

2020年度 卒業論文

ファッション通販サイトにおける
マルチモーダル
レコメンドシステムの提案

同志社大学文化情報学部文化情報学科

18160237

田中 真士

主担当教員 岩本 真裕子 准教授

副担当教員 伊藤 紀子 准教授

2020年12月

要 旨

近年では、オンラインの普及により、電子商取引市場は年々拡大している。物販だけでも 2019 年には 10 兆円規模の市場となっている。これは、ユーザがパソコンだけでなく、スマートフォンのように気軽にインターネットにアクセスできるようになった要因が背景にある。電子商取引 (以下、E コマースと呼ぶ) は実際の在庫を開示するものではなく、オンライン上にアップロードするだけで商品を販売することができる。アップロードされたアイテムは、常にオンライン上のショッピングサイトに残るため、ユーザにとって大量の商品から求めているアイテムを探し出すのは非常に困難である。E コマースが拡大していく現代において大量の商品からユーザにどのアイテムをレコメンドすれば購入に繋がるかはマーケティングにおける重要な課題であり、アイテムの適切なレコメンドシステムの開発は大変需要が高い。さらにファッションにおいては、体型や文化的背景、時代など考慮すべき様々な要因があり、特に新商品などについてはユーザにとって適切なアイテムをレコメンドすることが困難となっている。新商品はユーザのクリックした回数、いいね数、閲覧数などの嗜好が反映されていないアイテムであり、このような事前に準備のない状況をコールドスタートと言い、従来のレコメンドシステムの課題は、このコールドスタートを考慮した新しいレコメンドシステムを開発しなければいけない点が挙げられる。

そこで本研究ではユーザがアイテムを好む潜在的要因と視覚、言語、季節的時間軸などの様々な観点から細部の要因を含むマルチモーダルレコメンドシステム、MIBPR (Multimodal Interface Bayesian Personalized Ranking) を提案する。本システムは、自然言語処理と画像処理を用いてアイテムデータを数値化、季節に分類することを通してランキング学習を用いてユーザに最適なアイテムをレコメンドする。コールドスタートデータを用いて、新たに提案したシステムと従来のシステムを比較した評価実験を行った結果、本システムではユーザに合わせた最適なアイテムを提案する面で有意性があることが示された。

本論文は、以下のような構成となっている。第 2 章でファッションにおけるレコメンドシステム、コーディネートに関する先行研究を紹介し、第 3 章で先行研究を考慮した上で、アイテムの潜在的な詳細、キャッチフレーズ、季節的ダイナミクスを加えた新たな手法を提案する。第 4 章では、実際の E コマースのデータを用いて最適なアイテムのレコメンドに関する有意性について、先行研究との比較評価実験を行う。最後に第 5 章で結論と今後の展望を述べる。

目次

第1章 はじめに	1
第2章 先行研究	3
2.1 従来のレコメンドシステムのアプローチ	3
2.2 コールドスタートに対応した手法に関する研究	4
第3章 提案手法	8
3.1 視覚的要因と季節的要因	8
3.2 アイテムのタイトルとフレーズの自然言語処理	11
3.3 MIBPR の予測モデル	13
第4章 評価実験	15
4.1 データの概要	15
4.2 結果	17
4.3 考察	19
第5章 結論と今後の展望	20
謝辞	21
参考文献	22

第1章 はじめに

テクノロジーの革新が進むにつれて、私達の暮らしはますます便利になっている。私達の生活を大きく変えた革新技術と言えば、例えばテレビや固定電話、インフラ設備などが挙げられるだろう。それらのテクノロジーの進化の中でも、近年最も私達の注目を集めたのは1995年に発売されたWindows95ではないだろうか。

1970年代からコンピュータ自体はあったものの、このWindows95販売以来、多くのユーザがインターネットを利用するようになった。そして1990年代から販売されたコンピュータと同じくインターネットを利用できるスマートフォンが1990年代に販売された。2010年代に入り、スマートフォンは迅速なスピードで普及した。総務省の調査 [1] によれば、2018年には10代(13歳以上)から40代までの世代で、80%から90%程度がスマートフォンを使用しており、インターネット利用率についても90%以上と使用率が高い水準であり、これは2017年の使用率と比較しても、利用者数は上昇傾向にある。

インターネットの普及率が高くなるとともに、オンラインショッピングの利用率もますます上昇傾向にある。株式会社クロス・マーケティングが行った全国47都道府県20代から60代の男女20,772人を対象にしたオンラインショッピングに関する調査結果 [2](図1.1)ではインターネットでの生活用品の購入頻度は2017年と2018年を比べて、増えたと思うと回答した人が31%、今後増えると思うと回答した人が26%となっており、徐々にインターネットでの購入頻度が高まる兆しが見られている。このようなインターネットを介して行う売買を電子商取引(EコマースまたはEC)といい、オンラインショッピングでは、楽天¹やAmazon²などが有名である。つまり、私達の生活は、店舗購入型からEコマース購入型へと移行してきていることが分かる。

さらに昨今の新型コロナウイルス(COVID-19)の影響により、この傾向に拍車がかかっている。実際、日本経済新聞のニュースによれば、COVID-19の影響により、大手ファッションブランド「ZARA」などを展開するアパレル世界最大大手インディテックスが実店舗全体の16%に相当する最大1200店を2021年までに閉店する計画を打ち出しており、22年には売り上げの4分の1をEコマース上での販売とすることを目指すと述べている [3]。このように今後は、Eコマース上での取引がより当たり前の時代が来ると推測できる。

現在Eコマース上での「衣食住」に関する商品取引は多く、ファッションに関する商品では10代から30代を中心に利用者数の増加と共に購入率が高まっている。しかし、Eコマースと実店舗では大きな違いがある。それはEコマース上では、商品を店員が推薦・提

¹楽天 (<https://www.rakuten.co.jp/>)

²Amazon (<https://www.amazon.co.jp/>)

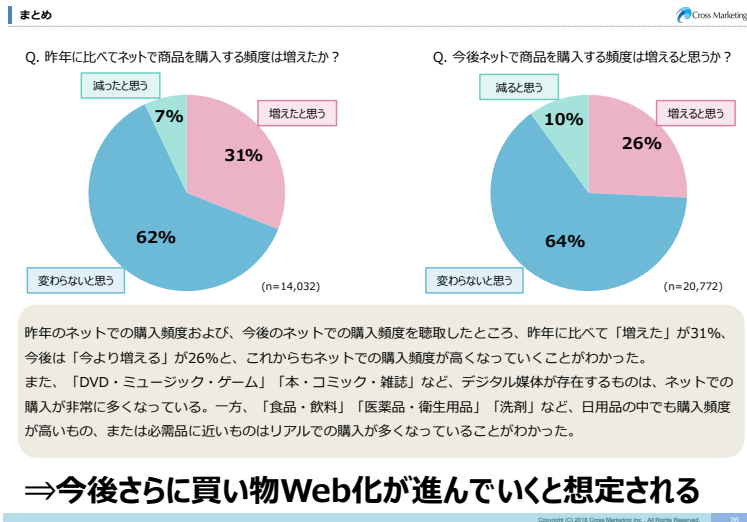


図 1.1: オンライン消費に関する調査：クロス・マーケティング [2] より転載。

案 (以下レコメンドと呼ぶ) することができないことである。実店舗では、店員がコーディネートした服をディスプレイしたり実際に顧客の希望を聞いて店員がレコメンドすることができる。これらのレコメンドにより顧客側は満足し、購買意欲にもつながるため E コマース上のレコメンドシステムはファッション業界では重要なものと考えられる。

しかしながら、急速に商品が変わるファッションにおいて、新商品やユーザの潜在的嗜好がわからない商品に関しては E コマース上でのレコメンドが困難であると言える。このように前提となる情報が少ない状況をコールドスタートと呼び、コールドスタートでのレコメンド手法の開発は大変需要が高い。また昨今のファッションに関するレコメンド手法では、アイテム同士の組み合わせにも配慮することがユーザへの適切なレコメンドであると述べられている [4] が未だ関連する研究は少なく、従来のモデルはアイテムのカテゴリ、視覚的な要因に着目したレコメンドシステムがほとんどである。そのため細部の要因を反映したレコメンドが困難である。そこで本研究では、視覚情報だけでなくアイテムの生地感、サイズ感などの潜在的な詳細、商品のキャッチフレーズ、金額的側面、季節的ダイナミクスをユーザの潜在的要因との類似性から選定し、ユーザが購入傾向の高い関連アイテムをレコメンドするマルチモーダルなシステムを提案する。本手法により、より詳細なアイテムをレコメンドし、現実世界のダイナミクスにも対応することで従来のシステムより購買につながるシステムを可能にする。

本論文は、以下のような構成となっている。第 2 章でファッションにおけるレコメンドシステム、コーディネートに関する先行研究を紹介し、第 3 章で先行研究を考慮した上で、アイテムの潜在的な詳細、キャッチフレーズ、季節的ダイナミクスを加えた新たな手法を提案する。第 4 章では、実際の E コマースのデータを用いて最適なアイテムのレコメンドに関する有意性について、先行研究との比較評価実験を行う。最後に第 5 章で結論と今後の展望を述べる。

第2章 先行研究

本章では先行研究として、従来のファッションレコメンドシステムについて様々なアプローチを紹介し、さらにコールドスタートに対応した VBPR 手法 (Visual Bayesian Personalized Ranking) [5] について述べる。

2.1 従来のレコメンドシステムのアプローチ

従来のレコメンドシステムにおいては、ユーザの購買履歴や閲覧履歴などを利用して、そのユーザと類似したユーザが同じ商品を購入する傾向にあると仮定し、片方のユーザが購入したアイテムをレコメンドするユーザベース協調フィルタリング [6] とユーザが購入または閲覧したアイテムの情報を利用して、アイテム間で類似性があるものをレコメンドするアイテムベースフィルタリングが主に利用されている [7]。また、ユーザの購買履歴、行動履歴などを分析にかけて好みを予測してから、ランキング形式でユーザにアイテムをレコメンドする Bayesian Personalized Ranking (以下、BPR と呼ぶ) [8] がある。

先行研究 [6] で用いられた、ユーザベース協調フィルタリングは、アイテムデータの情報を必要とせず、ユーザデータのみを利用する。例えば、あるユーザ A とユーザ B が同年代や同じ地域に住んでいる、または同じアイテムを購入する傾向にあれば、ユーザ A があるパンツを購入したとき、ユーザ B もユーザの特徴が似通っていることからそのパンツを購入するだろうとレコメンドするものである。E コマース上ではよく「この商品を買ったユーザはこの商品も購入している」と表示される。先行研究の Yang Hu et al. の研究 [9] でもユーザベースの観点も加えたレコメンドシステムを提案しており、Amazon や楽天などの膨大なファッションサイトでは、自分に合った服装を選ぶのは困難であるため、ユーザ個人の年齢、職業、文化的背景、居住地などの特徴量から、個人の好みを抽出し、服装を提案するシステムを開発している。しかし、現実世界においては Amazon などの膨大なデータがあつてこそ成立する手法であり、この手法を普遍的に利用することは困難であるという課題があつた。

一方、ユーザベース協調フィルタリングとは対照的に、先行研究 [7] で用いられたアイテムベースフィルタリングは、ユーザの詳細な情報を必要とすることなくアイテムをレコメンドすることができる。例えばあるユーザ A が購入したアイテムのカテゴリーが T-shirt の傾向であれば、T-shirt をレコメンドする。またアイテムベースは様々なアプローチから提案することが可能である。例えば Vivek Sembium et al. の研究 [10] では Amazon の購入だけでなく返品にも着目して提案している。実店舗では試着してサイズを合わせることができるが、オンライン上での商品の購入はサイズがわからないという大きな課題があり、実店

舗に比べ、返品率が高くなる。そこで、この研究では購入データと返品データのアイテムからユーザに適したサイズの提案をヒンジ損失で表現する提案を行っている。Eコマース上では、よく、「関連したアイテム」「この商品もおすすめ」と表示される。しかし、この手法ではアイテムの要因がカテゴリーだけであり、ファッションの全体、とくにコーディネートまで考慮したレコメンドは困難であった。また、デニムだけでも、ダメージデニム、ワイドデニムなど複雑なカテゴリとなっているため、単一の要因だけでなく様々な要因を考慮する必要がある。近年提案された BPR [8] はユーザベースとアイテムベースの両方を加えたハイブリッド型のシステムであり、統合した要素の中から最もユーザの嗜好に類似したアイテムを順番にレコメンドする手法である。しかし、このレコメンドシステムはコールドスタートにおいて関連性ある事前のデータがないため、適切にアイテムをレコメンドすることができない懸念が生じる。そのため、ユーザがクリック、購入したアイテムからコンテキストなどの情報を取り出し、コールドスタートに対応する手法で拡張する必要があった。

2.2 コールドスタートに対応した手法に関する研究

レコメンド手法には画像から特徴量を抽出する Convolutional Neural Network (以下, CNN と呼ぶ), または言語を数値化して分類を行う自然言語処理が主に拡張機能として利用することができる。

CNN には画像から直接色などの特徴をベクトルとして取り出すエンコーダや、事前にラベルの付いた大量の画像を学習させることでデータから、スーツであればスーツの特徴を、コートであればコートの特徴を簡単に数値として抽出する EfficientNet [11] などがある。EfficientNet のようなものを事前トレーニング済み CNN と言い、転移学習で新しく T-shirts 類を長袖や半袖、ノースリーブなどにも分類することも容易である。ファッションの画像処理を利用したシステムとしては、主にファッションのコーディネートに着目する研究が多く、これまでファッションに関する複数の商品のコーディネートを定量化した研究はあったが、現実的な状況を考慮できた提案システムはなかったため、Eric Kim et al. の研究 [12] では、現実存在する場面を画像データとして用い、その画像データとファッションの関係性、注意箇所を考慮した注意機構を用いたものを提案した。例えば、海の画像に対しては通常の服を提案するのではなく、水着を提案するといったものである。「インスタ映え」という言葉が流行語になる程、ファッションにおいてもそのユーザがいる場所とファッションの関係性を考慮したものは高いニーズがあるため、このようにコーディネートだけでなく場面に合わせてアイテムを提案するシステムはユーザにとって有益なものであると考えられる。

一方、自然言語処理は、パンツ、スニーカーなどのカテゴリ言語をベクトル化または確率的数値に変換することで、そのカテゴリ同士が類似性があるかを定量的に表す、または分類することができる。例えば、Yang Hu et al. [9] は自然言語処理の Latent Dirichlet Allocation (以下, LDA と呼ぶ) を利用して、ファッションアイテムのカテゴリーを 20 個のトピックであると仮定して、レコメンドするアイテムが 20 個のトピックの中ではどれに当てはまるか

を確率的な数値に変換することで要素として加えることを可能にした．他にはコーディネート観点で，Yuncheng Li et al. [13] が言語の分散表現手法，word2vec を利用して事前にファッションコーディネートされたアイテム群の各アイテム名をベクトル化し，コーディネートを定量化することで，新しいアイテムがそのコーディネートに合うかどうかの指標を作成した．Ruining He et al. の研究 [5] では，コールドスタートに対応するために，2.1 節で挙げた BPR に CNN を利用して視覚情報を加えた Visual Bayesian Personalized Ranking (以下，VBPR と呼ぶ) を提案した．この研究は先行研究 [9, 12, 13] とは異なり，実際のサイト上のデータを利用して購買を行うためのレコメンドシステムを提案している．またそのデータは，約 9 割が新商品のデータで構成されているためコールドスタートとして利用することができる．

VBPR は，最初に行列分解モデルを用いて予測式を立てる．行列分解モデルとは，行列分解 (Matrix Factorization) を用いて，次元削減を行う手法である．ユーザ数を m ，アイテム数を n とすると，各ユーザ u による各アイテム i に対しての評価データの $m \times n$ 行列表 X が得られる．ただし，ここで行列表 X の階数は c であり， $c \ll \min(m, n)$ を満たすものとする．このとき行列表 X を，

$$X \approx P^T Q \quad (2.1)$$

のように 2 つの行列表の積の形に分解して近似する．ここで P は $c \times m$ 行列表， Q は $c \times n$ 行列表である． P や Q のそれぞれの要素は，単なる潜在因子であり，ユーザとアイテムの相関を明示的に示しているものではない．行列分解では行列表 X と行列表 $P^T Q$ との誤差

$$J = \frac{1}{2} \|X - P^T Q\|^2 \quad (2.2)$$

が小さくなるように更新する．元々の評価データの行列表 X に欠損値がある場合，実際に観測されたユーザとアイテムのペア (u, i) の評価値 $x_{u,i}$ のみを用いて欠損値を含まない評価値行列表 \hat{X} を推定することができる．VBPR では，行列分解 $P^T Q$ をさらにいくつかの要因に分けて \hat{X} を予測している．ここで \hat{X} の (u, i) 成分 $\hat{x}_{u,i}$ は，ある特定のユーザ u のある特定のアイテム i に対する予測評価値を表しており，ここで i をユーザ u が購入したアイテムとすると，ユーザ u に対して，レコメンドするアイテム i の予測式 $\hat{x}_{u,i}$ は，

$$\hat{x}_{u,i} = \alpha + \beta_u + \beta_i + \gamma_u^T \gamma_i + \mathbf{V}_u^T \mathbf{V}_i \quad (2.3)$$

と表される．ここで， α はグローバルオフセット (定数)， β_u と β_i は定数であり，ユーザ u のバイアスとアイテム i のバイアスを表している． γ_u ， γ_i はそれぞれ次元 c の縦ベクトルであり，内積 $\gamma_u^T \gamma_i$ はユーザ u ，アイテム i の潜在的要因を表したスカラーとなる．潜在的要因とは例えばユーザ u が E コマース上でクリックした回数や閲覧数など，ユーザ u の嗜好を反映した要素を表している．つまり， $\gamma_u^T \gamma_i$ はユーザ u が購入した，または一度は興味を示したアイテム i の評価を表している．理論上，潜在的要因 $\gamma_u^T \gamma_i$ だけで予測を行うことは可能であるが，アイテムに関する潜在的要因を推定するには関連した観測が少ないため，

要因を潜在的要因と視覚的要因 $V_u^T V_i$ に分けている． V_u , V_i は、それぞれ次元 c の縦ベクトルであり、内積 $V_u^T V_i$ は、ユーザ u の嗜好に合わせたアイテム i の視覚的特徴量を表したスカラーである．この視覚的要因は python の Deep Learning を用いた CNN から抽出される．これらを用いた行列分解モデルから、予測評価値 $\hat{x}_{u,i}$ を導き出す．また、ユーザの嗜好がないアイテム j に関しても、予測評価値 $\hat{x}_{u,j}$ を求め、 $\hat{x}_{u,i}$ から $\hat{x}_{u,j}$ を引いた値を、

$$\hat{x}_{u,i,j} = \hat{x}_{u,i} - \hat{x}_{u,j} \quad (2.4)$$

と表し、VBPR の最終的な予測評価値となっている．また $\hat{x}_{u,i,j}$ を最適化するにあたり、次のように確率的勾配上昇法を取り入れた最適化フレームワーク、ランキング学習のペアワイズ法を取り入れた

$$\max_{\Theta} \sum_{(u,i,j) \in Ds} \ln \sigma(\hat{x}_{u,i,j}) - \lambda_{\Theta} \|\Theta\|^2 \quad (2.5)$$

を用いて最適化する．ここで $\sigma(\hat{x}_{u,i,j})$ は $\hat{x}_{u,i,j}$ を独立変数としたシグモイド関数を表す． λ_{Θ} はハイパーパラメータであり、グリッドサーチ探索を用いて、最良のパフォーマンスができる数値を探索する． Ds はトレーニングに用いるデータセットを

$$Ds = \{(u, i, j) | u \in U \wedge i \in I_u^+ \wedge j \in I \setminus I_u^+\} \quad (2.6)$$

で表している．この Ds はユーザ u に対して、嗜好を反映したアイテム i 、嗜好を反映していないアイテム j の組の集合を表す．パラメータ Θ は学習するすべてのパラメータを表し、次の式で更新を行う．

$$\Theta \leftarrow \Theta + \eta \cdot (\sigma(-\hat{x}_{u,i,j}) \frac{\partial \hat{x}_{u,i,j}}{\partial \Theta} - \lambda_{\Theta} \Theta) \quad (2.7)$$

ここで、 η は学習率である．これにより確率的勾配上昇法を用いて、優先的に順位の高い、つまり人気のアイテムを予測する．このシステムは、ユーザの嗜好を反映した潜在的視覚要因を考慮したアプローチであり、ユーザが紫の服の色を好む傾向にあれば紫の服を中心としたアイテムを自動的に提案することが可能となる．機械学習の最終的な評価推定を示す Area Under Curve (以下、AUC と呼ぶ) の結果では、従来の現実世界の E コマースに合わせた BPR [8] よりも有意性があることが分かった．

このように、これまで研究されてきたファッションのレコメンドシステムは、視覚情報またはユーザの潜在的要因のみを拡張して、レコメンドするものがほとんどであり、アイテムの詳細な部分を考慮しているものが存在しない．一般的に、ユーザが新商品を購入する要因は視覚情報だけでなく、値段や商品のコピーライティングなど別の要因もある．また特にファッションにおいては、フィット感やサイズ、生地の性質、モードファッション、ストリートファッションなど様々な購入傾向を決める要因があると考えられる．フィット感は視覚情報だけの CNN で考慮するには限界がある．また VBPR を改良した研究は他にもあるが、レコメンドされたアイテムが偏ってしまう傾向がある．例えば図 2.1 は Wang-Cheng et al. が提案した DVBPR [14] で、ユーザ A とユーザ B にレコメンドされたアイテムである．

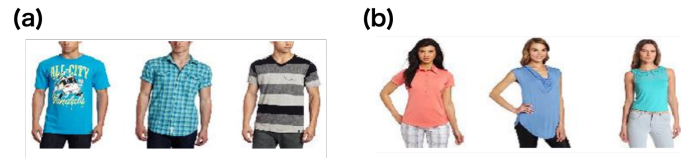


図 2.1: DVBPR [14] の結果.(a) ユーザ A に Recommend された上位 3 つのアイテム (b) ユーザ B に Recommend された上位 3 つのアイテム.

どちらも似通ったアイテムに偏ってしまっていることが見てとれる．また，E コマース上では季節変動などがあるため，その時期に合わせたカテゴリのアイテムを Recommend する必要性もあると考えられる．

そこで本研究では，ユーザの嗜好に合わせたアイテムの生地感，サイズ感などの詳細を加え，季節の時間軸に合わせたアイテムを Recommend するシステムを構築する．また従来の研究で考慮されなかったアイテムの売り文句，金額的な要因を考慮した新たな Recommend システムを提案する．

第3章 提案手法

本研究では、第2章で紹介した VBPR などの視覚的要因を考慮したレコメンドシステムに加え、ユーザの嗜好に合わせた、生地感などのアイテムの詳細や季節的時間軸を考慮した手法と、さらにマーケティングの観点から用いられるユーザが購入しやすい売り文句や金額的側面を考慮したマルチモーダルレコメンドシステム“Multimodal Interface Bayesian Personalized Ranking (以下、MIBPR と呼ぶ)”を提案する。これにより、ユーザの詳細な嗜好と購買率を促進する要素を反映させたレコメンドが可能となる。本章では、今回新たに提案する MIBPR の詳細について説明する。

3.1 視覚的要因と季節的要因

まずはじめに、このモデルで用いられる変数を表1に示しておく。先行研究の VBPR[5]をもとに、行列分解モデルを用いて予測式を立てる。

表 3.1: MIBPR で使用する変数.

変数	意味
U, I	ユーザセット, アイテムセット ($U = \{u_1, u_2, \dots, u_m\}$, $I = \{i_1, i_2, \dots, i_n\}$)
I_u^+	ユーザ u が購入したアイテムセット
$\hat{y}_{u,i}$	アイテム i に対するユーザ u の予測値
α	グローバルオフセット (定数)
β_u, β_i	ユーザ u のバイアス, アイテム i のバイアス (定数)
γ_u, γ_i	ユーザ u , アイテム i の潜在的要因
V_u, V_i	ユーザ u , アイテム i の視覚的要因
W_u, W_i	ユーザ u , アイテム i のタイトルの要因
θ_u, θ_i	ユーザ u , アイテム i のフレーズの要因
M_u, M_i	ユーザ u , アイテム i の金額的要因
S_u, S_i	ユーザ u , アイテム i の季節的要因
f_i	アイテム i の画像を線形変換した表示ベクトル

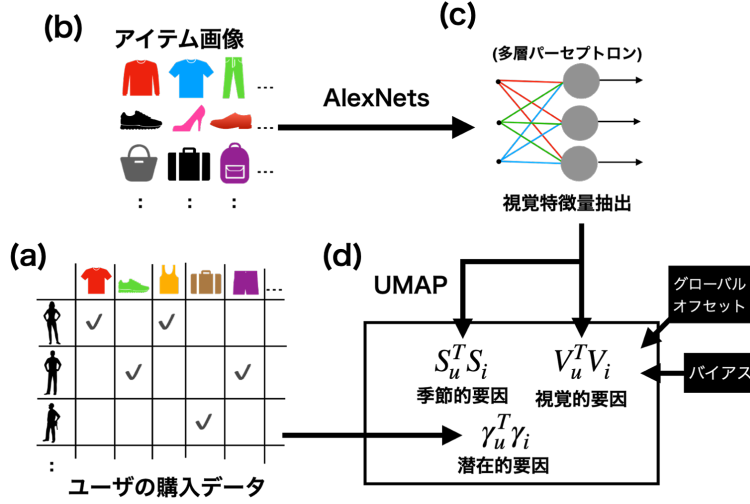


図 3.1: 潜在的要因 $\gamma_u^T \gamma_i$, 視覚的要因 $V_u^T V_i$, 季節的要因 $S_u^T S_i$ の概略図. (a) ユーザがアイテム i を購入したデータ. (b) アイテムセット I の画像データ. (c) 視覚的特徴量抽出の図. (d) a, b, c から取り出された要因

ユーザ数を m , アイテム数を n とすると, 各ユーザ u による各アイテム i に対しての評価データの $m \times n$ 行列 Y が得られる. ただし, ここで行列 Y の階数は c であり, $c \ll \min(m, n)$ を満たすものとする. このとき行列 Y を,

$$Y \approx P^T Q \quad (3.1)$$

のように2つの行列の積の形に分解して近似する. ここで P は $c \times m$ 行列, Q は $c \times n$ 行列である. 元々の評価データの行列 Y に欠損値がある場合, 実際に観測されたユーザとアイテムのペア (u, i) の評価値 $y_{u,i}$ のみを用いて欠損値を含まない評価値行列 \hat{Y} を推定することができる. 本研究では, 行列分解 $P^T Q$ をさらにいくつかの要因に分けて \hat{Y} を予測する. ここで \hat{Y} の (u, i) 成分 $\hat{y}_{u,i}$ は, ある特定のユーザ u のある特定のアイテム i に対する予測評価値を表しており, ここで i をユーザ u が購入したアイテムとすると, ユーザ u に対して, 購入したアイテム i の予測評価値 $\hat{y}_{u,i}$ は,

$$\hat{y}_{u,i} = \alpha + \beta_u + \beta_i + \gamma_u^T \gamma_i + V_u^T V_i \quad (3.2)$$

で推定できる. ここで, α はグローバルオフセット (定数), β_u と β_i は定数であり, ユーザ u のバイアスとアイテム i のバイアスを表している. 図 3.1 は, このモデルにおける視覚的要因, 潜在的要因と季節的要因の流れを表した概略図である. γ_u , γ_i はそれぞれ次元 c の縦ベクトルであり, 内積 $\gamma_u^T \gamma_i$ はユーザ u , アイテム i の潜在的要因を表したスカラーとなる. つまり図 3.1(a) のように各ユーザに対して, どのアイテムが購入されたかをデータ化することで, ユーザ u の嗜好を反映した要素を表している. V_u , V_i は, それぞれ次元 c の縦ベクトルであり, 内積 $V_u^T V_i$ は, ユーザ u , ユーザ u の嗜好に合わせたアイテム i の視覚的特徴量を表したスカラーである. アイテム画像 (図 3.1(b)) はそのままでは使用できないため, 事前トレーニング済み CNN の AlexNets [15] を用いて, 視覚的特徴量を抽出する (図

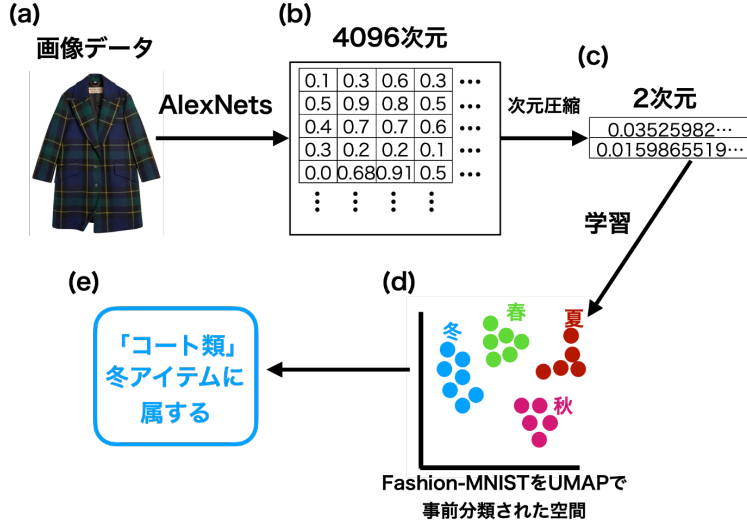


図 3.2: アイテム画像を季節にクラスタリングする流れ. (a) アイテム i の画像データ. (b) 4096 次元の視覚的特徴量. (c) 2 次元に圧縮された視覚的特徴量. (d) 視覚的特徴量から季節で分類された空間. (e) アイテム i が季節に分類された結果

3.1(c)). これにより視覚的要因 $\mathbf{V}_u^T \mathbf{V}_i$ を得る. 次に, 季節的要因の抽出方法について流れを図 3.2 に示す. まず, 画像データ (図 3.2(a)) から AlexNets [15] を用いて取り出した 4096 次元の視覚的特徴量 (図 3.2(b)) を主成分分析などを用いて 2 次元データ (図 3.2(c)) に圧縮する. 次に, アイテムのカテゴリが事前に付いた fashion-MNIST [16] データを利用して, 圧縮した 2 次元の特徴量を季節にクラスタリング (図 3.2(d)(e)) する.

今回, クラスタリング手法には, 多様体学習の 1 つである UMAP [17] を用いた. UMAP の特徴は, データをトレーニングデータとテストデータに分けることができ, 事前にトレーニングした空間にテストデータを付与することができる点である. 図 3.3 は, トレーニングデータとして三次元の空間に fashion-MNIST を付与したグラフである. x 軸, y 軸は UMAP を用いて抽出した 2 つの特徴量を表しており, z 軸は, 10 個のカテゴリごとの階層を表している. このグラフに, アイテムデータセット I をテストデータとして学習させて, カテゴリの ankle boots や coat など秋冬に限定されるアイテムを冬と定義し, T-shirts, Sandals など春夏と定義して用いる. Bag や Sneaker などは視覚的特徴量から季節を分類することは困難であるため, 本研究では全ての季節で使用できると仮定した. \mathbf{S}_u , \mathbf{S}_i はそれぞれ次元 c の縦ベクトルであり, 内積 $\mathbf{S}_u^T \mathbf{S}_i$ はユーザ u , アイテム i の指定した季節に分類された視覚的特徴量を表したスカラーとなる. この季節的要因を次のようにモデルに導入する.

$$\hat{y}_{u,i} = \alpha + \beta_u + \beta_i + \gamma_u^T \gamma_i + \mathbf{V}_u^T \mathbf{V}_i - \mathbf{S}_u^T \mathbf{S}_i \quad (3.3)$$

$\mathbf{S}_u^T \mathbf{S}_i$ は $\mathbf{V}_u^T \mathbf{V}_i$ を季節に分類したものであり, 事前に E コマースで使用する実際の季節とは異なる季節的要因を指定する. つまり, 5 月であれば, 秋冬を指定し, 12 月であれば春夏を指定することで, 数式 (3.2) の $-\mathbf{S}_u^T \mathbf{S}_i$ により提案すべきでない季節の要素が除かれる. これにより, レコメンドしたい季節のアイテムを優先的にレコメンドすることが可能になる.

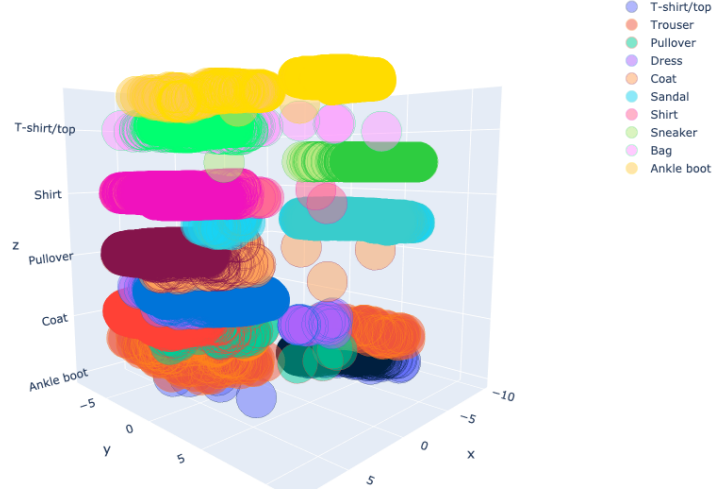


図 3.3: fashion-MNIST [16] を UMAP でクラスタリングしたグラフ. x 軸, y 軸は特徴量を 2 次元に圧縮した軸. z 軸はカテゴリごとの階層を表した軸.

ただし, 実際のデータでは V_i と S_i は 4096 次元と, このままでは次元が大きいため, 畳み込みにより 20 次元に次元圧縮した V_{fi} , S_{fi} に変換する. つまり変換後の $\hat{y}_{u,i}$ は,

$$\hat{y}_{u,i} = \alpha + \beta_u + \beta_i + \gamma_u^T \gamma_i + V_u^T V_{fi} - S_u^T S_{fi} \quad (3.4)$$

となる. このモデルにより, 先行研究の VBPR に加えて, 季節的要因を考慮することができる. 本研究ではさらに, アイテム名やキャッチフレーズ, 金額の情報をもとに, さらにユーザビリティの高いシステムを提案する.

3.2 アイテムのタイトルとフレーズの自然言語処理

本節では, アイテム名やキャッチフレーズなどのテキスト要素をモデルに加える手法について説明する.

アイテムのタイトルをベクトル化するために Jeffrey et al. の提案した Glove [18] を用いる. Glove は単語のベクトル表現を取得するための教師なし学習アルゴリズムであり, 事前トレーニングされたグローバルな 220 万個の単語全体から, 本研究で学習させる単語間の共起頻度を最適化関数を用いて単語一つにつき, 300 次元の共起行列に変換する. 最適化関数は,

$$\mathbf{W} = \sum_{p,q=1}^N f(X_{p,q})(\mathbf{w}_p^T \tilde{\mathbf{w}}_q + \mathbf{b}_p + \tilde{\mathbf{b}}_q - \log X_{p,q})^2 \quad (3.5)$$

と表す. N は単語の総数を表しており, $X_{p,q}$ は単語 p と q の共起頻度, \mathbf{w}_p は事前学習させた単語 p のベクトル, $\tilde{\mathbf{w}}_q$ は本研究で使用する単語 q のベクトル, \mathbf{b}_p , $\tilde{\mathbf{b}}_q$ はそれぞれ単語 p , q のバイアス項を表している. 関数 f は重みを加えるために使用される. また, 関数 f は a, in, with などの補助後は単語の重みを低くする役割も持つ. この手法で得たベクトル \mathbf{W} は,

それぞれユークリッド距離またはコサイン類似度で意味的類似性を表し、例えば、ice という単語に対して“solid, gas, water, fashion”は“1.water, 2.solid, 3.fashion, 4.gas”の順番に類似性があることを表すことができる。本研究では、アイテム i のタイトルを単語ずつ取り出し、取り出した全ての単語 q 個から n 個のアイテムの間に意味的類似性を表したベクトルに変換する。アイテム i のタイトルは複数の単語で構成されているため、得た単語ベクトルを加重平均したものを \mathbf{W}_i とする。 $\mathbf{W}_u^T \mathbf{W}_i$ は、ユーザ u 、アイテム i の単語の意味的類似性を表したスカラーである。この $\mathbf{W}_u^T \mathbf{W}_i$ を予測評価値推定のための行列分解モデルに加える。

次にアイテムのフレーズの処理について説明する。本研究でフレーズとは、キャッチコピーまたはアイテムの詳しい説明文のことを指す。フレーズは先ほど述べたタイトルの要因と違い、文章体であり、Glove を利用することが難しい。例えば、単語「fits」が使われる例で見ると、

- This jeans **fits** those who want to make their legs look slimmer.
このジーンズは足を細く見せたい人に**適しています**。
- This jeans **fits** yourself for coldness.
このジーンズなら寒さにも**耐えられる**。
- This jeans is easy to move and **fits** for work.
このジーンズは動きやすく、仕事でも**使いやすい** (動きの観点で)。
- Dress length is 68cm and A person with a height of 170 cm **fits** perfectly.
着丈が 68cm であり、170cm の方には**ぴったり合います**。

などが挙げられる。これらの4文はファッションに関連した文章体であり、“fits”という単語が全ての文章の中に含まれるが、それぞれ意味やニュアンスが異なることがわかる。このようにフレーズでは、1つの単語に対して、文脈に合うそれぞれの意味を考慮する必要がある。しかし、Glove などの単語分散表現では、文脈に合わせた意味は表さず、単語単体の意味で、単語間の類似性を表現するものである。特に注意すべき点は、ファッションにおいては一見関係のない単語が文章中では重要な相関がある場合が多いということである。そこで本研究のフレーズ処理では、文章中の単語の意味を考慮して文章を直接ベクトル化する Embeddings from Language Models (以下、ELMo と呼ぶ) [19] を用いる。ELMo の具体例を図 3.4 に示す。例えば、“wide pants is twice as long, but three times as wide.” というフレーズがあったとする。最初に、フレーズの単語を1つずつ取り出し、双方向言語モデル、bi-directional LSTM (以下、biLSTM と呼ぶ) で読み取り、分散表現として単語情報を得る。そして、次の単語情報を biLSTM で得ると共に、手前 (左の単語) の biLSTM で得た単語情報も学習させる。これを左から右 (図 3.4(a)) に繰り返し行うことでフレーズ内の意味を学習した ELMo を作成する。図 3.4(a) では、“wide” という単語の分散表現を biLSTM1 で学習させる。次に、biLSTM1 を biLSTM2 に学習させる。さらに学習させた biLSTM2 に “pants”

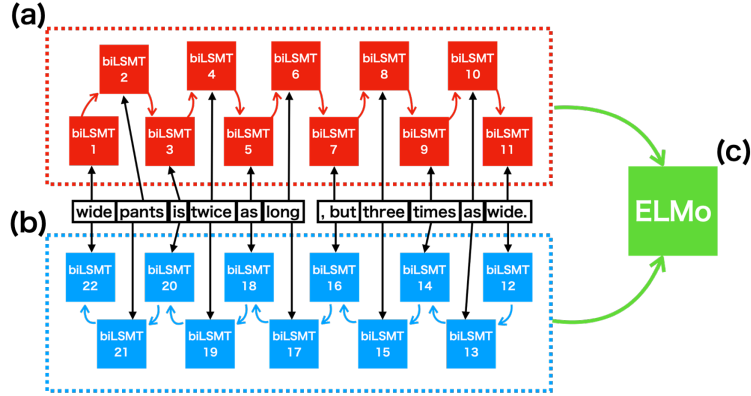


図 3.4: ELMo [19] のアルゴリズム. (a) フレーズを左から右に学習させた例. (b) フレーズを右から左に学習させた例. (c) 最終的にフレーズを学習させた分散表現

を学習させる. この工程を全ての単語で繰り返し, 最終的に全ての単語と biLSTM を学習させた biLSTM11 を得る. 同様に, フレーズを逆 (図 3.4(b)) から学習させた biLSTM22 を得る. この biLSTM11 と biLSTM22 を学習させたものが ELMo (図 3.4(c)) となる.

本研究は AllenNLP¹が提供する大量のコーパスを事前学習させた ELMo を用いて, 使用する文章をベクトルデータ θ_i に変換する. $\theta_u^T \theta_i$ は, ユーザ u , アイテム i のフレーズを表したスカラーであり, この $\theta_u^T \theta_i$ を予測評価値推定のための行列分解モデルに加える. このフレーズには, 配達のスPEEDであったり, アイテムのサイズ感, 売り文句, 割引など, ファッションだけでなく様々な詳細も含んでいる.

最後に, 金額的要因については, 説明変数として加える. アイテムデータ n 個の全ての金額を取り出し, スタージェスの公式より, 階級数を算出する. スタージェスの公式は

$$k = \log_2 n + 1 \quad (3.6)$$

となる. k は階級数を表しており, アイテム i に関する k 次元ベクトル M_i を作成する. M_i は, ユーザ u が購入したアイテム i の金額が, 階級数 k の中に含まれる場合は 1 を, 含まれない場合は 0 とする説明変数である. さらに M_u をユーザ u が購入する金額を表す k 次元ベクトルとして, 最終的に内積 $M_u^T M_i$ は, ユーザ u , アイテム i に関する金額的要因を表したスカラーである. この $M_u^T M_i$ を予測評価値推定のための行列分解モデルに加える.

3.3 MIBPR の予測モデル

3.2 節で述べたアイテム名, フレーズ, 金額に関する要因を式 (3.2) に加えたものが新たなモデル MIBPR である. つまり, 購入したアイテムの予測値 $\hat{y}_{u,i}$ は,

$$\hat{y}_{u,i} = \alpha + \beta_u + \beta_i + \gamma_u^T \gamma_i + (V_u^T V_{fi} - S_u^T S_{fi}) + W_u^T W_i + \theta_u^T \theta_i + M_u^T M_i \quad (3.7)$$

¹Allen NLP ([texttthttps://allennlp.org/](https://allennlp.org/))

と表される．式 (3.7) はユーザが購入したアイテム i に関する予測式であるが，同様にユーザ u が購入していないアイテム j についても $\hat{y}_{u,j}$ を求めることができる．さらに，アイテム i の予測値 $\hat{y}_{u,i}$ からアイテム j の予測値 $\hat{y}_{u,j}$ を差し引いた予測値 $\hat{y}_{u,i,j}$

$$\hat{y}_{u,i,j} = \hat{y}_{u,i} - \hat{y}_{u,j} \quad (3.8)$$

が，本研究の最終的な予測モデルである．この $\hat{y}_{u,i,j}$ を損失関数の引数として使用し，最適化フレームワーク，ランキング学習のペアワイズ法を用いることで $\hat{y}_{u,i,j}$ を最適化することができる．最適化は次式により計算する．

$$\max_{\Theta} \sum_{(u,i,j) \in Ds} \ln \sigma(\hat{y}_{u,i,j}) - \lambda_{\Theta} \|\Theta\|^2 \quad (3.9)$$

ここで， $\sigma(\hat{y}_{u,i,j})$ は損失関数であり，今回は，シグモイド関数を用いている． λ_{Θ} はハイパーパラメータであり，グリッドサーチ探索を用いて，最良のパフォーマンスを可能にする数値を探索する． Ds は，トレーニングに用いるデータセットを表し，

$$Ds = \{(u, i, j) | u \in U \wedge i \in I_u^+ \wedge j \in I \setminus I_u^+\} \quad (3.10)$$

と表せる． i はユーザ u に対して，購入したアイテム， j は購入していないまたは認識できなかったものである．式 (3.9) のパラメータ Θ は，実際には 12 個のパラメータ $\beta_u, \beta_i, \gamma_u, \gamma_i, V_u, V_{fi}, W_u, W_i, \theta_u, \theta_i, M_u, M_i$ が代入され計算される．このパラメータ Θ は，次の確率的勾配上昇法を用いた次の式で更新を行う．

$$\Theta \leftarrow \Theta + \eta \cdot (\sigma(-\hat{y}_{u,i,j}) \frac{\partial \hat{y}_{u,i,j}}{\partial \Theta} - \lambda_{\Theta} \Theta) \quad (3.11)$$

ここで η は学習率である．この更新式により，優先的にユーザの嗜好に合わせた順位の高いアイテムの順番に予測を行う．Google などの検索時に閲覧数が多い人気のサイト順に出す手法と同じである．

第4章 評価実験

第3章で述べた本研究のMIBPRの有用性を検証するために、システムに対する評価実験を行った。比較対象として、先行研究のVBPRシステムと、本研究のレコメンドシステムであるMIBPRをMathWorks¹が提供するMatlabとPythonを用いて構築し、それぞれに対するAUC評価推定を行い検証した。

4.1 データの概要

データセットは現実世界のファッション通販サイトのデータ、Tradesy [20] のデータを使用する。Tradesyは大手海外ファッションサイトであり、ユーザ数33,864人が購入したデータ、アイテム数256,996個のデータで構成されている。アイテムには、アイテム画像データ、実際のアイテムが販売されているURLのデータが含まれる。ユーザが購入したアイテムデータの数はユーザー一人あたり20個未満であり、また本研究では最終的な評価を行うテストデータとして、Wang-Cheng らのDVBPR [14] で用いられたユーザ1人あたりの購入アイテムを1つのみに再構築したデータを使用する。本研究で新たに加える視覚的要因、金額的要因、フレーズの要因、およびタイトルの要因について、実際のEコマースのサイトの具体例を図4.1に示す。それぞれの要因データはPythonを用いて、スクレイピングを行い、AlexNets [15]、Glove [18]、ELMo [19]、スタージェスの公式により数値化する。

季節的要因については、図3.3で示したUMAPで作成した空間にTradesyデータのアイテム画像の視覚的特徴量をテストデータとして学習させた。図4.2は、最終的にクラスタリングを試みた結果である。横軸と縦軸は2つの特徴量を表す。データが大きな1つのクラスタになっており、Tradesyのデータでは、UMAPを用いてカテゴリに分類することができなかった。UMAPを用いる際には様々な工夫が必要であるため、現時点では、アイテムを季節的要因に分類する実装には改善が必要である。また上手く実装されない原因として、Tradesyは様々なカテゴリが存在する(Tシャツであれば、半袖、長袖、ノースリーブなど)ことも考えられる。しかし、今回の評価実験ではTradesyを用いるため、まずは、季節的要因を除いて評価実験を行うとする。

¹MathWorks (<https://jp.mathworks.com/>)

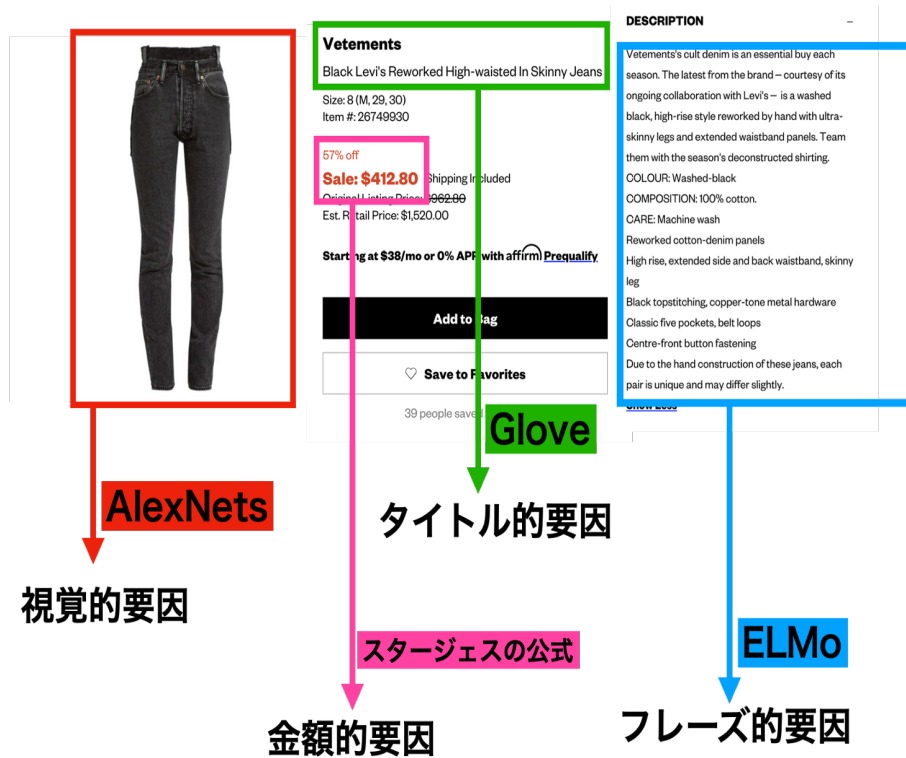


図 4.1: 実際の Tradesy サイト [20] のアプローチの流れ. アイテムの画像を AlexNets を用いて処理し, 視覚的要因として加える. アイテムのタイトルを GloVe を用いて処理し, 金額はスタージェスの公式を用いる. アイテムの詳細を記述した文章に ELMo を用いて処理し, フレーズ的要因として加える.

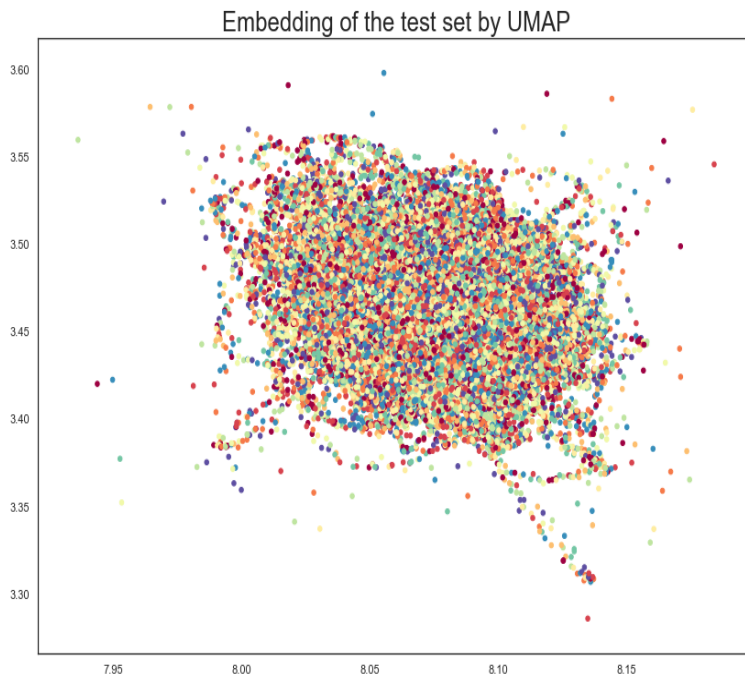


図 4.2: Tradesy データ [20] を用いた UMAP の結果. 横軸と縦軸は特徴量を 2 次元に圧縮した軸.

4.2 結果

本研究の評価推定には先行研究 [5] で用いられた AUC を用いる。AUC は

$$AUC = \frac{1}{|U|} \sum_u \frac{1}{|E(u)|} \sum_{(i,j) \in E(u)} \delta(\hat{y}_{u,i} > \hat{y}_{u,j}) \quad (4.1)$$

を用いて計算する。ここで $E(u)$ は

$$E(u) = \{(i,j) | (u,i) \in Tu \wedge (u,j) \notin (Pu \cup Vu \cup Tu)\} \quad (4.2)$$

であり、 Tu はテストデータ、 Vu は検証用データ、 Pu はトレーニングデータを表し、 $\delta(\hat{y}_{u,i} > \hat{y}_{u,j})$ は指示関数で、条件が合えば 1 を返すようになっている。この AUC の値が 1 に近ければ近いほど最適化していることを表す。

評価の結果を図 4.3 に示す。横軸は Epoch 数、縦軸は AUC の値を表している。先行研究の VBPR [5] を Visual (黒線)、MIBPR のすべての要素を合わせたものを All (赤線)、フレーズと視覚的要因を合わせたものを Phrase and Visual (青線)、タイトルと視覚的要因だけ組み合わせたものを Title and Visual (紫線)、タイトルとフレーズと視覚的要因だけ組み合わせたものを Title and Phrase and Visual (緑線) で示している。またデータ量は十分多いため、学習率 $\eta = 0.005$ と低くし、Epoch 数を 80 回に設定している。図 4.3 からタイトルを入れた 3 つの手法は、Epoch30 以降、先行研究よりも AUC の値は高い結果となった。また、図 4.4 は、Epoch80 の AUC 評価値の値を対数で表した棒グラフである。図 4.4 からわかる通り、先行研究の視覚的要因のみのシステムに比べて、新しい要素を加えた 4 つのシステムは有意性があることがわかった。しかし、すべての要因を加えた MIBPR と金額的要因の組み合わせを取り除いたシステム、つまりタイトル、フレーズ、視覚の 3 つを組み合わせたシステムでは AUC の値に大きな差はなかった。

次に、コールドスタートにおける評価実験について Epoch 数 10 毎の AUC の値を表 4.1 に示す。All items は図 4.3, 4.4 でも用いた Tradesy データを表し、Cold items は DVBPR [14] の研究で用いられたユーザ 1 人あたりが購入したアイテムが 1 つのデータで構成されている。表 4.1 から Cold items では、タイトルの要因、フレーズの要因、視覚的要因の組み合わせが、AUC の値が高い結果であった。

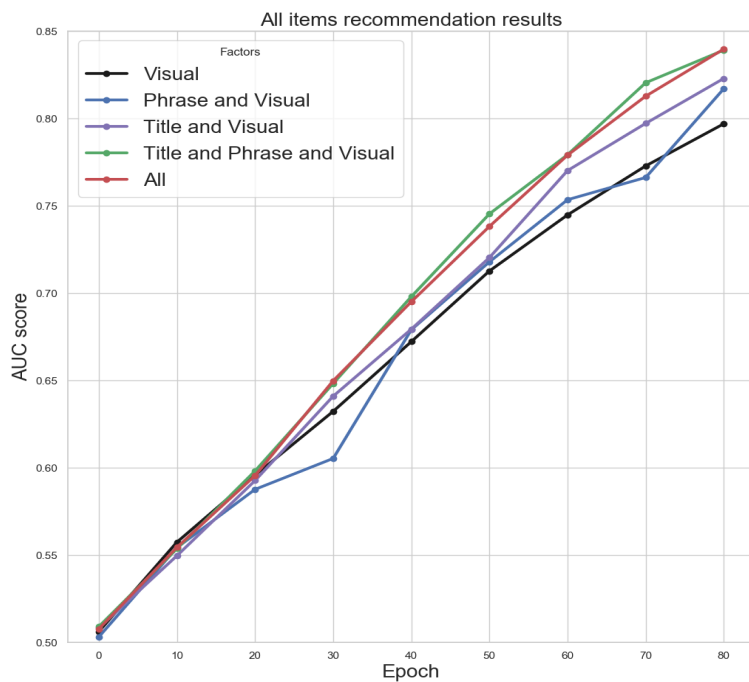


図 4.3: AUC 評価推定の結果. 横軸は Epoch 数, 縦軸は AUC の値を表す. 黒線は視覚的要因, 青線はフレーズと視覚的要因, 紫線はタイトルと視覚的要因, 緑線はタイトルとフレーズと視覚的要因, 赤線は全ての要因を用いた結果.

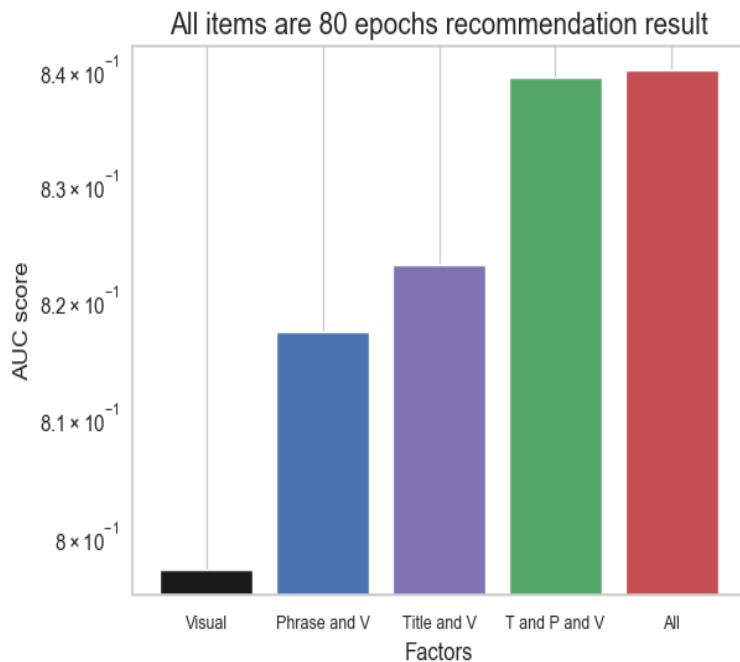


図 4.4: 80Epoch 目の All items の結果. 縦軸は AUC の値, 横軸は要因の組み合わせを表す. 黒棒は視覚的要因, 青棒はフレーズと視覚的要因, 紫棒はタイトルとフレーズと視覚的要因, 赤棒は全ての要因を用いた結果.

表 4.1: AUC : Epoch 数ごとの評価推定結果.

Factors	Item	Epoch0	Epoch10	Epoch20	Epoch30
Visual	All items	0.50624	0.55747	0.59583	0.63237
	Cold items	0.50894	0.55368	0.59080	0.62925
Phrase and Visual	All items	0.50320	0.55433	0.58770	0.60541
	Cold items	0.50903	0.55498	0.58348	0.60359
Title and Visual	All items	0.50826	0.54971	0.59273	0.64106
	Cold items	0.50915	0.54659	0.58991	0.63763
Title and Phrase and Visual	All items	0.50911	0.55372	0.59824	0.64832
	Cold items	0.51030	0.55005	0.59774	0.64133
All	All items	0.50777	0.55454	0.59558	0.64989
	Cold items	0.51396	0.55480	0.59233	0.64694

Factors	Item	Epoch40	Epoch50	Epoch60	Epoch70	Epoch80
Visual	All items	0.67229	0.71277	0.74487	0.77291	0.79705
	Cold items	0.67127	0.70786	0.74019	0.76736	0.79491
Phrase and Visual	All items	0.67909	0.71798	0.75356	0.76629	0.81715
	Cold items	0.67440	0.71164	0.74878	0.75882	0.81366
Title and Visual	All items	0.67944	0.72044	0.77025	0.79726	0.82301
	Cold items	0.68054	0.71810	0.76647	0.79606	0.82025
Title and Phrase and Visual	All items	0.69805	0.74546	0.77934	0.82054	0.83923
	Cold items	0.69702	0.74368	0.77365	0.81635	0.83607
All	All items	0.69515	0.73828	0.77908	0.81285	0.83984
	Cold items	0.69150	0.73405	0.77651	0.80944	0.83259

4.3 考察

本研究では、視覚的要因とタイトルの要因、フレーズ的要因、金額的要因を加えることにより、ユーザに対して、より適切にアイテムをレコメンドすることが可能となった。しかし、表 4.1 より、すべての要因を加えたシステムと金額的要因を取り除いたシステム、つまり視覚的要因、タイトルの要因、フレーズ的要因の組み合わせは Epoch80 の時点では大きな差はなく、Epoch 数 10 ごとの結果では、視覚的要因、タイトルの要因、フレーズ的要因の組み合わせの方が総合的に AUC の値が高いことが見られる。Cold items に関しても視覚的要因、タイトルの要因、フレーズ的要因の方が、AUC の値は総合的に高い。これは、金額的要因が、ユーザが購入したアイテムの金額以外は全て関心がないと定義したため AUC の値が低下したと考えられる。また、Cold items のレコメンドにおいて、視覚的要因、タイトルの要因、フレーズ的要因の組み合わせが最も良いパフォーマンスを行えたことから、要因を増やせば必ずしも精度が上がるわけではなく、適切な要因の組み合わせを行う必要がある。

第5章 結論と今後の展望

本研究では、従来の視覚的要因を考慮したレコメンドに加え、さらに、季節的要因、ユーザの購入傾向にあるフレーズ、アイテムのカテゴリ、詳細な内容、金額的要因を反映させたレコメンドシステムの手法を提案した。具体的には、それぞれの要因をランキング学習のペアワイズ手法に組み込んだ。これにより、ユーザの欲求やより求めるアイテムを自動的にシステムに反映することが可能となる。しかし、現時点では、季節的要因は視覚的アプローチでは困難であり、自然言語の観点からさらにクラスタリングを行う必要がある課題が残った。評価実験結果において金額的要因を加えたシステムよりも除いたシステムの方が評価推定が高かったのは、ユーザが必ずしもその値段帯のアイテムだけを購入するとは限らないためである可能性が高い。さらにファッションにおいては季節だけでなく、時代によって流行、好まれるアイテムが異なってくる。

今後、本研究では金額的要因の正しい定義と、季節的要因にクラスタリングするアプローチ、考慮できなかった時系列ダイナミクスの要因を加える改良を行なっていくことで、実際のEコマース上で、ユーザに対してより適切なアイテムをレコメンドすることが期待できる。最終的には、アイテムの適切なトレンド考慮を実現する時系列モデルを加えたレコメンドシステムの構築を目指す。

謝 辞

本論文を作成するにあたり，多大なるご指導を賜りました岩本 真裕子 准教授，伊藤 紀子 准教授，島根大学修士 藤原 魁人さんに厚く御礼申し上げます．最後に，卒業に至るまでの間ご指導いただいた諸先生方ならびに研究室メンバの皆様に心からの感謝の意を表す．

参考文献

- [1] 総務省,『第 1 部 特集 人口減少時代の ICT による持続的成長』,
<https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h30/html/nd142110.html>,
(参照 2020.06.20).
- [2] 株式会社 クロス・マーケティング,『オンライン消費に関する調査』, <https://www.cross-m.co.jp/report/it/os20180515/>, (参照 2020.06.20).
- [3] 日本経済新聞,『ZARA1200 店閉鎖、世界の 2 割弱 EC 比率を 25%に』,
<https://www.nikkei.com/article/DGXMZO60221280R10C20A6TJC000/>, (参 照
2020.06.20).
- [4] Ruiping Yin, Kan Li, Jie Lu, “Enhancing Fashion Recommendation with Visual Compatibility Relationship”, *Proceeding WWW ‘19 The World Wide Web Conference*, 3434-3440, 2019.
- [5] Ruining He, Julian McAuley. , “VBPR: Visual bayesian personalized ranking from implicit feedback” *Thirtieth AAAI Conference on Artificial Intelligence*. 144-150, 2016.
- [6] Brent Smith, Greg Linden, “Two decades of recommender systems at Amazon.com”, *IEEE Internet Computing*, **21** (3) , 12-18 , 2017.
- [7] Doug Riecken, “Personalized Views of Personalization”, *Communications of the ACM*, **43** (8), 26-28, 2000.
- [8] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, Lars Schmidt Thieme, “BPR : Bayesian personalized ranking from implicit feedback.”, *In Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence. AUAI Press*, 452-461, 2009.
- [9] Yang Hu, Xi Yi, Larry S. Davis, “Collaborative Fashion Recommendation: A Functional Tensor Factorization Approach” , *Proceedings of the 23rd ACM international conference on Multimedia*, 129-138, 2015.
- [10] Vivek Sembium, Rajeev Rastogi, Atul Saroop, Srujana Merugu, “Recommending Product Sizes to Customers”, *RecSys ’17: Proceedings of the Eleventh ACM Conference on Recommender Systems*, 243-250, 2017.

- [11] Mingxing Tan, Quoc Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks”, *Proceedings of the 36th International Conference on Machine Learning* (97), 6105-6114, 2019.
- [12] Wang-Cheng Kang, Eric Kim, Jure Leskovec, Charles Rosenberg, Julian McAuley, “Complete the Look: Scene-Based Complementary Product Recommendation”, *Computer Science 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1-10, 2019.
- [13] Yuncheng Li, LiangLiang Cao, Jiang Zhu, Jiebo Luo, “Fashion Outfit Composition Using An End-to-End Deep Learning Approach on Set Data”, *IEEE Transactions on Multimedia*, **19** (8), 1946-1955, 2017.
- [14] Wang-Cheng Kang, Chen Fang, Zhaowen Wang, Julia McAuley, “Visually-Aware Fashion Recommendation and Design with Generative Image Models.”, *In Proceedings of IEEE International Conference on Data Mining(ICDM’17)*, 1-10, 2017.
- [15] Alex Krizhevsky, Ilya Sutskever, Geoffrey E.Hinton, “Imagenet classification with deep convolutional neural net-works.”, *NIPS’12: Proceedings of the 25th International Conference on Neural Information Processing Systems*, 1097-1105, 2012.
- [16] Han Xiao, Kashif Rasul, Roland Vollgraf, “Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms”, *Mathematics and Computer Science’17* (8), 1-6, 2017.
- [17] Leland McInnes, John Hearly, James Melville, “UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction”, *The journal of Manifold Approximation and Projection*, **3** (29), 1-63, 2018.
- [18] Jeffrey Pennington, Richard Socher, Christopher Manning, “GloVe: Global Vectors for Word Representation”, *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532-1543, 2014.
- [19] Matthew Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, Luke Zettlemoyer, “Deep Contextualized Word Representations”, *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, **1**, 2227-2237, 2018.
- [20] Julian McAuley-UCSD CSE, 『Tradesy data』, <http://jmcauley.ucsd.edu/data/tradesy/>, (参照 2020.09.28).