

特定アイテムへのフィードバックを重視した 協調フィルタリング手法の提案

平井 遥斗

富山県立大学 情報システム工学科

2024 年 4 月 21 日

はじめに

近年、インターネット利用者の増加に伴い、大量の情報がインターネット上に流通するようになった。

一般に、精度の高い推薦にはユーザの情報が大量に必要なが、近年ではプライバシー保護の観点からユーザ情報の取得に対する目が厳しくなっている。

本研究の目的

本研究では、特定の少数アイテムへのフィードバックを重視した協調フィルタリングの学習アプローチを検討する。

人気アイテムを重視した場合とランダムに選択したアイテムを重視した場合を比較する評価実験を行い、提案手法の有用性について考察する。

協調フィルタリング

現在、協調フィルタリングはインターネットショッピングをはじめとする多くのオンラインシステムに実装されている。

協調フィルタリング 嗜好が類似しているユーザを発見し、その類似ユーザが好むアイテムを対象ユーザに推薦する。

Matrix Factorization(MF)

ユーザ i の因子ベクトル： p_i アイテム j の因子ベクトル： q_j
予測値 \hat{r}_{ij}

$$\hat{r}_{ij} = p_i \times q_j \quad (1)$$

目的関数 L

$$L = \sum_{i,j} (r_{ij} - \hat{r}_{ij})^2 \quad (2)$$

概要

プローピングアイテムを重視した学習方法を提案する。

バイアス項を用いることで、評価を付けられやすいアイテムや高評価をつけやすいユーザといった、ユーザやアイテムの傾向を表現することができ、評価の偏りを考慮することができる。

Baised-Matrix Factorization(B-MF)

全評価値の平均： μ ，重み： w_j ，正規化係数： λ

i 番目のユーザ、 j 番目のアイテムのバイアス： a_i 、 b_j

予測値 \hat{r}_{ij}

$$\hat{r}_{ij} = p_i \times q_j + \mu + a_i + b_j \quad (3)$$

目的関数 L

$$L = \sum_i \left\{ \sum_j e_{ij}^2 * w_j \right\} + \lambda (\|P\|^2 + \|Q\|^2 + \|A\|^2 + \|B\|^2) \quad (4)$$

$$e_{ij} = r_{ij} - \hat{r}_{ij} \quad (5)$$

実験概要

データセットとして Movielens を用いて提案手法を適用する。

人気アイテムを重視した場合とランダムに選択したアイテムを重視した場合で予測精度や多様性、人気バイアスといった観点で違いがみられるか比較を行い、提案手法の有用性について考察する。

実験

評価実験では、「人気度の高いアイテム 100 件」と「ランダムに選択したアイテム 100 件」をそれぞれプロービングアイテムとみなしてモデルを構築した。

表 1: 評価実験で設定したパラメータ

学習率 γ	0.02
正規化係数 λ	0.02
因子数 k	100
エポック数	30

実験結果

予測精度として平均絶対値誤差 (MAE) を用いた。

MAE : 実際の評価値と予測平均値の差を表している。

$$MAE = \frac{1}{|R_{test}|} \sum_{i,j \in R_{test}} |r_{ij} - \hat{r}_{ij}| \quad (6)$$

表 2: 人気度の高いアイテム 100 件をプロービングアイテムとしたモデルの実験結果

#other_item	100	1,000	ALL
MAE	0.896	0.769	0.754
corr_num	-0.142	-0.135	-0.140
corr_rate	-0.153	-0.114	-0.097

表 3: ランダムに選択したアイテム 100 件をプロービングアイテムとしたモデルの実験結果

#other_item	100	1,000	ALL
MAE	0.922	0.772	0.754
corr_num	-0.091	-0.070	-0.077
corr_rate	-0.067	-0.020	-0.014

実験結果

表 2, 3 において *corr_num*, *corr_rate* はプロービングアイテムが含まれる個数あるいは割合と MAE の Pearson 相関を表している.

- *#other_item* が 100 のときに, 表 2 の方が MAE の増加が抑えられていることから, 人気アイテムの方がプロービングアイテムとして適している可能性がるといえる.
- *corr_num*, *corr_rate* においては表 2 の方が負の相関が強いことがわかる.

まとめ

協調フィルタリングにおいてプロービングアイテムを重視する学習方法を提案した。

評価実験では、プロービングアイテムの割合を高めることでそれを重視したモデルが構築されていることを示した。

今後はプロービングアイテムの決定方法について検討する予定である。