

説明生成可能な推薦手法へのレビューの有用性データの利用方法

城戸 俊亮† 坂本 竜二† 有次 正義††

† 熊本大学大学院自然科学教育部 〒 860-8555 熊本県熊本市中央区黒髪 2 丁目 39 番 1 号

†† 熊本大学大学院先端科学研究部 〒 860-8555 熊本県熊本市中央区黒髪 2 丁目 39 番 1 号

E-mail: †kidos@st.cs.kumamoto-u.ac.jp, ††131t6802@gmail.com, ††aritsugi@cs.kumamoto-u.ac.jp

あらまし 説明付き推薦手法はアイテムだけでなく、なぜそのアイテムを推薦したのかなどの説明文も提示することができる推薦手法である。レビューの内容からユーザの嗜好を抽出することで説明生成を行う。そのため、レビューの質や量が推薦精度や説明生成に大きく影響し、レビューの質が低い、もしくは数が少ないユーザに対しては、誤った推薦や説明生成が起りやすいという問題点がある。本研究では、既存の説明付き推薦手法においてレビューの有用性に関するデータを考慮することで、アイテムの推薦精度と説明の質の向上を目指す。複数の既存手法に適用し、提案手法が既存手法に汎用的に適用可能であることを示す。

キーワード 推薦システム, 感情分析, レビューの有用性

1 はじめに

推薦システムの分野において、説明付き推薦手法が注目されている。協調フィルタリング [1] を用いた既存の推薦手法はアイテムを推薦するのみであったが、説明付き推薦手法は推薦に関する説明も提示することができる。推薦に関する説明とは、推薦システムがユーザに対して特定のアイテムを推薦する理由や推薦する経緯などである。推薦の過程で説明を生成することで、推薦システムは透明性を持ち、ユーザが誤りに気づくことができる (scrutability)、ユーザの信頼性を高める、効果的 (effectiveness) かつ効率的 (efficiency) な決定を下す、ユーザに購入を促す (persuasiveness)、ユーザの満足度 (satisfaction) を向上させるといった利点がある [2]。

既存の説明付き推薦手法はユーザが記載したレビュー文から、ユーザの嗜好を抽出することで説明生成を行う。これらの手法で扱うレビューデータには 3 つの問題点がある。1 点目はレビューの疎性である。ユーザのレビュー数が少ない場合、嗜好を正確に捉えることができず、誤った推薦や説明を生成する可能性がある。2 点目はレビューデータの質である。レビューの記載方法は基本的には自由形式のため、文章の長さや内容には個人差が生じる。そのため、質が低いレビューから嗜好の抽出をすることは困難である。3 点目はレビュー文がないデータである。ユーザのレビューにはアイテムの評価値のみでレビュー文が記載されていないデータもあり、レビュー文がないデータからは説明生成を行うことは難しい。

そこで、上記の問題を改善するためにレビューの有用性データを考慮する。レビューの有用性データとはユーザが他人のレビューを参照したときにどの程度役に立ったのかを評価したデータである。ユーザはアイテムに関して注目している情報を得るために他人のレビューを判断材料にする。そのため、レビューの有用性データからもユーザの嗜好を抽出できると考える。本研究ではレビューの有用性データを既存の説

明付き推薦モデルに適用することで、推薦精度と説明の質の向上を目指す。

本論文は以下のように構成する。まず第 2 章では本研究と関連のある既存研究について述べる。第 3 章では既存手法のモデルとレビューの有用性データを考慮したモデルを解説する。第 4 章では提案手法の推薦と説明生成における有効性を評価するための実験内容と実験結果について述べる。最後に第 5 章では本論文のまとめと将来展望について述べる。

2 関連研究

従来の推薦システムはユーザがどのアイテムを購入するか予測するための研究がなされてきた。中でも協調フィルタリングを用いる手法は高い推薦精度を出している。アイテムの評価値から潜在要因を学習することで、評価値の欠損値を予測し、ユーザに推薦するアイテムのリストを作成する。しかし、潜在要因を学習するだけではアイテムの特定の特徴に基づいて推薦することは困難である。さらに、ユーザに推薦するアイテムに対する説明を生成することも難しい。

近年では推薦の説明生成についても注目されている。推薦の説明を生成することは、推薦の透明性を向上させ、ユーザが推薦システムを評価する上で重要な役割を果たす。[3] はユーザが意思決定をする際に注目するアイテムの特徴と特徴に対する感情極性を MF に組み込んだ Explicit Factor Model (EFM) を提案した。レビューからユーザの嗜好を表す特徴とその特徴に対するユーザの感情極性を抽出する [4]。‘The actor is good’ の場合は、抽出される特徴は ‘actor’ であり、感情極性は ‘+1’ となる。ユーザが嗜好する特徴を表現した行列とアイテムの特徴に対する品質を表現した行列と評価行列を同時に行列分解することで、高い推薦精度を維持しつつ、説明生成を行う。[5] は EFM の式に BPR [6] を導入した Learning to Rank user Preferences based on Phrase-level sentiment analysis across Multiple categories (LRPPM) を提案した。LRPPM はユーザ

が嗜好する特徴の行列を計算する際に、BPRによってレビューに記載した特徴と記載していない特徴の嗜好度の差でモデルを学習することで、EFMと比べて推薦精度を向上させている.[7]はアイテムの特徴と感情極性のほかに、ユーザの意見を組み込んだテンソルを構築するMulti-Task Explainable Recommendation(MTER)を提案した. ‘The actor is good’の場合、アイテムの特徴‘actor’に対するユーザの意見として‘good’が抽出される. ユーザが特定のアイテムの特徴に対する嗜好度を表現したテンソル、ユーザが目目する各特徴に対する意見との関連度を表したテンソル、アイテムの各特徴に対する意見との関連度を表したテンソルを構築し、同時にテンソル分解[8][9]をすることで推薦とユーザがよく使う意見をういた説明を生成することができる. ユーザ調査による実験では、既存手法と比べて説明の効果を向上させた.

他人のレビューはアイテムの品質や評判を知るために利用されており、アイテムを購入するかどうかの意思決定をする上で重要な判断材料となる[10]. そのため、ユーザがどのようなレビューを参照したのか、評価したのかといったレビューの有用性データはユーザの嗜好を予測する情報源になり得る. レビューの有用性データに関する研究は、有用性データをもとにユーザがまだ参照していないレビューの中から、最適なレビューを推薦する研究が行われてきた[11].

しかし、これらの研究では、アイテムを購入、参照する過程には共通のユーザが含まれるにも関わらず、双方のデータを活用していない.[12]は他人のレビューの参照データを有益であると考え、推薦の際にレビューの参照データを考慮することでEFMの推薦精度を向上させている. 本論文では他人のレビューの参照データを複数のモデルに適用させる.

3 提案手法

有用性データの利用方法を述べた上で、既存手法であるEFMとMTERを基に有用性データを適用したモデルを構築する.

3.1 レビューの有用性データの利用方法

レビューの有用性データにはユーザが他人のレビューを参照した履歴とそのレビューの評価値があり、本研究では、レビューの参照履歴を使用する. 他人のレビューに含まれるアイテムの特徴はレビューの推薦に関する研究において推薦精度に影響を与えるとされており[10], 参照したレビューに含まれる特徴もユーザの嗜好する特徴として有効であると考え. ユーザが参照したレビューから得られた特徴の集合を式(1)で表す.

$$G = \{(u, u', f) \mid u \in U, u' \in U, f \in F\} \quad (1)$$

ユーザ u が ユーザ u' の記載した特徴 f を参照 }

ユーザが自らレビューに記載した特徴に加え、ユーザが参照したレビューに記載された特徴も、ユーザが目目する特徴として追加する.

モデルに有用性データを適用する方法としては、ユーザがレビューに記載した特徴と参照したレビューに記載されている特

徴では、ユーザの興味の度合いに差があると考え、ユーザの特徴に対する興味の度合いを計算する過程ではその差を考慮する必要がある.[5]は注目行列の計算にBPR[6]を導入することで、レビューに記載した特徴と記載していない特徴の注目度の差でモデルを学習させた. 本論文ではBPRを用いてユーザが記載した特徴と参照した特徴と記載も参照もしていない特徴における興味の度合いにそれぞれ差がつくように学習を行う. つまり、参照した特徴は記載した特徴よりも興味の度合いが低いが、記載も参照もしなかった特徴よりは興味の度合いが高いと仮定し、既存のモデルに参照した特徴を考慮するBPRの式を追加する. 既存の説明付き推薦手法に有用性データを考慮した2通りのモデルについて説明する.

3.2 EFMへの適用

3.2.1 Explicit Factor Model(EFM)

レビューから明示的な特徴とその感情極性を抽出し、入力である評価行列 A と注目行列 X と品質行列 Y を構築する. 出力は行列 H_1, U_1, V, U_2, H_2 である. U_1, U_2, V は明示的な特徴に関する行列と呼ばれ、 H_1, H_2 は暗黙的な特徴に関する行列と呼ばれる.

まずは、ユーザと特徴に関する注目行列 X を構築する. ユーザは特徴に対する興味の度合いが高いほど、レビューに記載する傾向にあり、ユーザごとに興味がある特徴は異なると仮定する. レビューから抽出された特徴集合を $F = \{F_1, F_2, \dots, F_p\}$, m 人のユーザ集合を $U = \{u_1, u_2, \dots, u_m\}$ と定義する. ユーザが特徴に関してレビューに記載したときの X の要素は以下となる.

$$X_{ij} = \begin{cases} 0, (u_i \text{ が } F_j \text{ の記載なし}) \\ 1 + (N-1) \left(\frac{2}{1 + e^{-t_{ij}}} - 1 \right), (\text{記載あり}) \end{cases} \quad (2)$$

N は評価値の最大値である. 上記の式の値はAmazonやYelpで用いられる評価値と同じ範囲の $[1, N]$ に変換している.

次に、アイテムと特徴に関する品質行列 Y を構築する. Y の各要素 Y_{ij} はアイテムの特徴に対する品質を表現する. n 個のアイテム集合を $P = \{p_1, p_2, \dots, p_n\}$ と定義する. アイテム p_i について特徴 F_j を k 回記載されたとき、感情極性の値 s_{ij} を用いて Y の要素は以下となる.

$$Y_{ij} = \begin{cases} 0, (p_i \text{ のレビューに } F_j \text{ の記載なし}) \\ 1 + (N-1) \left(\frac{N-1}{1 + e^{-ks_{ij}}} - 1 \right), (\text{記載あり}) \end{cases} \quad (3)$$

上記の式の値はAmazonやYelpで用いられる評価値と同じ範囲の $[1, N]$ に変換している.

3.2.2 EFMに有用性データを考慮したモデル

ユーザ u_i が注目する特徴集合を F_i^+ とすると、学習データは式(4)に表される.

$$D = \{(i, f_A, f_B) \mid f_A \in F_i^+ \wedge f_B \in F \setminus F_i^+\} \quad (4)$$

ユーザ u_i が記載した特徴 f_A と記載しない特徴 f_B の注目度を \hat{X}_{if_A} と \hat{X}_{if_B} とする. 2つの注目度の差 $\hat{X}_{if_A f_B}$ を式 (5) で表す.

$$\begin{aligned}\hat{X}_{if_A f_B} &= \hat{X}_{if_A} - \hat{X}_{if_B} \\ &= (\mathbf{u}_{1i}^T \mathbf{v}_{f_A} - \mathbf{u}_{1i}^T \mathbf{v}_{f_B})\end{aligned}\quad (5)$$

\mathbf{u}_{1i} は U_1 のユーザ u_i に関するベクトルであり, $\mathbf{v}_{f_A}, \mathbf{v}_{f_B}$ は V の特徴 f_A, f_B に関するベクトルである. 式 (5) の f_B はユーザ u_i が記載しない特徴全てに該当する. しかし, アイテムの評価と同様に未評価の特徴, つまり欠損値の特徴が存在すると考えられる. Multi Pairwise Ranking(MPR) [13] は評価されていないアイテムは低い評価のアイテムよりも好むことを考慮した. そこで, 参照したレビューから得られた特徴の中でユーザ自身が記載していない特徴は気にしない特徴として扱わず, 記載した特徴よりも注目度が低い特徴と仮定する. ユーザ u_i が参照した特徴集合を $F_i^+ = \{f_C \mid (i, i', f_C) \in G \wedge f_C \in F \setminus F_i^+\}$ としたときユーザが参照した特徴 f_C を考慮したときの学習データは式 (6) に表される.

$$\begin{aligned}D' &= \{(i, f_A, f_B, f_C, f_D) \mid \\ &f_A \in F_i^+ \wedge (f_B, f_D) \in F \setminus (F_i^+ \cup F_i'^+)\} \wedge f_C \in F_i'^+\end{aligned}\quad (6)$$

D' から各特徴間の注目度の差 $\hat{X}_{if_A f_B f_C f_D}$ を式 (7) で表す.

$$\hat{X}_{if_A f_B f_C f_D} = (\hat{X}_{if_A} - \hat{X}_{if_B}) - (\hat{X}_{if_C} - \hat{X}_{if_D}) \quad (7)$$

注目行列 X について有用性データを考慮した学習モデルは式 (8) で表す.

$$\begin{aligned}&\underset{U_1, U_2, V, H_1, H_2}{\text{minimize}} \left\{ \sum_{(i,j) \in K} (a_{ij} - \mathbf{h}_{1i}^T \mathbf{h}_{2j} - \mathbf{u}_{1i}^T \mathbf{u}_{2j})^2 \right. \\ &- \lambda_x \sum_{(i,j) \in K} \sum_{f_A \in F_i^+} \sum_{f_B, f_D \in F \setminus (F_i^+ \cup F_i'^+)} \sum_{f_C \in F_i'^+} \ln \sigma(\hat{X}_{if_A f_B f_C f_D}) \\ &+ \lambda_y \sum_{(i,j) \in K} \sum_{f_Q \in F_j} (y_{jf_Q} - \mathbf{u}_{2j}^T \mathbf{v}_{f_Q})^2 + \lambda_h \sum_{r'} \sum_{(i,j) \in K} (h_{1ir'}^2 + h_{2jr'}^2) \\ &+ \lambda_v \sum_r \left(\sum_{f_A \in F_i^+} v_{f_A r}^2 + \sum_{f_B, f_D \in F \setminus (F_i^+ \cup F_i'^+)} (v_{f_B r}^2 + v_{f_D r}^2) \right. \\ &\quad \left. + \sum_{f_C \in F_i'^+} v_{f_C r}^2 \right) \\ &\quad \left. + \lambda_u \sum_r \sum_{(i,j) \in K} (u_{1ir}^2 + u_{2jr}^2) \right\} \\ &\text{s.t. } U_1 \in \mathbb{R}_+^{I \times R}, U_2 \in \mathbb{R}_+^{J \times R}, V \in \mathbb{R}_+^{P \times R}, \\ &H_1 \in \mathbb{R}_+^{I \times R'}, H_2 \in \mathbb{R}_+^{J \times R'}, P = [U_1, H_1], Q = [U_2, H_2]\end{aligned}\quad (8)$$

λ_h, λ_v は正則化パラメータであり, λ_x, λ_y はトレードオフパラメータである.

3.3 MTER への適用

3.3.1 Multi-Task Explainable Recommendation(MTER)

入力としてレビューから抽出した特徴, 意見からユーザとアイテムと特徴の関連を表現するテンソル X , ユーザと特徴と意見の関連を表現するテンソル Y^U , アイテムと特徴と意見の関

連を表現するテンソル Y^I を構築する. 出力はユーザベクトル U , アイテムベクトル I , 特徴ベクトル F , 意見ベクトル O , コアテンソル G_1, G_2, G_3 である.

まずは, ユーザとアイテムと特徴の関連度を表現するテンソル X を構築する. ユーザ u_i がアイテム p_j の特徴 F_k に関して記載されたとき, 感情値 s_{ijk} を用いて X の要素は以下となる.

$$X_{ijk} = \begin{cases} 0, & (u_i \text{ の } p_j \text{ のレビューに } F_k \text{ の記載なし}) \\ 1 + \frac{N-1}{1+e^{-s_{ijk}}}, & (\text{記載あり}) \end{cases}\quad (9)$$

次に, ユーザと特徴と意見の関連度を表現するテンソル Y^U を構築する. ユーザ u_i が特徴 F_k に対して意見 w を Γ 回記載したとき, Y^U の要素は以下となる.

$$Y_{ikw}^U = \begin{cases} 0, & (u_i \text{ が } F_k \text{ に対して } w \text{ の記載なし}) \\ 1 + (N-1) \left(\frac{2}{1+e^{-\Gamma}} - 1 \right), & (\text{記載あり}) \end{cases}\quad (10)$$

最後に, アイテムと特徴と意見の関連度を表現するテンソル Y^I を構築する. アイテム p_j の特徴 F_k に対して意見 w を Ω 回記載されたとき, Y^I の要素は以下となる.

$$Y_{jkw}^I = \begin{cases} 0, & (p_j \text{ の } F_k \text{ に対して } w \text{ の記載なし}) \\ 1 + (N-1) \left(\frac{2}{1+e^{-\Omega}} - 1 \right), & (\text{記載あり}) \end{cases}\quad (11)$$

3.3.2 MTER に有用性データを考慮したモデル

入力の学習データには式 (6) を使用し, EFM と同様に有用性データを考慮したモデルの拡張を行う. D' から各特徴間の注目度の差 $\hat{X}_{if_A f_B f_C f_D}$ を式 (12) で表す.

$$\hat{X}_{if_A f_B f_C f_D} = (\hat{X}_{if_A} - \hat{X}_{if_B}) - (\hat{X}_{if_C} - \hat{X}_{if_D}) \quad (12)$$

テンソル X について有用性データを考慮した学習モデルを式 (13) で表す.

$$\begin{aligned}&\underset{\hat{X}, \hat{Y}^U, \hat{Y}^I}{\text{minimize}} \|\hat{X} - \tilde{X}\|_F + \|\hat{Y}^U - Y^U\|_F + \|\hat{Y}^I - Y^I\|_F \\ &- \lambda_B \sum_{i=1}^m \sum_{(j,l) \in D_i^*} \ln \sigma(\hat{x}_{ij(p+1)} - \hat{x}_{il(p+1)}) \\ &- \lambda_C \sum_{(n,m) \in K} \sum_{f_A \in F_n^+} \sum_{f_B, f_D \in F \setminus (F_n^+ \cup F_n'^+)} \sum_{f_C \in F_n'^+} \ln \sigma(\hat{X}_{if_A f_B f_C f_D}) \\ &\quad + \lambda_F (\|U\|^2 + \|I\|^2 + \|F\|^2 + \|O\|^2) \\ &\quad + \lambda_G (\|G_1\|^2 + \|G_2\|^2 + \|G_3\|^2) \\ &\quad \text{s.t. } \hat{X} = G_1 \times_a U \times_b I \times_c \hat{F}, \\ &\quad \hat{Y}^U = G_2 \times_a U \times_c F \times_d O, \\ &\quad \hat{Y}^I = G_3 \times_b I \times_c F \times_d O, \\ &\quad U \geq 0, I \geq 0, F \geq 0, O \geq 0, G_1 \geq 0, G_2 \geq 0, G_3 \geq 0.\end{aligned}\quad (13)$$

λ_F, λ_G は正則化パラメータであり, λ_B, λ_C はトレードオフパラメータである.

4 実 験

本実験では既存手法に有用性データを適用することで, 推薦精度と説明の質が向上するかを確かめる. 推薦精度の評価はランキング予測によって行う.

4.1 ランキング予測

本実験ではデータセットに [14] で用いた Ciao を使用する。既存手法として EFM と MTER を使用し、各手法に有用性データを考慮した場合の精度比較を行う。データセットの評価値が非常に疎であるため、5 つ以上のアイテムを評価したユーザと 2 つ以上の評価されたアイテムのみをデータに用いた。表 1 に元の Ciao データと実験で使用する Ciao データの統計情報を示す。本研究の上記の条件下での実験データではユーザのアイテム評価数のうち約 40% のアイテムに関するレビューを参照している。また表 2 では実験で記載した特徴数の数が 5 個以下、つまり $|F_{A_i}^+| \leq 5$ を満たすユーザ i が参照したレビューから得られた特徴 f_C の数を表す。どのユーザも 5 個以上の特徴を参照していることから参照した特徴を考慮することで学習データを増加できる。これらのことから、レビューの有用性データは嗜好データを増加する見込みがあると考えられる。全データのうち 20% をテストデータ、ハイパーパラメータは 5 交差検定でグリッドサーチを用いて決定した。上位 10 件のランキング予測問題として NDCG(Normalized Discounted Cumulative Gain) を用いる。

表 1 データセットの統計情報

データセット	ユーザ数	アイテム数	アイテム評価数	レビュー数	レビュー評価数
Ciao	43,666	11,2804	302,232	302,232	8,894,899
Ciao(実験)	101	368	3109	3109	5400

表 2 ユーザごとの特徴数に関するデータ

ユーザ i	記載した特徴数 $ F_{A_i}^+ $	参照から得られた特徴数 $ F_{C_i} $
3	4	75
16	3	8
27	4	14
61	2	39
62	5	37
69	2	22
74	1	20
78	1	18
83	3	33
92	3	34
97	5	35
98	4	26

4.1.1 実験結果

ランキング予測による実験結果を表 3 に示す。MTER は EFM にユーザの意見という要素を追加するために、行列からテンソルにしたモデルであり、推薦精度において EFM を上回ることについては [7] で検証されている。EFM_MPR は EFM に式 (8) を適用したモデルであり、MTER_MPR は MTER に式 (13) を適用したモデルである。MTER_MPR が最も推薦精度が高いことと、有用性データを考慮した 2 つの既存モデルは推薦精度が向上していることを示す。このことから記載した特徴は参照した特徴より注目度が高く、参照して得られた特徴は記載も参照もされていない特徴よりも注目度が高いという仮説が今回適用したモデルの学習において有効だったといえる。

表 3 ランキング予測における NDCG

	EFM	EFM_MPR	MTER	MTER_MPR
NDCG	0.06166	0.13372	0.131268	0.151667

4.2 説明生成

4.2.1 実験設定

データセットは Ciao をフィルタリングした表 1 のデータを使用する。モデルによって生成された説明の質に関して 2 点の評価を行う。

1 点目は一貫性の評価である。この評価は深層学習における判断根拠を理解しようとする研究に用いられる [15]。深層学習における「説明」とはネットワークの出力に貢献を行っている特徴量を人間が理解できる画像やテキストといった表現に変換することである。学習の中身を説明するために、さまざまな手法がある。これらの手法で生成された説明に対する評価の 1 つに説明の一貫性が用いられる [16]。説明の一貫性とはある入力に対する説明は、その入力に近いデータに対する説明と近しければ、一貫性が担保されており、説明の質が高いと考える。本実験では説明の一貫性の評価を用いてモデルの説明生成の質を評価する。入力に必要な近いデータとしては表 1 の Ciao のデータからユーザごとのレビュー数を一部削除し、削除する前と削除した後でどの程度説明が変化したかを確認する。削除する割合は今回は 1/3 にした。EFM と MTER はユーザが嗜好する特徴を予測し、その特徴をテンプレートの説明文に当てはめることで説明を行う。そのため、ユーザの特徴に対する嗜好度をランキング化した特徴ランキングの変化を NDCG で確かめることで説明の変化を確認する。

2 点目は嗜好する特徴の予測精度の評価である。テストデータに記載された特徴は、モデルを用いて訓練データから作成した嗜好する特徴のランキング内でも順位を調べることで評価を行う。

実験を始める前に EFM と MTER に有用性データを用いることで説明が変化しているかどうかを同じユーザで確認する。それぞれのモデルを用いて説明文を生成した結果を以下に示す。

- EFM: you might be interested in [acting] ,on which this product performs well.
- EFM_MPR: you might be interested in [cast] ,on which this product performs well.
- MTER: Its [casting] is [great].
- MTER_MPR: Its [character] is [great].

EFM と EFM_MPR のモデルは [feature] のみを出力するので、説明生成にはテンプレートの文章を使用した。MTER と MTER_MPR は [feature] と [opinion] を出力する。それぞれのモデルで異なる特徴が出力されており、有用性データを考慮することで生成される説明が変化することを確認できた。

4.2.2 実験結果

一貫性の評価における実験結果を表4に示す。

表4 一貫性の評価におけるNDCG

	EFM	EFM_MPR	MTER	MTER_MPR
NDCG	0.851728	0.903029	0.921502	0.929526

MTER_MPRが最も高い値を出しており、データセットの量を減らした場合も、元のデータセットの場合と最も近い特徴のランキングを生成することができたことを示した。また、生成される特徴のランキングが一貫性を保つためにランキング損失を考慮することとユーザの意見というデータの要素を増やすことは有効であった。

次に、嗜好する特徴予測における実験結果を表5に示す。

表5 嗜好する特徴予測における順位

	EFM	EFM_MPR	MTER	MTER_MPR
NDCG	47.311020	41.754084	39.145311	38.103622

各手法の値は、モデルによって作成した嗜好する特徴のランキング内で、テストデータに含まれる特徴の平均順位である。順位の高いほど、ユーザが実際にレビューに記載している特徴を予測することができることを示す。MTER_MPRが最も高い順位を出しており、EFM_MPRもEFMと比べて向上している。

5 まとめ

本論文では、既存の説明付き推薦手法にレビューの有用性データを考慮したモデルを構築した。ユーザが自らレビューに記載した特徴としない特徴、参照したレビューに記載してある特徴と記載していない特徴の差をモデルに学習させることで、レビューから得られる暗黙的なデータに対応した拡張を行った。実験結果としては、従来手法と比べ、ランキング予測の精度と説明の質が向上した。今後はレビューの参照履歴だけでなく、レビューの評価を考慮したモデルを構築する必要がある。また、今回は適用しなかった既存手法にも有用性データを適用し、その効果をさらに検証する必要があると考える。

文 献

- [1] Yehuda Koren, Robert Bell, and Chris Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, pp. 30-37(2009).
- [2] Nava Tintarev and Judith Masthoff, "Effective Explanations of Recommendations: User-Centered Design" In *RecSys*, 2007.
- [3] Yongfeng Zhang and Guokun Lai, "Explicit Factor Models for Explainable Recommendation based on Phrase-level Sentiment Analysis" In *SIGIR*, 2014.
- [4] Yue Lu, Malu Castellanos, Umeshwar Dayal, and Chengxiang Zhai, "Automatic Construction of a Context-aware Sentiment Lexicon: An Optimization Approach," In *WWW*, 2011.
- [5] Xu Chen, Zheng Qin, Yongfeng Zhang, and Tao Xu, "Learning to Rank Features for Recommendation over Multiple

Categories," In *SIGIR*, 2016.

- [6] Steen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme, "BPR: Bayesian Personalized Ranking from Implicit Feedback," In *UAI*, 2009.
- [7] Nan Wang, Hongning Wang, Yiling Jia, Yue Yin, "Explainable Recommendation via Multi-Task Learning in Opinionated Text Data," In *SIGIR*, 2018.
- [8] Alexandros Karatzoglou, Xavier Amatriain, Linas Baltrunas, and Nuria Oliver. "Multiverse recommendation: n-dimensional tensor factorization for contextaware collaborative filtering." In *RecSys*, 2010.
- [9] Tamara G. Kolda and Brett W. Bader, "Tensor Decompositions and Applications," *SIAM Rev.* 51, 3, 455-500, 2009.
- [10] Y. Lu, P. Tsaparas, A. Ntoulas, and L. Polanyi, "Exploiting social context for review quality prediction," In *WWW*, 2010.
- [11] Yu Hong, Jun Lu, Jianmin Yao, Qiaoming Zhu, and Guodong Zhou, "What reviews are satisfactory: novel features for automatic helpfulness voting," In *SIGIR*, 2012.
- [12] 坂本竜二, 有次正義, "レビューの有用性データを考慮した説明生成可能な推薦手法" In *DEIM*, 2019.
- [13] Runlong Yu, Yunzhou Zhang, Yuyang Ye, Le Wu, Chao Wang, Qi Liu, and Enhong Chen, "Multiple Pairwise Ranking with Implicit Feedback," In *CIKM*, 2018.
- [14] Jiliang Tang, Huiji Gao, Xia Hu, and Huan Liu. *Context-Aware Review Helpfulness Rating Prediction*. In *RecSys*, 2013.
- [15] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhutdinov, Richard Zemel, Yoshua Bengio, "Show, Attend and Tell: Neural Image Caption Generation with Visual Attention," In *PMLR*, vol. 37, pp. 2048-2057(2015).
- [16] Gregoire Montavon, Wojciech Samek, Klaus-Robert Muller, "Methods for Interpreting and Understanding Deep Neural Networks" In *ICASSP*, 2017.