

# 複合型アルゴリズムを用いた映画推薦システムの実装と評価

坂本竜太<sup>†</sup> スティーラ・パンティラヌラク<sup>‡</sup> 吉見哲哉<sup>†</sup> 後藤慎平<sup>†</sup> 山谷一貴<sup>†</sup> 伊藤壘<sup>†</sup>

柿崎淑郎<sup>††</sup> 辻秀一<sup>‡</sup>

<sup>†</sup>東海大学電子情報学部 <sup>‡</sup>東海大学情報理工学部 <sup>††</sup>東海大学連合大学院理工学研究科

## 1 はじめに

インターネットの普及によりユーザが出会う情報の量が日々増加している。そんな中、推薦システムが多くの企業で採用され始め、その技術が多く研究されているのは、ユーザが有益な情報を取捨選択する際に補助を必要と感じ、電子商取引サイト等のインターネット情報提供サービスにおいても情報推薦システムは重要不可欠な要素であることを現していると言える。

本研究では複合型アルゴリズム[1]を用いて映画は比較的評価しやすい情報であると考え、映画推薦に特化したシステムを構築した。ユーザが情報の発信と推薦をスムーズに行えるよう工夫したシステムを構築し、運用実験を行い、情報発信・推薦の容易なインターフェースと検索方法を模索し、現実に利用できるシステムの実装を行う。

## 2 従来方式と実装方式

推薦システムとは情報収集を行うユーザに対し、システムに登録されたルールあるいは興味・関心の傾向に応じて、そのユーザの嗜好に合う情報を推薦するシステムである。推薦情報をユーザの嗜好情報を基に取捨選択する際に情報フィルタリングが必要であり、一般的には、コンテンツに基づくフィルタリング(内容フィルタリング)(Content-based Filtering)と協調フィルタリング(Collaborative Filtering)の2種類がある[2]。しかし、これらの従来方式では以下のような問題がある。

- 内容フィルタリング  
推薦情報の基準はユーザが過去に入力した情報に関連したものであるため、推薦情報が単純化してしまう。
- 協調フィルタリング

利用者から発信される情報によって推薦情報の決定をするため、発信される情報量が不足すると、推薦情報の精度が低くなってしまふ。

本システムは上記の課題を解決するために、上記2つのフィルタリングを組み合わせた複合型アルゴリズム[1]を用い、実装方式とする。複合型アルゴリズムは協調フィルタリングを軸にし、ユーザからの発信情報不足による欠損データを内容フィルタリングによって補完することで利用者数の少ない場合でも精度の高い推薦を可

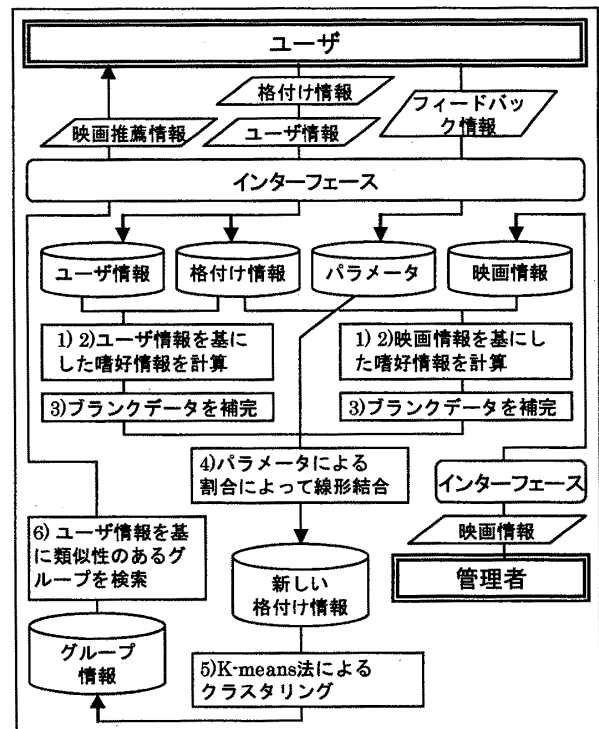


図1 システム全体概要

能とする特徴を持つアルゴリズムである。

本実装方式では以下の4種類の情報をを用いる。

- ユーザの個人情報(格納したユーザ情報)
- 映画の詳細情報を格納した映画情報
  - ユーザが映画を格付け評価した格付け情報
- 推薦された映画に対し期待との一致度を割合変換したパラメータ

処理概要は以下のとおりである。また、図1の1)~6)に流れを示す。

### 1) 類似度計算

ユーザ情報と映画情報をクラスタリングする。クラスタリングされた各データ間、格付け情報からのユーザ間と映画間の類似度を求める。

### 2) 線形結合

クラスタリングデータの近似値と、格付け情報のデータの近似値を線形結合し、ユーザベースの嗜好情報、映画ベースの嗜好情報を求める。

### 3) 予測処理

線形結合したデータに対し、不足する格付けデータ(ブランクデータ)の予測処理を行い、ユーザベースの嗜好情報、映画ベースの嗜好情報を補完する。

### 4) パラメータによる線形結合

3)で求めた2つの格付け情報をユーザ別のパラメータによる割合で線形結合し、新しい格付け情報を作成する。

Development and Evaluation of Movie Recommendation System with Hybrid Algorithm

<sup>†</sup>Sakamoto,R., <sup>‡</sup>Sutheera,P., <sup>†</sup>Yoshimi,T., <sup>†</sup>Goto,S., <sup>†</sup>Yamatani,K., <sup>†</sup>Ito,R., <sup>††</sup>Kakizaki,Y., <sup>‡</sup>Tsuji,H.

<sup>†</sup>School of Information Technology and Electronics, Tokai Univ.

<sup>‡</sup>School of Information Science and Technology, Tokai Univ.

<sup>††</sup>Graduate School of Science and Technology, Tokai Univ. United Graduate School

### 5) クラスタリング

新しい格付け情報を K-means 法によるクラスタリングし、グループ情報を作成する。

### 6) 類似グループ抽出処理

ユーザの格付け情報をグループ別に類似度計算し、ユーザと一番嗜好の近いグループを発見する。

## 3 実装する映画推薦システム

複合型アルゴリズムを用いた映画推薦システムの利用の流れと機能概要を説明する。開発環境を表 1 に示す。

表 1 開発環境

ソフトウェア	ソフトウェア構成内容
OS	Vine 3.2
プログラミング言語	PHP 5.2.0
データベースサーバ	MySQL 4.0.25
Web サーバ	Apache 2.0.50

ユーザは本システムを利用する際、ユーザ情報としてニックネーム、メールアドレス、パスワード、年齢層、性別、職種を登録する必要がある。次回以降、メールアドレスとパスワードで認証を行いログインする。ログイン後、メンバーページを中心に、映画推薦に関する機能を利用できる。

本システムの映画推薦に関する機能は以下のとおりである。

#### (1) 格付け評価推進機能

推薦精度に関連する発信情報量に影響を及ぼすため、ユーザが今までに観た映画を思い出させるきっかけを作り、発信情報量をなるべく増やすことを目的とした機能である。ユーザが一番閲覧する機会があるメンバーページの一部分に設置されている。以下の各機能が設置したものである。

- ユーザが評価した映画のタイトルや詳細をランキング形式で見ることができ評価できる。
- ユーザが評価した映画タイトルを年代別にランダムに表示するシステムを設置し、ユーザが昔見た記憶に薄い映画タイトルを見つけることができる。

#### (2) 格付け評価機能

映画の評価は、最大 5 タイトルまで同時に評価できる。映画のタイトル検索はタイトルの一部を入力すると、その文字が含まれた映画タイトルの候補がページに羅列される、"suggest.js"を用いた。そこから評価するタイトルを選択する方式を採用した。この機能を追加することにより、ユーザが登録する映画タイトルの揺らぎをなくすことができる。評価項目は点数(5段階評価)、その映画を見た状況「いつ」「どこで」「だれと」観たのかを選択し評価する。入力された情報は各ユーザごとに格付け情報テーブルに記録される。

#### (3) 推薦映画表示機能

ユーザ情報と映画格付け情報を基に類似したユーザごとにクラスタリングを行い、分類されたグループで高評価の映画をそのグループ内のユーザに推薦する。

#### (4) 推薦映画フィードバック機能

ユーザは推薦された映画に対し、ユーザ自身の興味との一致具合を 3 段階のフィードバック情報としてシステムに返す。フィードバック情報は、ユーザベースの格付け情報と映画ベースの格付け情報の推薦精度を割合として算出させる。算出されたデータはパラメータテーブルに格納され、新しい格付け情報テーブルを作成する際に利用される。

#### (5) 管理者機能

映画管理の面で、新しい映画タイトルを追加する際、同じ映画タイトルで全くの別の作品を二重に登録しないように、映画の制作年号をタイトルの後ろに付加し、識別できるようにした。これにより、管理の面でも同じタイトルの重複ミスを減少させる。

## 4 評価

### • 評価項目

ユーザインターフェースおよび推薦された映画全体と映画個別の推薦精度、推薦映画フィードバック機能を評価対象とする。

### • 評価方法

実装したシステムで運用実験を行って、ユーザに利用してもらい、ユーザインターフェースと推薦精度を評価する。

## 5 おわりに

本稿では複合型アルゴリズムを用いた映画推薦システムを実装・評価した。

格付け評価を推薦する機能類を実装したが、映画を多く観ているのにも関わらず、思い出せないという状況から完全に脱却させるには至らなかった。また、評価実験の被験者が幅広い層でなかったため、格付けされる映画が一部に集中することで、推薦映画が偏ってしまう問題が存在し、評価実験に関する信憑性に疑問が残る結果となってしまった。

膨大な情報を模索するユーザに有益な情報を取捨選択するシステムとして、従来方式の問題を解決したアルゴリズムを採用した本システムは映画を媒体として実現したが、今後、情報発信・推薦の容易なインターフェースと検索方法を模索し、現実に利用できるシステムを構築したい。

## 参考文献

- [1] Suthera, P. Tsuji, H. A Multi-Clustering Hybrid Recommender System, Proc. of the 7th IEEE International Conference on Computer and Information Technology, pp.223-234, 2007.
- [2] 土方嘉徳. 嗜好抽出と情報推薦技術. 情報処理, Vol.48, No.9, pp.959-961, 2007.

\*<http://www.enjoyxstudy.com/javascript/suggest/>