

セレンディピティを考慮した情報推薦システムの提案

Proposal of Collaborative Filtering based Recommendation System in Consideration of Serendipity

打矢 隆弘† 吉久 宗一† 内匠 逸†
Takahiro Uchiya Soichi Yoshihisa Ichi Takumi

1. まえがき

ユーザにとって有益である情報・商品を提示するためのツールとして、現在、多くの e-コマースサイトで情報推薦システムが活用されている。情報推薦システムでは協調フィルタリング方式が多く用いられており、ユーザの閲覧・購入履歴、アイテムの評価などからユーザの嗜好を推測し、好みそうなアイテムを推薦する。しかし、既存の情報推薦システムでは、推薦したアイテムがユーザの嗜好と合致する可能性は高いが、ユーザが一度商品を購入すると似たようなアイテムばかりが推薦されるといった問題がある。それにより、ユーザの推薦に対する「飽き」が生じ、推薦効果が著しく低下している。

ユーザにとって常に新鮮であり新しい発見のある推薦を実現するために、本研究では情報推薦において意外性を表す指標である「セレンディピティ」について着目し、ユーザにとって有益な情報推薦を行うシステムを提案する。従来の研究では推薦精度のみに着目することが多かったが、本研究では知名度やソーシャルグラフといった従来の推薦にはない要素を付加し意外性の高い推薦を行うことによって、推薦効果を大幅に向上することができると考える。特にユーザが e-コマースサイトを継続的に利用している場合はセレンディピティの高い推薦は非常に効果的である。

以下、2章では、協調フィルタリング方式などの従来の推薦システムの手法、またその利点・欠点について述べ、各手法のセレンディピティの観点からの考察を述べる。3章では、セレンディピティを考慮した推薦システムの設計・実装について述べる。4章では、MovieLens のデータを用いた検証実験とその考察について述べる。5章では、関連研究について述べる。6章では、まとめと今後の課題について述べる。

2. 従来方式

2.1 セレンディピティの定義

情報推薦におけるセレンディピティとは、推薦されることによりユーザの潜在的な興味を引き出し、思いがけない(意外な)発見を与えるものである。

本稿では、セレンディピティを「意外性」と定義する。情報推薦システムは、ユーザにとって有益である情報・商品を推薦アイテムリストという形でユーザに提供するが、このリスト内に「意外性」のあるアイテムが含まれていることをセレンディピティの高い推薦と呼ぶ。

† 〒466-8555 愛知県名古屋市中区御器所町名古屋工業大学 大学院 工学研究科 情報工学専攻 Nagoya Institute of Technology, Graduate School of Engineering, Gokiso, Showa, Nagoya, Aichi, 466-8555 Japan.

2.2 従来方式とセレンディピティ

従来の推薦方式は、

- ① 運営者が設定した独自のルールに基づき推薦するアイテムを決定する**ルールベース方式**
 - ② アイテムに付与されているメタデータを利用して類似アイテムを推薦する**内容に基づくフィルタリング方式**
 - ③ 類似した嗜好をもつユーザの評価が高いアイテムを推薦する**協調フィルタリング方式**
- の3つに大別される。従来手法とセレンディピティの関係を図1に示す。

推薦方式	分析情報	推薦方法	メリット	デメリット	セレンディピティ
協調フィルタリング	ユーザの行動、購入履歴	嗜好の類似したグループを形成し、推薦アイテムを決定	事前の準備が不要。コンテンツに依存しない。	ユーザの行動履歴が集まらないと推薦ができない	コンテンツに依存しないためグループの作り方によっては高くなる場合がある
内容に基づくフィルタリング	コンテンツのもつ特徴表現、ユーザの興味表現	特徴、興味表現を分析し、コンテンツ間の関連性から推薦アイテムを決定	嗜好とかけ離れた推薦をしない	管理コストがかかる。適切なメタデータの付加が必要	コンテンツの関連性から推薦するため常に似たアイテムのみが推薦されるため低い
ルールベース	過去の実績やマーケティングデータ	設定された独自のルールで推薦アイテムを決定	運営者の意思を反映できる	管理コストがかかる。ユーザの嗜好が反映されない	個々のユーザの嗜好を全く考慮していないため低い

図1: 従来方式とセレンディピティ

セレンディピティの観点から従来方式を分析した場合、①の方式は「個々のユーザの嗜好に適した推薦ができない」、②の方式は「基本的に以前購入したアイテムと類似したアイテムしか推薦されない」ことから、セレンディピティは低いと考えられる。一方、③の方式は「アイテム間の類似性を一切考慮せず推薦を行う」ため、嗜好が似ている他のユーザが様々なアイテムを購入している場合は、セレンディピティの高い推薦を行える可能性がある。そこで本研究では「協調フィルタリング方式」を軸とした情報推薦システムの構築を行う。

2.3 協調フィルタリング方式

協調フィルタリング方式は、ユーザベース方式(user-based method)、アイテムベース方式(item-based method)、モデルベース方式(model-based method)の3つに大別される。ユーザベース方式[1]では、以下の処理を行う。

1. 被推薦者和其他のユーザとの**類似度**を計算
2. まだ被推薦者が評価していないアイテムの評価の**推定値**を過去の評価値から計算
3. 最も推定値の高いアイテムを推薦

類似度の計算にはピアソン相関係数を用い、被推薦者 a と、ユーザ u 間の類似度 r_{au} を式(1)に示す。ここで、 σ_a 、 σ_u はユーザ a 、 u の標準偏差、 \bar{a} 、 \bar{u} は評価値の平均とする。

$$r_{au} = \frac{\sum_i (a_i - \bar{a})(u_i - \bar{u})}{\sqrt{\sum_i (a_i - \bar{a})^2} \sqrt{\sum_i (u_i - \bar{u})^2}} \quad (1)$$

推定値は以下のように求める。被推薦者 a が未評価のアイテム i を、いずれかのユーザ $u \in S(a)$ が評価していた場合、その推定値 a_i を式(2)に示す。

$$a_i = \bar{a} + \frac{\sum_{u \in S(a)} (u_i - \bar{u}) r_{au}}{\sum_{u \in S(a)} |r_{au}|} \quad (2)$$

3. 提案システム

3.1 設計指針

本研究では、以下の指針に沿ってセレンディピティの高い情報推薦システムを構築した。

(指針 1) ユーザが意外であると感じるアイテムは、これまでに購入したコンテンツと類似していないが、ユーザの興味を惹くコンテンツであると考えられる。そこで、「ユーザベースの協調フィルタリング方式」を基本アルゴリズムとして採用する。この方式では嗜好の類似したグループを形成し、ユーザが未評価(未購入)でかつ他者の評価の高いアイテムを提示するため、ユーザの嗜好に一致しないアイテムが推薦される可能性は低い。これにより、推薦されるアイテムの「**正確性**」を確保する。

(指針 2) 多様性のある推薦アイテムリストが作成できれば、そのリスト内に含まれているアイテムが意外性をもつ可能性が高まる。そこで、多様性に寄与すると考えられる下記のアイデアを提案システムに組み込む。これにより、推薦されるアイテムの「**多様性**」を確保する。

- ◇ ユーザ間のソーシャルグラフを推薦に活用
- ◇ ユーザの興味のある分野において、マイスターユーザ(専門性が高いユーザ)の推薦を導入
- ◇ アイテムの popularity (知名度) を計算し、知名度の低いアイテムを推薦に活用
- ◇ アイテムに話題性、新商品などの付加価値を付与

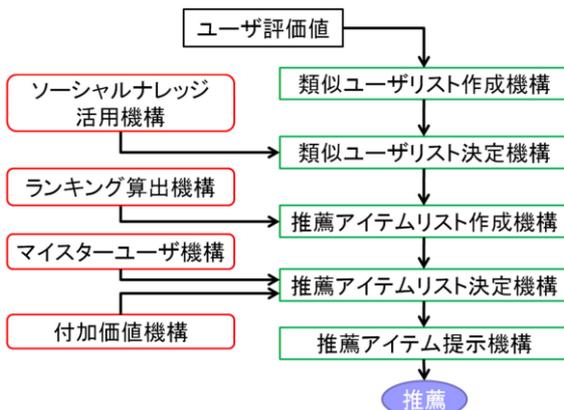


図 2: システム構成

3.2 設計

図 2 にシステム構成を示す。従来の協調フィルタリング方式(図中の右フロー)に加え、ソーシャルグラフを活用することで類似ユーザリストにゆらぎを与える「ソーシャ

ルナレッジ活用機構」、及び、知名度を活用しアイテムの推薦アイテムリストにゆらぎを与える「ランキング算出機構」を導入する。また、アイテムに話題性、新商品などの付加価値を付与する「付加価値機構」、専門性の高いマイスターユーザの高評価のアイテムを推薦させる「マイスターユーザ機構」を導入する。以上の機構により、**正確性**と**多様性**を考慮した推薦を行う。

3.3 内部設計

以下に各機構の役割を示す。

[類似ユーザリスト作成機構] 協調フィルタリング方式のアルゴリズムに基づき、ピアソン相関係数を用いて類似するユーザのリストを作成する(図 3)。ピアソン相関係数は以下の式に基づき算出する。

$$r_{AB} = \frac{\sum_{i=1}^n (A_i - \bar{A})(B_i - \bar{B})}{\sqrt{\sum_{i=1}^n (A_i - \bar{A})^2} \sqrt{\sum_{i=1}^n (B_i - \bar{B})^2}}$$

ここで A, B はユーザ、 A_i, B_i はアイテム i に対する評価、 \bar{A}, \bar{B} は各ユーザの評価の平均とする。

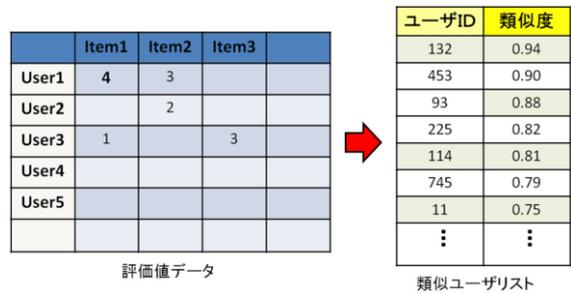


図 3: 類似ユーザリストの作成

[ソーシャルナレッジ活用機構] 現実社会において、友人から口コミで有用な情報・アイテムを教えてもらうことはよくあることである。本機構では、友人関係、世代、コミュニティといったソーシャルグラフを用いることで、ユーザ間の類似度に重みづけすることにより類似ユーザリストを変化させる。具体的にはソーシャルネットワークサービス(mixi, twitter 等)からソーシャルグラフを抽出し、重みづけを行う。

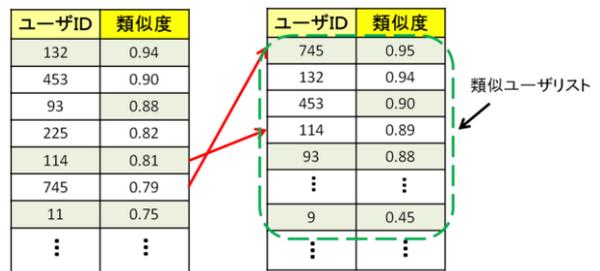


図 4: ソーシャルナレッジ活用機構による類似ユーザリストの変化

[類似ユーザリスト決定機構] 類似ユーザ作成機構にソーシャルナレッジ活用機構の情報を付加し、上位 L 件のユーザを類似ユーザリストとして保持する(図 4)。

[推薦アイテムリスト作成機構] 従来の協調フィルタリング方式のアルゴリズムに基づき、個々のアイテムの推定値を下記の式より算出し推薦アイテムリストを作成する。ここでは、ユーザ A に対してアイテム 1 の推定値を算出するものとし、 $\{J | J \in \text{類似ユーザリストの中でアイテム 1 を評価したユーザ}\}$ とする。

$$A_1 \text{推定値} = \bar{A} + \frac{\sum_J (J_1 - \bar{J}) r_{AJ}}{\sum_J |r_{AJ}|}$$

[ランキング算出機構] 従来の協調フィルタリング方式では、新規に追加されたアイテムや評価の少ないアイテムは推薦されにくい(再生起問題)。そこで、ランキング算出機構では、アイテムの評価回数を知名度とみなし、この順位をランキング情報として保持、知名度の低いアイテム(掘り出し物)を優先して推薦されるように重みづけを行い、予測評価値の高い上位 M 件のアイテムを推薦アイテムリストとして保持する(図 5)。

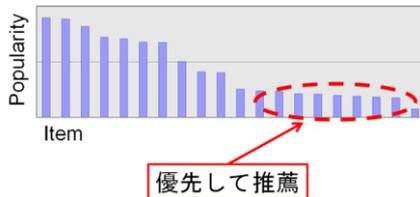


図 5: ランキング算出機構

[マイスターユーザ機構] マイスターユーザ機構では、被推薦ユーザが特に興味がある分野を推測し、その分野での専門性の高いマイスターユーザの高評価アイテムをリストに加える。具体的には、アイテムにあらかじめ付加した各属性毎にマイスターユーザを選出し、推薦アイテムリストを作成する。次に被推薦ユーザが評価したアイテムの属性の和を算出し、最も評価の多い分野をユーザの興味のある分野とする。その分野でのマイスターユーザの推薦アイテムリストと、推薦アイテムリスト作成機構で出力したリストを比較し、一致するアイテムに重みづけを行う。

[付加価値機構] 付加価値機構では、話題性・新商品などといったアイテムの付加価値情報を保持し、優先して推薦されるように重みづけを行う。付加価値情報はマーケティングデータやあらかじめ得られた知見などに基づき、運営者が独自に設計を行う。この手法はルールベース方式に近い手法であるが、従来のルールベース方式ではユーザの嗜好を加味しないという欠点があった。本手法では協調フィルタリング方式を用いて、あらかじめ作成した推薦アイテムリストから一致するアイテムのみを重みづけすることにより、個々のユーザに適した重みづけを行うことが出来ると考える。

[推薦アイテムリスト決定機構] マイスターユーザ機構、付加価値機構からの評価値情報を統合し、推薦アイテムリストの評価値を変化させる。推薦アイテムリストの上位 N 件を推薦アイテムとして決定する。

[推薦アイテム提示機構] 被推薦ユーザに対し、推薦アイテムを提示する(図 6)。ソーシャルナレッジ活用機構からアイテムのレビュー等が取得出来る場合、併せて提示を行う。

```
RecommendedItem[item:1449, value:4.7130356]
RecommendedItem[item:1463, value:4.532252]
RecommendedItem[item:1594, value:4.502109]
RecommendedItem[item:1398, value:4.4997554]
RecommendedItem[item:1642, value:4.4975643]
RecommendedItem[item:408, value:4.4796996]
RecommendedItem[item:318, value:4.4761853]
RecommendedItem[item:483, value:4.4615626]
RecommendedItem[item:603, value:4.3749685]
RecommendedItem[item:1639, value:4.3277173]
```

図 6: 推薦アイテムリスト

3.4 実装

提案システムのプロトタイプとして、クライアント環境で動作する情報推薦アプリケーションを作成した。協調フィルタリング方式の実装においては、Apache Mahout プロジェクト[2]の Collaborative Filtering パッケージを利用した。Collaborative Filtering パッケージには、ユーザベースの協調フィルタリングを行うための Java API が整備されており、これを用いて協調フィルタリング処理の実装を行った。また、新規導入の「ソーシャルナレッジ活用機構」、「ランキング算出機構」、「付加価値機構」、「マイスターユーザ機構」はそれぞれ独立した新規 Java クラスとして実装した。

4. 実験と評価

実際の評価値データを用いて、提案システムで導入した各機構の有用性の確認を行った。

4.1 実験概要

[実験データ]

実験データとして、MovieLens[3] と呼ばれる映画の評価情報が記述されたデータセットを利用した。データサイズとして、ユーザ数 943、アイテム数 1682、総評価回数 90570、ユーザはアイテムに対し 5 段階での評価をし、最低 20 個以上のアイテムを評価している。また、各ユーザは年齢、性別、職業のタグが付与されており、各アイテムにはそれぞれアイテムの属性(映画のジャンル)を示すいくつかのタグが与えられている。今回の実験では年齢、性別、職業のデータから擬似的なソーシャルグラフ、マイスターユーザを作成し、検証を行った。

[実験方法]

943 名の各ユーザに対し、以下の 6 つの方式を用いてそれぞれ 10 件の推薦アイテムリストを作成し、各ジャンルの出現頻度、各推薦アイテムの出現頻度を算出し、検証を行った。本実験では、パラメータ $L=200$, $M=50$, $N=10$ を用いた。

- ① 協調フィルタリング(CF)のみ
- ② CF + ソーシャルナレッジ活用機構

- ③ CF + ランキング算出機構
- ④ CF + マイスターユーザ機構
- ⑤ CF + 付加価値機構
- ⑥ CF + ソーシャルナレッジ活用機構 + ランキング算出機構 + マイスターユーザ機構 + 付加価値機構

また、⑥については付加した機構の重みを変化させ、どのように推薦アイテムリストに影響を与えるかを検証した。

4.2 実験結果と考察

②～⑥それぞれについて、各ジャンル(18ジャンル)の出現頻度を①と比較した結果を図7, 図9, 図11, 図13, 図15に示す。また、②～⑥それぞれについて、各推薦アイテムの出現頻度を降順にソートし、①と比較した結果を図8, 図10, 図12, 図14, 図16に示す。この結果から、それぞれの機構によって推薦アイテムにばらつきが大きくなっていることが確認できる。以下に各機構についての考察を述べる。

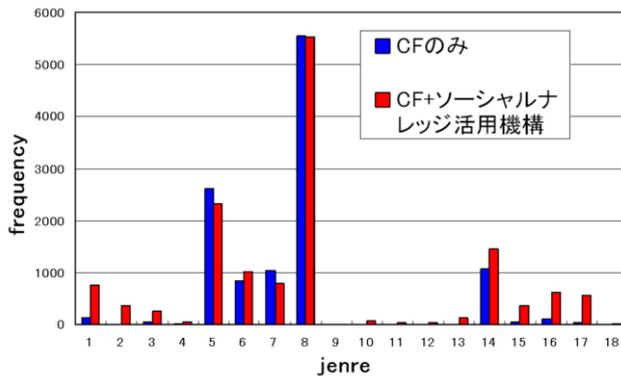


図6: ソーシャルナレッジ活用機構を付加した場合の各ジャンルの出現頻度

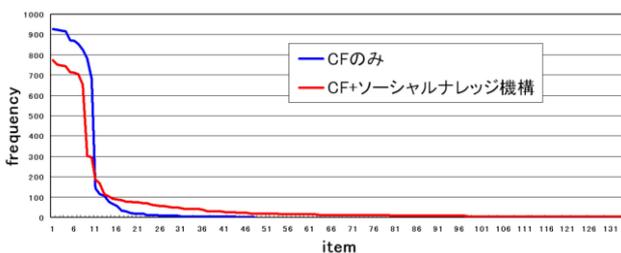


図7: ソーシャルナレッジ活用機構を付加した場合の各アイテムの出現頻度

[ソーシャルナレッジ活用機構に対する考察(図6, 図7)]

本機構ではユーザの類似度に変化を与える為、従来のCFとジャンルの出現傾向は似ている結果となったが、ジャンル1, 2, 3, 15, 16, 17などCFでは推薦されにくいジャンルの出現頻度が高くなったことが確認できた(図6)。また、各アイテムの出現頻度ではロングテールの傾向が見られたことから、出現アイテムの多様性が大きく向上することが確認できた(図7)。従来の協調フィルタリング方式では、多くのアイテムを評価しているユーザは類似ユ

ーザに選出され易い。その結果、そのユーザが高評価を与えたアイテムは常に高い予測値になってしまい、似たようなアイテムが推薦されてしまうという問題があった。ソーシャルナレッジ機構では、類似ユーザリストのばらつきを大きくさせることで、推薦されるアイテムの多様性が向上したと考えられる。

今回の実験では職業や世代などから1人に対し3～5人程度の友人を擬似的に作成したが、実際のソーシャルグラフでは、友人数のばらつきが大きくなるのが考えられる為、今後検証を行っていく必要がある。

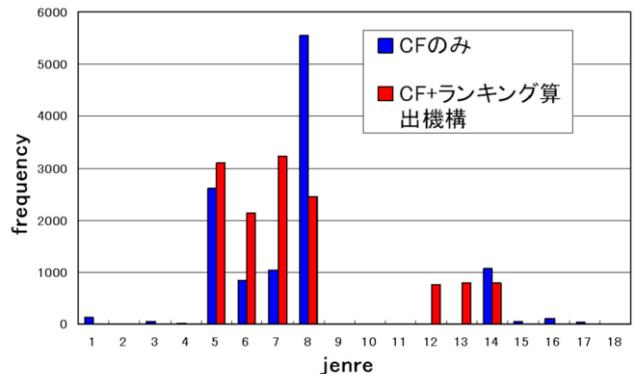


図8: ランキング算出機構を付加した場合の各ジャンルの出現頻度

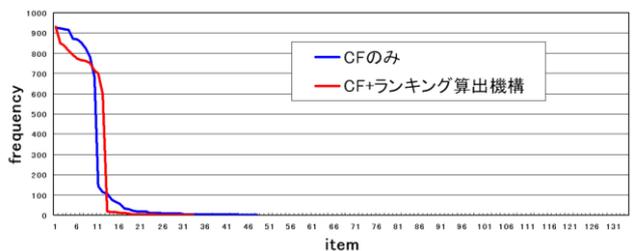


図9: ランキング算出機構を付加した場合の各アイテムの出現頻度

[ランキング算出機構に対する考察(図8, 図9)]

ランキング算出機構ではアイテム全体に対し重みづけを行う為、全体のアイテム属性の分布に比例したジャンルの出現傾向となった(図8)。また、ランキング算出機構では評価数に反比例するように重みづけを行っている。今回用いたデータセットでは、個々のアイテムの評価回数の差が大きく、ほとんど評価されていないアイテムに大きく重みづけがされる。その結果、推薦アイテムが偏ってしまうということが確認できた(図9)。実際のeコマースサイトでも、アイテムに対する評価が少ない問題(評価値疎ら問題)が考えられるため未評価のアイテムばかりが推薦されてしまうことが考えられる。現在、ランキング算出機構では、一意に重みづけを行っているが、多様性の高い推薦を行うには評価回数の分布に応じた適切な重みづけが必要である。

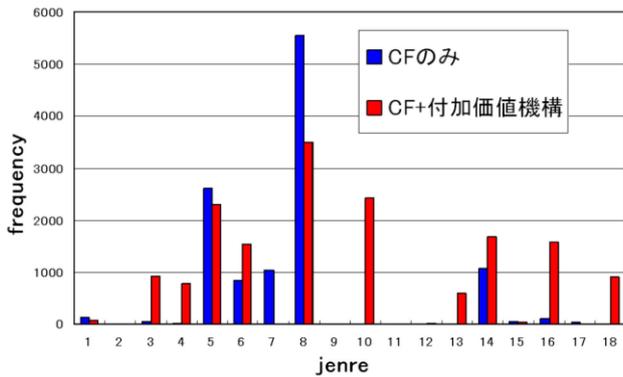


図 10: 付加価値機構を付加した場合の各ジャンルの出現頻度

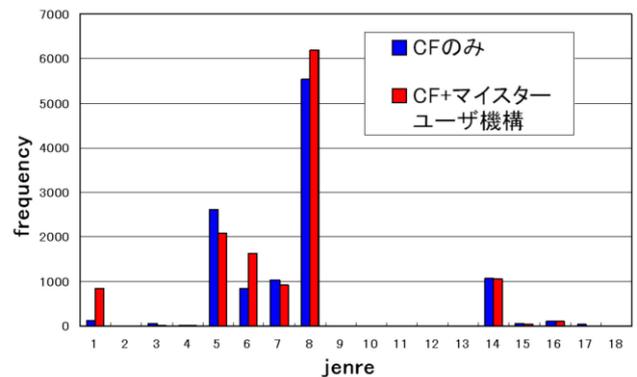


図 12: マイスターユーザ機構を付加した場合の各ジャンルの出現頻度

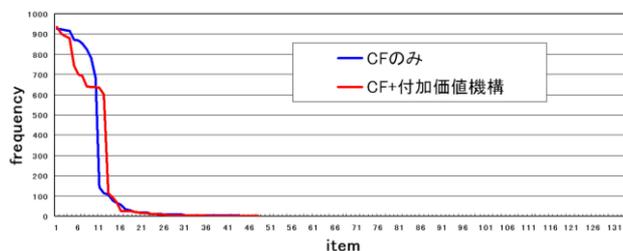


図 11: 付加価値機構を付加した場合の各アイテムの出現頻度

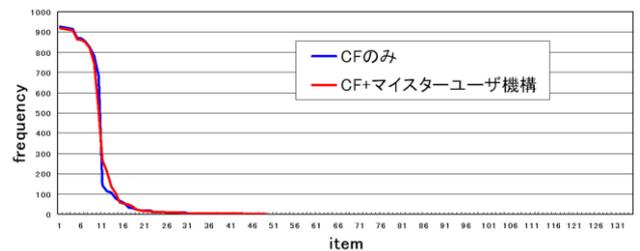


図 13: マイスターユーザ機構を付加した場合の各アイテムの出現頻度

[付加価値機構に対する考察(図 10, 図 11)]

今回は、各ジャンル3つのアイテムを付加価値機構の付加価値アイテムリストに加えた。

付加価値機構では運営者が独自にルールを設計する為、ルール設計次第では、ある程度任意にジャンルにばらつきを持たせることができることが確認できた(図 10)。また、本システムでは、協調フィルタリング方式によって作成した推薦アイテムリスト内のアイテムと一致するアイテムのみ付加価値をつけるため、ユーザの嗜好とかけ離れることが少ないと考えられる。しかし、今回の実験ではルールが少なかった為、同ジャンル内の同じアイテムが推薦されてしまい、推薦アイテムに偏りが出ていることが確認できた(図 11)。実際に運用するには、付加価値機構の重みを下げることや、付加価値アイテムリスト内のアイテム数を増やすことが必要である。

[マイスターユーザ機構に対する考察(図 12, 図 13)]

マイスターユーザ機構では、ユーザの最も興味のあるジャンルを推測し、そのジャンルでのマイスターユーザの高評価アイテムを推薦する為、従来の手法よりジャンルの多様性は低くなっていることが確認できた(図 12)。一方、各アイテムの出現頻度は CF と大きな差は見られなかった(図 13)。これらの結果から、ユーザの興味のあるジャンル内での高評価アイテムの提示に本機構が役に立つことが判明した。すなわち、多様性を犠牲にして推薦の「正確性」の向上を達成することができると考えられる。

[全機構を付加した場合に対する考察(図 14, 図 15)]

図 14, 図 15 の提案手法では、従来の協調フィルタリング方式に加えこれまで述べた4つの機構を付加している。

提案手法1では、付加した各機構の重みを低くすることで、従来の協調フィルタリング方式の影響が大きくなるように設計した。また、提案手法2では付加した各機構の重みを提案手法1に対し大きくすることで、協調フィルタリングの影響が小さくなるように設計し、提案手法3では、提案手法2より更に付加した各機構の重みを大きくすることで、協調フィルタリング方式の影響を更に小さくするように設計した。

その結果、新たに付加した機構によって多様性が高くなることが確認できた。CF < 提案手法1 < 提案手法2 < 提案手法3 の順に各ジャンルの出現頻度のばらつきが大きくなっていることがわかる(図 14)。また、従来の手法では全アイテム中、49種類のアイテムしか推薦されなかったが、提案手法3では133種類のアイテムが推薦され、大きく性能を向上させることができた(図 15)。また、新たに追加した機構と、推薦精度が高いとされる従来の協調フィルタリングとの重みを変化させることで、推薦の意外性の度合いを調節できることが確認できた。

一般に e-コマースサイトを閲覧するユーザは、特定の商品を購入する意思があるユーザとないユーザに分類できると考えられる。本システムではユーザの特定商品の購入意思の有無によって、購入意思のあるユーザには推薦精度を優先する推薦(提案手法1)、購入意思のないユーザには意外性を優先する推薦(提案手法3)を行うことで、

ユーザの満足度を向上させる推薦を行うことが期待できる。

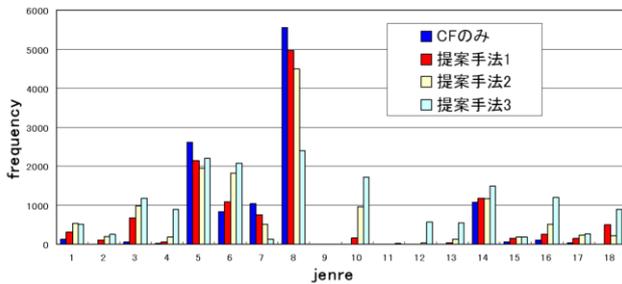


図 14: 全機構を付加し重みづけを変化させた場合の各ジャンルの出現頻度

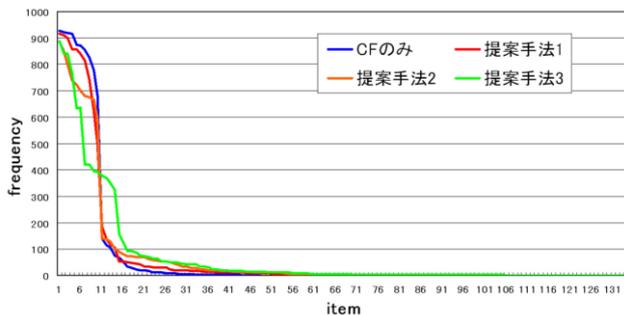


図 15: 全機構を付加し重みづけを変化させた場合の各アイテムの出現頻度

[全体の考察]

提案手法では、新たな4つの機構を付加することにより、ジャンル、推薦アイテムとも出現頻度にばらつきが大きくなり、それを統合することで、より多様性の高い推薦を行うことができることが確認できた。これにより、従来の推薦システムの問題点であった似たようなものが推薦されるという問題点が改善され、意外性の高い推薦を行うことが出来ると考えられる。

5. 関連研究

意外性を考慮した情報推薦システムの研究開発事例について述べる。

5.1 加藤らの研究

加藤ら[4]は、インターネットを利用したオンラインショッピングを対象に、推薦の正確性と意外性のバランスを考慮した推薦システムを提案している。推薦に正確性と意外性を持たせる手段として、

- ① ユーザ特徴ベクトル作成機能により、遺伝的アルゴリズムを用いてユーザの嗜好の変化に応じた特徴ベクトルを生成
- ② フィルタリング機能により、商品ベクトルのクラスタリングを行いもっともユーザの嗜好にあったカテゴリが属するクラスタのみを推薦対象とすることにより、ユーザが全く興味を持たないデータを推薦結果から除く

- ③ マッチング機能により、商品特徴ベクトルとユーザ特徴ベクトルのマッチングにおいて、順マッチングと交差マッチングの回数を確率的に変化させることによる正確性と意外性の調和のとれた推薦処理の3つの機能を実現している。

アルゴリズムの流れを以下に説明する。①は推薦の正確性のみを高めるための機能であるので①の説明は省略する。

②において、ユーザ特徴ベクトルの要素から、ユーザが最も好むカテゴリ C_p を決定する。そしてカテゴリ別のユーザ特徴ベクトル間のユークリッド距離を求め、カテゴリのクラスタリングを行う。そして、意外性のある推薦を行うための商品として、 C_p が属するクラスタに含まれるカテゴリを抽出する。その後は、推薦された商品をユーザが選択するたびに、カテゴリ別のユーザ特徴ベクトルを更新し、推薦対象となるカテゴリを抽出しなおす。すなわち、②は意外性のある推薦を行うための商品候補を抽出する機能をもっている。

次に③で正確性と意外性の調和を取るための処理の流れを説明する。順マッチングは、購買履歴（または選択履歴）があるカテゴリ A 内のユーザ特徴ベクトルと、カテゴリ A 内の商品特徴ベクトル（あらかじめ商品ごとに設定されているものとする）とを比較することにより推薦商品を決する方法である。一方、交差マッチングは、購買履歴のあるカテゴリ A 内のユーザ特徴ベクトルと、 A 以外のカテゴリ B 内の商品特徴ベクトルとを比較することにより推薦商品を決する方法である。

それぞれのマッチングは、ユーザの選択履歴をもとに、よく選択する傾向にあるカテゴリから商品を選択した頻度と、それ以外のカテゴリから商品を選択した頻度をもとに、推薦結果の個数を確率的に求めることで使い分けられる。以下に手順を示す。

1. フィルタリング機能より、ユーザが最も好む傾向にあるカテゴリと交差マッチングにおいて選択対象となるカテゴリを選択する。
2. 順マッチングを用いて推薦する確率 P_o と交差マッチングを用いて推薦する確率 P_c を算出する。

$$P_o = \frac{I_o + N_o}{I_o + I_c + N_o + N_c}$$

$$P_c = 1 - P_o = \frac{I_c + N_c}{I_o + I_c + N_o + N_c}$$

ここで I_o , I_c は正確性を優先した推薦結果の個数と意外性を優先した推薦結果の個数の初期値であり、ユーザが初期設定値としてシステムに与える値である。また、 N_o , N_c は過去の履歴において、正確性を優先した推薦結果と意外性を優先した推薦結果が何個ずつ選択されたかを表す値であり、ユーザの選択結果に従って更新されていく。

3. 算出された選択確率から順マッチングを用いて推薦する個数と交差マッチングを用いて推薦する個数を決定する。

この研究では、上記のアルゴリズムを実験システムに実装し、20代の男性10名による主観評価実験によってその有効性と妥当性を検証している。実験は、遺伝的アル

ゴリズムのパラメータを決定するための実験と推薦結果を評価する実験の2段階に分けて行われている。それぞれの実験では、被験者は提示された推薦結果に対して、正確性(好みの商品であり、推薦されることが予想できる商品)と意外性(予想外の推薦結果であるが、好みにあう商品)に対して5段階評価(良い/やや良い/ふつう/やや悪い/悪い)を行っている。すなわち、正確性及び意外性に関する全ユーザの5段階評価の平均値を評価指標として、正確性及び意外性の評価を行っていることになる。通常、推薦システムでは、正確性と意外性とはトレードオフの関係にあるが、この実験では、推定処理のたびに順マッチングと交差マッチングを適切な割合で組み合わせて推定を行っているため、正確性、意外性の双方共に高い評価値を示す結果となった。

この研究では、ユーザ特徴ベクトルの初期設定において、カテゴリ毎にランダムに選ばれた商品に対する評価値の入力をユーザが手作業で行わなければならないという課題が存在する。Amazonのような多数のカテゴリを扱うオンラインショップで同手法を適用した場合、カテゴリ数が莫大なため、ユーザの負担は非常に大きい。

5.2 清水らの研究

清水ら[5]は、協調フィルタリング手法を用いた推薦システムの意外性(発見性: Novelty)によるユーザ満足度の向上について検討している。提案方式は、従来の協調フィルタリング手法において用いられていたユーザの嗜好のプロファイルに加えて、ユーザがどのようなアイテムを知っているかについてのプロファイルを利用することで意外性の向上を図るものである。意外性を実現するためのアルゴリズムとして3つの方式が提案されているが、以下では、結果として意外性を実現するのに最も適したアルゴリズムであると判断した方式についてのみ述べる。

論文中では、アルゴリズムの説明にあたり、まず嗜好と不既知(知らない)の性質の違いについて述べている。嗜好に関しては、どれくらい好きかという嗜好の程度が推薦システムの推薦結果に反映されるべきであるが、不既知に関しては、程度の違いはないため、順位付けすることよりも、知らない商品集合を特定することのほうが重要であるというものである。

以上の想定を元に、この研究で提案されている嗜好と不既知の性質の違いを考慮した推薦アルゴリズムを図16に示す。このアルゴリズムでは、知らない商品集合を特定し、その中に含まれるアイテムを嗜好の予測値により順位付けするという推薦方式をとっている。ここで、嗜好のプロファイルに協調フィルタリングを適用して作成される商品リスト L_{pi} と、既知・不既知のプロファイルに協調フィルタリングを適用して作成される商品リスト $L_{pi}^{unknown}$ を結合することを考えたとき、それぞれのリストが、結果として生じるリストにどれだけ影響を及ぼすかを表す重みを $\alpha \in [0, 1]$ とする。

まず、既知・不既知のプロファイルに協調フィルタリングアルゴリズムを適用する。それにより得られる既知・不既知の予測評価値 $P_i^{know}(b)$ が閾値 $\alpha \in [0, 1]$ より小さい商品の集合を知らない商品集合 B'_i とする。次に、知らない商品集合 B'_i に含まれる商品を嗜好の予測評価値 P_i で順位付けすることで推薦リスト L_{pi} を作成する。

この研究では、図16のアルゴリズムを実験システムに実装し、10代~50代以上の計100名による主観評価実験によってその有効性と妥当性を検証している。なお、各商品に嗜好の評価値と既知・不既知の評価値の両方が必ずついている場合を想定しての実験(i)と、両方の評価値ともついていない場合または不既知の評価値のみがついている場合を想定しての実験(ii)の2種類が行われている。それぞれの実験では、各被験者は合計1000曲からなる音楽データベースの中からランダムに選択された200曲に対して、嗜好の5段階評価値(1~5)と既知・不既知の評価値(1 or 0)をつける。200曲のうち100曲は訓練事例、残りの100曲はテストセットとして使用される。そして、 $\alpha \in [0, 0.1, \dots, 0.9, 1.0]$ を変化させたときに作成される上位5個の推薦リストの各指標の変化を調べることにより、推薦システムの意外性の評価を行っている。なお、実験(ii)では、不既知の評価がついている曲の50%に嗜好の評価値がついているデータセットを作成する。その結果、実験(i)では、 $\alpha = 0.3$ のときに Novelty がもっとも高い値を示しており、従来方式よりも高い精度を示している。実験(ii)では、推薦リスト中の約60%に不既知の評価がついていた場合に Novelty がもっとも高い値を示しており、不既知の評価値のみがついている曲だけでなく、全く評価値がついていない曲からも多くの推薦を受けたい場合には、図16のアルゴリズムは従来方式よりも有効であるということが示されている。

この研究は不既知アイテムを選別したのち、それらの中から有益なアイテムを推薦することから、意外性の高い推薦が実現できると考えられる。しかし、事前に多数のアイテムに対して、ユーザ自らが既知・不既知の情報を付加してゆくことは非常にコストがかかり現実的に困難である。

```

procedure discovery3( $B_i, \alpha$ ) {
   $b \in B_i$  : compute  $p_i^{know}(b)$ ;
  set  $B'_i \leftarrow \{B_i \mid p_i^{know}(b) \leq \alpha\}$ ;
  for all  $b \in B'_i$  do
    compute  $p_i(b)$ ;
  end do
  compute  $L_{pi} : \{1, 2, \dots, |B'_i|\} \rightarrow B_i$  using  $p_i$ ;
  return  $L_{pi}$ ;
}

```

図16: 嗜好と発見性を考慮した協調フィルタリングアルゴリズム

5.3 小川らの研究

小川ら[6]は、正確性と多様性の両立を実現するために既存の協調フィルタリングによる Top-N 推薦リストの計算アルゴリズムを改善し、新たな推薦アルゴリズムを提案している。推薦の正確さについては従来の協調フィルタリングを改良したものをを用い、多様性については共購買アルゴリズムを用いる。

[協調フィルタリングの改良] 類似度計算と予測評価値計算の2つの手順で行われる。類似度計算については Adjusted Weight Sum を用いる。予測評価値計算では、値を類似度の合計値で正規化しない Simple Adjusted Summation 手法(SAS 手法)を用いる。

$$P_{u,i}^{SAS} = \bar{r}_u + \sum_{m \in M} sim(u,m)(r_{m,i} - \text{standardvalue})$$

ここで、 $r_{m,i}$ はユーザ i のアイテムへ投票した評価値、 \bar{r}_u はユーザ u の投票したすべての評価値の平均、 $\Sigma_{m \in M}$ はユーザ u の類似度の高い上位 M 人の近傍ユーザ、 standardvalue は評価値の基準値(評価値が 1~5 の場合基準値は 3)とする。上記の式により、近傍ユーザの多くが高く評価したアイテムほど値が高くなり、推薦リストの上位に現れやすくなる。これはつまり、多数のユーザが推薦していたとしても、全体の割合で正規化することで意外性のあるものの評価値が下がってしまうことや、全体で平均化することで埋もれてしまうことを防ぐことが出来る。

[多様化アルゴリズムの導入] 多様化アルゴリズムとして、アイテムの共購買のネットワークから求められる指標(ネットワーク径、次数中心性、PageRank など)を多様化の指標として用いる。これらの指標をもとに、既存の推薦リストの推薦順位を入れ替えることで、多様化を行った新たな推薦リストを作成する。

以上のアルゴリズムを用いることで、従来の協調フィルタリングを用いる方法よりも意外性が高くなることを示している。

この研究は、本稿で提案している手法と同様に、様々な指標を考慮して推薦リスト内のアイテムの推薦順を入れ替えることで多様化を図っている。提案されている多様化の指標は非常に興味深いものであるが、多くの計算を必要とする指標を増やせば増やすほど、多様化アルゴリズムが複雑化し、処理速度が低下する恐れがある。

6. おわりに

本研究では情報推薦システムにおけるセレンディピティという新しい概念について着目し、ユーザにとって常に新鮮であり新しい発見のある推薦の実現を目指した。

本稿ではまず、従来の推薦方式について説明し、セレンディピティの観点からの評価を述べた。また、セレンディピティを考慮した情報推薦システムのモデルを提案し、プロトタイプシステムの設計・実装について説明した。そして、実データを用いた評価実験を実施し、評価・考察を行った。従来のユーザベースの協調フィルタリングのみの推薦ではなく、他の要素を付加させることで推薦アイテムリストにブレを生じさせ、その結果、推薦アイテムリストの多様性を向上させることが出来た。これにより、従来の推薦システムの似たようなアイテムばかりが推薦されるという問題点が改善され、多様性が高くなることで、意外性のある推薦を行うことが出来ると考えられる。

今後の課題を以下に示す。

- ・ Webアプリケーションとしての実装
現在のシステムはクライアントアプリケーションであるが、これを Web アプリケーションとして実現する。これにより、Web を用いた大規模な検証実験を可能とする。
- ・ 各機構のパラメータ設定
推薦システムを適用する環境や対象とするユーザに合わせた、各機構の最適なパラメータ設定を検討する。
- ・ セレンディピティの評価

セレンディピティは数値化することが困難である。そこでシステムをユーザに使用してもらい、アンケートなどで推薦システムのセレンディピティを評価する。

参考文献

- [1] 土方嘉徳, 「嗜好抽出と情報推薦技術」, 情報処理, Vol.48, No.9, pp963-964, 2007.
- [2] Apache Mahout プロジェクト <http://mahout.apache.org/>
- [3] MovieLens <http://www.movielens.org/>
- [4] 加藤由花, 川口賢二, 箱崎勝也, “オンラインショッピングを対象とした正確性と意外性のバランスを考慮したリコメダシステム”, 情報処理学会論文誌: データベース, Vol. 46, No. SIG13(TOD27), pp. 53-64, 2005.
- [5] 清水拓也, 土方嘉徳, 西田正吾, “発見性を考慮した協調フィルタリングアルゴリズムに関する複数方式の検討”, DEWS2007 L2-2, 2007.
- [6] 小川祐樹他, “多様性のある協調フィルタリングアルゴリズムの提案 -Amazon.com の関連商品ネットワークを用いて-”, 日本経営情報報告会 2007 年春季全国研究発表大会予稿集, pp108-111, 2007.