

推薦システムのための潜在的嗜好推定方法について

前田 優[†] 白川 真一^{††} 大原 剛三^{††}
[†] 青山学院大学大学院理工学研究科 ^{††} 青山学院大学理工学部

1. はじめに

インターネットの普及に伴い Web 空間上には大量の情報が存在し、推薦システムという技術が重要視されている。既存の推薦システムではユーザの嗜好を考慮して推薦アイテムを決定しているが、それがユーザにとって思いがけないアイテムであるかという意外性も必要であると指摘されている[1]。そこで著者らはこれまでにユーザ自身が気付いていない、もしくは意識していないアイテムの特徴に対する嗜好を潜在的嗜好と定義し、潜在的嗜好を用いて意外性の高いアイテムを推薦する方法を提案している[2]。文献[2]では、被験者実験によって提案手法と既存の協調フィルタリングの推薦結果を比較したが、提案手法の大きな有用性を示すには至っていない。本稿では提案手法における潜在的嗜好の推定精度を求め、提案手法の特性を詳細に検討する。

2. 潜在的嗜好推定方法

提案手法では、ユーザのアイテム評価履歴から作成したユーザプロフィールにおいてユーザの潜在的嗜好を推定する。ユーザプロフィールはユーザがアイテムに与えた評価値と、評価値の定義域の中央値との差をユーザの嗜好の度合い(以下、嗜好度)として特徴ごとに累積し、-1 から 1 の範囲で正規化する。次に、ユーザプロフィールからピアソン相関係数を用いてユーザ間類似度を計算し、協調フィルタリングで利用されている評価値推定方式を用いて潜在的な嗜好度を推定する。

3. 実験設定

潜在的嗜好の推定精度を検証するために、MovieLens Data sets 100k[3]を用いた実験を行った。まず、全データを用いてユーザのプロフィールを作成する。次にあるユーザの 1 つの特徴(ここでは、ジャンル)に対する嗜好度を 0 に置き換え、提案手法を用いて推定する。推定結果と元の値との差を誤差とし、推定対象ユーザにおける全ジャンルに対する誤差を計算する。943 名の 1,682 タイトルの映画に対する 1 から 5 の 5 段階の評価値からなる本データセットのうち、今回は評価数が 40 件以上の 644 名のデータを用いた。また推定対象ユーザは全ジャンルを視聴しているユーザ 89 名とし、18 個のジャンルに対する嗜好度を推定し誤差を計算する。

4. 実験結果

推定値全体の平均誤差は約 0.17 となった。また正解の嗜好度ごとの平均誤差を図 1 に、ジャンルごとの平均

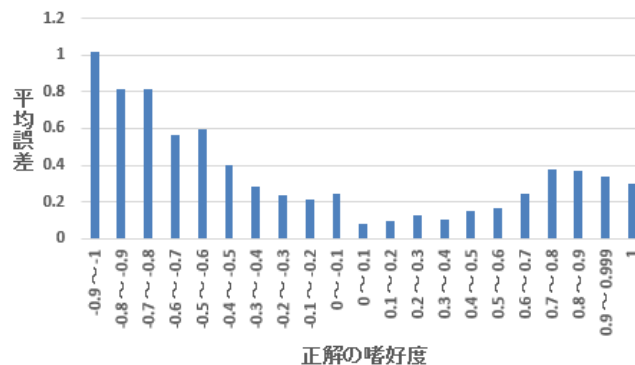


図 1. 正解の嗜好度ごとの推定嗜好度の平均誤差

表 1. ジャンルごとの推定嗜好度の平均誤差

Action	0.22	Documentary	0.16	Mystery	0.11
Adventure	0.13	Drama	0.27	Romance	0.14
Animation	0.14	Fantasy	0.27	Sci-Fi	0.13
Children's	0.23	Film-Noir	0.11	Thriller	0.18
Comedy	0.29	Horror	0.17	War	0.12
Crime	0.13	Musical	0.18	Western	0.11

誤差を表 1 に示す。図 1 より、嗜好度が -1 に近くなるにつれ誤差が大きくなるのがわかる。これは、全体としてプロフィールの嗜好度が正の値に偏っているため、少ない負の嗜好度の推定精度が下がったものと考えられる。しかし、-1 付近の嗜好度は推薦の際に考慮されにくいいため、実際に推薦する際にはこの誤差の影響は小さいと考えられる。また、表 1 より、ジャンルごとの平均誤差はプロフィールの値ごとの誤差よりも大きな差は見られないが、他と比べて誤差の大きいジャンルが幾つかあるのがわかる。データを確認したところ、誤差が大きいジャンルはそのジャンルを含むアイテムに対する評価数が少ないという傾向が見られた。

5. まとめ

本研究では推薦システムの意外性向上を目的に、ユーザの潜在的嗜好の推定手法を提案している。今回の実験で潜在的嗜好の推定の問題点を明らかにした。今後はこの問題を解決し、潜在的嗜好を用いた推薦手法の意外性を向上させることが課題となる。

参考文献

- [1] 神鷹敏弘, “推薦システムのアルゴリズム”, 人工知能学会誌, Vol. 22, No. 6, pp. 826-837 (2007).
- [2] 前田優, 白川真一, 大原剛三, “推薦システムにおける意外性向上のための潜在的嗜好の抽出”, FIT2013(第12回情報科学技術フォーラム), F-015, pp. 317-320 (2013).
- [3] GroupLens, <http://grouplens.org/datasets/movielens/>