

価値観に基づくユーザモデルによる協調フィルタリングの拡張手法の提案

Proposal of Extended Collaborative Filtering by Personal-value-based User Model

三澤 遼理*¹
Ryori Misawa

服部 俊一*¹
Shunichi Hattori

高間 康史*¹
Yasufumi Takama

*¹ 首都大学東京大学院システムデザイン研究科
Graduate School of System Design, Tokyo Metropolitan University

This paper proposes to extend collaborative filtering by introducing personal value-based user model. A personal value-based user model consists of attributes on which a user put high priority when evaluating an item. This paper employs this model for calculating similar users in collaborative filtering. Effectiveness and characteristics of the proposed method is shown by applying it to dataset of actual review sites.

1. はじめに

本稿では、価値観に基づくユーザモデルを協調フィルタリングに適用した推薦手法について提案する。オンラインショッピングサイトなどにおいて、ユーザに有用な情報を提示する手法として協調フィルタリングが広く用いられている [Resnick 94]。既存の推薦手法に共通する課題として、新規ユーザや新規アイテムに対するシステム利用履歴の不足から適切な推薦が不可能となる、cold-start 問題が指摘されている [Schein 02]。一方、個人の嗜好や消費行動に影響を与える要素として価値観が挙げられる。情報推薦における価値観はアイテムの属性に対するこだわりの強さとして表れると考えられ、こだわりを持つ属性ほど、より少ない情報でモデリング可能であることが示されている [Hattori 13]。類似した価値観を持つユーザは、アイテムを選択する際に共通する視点を持つことが期待されるため、協調フィルタリングにおける類似ユーザの計算に価値観を利用する事で、少数の利用履歴しか持たないユーザへの推薦が可能となる事による cold-start 問題の解決や、ユーザのこだわりとの関連が強い事が想定される、極性の強い評価を受けたアイテムへの推薦精度の向上が期待できる。フォートラベルからデータを取得し作成したデータセットに適用し、提案手法の有効性を示す。

2. 関連研究

情報推薦システムで用いられる代表的手法として協調フィルタリングが挙げられる [Resnick 94]。協調フィルタリングは嗜好が類似する他のユーザの嗜好情報を用いてユーザの嗜好を予測する手法であり、利点としてユーザの嗜好情報のみで推薦が行えることや処理が手軽なことが挙げられる。しかし、協調フィルタリングには新規ユーザ、新規アイテムに対してシステムの利用履歴が少ないことから推薦の精度が低くなってしまう cold-start 問題が指摘されている [Schein 02]。

協調フィルタリングはメモリベース法とモデルベース法に分類することができる [神島 08]。メモリベース法では推薦を行う時点で蓄積された嗜好データを利用して推薦を行う。モデルベース法では事前に構築しておいたモデルと推薦を受けるユーザの嗜好データをもとに推薦を行う。さらに、メモリベース法は大きくユーザベースとアイテムベースの 2 つに分類できる [神島 08]。ユーザベースは嗜好が類似するユーザが好むものを推薦する

が、アイテムベースはユーザが好むアイテムと類似するアイテムを推薦する。提案手法のアルゴリズムはユーザベースのアルゴリズムに一部変更を加えたものとなっている。

この他、提案手法と関連の強い研究として、アイテムが持つ属性に対する評価を利用して商品の推薦を行なった伊藤らの研究が挙げられる [伊藤 10]。伊藤らの研究では主成分分析を用いて求めた評価傾向の類似するユーザの情報を用いて推薦を行なっており、属性が評価に与える影響を利用する提案手法とは目的が異なる。

3. 価値観に基づくユーザモデルを適用した協調フィルタリング

3.1 価値観に基づくユーザモデル

価値観の定義の一つとして、ユーザが抱く属性へのこだわりの強さに着目し、ユーザモデリングを行う手法が提案されている [Hattori 13]。価値観に基づくユーザモデルの構築には評価一致率と呼ばれる指標を用いる。ユーザのアイテムに対する評価極性(好評または不評)とアイテムの各属性に対する評価極性を用いて、属性ごとの評価一致率を計算しユーザモデルを構築する。ユーザ u_i のあるアイテム j への評価極性 p_{ij} とアイテムの属性 a_k への評価極性 p_{ijk} が一致するかどうかを調べ、一致する回数を O_{ik} 、一致しない回数を Q_{ik} とする。この時、ユーザ u_i の属性 a_k の評価一致率 P_{ik} は式(1)で算出される。価値観に基づくユーザモデルは属性数を l とすると、 l 次元のベクトルとして表される。評価一致率が高い属性は評価への影響度が高い、こだわりの強い属性と推論される。

$$P_{ik} = \frac{O_{ik}}{O_{ik} + Q_{ik}} \quad (1)$$

3.2 類似度計算と評価値の予測

提案手法は次の 3 つの手順から構成される。

1. 全ユーザの価値観に基づくユーザモデル作成
2. ユーザモデルを用いたユーザ間類似度の計算
3. ユーザのアイテムへの予測評価値を、ユーザ間の類似度を重みとした評価値の加重平均によって計算

手順 1 では、式(1)を用いて、全ユーザの価値観に基づくユーザモデルを作成する。手順 2 では、作成されたユーザモデルを用いてユーザ間の類似度を計算する。類似度計算手法には Pearson 相関係数を採用した。ユーザ u_i のアイテム j への評価値を s_{ij} とすると、推薦対象ユーザ u_t とユーザ u_i の Pearson 相関

係数 $Pear(t, i)$ は式(2)で表される。ただし、 I_{ti} は 2 人のユーザ u_t, u_i が共通に評価したアイテムの集合であり、 $\bar{s}_t' (\bar{s}_i')$ は 2 人のユーザが共通に評価しているアイテムの、 $u_t (u_i)$ の総合評価値の平均値である。

$$Pear(t, i) = \frac{\sum_{ik \in I_{ti}} (s_{tk} - \bar{s}_t') (s_{ik} - \bar{s}_i')}{\sqrt{\sum_{ik \in I_{ti}} (s_{tk} - \bar{s}_t')^2} \sqrt{\sum_{ik \in I_{ti}} (s_{ik} - \bar{s}_i')^2}} \quad (2)$$

手順 3 では、手順 2 で求めたユーザ間類似度を用いて予測評価値を計算し、アイテム推薦を行う。ユーザ u_t のアイテム i_j への予測評価値 \hat{s}_{tj} は式(3)で表される [神 野 08]。ただし、 U_j はアイテム i_j を評価済みのユーザの集合である。式(3)で計算された予測評価値を元にアイテムの推薦候補リストを作成し、アイテムの推薦を行う。

$$\hat{s}_{tj} = \frac{\sum_{u_i \in U_j} Pear(t, i) s_{ij}}{\sum_{u_i \in U_j} Pear(t, i)} \quad (3)$$

提案手法ではユーザ間の類似度計算に、価値観に基づくユーザモデルを用いる。価値観に基づくユーザモデルはユーザがこだわりを持つ属性について、より少ない情報でモデリングすることが可能であることが示されており [Hattori 13]、提案手法ではこの特性を利用する事で、システムの利用履歴が少ないユーザに対するアイテムの推薦が可能になることや、評価の極性が極端なアイテムの推薦精度の向上が期待できる。

4. フォートラベルデータセットを用いた評価実験

4.1 実験概要

本稿では、極性の強い評価を受けたアイテムに対する推薦精度の評価実験結果を示す。実験には、旅行口コミサイトであるフォートラベルのレビューをスクレイピングによって収集し、構築したデータセットを用いた。このデータセットはユーザ 15,948 名、アイテム(施設) 11,825 個、評価データ 83,452 件を含んでおり、アイテムへの評価値と属性への評価値の範囲は 0.5 刻みで 1 から 5 である。また、属性への評価値は一部欠落している場合もある。

フォートラベルデータセットは 1 件しか評価していないユーザが全体の約半数を占めるなど、実験には不要なデータが含まれているため、データセットの加工を行った。評価件数が 2 以下のユーザ、評価を付与された件数が 1 以下のアイテムを削除した。また、特定の評価値の割合が 90%以上であるユーザ、アイテムも、協調フィルタリングによる推薦に適さないと考え削除した。

このデータセットに対して交差確認を適用して、式(2)によって計算されたアイテムへの予測評価値と、実際の評価値との絶対誤差を計算し、その平均値である MAE(平均絶対誤差)により提案手法を以下の 3 手法と比較する。

1. ユーザベースの協調フィルタリング [Resnick 94]
2. アイテム平均評価値
3. ランダム推薦

1, 2 は推薦システムを評価する際の比較対象としてよく用いられるアルゴリズムである。ランダム推薦は、ユーザ間の類似度を必ず 1 としてユーザベースの協調フィルタリングを行う推薦手法であり、提案手法によるユーザ間の類似度計算の妥当性を検証するために比較対象としている。

評価の極性が強いアイテムとは、評価がある値に偏っておらず、1 または 5 近辺の評価が多く付けられているアイテムとする。そのような評価には、ユーザのこだわりが強く反映されると考えられるため、提案手法で利用するユーザモデルが有効に機能

することが期待できる。そこで、実際にユーザがアイテムへ付与していた評価値毎に 4 手法の MAE を計算し、比較を行う。

4.2 実験結果

図 1 に 4 つの手法の評価値毎の MAE のグラフを示す。横軸はユーザがアイテムへ付与した実際の評価値、縦軸は MAE である。全ての評価値において提案手法の MAE が最も低く、特に評価値 1 から 2.5 において差が大きいことがわかる。これより、提案手法ではユーザのこだわりを反映した推薦が可能といえる。

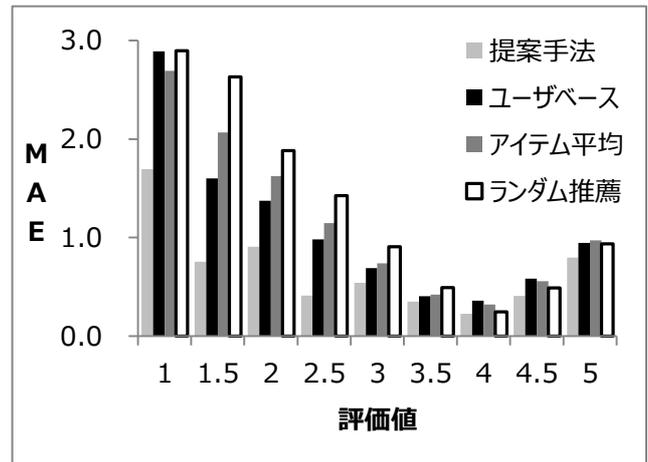


図 1: 評価値毎の MAE の比較

5. 結論

本稿では、価値観に基づくユーザモデルを協調フィルタリングに適用し、類似する価値観を持つユーザの情報を用いた推薦手法を提案した。本稿では極性の強い評価を受けたアイテムに対する提案手法の有効性を示したが、cold-start 問題の観点でも良好な結果を得ている。今後は、評価履歴およびユーザモデルのクラスタリングや、ユーザ、アイテムごとの予測評価値の MAE の計算などを行い、提案手法の性質についてより詳細な分析を行うことにより、提案手法に適した対象や応用が明らかになることが期待できる。また、より多様なデータセットでの検証も、提案手法の特性を明らかにする上で有効と考える。

参考文献

- [Hattori 13] Hattori, S., and Takama, Y. : Proposal of User Modeling Method Employing Reputation Analysis on User Reviews Based on Personal Values, in *JSAI2013*, 1A3-IOS-3a-4 (2013)
- [Resnick 94] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. and Riedl, J. : GroupLens: An Open Architecture for Collaborative Filtering of Netnews, in *Proc. of CSCW*, pp. 175-186 (1994)
- [Schein 02] Schein, A.-I., Popescul, A., Ungar, L.-H., and Pennock, D.-M. : Methods and metrics for Cold-Start Recommendations, in *Proc. of SIGIR*, pp. 253-260 (2002)
- [伊藤 10] 伊藤 ゆかり, 波多野 賢治, 松本 尚宏: 商品の価値とユーザの嗜好を考慮した商品推薦システムの提案, 電子情報通信学会第二種研究会技術報告, WI2-2010-27, pp. 101-106 (2010)
- [神野 08] 神野 敏弘: 推薦システムのアルゴリズム(2), 人工知能学会誌, Vol. 23, No. 1, pp. 89-103 (2008)