

親しさを考慮しての推薦リストの有効性向上と融合評価

呉 佳超[†] 高山 毅[†] 佐藤 永欣[†] 村田 嘉利[†]

[†] 岩手県立大学大学院ソフトウェア情報学研究科 〒020-0193 岩手県岩手郡滝沢村菓子 152-52

E-mail: [†] g231j201@s.iwate-pu.ac.jp, {takayama, nobu-s, y-murata}@iwate-pu.ac.jp

あらまし 近年、情報推薦において、推薦の精度のみならず満足度等の有効性を向上させることに関心が高まっている。著者らはこれまで、ユーザ間の親しさを考慮して有効性を向上させることに取り組んでいる。しかし、これまで提案してきた手法では、「①コールドスタート問題」、「②情報推薦のための既存技術として代表的な協調フィルタリングとの融合方式」についての検討が、充分ではない。本稿では、①、②の解決のために、「親しさを考慮する推薦方式」と「協調フィルタリング」との適切な融合方式を提案するとともに、有効性の拡充を目指す。テレビ番組を商品アイテムとして評価実験を行なったところ、提案方式は有効との結果が得られた。

キーワード アイテム推薦, 協調フィルタリング, 有効性, 親しさ, テレビ番組

1. はじめに

本稿では、顧客間の親しさを考慮して作成する推薦リストについて、有効性をより向上させるための提案を行なう。具体的には、協調フィルタリングとの適切な融合方式を、実験的に検討する。近年、レコメンデーションに関する研究が、活発化している([1]-[3])。中でも、推薦リストの有効性を向上させる研究が、活発化している。具体的には、「discovery ratio (発見性)」[4]、「serendipity (意外性)」[5]等を考慮したアプローチが代表的である。そこでは、基礎データとして顧客の購買履歴や嗜好データを用いているが、それらの顧客間の親しさの考慮は、充分とはいえない。

一般に、人間は自分と親しい人が好む商品に興味を惹かれる場合がある。著者らの研究グループでは、顧客間の親しさを考慮することにより、商品の推薦リストの有効性を向上させることを目指している[6]。しかし、これまで提案してきた手法では、「①コールドスタート問題」、「②協調フィルタリングとの融合方式」についての検討が、充分ではない。本稿では、①、②の解決のために、「親しさを考慮する推薦方式」と「協調フィルタリング」との適切な融合方式を提案するとともに、有効性の拡充を目指す。

これ以降、本稿は以下のように構成されている。まず、2章で関連研究として、協調フィルタリングと推薦リストの有効性を向上させる取り組みを紹介する。3章で、親しさを考慮しての推薦リストの有効性をより向上させるための、本稿での提案手法を説明する。4章で放送番組の推薦を題材として、提案手法の評価実験について述べる。最後に5章で、結論と今後の展望を述べる。

2. 関連研究

2.1 協調フィルタリング

協調フィルタリングの基本的な考え方は、推薦リストの送付先であるアクティブユーザと好みの似たユーザグループが

好むアイテムを、そのアクティブユーザに推薦するというものである。アイテムの内容は見ず、ユーザがどのアイテムにどの評価値を付けたかという情報のみから好みの近いユーザを同定し、アクティブユーザの未評価のアイテムの評価値を予測する。ここで変数を定義する。アクティブユーザ u_a を含む n 人の顧客集合 U を、 $U = \{u_1, u_2, \dots, u_a, \dots, u_n\}$ 、 m 個の商品アイテムの集合 I を、 $I = \{i_1, i_2, \dots, i_m\}$ とする。また、 s_{ij} は、ユーザ u_i のアイテム i_j に対する評価値とする。アクティブユーザ u_a と他のユーザ u_i の類似度 ρ_{ai} は、共通に評価しているアイテムについてのPearson相関で(1)式で測る。ここで y_{ai} は二人が共通に評価したアイテムの集合、 \bar{s}_i 、 \bar{s}_a は、共通に評価したアイテムに対してのそれぞれ、 u_i 、 u_a の平均評価値である。

$$\rho_{ai} = \frac{\sum_{k \in y_{ai}} (s_{ak} - \bar{s}_a)(s_{ik} - \bar{s}_i)}{\sqrt{\sum_{k \in y_{ai}} (s_{ak} - \bar{s}_a)^2} \sqrt{\sum_{k \in y_{ai}} (s_{ik} - \bar{s}_i)^2}} \quad (1)$$

(1)式を用い、 u_a が未評価の i_j の評価値は、各ユーザの i_j への評価値の加重平均で(2)式で予測する。ここで x_j はアイテム j を評価済みのユーザの集合である。

$$s_{aj} = \bar{s}_a + \frac{\sum_{i \in x_j} \rho_{ai}(s_{ij} - \bar{s}_i)}{\sum_{i \in x_j} |\rho_{ai}|} \quad (2)$$

2.2 推薦リストの有効性を向上させる取り組み

清水らは、発見性(discovery ratio)を考慮した協調フィルタリングアルゴリズムを提案している[4]。ユーザから見て、アイテムの既知・不既知の概念を導入し、ユーザ間とアイテム間、双方の類似度を計算する。そして、ユーザが知らずかつ好むアイテムを推測して、推薦リストの有効性を高めている。

村上らは、推薦の意外性(serendipity)向上のための手法を提案している[5]。そこでは嗜好モデルに加え習慣モデルを導入し、習慣的にアクセスしている可能性のあるアイテムには意外性が低いと考える。そして「嗜好モデルによる予測結果」から「習慣モデルにより意外性の低いアイテム」を除き、推薦リストの有効性を高めている。

しかしながら、これらの先行研究はユーザ間の親しさの考慮が充分とは言えない。

近年、Twitter、Facebook他のソーシャルメディアで、個人の意見を述べたり、友人がそれに反応することが活発化している。本研究では前提として、ソーシャルメディアから「各ユーザの嗜好」、「ユーザ間の親しさ」が得られると仮定する。この仮定の妥当性は、2.3節で述べる。親しさについては具体的に、中立的な状態を3とし、「良い：5」～「悪い：1」までの5段階のデータが得られると仮定する。

著者らの研究グループは、ソーシャルメディア等から取得可能な「ユーザ間の親しさ」に基づいて、推薦リストの有効性を向上させることを提案している[6]。そこでは、従来の協調フィルタリングへ顧客間の親しさを以下の方法1、2のいずれかによって盛り込んでいる。

方法1は、親しいユーザのデータのみ使う方式である。具体的には、親しさの閾値を設定し、前節の(1)、(2)式でアクティブユーザから見て閾値以上のユーザのデータのみ利用する。

方法2は、アクティブユーザから見ての他の各ユーザの重要度を、親しさに基づいて重み付けする方式である。まず、以下の式により親しさ F を0.0~1.0に正規化して、正規化された親しさ F' を得る：

$$F' = \frac{F-1}{4} \quad (3)$$

次に(1)式の ρ_{ai} に F' を掛け、重み付き類似度 ρ'_{ai} を得る：

$$\begin{aligned} \rho'_{ai} &= F_{ai}' \times \rho_{ai} \\ &= \frac{F_{ai}' \times \sum_{k \in y_{ai}} (s_{ak} - \bar{s}_a)(s_{ik} - \bar{s}_i)}{\sqrt{\sum_{k \in y_{ai}} (s_{ak} - \bar{s}_a)^2} \sqrt{\sum_{k \in y_{ai}} (s_{ik} - \bar{s}_i)^2}} \quad (4) \end{aligned}$$

この ρ'_{ai} を、(2)式の嗜好の予測式で ρ_{ai} の代わりに導入する。

試作システムを用いて、i) 方法1に基づく推薦、ii) 方法2に基づく推薦、iii) 通常の協調フィルタリングに基づく推薦、の三つを相対比較する実験を行なった。その結果、方法1、2のいずれも、通常の協調フィルタリングより有効との評価が得られた。

しかし方法1は、ユーザを限定するので、通常の協調フィルタリングよりコールドスタート問題が発生しやすい欠点があ

る。また方法2では、 ρ と F' を掛けるというアプローチを採ったが、融合方式として最善かは、明らかではない。

2.3 ソーシャルメディアとレコメンデーション

近年、ソーシャルメディアから、レコメンデーションのために有意義な情報を取り出す研究が、活発化している。

2.3.1 ソーシャルメディアからの親しさ情報の取得

文献[7]は、SNSで密接な関係にある、'link prediction/ friendship prediction' と 'interest targeting/ service recommendation' を、一つの枠組みFIP(Friendship-Interest Propagation)モデルでモデル化する過程で、ソーシャルメディアからユーザ間の親しさに関する情報を得ている。彼らの研究では、link predictionはlinkがあるか否かのバイナリの予測に留まっている。我々は彼らの考え方で親しさの程度まで多段階で表現した場合、その成果をどう利用できるか、という視点で進めていく。

実際、文献[8]は、Facebookからユーザ間のtie(繋がり)の強弱を予測するモデルを提案している。彼らの実験結果によれば、約85%の精度を達成している。彼らは、この研究成果に基づき、「ソーシャルメディアの設計」、「プライバシー制御」、「メッセージルーティング」、「友達紹介」、「情報の優先付け」が可能である旨、主張している。我々は更なる応用領域として、アイテム推薦を検討している。

2.3.2 レコメンドするものの粒度

文献[1]は、ソーシャルネットワーク上でユーザが過去に取ったアクション、具体的には、view, friending, browsing, interacting with other usersに基づいて、興味を持ちそうな投稿を推薦する。各行動のregularity(規則性)に基づき評価し、普段前向きなアクションを取ることが多いもの(仲が良い人の投稿等)がfeedされたら、推薦している。また、文献[2]では、Flickrを対象としての、投稿された写真のレコメンデーションを行なっている。彼らはタグに基づき各写真をカテゴリ分けし、カテゴリごとに信頼度を付与して、メタデータとコメントに基づき推薦アイテムを決定している。

これらの研究と我々の研究の違いは、以下のようである。すなわち、文献[1]や[2]では、レコメンドされるアイテムは、投稿されたものそのものである。一方、我々の研究では、投稿されたもの自体ではなく、ソーシャルネットワーク上で話題に挙がっているものをレコメンドすることを目指している。

2.3.3 プライバシー侵害とのトレードオフ

文献[3]は、ソーシャルメディアからレコメンデーションのパーソナライゼーションをする場合に、「プライバシーの侵害」と「レコメンデーション精度」のトレードオフ関係を、定量的に分析している。レコメンデーションの根拠となるデータを隠すほどプライバシーは守られるが、アクティブユーザから見ると、関心と呼ばないレコメンデーションになって

しまう。一方で、レコメンデーションの根拠となるデータとして、どのユーザのデータであるかを開示することは、プライバシー保護の面で問題がある。一般に、あるアクティブユーザの、あるアイテムに対する予測評価値が高ければ即、そのアクティブユーザが購入することに結び付くわけではない。そこで我々は、特定のユーザを明示しないまでも、親しいユーザの好みを考慮していることを前面に出して、行動に繋げることを目指している。

検討が必要なのは、ソーシャルメディアからユーザ間の親しさに関する情報を得る際に、プライバシーの保護をどう考えるかである。文献[9]は、会社内でのSNS上での行動分析から、従業員間の仕事上および私的な繋がりや強弱を見出すことに取り組んでいる。近年、Best Buy, Deloitte, Microsoft, IBM等の会社のように、SNSを社内限定で運用して、社内の組織を横断するcommunicationやcollaborationを活性化させようとする動きが活発化している。これらの社内SNSでは、匿名性が確保されにくいデメリットがある一方で、戦略的なチーム編成、人的資源の効果的な配置等でメリットが得られることが知られている。我々も、匿名性の確保については別途議論することにして、SNSからユーザ間の親しさを抽出してアイテム推薦に利用すると、どの程度のメリットが得られるかを吟味していく。

3. 提案手法

2.2節で述べた文献[6]の方式では、類似度 ρ と正規化した親しさ F' を掛けた結果 p' は、 ρ と F' の双方が0.0~1.0の各定義域内で高めの値を取らない限り、 p' の定義域0.0~1.0の中で、相対的に低めの値に集中しやすいという傾向がある。その結果、推定評価値に差が付きにくく、推薦リストの順位が適切でない危険性がある。実際、文献[6]の実験では、31被験者中12名は、三つの方法(通常の協調フィルタリング, 方法1, 方法2)に基づく各推薦リストが一致してしまった。

そこで本稿では、類似度 ρ と正規化した親しさ F' を積ではなく重み付き和で融合することを考える。具体的には、「類似度」と「正規化した親しさ」の各重み α, β を以下のように変えてみて、いずれが適切か探っていく：

- (1) $0.5 \times \text{類似度 } \rho + 0.5 \times \text{正規化した親しさ } F'$
- (2) $1.0 \times \text{類似度 } \rho + 0.0 \times \text{正規化した親しさ } F'$
- (3) $0.0 \times \text{類似度 } \rho + 1.0 \times \text{正規化した親しさ } F'$

(1)は ρ と F' の重みを同じにするものである。以降では、「親しさを中程度考慮する方式」と呼ぶ。(2)は ρ のみを考慮し、 F' は考慮しない。これは、既存の通常の協調フィルタリングに相当する。(3)は F' のみを考慮し、 ρ は考慮しない。以降では、「親しさを最大限考慮する考慮方式」と呼ぶ。本稿では、以上の(1), (3)の二つの方式を提案し、(1)~(3)のいずれが相対的に優れているかを実験を通じて評価する。以降では特に断わ

らずに「重み」と書いた場合には類似度重み α を指すものとする。 β はその定義から、「 $\beta = 1.0 - \alpha$ 」である。

4. 評価実験

文献[5]と同様、テレビ番組を商品アイテムとして用い推薦リストを作成する。被験者は、IT系の学部の研究室に所属する1~4年生と大学院生の計39名である。

4.1 第一回アンケート

4.1.1 実験方法

第一回アンケートでは、実験で用いる番組アイテムを決定する。具体的には、被験者群へよく見る番組を調査する(図1)。一日でも番組数は膨大なので、「嗜好の評価値」をデータ収集する番組を絞り込む。

Q. ほぼ毎回見るか録画する番組を、最大五つまで書いてください。

番組名	放送局	曜日	時間帯

図1 第一回アンケートの内容。

4.1.2 実験結果

延べ176の番組が得られた。文献[10]に基づきそれらをジャンル分けすると、各ジャンルごとの番組数は、重複を省いて計32番組である(表1)。なお、番組の中には、「ニュース/報道」と「情報/ワイドショー」のように、複数のジャンルが付与されているものがある。

表1 ジャンルごとの番組数

ジャンル名	番組数
① ニュース/報道	6
② スポーツ	2
③ 情報/ワイドショー	1
④ ドラマ	2
⑤ バラエティ	5
⑥ 映画	2
⑦ アニメ/特撮	4
⑧ 音楽	2
⑨ 劇場/公演	0
①かつ③	4
④かつ⑥	1
④かつ⑦	1
⑤かつ⑦	1
⑤かつ⑨	1
計	32

4.2 第二回アンケート: 各番組アイテムの評価値とユーザ間の親しさの調査

4.2.1 実験方法

2.2節で述べたように、本研究ではソーシャルメディアから「各ユーザの嗜好」、「ユーザ間の親しさ」の情報を得られると仮定している。本稿では、より簡易にこれらのデータを、アンケート調査で取得する。Q1では、第一回アンケートから定めた32番組の評価値の付与を依頼する(図2)。各番組の評価値は、「5: 好きである, ~ 3: これを知ってはいるが、好きでも嫌いでもない, ~ 1: 好きではない, *: 知らない/わからない」の6段階とする。

Q2では、ユーザ u_a から見たユーザ u_i の親しさ F_{ai} を、「会う頻度」と「親しさの濃度」の二尺度で尋ねる。ここで、「会う頻度(会話はしなくても良い)」は、「5: ほぼ毎日, ~ 3: 週に1~2日, ~ 1: 会う可能性はほぼない, *: 面識がない」の6段階とする。また、「親しさの濃度」は「5: 非常に仲が良い/たくさん話をする, ~ 3: 普通/世間話程度に話をする, ~ 1: 仲が悪い/ほとんど話をしない, *: 面識がない」の6段階とする。

- Q1. 以下の各番組に対して、評価値を付けてください。一番近いものに○を付けて下さい。
 <評価値> <時間帯>
 5. 好きである 朝(日の出~10:00)
 4. 少し好きである 昼(10:00~16:00)
 3. これを知ってはいるが、好きでも嫌いでもない 夕方および夜
 2. あまり好きではない (16:00~日の出)
 1. 好きではない
 *. 知らない/分からない

ジャンル	番組名	放送局名	曜日	時間帯	評価値
ニュース/報道	ニュースZERO	日本テレビ	月~金	夕・夜	5 4 3 2 1 *
	パンキシャ	日本テレビ	日	夕・夜	5 4 3 2 1 *
	ニュースウオッチ9	NHK	月~金	夕・夜	5 4 3 2 1 *
	ニュース7	NHK	月~金	夕・夜	5 4 3 2 1 *
	おはよう日本	NHK	月~金	朝	5 4 3 2 1 *
スポーツ	クロズアップ現代	NHK	月~木	夕・夜	5 4 3 2 1 *
	やべっちfe	テレビ朝日	日	夕・夜	5 4 3 2 1 *
情報/ワイドショー	すぽると	フジテレビ	月	夕・夜	5 4 3 2 1 *
	ヒルナンデス	テレビ岩手	月~金	昼	5 4 3 2 1 *
ドラマ	相棒	テレビ朝日	水	夕・夜	5 4 3 2 1 *
	平清盛	NHK	日	夕・夜	5 4 3 2 1 *
バラエティ	めっちゃ×2イケてるッ	フジテレビ	土	夕・夜	5 4 3 2 1 *
...

図2 第二回アンケートのQ1の内容。

Q2. 以下は一番近いものに○を付けてください。可能な範囲での回答で結構です。

- <会う頻度(会話はしなくても良い)> <親しさの濃度>
 5. ほぼ毎日 5. 非常に仲が良い/たくさん話をする
 4. 週に3~4日 4. 仲が良い/ほどほど話をする
 3. 週に1~2日 3. 普通/世間話程度に話をする
 2. 時々/たまに 2. 仲が良いとは言えない/あまり話をしない
 1. 会う可能性はほぼない 1. 仲が悪い/ほとんど話をしない
 *. 面識がない *. 面識がない

学年	氏名	会う頻度	親しさの濃度
M2	S.S	5 4 3 2 1 *	5 4 3 2 1 *
M2	J.W	5 4 3 2 1 *	5 4 3 2 1 *
M1	S.T	5 4 3 2 1 *	5 4 3 2 1 *
4年	C.B	5 4 3 2 1 *	5 4 3 2 1 *
4年	M.S	5 4 3 2 1 *	5 4 3 2 1 *
...

図3 第二回アンケートQ2の内容。

アンケート結果に基づき、 F_{ai} は以下の式で評価する。

$$F_{ai} = \frac{(\text{会う頻度}) + (\text{親しさの濃度})}{2} \quad (5)$$

ここで(5)式の例外として、親しさの濃度、或いは、会う頻度、いずれかが*の場合は $F_{ai} = 0$ とする。

4.2.2 実験結果

Q1では、39被験者×32番組=1248の100%に当たる1248番組について、評価値を得た(表2)。

表2 各番組に対する評価値の分布

評価値	番組数	割合(%)
5	109	8.7
4	282	22.6
3	321	25.7
2	167	13.4
1	153	12.3
*	216	17.3
合計	1248	100.0

Q2では、39名の被験者へ自分以外の38名への親しさの付与を依頼し、39*38=1482の100%に当たる1482の親しさデータを得た(表3~4)。

表3 会う頻度の分布表

会う頻度	人数	割合(%)
5	186	12.6
4	396	26.7
3	273	18.4
2	359	24.2
1	181	12.2
*	87	5.9
合計	1482	100.0

表4 親しさの濃度の分布表

親しさの濃度	人数	割合(%)
5	137	9.2
4	288	19.4
3	447	30.2
2	232	15.7
1	248	16.7
*	130	8.8
合計	1482	100.0

4.3 第三回アンケート: 提案方式の評価

第二回アンケートの結果によれば、被験者39名中7名は実験対象の32番組中に知らない番組がなく、推薦リストを作成できない。そのため第三回アンケートでは、残りの32名の被験者へ推薦リストを提示して、評価を進める。

1章で述べたように、人間は自分の親しい人が好む商品に興味を惹かれる場合がある。これをより詳しく検討すると、以下の明示的/暗示的の二つの理由が考えられる:

- 「明示的理由」: 親しい人が好んでいることが分かったからこそ、興味を惹かれる。
- 「暗示的理由」: 嗜好が似ているために親しくなっており、言わば無意識的に各商品アイテムに類似の評価値を付与する。

以上を考慮して、第三回アンケートでは32被験者を16名ずつ二つのグループに分けて、実験を進める。

4.3.1 第一グループの実験について

図4は、第一グループに対する第三回アンケートの内容を示している。

以下の各番組をどの程度視聴したいですか？

5. ぜひ視聴したい
4. 視聴しても良い
3. どちらとも言えない
2. あまり視聴したいとは思わない
1. 視聴したくない

1. 親しさを考慮しない推薦リスト(類似度重み = 1.0)

おススメ順位	番組ID	今回評価値	次回評価値
1		5 4 3 2 1	5 4 3 2 1
2		5 4 3 2 1	5 4 3 2 1
3		5 4 3 2 1	5 4 3 2 1
4		5 4 3 2 1	5 4 3 2 1
5		5 4 3 2 1	5 4 3 2 1

2. あなたが親しい人が見ている番組を中程度考慮した推薦リスト(類似度重み = 0.5)

おススメ順位	番組ID	今回評価値	次回評価値
1		5 4 3 2 1	5 4 3 2 1
2		5 4 3 2 1	5 4 3 2 1
3		5 4 3 2 1	5 4 3 2 1
4		5 4 3 2 1	5 4 3 2 1
5		5 4 3 2 1	5 4 3 2 1

3. あなたが親しい人が見ている番組を最大限考慮した推薦リスト(類似度重み = 0.0)

おススメ順位	番組ID	今回評価値	次回評価値
1		5 4 3 2 1	5 4 3 2 1
2		5 4 3 2 1	5 4 3 2 1
3		5 4 3 2 1	5 4 3 2 1
4		5 4 3 2 1	5 4 3 2 1
5		5 4 3 2 1	5 4 3 2 1

図4 第一グループに対する第三回アンケートの内容。

4.3.1.1 実験方法

被験者には、

- (1) $\alpha=1.0$ (=通常の協調フィルタリング)
- (2) $\alpha=0.5$

(3) $\alpha=0.0$

に基づく三種類の推薦リストを、各上位5件で評価させる。ここで順序効果を考慮して、推薦リストの提示順序は公平にシャッフルする。図5が、試作システムの推薦リスト提示画面の例である。

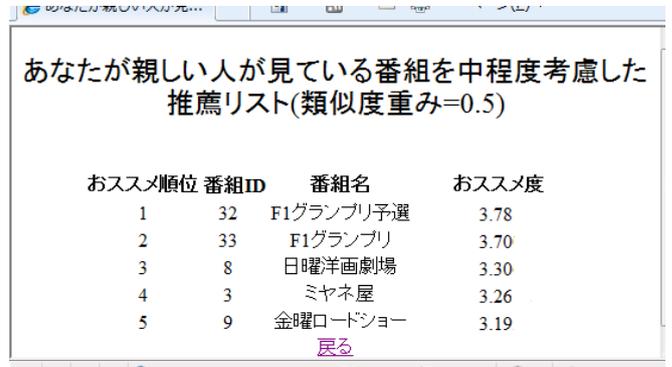


図5 試作システムの推薦リスト提示画面の例。

被験者に、

①EPG(Electronic Program Guide; 電子番組表)に基づく番組情報(紙媒体)

②CM直前の、CM明けの内容予告(映像)

③番組終了時の、次回の内容予告(映像)

の三つを見せ、各番組をどの程度視聴したいかを5段階評価させる。①、②から「今回評価値」を、③から「次回評価値」を得、その相加平均を番組の視聴したい度合とする。

番組映像は、「④番組の内容開始部分」を含めて、③、②、④の優先順で、合計視聴時間が30秒になるようにする。優先順位を設けるのは、②と③は、番組によっては存在しないためである。注意点として、他の推薦リストで出てきた番組は、前に書いたのと同じ評価値を付けるよう、依頼する。



図6 番組映像の区分。

以上で得られた、「番組に対する評価値」に基づき、推薦リストをDCG(Discounted Cumulative Gain)[11]で分析する。DCGはもともと、検索エンジンの検索結果を評価する指標だが、近年、推薦リストの有効性の評価にも使われている。

$$DCG(i) = \begin{cases} G(1) & (if\ i = 1) \\ DCG(i-1) + \frac{G(i)}{\log_2 i} & (otherwise) \end{cases} \quad (6)$$

ここで $G(i)$ は、第 i 位にランク付けされたアイテムの適合性を示す。第二式の狙いは、順位が下がるにつれて重みを軽く

することで、適合性が高いアイテムが上位に出現するほど、DCGの値は大きくなる。

4.3.1.2 実験結果

推薦リストの評価結果をDCGで分析したのが図7である。方法ごとにDCGの平均値を比較している。通常の協調フィルタリングより、重み0.5と重み0.0に基づく推薦リストの方が良い値が得られた。

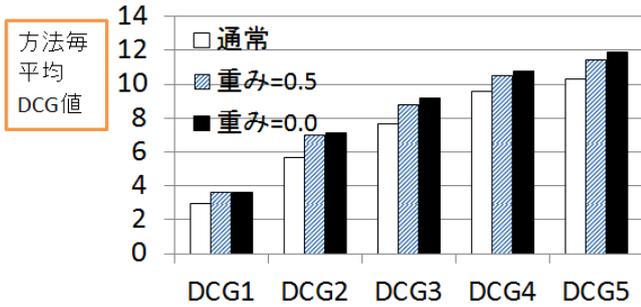


図7 第一グループの DCG

図7から、分散分析[12]より有意差の有無の検定を行なった(表5)。考察として、DCG(1)~(5)のすべてで有意差が出た。

Bonferroniの多重比較によって、どの二者間に有意差があるのかを検定したのが表6の(a)~(c)行である。なお、「>」は有意差があり左の方が右より優位であることを、「≒」は左と右で有意差がないことを示している。「重み0.0 > 重み0.5 > 重み1.0」なる結果が得られている。

表5 図7の分散分析の結果

※有意水準5%での境界値=3.20

DCG	DCG1	DCG2	DCG3	DCG4	DCG5
F 値	5.06	4.98	5.42	5.11	6.33

4.3.2 第二グループの実験について

第一グループの実験では、各推薦リストの命名はどれが提案手法でどれが従来手法か、被験者側から推定できる危険性がある。これは、4.3節冒頭で述べた「明示的理由」を考慮し、敢えてそうしている。第二グループの実験では、明示的理由を外し暗示的理由のみを頼りにした場合の効果を分析する。

4.3.2.1 実験方法

図4のアンケート中の1.~ 3.の三つの推薦リストの見出しを、それぞれ以降のように書き直して進める。

1. 推薦リスト1(重み=1.0)
2. 推薦リスト2(重み=0.5)
3. 推薦リスト3(重み=0.0)

それ以外は、第一グループの実験方法と同じである。

表6 Bonferroniの多重比較に基づく優位性の検定結果

‘>’: 左が右より優位, ‘<’: 右が左より優位, ‘≒’: 有意差がないことを示す

グループ	DCG 重み x vs y	(1)	(2)	(3)	(4)	(5)
		第一(明示+暗示)				
第一(明示+暗示)	(a)0.5 vs 1.0	>	>	≒	≒	≒
	(b)0.0 vs 1.0	>	>	>	>	>
	(c)0.5 vs 0.0	≒	≒	≒	≒	≒
	(d)0.25 vs 1.0	≒	≒	≒	≒	≒
	(e)0.25 vs 0.5	≒	≒	≒	≒	≒
	(f)0.25 vs 0.0	≒	≒	≒	≒	≒
第二(暗示のみ)						
第二(暗示のみ)	(g)0.5 vs 1.0	≒	≒	≒	≒	>
	(h)0.0 vs 1.0	>	>	>	>	>
	(i)0.5 vs 0.0	≒	≒	≒	≒	≒
	(j)0.25 vs 1.0	≒	≒	≒	≒	>
	(k)0.25 vs 0.5	≒	≒	≒	≒	≒
	(l)0.25 vs 0.0	≒	≒	≒	≒	≒
全体						
全体	(m)0.5 vs 1.0	>	>	>	≒	≒
	(n)0.0 vs 1.0	>	>	>	>	>
	(o)0.5 vs 0.0	≒	≒	≒	≒	≒
	(p)0.25 vs 1.0	≒	>	>	>	>
	(q)0.25 vs 0.5	≒	≒	≒	≒	≒
	(r)0.25 vs 0.0	≒	≒	≒	≒	≒

4.3.2.2 実験結果

推薦リストの評価結果をDCGで分析した(図8)。第一グループと同様に、通常の協調フィルタリングより、提案手法の重み0.5と重み0.0に基づく推薦リストの方が良い値が得られた。

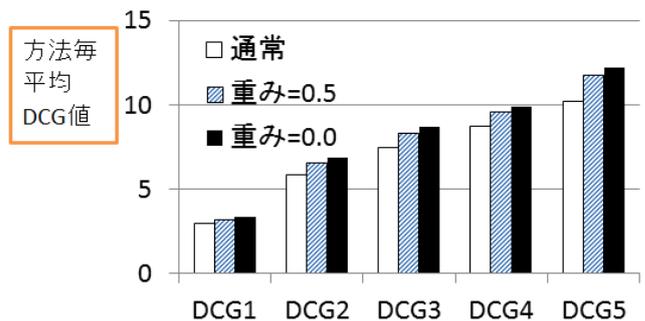


図8 第二グループの DCG

図8から、分散分析により有意差の有無の検定を行なった結果が表7である。考察として、DCG(1)~(5)で有意差が出た。Bonferroniの多重比較によって、どの二者間に有意差があるのかを検定したのが、前掲の表6の(g)~(i)行である。ここでも、「重み0.0 > 重み0.5 > 重み1.0」なる結果が得られた。

表7 図8の分散分析の結果
※有意水準5%での境界値=3.20

DCG	DCG1	DCG2	DCG3	DCG4	DCG5
F値	3.28	5.41	5.80	5.00	8.50

4.3.3 全被験者の実験結果

ここでは、第一、第二グループを併せた全被験者での結果を分析する(図9, 表8)。ここでも通常の協調フィルタリングより、提案手法の $\alpha=0.5$ と 0.0 に基づく推薦リストの方が良い値が得られ、DCG(1)~(5)で有意差が出た。Bonferroniの多重比較によって、どの二者間に有意差があるのかを検定したが、前掲の表6の(m)~(o)行である。全体としても、「重み $0.0 >$ 重み $0.5 >$ 重み 1.0 」なる結果が得られた。

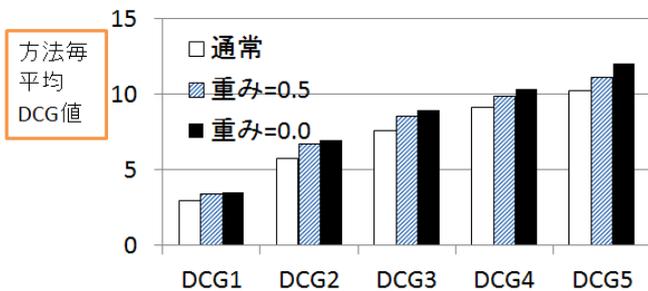


図9 全被験者の DCG.

表8 図9の分散分析の結果
※有意水準5%での境界値=3.11

DCG	DCG1	DCG2	DCG3	DCG4	DCG5
F値	7.75	10.18	11.75	10.90	15.32

4.3.4 類似度重み=0.25の場合のシミュレーション

前項までで、重み 0.5 と 0.0 が通常の協調フィルタリングより優れ、なおかつ、その二者では重み 0.0 の方が相対的に優れるという実験結果を得た。しかし、重みの最適値は 0.0 と 0.5 の間に存在する可能性もある。そこで、ここでは重み 0.25 とした場合の結果を紙上でシミュレートする。データは、実験時に得たものを再利用する。実際にシミュレートしたところ、推薦リストにあらわれた番組は、推薦順位が変化するのみで、すべてこれまでに評価値を付与した番組であった。推薦リストの評価結果をDCGで分析し(図10)、また、分散分析からF値を得た(表9)。

紙幅の都合により、第一、第二グループでのグラフとF値の掲載は省く。それらを含め、重み 0.25 でも通常の協調フィルタリングより、良い値が得られた。Bonferroniの多重比較によって、どの二者間に有意差があるのかを検定したが、前掲の表6の(d)~(f), (j)~(l), (p)~(r)行である。

予想に反し、重み 0.25 は、 0.0 および 0.5 のいずれよりも劣る結果となった。これは、重みを $0.2/0.1/0.01/0.75$ とした場合も同様であった。

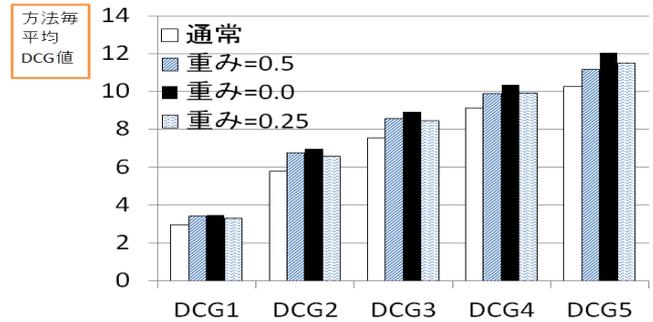


図10 (重み0.25を含む)全被験者の DCG

表9 図10の分散分析の結果
※有意水準5%での境界値=2.68

DCG	DCG1	DCG2	DCG3	DCG4	DCG5
F値	5.81	8.04	9.54	8.00	11.96

4.3.5 和と積の優位性の相対評価

2.2節で述べた、積を用いる方法2[6]と、本稿で提案した和を用いる方法を相対評価する。具体的には、和の方では重みとして $0.0/0.1/0.5$ の三通りを採用し、方法2を加えた四者比較とする。重み 1.0 は、これら四者より劣る結果が既に得られているので外す。データは、本稿/文献[6]の双方を用いる。

まず、本稿のデータで分析した結果が、図11と表10である。図11より、方法2よりも和の三つの方が良い値が得られた。検定結果として、DCG(1)~(5)のすべてで有意差が出た。

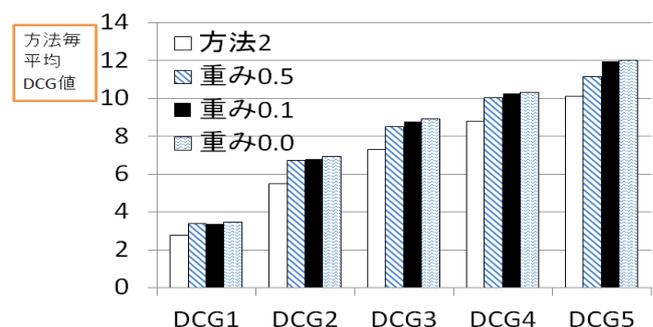


図11 本稿のデータでの DCG

表10 図11の分散分析の結果

DCG	DCG1	DCG2	DCG3	DCG4	DCG5
F値	8.28	12.59	12.12	14.21	13.66

※有意水準5%での境界値=2.68

Bonferroni の多重比較によって、どの二者間に有意差があ

るのかを検定したところ、表 11 の(a)~(c)の結果が得られた。すなわち、DCG(1)~(5)で、和が積より良いとの結果が得られた。

紙幅の都合により詳細なデータは省くが、文献[6]のデータで同様の検定を行なったところ、DCG(5)のみで有意差が出た。Bonferroni の多重比較によって、どの二者間に有意差があるのかを検定した結果が、表 11 の(d)~(f)である。DCG(5)のみで、和が積より良いとの結果が得られた。有意差が出にくかった原因としては、「文献[6]のデータではアニメ番組が突出して多かった」ことが考えられる。

以上より、本稿/文献[6]のいずれのデータを用いても、積より和の方が相対的に良い結果が得られた。

表 11 和と積の優位性の検定結果
 ‘>’: 左が右より優位, ‘<’: 右が左より優位,
 ‘≐’: 有意差がない ことを示す

データ	DCG					
	重み x vs 積	(1)	(2)	(3)	(4)	(5)
本稿	(a)0.5 vs 方法2	>	>	>	>	>
	(b)0.1 vs 方法2	>	>	>	>	>
	(c)0.0 vs 方法2	>	>	>	>	>
文献 [6]	(d)0.5 vs 方法2	≐	≐	≐	≐	>
	(e)0.1 vs 方法2	≐	≐	≐	≐	>
	(f)0.0 vs 方法2	≐	≐	≐	≐	>

5. 結論と今後の展望

本稿では、顧客間の親しさを考慮して作成する推薦リストで、有効性をより向上させるための提案を行なった。具体的には、協調フィルタリングとの融合時の適切な融合方式を実験的に検討した。ユーザを限定しないことでコールドスタート問題発生の危険性は緩和される。また、試作システムを用いた実験で、提案する二つの融合方式はいずれも、通常の協調フィルタリングより良い評価を得られた。二つの融合方式では、親しさを最大限考慮する方式は中程度考慮する方式より、良い結果が得られた。更に、文献[6]の積を用いる方式より、本稿の和を用いる方式は良い評価を得た。

本稿で得られた実験結果は、融合に際し協調フィルタリングをまったく用いず、親しさデータのみを用いた方が有効である可能性まで示唆している。しかし 4.2 節で述べたように、本稿の評価実験では「親しさのデータ」「ユーザの嗜好データ」を、アンケート調査を用いて効率的に獲得している。2.2 節で述べたように、それらはソーシャルメディア等から獲得できる[7][8]ことを前提としている。実際にソーシャルメディア

から獲得してレコメンデーションを行なった場合の有効性については、更なる検討が必要である。

また、本稿のアプローチと関連する取り組みとして、trust-aware (信頼性指向)推薦システムが挙げられる[13]。文献調査の結果、文献[14]がその代表的な取り組みの一つで、本稿のアプローチとの類似性/相違点を検討する必要がある。

その他の今後の展望として、以下の二点が挙げられる: i)被験者実験での被験者人数の拡充, ii)提案手法と発見性[4], 意外性[5]を考慮するアプローチとの相対評価。

参考文献

- [1] J. Freyne, S. Berkovsky, E. Daly, and W. Geyer: “Social Networking Feeds: Recommending Items of Interest,” Proc. of the 4th International Conference on ACM Recommender Systems (RecSys2010), pp. 277-280, 2010.
- [2] A. Gürsel, S. Sen: “Producing Timely Recommendations From Social Networks Through Targeted Search,” Proc. of the 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS09), pp. 805-812, 2009.
- [3] A. Machanavajjhala, A. Korolova, A. D. Sarma: “Personalized Social Recommendations - Accurate or Private?,” Proc. of the 37th International Conference on VLDB (VLDB2011), pp. 440-450, 2011.
- [4] 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, 電子情報通信学会論文誌, Vol.91-D, No.3, pp.538-550, 2008.
- [5] 村上知子, 森紘一郎, 折原良平: 推薦の意外性向上のための手法とその評価, 人工知能学会論文誌, Vol.24, No.5, pp.428-436, 2009.
- [6] J. Wu, T. Takayama, N. Sato and Y. Murata: “Improving Effectiveness of Recommendation List and its Evaluation,” Proc. of the 14th WSEAS International Conference on Mathematical and Computational Methods in Science and Engineering (MACMESE '12), pp.53-56, 2012.
- [7] S. H. Yang, B. Long, A. Smola, N. Sadagopan, Z. Zheng, H. Zha: “Like Like Alike - Joint Friendship and Interest Propagation in Social Networks,” Proc. of the 20th International Conference on WWW (WWW2011), pp. 537-546, 2011.
- [8] E. Gilbert, K. Karahalios: “Predicting Tie Strength with Social Media,” Proc. of the 27th International Conference on ACM CHI (CHI2009), pp. 29-38, 2009.
- [9] A. Wu, J. M. DiMicco, D. R. Millen: “Detecting Professional Versus Personal Closeness Using An Enterprise Social Network Site,” Proc. of the 28th International Conference on ACM CHI (CHI2010), pp. 1955-1964, 2010.
- [10] 社団法人電波産業会: “デジタル放送に使用する番組配列情報標準規格”, ARIB STD-B10 4.6 版, 2008.
- [11] K. J'arelin, and, J. Kek'al'ainen.: IR Evaluation Methods for Retrieving Highly Relevant Documents, Proc. 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp.41-48, 2000.
- [12] A. E. Watkins: Statistics in Action, Analysis of Variance and Anova Cd Package : Understanding a World of Data, John Wiley & Sons Inc, 2008.
- [13] *Private Communication*, 立命館大学情報理工学部・奥健太助教, DEIM2013 本稿の発表会場にて, 2013-3-4.
- [14] P. Massa, P. Avesani: Trust-aware Recommender Systems, Proc. of the 1st International Conference on ACM Recommender Systems (RecSys2007), pp. 17-24, 2007.