

# 協調フィルタリングを用いた本推薦プログラムの試作

A prototype of book recommendation system that uses cooperated filtering

谷 研究室 榎本雄次  
Enomoto Yuji

## 概要

アイテムベース, ユーザーベースの協調フィルタリングを用いた推薦対象ユーザーへのアイテムを表示するプログラムを PHP を用いて試作した. 本プログラムはデータベースを利用している.

## 1 はじめに

いまや急激なインターネットの普及によって必要な情報を手軽に入手できるようになった. しかしその一方で大量の情報の中からユーザーが必要な情報だけを取得するには多くの時間と手間がかかるようになってしまった.

近年, このような大量の情報の中から必要な情報を抽出する例として, レコメンドシステムが挙げられる. 推薦システムとは, ユーザーの趣味嗜好, 購入履歴からユーザーがまだ購入, レンタルしたことがなく, きっとそのユーザーが満足するであろう商品を推薦, 提案しようというものである.

レコメンドシステムでは, 協調フィルタリングなどの技術が使われている. Amazon や楽天, livedoor など数々のオンラインショッピングサイトで使われている. その予測にはレコメンデーションの手法の一つである協調フィルタリングが用いられる. 協調フィルタリングとは自分の嗜好に関するデータをネットワーク上で他人と共有することで自分と趣味が近い人を捜し, その人たちの情報選択状況を勘案してフィルタリングする手法である. 趣味を同じくする友人の本棚には自分にとって興味深い本が並んでいると思われる. 計算機上でこのような手法を利用すれば有用な場面が多いと思われる.

昨年の本研究室卒業研究 [5] では, ユーザーベース, アイテムベースのプログラムの実装, そして評価値に対してユーザーのジャンルの好みを加えた方法に対して評価を行った. 評価に用いたデータは MovieLens の映画情報データセットである.

今回はこの技術を利用して, Book-Crossing Dataset [3] という洋書のデータセットを基にプログラムの実装, そしてそのプログラムを実行するインターフェイスの開発を試みた.

## 2 協調フィルタリングとは

協調フィルタリング ( Collaborative Filtering, CF ) とは, 多くのユーザの嗜好情報を蓄積し, あるユーザと嗜好

の類似した他のユーザの情報を用いて自動的に推論を行う方法論である. 趣味の似た人からの意見を参考にするという口コミの原理に例えられることが多い.

例えば, ユーザ A がアイテム X を好むとすると, アイテム X を好む別のユーザ B が好むアイテム Y を探し出し, ユーザ A もアイテム Y を好むのではないかと, という推論をコンピュータによって自動的に行う. 実装にはユーザ同士の類似度を, 同じアイテムにつけた評価の相関係数などによって表して類推に利用することが多い.

Amazon の推薦機能に協調フィルタリングが採用されたことによって有名になった技術である. 1992 年に David Goldberg らの論文 [4] で登場した.

### 2.1 アイテムベース

アイテムベースとは, 各アイテム間の類似度を求め, 推薦対象のユーザーの未評価値を他のアイテムとの類似度とそのアイテムの評価値を用いて予測するもので, 2005 年に Daniel Lemire [1] らによって提案された.

### 2.2 ユーザーベース

ユーザーベースとは, 推薦対象のユーザーと他のユーザー間の類似度を求め, 推薦対象のユーザーの未評価値を他のユーザーとの類似度とそのユーザーの評価値を用いて予測するもので, Badrul Sarwar [2] らの論文で発表された.

## 3 実験

### 3.1 研究概要

協調フィルタリングのアイテムベース, ユーザーベースを用いて本推薦プログラムのインターフェイスを開発.

### 3.2 データセット

	データベース
評価数	500000 件
アイテム数	203581 件
ユーザー数	46517 件

今回は Book-Crossing Dataset[3] のデータセットを利用し、データベースを作成。評価数は 5 段階。評価数が 1 以上のユーザーに関してユーザー ID, アイテム ID, 各アイテムに対する評価を基本データセットとしている。

### 3.3 実験環境

使用言語	PHP
サーバー環境	Apache HTTP Server 2.2
データベース	MySQL

### 3.4 実験方法

html を利用しある特定のユーザー ID を入力し PHP を用いて協調フィルタリングのアイテムベース, ユーザー

ベースを利用した外部プログラムを実行する。

データベースを利用しそのユーザーの嗜好に合わせた推薦アイテムの ISBN, 作品名, 作者名, 発行年, 出版社を表示。

## 4 結果・考察

表示やデータベース間に異常はなかったが,PHP を実行する際にデータ量が多いと html で表示するのに時間がかかる。

また表示画面の装飾が乏しかったのもう少し html を利用して見た目に力を入れてもよかった。

## 参考文献

- [1] Daniel Lemire, Anna Maclachlan: Slope One Predictors for Online Rating-Based Collaborative Filtering(2005)
- [2] Badrul Sarwar, George Karypis, Joseph Konstan, John Riedl: Item-Based Collaborative Filtering Recommendation Algorithms(2001)
- [3] Book-Crossing Dataset <http://www.informatik.uni-freiburg.de/cziegler/BX/>
- [4] David Goldberg, David Nichols, Brian M. Oki, Douglas Terry: Using collaborative filtering to weave an information tapestry(1992)
- [5] 谷 研究室 田中健太 協調フィルタリングを用いた映画予測プログラムの試作  
<http://www.tani.cs.chs.nihon-u.ac.jp/g-2009/tananori/dissertation/dissertation/resume/resume-presenter.pdf>