原著論文

アイテムが選択された日時を用いた協調フィルタリングと着信メロディのダウンロード履歴を用いた評価†

小澤 順*・内藤 栄一*・コストブ ブラホ*

協調フィルタリングによる情報推薦の手法においては、類似ユーザの検出と、検出された類似ユーザから推薦するアイテムの決定、との2つのステップから構成されている。本論文では、ユーザにアイテムが選択された日時情報を考慮した協調フィルタリングの手法を提案する。提案手法では、所定の期間におけるユーザ間の類似度により、嗜好の類似したユーザを検出する。さらに、推薦するアイテムの決定においても、類似したユーザの所定の期間に選択されたアイテムから求める。これらの期間の長さを変更することにより、協調フィルタリングの推薦性能への影響を実験的に検証する。本実験では、11ヶ月の間に約25万人のユーザによってダウンロードされた携帯電話向けの着信メロディの履歴を用いる。類似ユーザの選択における期間、推薦アイテムの決定における期間の各長さを変更することにより、推薦結果の再現率を用いて推薦性能への影響を評価する。本実験の結果、類似ユーザ選択においては、期間の長さの再現率への影響は少ないが、推薦アイテムの決定においては、期間の長さが影響することがわかった。

キーワード：協調フィルタリング、日時情報、ダウンロード履歴、着信メロディ

1. はじめに

近年、インターネットの普及により、ユーザが膨大なコンテンツにネットワークを通じてアクセスできる環境が整備されてきている。そのため、インターネット上の膨大な情報から必要な情報をユーザが簡単にフィルタリングできることが望まれている[1]。このような課題において、コンテンツに対してユーザが評価した結果を利用して、そのユーザ評価と類似した評価を行った他のユーザを検索し、検索された類似ユーザの評価を基に必要な情報をフィルタリングする協調フィルタリングという手法がある[2]。この手法は、アイテムに対する多くのユーザの評価情報を収集する必要があるが、近年においては、アイテムへのアクセス履歴から自動的に取得し、その履歴を基に推薦を行う手法が提案されている。このように、推薦するアイテムの種類に依存することなく、汎用的なアーキテクチャで情報推薦を実現することが可能になってきている。

協調フィルタリングを実現するためには、アイテム

に対する評価情報からユーザ間の類似度を算出する必要がある。そこで、様々な類似度の算出方法が提案されている[3-5]。しかしながら、アイテムの種類によっては、時間と共に推薦すべき情報が異なる場合がある。例えば、音楽情報に関しては、流行等があるために、ある日時において選択されていたものであっても、時間が経過すると同一のユーザであっても選択すべきでない場合がある。このような課題に対して、協調フィルタリングの分野では、アイテムに対するユーザの嗜好の動的な変化に対応させる手法の提案がなされている。例えば、ユーザの嗜好の抽出を定常的にを行い、推薦を行うときのユーザの嗜好と類似しているユーザを抽出することで情報を推薦するものがある[17]。また、アイテムを評価した日時情報を用いてユーザ間の類似度を算出し、ユーザのアイテムに対する評価値を推定するものがある[18]。

本論文では、類似ユーザを選択するときに、ユーザがアイテムを選択した日時を考慮して嗜好が類似したユーザを選択する。さらに、類似ユーザがすでにアクセスしたアイテムから推薦するアイテムを選択する際にも、そのアイテムにアクセスした日時を考慮して推薦するアイテムを決定する。このように、従来までの協調フィルタリングでの二つの処理において、日時情報を利用する手法を提案する。さらに、各処理においてアイテムにアクセスした日時の情報を利用することにより、推薦性能への各影響度合について、携帯電話

† Collaborative Filtering based on Date of Item Selection and its Evaluation by Ringing Tone Download History
Jun OZAWA, Eiichi NAITO and Vlaho KOSTOV

* 松下電器産業（株）先端技術研究所 知能情報技術研究所
Intelligent Information Processing Technology Laboratory, Advanced Technology Research Laboratories, Matsushita Electric Industrial Co., Ltd.

向け着信メロディの実ダウンロード履歴を用いて評価する。

協調フィルタリングの手法の計算量は、ユーザ数とアイテム数に大きく依存する。そのため、あるユーザに対して推薦するアイテムを決定するとき、そのユーザと類似した嗜好をもったユーザの数が増大すると、計算量も増大する。そこで、ユーザがアイテムに対してアクセスした日時や、推薦アイテムを決定するとき日時情報を用いることにより、計算量を削減することも可能である。また、協調フィルタリングにおいては、一般的に、アイテムを推薦するユーザと類似した全てのユーザのアイテムへのアクセス履歴を用いて推薦アイテムを決定する。本稿では、さらに、類似ユーザの数を制限することにより、計算量を削減したときに、推薦性能に与える影響についても、実データによる評価を行う。

第2章において、本稿で提案する日時情報を用いた協調フィルタリングの定式化について述べる。第3章においては、携帯電話向け着信メロディのダウンロード履歴を用いて、協調フィルタリングにおける日時情報が推薦性能に与える影響について実験的に考察する。第4章では、本稿で提案した手法に対して、従来までの嗜好の変化に対応させる情報フィルタリングとの比較を行うことで本研究の位置付けを明確にする。最後に、第5章においてまとめる。

2. 日時情報を考慮した協調フィルタリング

協調フィルタリングの手法は2つのステップで行われる。1つ目のステップは、アイテムを推薦する対象ユーザと類似するユーザを見つけ出すことである。2つ目のステップは、類似ユーザの評価に基づいて対象ユーザに推薦するアイテムを決定することである。本章では、各ステップにおける日時情報の利用方法について説明する。

2.1 評価データとユーザ間類似度

協調フィルタリングでは、Webサイトやテレビ番組、映画等の各アイテムに対して多くのユーザに評価させる。この評価データを基にユーザ間の類似度を算出し、類似した嗜好をもつユーザを見つけ出す[4]。評価データはアイテムに対するユーザの評価を取得する方法によって2種類のデータがある。一つは、ユーザの普段の操作から評価値を取得する暗黙的評価データである。例えば、Webサイトの訪問履歴から、所定のユーザが評価アイテムであるページに対して訪問した場合には1、訪問していない場合には0という評価

値が与えられる。もう一つの評価データは、ユーザがアイテムに対して直接評価値を入力する明示的評価データである。例えば、評価アイテムの映画に対して、ユーザの嗜好によって0から5までの評価値が直接入力される[6]。

このようにして得られた評価データに基づき、ユーザ間類似度が定義されるが、さらに、ユーザ間類似度の様々な拡張やその有効性を検証するための比較に関する研究も行われている[3]。また、共起関係行列の固有値分解によるファジィクラスタリング[7]を協調フィルタリングのユーザ間類似度へ適用する研究[8]や、簡略ファジィ推論モデルを協調フィルタリングへ適用する試み[9]もある。

しかしながら、アイテムの種類によっては、ユーザに推薦すべきアイテムが日時に依存する場合がある。特に、アイテムが楽曲や映画等の種類の場合には、ユーザの嗜好はアイテムの流行に依存する。例えば、楽曲の場合、楽曲が発表されてすぐに選択するユーザと、楽曲が発表されて十分な期間がたってからその楽曲を選択するユーザとでは、必ずしも嗜好が類似しているとはいえない。また、ある1ヶ月間のアイテムの選択履歴においては、嗜好が類似していたユーザであっても、1年間のアイテムの選択履歴からは、嗜好が類似していない場合もある。このように、ユーザ間類似度を定量化する場合においては、アイテムが選択された日時の情報を考慮して、ユーザ間類似度を算出する必要がある。このユーザ間類似度は、これまでに多く提案されている類似度を算出する手法と融合することが可能である。

本稿では暗黙的評価データとして、 i 番目のユーザがアイテム p_k ($k=1, 2, \dots, K$) に対してダウンロードまたは情報アクセスを行った場合に評価値 $u_{i,k} = 1$ 、その他の場合は評価値 $u_{i,k} = 0$ とする。これにより、 i 番目のユーザに対しては K 個の $\{0, 1\}$ の要素からなる評価ベクトル u_i で評価値データが表現される。例えば、 i 番目のユーザと j 番目のユーザの評価ベクトルは、

$$u_i = (p_1, p_2, p_3, \dots, p_k, \dots, p_K) = (1, 0, 1, \dots, 1, \dots, 0) \quad (1)$$

$$u_j = (0, 1, 0, \dots, 0, \dots, 0) \quad (2)$$

のように表現される。

次に、 $\{0, 1\}$ の要素からなる2つのベクトルの類似度を算出する指標として、Jaccard類似度がある[10]。そこで、 i 番目ユーザと j 番目のユーザの2つの評価の類似度 $S(u_i, u_j)$ をJaccard類似度を用いて、

$$S(u_i, u_j) = \sum_k \min(u_{i,k}, u_{j,k}) / \sum_k \max(u_{i,k}, u_{j,k}) \quad (3)$$

$$(u_{i,k} \neq 0 \text{ or } u_{j,k} \neq 0)$$

により定義する。\$S(u_i, u_j)\$は、0 から 1 までの実数値をとり、類似度が高いほど 1 に近い値となる。

本稿では、(3)式を拡張し、ユーザのアイテムへのアクセス期間を考慮した類似度を定義する。いま、期間TLにおいて、ユーザがアクセスしたアイテムのベクトルを\$u_i^{TL}, u_j^{TL}\$とする。このとき、2人のユーザ\$u_i, u_j\$間の類似度\$S^{TL}(u_i, u_j)\$は、

$$S^{TL}(u_i, u_j) = \sum_k \min(u_{i,k}^{TL}, u_{j,k}^{TL}) / \sum_k \max(u_{i,k}^{TL}, u_{j,k}^{TL}) \quad (4)$$

$$(u_{i,k}^{TL} \neq 0 \text{ or } u_{j,k}^{TL} \neq 0)$$

により算出される。これにより、期間TLにおけるユーザ間の類似度が算出され、期間毎の類似度が定義される。所定の期間において類似していた2人のユーザでも、別の期間においては類似していなかったり、短期的には類似していたユーザでも、長期的な観点においては類似していなかったり、その逆も存在する。本稿では、この期間が、協調フィルタリングのアイテムの推薦結果へ及ぼす影響について考察する。

2.2 推薦アイテムの決定方法

協調フィルタリングにおける2つ目のステップである、推薦アイテムの決定方法について述べる。前節において、類似ユーザの選択を行い、あるi番目のユーザに対して推薦アイテムを決定する場合には、i番目のユーザと類似している別のユーザの集合を選択し、その集合に含まれるユーザにおいて、i番目のユーザが選択していないアイテムを抽出することで推薦アイテムを決定する。具体的には、i番目のユーザに対してアイテム\$P_k\$を推薦すべき度合\$W(u_i, P_k)\$は、

$$W(u_i, P_k) = \sum_{j \neq i} (u_{j,k} \times S(u_i, u_j)) \quad (5)$$

$$(u_{i,k} \neq 1)$$

によって算出される。各\$P_k(k=1, 2, \dots, K)\$の値のうち、i番目のユーザがすでに選択していないものの中で、値の大きい\$P_k\$を推薦アイテムとして決定する。これにより、i番目のユーザと類似したユーザがよく選択しているアイテムを自動的に選択することが可能となり、情報のフィルタリングに利用することができる。

しかしながら、ユーザに推薦するアイテムもまたアイテムにアクセスした日時情報に依存する場合がある。2.1節で述べたのと同様に、楽曲等のアイテムに

おいては、アイテムにアクセスされた日時が新しい方が好まれる場合が多い。このように、ユーザに推薦するアイテムを選択するときもアイテムの日時情報を考慮すべき場合がある。

そこで、本稿では、(5)式を拡張し、推薦するアイテムを、ユーザがアイテムを選択した日時の情報を考慮して決定する。j番目のユーザの期間TRにおける評価ベクトル\$u_j^{TR}\$を定義することで、i番目のユーザに対してアイテム\$P_k\$を推薦すべき度合\$W^{TR}(u_i, P_k)\$は、

$$W^{TR}(u_i, P_k) = \sum_{j \neq i} (u_{j,k}^{TR} \times S(u_i, u_j)) \quad (6)$$

$$(u_{i,k} \neq 1)$$

と算出される。

さらに、(4)式の所定期間TLにおけるユーザ間の類似度\$S^{TL}(u_i, u_j)\$を用いることで、i番目のユーザに対してアイテム\$P_k\$を推薦すべき度合\$W^{TR, TL}(u_i, P_k)\$は

$$W^{TR, TL}(u_i, P_k) = \sum_{j \neq i} (u_{j,k}^{TR} \times S^{TL}(u_i, u_j)) \quad (7)$$

$$(u_{i,k} \neq 1)$$

となる。これにより、期間TLに類似したユーザの嗜好で期間TRでの推薦するアイテムを選択することが可能になる。

3. 着信メロディのダウンロード履歴を用いた評価

2章で定義した日時情報を考慮した協調フィルタリングが推薦結果に与える影響を調べる。本稿では、ユーザが気軽にダウンロードを行い、多くの履歴データを収集できるアイテムとして携帯電話における着信メロディを選択し、その履歴データを用いて推薦性能への影響を調べる。

3.1 携帯電話向け着信メロディダウンロードシステム

本稿で実験に利用した携帯電話向けの着信メロディダウンロードシステムの構成を図1に示す。図1に示すように、携帯電話に対するコンテンツサービスは、サーバ1とサーバ2の2つのサーバから構成されている。サーバ1は、携帯電話がWebサービスにアクセスする際に、はじめに接続するサーバであり通信会社等により運営されている。サーバ1では、携帯電話の所有者に関する「個人DB」を保有し、各個人に対する課金に関するデータベースも保有している。サーバ1のアクセス管理モジュールにより、携帯電話で利用されたパケット量を監視し、個人に対する課金も行っている。サーバ2はコンテンツのサービスを行うものであり、会員にコンテンツのダウンロードが発生した場合には、サーバ1の課金データベースへ課金情報の登録

を行う。同時に、アクセス履歴としてサーバ2のアクセスログDBへも登録される。本稿では、着信メロディのダウンロード履歴が蓄積されているアクセス履歴を用いて、前章で提案した協調フィルタリングの性能検証を行う。

図1 携帯電話向けサービスのシステム構成

実験に用いた着信メロディのダウンロード履歴は2001年11月29日から2002年10月31日までの11ヶ月間蓄積されたものである。ダウンロード履歴における着信メロディの種類は10,478曲、ダウンロードを行ったユーザの数は257,837人、延べダウンロード回数は1,992,417回であった。しかしながら、ダウンロードを行ったユーザのうち大部分のユーザは少数の曲しかダウンロードを行っていないため、協調フィルタリングの性能評価のデータに適さない。そこで、全ユーザの中から、11ヶ月の間に10曲以上ダウンロードを行っていた741人のユーザを実験対象として選択した。

3.2 推薦アイテムの評価指標

前節でも述べたように、本稿で利用する履歴データは全体で11ヶ月のデータである。これらのデータに対して、ユーザによる最新の1ヶ月の間にダウンロードされたアイテムを、過去の10ヶ月の履歴を用いて推定できるか否かを判定する。図2に本稿の履歴の利用概念図を示す。本稿での実験は、類似ユーザを評価するための期間を2、5、10ヶ月の3つの期間を用いて協調フィルタリングの性能の検証を行う。また、推薦アイテムを決定するための期間は、1、2、3、11ヶ月の4つの期間を用いて性能の検証を行う。これらの期間に対して、最新の1ヶ月のダウンロードの履歴を用いて評価する。

図2 実験における期間情報

本実験においては、ユーザから所定の期間にダウンロードされた着信メロディを、それ以前のダウンロードの履歴を用いて再現できる割合によって協調フィルタリングの性能を評価する。一般に、情報検索における検索性能を定量化する指標として再現率と適合率が用いられる[11]。Sをシステムが推薦するアイテムの集合、Qをユーザが選択したアイテムの集合とした場合、システムが提案する推薦アイテムの中に、ユーザが選択したアイテムが含まれる割合を適合率とよび

$$\text{Precision} = |S \cap Q| / |S| \quad (8)$$

で定義される。ただし、 $S \cap Q$ は、SとQの積集合を、 $|\cdot|$ はその集合の要素数を示す。一方、ユーザが選択したアイテムの中に、システムが選択したアイテムが含まれる割合を再現率とよび

$$\text{Recall} = |S \cap Q| / |Q| \quad (9)$$

で定義される。適合率、再現率とも0から1までの値をとり、大きい値であるほど検索の性能が高いことを示している。一般に、再現率と適合率とは互いにトレードオフの関係にあり、一方が高い値をとると、他方の値が低くなる。再現率を測定するためには、被験者が蓄積されている全てのデータを確認し、各被験者の嗜好にあったアイテムをあらかじめ選択しておき、システムの出力と比較する必要がある。一方で、適合率を測定するためには、システムが提案するアイテムに対して、ユーザがアンケート等により、ユーザの好みに合うか否かを判定させることにより測定が可能である。

本稿で提案する協調フィルタリングでは、ユーザの

ダウンロードの履歴を利用して、ユーザ毎に推薦するアイテムからなるページを生成し、ユーザがそのページを表示することで、簡単に着信メロディをダウンロードできるサービスを目指している。そのページにおいて推薦されるアイテムの数は、通常のユーザがダウンロードする着信メロディの数より多い値にあらかじめ設定されており、その中にユーザの嗜好にあった着信メロディが多く含まれることにより、ユーザによる情報アクセスを喚起させ、推薦がなかったときに比べ多くの着信メロディのダウンロードをさせようとするものである。

一方、本稿で対象としている着信メロディのダウンロードの履歴においては、所定の期間にユーザによってダウンロード(選択)される数が大きく異なる。各ユーザの所定の期間において選択されたアイテムを協調フィルタリングで再現するためには、ユーザ毎にダウンロードするアイテム数も推定する必要がある。しかしながら、本手法では推薦する数を考慮した手法となっておらず、推薦するアイテムの数は一定としている。推薦するアイテム数を一定とした場合、適合率は評価期間においてユーザが選択したアイテムの数に大きく依存する。

例えば、システムが推薦するアイテム数が40個の場合、ユーザが評価期間に選択したアイテム数が1個と10個の各場合には、適合率の最大値は各0.025と、0.25と大きく値が異なる。一方、再現率は、ユーザの評価期間に選択したアイテム数に依存せず、0から1までの実数値で評価できる。よって、本稿では、推薦するアイテム数を一定とし、再現率でシステムの推薦性能を評価することとする。

3.3 ユーザ間類似度の期間情報の推薦性能への影響

推薦するアイテムを選択する期間(TR)を1ヶ月または2ヶ月と固定した場合、類似度を算出する期間(TL)を変化させたときに、協調フィルタリングの性能への影響を図3に示す。ただし、システムが推薦するアイテム数は40個とし、所定の期間にユーザがダウンロードした全アイテムに対して、何個のアイテムがシステムによって選択されているか(再現率)を評価指標としている。図3に示すように、類似度を算出する期間が長くなるにつれて、推薦性能を示す(9)式の再現率は向上している。しかしながら、その向上の度合はあまり大きくなく、再現率が約23%から24%へ約1ポイント向上したにすぎない。この傾向は、推薦アイテムを選択する期間が1ヶ月の場合も2ヶ月の場合も同様の傾向である。これより、類似ユーザを選択する際には、長期的な類似ユーザ選択、短期的な

図3 類似度を算出する期間と再現率

類似ユーザ選択のいずれの期間選択であっても、推薦性能に変化があまり見られないことがわかる。長期間において嗜好が類似するユーザを選択するためには計算量は増大する。しかしながら、計算量の増大に伴う推薦性能の向上はあまり見られない。

また、携帯電話の着信メロディの分野においては、嗜好の類似度は短期的に類似していれば、長期的にも類似していることを示している。また、各個人毎に嗜好の変化はあるが、所定の期間で類似した嗜好をもっていたユーザは、同様の傾向に嗜好が変化していると考えられる。

3.4 推薦アイテム決定の期間情報の推薦性能への影響

ユーザ間類似度を算出する期間(TL)を10ヶ月に固定し、推薦するアイテムを決定する期間(TR)を変化させた場合の、協調フィルタリングの推薦性能への影響を図4に示す。なお、評価指標は、前節と同様にシステムが推薦するアイテム数を40個とした場合の再現率を用いる。図4に示すように、推薦するアイテムを選択する期間を長くすると(9)式の再現率が下がっているのがわかる。これは、推薦期間を長期間にすると、推薦するアイテムにおいて、過去のアイテムが多く含まれるようになる。しかしながら、本評価で利用している着信メロディのアイテムにおいては、新曲のアイテムが多く選択されているため、過去のアイテムを推薦結果として選択すると、再現率が低下する傾向にある。推薦するアイテムを選択する期間を短くする

図4 アイテムを決定する期間と再現率

と、再現率は向上するが、推薦できるアイテムの数が限定されてくる。

3.5 類似ユーザ数による推薦性能への影響

協調フィルタリングにおいては、ユーザ数の増加とアイテムの増加によって計算量が増大する。協調フィルタリングの計算量を削減する研究として、アイテム間の相関を利用したもの[12]、ユーザのクラスタリングによるもの[13]がある。このように、協調フィルタリングにおいては、アイテムに対して2次情報を付加する必要はないが、ユーザ数やアイテム数が増加すると、類似ユーザの検索や、推薦アイテムの順序付けに、多くの計算量が必要になるという実用上の課題がある。

本論文で提案しているように、日時情報を利用した類似度の算出や、日時情報を利用した推薦アイテムの選択により、協調フィルタの計算量の削減を行うことが可能になる。他に計算量を削減する方法として、どれだけの類似ユーザの履歴を用いて推薦アイテムを決定するかによって、計算量を削減することが可能になる。しかしながら、類似ユーザ数を削減すると、推薦アイテムの種類が少なくなり、推薦性能が低くなる可能性がある。そこで、類似ユーザの数が、推薦性能に与える影響についての実験を行った。

ユーザ間類似度を算出する期間を10ヶ月に固定する。次に、推薦アイテムを決定する期間を1, 2, 3, 11ヶ月に変化させる。推薦アイテムを決定する各期間に対して、推薦アイテムを決定するのに利用する類似ユーザの数を変化させた推薦結果を算出する実験を行い、その結果を図5に示す。図5では、推薦アイテムを決定する期間に寄らず、類似ユーザの数に対して一定の傾向を示している。例えば、推薦してアイテムを決定する期間が11ヶ月の場合、類似ユーザの数が約2000人の時点で最大値をとり、それよりも類似ユーザの数を増加させると、推薦における再現率の低

下が見られる。これは、推薦アイテムを決定する期間が1, 2, 3ヶ月の場合においても、類似ユーザの数が所定の値よりも増加させることで、再現率の低下が見られる。

一般に、類似ユーザ数が多ければ多いほどより良い結果が出ることが期待されるが、この場合は、類似ユーザ数が増加すると、すべてのユーザに共通に人気のある曲(売れ筋の曲)が推薦されがちになり、品質が悪化するという結果が得られた。一方、類似ユーザ数が少なすぎても、少数の類似ユーザ特有の嗜好が現れがちになり、やはり品質は悪化する。すなわち、アイテムの種類によって、推薦の再現率を高めるためには、所定の値の類似ユーザ数が存在することになる。その値の類似ユーザ数よりも類似ユーザ数を増加させても、計算量が増大するだけで、推薦性能が悪化することを示している。

4. 関連研究

本章では、従来までの嗜好の変化に追従させる情報フィルタリングと本稿で提案した日時情報を利用した協調フィルタリング手法との比較について考察する。

4.1 嗜好の変化に追従させる情報フィルタリング

情報へのアクセス履歴から、ユーザが好むコンテンツを推定することで情報のフィルタリングを行うシステムが提案されている[14]。これらのシステムにおいては、ユーザの好みを情報のアクセス履歴から、キーワードとその重みとで表現し、情報のフィルタリングを行うシステムであるため、アクセス履歴を所定の期間に限定することで、ユーザの嗜好に追従させることが可能になる。協調フィルタリングの分野においても、ユーザの嗜好の変化を考慮した手法が提案されている。例えば、ユーザのプロファイルをキーワードとその重みで表現し、ユーザの興味の推移に伴って、キーワードやその重みを動的に変化させるものである[17]。さらに、変化後の興味と類似しているユーザを抽出することで、協調フィルタリングの機能を実現するものである。これらのシステムが対象としているものは、テレビ番組やWebアクセスなどのテキスト情報で情報の特徴を表現できるものを対象としており、ユーザの好みをキーワードとその重みで表現できるものである。これに対して本研究は、携帯電話向け着信メロディ等のキーワードで情報の内容を表現しにくいものを対象にした協調フィルタリング手法である。

また、ユーザの好みの変化をあらかじめ嗜好の推移モデルとして構築し、そのモデルを用いてユーザの好みの推移を予測する検索システムが提案されている。

図5 類似ユーザ数と再現率

例えば、ユーザの好みに対して、静的な好みと動的な好みに分類し、ある属性で検索したユーザは、次に別の属性で情報を検索するという状態推移モデルを定義し、そのモデルを用いて、楽曲を検索するシステムを開発している[15]。また、ユーザがキーワードを用いて検索するシステムにおいて、ユーザから入力されたキーワードの履歴を用いて、仮説推論を行い、ユーザの興味の変化を抽出する研究もある[16]。このように、あらかじめ嗜好の推移モデルを構築するためには、対象に応じた分野に関する知識が必要になる。これに対して、本稿での協調フィルタリングでは、日時を考慮する期間等のパラメータを設計的に決定する必要があるが、分野に依存した知識を構築する必要はない。

4.2 アイテムの選択日時を考慮した協調フィルタリング

ユーザが入力した評価値の類似度の算出に、評価が入力された日時情報を利用する協調フィルタリングの手法が提案されている[18]。このシステムにおいては、評価された日時の差が大きいくほど、重みが低くなる指数関数 $f(t)$ を定義し、この指数関数を用いてユーザ間類似度を算出している。例えば、アイテムに対して評価を行った日時の差異 t に対して、 $f(t)$ は定数 λ によって、

$$f(t) = \exp(-\lambda t) \quad \lambda > 0, t > 0 \quad (11)$$

という式により、ユーザ間の類似度を定義している。ここで、 $f(t)$ は、 $t=0$ のときに、1.0となり、評価された日時の差異 t の値が大きくなるにつれて、値は小さくなり、類似度が減少する。(11)式の類似度を用いて、あるユーザと類似した評価を入力した類似ユーザを抽出し、そのユーザが未入力の評価値を算出するものである。このようにして、算出された評価値と、実際に入力した評価値とを比較することにより、手法の性能測定を行っている。評価対象として、EachMovie[6]のデータと、GroupLens[4]で利用されているデータに適用し、最適な λ の値を求めることで、手法の有効性を確認している。この手法は、協調フィルタリングの結果を出力する上で、ユーザが評価を入力した日時情報を利用するという点で、本研究との類似点が多い。しかしながら、本研究とは、次の3つの点において異なっている。

(1) 日時情報の反映のさせ方と計算量について

指数関数によってユーザ間類似度を算出する協調フィルタリングと本稿での協調フィルタリング手法で

のユーザ間類似度の算出のユーザ間類似度への日時情報の影響度合を図6に示す。指数関数によってユーザ間類似度を算出する協調フィルタリングにおいては、評価を入力した日時の差異の長さによって、類似度への反映のさせ方が異なっている。日時情報が大きく離れているユーザの評価も小さい割合であるが考慮する類似度の算出方法となっている。一方、本手法においては、所定の期間(0~TL)において評価された値に対しては、類似度に対して同等に評価している。所定の値(TL)以上に日時情報が離れているユーザに対しては評価の値を全く考慮しないようにしている。このように、日時情報の影響度合い(重み)が異なっている。

図6 日時情報の考慮度合の比較

協調フィルタリングにおいては、評価値の出力または推薦結果を出力する上で、多くの計算量を必要とし、計算量を削減するための工夫がなされている。N人のユーザがK種類のアイテムに対して評価した値を用いて出力を計算するためには、評価値の比較演算を $O(K \times N)$ 回のオーダーで計算する必要がある。指数関数を用いたユーザ間類似度を算出するためには、すべてのユーザに対して日時情報を考慮するため、 $O(K \times N)$ に対して日時情報を用いて実数値演算を行う必要がある。一方、本手法においては、対象となるユーザがK個のアイテムに対して評価を入力した日時情報を用いて、所定期間内のユーザの評価値を考慮し、期間外の評価値は考慮しない。よって、協調フィルタリングの類似ユーザを抽出する処理においては、本手法の計算量の方が少なくなる。

(2) 手法における日時情報の利用について

文献[18]の協調フィルタリングの手法においては、日時情報をユーザ間の類似度を算出する処理にのみ利用していた。これに対して、本手法においては、(4)式のユーザ間類似度を算出する処理と、(6)式を用いて類似ユーザの評価を基に新たな評価を出力する協調

フィルタリングの2つの処理において、日時情報を利用している。この点において異なっている。さらに、前章でも述べたように、ユーザ間類似度の算出において日時情報を利用しても、再現率の向上は少なく、それよりは、推薦コンテンツを出力する処理において、日時情報を利用した方が、再現率が向上することを示している。すなわち、文献[18]で日時情報をユーザ間類似度の算出に用いていることで、推薦性能の向上が報告されているが、それよりは、推薦コンテンツを決定する処理で日時情報を利用した方が、携帯電話向け着信メロディのコンテンツでは有効であった。

(3) 有効性の評価実験対象の違いについて

文献[18]では、各アイテムに対して0から5までの値をユーザが入力し、その評価値を予測できるか否かを判定している。これに対して、本研究においては、ユーザがダウンロードしたか否かという、ユーザが意図せずに行った評価対象について、日時情報の利用の有効性を示した。ユーザが入力した値をシステムが推定できるか否かを評価として利用している点では同一であるが、2章で述べたように、文献[18]ではユーザの明示的評価の値に対する評価実験であるのに対して、本論文では暗黙的評価に対する評価実験を行っている。

5. おわりに

協調フィルタリングのユーザ間類似度の算出と、推薦アイテムの決定において、ユーザのアイテムへのアクセス日時の情報を利用した手法を提案した。さらに、携帯電話向け着信メロディのダウンロードの履歴を用いて、本手法による推薦性能への影響度合を実験した。実験の結果、類似ユーザの選択においては、ユーザ間を比較する期間を長くすると、推薦性能は向上するが、大きな優位差は確認できなかった。しかしながら、推薦アイテムを決定することにおいては、期間を限定することにより、推薦の再現率の向上が確認できた。また、類似ユーザの数を所定の数以上に増加させても推薦性能への影響は少ないことがわかった。これらにより協調フィルタリングの性能を落とすことなく、システムを実現するための計算量を削減することが可能になる。

本稿においては、アイテムが選択された日時の情報を用いた協調フィルタリングと、その推薦性能への影響について携帯電話の着信メロディのダウンロード履歴を用いて評価した。しかしながら、この影響は、アイテムの種類によっても大きく傾向が異なることが予想される。協調フィルタリングが適用される書籍や映

画やテレビ番組等のコンテンツにおける評価も必要である。また、本システムを実運用することで、ユーザの利便性向上の度合を定量的に分析し、推薦のためのパラメータの調整やユーザインタフェースの評価を行う必要がある。

参 考 文 献

- [1] 土方：情報推薦・情報フィルタリングのためのユーザプロファイリング技術，人工知能学会誌，Vol.19，No.3，pp.365-372，2004.
- [2] 梅木：ネットワークコミュニティ形成支援技術，人工知能学会誌，Vol.14，No.6，pp.943-950，1999.
- [3] Breese, J.S., Heckerman D., and Kadie, C.: Empirical Analysis of Predictive Algorithms for Collaborative Filtering, Proc. of 14th Conf. of Uncertainty in Artificial Intelligence, pp.43-52, 1998.
- [4] Resnick, P., Iacovou, N., Sushak, M., Bergstrom, P., and Riedl, J.: GroupLens: An Open Architecture for Collaborative Filtering of Netnews, Proc. of ACM Conference, Computer Support Cooperative Work (CSC94) pp.175-186, 1994.
- [5] 大杉，門田，森崎，松本：協調フィルタリングに基づくソフトウェア機能推薦システム，情報処理学会論文誌，Vol.45，No.1，pp.267-278，2004.
- [6] <http://www.research.digital.com/SRC/EachMovie>
- [7] 井上，浦浜：共起関係行列に基づくファジークラスタリング，電子情報通信学会論文誌，Vol.J84-D-II，No.12，pp.2703-2707，2001.
- [8] 井上，浦浜：ファジークラスタリングに基づく協調フィルタリング，電子情報通信学会論文誌，Vol.J84-A，No.4，pp.559-561，2001.
- [9] 片山，渡邊：ファジィ推論モデルによる協調フィルタリング，第21回ファジィシステムシンポジウム，pp.237-238，2005.
- [10] Salton, G. et al. "Introduction to Modern Information Retrieval," McGraw-Hill, 1983.
- [11] 徳永：情報検索と言語処理，東京大学出版会，1999.
- [12] 井上，浦浜：品目間相関に基づく協調フィルタリング，電子情報通信学会論文誌，Vol.J85-D-II，No.3，pp.533-536，2002.
- [13] 井上，浦浜：人のクラスタリングによる協調フィルタリングの高速化，電子情報通信学会論文誌，Vol.J84-D-II，No.12，pp.2700-2702，2001.
- [14] 隆，渡辺，樽口：履歴情報を用いたTV番組選択支援エージェント，情報処理学会論文誌，Vol.42，No.12，pp.3130-3143，2001.
- [15] 井原，金田，上野，金山：情報がもつ属性を用いたユーザ動的好み推定法の提案およびその評価，電子情報通信学会論文誌，Vol.J80-A，No.4，pp.700-707，1997.
- [16] 大澤，須川，谷内田：ユーザの変化する興味を理解し表現する文献検索支援システムIndex Navigator，人工知能学会論文誌，Vol.13，No.3，pp.461-469，1998.
- [17] Uchiyigit G., and Clark K. L.: A Multi-Agent Architecture for Dynamic Collaborative Filtering, Proc. of the 5th International Conference on Enterprise Information Systems, pp.363-368, 2003.
- [18] Ding Y., and Li X.: Time Weight Collaborative Filter-

ing, Proc. of the 14th ACM International Conference on Information and Knowledge Management, pp.485-492, 2005.

(2005 年12月 4 日 受付)

(2006 年 7 月16日 採録)

[問い合わせ先]

〒619-0237 京都府相楽郡精華町光台3-4

松下電器産業(株) 先端技術研究所

小澤 順

TEL: 0774-98-2507

FAX: 0774-98-2587

E-mail: ozawa.jun@jp.panasonic.com

著者紹介

おざわ じゅん
小澤 順 [正会員]

1988年大阪府立大学工学部機械工学科卒業。1990年東京工業大学総合理工学研究科システム科学専攻修了。同年、松下電器産業(株)入社。1993年4月より2年間国際ファジィ工学研究所へ出向。現在、松下電器産業(株)先端技術研究所に勤務。情報検索インタフェース、個人適応モデル、ネットワークエージェントに関する研究に従事。工学博士。1992年電気関係学会関西支部連合大会講演会奨励賞受賞。1999年大阪電気通信大学客員助教授。2000年から2001年 理化学研究所脳科学総合研究センター非常勤研究員を兼任。情報処理学会、人工知能学会、ヒューマンインタフェース学会等の会員。

Vlaho Kostov [非会員]

He received a Ph.D. in Engineering from the Tokyo Metropolitan Institute of Technology, Department of Production, Information and Systems Engineering in 2002. He worked at United Nations Communications Department 1995-1996, T-Mobile-Germany 1996 and at ASCOM-Switzerland 1996-1997. He received a Monbusho Grant (Japanese Ministry of Science) 1997-2001 and Atsumi Foundation Grant 2001-2002. He is a member of the IEEE and JSKE. His research interests include Wearable computing, Human-machine interfaces, Emotional Engineering, non-verbal communication, Ambient Intelligence and IP services. Currently he is working at Panasonic Corporate R&D in Japan on new product planning and development.

うへはら えいいち
内藤 栄一 [非会員]

1988年東京大学工学部金属材料科学科卒業。同年、松下電器産業(株)入社。中央研究所、本社ネットワーク事業推進本部などを経て、現在同社先端技術研究所に勤務。ファジィ情報検索、ユーザモデリング、個人適応の研究に従事。1993年電気関係学会関西支部連合大会講演会奨励賞受賞。

Collaborative Filtering based on Date of Item Selection and its Evaluation by Ringing Tone Download History

by

Jun OZAWA, Eiichi NAITO and Vlaho KOSTOV

Abstract :

Collaborative filtering method consists of two steps. One is detection of similar users, and the other is selection of recommended items through the detected users. In this paper, we propose a method of collaborative filtering that considers the date of item selection by the relevant users. The method detects the relevant users through the similarity of users in a fixed period. Furthermore, recommended items are determined from the sets selected in another fixed period. Influence on recommendation accuracy is examined by changing the said two periods. The ringing tone download history data by about 250 thousand users in 11 months are employed for the examination. The influence is evaluated using recall by changing the periods for similar users detection and recommended items selection. As the results, it was shown that the influence of the period for similar users is low, and that of the period for items selection is not low.

Keywords : collaborative filter, item selection date, download history, ringing tone

Contact Address : **Jun OZAWA**

Advanced Technology Research Laboratories, Matsushita Electric Industrial Co., Ltd.

3-4 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0237, JAPAN

TEL : 0774-98-2507

FAX : 0774-98-2587

E-mail : ozawa.jun@jp.panasonic.com