

# 閲覧者の少ない映画のためのハイブリッド型推薦システムの研究 A hybrid recommender system for movies with few viewers

渡邊 郁美<sup>†</sup> 今野 将<sup>‡</sup>

Ikumi Watanabe<sup>†</sup> Susumu Konno<sup>‡</sup>

## 1 はじめに

近年インターネットの発展により映画の配信サービスが急速に普及している [1]. それに伴い利用者数も年々増加傾向にあり, さらなる需要が見込まれる. 利用者の多くは, キーワード検索を用いて出てきたものや, 他者のレビューを参考に閲覧する映画を判断している.

一方, 各サービスにて登録されている映画の数は数千~数万件と膨大であり, 全ての映画情報を利用者自身が検索し網羅することは困難であると言える. そこで多くの配信サービスで, 利用者に最適な映画を推薦する推薦システムが活用されている. 推薦システムの特徴として設計や実装には現在4つのアプローチの仕方があり, 中でも主流とされているものが協調フィルタリング, ハイブリッド型の推薦システムである.

協調フィルタリングは, ユーザの行動履歴をもとにユーザとユーザ間, アイテムとアイテム間, もしくはユーザとアイテム間の類似度を計算し, 推薦対象のユーザに未評価アイテムを推薦する技術である.

しかしながら, 既存のシステムにはいくつかの問題点があり, 代表的な例としては小カバー率問題が挙げられる. 小カバー率問題とはユーザ評価の少ないアイテムは推薦が難しくなり, 推薦内容が評価の多いものに偏ってしまう問題である. この問題の解決策として複数のアルゴリズムによるハイブリッド型のシステムが提案されている. 既存のハイブリッド型システムの研究はマイナーアイテムの除去を目的としたものが主流とされ, 推薦には焦点を当てていない.

本研究ではユーザの行動履歴のうち, 特に閲覧数が少ないマイナー映画でも推薦されるシステムを考案し, 有用性について検証することを目的とする. また上記の目的を達成するために, モデルベース協調フィルタリング, およびコンテンツベースフィルタリングを組み合わせたハイブリッド型推薦システムに対し, マイナー表現のパラメータを模索し, 既存の推薦システムとの精度の比較を行う.

## 2 ハイブリッド型推薦システムの設計

提案するシステムの具体的な流れを図1に示す. はじめに, ユーザの行動履歴と映画のメタデータをそれぞれ

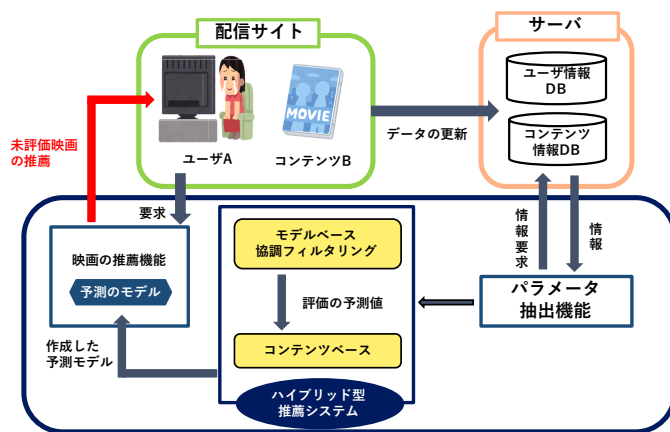


図1: ハイブリッド型推薦システムを用いた映画推薦

れ個別のデータベースに保存する. ここで保存されるユーザの行動履歴は映画に対する評価, ユーザが印象としてつけるタグデータである. また映画のメタデータとして, 各映画のID, タイトル, ジャンル, 評価数が保存されている. 次に各データベースに対し, パラメータ抽出機能が一定時間ごとに情報要求を出し, データベース上から情報を受け取る. 受け取った情報から必要なパラメータを抽出し, 各推薦システムでユーザ間, またアイテム間の類似度等の計算を行い, 推薦のモデルを作成する. 作成したモデルをもとに, ユーザがサービスにおいてログイン等の行動を起こした際に, 未評価の映画をトップ画面等に推薦する.

## 3 ハイブリッド型推薦システムの試作

本研究ではハイブリッド型システムの精度の比較のために, モデルベースの推薦システム, コンテンツベースの推薦システム, ハイブリッド型推薦システムの3種類のシステムの設計を行った. ここでのハイブリッド型推薦システムとは, 図2のようにモデルベース協調フィルタリングとコンテンツベースフィルタリングを組み合わせたものである. なお開発にはpython3.7.3を使用し, 各学習データ, および検証データとしてMovieLens[2]のデータセットを用いる.

モデルベース協調フィルタリングは, WNMFを用いてユーザが未評価の映画に対しどのような評価値をつけるか評価の予測値を計算する. このうち未評価の予

<sup>†</sup>千葉工業大学大学院先進工学研究科知能メディア工学専攻

<sup>‡</sup>千葉工業大学先進工学部知能メディア工学科

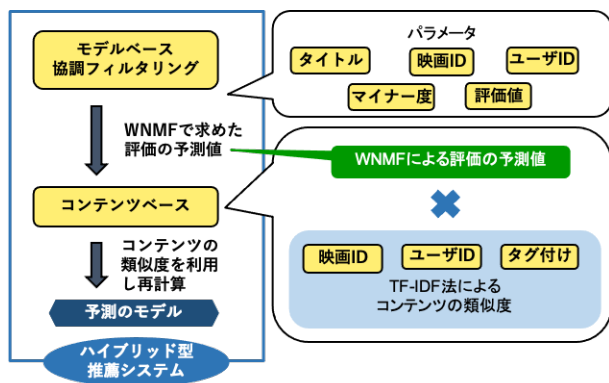


図2: ハイブリッド型推薦システム

測値を全て降順で返し、上位10個の映画を推薦するシステムである。ここで用いるパラメータは、映画のタイトル、映画ID、ユーザID、評価値、マイナー表現の閾値の5つとする。

コンテンツベースフィルタリングは、映画のメタデータからTF-IDF法を用いて各映画の特徴ベクトルを算出し、映画同士の類似度をコサイン類似度によって計算する。ここでは、映画タイトル、ユーザID、ユーザが各映画に対してつけたタグデータをパラメータとして用いている。

ハイブリッド型推薦システムでは、Jannach[3]らがまとめている手法のうちパイプライン型を用いる。パイプライン型の設計では、図2に示す通りモデルベース協調フィルタリングで算出した評価の予測値を、コンテンツベースで算出した各映画の類似度の重みをかけて計算することで、ユーザの好みを反映した上での最終的な予測評価値を算出する。このうち未評価の予測値を全て降順で返し、上位10個の映画を推薦する。

#### 4 試作システムの検証

試作したシステムのうち、モデルベース協調フィルタリングのみの精度とハイブリッド型を用いた時の映画の推薦精度の比較実験を行った。検証方法として、以下の2種類のデータセットを作成し、推薦された映画のタイトルを比較する。

- dataA: 評価数を一定数で足切りしたデータ
- dataB: dataA の評価の総数が7割になるまで削除したデータ

上記のデータは全て学習用とするため、システムの精度評価を行うために評価値データの多い正解のデータと、評価値のデータ量を減らしマイナー映画を増やし

表1: 推薦精度の比較

	協調フィルタリング	ハイブリッド型推薦
ID: 2	20%	90%
ID: 11	0%	70%

た比較用のデータを作成している。ただしデータのスパース性の問題から、dataA の評価数は一定以上の値で足切りを行なっている。この時、各システムにおけるdataA の推薦内容を疑わないものとし、dataA の推薦内容とdataB の内容の一致率を計る。一致率は式(1)の通りとし、値が高いほど精度が良いものとする。

$$\text{一致率 (\%)} = \frac{\text{一致した映画の個数}}{\text{推薦された映画の個数} * 100} \quad (1)$$

なお比較するユーザは協調フィルタリングのみの精度を測るときにランダムに2人選出し、ハイブリッド型の精度では協調フィルタリングと同様の2名を選出する。

検証の結果、各推薦精度は表1の通りである。

表1より、モデルベース協調フィルタリングのみを用いた時より、提案システムを用いた時の方が両者共に一致率が高いことが見て取れる。よって、ハイブリッド型の推薦システムの有用性が確認できた。

#### 5 終わりに

現在までにマイナー映画を推薦するシステムとして、ハイブリッド型推薦システムの有用性の検証を行った。今後の課題として、本研究で用いた評価方法の妥当性の検証や、評価数の足切りを行わないデータでの検証が必要であると考えられる。また、マイナー表現をパラメータとして組み込んだ際の比較検証も行う必要がある。

#### 参考文献

- [1] 株式会社ICT総研. 2019年有料動画配信サービス利用動向に関する調査. <https://japan.zdnet.com/release/30299626/>.
- [2] MovieLens. <https://eigahitottobi.com/article/76711/>.
- [3] Sheng Zhang, Weihong Wang, James Ford, and Filia Makedon. Learning from incomplete ratings using non-negative matrix factorization. In *IN PROC. OF THE 6TH SIAM CONFERENCE ON DATA MINING (SDM)*, pp. 549–553, 2006.