

複合型アルゴリズムを用いた映画推薦システムの提案

伊藤 塁† 吉見 哲哉† スティーラ・パンティラヌラク‡ 坂本 竜太†‡
柿崎 淑郎‡‡ 辻 秀一†

† 東海大学 ‡ モンクット王ラカバン工科大学 †‡ NEC ‡‡ 東京理科大学

概要：情報推薦システムの実現方式としてコンテンツベースフィルタリング方式と協調フィルタリング方式の2つがある。前者の方式では推薦情報が単純化してしまうという問題があり、後者には事前に一定以上のユーザ情報が入っていないと推薦精度が落ちる問題があった。本報告では、これらの問題を解決するための複合型アルゴリズムを用いた映画推薦システムを提案する。

A Proposal of Movie Recommendation System Using Hybrid Recommender Method

Rui Ito† Tetsuya Yoshimi† Sutheera Puntheeranurak‡ Ryota Sakamoto†‡
Yoshio Kakizaki‡‡ Hidekazu Tsuji†

† Tokai University ‡ King Mongkut's Institute of Technology Ladkrabang
†‡ NEC ‡‡ Tokyo University of Science

Abstract : There are two methods, Content-based Filtering and Collaborative Filtering for realizing recommender system. The first method has the problem that a simple recommender-information is provided. The second method has the problem that the precision of the recommendation is low in case that there is few user information. We propose movie recommendation system using hybrid recommender method to solve the problem.

1. はじめに

現在、インターネットの普及によりユーザが
出会う情報の量は日々増加しており、ユーザは
この膨大な情報を扱うことが難しくなっている。
この問題の解決案の一つとして、情報推薦シ
ステムと呼ばれる情報を自動的に取捨選択し提示
するシステムがある、また、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な要素となっ
ている。しかし、情報推薦システムによる推
薦情報が必ずしもユーザの望むものではな
いのが現状である。

本報告では、筆者らが提案した複合型アル
ゴリズム[1]を用いて、ユーザが情報の発信
と推薦をスムーズに行える映画推薦システ
ムを提案する[2]。

2. 従来方式

情報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[3]。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 Fil
tering)と協調フィルタリング(Collaborative
Filtering)の2種類がある。しかし、これら
の従来方式では以下のような利点と問題点
がある。

2.1 内容フィルタリング

登録時のユーザ嗜好情報を基にして推薦
を行うため、システムを利用し始めたユー
ザにも一定の情報推薦を行うことができ
る利点がある。しかし、ユーザが事前に
設定した嗜好情報を元に推薦情報を決定
するため、ユーザの嗜好変化に追従する
ことが難しく、推薦情報が単純化し
てしまう問題点がある。

2.2 協調フィルタリング

情報推薦の際、参照されるのはユーザ自
身の嗜好情報と類似している他ユーザの
嗜好情報であるため、自分が想像もして
いないような推薦結果を得ることができ
る利点がある。しかし、複数のユーザか
ら発信される嗜好情報を元に推薦情報
を決定するため、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いう問題点がある。

2.3 従来の映画推薦システム

従来の映画推薦システムとして、2008年2月現在運用されている映画推薦システムであるFANCAST[4]とmovielens[5]、au one ラボ 映画推薦ツール[6]の3つを挙げる。各システムの特徴を以下に述べ、表1に示す。

(1) FANCAST

FANCASTは検索した情報やお気に入り登録した行動履歴から無条件に映画が推薦される。推薦対象にテレビ番組も含まれており、また推薦理由が提示される特徴がある。

(2) movielens

movielensはランダムに出てくる15件の映画を10段階で評価し推薦映画が映画属性で条件表示される。格付け評価する際に予想格付けが表示される特徴がある。アルゴリズムは協調フィルタリングを用いている。

(3) au one ラボ 映画推薦ツール

au one ラボ 映画推薦ツールは映画推薦のために、個人の気分に合わせて、感情別項目に5段階評価を入力するといった特徴があり、希望する気分や映画属性で条件表示することができる。また、事前に見たい、見たくない、リストが作れるため、無駄な推薦情報を取り除かれ、スムーズに映画が推薦される。アルゴリズムはベイジアンネットワーク[7]を用いている。

表1 従来映画推薦システムの特徴

	FANCAST	Movielens	au one ラボ 映画推薦ツール
入力データ	行動履歴 (検索やお気に入り登録)	10段階格付け評価 (最低格付け数を定義)	感情別 5段階格付け評価
出力データ	条件無く表示	映画属性で条件表示	希望する気分や映画属性で条件表示
アルゴリズム	不明	協調フィルタリング	ベイジアンネットワーク
特徴	TV番組も推薦対象、推薦理由を提示	格付け評価する際に予想格付けを表示	事前に、見たい、見たくないリストを作成できる。

従来方式の情報フィルタリングの問題点は、そのまま従来の映画推薦システムの問題点となっている。

上記で述べた movielens は協調フィルタリングを使用しているため、ユーザ登録時、15件もの映画を格付けしなければならない。これは一定のユーザの嗜好情報が蓄積を必要とする協調

フィルタリングの問題点をそのまま表しているといえる。また、au one ラボ 映画推薦ツールは条件入力を必要とするなど、推薦情報表示に手間がかかる。気分や感情など細かい設定をするのもひとつの方向性ではあるが、入力を煩わしく感じるユーザもいるはずである。

以上のように従来の映画推薦システムではいくつかの問題点がある。

3. 提案方式

本研究では、従来の情報推薦システムの問題点を解決する複合型アルゴリズムによる映画推薦システムを提案する。

3.1 複合型アルゴリズム

複合型アルゴリズムとは、コンテンツに基づくフィルタリングと協調フィルタリングを組み合わせる情報推薦を行う方式である。コンテンツに基づくフィルタリングは推薦情報が単純化してしまう点があり、協調フィルタリングはユーザからの発信情報が少ないと推薦精度が落ちてしまうという問題点がある。複合型アルゴリズムは協調フィルタリングを軸にユーザからの発信情報不足による欠損データをコンテンツに基づくフィルタリングによって補完することで、ユーザからの発信情報が少ない場合でも精度の高い推薦を行うことができるという特徴を持つアルゴリズムである。

本方式では以下の3種類の情報を用いる。

- ・ユーザ情報は、性別、年齢層、職種など。
- ・映画情報は、タイトル、ジャンル、俳優など。
- ・格付け情報は、タイトル、得点など。

複合型アルゴリズムの処理構成を図1に示す。処理は以下の(1)～(5)のように行われる。

(1) k-means クラスタリング

1a)と1b)で、ユーザ情報と映画情報をバイナリ化した後、k-means法によりクラスタリング処理を行う。

(2) 類似度計算

2a)と2d)で、クラスタリング処理されたデータでユーザ間、映画間の類似度を計算する。また、2b)と2c)で格付け情報を基にユーザ間、映画間の類似度を計算する。

(3) 線形結合

3a)と3b)で、2a)と2b)によって算出されたユーザ間類似度、2c)と2d)によって算出された映画間類似度を線形結合する。

(4) 予測処理

4a)と4b)で、結合したユーザ間および映画間の類似度からそれぞれの類似度を基に格付け情報内に存在する空白データの格付け点数を予測し、空白データを埋める。空白デー

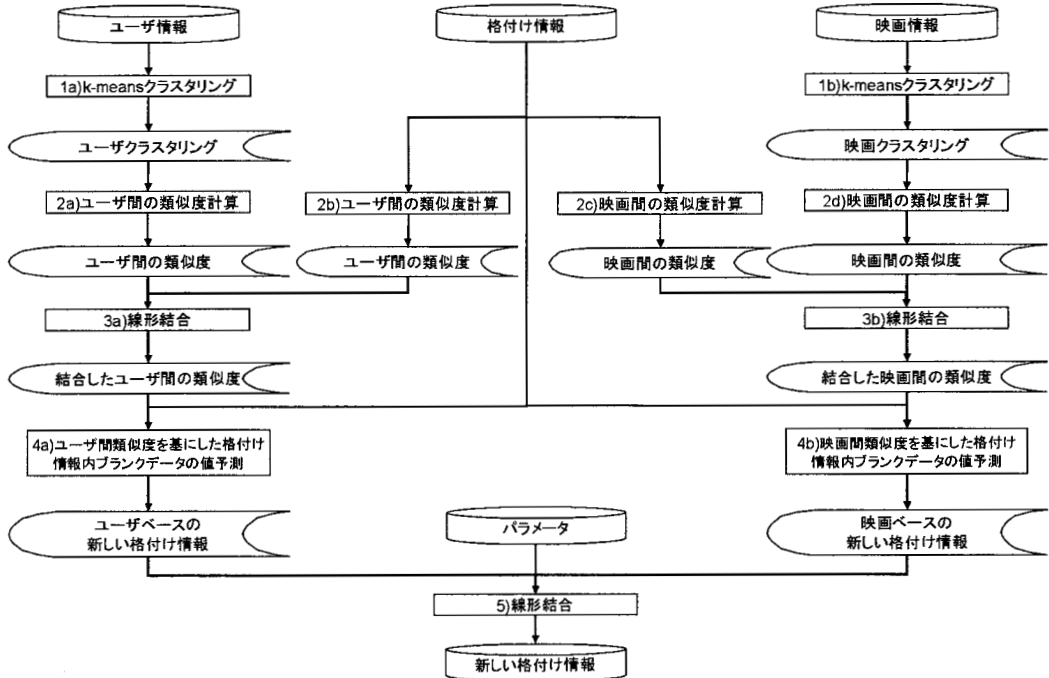


図1 複合型アルゴリズムの構成

タとは、あるユーザ A がある映画 B に対し格付けを行わなかった際の格付け情報の座標 (A, B) を指す。空白データが埋まった格付け情報を、ユーザ、映画それぞれを基にした新しい格付け情報とする。

(5) 補完データによる線形結合
ユーザおよび映画を基にした新しい格付け情報を線形結合し、新しい格付け情報を作成する。

3.2 提案システムの全体構成

本システムで推薦に用いる情報はユーザ情報、格付け情報、映画情報、フィードバック情報の4つである。それぞれの情報は、ユーザ情報登録処理、格付け情報登録処理、映画情報登録処理、フィードバック処理から、それぞれ入力される。4つの情報を入力情報として複合型アルゴリズムで計算した結果は新しい格付け情報として出力される。新しい格付け情報はクラスタリングされ、結果はグループ情報として出力される。グループ情報は類似グループ抽出処理による推薦情報を要求するユーザと類似したグループを抽出するために用いられる。類似グループ抽出処理は類似したグループが見つかったら推薦映画情報出力処理へ推薦映画の情報を渡し、推薦情報出力処理によって推薦映画情報を基に推薦映画を出力する。

図2に本システムの全体図を示す。

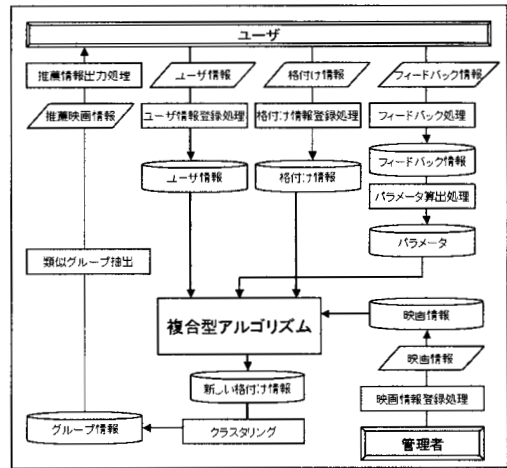


図2 映画推薦システム全体図

提案システムの各処理は以下のようなものである。

(1) ユーザ情報登録処理

初めて本システムを利用するユーザが行うのがユーザ情報登録機能である。ユーザ情報登録画面で“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”を入力する。入力された6つの情報はユーザ情報テーブルに保存される。図3にユーザ情

報登録インターフェース例を示す。“ニックネーム”、“メールアドレス”、“パスワード”はテキストフォームに直接入力する方式である。“性別”はラジオボタンで男性か女性を選択する方式である。“年齢層”、“職種”はプルダウン選択方式である。

図3 ユーザ情報登録インターフェース例

(2) 格付け情報登録処理

映画格付け情報登録画面から“タイトル”、“得点”、“いつ”、“どこで”、“だれと”の5つの情報をユーザに入力してもらう。5つの情報は格付け情報テーブルに保存される。図4に映画格付け情報登録インターフェース例を示す。“得点”は1点～5点の5段階である。同時に“いつ”、“どこで”、“だれと”その映画を観たシチュエーションも入力する。シチュエーションは不明とすることもできる。“タイトル”はサジェスト方式[8]を使用した。これは、ユーザが文字をテキストフォームに入力すると候補が画面に羅列され、その中からユーザが選択する方式である。“得点”、“いつ”、“どこで”、“だれと”はプルダウン選択方式である。

図4 映画格付け情報登録インターフェース例

(3) 映画情報登録処理

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) フィードバック処理

本システムのフィードバックは、ユーザが推薦された映画に対する期待との一致と興味を示す

ためにフィードバック情報として管理者に情報を送信する処理である。図5はフィードバックインターフェース例である。フィードバックはユーザが推薦された映画に対しての期待との一致具合を“ぜひ観たい! ”、“興味はある”、“興味なし”の3段階から選択し、送信する処理である。ユーザが選択したフィードバック情報は「ぜひ観たい!」を1、「興味はある」を2、「興味なし」を3、と1～3の数値に変換され、フィードバック情報テーブルに保存される。フィードバック情報テーブルに保存された情報はパラメータ算出処理に使用する。パラメータは新しい格付け情報算出時のユーザベースの新しい格付け情報と映画ベースの新しい格付け情報の依存割合である。すべてのユーザはPというパラメータ情報を持っている。パラメータの範囲は0～100で、初期値は50である。パラメータが100に近づくほど新しい格付け情報テーブル算出時、映画ベースの新しい格付け情報に依存する割合が高くなる。端的には、パラメータが100のときは、映画ベースの新しい格付け情報からのみで新しい格付け情報テーブルを算出し、パラメータが0のときは、ユーザベースの新しい格付け情報からのみで新しい格付け情報テーブルを算出する。

図5 フィードバックインターフェース例

(5) 推薦情報出力処理

ユーザ情報、格付け情報、フィードバック情報を複合型アルゴリズムで算出した結果、得られる推薦映画情報を推薦映画情報表示画面に出力する機能である。図6は推薦映画情報出力インターフェース例である。タイトル、英語タイトル、公開年、配給会社、ジャンル、監督、出演俳優を表示する。推薦度の高い順番に30件の映画を表示する。

図6 推薦映画情報出力インターフェース例

4. 試作・実験

複合型アルゴリズムを用いた映画推薦システムを下記表2のソフトウェア開発環境で試作した。

表2 ソフトウェア開発環境

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

ユーザに本システムを以下のように利用してもらい、実験を行った。まず初めに、ユーザ情報としてニックネーム、パスワード、年齢層、性別、職種を登録する。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、メンバーズページを中心に、先に示した映画推薦に関する機能を利用する。本システムを一ヶ月運用後、メンバーズページに実装した「アンケート機能」により、ユーザにシステムの利便性などのアンケートに回答してもらった。WEBアンケートの設問と回答結果を、表3に示す。なお、各設問は1から5点の5段階で答えてもらう形式になっている。

実施した結果、全ユーザ数113人中、20人から回答を得られた。回答者のユーザ属性は「男性」が71%、「21歳～25歳」が85%、「大学(院)生(理・工・情報系)」が81%である。

表3 アンケートの回答結果

設問	最小値(点)	最大値(点)	平均(点)	標準偏差(点)
問1.システム全体の使いやすさ	1	5	3.60	1.10
問2.推薦された映画の期待との一致具合	2	4	2.75	0.64
問3.デザインや文字の読みやすさ	1	5	3.85	1.14
問4.登録までの流れのよさ	3	5	4.15	0.88
問5.収録されているデータ量	1	5	3.50	1.32
問6.映画タイトルの検索のしやすさ	2	5	3.65	1.14

5. 評価

本研究はよりの確な情報推薦システムを目指した提案である。その評価として、推薦精度と使用性についてユーザからのアンケートを基に評価する。そして従来の映画推薦システムと本システムの比較を行う。

5.1 システムの使用性

表3のWEBアンケートの回答結果より、システムの使用性に関する設問においては、「登録までの流れのよさ」が平均4.15点と良好であり、以下「デザインや文字の読みやすさ」が3.85点、「映画タイトルの検索のしやすさ」が3.65点、

「システム全体の使いやすさ」が3.60点と、軒並み3点以上となった。よって、WEBシステムとしてのシステムの使用性は良いと評価されたといえる。

5.2 推薦精度

上記同様のWEBアンケート上での「推薦された映画の期待との一致具合」は、平均2.75点となり、平均点を基準で考えると芳しい結果とはいえない。しかし、1点と5点を付けた回答者がいないことと、回答者の55%が3点であった。また、フィードバック情報の集計は平均1.46と、推薦された映画に対して「興味ない」と判断したユーザが半数を占めた。この結果から、推薦された映画はユーザの期待に沿うタイトルもあり、逆に沿わないタイトルも表示されたと考えられ、協調フィルタリングの利点である意外性のある推薦が行えたといえる。そして、コンテンツに基づくフィルタリングの問題点である、推薦の単純化に対しある一定の効果を示したといえる。

5.3 従来の映画推薦システムと本システムの比較

比較対象の従来の映画推薦システムは“FANCAST”、“movielens”、“au one ラボ映画推薦ツール”である。システムの使用性は「ユーザ情報登録のインターフェース」、「映画格付け情報登録のインターフェース」、「推薦情報表示のインターフェース」、「フィードバックの有無」、「推薦精度」の5項目を比較評価した。

(1) ユーザ情報登録インターフェースの比較評価

FANCASTのユーザ情報入力には“メールアドレス”、“パスワード”の2項目である。movielensは“メールアドレス”、“ニックネーム”、“リアルネーム”、“パスワード”、“郵便番号”の5項目である。ユーザ情報登録時、最低15件の映画格付け情報登録が必要である。au one ラボ映画推薦ツールは“ニックネーム”、“生年”、“性別”、“未婚・既婚”、“職業”の5項目である。ユーザ情報登録時、最初にメールアドレスを入力し、送信されてきたメールからユーザ情報登録を行う。本システムは“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”の6項目である。入力項目は最も多くなってしまったが、アンケート結果より「登録までの流れのよさ」は平均4.15点と非常に高い評価を受けることができた。

(2) 映画格付け情報登録のインターフェースの比較評価

FANCAST は行動履歴から推薦を行うため格付け情報登録はない。movielens は格付けをJavaScript で視覚的に確認することができるためスムーズである。au one ラボ 映画推薦ツールは検索→格付けとページの移動が多く、スムーズとはいえない。本システムは suggest 方式を使用した。アンケート結果より「映画タイトルの検索のしやすさ」で平均3.65点とまずまずの評価を受けることができた。

(3) 推薦情報表示のインターフェースの比較評価

FANCAST は推薦された映画がタイトルだけでなくサムネイルを用いているため視認性も高く、映画の雰囲気等が、一目で伝わる。movielens は推薦された映画の推薦理由が記載されているため、ユーザが推薦された映画に対して「見る」、「見ない」という判断をより深く考えることができる。au one ラボ 映画推薦ツールは推薦された映画を「見たい」、「見たくない」というリストに振り分けられるため、無駄な推薦情報を省くことができる。本システムは、すべての推薦結果が表示され、また表示された各推薦映画を、そのページ上でフィードバックすることが可能でユーザの手間が軽減される利点がある。しかし、すべての推薦情報が羅列されるため、推薦情報全体としては見にくく、快適に情報が伝わりづらい。

(4) フィードバックの有無の比較評価

FANCAST、movielens はフィードバック機能を有しておらず、ユーザの格付け評価から映画が推薦される仕様となっている。au one ラボ 映画推薦ツールはフィードバック機能を有しており、推薦された映画からユーザが感情別の8項目、5段階の評価を返すことができる。本システムにおいても、推薦された映画に3段階の評価を返すことができる。

(5) 推薦精度の比較評価

FANCAST、movielens はアメリカのサイトなので推薦される映画も洋画が中心となるため、日本のサイトと同様に評価することはできなかった。au one ラボ 映画推薦ツール、本システムは登録ユーザ情報を統一し、15件の映画に同じ格付けをしたものを入力情報として、出力される推薦情報を主観的に評価した。au one ラボ 映画推薦システムは興味を持つ作品が多数、推薦された。本システムも興味を持つ作品が推薦されたが、興味のない映画の割合が au one ラボ 映画推薦ツールを上回ってしまった。ただ、推薦精度の評価は主観であること、4つのうち2つがアメリカのサイトであったことからあまり効率的な比較評価を行うことができなかった。

表4に比較した項目をまとめた表を示す。

表4 従来映画推薦システムと本システムの比較

比較項目	提案システム	FANCAST	movielens	au one ラボ 映画推薦ツール
ユーザ情報登録のインターフェース	○	◎	×	△
映画格付け情報登録のインターフェース	◎	-	○	△
推薦情報表示のインターフェース	△	○	◎	○
フィードバックの有無	○	×	×	○
推薦精度	○	×	△	◎
日本語サイト	○	×	×	○

6. おわりに

本研究では複合型アルゴリズムを用いた映画推薦システムを提案し、試作・評価を行った。ユーザからのアンケートデータを基に評価を行い、一定の有効性が示された。また、従来の推薦システムと比較し、提案システムは一定レベルの推薦精度を持つことや、映画格付けインターフェースに優れていることなどが示された。しかし、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た。また、格付け評価を入力する支援機能を装備したが、映画を多く観ているにも関わらず、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。

今後、映画推薦の容易なインターフェースと検索方法を模索し、より良いシステムを構築したい。

参考文献

- [1] Sutheera, P., Tsuji, H., A Multi-Clustering Hybrid Recommender System, In proceedings of 7th IEEE International Conference on Computer and Information Technology, pp.223-234, October 16-19 2007.
- [2] 坂本竜太, スティーラ・パンティラヌラク, 吉見哲哉, 後藤慎平, 山谷一喜, 伊藤壘, 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, 情報処理学会第70回全国大会, 6R-7, 2008年3月.
- [3] 特集 利用者の好みをとらえ活かす -嗜好抽出技術の最前線-, 情報処理, Vol48.No.9, 2007.
- [4] FANCAST, <http://www.fanecast.com/>,

Comcast.

[5] movielens, <http://movielens.umn.edu/>,
GroupLens Research.

[6] au one ラボ 映画 推薦 ツール,
https://labs.auone.jp/CGI/research/apl_top.cgi?app_id=01, KDDI.

[7] 小野智弘, 本村陽一, 麻生英樹, ペイジア
ンネットによる映画コンテンツ推薦方式の検討,
電子情報通信学科技術報告, vol.104, no.348,
pp.55-60, 2004.

[8] Enjoy*Study,
<http://www.enjoyxstudy.com/javascript/suggest/>