

アイテムベース協調フィルタリングにおける 意外性のあるアイテム推薦手法の検討

伊藤 寛明[†] 吉川 大弘[†] 古橋 武[†]
名古屋大学[†]

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 [1]。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 [2][3]。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 [1][3]。

従来のアソシエーション分析を適用した協調フィルタリングでは、推薦を行うユーザと同じ評価をしたユーザの嗜好情報をもとにアイテムを推薦していた。本稿では、この協調フィルタリングによる推薦システムにおいて、推薦を行うユーザと異なる評価をしたユーザの嗜好情報も合わせて用いることで、推薦システムにおける意外性の向上を図る。

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。このルールは、 A という事象が生じたときに、 B という事象が生じるという意味をもつ。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。

A study on unexpected item recommendation method in item-based collaborative filtering

Hiroaki Ito[†], Yoshikawa Tomohiro[†], Takeshi Furuhashi[†], [†]Nagoya University

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

$N(A)$, $N(A \cap B)$ はそれぞれ、条件部 A 、条件部 A と結論部 B を同時に満たすデータの件数（本稿においてはユーザ数）である。

2.2 提案手法

推薦を行うユーザ（以降、“対象ユーザ”と呼ぶ）の評価、さらにその反対の評価をアソシエーションルールの条件部に用いて、結論部に各アイテムに対する評価「Like」を当てる。対象ユーザの評価履歴を A とし、それを条件部に用いる場合の *confidence* を $confidence_{(A \Rightarrow B)}$ 、同様に反対の評価 \bar{A} を条件部に用いる場合を $confidence_{(\bar{A} \Rightarrow B)}$ と表記する。

両者の差 d を式 (2) で定義する。

$$d = confidence_{(A \Rightarrow B)} - confidence_{(\bar{A} \Rightarrow B)} \quad (2)$$

d は $confidence_{(A \Rightarrow B)}$ が高く、反対の評価をしたユーザが「Like」と評価していない場合に大きくなるため、嗜好性の違いの大きさ、すなわち B が「Like」となるために \bar{A} でなく A であることの情報量の大きさを表しているといえる。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、 B の全体からの「Like」割合が高ければ、式 (2) の右辺はどちらの項も値が大きくなるため、 d の値は小さくなる。よって、 d の値が大きいことは、対象ユーザに特化した意外な推薦に結びつくと考えられる。しかし、 d の値のみでは $confidence_{(A \Rightarrow B)}$ の値そのものが考慮されない。そこで推薦に用いるスコアを式 (3) で定義する。

$$s_B = \begin{cases} confidence_{(A \Rightarrow B)}^\alpha * d & \text{if } d \geq 0 \\ confidence_{(A \Rightarrow \bar{B})}^\alpha * d & \text{otherwise} \end{cases} \quad (3)$$

α は精度に対する重みであり、 α が 0 の場合に、 $s_B = d$ となるため、 α が 0 に近いほど意外性のあるアイテムを推薦することができると考えられる。対象ユーザが条件部のアイテムに対して評価を行い、結論部のアイテムが未評価であるとき、全ユーザの評価情報により

求められる s_B を、結論部のアイテム B のスコアに加算する。対象ユーザのすべての評価履歴をもとに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。

3 実験

3.1 使用データ

実験には、GroupLens[4] が公開している MovieLens[5] の映画に対する評価データを用いた [6][7]。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。10-fold cross-validation を 10 試行行ったときの平均値を求めた。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 [3]。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。

$$t_i = \begin{cases} 1 & \text{if } e(I_i) = 1 \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

b) Novelty [3]

式 (5) 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。

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

c) Personalizability [3]

式 (6) 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。

$$t_i = \begin{cases} \log_2 \frac{1}{P(e(I_i)=1)} & \text{if } e(I_i) = 1 \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

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。これらの指標をもとに、従来手法である Weighted Sum 法 [8] と提案手法を比較した。従来手法におけるアイテムの近傍数は、精度が最も高くなった 610 とした。図 1 に結果を示す。α が 0 に近いときは、提案手法は従来手法よりも意外性が高く、1 に近い場合は精度が高くなることが確認できた。また、α が 0.2 から 0.4 で精度・意外性共に提案手法が上回る結果となった。

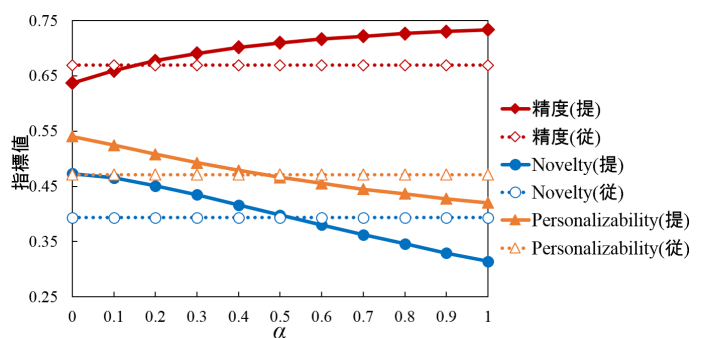


図 1: α による精度・意外性の変化

4 おわりに

本稿では、アソシエーションルールを用いた推薦手法において、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、推薦システムの意外性を向上させることができること、また調整項 α により、精度と意外性との重みを調整できることを示した。

参考文献

- [1] 神島 敏弘：推薦システムのアルゴリズム (1) ~ (3), 人工知能学会誌, Vol.22, No.6~Vol.23, No.2, 2007-2008 in large databases, 20th VLDB, pp.487-499, 1994
- [2] 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
- [3] 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, 情報処理学会誌 数理モデル化と応用, Vol.6, No.1, pp.111-118, 2013
- [4] the original MovieLens dataset from GroupLens research group: <http://www.grouplens.org>
- [5] Cantador, I., et al.: HetRec 2011, 5th ACM RecSys, pp.387-388, 2011
- [6] IMDb website: <http://www.imdb.com>
- [7] Rotten Tomatoes website: <http://www.rottentomatoes.com>
- [8] Sarwar, B., Karypis, G. Konstan, J., et al.: Item-based collaborative filtering recommendation algorithms, 10th WWW, pp.285-295, 2001