DEIM Forum 2021 F33-3

商品購入履歴中の異カテゴリ商品対を用いた機械学習によるクロスカテゴリ推薦

右原　将吾†　　北山　大輔†

† 工学院大学情報学部システム数理学科　〒 163-8677 東京都新宿区西新宿 1-24-2
E-mail: †j317041@ns.kogakuin.ac.jp, ††kitayama@cc.kogakuin.ac.jp

あらまし　近年，ユーザに関する情報を獲得するドメインと，推薦対象のアイテムが属するドメインが異なる環境の情報推薦であるクロスドメイン推薦が研究されている．従来のクロスドメイン推薦の研究では，一つのドメインのコンテンツに対して，他の一つのドメインのコンテンツを推薦する物や，コンテンツの特徴量としてレビューを用いているものは少ない．そこで我々は，コンテンツの特徴量としてレビューを用い，一つのドメインのコンテンツに対し複数のドメインのコンテンツを推薦する手法を提案する．構築された予測モデルの精度をもとにどのような学習を行うのが効果的かの検証を行った．結果として，効果的な条件として，学習データのバリエーションを増やす，中間層のノード数を 25 から 50 の間に設定する，同一商品をひとまとめにしてレビュー数を増やすような仕組みが有効であることが判明した．

キーワード　商品推薦，機械学習，クロスドメイン推薦

1　はじめに

近年，楽天市場 [1]などのオンラインショッピングサイトの需要の増加により，多くの商品が出品され，様々なユーザに購入されている．ユーザに商品を推薦する際に，オンラインショッピングサイトでは協調フィルタリング，コンテンツベースフィルタリングというユーザ間や商品間の類似度を測り推薦する手法が一般的である．しかし，これらの手法では一般に人気な商品が推薦されてしまう問題や，類似商品ばかり推薦されてしまう問題がある．

そこで我々は，購入された商品対の特徴を学習することである特徴を持つ商品とともに買われやすい商品特徴を予測する手法を提案する．同時にカテゴリ対についても学習することで，入力カテゴリと出力カテゴリの組で推薦商品を決定できるクロスカテゴリ推薦を実現する．

ある特徴を持つ商品とともに買われやすい商品特徴を例を用いて説明する．例えば,「コク」という特徴を持つ「お酒」と「映える」という特徴を持つ「グラス」が多数の人の購入履歴に共に出現するとする．このとき,「コク」と「映える」というのは，共に買われやすい特徴であると考えられる．この特徴を利用することで，商品推薦を行う．具体的には，ユーザ A が「酒」の「魔王 芋焼酎」を購入した際に推薦してほしいカテゴリを「食器」を選択した場合に推薦商品として食器のグラスである「有田焼 陶器焼酎グラス・ロックグラス」を推薦するというものである．この時，購入商品には「コク」，推薦商品には「映える」という特徴があり，予測モデルはこの二つの特徴は共に買われやすいと学習されており推薦商品として出力する．本稿では，レビューがユーザから見た商品特徴を表していると考え，レビューの分散表現をもとに商品特徴を表現する．

本研究の貢献は以下のとおりである．

- クロスカテゴリ推薦に関して，ニューラルネットワークを用いた予測モデルによるアプローチを提案する．
- レビューから仮想的な購買履歴を作成し，それを用いた推薦結果を示し，提案手法の性質を明らかにする．

本稿の構成は以下の通りである．まず，2 節では，本稿の関連研究について述べる．3 節では，本研究の概要について述べる．4 節では，特徴ベクトルの生成と予測モデルについて述べる．5 節では，実験について述べる．6 節では，システムの出力例について述べる．7 節では，本稿のまとめと今後の課題について述べる．

2　関連研究

富士谷ら [1] は，コンテンツの多様性を持つドメインとして，テレビ番組を対象に放送ごとの番組に適した書籍推薦を行った．テレビ番組の内容は多様であるため，有効な特徴量も一様では無い．冨士谷らは多様性を考慮して，TF-IDF と LDA の特徴量を併用し，クロスドメイン推薦を行った．

石塚ら [2] は，図書館の書籍貸出システムの貸出履歴から各ユーザの嗜好の傾向を分析し，ユーザが興味を示しそうな最新情報が掲載された Web コンテンツを推薦する手法を提案している．ユーザが読んでいる書籍の題材として，大学図書館の貸出システムから学部・学科ごとに読まれている書籍の傾向を読み取り，嗜好を Wikipedia のトピックモデルを用いて分類・推定している．

中辻ら [3] は，既存の協調フィルタリングでは，被推薦ユーザが評価した事の無いドメインのアイテムを推薦するのは困難であるとし，評価した事の無いドメインのアイテムを MSE 良

1：https://www.rakuten.co.jp/

く推薦可能とする手法を提案している．富士谷らの手法を用いる事で複数のドメインのアイテムを持つプロバイダは，ドメインを跨る推薦を実現でき，ドメインを跨るユーザの回遊や購買を促進できるとしている．

鈴木ら [4] は，レビュー文に現れる顧客特有の属性と各ドメインからその人の嗜好との関係を捉えた消費行動モデルを構築を目的として，AmazonReview データにおいてクロスドメイン推薦モデルを構築し，ドメイン間のユーザの嗜好に一貫性があることを明らかにした．

荒澤ら [5] は，社会的に影響を与えている人物のようなユーザに共通したインフルエンサのだけでなく，身近な友人のような個々人のインフルエンサを推定する技術は，より高度にパーソナライズされた情報推薦システムへの発展につながるとした．そこで他者に影響されているであろうことがうかがえる SNS 上の反応や関心を分析することでユーザごとに影響を受けている人物を推定する手法を複数提案しそれらの推定性能に関する比較実験を行った．

中本ら [6] は，2013 年以降，コンテンツ産業内部アニメ業界にて売上増加が続いており，コンテンツ産業内の活気が他の産業に向ける事ができれば産業全体の市場活性化が期待できると考えた．日本コンテンツ産業の消費者に，従来の推薦よりも消費を促すことを研究の目的とした．そこで映画・漫画・アニメ・小説・楽曲・ドラマ・ゲーム作品のクロスドメイン推薦を実現し評価するシステムの構築を行った．

吉井ら [7] は，従来手法として，因子分解により要素間の相互作用を考慮して回帰予測を行うことのできる手法である FactorizationMachines(FM) という手法があり，この手法の改良を行った．

これらの研究は，特定のドメインのコンテンツを対象とした推薦であるのに対し，本研究は，汎用的なクロスドメイン推薦を目指しており，そのステップとしてカテゴリをドメインと見なしてモデル化している点で異なる．

3　研究の概要

提案手法では，クロスカテゴリ推薦を実現するために，購入商品対の特徴を入力，出力とした学習器で学習して，予測器を構築する．例えば，あるユーザが商品 A，B，C，D を購入していた場合，入力を A，出力を B として学習，入力を A，出力を C として学習というように，ある商品とともに購入される商品の関係を学習させる．このとき，同時にカテゴリの関係も学習させるため，商品 A のカテゴリである C_A と商品 B のカテゴリである C_B も入力として用いる．出力商品のカテゴリも入力として用いるのは，推薦時に「閲覧商品」と「任意のカテゴリ」を入力として推薦を行うためである．

提案手法による推薦の流れについて説明する．まず，ユーザは商品を購入する際に推薦してほしいカテゴリを選択する．選択した商品とカテゴリは後述する方法でベクトルの生成を行う．そして，予測モデルに入力として「購入商品カテゴリベクトル」,「購入商品特徴ベクトル」と「推薦商品カテゴリベクトル」が入る．予測ベクトルと類似度が高い商品を推薦商品として推薦する．本研究が提案するシステムの構成を図 1 に示す．

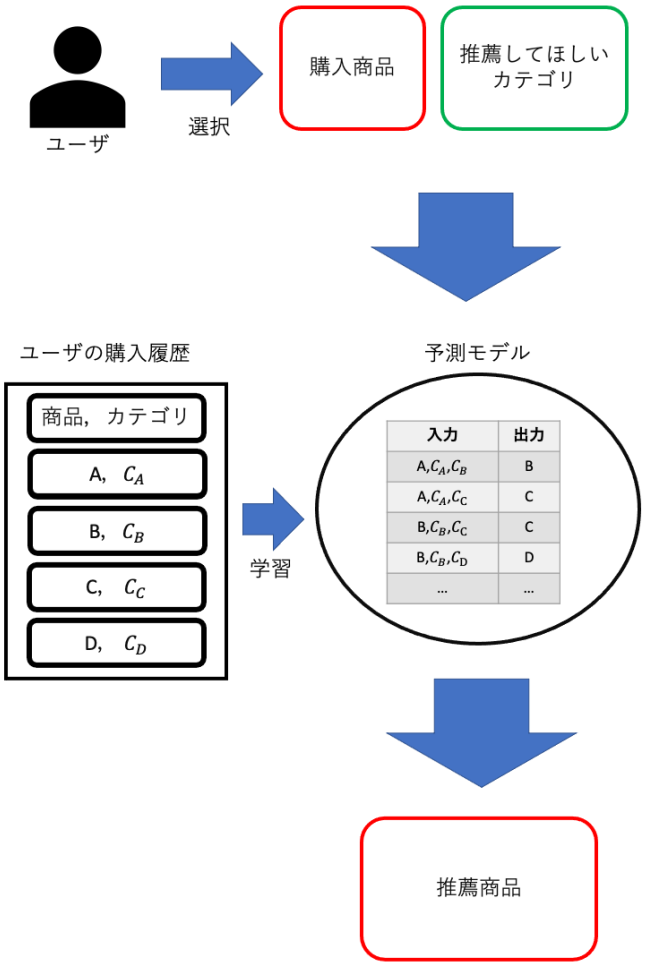

図 1　システムの概要

4　クロスカテゴリ商品予測モデル

ここでは商品特徴ベクトルの生成法を述べた後，予測モデルの生成について述べる．

4.1　商品特徴ベクトルの生成

ここでは商品ベクトルの生成について説明する．ベクトル生成には，Simple Word-Embedding-based Methods(SWEM) という手法を用いる．この手法は Shen ら [8] が提案した手法であり，単語の各次元の最大値や平均値を文書ベクトルとして採用する手法である．手法はキーワードベクトルに加えて文章ベクトルを得るためのニューラルネットワーク自体を，大規模コーパスから学習させる必要があるが，SWEM は学習パラメータを必要とせず，計算コストも低い点で優れている．

次に，商品につけられているレビューを使ってベクトルの生成を行う．レビューを形態素解析器 MeCab [9] を用いて，分かち書きし，Word2Vec [10] で単語ベクトルを算出する．まず，SWEM を用いて，レビューごとの文書ベクトルを算出する．次に同一商品のレビューベクトルの平均を商品の特徴ベクトルとして用いる． 本稿での SWEM は平均を用いる．

4.2 予測モデルの生成

本稿での予測モデルは入力層，中間層，出力層の 3 層からなるニューラルネットワークを用いる．入力層には入力商品カテゴリベクトル，入力商品ベクトル，出力商品カテゴリベクトルを連結したものが入り，出力には出力商品ベクトルが入る．各層の活性化関数は，中間層で relu，出力層で tanh を用いる．

学習データとしては，商品の購入履歴を用い．入力商品，出力商品は同ユーザの購入商品対を用いる．カテゴリベクトルは One-hot ベクトルであり，商品ベクトルは 4.1 節で生成したものを用いる．

学習データの作成例を表 1 に示す．まず，購入履歴中の商品を $[i_n{:}C_m]$ で表記し，i_n は商品番号，C_m はそのカテゴリとする．ここで仮に，$[i_1{:}C_1]$，$[i_2{:}C_1]$，$[i_3{:}C_2]$，$[i_4{:}C_3]$ という 4 つの購入履歴があるとする．4 つのペアから 2 つ取り出し，$[i_1, C_1, C_2]$ と $[i_3]$ のように 2 つに分ける．ここで，$[i_1, C_1, C_2]$ をそれぞれ [商品，購入商品カテゴリ，推薦カテゴリ] として入力し，$[i_3]$ は [推薦商品] を推薦カテゴリ C_2 と商品対になっている商品 i_3 を出力として学習させる．この予測モデルを用いて入力商品を推薦カテゴリより出力された商品ベクトルとコサイン類似度が高い商品ベクトルを持つ商品を推薦商品とする．

5 実　　験

5.1 実験用データと前処理

本稿では楽天市場が提供している「楽天市場：商品情報，みんなのレビュー・口コミ情報」の「商品レビュー」を使用している．「商品レビュー」は 2018 年，2019 年のものを使用する．商品に付けられているジャンルをカテゴリとする．購入履歴のデータがないため，商品レビューを購入履歴として扱う．各データの項目を表 2 にまとめる．次に，同一商品の集約処理について述べる．学習の際に商品単位での特徴ベクトルが必要になる．本稿では，単に同一 ID の商品を同一商品とみなして商品ベクトルを生成した．しかし，楽天市場のデータは同一商品であっても店舗ごとに商品 ID が異なる．そのため，ID の異なる同じ商品をまとめ，一商品あたりのレビュー数を増やすことが商品ベクトル生成に効果的と考えられるが，今後の課題とする．

5.2 実 験 方 法

楽天市場の商品レビューより購入履歴が 100 件以上あるユーザを抽出し，対象とする．上記のユーザより仮想的な購入履歴を作成し，予測モデルの構築を行う．構築されたモデルの精度をもとにどのような学習を行うのが効果的なのかの検証を行う精度とは学習の際に，評価データにおける入力に対して正解の出力ベクトルと予測ベクトルの平均二乗誤差 (MSE) を指す．この値が小さい方が良いモデルと言える．以下の項目に関して条件を変えて比較を行う．

- 学習データ
- 中間層のノード数
- 用いる商品

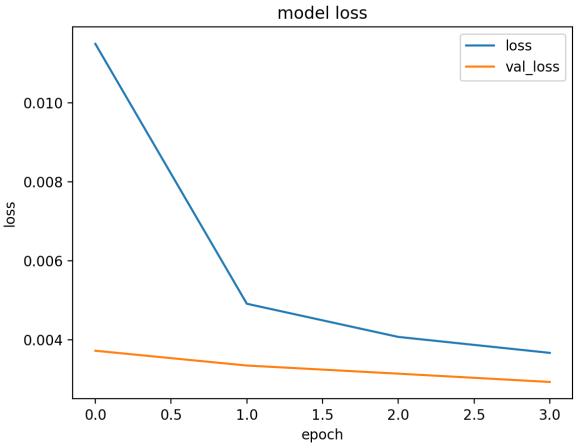

図 2　直近 1 件までとした場合

各項目の条件についてまとめたものを表 3 に示す．学習データに関しては，ある商品の直近 1 件の商品を組み合わせる場合と直近 5 件の商品を組み合わせる場合で MSE の比較を行う．この際，中間層のノード数は 50，用いる商品はレビュー数が 10 件を超える商品に限定し行う．中間層のノード数に関しては，予測モデルの中間層のノード数を 10，25，50，125 に変更し，MSE の比較を行う．この際，学習データの作成方法は直近 5 件の商品の組み合わせ，用いる商品はレビュー数が 10 件を超える商品に限定し行う．用いる商品に関しては，同一商品集約の際にレビュー数が 10 件を超える商品に限定した場合とレビュー数で絞らない場合で MSE の比較を行う．この際，学習データの作成方法は直近 5 件の商品の組み合わせ，中間層のノード数を 50 に設定し行う予測モデルの入力側の次元としては，入力商品カテゴリベクトルと出力商品カテゴリベクトルの 78 次元，入力商品ベクトルが 300 次元の合計 378 次元である．

5.3 実験結果と考察

三つの項目に関しての結果と考察を述べる

5.3.1 学習データ

直近 1 件までにした際の MSE の変化を図 2，直近 5 件までにした際の MSE の変化を図 3 に示す．直近 1 件までとした場合の MSE は 0.0033，直近 5 件までとした場合の MSE は 0.0026 と，直近 5 件までとした場合の方が MSE が良いことが確認できた．直近 5 件までとした時に増えるデータは入力ベクトルのカテゴリベクトル部分が異なるだけの類似したものである．このような類似する入力ベクトルであっても異なる出力ベクトルであれば，MSE の向上が見られた．この結果より，学習データのバリエーションを増やした方が効果的であると考えられる．

5.3.2 中間層のノード数

ノード数を 10，25，50，125 にした際の MSE の変化をそれぞれ図 4，5，6，7 に示す．ノード数 10 の MSE は 0.0027，ノード数 25 の MSE は 0.0026，ノード数 50 の MSE は 0.0026，ノード数 125 の MSE は 0.0027 であった．結果としてノード数は 25，50 で MSE が良くなることが確認できた．

表 1　学習に用いる購入履歴データの例

商品対	入力 (購入商品カテゴリ，商品，推薦カテゴリ)	出力 (推薦商品)
$[i_1{:}C_1]$，$[i_3{:}C_2]$	C_1，i_1，C_2	i_3
$[i_2{:}C_1]$，$[i_3{:}C_2]$	C_1，i_2，C_2	i_3
$[i_1{:}C_1]$，$[i_4{:}C_3]$	C_1，i_1，C_3	i_4
$[i_3{:}C_2]$，$[i_4{:}C_3]$	C_2，i_3，C_3	i_4
$[i_3{:}C_2]$，$[i_2{:}C_1]$	C_2，i_3，C_1	i_2
....	....	....

表 2　各データの項目

商品レビュー
投稿者 ID
店舗名
店舗 ID
商品名
商品 ID
商品ページ URL
商品ジャンル ID
商品ジャンル ID パス
使い道
目的
頻度
評価ポイント
レビュータイトル
レビュー内容
参考になった数
レビュー登録日時

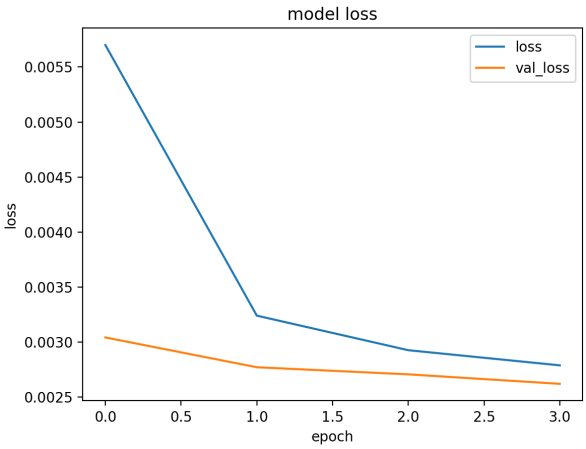

図 3　直近 5 件までとした場合

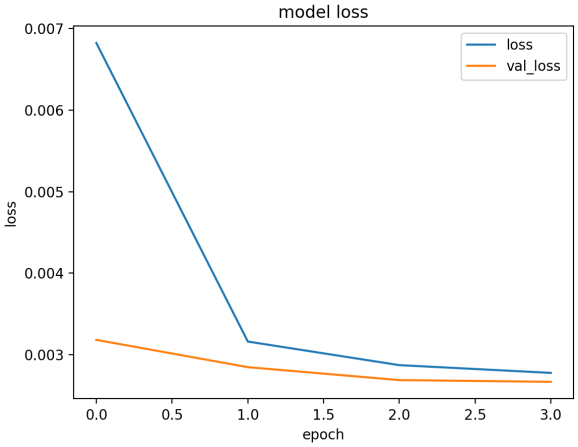

図 4　ノード数:10

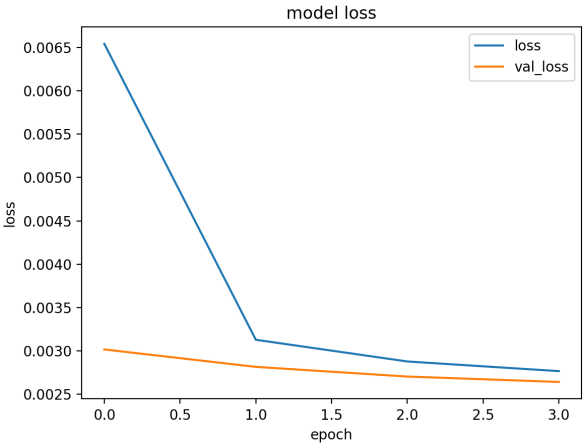

図 5　ノード数:25

5.3.3　商品ベクトル

レビュー数で商品を絞った際の MSE の変化を図 8，絞らなかった際の MSE の変化を図 9 に示す．レビュー数が 10 件を超える商品に絞った場合の MSE は 0.0026，絞らなかった場合は 0.0187 となっており，レビュー数で商品を絞った方が MSE が良くなっていることが確認できた．レビュー数で商品を絞ることで学習データ数が減るにもかかわらず，MSE が向上している．レビュー数が多くなることで，商品ベクトルの生成精度が良くなっていると考えられる．したがって，高精度なベクトルを作るために，同一商品をひとまとめにしてレビュー数を増やすような仕組みが有効であると考えられる．

6　推薦商品の出力の例

実際に購入商品と推薦カテゴリを予測器に入力し，推薦商品の出力を行った．出力結果をまとめたものを表 4 に示す．表 4 より，4 つ全ての商品に関して想定する推薦商品が出力できていないことがわかる．しかし，スノーボードの想定する推薦商品はスキーウェアであったがスキー場に行くまでの服装としたら適した推薦が出来ていると考えられる．ある特定の同一商品が他の入力商品においても推薦商品として出力される結果が多

表 3　各項目の条件

	学習データの作成法	中間層のノード数	用いる商品
学習データの比較	直近 1 件	50	レビュー数 11 以上
	直近 5 件	50	レビュー数 11 以上
中間層のノード数の比較	直近 5 件	10	レビュー数 11 以上
	直近 5 件	25	レビュー数 11 以上
	直近 5 件	50	レビュー数 11 以上
	直近 5 件	125	レビュー数 11 以上
用いる商品の比較	直近 5 件	50	レビュー数 11 以上
	直近 5 件	50	レビュー数 1 以上

表 4　推 薦 商 品

購入商品	推薦カテゴリ	想定する推薦商品	実際の推薦商品
ランドセル	バッグ・小物・ブランド雑貨	レインカバー	20 色のストールマフラー
マリンシューズ	レディースファッション	海水浴に適した服	ニットワンピース 長袖
ダイエット サプリメント	食品	健康食品	淡路島玉ねぎ 3 キロ
スノーボード	レディースファッション	スキーウェア	ニットワンピース 長袖

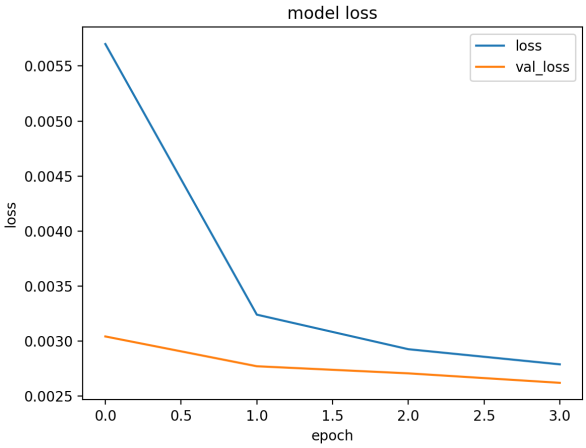

図 6　ノード数:50

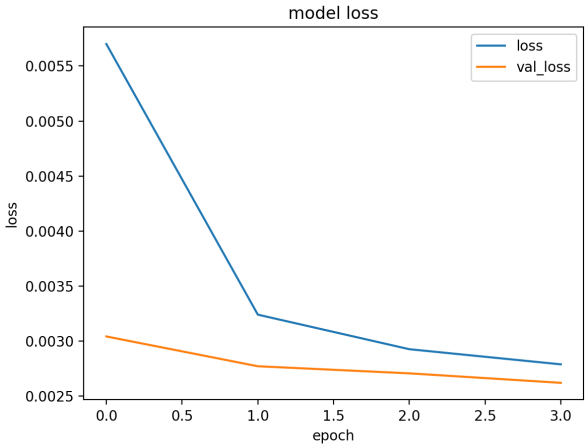

図 8　レビュー数制限:有り

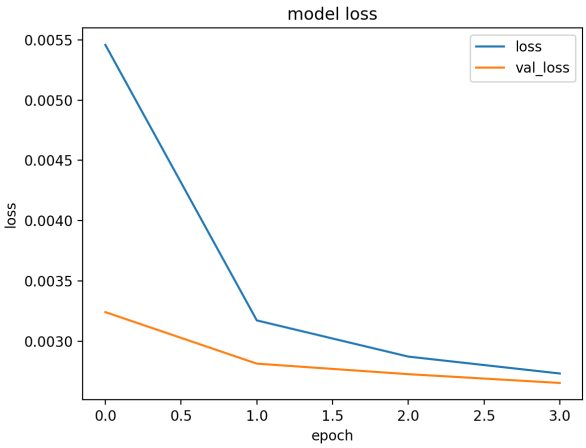

図 7　ノード数:125

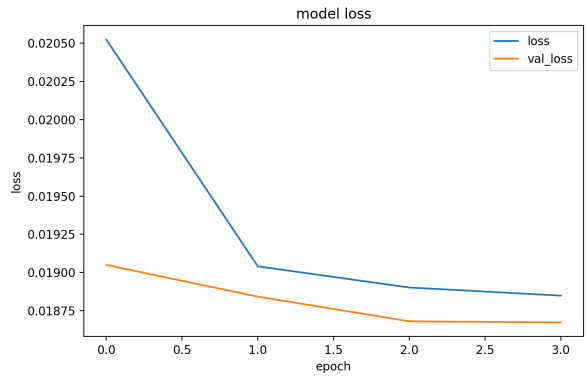

図 9　レビュー数制限:無し

く見られた．このことから，予測ベクトルが類似する傾向にあり，推薦システムとして多様性を担保する仕組みが必要と考えられる．

7　おわりに

従来の EC サイトでは協調フィルタリングやコンテンツベースといったユーザ間やアイテム間の類似度を測る手法が一般的となっている．しかしこのような手法だと他ユーザとの嗜好に

差があった場合興味のない商品が推薦されることがあるという問題があった．それに対し本稿では，購入履歴中から一緒に買われやすい商品対の特徴を学習器で学習し，予測モデルを構築することで，クロスカテゴリの推薦を行う手法を提案した．商品特徴は商品につけられたレビューを MeCab と SWEM を用いて表現し，商品のカテゴリは ohe-hot ベクトルを用いて表現した．また，既存のクロスドメイン推薦では，1 つのドメインのコンテンツに対し 1 つのドメインのコンテンツの推薦を扱うことが多いが，本手法により，カテゴリやドメインの組み合わせを問わず，汎用的なクロスドメイン推薦が可能になる．実験により効果的な学習を行う条件として，学習データに関してはバリエーションを増やす，中間層のノード数を 25 から 50 の間にする，用いる商品は同一商品をひとまとめにしてレビュー数を増やすことが有効である．今後の課題として，学習じのエポック数の検討，学習データの作成方法の再検討，中間層のノード数の再検討，商品ベクトルの生成方法の再検討，同一商品の集約があげられる．

謝　　辞

本研究では，国立情報学研究所の IDR データセット提供サービスにより楽天株式会社から提供を受けた「楽天データセット」を利用しました．また，本研究の一部は，2020 年度科研費基盤研究 (C)(課題番号：18K11551) によるものです．ここに記して謝意を表すものとします．

文　　献

[1] 富士谷康, 村尾和哉, 望月祐洋, 西尾信彦. コンテンツの多様性を考慮したクロスドメイン推薦. 情報処理学会論文誌, Vol. 57, No. 10, pp. 2210–2221, oct 2016.

[2] 石塚大貴, 中沢実. トピックモデルによる書籍から web コンテンツのクロスドメイン推薦方式の実装. 研究報告マルチメディア通信と分散処理（DPS）, Vol. 2017-DPS-169, No. 12, pp. 2210–2221, jan.

[3] 中辻真, 藤原靖宏, 内山俊郎. ユーザグラフ上のランダムウォークに基づくクロスドメイン推薦. 人工知能学会論文誌, Vol. 27, No. 5, pp. 296–307, 2012.

[4] 鈴木凱亜, 大知正直, 坂田一郎. Ec サイトのレビュー文を用いた商品ドメイン間のユーザの購買傾向の相関と背景因子の分析. 研究報告ドキュメントコミュニケーション, Vol. 2019-DC-114, No. 9, sep 2019.

[5] 荒澤孔明, 服部峻. Sns における反応と関心に基づくインフルエンサ推定の個人化. 情報処理学会論文誌データベース（TOD）, Vol. 13, No. 2, pp. 1–18, apr 2020.

[6] 中本昌吾, 宮治裕. 自然言語処理を用いたコンテンツ業界作品のクロスドメイン推薦. 第 81 回全国大会講演論文集, 第 2019 巻, pp. 441–442, feb 2019.

[7] 吉井和輝, 立間淳司, 青野雅樹. クロスドメイン推薦に向けたユーザ嗜好の予測手法の提案. 第 77 回全国大会講演論文集, 第 2015 巻, pp. 593–594, mar.

[8] Dinghan Shen, Guoyin Wang, Wenlin Wang, Martin Renqiang Min, Qinliang Su, Yizhe Zhang, Chunyuan Li, Ricardo Henao, and Lawrence Carin. Baseline needs more love: On simple word-embedding-based models and associated pooling mechanisms. *CoRR*, Vol. abs/1506.06726, pp. 1–13, 2018.

[9] 工藤拓, 山本薫, 松本裕治. Conditional random fields を用いた日本語形態素解析. 情報処理学会研究報告, Vol. 2004, No. 47, pp. 89–96, may 2004.

[10] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *ICLR*, Vol. arXiv:1301.3781, pp. 1–12, sep 2013.