293, pp. 235–250, 2015.

[94] Naoki Nishimura, Noriyoshi Sukegawa, Yuichi Takano, and Jiro Iwanaga. A latent-class model for estimating product-choice probabilities from clickstream data. *Information Sciences*, Vol. 429, pp. 406–420, 2018.

[95] Rainer Olbrich and Christian Holsing. Modeling consumer purchasing behavior in social shopping communities with clickstream data. *International Journal of Electronic Commerce*, Vol. 16, No. 2, pp. 15–40, 2011.

[96] Michael P. O'Mahony, Neil J. Hurley, and Guénolé Silvestre. Detecting noise in recommender system databases. In *Proceedings of the 11th International Conference on Intelligent User Interfaces*, pp. 109–115. ACM, 2006.

[97] Denis Parra, Alexandros Karatzoglou, Xavier Amatriain, and Idil Yavuz. Implicit feedback recommendation via implicit-to-explicit ordinal logistic regression mapping. *Proceedings of the CARS-2011*, 2011.

[98] Hau Xuan Pham and Jason J. Jung. Preference-based user rating correction process for interactive recommendation systems. *Multimedia Tools and Applications*, Vol. 65, No. 1, pp. 119–132, 2013.

[99] Nikolaos Polatidis and Christos K. Georgiadis. A multi-level collaborative filtering method that improves recommendations. *Expert Systems with Applications*, Vol. 48, pp. 100–110, 2016.

[100] Jiangtao Qiu. A predictive model for customer purchase behavior in e-commerce context. In *PACIS*, p. 369, 2014.

[101] Werner J. Reinartz and Vijay Kumar. On the profitability of long-life customers in a noncontractual setting: An empirical investigation and implications for marketing. *Journal of marketing*, Vol. 64, No. 4, pp. 17–35, 2000.

[102] Werner J. Reinartz and Vita Kumar. The impact of customer relationship charac-

teristics on profitable lifetime duration. *Journal of marketing*, Vol. 67, No. 1, pp. 77–99, 2003.

[103] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. Grouplens: an open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, pp. 175–186, 1994.

[104] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. Grouplens: an open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, pp. 175–186. ACM, 1994.

[105] Francesco Ricci, Lior Rokach, and Bracha Shapira. Introduction to recommender systems handbook. In *Recommender systems handbook*, pp. 1–35. Springer, 2011.

[106] Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor. *Recommender Systems Handbook*. Springer, 2015.

[107] T. Robertson, F. Wright, and R. Dykstra. Order restricted statistical inference, chichester: John wiley and sons. *RobertsonOrder Restricted Statistical Inference1988*, 1988.

[108] Robin Roundy. A 98%-effective lot-sizing rule for a multi-product, multi-stage production/inventory system. *Mathematics of operations research*, Vol. 11, No. 4, pp. 699–727, 1986.

[109] Aghiles Salah, Nicoleta Rogovschi, and Mohamed Nadif. A dynamic collaborative filtering system via a weighted clustering approach. *Neurocomputing*, Vol. 175, pp. 206–215, 2016.

[110] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*, pp. 285–295. ACM, 2001.

[111] Shota Sato and Yumi Asahi. The model of purchasing and visiting behavior of customers in an e-commerce site for consumers. *International Proceedings of Economics Development & Research*, Vol. 52, , 2012.

[112] David C. Schmittlein and Robert A. Peterson. Customer base analysis: An industrial purchase process application. *Marketing Science*, Vol. 13, No. 1, pp. 41–67, 1994.

[113] Hinrich Schütze, Christopher D. Manning, and Prabhakar Raghavan. *Introduction to information retrieval*, Vol. 39. Cambridge University Press Cambridge, 2008.

[114] Catarina Sismeiro and Randolph E. Bucklin. Modeling purchase behavior at an e-commerce web site: A task-completion approach. *Journal of marketing research*, Vol. 41, No. 3, pp. 306–323, 2004.

[115] J. Spouge, H. Wan, and WJ Wilbur. Least squares isotonic regression in two dimensions. *Journal of Optimization Theory and Applications*, Vol. 117, No. 3, pp.

585–605, 2003.

[116] David J. Stang, Edward J.O' Connell. The computer as experimenter in social psychological research. *Behavior Research Methods & Instrumentation*, Vol. 6, No. 2, pp. 223–231, 1974.

[117] Quentin F. Stout. Optimal algorithms for unimodal regression. *Computing science and statistics*, Vol. 32, pp. 348–355, 2000.

[118] Quentin F. Stout. Isotonic regression for multiple independent variables. *Algorithmica*, Vol. 71, No. 2, pp. 450–470, 2015.

[119] Qiang Su and Lu Chen. A method for discovering clusters of e-commerce interest patterns using click-stream data. *Electronic Commerce Research and Applications*, Vol. 14, No. 1, pp. 1–13, 2015.

[120] Raciel Yera Toledo, Yailé Caballero Mota, and Luis Martínez. Correcting noisy ratings in collaborative recommender systems. *Knowledge-Based Systems*, Vol. 76, pp. 96–108, 2015.

[121] E. Turban, D. King, J. Lee, T.-P. Liang, and D.C. Turban. *Electronic commerce: A managerial and social networks perspective [8th ed.]*. Springer, 2015.

[122] Dirk Van den Poel and Wouter Buckinx. Predicting online-purchasing behaviour. *European journal of operational research*, Vol. 166, No. 2, pp. 557–575, 2005.

[123] Norha M. Villegas, Cristian Sánchez, Javier Díaz-Cely, and Gabriel Tamura. Characterizing context-aware recommender systems: A systematic literature review. *Knowledge-Based Systems*, Vol. 140, pp. 173–200, 2018.

[124] Geoffrey S. Watson. Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, pp. 359–372, 1964.

[125] Liu Xiaojun. An improved clustering-based collaborative filtering recommendation algorithm. *Cluster Computing*, Vol. 20, No. 2, pp. 1281–1288, 2017.

[126] Yang Xu, Xiaoguang Hong, Zhaohui Peng, Guang Yang, and Philip S. Yu. Temporal recommendation via modeling dynamic interests with inverted-u-curves. In *Database Systems for Advanced Applications - 21st International Conference, DASFAA 2016, Dallas, TX, USA, April 16-19, 2016, Proceedings, Part I*, pp. 313–329, 2016.

[127] Kai Yu, Xiaowei Xu, Jianhua Tao, Martin Ester, and Hans-Peter Kriegel. Instance selection techniques for memory-based collaborative filtering. In *Proceedings of the 2002 SIAM International Conference on Data Mining*, pp. 59–74. SIAM, 2002.

[128] Robert B. Zajonc. Attitudinal effects of mere exposure. *Journal of personality and social psychology*, Vol. 9, No. 2p2, p. 1, 1968.

[129] Robert B. Zajonc, Philip Shaver, Carol Tavris, and David Van Kreveld. Exposure, satiation, and stimulus discriminability. *Journal of Personality and Social Psychology*, Vol. 21, No. 3, p. 270, 1972.

[130] Jia Zhang, Yaojin Lin, Menglei Lin, and Jinghua Liu. An effective collaborative

filtering algorithm based on user preference clustering. *Applied Intelligence*, Vol. 45, No. 2, pp. 230–240, 2016.

[131] Sheng Zhang, Weihong Wang, James Ford, and Fillia Makedon. Learning from incomplete ratings using non-negative matrix factorization. In *Proceedings of the 2006 SIAM International Conference on Data Mining*, pp. 549–553. SIAM, 2006.

[132] Yongzheng Zhang and Marco Pennacchiotti. Predicting purchase behaviors from social media. In *Proceedings of the 22nd international conference on World Wide Web*, pp. 1521–1532, 2013.

[133] Andrew Zimdars, David Maxwell Chickering, and Christopher Meek. Using temporal data for making recommendations. In *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*, pp. 580–588. Morgan Kaufmann Publishers Inc., 2001.

[134] 阿部誠. CRM のデータ分析に理論とモデルを組み込む消費者行動理論にもとづいた RF 分析. *Discussion paper series. CIRJE-J.*

[135] 阿部誠. RFM 指標と顧客生涯価値: 階層ベイズモデルを使った非契約型顧客関係管理における消費者行動の分析. 日本統計学会誌, Vol. 41, No. 1, pp. 51–81, 2011.

[136] 阿部誠. RFM データを用いた顧客生涯価値の算出—既存顧客の維持介入と新規顧客の獲得—. マーケティングジャーナル, Vol. 34, No. 1, pp. 73–90, 2014.

[137] 加藤直樹, 羽室行信, 矢田勝俊. データマイニングとその応用. 朝倉書店, 2008.

[138] 岩永二郎, 鍋谷昴一, 梶原悠, 五十嵐健太. 関心度と忘却度に基づくレコメンド手法 : 単調性制約付きレコメンドモデルの構築 (特集 データ解析コンペティション : インフォミディアリ・データの分析). オペレーションズ・リサーチ, Vol. 59, No. 2, pp. 72–80, feb 2014.

[139] 宮本聡介, 太田信夫. 単純接触効果研究の最前線. 北大路書房, 2008.

[140] 全眞嬉, 加藤直樹, 徳山豪. 高次元ピラミッド構築問題とデータマイニングへの応用. 情報処理学会研究報告アルゴリズム (AL), Vol. 2003, No. 3 (2002-AL-088), pp. 71–78, 2003.

[141] 竹澤邦夫. R によるノンパラメトリック会期の入門講義. メタ・ブレーン, 2009.

論文リスト

査読付き論文（筆頭著者）

（１） Jiro Iwanaga, Naoki Nishimura, Noriyoshi Sukegawa, Yuichi Takano, “Improving collaborative filtering recommendations by estimating user preferences from clickstream data”, Electronic Commerce Research and Applications, Vol.37, 2019, 100877

（２） Jiro Iwanaga, Naoki Nishimura, Noriyoshi Sukegawa, Yuichi Takano, “Estimating product-choice probabilities from recency and frequency of page views”, Knowledge-Based Systems, Vol.99, 2016, pp.157-167

（３） 岩永二郎, 鍋谷昴一, 梶原悠, 五十嵐健太,”関心度と忘却度に基づくレコメンド手法: 単調性制約付きレコメンドモデルの構築”, オペレーションズ・リサーチ, Vol.59, No.2, 2014, pp.72-80

その他、査読付き論文

（１） Teppei Sakamoto, Haruka Yamashita, Masayuki Goto, Jiro Iwanaga, “Model for Relational Analysis of Posted Articles and Reactions on Restaurant Guide Sites”, Industrial Engineering & Management Systems, Vol.19, No.3, 2020, pp.669-679

（２） 劉佩潔, 山下遥, 岩永二郎, 樽石将人, 後藤正幸, “グルメサービスにおけるレストラン推薦投稿へのリアクション数増加を目的とした潜在クラスモデル分析”, 情報処理学会論文誌, Vol.59, No.1, 2018, pp.211-226

（３） Naoki Nishimura, Noriyoshi Sukegawa, Yuichi Takano, Jiro Iwanaga, “A latent-class model for estimating product-choice probabilities from clickstream data”, Information Sciences, Vol.429, 2018, pp.406-420

査読なし発表論文

（１） 西村直樹, 鮭川矩義, 高野祐一, 岩永二郎, “形状制約モデルによる顧客の商品選択行動の予測”, オペレーションズ・リサーチ, Vol.65, No.6, 2020, pp.328-333

（２） 竹野峻輔, 氏原淳志, 岩永二郎, “優先度学習による推薦文からの見出し抽出”, オペレーションズ・リサーチ, Vol.62, No.11, 2017, pp.731-736

（３） 西村直樹, 鮭川矩義, 高野祐一, 岩永二郎, 水野眞治, “EC サイトの商品特性を考慮した 2 次元確率表による購買予測”, オペレーションズ・リサーチ, Vol.60, No.2, 2015, pp.69-74