

協調フィルタリングにおけるアイテムベースとユーザベースの 統合型推薦システムに関する検討

A Study on Recommendation Combining Item-based and User-based Method in Collaborative Filtering

伊藤寛明 吉川大弘 古橋武
Hiroaki Ito Tomohiro Yoshikawa Takeshi Furuhashi

名古屋大学大学院工学研究科
Graduate School of Engineering Nagoya University

Collaborative filtering is one of the typical methods in recommendation system. Memory-based collaborative filtering can be classified into item-based and user-based method. We applied the association analysis which is one of the data mining techniques to recommendation system and gained the knowledge that item-based method was superior in Accuracy and user-based method was superior in Serendipity. In this paper, we study on the combination of them to improve the performance, and propose a new recommendation method to improve Serendipity based on the item-based method.

1. はじめに

近年、インターネットの普及により電子商取引が増加しており、それに伴い EC サイトでは膨大な数の商品を扱うようになってきている。そのため、それらの商品の中から、ユーザの嗜好にあった商品をユーザ自身で探し出すことが困難となり、推薦システムの利用が期待されている [神鷲 08]。一方、大量にあるデータの中から、価値のある情報を抽出するデータマイニング手法の一つにアソシエーション分析がある。この手法をユーザの評価履歴に対して適用し、協調フィルタリングによるアイテム推薦に用いた研究が報告されている [Lin 02][吉川 13]。また一方推薦システムにおいて、推薦したアイテムがユーザに好まれた割合を表す“精度”は、最も重要な評価指標の 1 つである。しかし近年、ユーザ満足度の観点から、精度に加えて、“意外性”に対する評価の必要性が指摘され始めている [神鷲 08][吉川 13]。

著者らはこれまでに、アソシエーション分析に基づく推薦システムに対する検討を行い、アイテムベースは精度が高く、ユーザベースは意外性が高いという知見を得た [伊藤 13]。本稿では、これら両者を統合することを検討し、推薦システムにおける総合的な性能向上を図る。また、アイテムベースの高い精度を保ちつつ、意外性を向上させるための手法を提案する。

2. 推薦システム

2.1 アソシエーション分析

アソシエーション分析とは、データの中から価値のある組み合わせ（アソシエーションルール）を見つけ出す手法である。アソシエーションルールは、 $A \Rightarrow B$ と表され、 A は条件部、 B は結論部と呼ばれる。代表的なアソシエーションルールの評価指標として *confidence* がある。

$$confidence = \frac{N(A \cap B)}{N(A)} \quad (1)$$

$N(A)$ は条件部 A 、 $N(A \cap B)$ は条件部 A と結論部 B を同時に満たすデータの件数である。本稿においては、ユーザ数またはアイテム数となる。

連絡先: 伊藤寛明, 名古屋大学大学院工学研究科, 名古屋市千種区不老町, 052-789-2793, 052-789-3166, itou@cplx.cse.nagoya-u.ac.jp

2.2 アイテムベース協調フィルタリング

推薦を行うユーザ（以降、“対象ユーザ”と呼ぶ）の評価履歴をアソシエーションルールの条件部に用いて、結論部に各アイテムに対する評価「Like」を当てる。例えば、対象ユーザがアイテム A に対して「Like」と評価し、アイテム B が未評価であるとき、全ユーザに対して求められる「アイテム A =Like \Rightarrow アイテム B =Like」の *confidence* を、アイテム B のスコアに加算する。対象ユーザのすべての評価履歴により未評価のアイテムのスコアを求め、最もスコアの高いアイテムを推薦する [吉川 13]。

2.3 ユーザベース協調フィルタリング

ユーザベースでは、各ユーザの評価をアソシエーションルールの条件部に用いて、結論部には対象ユーザの評価「Like」を当てる。例えば、対象ユーザとユーザ 1 の評価履歴に基づいて求められた「ユーザ 1=Like \Rightarrow 対象ユーザ=Like」というルールに対する *confidence* が高ければ、ユーザ 1 が「Like」と評価をしたアイテムは対象ユーザにとっても「Like」となる可能性が高いとして、*confidence* をアイテムのスコアに加算する。これは、「ユーザ 1=Don't Like \Rightarrow 対象ユーザ=Like」に対しても計算する。これをすべてのユーザ、すべての対象ユーザの未評価アイテムについて求め、最もスコアの高いアイテムを推薦する [Lin 02]。

2.4 統合手法

アイテムベース、ユーザベースにおけるアイテムのスコアをそれぞれ I_{score} 、 U_{score} とし、両者を統合する。統合型推薦システムにおけるアイテム I_i のスコアを式 (2) で定義する。

$$Score_{I_i} = \alpha * I_{score}_{I_i} + (1 - \alpha) * U_{score}_{I_i} \quad (2)$$

$\alpha = 0$ はユーザベースによる推薦、 $\alpha = 1$ はアイテムベースによる推薦を意味する。対象ユーザの未評価アイテムのうち、式 (2) で求められたスコアの最も高いアイテムを推薦する。

2.5 提案手法

本稿では、意外性の向上を目的としたアイテムベースの推薦手法を提案する。ここでは、アソシエーションルールの条件部における対象ユーザの評価だけでなく、その反対の評価を条件部としたルールの情報を用いる。対象ユーザの評価履歴を A とし、それを条件部に用いた場合を *confidence*_($A \Rightarrow B$)、反対

の評価 \bar{A} を用いた場合を $confidence_{(\bar{A} \Rightarrow B)}$ と表記する. 両者の差 d を式 (3) で定義する.

$$d = confidence_{(A \Rightarrow B)} - confidence_{(\bar{A} \Rightarrow B)} \quad (3)$$

精度に結びつくと考えられる $confidence_{(A \Rightarrow B)}$ が高い場合であっても, そもそも B が全体からの「Like」割合が高ければ, d の値は小さくなる. よって, d の値が大きいことは, 対象ユーザの評価が A であることの情報量が大い, すなわち, 対象ユーザに特化した意外な推薦に結びつくと考えられる. 提案手法では, 推薦に用いるスコアを式 (4) で定義する.

$$s_B = \begin{cases} confidence_{(A \Rightarrow B)}^\beta * d & \text{if } d \geq 0 \\ confidence_{(A \Rightarrow \bar{B})}^\beta * d & \text{otherwise} \end{cases} \quad (4)$$

β は精度に対する重みであり, β が大きいときは精度重視, 小さいときは意外性重視の推薦となる. 対象ユーザのすべての評価履歴に基づき, 未評価のアイテムのスコアを式 (4) で求め, 最もスコアの高いアイテムを推薦する.

3. 実験

3.1 使用データ

実験には, GroupLens^{*1} が公開している MovieLens[Cantador 11] の映画に対する評価データを用いた^{*2*}. 映画に対する 10 段階の評点 (0.5~5.0 の 0.5 刻み) のうち, 0.5 から 3.5 を「Don't Like」, 4 から 5 を「Like」として実験を行った. ただし, 「Like」と「Don't Like」をそれぞれ 51, 50 回以上評価したユーザ 1118 人, 300 人以上に評価された 611 のアイテムを対象とした.

3.2 推薦システムの評価

本実験では, 対象ユーザにおける評価履歴「Like」のアイテムの中から, ランダムに選択された 1 つを評価済みアイテムとして与えた状態から, その他の評価済みアイテムを「未評価」とみなしてアイテムの推薦を 50 回行った. 推薦システムの評価指標 ($= \frac{1}{N} \sum_{i=1}^N t_i$) を以下に示す [吉川 13]. 各指標について, 10-fold cross-validation を 10 試行行ったときの平均値を求めた. 推薦回数を N , 推薦アイテムの集合を $I = \{I_1, I_2, \dots, I_N\}$, I_i に対する評価履歴を $e(I_i) = 1 / -1$ (Like/Don't Like) とする.

a) 精度

精度は, 対象ユーザが推薦されたアイテムに対して「Like」と答えた割合である.

$$t_i = \begin{cases} 1 & \text{if } e(I_i) = 1 \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

b) Novelty [Chandrashekar 11]

式 (6) で, I_{NP} は Non-Personalized 法における推薦アイテムの集合であり, Novelty は推薦アイテムが「Like」, かつ Non-Personalized な推薦には現れない割合である.

$$t_i = \begin{cases} 1 & \text{if } e(I_i) = 1 \text{ and } I_i \notin I_{NP} \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

c) Personalizability [吉川 13]

式 (7) において, $P(e(I_i) = 1)$ は, 全ユーザにおけるアイテム I_i に対する「Like」割合である. Personalizability は, 推薦アイテムの「Like」の割合の低さを情報量にしたもので, 推

薦されたアイテムが「Like」, かつそのアイテムの「Like」割合が小さいほど大きな値をとる.

$$t_i = \begin{cases} \log_2 \frac{1}{P(e(I_i)=1)} & \text{if } e(I_i) = 1 \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

b), c) は, 従来研究において意外性の指標として提案されている. 図 1 に, 2.4, 2.5 でそれぞれ示した手法に対する, 上述した a)~c) の指標値の変化を示す. 本来, α と β は同じ値で比較すべきものではないが, 共に α, β 大 \Rightarrow 精度重視となるパラメータであるため, 同じ軸で表している. 統合手法により, 精度または意外性の重要度を調整可能であることがわかる. さらに提案手法により, 精度・意外性共に高い推薦システムの実現可能性が示された.

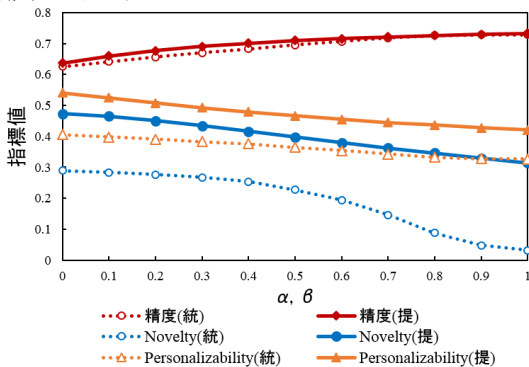


図 1: α, β による指標値の変化

4. おわりに

本稿では, 推薦システムにおける総合的な性能向上を目的として, アイテムベースにおいて, 対象ユーザとは異なる評価をしたユーザの嗜好情報を用いることで, 精度を保ちつつ, 意外性の向上が可能となる手法を提案した. 今後の課題として, ユーザベースの性能向上, そして統合型推薦システムの検討が挙げられる.

参考文献

[Cantador 11] Cantador, I., Brusilovsky, P., and Kuflik, T.: HetRec2011, pp. 387–388 (2011)

[Chandrashekar 11] Chandrashekar, H. and Bhasker, B.: Personalized Recommendation Systems: Entropy Based Collaborative Filtering Recommender Techniques., *Journal of Electronic Commerce Research*, Vol. 12, No. 3 (2011)

[Lin 02] Lin, W., Alvarez, S. A., and Ruiz, C.: Efficient adaptive-support association rule mining for recommender systems, *Data mining and knowledge discovery*, Vol. 6, No. 1, pp. 83–105 (2002)

[伊藤 13] 伊藤 寛明, 吉川 大弘, 古橋 武: アソシエーションルールを用いたアイテム推薦におけるアイテムベースとユーザベースの性能比較, 人工知能学会全国大会論文集, Vol. 27, pp. ROMBUNNO.3E1-3 (2013)

[吉川 13] 吉川 大弘, 森 貴章, 古橋 武: Personalizability を考慮した推薦システムの提案, 情報処理学会論文誌. 数理モデル化と応用, Vol. 6, No. 1, pp. 111–118 (2013)

[神嶋 08] 神嶋 敏弘: 推薦システムのアルゴリズム (1)~(3), 感性工学研究論文集, Vol. 22, 23, No. 6, 2 (2007–2008)

*1 GroupLens research group <http://www.grouplens.org>

*2 IMDb Website <http://www.imdb.com>

*3 Rotten Tomatoes website <http://www.rottentomatoes.com>