

商品レビューからの自己符号化器を用いた 情報補完による評価値推定精度の向上

Improvement of rating prediction accuracy by information supplementation using auto-encoder from product review

林 知範^{1*} 高木 友博²
Tomonori Hayashi¹ Tomohiro Takagi²

¹ 明治大学理工学研究科基礎理工学専攻
¹ Department of Computer Science, Meiji University

Abstract: Recent studies have shown that incorporating context data such as text, time, and location in recommender systems improves the accuracy of the recommendation. In this paper, we focus reviews of item that user writes for considering user's preference of product category. We aim to improve the prediction accuracy of ratings by learning the ratings and reviews of item that user writes.

1 はじめに

近年、インターネットやモバイルデバイスの普及に伴い、人々は日常的に多くの情報を発信することが可能となった。その結果、情報量は膨大となりユーザーにとって何が重要な情報なのかを判断するのが困難となった。そこでユーザーの過去のデータを分析することでパーソナライズし、適した情報を提供することが可能な「情報推薦システム」が注目されている。

情報推薦システムの1つに e-commerce などで行われる「商品推薦システム」がある。ユーザーの購入履歴を分析することで、ユーザーの嗜好にあった商品を推薦するものである。この商品推薦システムの問題点の1つとして、データ数が少ない商品に関して十分な学習が行えないう最適推薦ができないという問題がある。この問題は、現実的に起こりやすく解決すべき大きな課題とされている。

本研究は、e-commerce で利用される商品推薦システムにおいて、データのスパース性による精度低下に対し、商品レビューから自己符号化器を用いて抽出した情報で評価回数が少ない商品情報を補完することにより、商品の評価予測精度を向上させることを目的とする。

2 関連研究

情報推薦システムに広く使われている推薦手法に協調フィルタリングがある。ユーザーが商品につけた評価値から直接ユーザー間または商品間の比較を行うメモリベース法や評価値を利用して推薦を行うためのモデルを学習するモデルベース法などがある。中でも、ユーザーと商品の情報から評価値行列を生成し、元の評価値行列を複数の低次元の行列に分解することでユーザーや商品が持つ潜在因子を利用するモデルが高い性能を発揮することが知られている。これらは行列因子分解モデルと呼ばれる。

2.1 行列因子分解適用での課題

行列因子分解は、上記の通り商品推薦システムにおいて広く利用されている手法である。本研究でも行列因子分解をベースに提案手法を考えている。ユーザーが購入した商品につける評価値で行列を構成し、評価値を元にユーザーの特徴と商品の特徴を反映した2つの潜在因子行列を生成する。ユーザーを表すベクトルを u_i 、商品を表すベクトルを v_j として (1) 式で評価値を予測する。

$$\hat{r} = u_i^T v_j \quad (1)$$

さらに (2) 式の目的関数を最小化するようにユーザーベクトルと商品ベクトルを更新する。

* 連絡先： 林知範, 明治大学理工学研究科基礎理工学専攻,
〒 214-8751 神奈川県川崎市多摩区東三田 1-1-1,
E-mail: t-hayashi@cs.meiji.ac.jp

$$L = \sum_i^N \sum_j^M (r_{ij} - \hat{r}_{ij})^2 + \lambda_u \sum_i^N \|u_i\|^2 + \lambda_j \sum_j^M \|v_j\|^2 \quad (2)$$

(2) 式の目的関数は、実際の評価値 r とユーザー行列とアイテム行列によって推定した予測評価値 \hat{r}_{ij} の差の損失関数と正則化項によって構成される。

この目的関数からわかるように商品推薦システムにおける行列因子分解の商品ベクトルの更新は、学習期間での商品毎の評価回数に依存することがわかる。つまり、評価回数が多い商品は多くのユーザーの好みを反映することができるが、評価回数が少ない商品は特定のユーザーの好みしか反映できていないといった学習が不十分な状態であることが考えられる。

そこで、学習期間における評価回数が少ない状況下でもよりユーザーの好みを反映させるために、ユーザーが書いた商品レビューからユーザーの好みを抽出し、学習が不十分な商品の情報補完に適用することで評価値予測の精度の向上に取り組む。また、ただ好みを抽出するのではなく、ユーザーが書いた商品レビューを商品カテゴリで分類し好みを抽出することで、ユーザーの商品カテゴリへの好みを反映させたベクトルを生成する。

2.2 最新の動向

商品推薦における評価値予測問題に商品レビューを利用するアプローチは、今までにいくつかの研究で取られている。[Ganu 2009] や [岡田 2012] らは商品レビューの極性情報や tf-idf といった商品レビューの表層語に注目し、生成した特徴量を推薦手法に適用している。また、[北原 2016] らは、商品レビューに出現する表層語に注目する既存の手法に対して、商品レビューの表層に出現しない潜在的な意味に注目し、ハイブリッド型の協調フィルタリングに組み込むことで予測精度の向上を実現している。

3 提案手法

本研究では、ユーザーが書いた商品レビューから商品カテゴリへの好みといった特徴量を抽出し、その抽出した情報によって商品の情報を補完する手法を提案する。抽出した特徴量で情報が少ない商品に対して補完を行うことで、ユーザーが商品に対してつける評価値の予測精度を向上させる。

3.1 特徴量の抽出

抽出する特徴量は、ユーザーの各商品カテゴリへの潜在的な好みである。ユーザーが書いた商品レビューをカテゴリごとにまとめ、Denoising Auto-Encoder(DAE)によって抽出をする。DAE は中間層が一層のニューラルネットワークであり、以下のように定式化される

$$\tilde{x} = C(\tilde{x}|x) \quad (3)$$

$$h = f(W\tilde{x} + b) \quad (4)$$

$$y = f(\tilde{W}h + \tilde{b}) \quad (5)$$

x は入力ベクトル、 $C(\cdot|\cdot)$ は入力に対してある確率分布に従ってノイズを加える関数である。 x に対してノイズを加えた \tilde{x} を入力として、重み行列 W, \tilde{W} とバイアスベクトル b, \tilde{b} 、活性化関数 $f(\cdot)$ で構成されるニューラルネットワークとする。

入力ベクトルは、ユーザーごとに商品カテゴリでまとめたレビューに出現する単語で構成されている。使用する単語は、学習期間中の商品レビューに出現する頻度上位 1 万語の名詞と形容詞を採用した。

ここで生成した特徴量ベクトルは、各ユーザーの各商品カテゴリへの潜在的な好みを商品レビューの単語情報から抽出したものとなっている。

3.2 ユーザーの商品カテゴリへの嗜好の反映

上記で抽出した特徴量を行列因子分解に組み込むことで、評価値予測の精度の向上を実現する。

一般的な行列因子分解による評価値予測では、学習期間での評価回数によって学習後の商品ベクトルの状態が変化すると考えられる。評価回数が多ければ、その分だけ学習後のベクトルは様々なユーザーの好みを反映できていると考えられる。しかし、学習期間中の評価回数が極端に少ない商品の場合には、学習後のベクトルは特定のユーザーの好みを反映させたベクトルというような学習が不十分な可能性がある。

そこで、評価値を予測する際に商品レビューから抽出した特徴量を組み込むことで、ユーザーの商品カテゴリへの好みを反映させ、学習が不十分な商品の情報を補完する。本提案手法は (6) 式の目的関数を最小化する。

$$L = \sum_i^N \sum_j^M (r_{ij} - u_i^T (v_j + category_{ic}))^2 + \lambda_u \sum_i^N \|u_i\|^2 + \lambda_j \sum_j^M \|v_j\|^2 + \lambda_k \sum_i^N \sum_c^C \|(category_{uc})\|^2 \quad (6)$$

表 1 学習期間のデータ量の詳細値

| 学習比率 | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% |
|---------|---------|--------|---------|---------|---------|---------|---------|---------|---------|
| 総学習データ量 | 15594 | 32225 | 48487 | 64815 | 81246 | 97328 | 113480 | 129640 | 134586 |
| スパース率 | 99.998% | 99.97% | 99.962% | 99.949% | 99.936% | 99.923% | 99.911% | 99.985% | 99.895% |

(2) 式の商品ベクトル v_j に各ユーザーの各商品カテゴリへの潜在的な好みを抽出した特徴量 $category_{uc}$ を加えて最適化を行う。

$u_i, v_j, category_{ic}$ の更新は以下により行う。

$$e = -(r_{ij} - u_i^T(v_j + category_{ic})) \quad (7)$$

$$u_i = u_i - lr(e(v_j + category_{ic}) + \alpha u_i) \quad (8)$$

$$v_j = v_j - lr(eu_i + \alpha(v_j + category_{ic})) \quad (9)$$

$$category_{ic} = category_{ic} - lr(eu_i + \alpha(v_j + category_{ic})) \quad (10)$$

(7) 式では実際の評価値と商品カテゴリへの好みを反映させた予測評価値との誤差を求めている。また、(8)~(10) 式によってユーザーベクトル u_i , 商品ベクトル v_j , 各ユーザーのカテゴリごとの好みを反映させたカテゴリベクトル $category_{ic}$ を (7) 式で算出した誤差を用いて更新している。

4 実験

4.1 実験環境

実験では、楽天データセットの楽天市場を用いる。本データセットでは、2011年4月から2014年3月の楽天市場での購入履歴を利用している。本データセットを図1のように学習量を変化させることでデータのスパース性に対して、どのような有用性があるかを確認する。実験では、テスト期間にユーザーが購入した商品の評価値を予測し RMSE(平均二乗誤差) によって評価する。

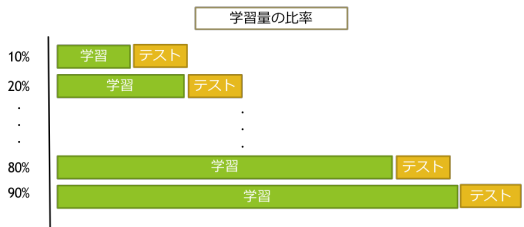


図 1 データセットの構成

4.2 各種パラメーター

本実験を行う前の予備実験により各種パラメーターを決定する。本提案手法のベースとなる行列因子分解

の次元数に関して考察を行った。本実験同様、学習量を10%から90%に変化させて各 RMSE の平均によって次元数を決定した。

表 2 各次元数での RMSE による評価

| 学習比率 | 10% | 20% | ... | 80% | 90% | 平均 |
|---------|------|------|-----|------|------|------|
| 次元数 50 | 1.39 | 1.36 | ... | 1.2 | 1.19 | 1.28 |
| 次元数 100 | 1.41 | 1.36 | ... | 1.19 | 1.23 | 1.3 |
| 次元数 200 | 1.42 | 1.38 | ... | 1.26 | 1.25 | 1.35 |

表 2 の結果より RMSE が最低となる分解次元数は 50 付近だということがわかる。

また、商品レビューより各ユーザーの各カテゴリの潜在的な好みを抽出に用いる DAE のパラメーターもこの分解次元数に合わせる必要があるため、中間層を 50 次元として入力層に dropout 関数を導入している。

4.3 RMSE による評価

評価値予測の精度における評価指標は、RMSE を用いる。テスト期間における評価総数を L 、予測対象の評価値を r_{ij} とし、予測評価値を \hat{r}_{ij} とすると、RMSE は (11) 式で定義される。

$$RMSE = \sqrt{\frac{1}{L} \sum_i^N \sum_j^M (r_{ij} - \hat{r}_{ij})^2} \quad (11)$$

4.4 対象手法

本実験における対象手法は、以下 3 手法とした。

- ・ベースライン手法としての行列因子分解
 - ・商品の情報補完にユーザーが書いた商品レビューから各商品カテゴリの好みを抽出した特徴量を用いた本研究の提案手法
 - ・ユーザーが書いた商品レビューから各カテゴリではなく単一の特徴量を生成し商品の情報補完に用いる手法を比較手法
- ベースライン手法と比較することで商品レビューを情

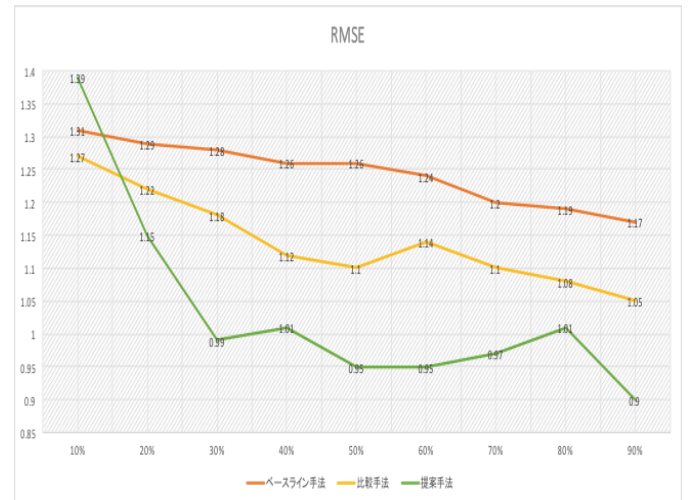
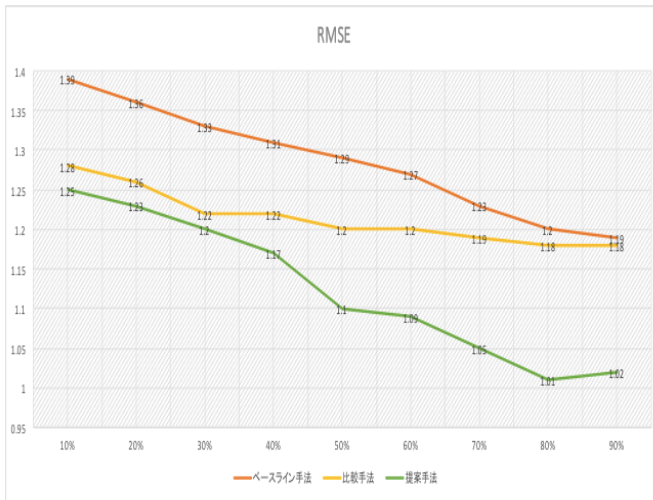


図2 (左) 全カテゴリにおける評価実験,(右) 単一カテゴリにおける評価実験

報補完に用いる有用性を示し、比較手法と比較することで商品カテゴリ毎に特徴量を生成する有用性を示す。

4.5 実験結果・考察

本実験では楽天市場の全カテゴリを用いた実験とその中の1つのカテゴリ(単一カテゴリ)を用いた実験の2つを行った。全カテゴリを用いた実験での総カテゴリ数は35を採用し、単一カテゴリでの実験では総カテゴリ数は16を採用した。それぞれの実験結果を図2に示す。両評価実験の結果から提案手法の以下2点を示すことができた。

- ・商品レビューを情報補完に用いる有用性
- ・商品カテゴリ毎に特徴量を生成する有用性

また、図2(左)では、RMSEが1.0を下回らなかったものの学習比率50%以降で大きな精度向上を実現することができた。これは評価回数が少ない商品が多く出現しているデータセットでも、提案手法により商品カテゴリへの潜在的な好みを情報補完に適用することで、そのような商品にも対応できていると考えられる。

また、図2(右)では学習比率10%ではベースライン手法と比較手法よりも精度が劣る結果となってしまった。これは情報補完として適用した特徴量がノイズになり精度の低下をまねいた可能性が考えられる。この場合の解決策として、商品の評価回数に応じて商品ベクトルとユーザーのカテゴリへの好みを反映した特徴量ベクトルの割合を変化させるようなパラメータを導入すること状況に応じて柔軟に対応出来ると考えられる。この点を今後の課題の1つとして取り組んでいきたい。

5 おわりに

本稿では、ユーザーが書いた商品レビューを商品カテゴリによって分類し、その分類した商品レビュー群からユーザーのカテゴリへの潜在的な好みを抽出し、商品情報の情報補完として適用することで評価値予測精度の向上を確認した。商品レビューからの商品カテゴリへの好みの抽出にはDenoising Auto-Encoderを適用し従来手法よりもより良い特徴量の抽出に成功した。また、商品カテゴリへの潜在的な好みを情報補完として用いることで、学習が不十分な商品に対してもより最適な評価値の予測を実現することができた。

参考文献

- [Ganu 2009] Ganu,G.,Elhadad,N.,and Marian,A.: Beyond the Stars Improving Rating Predictions using Review Text Content, WebDB 2009.
- [岡田 2012] 岡田瑞穂, 藤井敦: レビューテキスト間の類似度を用いた協調フィルタリング, NLP 2012.
- [北原 2016] 北原将平, Rafal Rzepka, 荒木健治: レビューテキストを対象としたハイブリッド型推薦手法におけるトピックモデルの有効性について, JSAI 2016.