

推薦と格付けの相互作用に基づく 推薦アルゴリズムの性能評価

山下 晃 弘^{†1,†2} 川村 秀 憲^{†1}
飯塚 博 幸^{†3} 大内 東^{†1}

Web上の情報量拡大にともなって、オンラインストアなどでは推薦システムを導入する事例が増加している。同時に、これまで様々な推薦アルゴリズムが提案、分析、評価されてきた。システムによる推薦結果は、次にユーザが閲覧するアイテムとなり、結果としてどのアイテムが次に評価されるかに影響を及ぼす。したがって、推薦システムとユーザは相互作用を及ぼしあう関係にあるといえる。しかし、先行研究による推薦アルゴリズムの評価・分析では、一時的な推薦精度のみに着目し、このような相互作用は考慮されてこなかった。そこで本研究では、ユーザをエージェントとしてモデル化することで、推薦システムとユーザの相互作用を考慮した、新たな評価・分析モデルを提案する。さらに、本論文では、基本的な推薦アルゴリズムである「ユーザ間協調フィルタリング」と、実システムとしてすでに実用化されている「アイテム間協調フィルタリング」を例として取り上げ、その特性について提案モデルを用いて分析した。その結果、既存の評価・分析モデルの枠を超え、より有効な推薦への指針を検証可能であることが示された。

Performance Evaluation of Recommendation Algorithms Based on Rating-recommendation Interaction

AKIHIRO YAMASHITA,^{†1,†2} HIDENORI KAWAMURA,^{†1}
HIROYUKI IZUKA^{†3} and AZUMA OHUCHI^{†1}

The problem of information overload spreading across the Internet has been causing serious inefficiency in browsing and searching for information. As a way to overcome the problem, the recommender systems are recently used in many E-commerce sites. Many algorithms have been proposed to improve the accuracy of recommendation based on user ratings. The relation between recommender systems and users is rather interactive in the sense that recommendations decides which items are recommended to users and the results of ratings by users will affect the next recommendations. However, conventional studies

have not considered the interactive aspects so much. Therefore, our aim of this paper is to propose a new evaluation model using multiagent modeling where the recommender system and agents (as users) interacts with each other. The properties of typical recommendation algorithms such as user-based and item-based collaborative filtering will be analyzed with our proposed model. Our results also suggest the possibilities to propose a novel and effective recommendation algorithm.

1. はじめに

ウェブ上における情報量の拡大にともなって、素早く必要な情報を探し出すことが難しくなっている。そこで、Amazon.com^{*1}、MovieFinder.com^{*2}、Ebay^{*3}など、多くのオンラインストアでは推薦システム¹⁾を導入している²⁾。推薦システムは、ユーザの興味に適應した情報を積極的に提示してユーザの利便性向上を図るとともに、オンラインストアの運営側にとっては宣伝効果による収益拡大が見込めるため、双方にメリットのあるシステムである。特に Amazon.com は推薦システムを効果的に利用した典型的な例である³⁾。

推薦システムは、1992年に協調フィルタリング (Collaborative Filtering, 以下 CF) の基本概念⁴⁾が初めて示されて以降、k-nearest neighbor 法を初めて導入した GroupLens プロジェクト⁵⁾など、多くの改良、分析、評価がなされてきた⁶⁾⁻⁹⁾。CFに基づく推薦システムは、ユーザのアイテムに対する評価を基に、推薦するアイテムを決定する。推薦を受けたユーザは、その推薦結果に影響を受けながら、次の購買・評価を行う。つまり、推薦システムは、ユーザにアイテムを推薦し、その結果得られるユーザの評価をフィードバックとして、その有効性を高めていく。

CFに基づく様々な推薦アルゴリズムについて、実用に即した評価を行う場合、ユーザから収集した評価サンプルは、それ以前の推薦に影響を受けていることを考慮する必要がある。しかし、従来研究において、ユーザの評価は、それが形成される過程とは独立であると

†1 北海道大学大学院情報科学研究科

Graduate School of Information Science and Technology, Hokkaido University

†2 日本学術振興会

JSPS, Japan Society for the Promotion of Science

†3 大阪大学大学院工学研究科

Graduate School of Engineering, Osaka University

*1 <http://www.amazon.com/>

*2 <http://www.moviefinder.com/>

*3 <http://www.half.ebay.com/>

仮定して収集し、システムの有効性を検証することが多かった。この方法は、ユーザとシステムの相互作用を考慮しない、静的な評価手法ととらえられる。そこで、より実用段階を考慮して推薦アルゴリズムを評価する場合、ユーザによる購買・評価と、システムによる推薦の間で繰り返される相互作用を考慮した、動的な評価手法が必要である。

本論文では、ユーザの購買・評価行動を、エージェントとしてモデル化し、推薦システムとユーザ間の相互作用を考慮した、動的な評価手法を提案する。さらに、本論文では、CFに基づく基本的なアルゴリズムであるユーザ間CF (User-based CF) と、多くの利用実績を持つアイテム間CF (Item-based CF) について、提案手法を用いて評価・分析を行った。その結果、相互作用を考慮しない従来の静的な評価手法の枠を超えて、より有効にCFを利用するための指針について、検証可能であることを示す。

2. 関連研究

推薦システムの目的は、被推薦者が高い効用を得られるアイテムを、そのユーザの過去の評価に基づいて推測し、提示することである。推薦アルゴリズムは、アイテム自身の特徴を利用する「内容に基づくフィルタリング」と、アイテムの内容は利用せずユーザの評価のみで推薦を行うCFに大別される。さらにCFは、評価データベース全体を用いる、「メモリベース方式」と、あらかじめ評価データベースを確率モデルなどでモデル化し、それを用いて推薦を行う「モデルベース方式」に分類できる¹⁰⁾。GroupLensプロジェクトで用いられたユーザ間CF⁵⁾や、Amazon.comで用いられるアイテム間CF³⁾は、メモリベース方式の典型例である。また、モデルベース方式の例として、クラスタリングを用いた手法¹¹⁾などがあげられ、さらに、両手法を融合する試みもなされている¹²⁾。

メモリベース方式では、ユーザから得た評価が瞬時に次の推薦結果に影響を及ぼす。一方、モデルベース手法では、一般化したモデルを作成して推薦を行うため、モデルを更新するまでは、ユーザの評価は推薦結果に影響しない。本論文では、推薦システムとユーザによる相互作用の分析が目的であり、互いの影響を強く受けるメモリベースのCFを扱う。

推薦システムの客観的な評価は困難であるといわれている^{6),13)}。これまで多くの評価指標が提案されているが、一般的な指標は予測精度である。多くの研究では、実際のWebサイトなどで収集した実データをサンプルとして精度を測定している^{14),15)}。実データによる推薦精度の評価では、まず実データを訓練データと検査データに分割し、訓練データで推薦アルゴリズムを学習させた後、検査データの予測精度を測定する。しかしこの手法では、ある一時に得られる評価データを用いて検証を行うため、ユーザが推薦結果を閲覧し、それ

が次に評価するアイテムに影響を与えるという相互作用は考慮されない。したがって、相互作用に基づく分析を行うためには、推薦と評価を何度も繰り返す必要がある。しかし、そのような実験には多くのコストと時間が必要なだけでなく、複数の推薦アルゴリズムを、同条件下で検証することは困難である。

そこで本研究では、ユーザと推薦システムの相互作用に基づく分析のための、マルチエージェントを用いた新たな評価手法を提案する。また、その提案手法を用いて、既存アルゴリズムの特徴分析を行った。

本研究の目的は、特定のユーザやアイテムを仮定しない単純なモデルを使用し、推薦アルゴリズムが持つ基本的特性を分析することである。岩田ら¹⁶⁾は、実際にユーザの購買データを用いて推薦結果が購買行動に与える影響のモデル化を試みている。このような実データに基づいたモデルと組み合わせることでより現実に即した検証も可能である。

3. 推薦と格付けの相互作用に基づく性能評価のフレームワーク

3.1 構成要素

本研究で構築したモデルは、ユーザをモデル化した購入・格付けエージェント(以下エージェント)とアイテム、推薦システムの3つの要素から構成される。

推薦システムは N_{user} 人のエージェント集合 $U = \{i | i = 1, 2, \dots, N_{\text{user}}\}$ の各エージェントに対して、 N_{item} 個のアイテム集合 $C = \{j | j = 1, 2, \dots, N_{\text{item}}\}$ の中から適切なアイテムを推薦する。各エージェントは、個人の嗜好を持ち、本モデルではその嗜好を嗜好ベクトル $\mathbf{p}_i = (p_{i1}, p_{i2}, \dots, p_{id})$ として表現する。同様に、各アイテムはそれぞれの特徴を持ち、本モデルではその特徴を特徴ベクトル $\mathbf{v}_j = (v_{j1}, v_{j2}, \dots, v_{jd})$ として表現する。ただし、 \mathbf{p}_i と \mathbf{v}_j はともに d 次元のベクトルであり、各要素の範囲は $[-1, 1]$ とする。

エージェント i は、アイテム j を推薦されたとき、それがどの程度自分の嗜好に一致しているかを表す効用 $s_{i,j}$ を計算する。効用は、推薦に対するエージェントの満足度を数値化したものであり、高い効用を与える推薦が効果的な推薦である。その後エージェント i は、効用 $s_{i,j}$ を基に推薦されたアイテムの評価を行い、推薦システムに入力する。本論文では、この評価を、エージェント i によるアイテム j の格付け $r_{i,j}$ と呼ぶ。格付けは、エージェント-アイテムの行列として表現できる。本論文ではこの行列を格付け行列と呼ぶ。

3.2 効用

効用 $s_{i,j}$ は、 \mathbf{p}_i と \mathbf{v}_j から計算される。

$$s_{i,j} = f_{\text{utility}}(\mathbf{p}_i, \mathbf{v}_j).$$

\mathbf{p}_i は、エージェント i の嗜好であり、好みのアイテムの特徴を表現する。したがって、本モデルでは、 \mathbf{p}_i と \mathbf{v}_j の距離が近いほど、効用 $s_{i,j}$ が大きくなるよう設定する。つまり、関数 f_{utility} は、 \mathbf{p}_i と \mathbf{v}_j の距離に対して単調減少関数でなければならない。

$s_{i,j}$ の具体的な値は、 f_{utility} の設計次第で変化するが、単調減少関数であれば、その大小関係は変化しない。つまり、推薦アルゴリズムの優劣の評価では、任意の単調減少関数を f_{utility} として設定可能である。

単純な単調減少関数の 1 つに、反比例があげられる。しかし、 f_{utility} を反比例とする場合、距離が 0 のときに $s_{i,j}$ が無限大となり、好ましくない。また別な関数として、線形で減少する関数が考えられる。つまり、 \mathbf{p}_i と \mathbf{v}_j の距離が 0 のときに $s_{i,j} = 1$ となり、距離が最大のときに $s_{i,j} = 0$ となる直線の関数である。3.1 節で述べたように、本モデルでは \mathbf{p}_i および \mathbf{v}_j が存在する空間の範囲は、各軸独立に $[-1, 1]$ である。そのため、2 つのベクトルが空間上で対角の位置に存在する場合に距離が最大となるが、その値は次元数 d に依存する。つまり、線形関数は x 軸との交点を持つために、次元数 d に応じて傾きを変える必要がある。そこで、本モデルでは、負の傾きを持つ次の指数関数を用いた。

$$f_{\text{utility}}(\mathbf{p}_i, \mathbf{v}_j) = \exp(-\alpha \|\mathbf{p}_i - \mathbf{v}_j\|) \quad (1)$$

この式は、効用の値域が $0 \leq s_{i,j} < 1$ となり、また x 軸との交点もないため、次元数 d に応じて f_{utility} を変更する必要はない。また α は、関数の傾きを調整するパラメータであり、本論文では、 $s_{i,j}$ の分散が大きくなるよう調節した結果、 $\alpha = 0.5$ とした。

3.3 格付け

本論文では商用 Web サイトなどで一般的な 5 段階による格付けを用いた。格付け $r_{i,j}$ は、エージェント i が効用 $s_{i,j}$ を基に計算する。実際には、高い格付けばかりをするユーザーが存在するなど、格付けが効用を正確に反映するとは限らない。しかし、本論文では単純化のため、5 段階の格付けを偏りなく行う関数 $f_{\text{rating}}(s_{i,j})$ を設定し、全エージェントがその関数に基づいて格付けを行うものとする。

f_{rating} は効用 $s_{i,j}$ から 5 段階の格付け $r_{i,j}$ を求めるための 4 つの閾値を持つ関数である。各閾値は、 \mathbf{p}_i および \mathbf{v}_j を一様分布で与えた場合に、5 段階の格付けが同じ頻度で行われるようあらかじめ予備実験により決定する。

\mathbf{p}_i および \mathbf{v}_j の次元数 d が変化すれば、ベクトル間の距離から計算される効用 $s_{i,j}$ の値も変化するため、閾値はベクトルの次元数 d に応じて個別に設定する。

式 (2) に $d = 5$ の場合について求めた関数を示す。

$$r_{i,j} = f_{\text{rating}}(s_{i,j}) = \begin{cases} 1 & (s_{i,j} \leq 0.3358) \\ 2 & (0.3358 < s_{i,j} \leq 0.3878) \\ 3 & (0.3878 < s_{i,j} \leq 0.4417) \\ 4 & (0.4417 < s_{i,j} \leq 0.5160) \\ 5 & (0.5160 < s_{i,j}) \end{cases}, \quad (2)$$

ただし、閾値は \mathbf{p}_i および \mathbf{v}_j が一様分布の場合を基準に設定するため、 \mathbf{p}_i の分布が変われば 5 段階の各格付けの出現頻度は変化する。

ここで、ベクトルの次元数 d が推薦結果に与える影響について考察する。前節の式 (1) で示したとおり、効用 $s_{i,j}$ は \mathbf{p}_i と \mathbf{v}_j の距離から計算されるため、ベクトルの次元数 d に依存して値が変化する。しかし、格付け $r_{i,j}$ は、5 段階の各格付けが同じ頻度で行われるため、 \mathbf{p}_i と \mathbf{v}_j 間の距離が近いアイテムから順番に高い格付けとなり、次元数 d には依存しない。また、推薦アルゴリズムは、格付け行列から推薦するアイテムを決定するため、 d には依存しない。よって、次元数 d は、推薦システムによる推薦結果には影響を与えず、エージェントが得る効用 $s_{i,j}$ におよび影響する。

次元数が増加すれば空間が広がり、 \mathbf{p}_i と \mathbf{v}_j 間の距離が全体的に増加する。その結果、 $s_{i,j}$ は小さい値になる。しかし、次元数 d を固定して推薦アルゴリズムを比較すれば、その順序関係は変わらない。したがって、本論文では $d = 5$ として実験を行った結果を示すが、この次元数を増減させても、結果は変化しない。

3.4 全体の流れ

図 1 にシミュレーションの流れを示す。シミュレーションは初期設定の後、以下に示す処理 A から処理 D を、1 シミュレーションステップとして繰り返す。

初期設定：パラメータとして、 N_{user} 、 N_{item} 、 \mathbf{p}_i の分布、 \mathbf{v}_i の分布、推薦アルゴリズムを決定する。

処理 A：推薦システムは、格付け行列に基づき、推薦アルゴリズムに従って格付け行列の未入力部分を推定する。初期状態では、格付け行列には何も入力されていないので、推薦アルゴリズムによらずランダムな推定になる。

処理 B：各エージェントに、格付けの推定値が最も高いアイテムを 1 つ推薦する。推定値が最も高いアイテムが複数存在する場合は、その中からランダムに推薦する。

処理 C：各エージェントは、推薦された商品の格付けを計算する。このとき、全エージェ

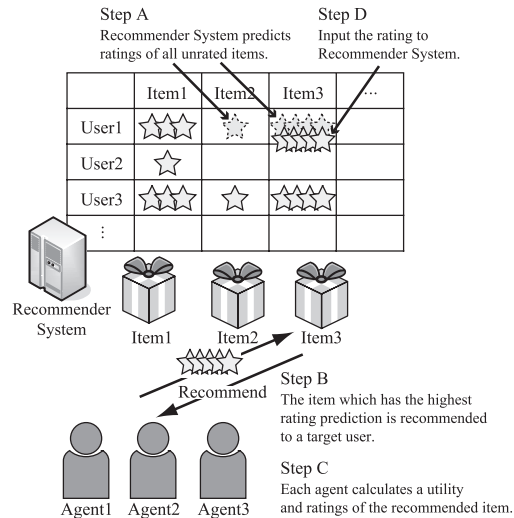


図1 シミュレーションの流れ
Fig. 1 Simulation process.

ントの効用の平均値 (utility average) はそのステップにおける全体の効用ととらえられる。したがって、ステップごとに効用の平均値を測定すれば、全体の効用の推移が観察できる。

処理 D: 各エージェントは、処理 C で求めた格付けを推薦システムに入力する。したがって、ステップ t において、各エージェントは t 回目の格付けを行うことになる。その後、処理 A に戻り繰り返す。

4. 実験設定

4.1 嗜好ベクトル p_i の分布

p_i を様々な分布に従って生成することで、全体の嗜好のばらつき具合 (トレンド) が変化する。

実世界の嗜好の分布は、様々な嗜好を持つユーザが偏りなく存在する状況 (トレンドがない状況) から、似た嗜好を持つ多数のユーザと、ごく少数のニッチからなるロングテール構造 (トレンドが強い状況) まで様々な状況が想定される。ユーザの嗜好の分布は、推薦システムが対象とするコンテンツや、ユーザ層にも依存し、一概に決めることはできない。

そこで本論文では、提案する評価手法の有効性を検証することを目的とし、次に示す 3 つの典型的な分布を設定してその違いを比較した。

一様分布 (Uniform distribution): p_i が一様分布に従う場合、 p_i の各要素独立に範囲 $[-1, 1]$ の一様乱数で与えた。この分布は、本モデルの中で最もトレンドが弱い分布となる。

多変量正規分布 (Multivariate normal distribution): p_i が多変量正規分布に従う場合、エージェント全体の嗜好に偏りが生じる。多変量正規分布に基づく乱数 x は、次元数 d のとき、各要素間で独立な d 変量正規乱数 $z = (z_1, z_2, \dots, z_d)$ と、分散共分散行列 Σ をコレスキー分解 $\Sigma = LL^T$ した行列 L の積で求まる。つまり、平均ベクトルを μ とすれば、

$$x = Lz + \mu \sim N(\mu, \Sigma),$$

で計算できる。

本論文では、提案手法の有効性検証を目的とするため、極力シンプルな分布を設定した。つまり、 $\mu = 0$ とし、 Σ は、式 (3) に示すように、対角要素がすべて v で、それ以外すべて 0 という単純な行列 Σ_v を用いた。したがって、 v が小さいほど強いトレンドを持つ分布である。

$$\Sigma_v = \begin{cases} \sigma_{ij} = v & \text{where } i = j \\ \sigma_{ij} = 0 & \text{otherwise} \end{cases} \quad (3)$$

2 つの頂点を持つ分布 (Two-peak distribution): 嗜好の偏りは、1 つに集中するとは限らず、似たグループが複数形成される状況も想定される。そこで、2 つの頂点を持つ分布を生成し検証を行った。2 つの頂点を持つ分布は、分散共分散行列 Σ が等しく平均ベクトル μ が異なる 2 つの多変量正規分布を作成し、等確率でどちらかの分布に従うようなベクトル乱数により発生させた。本論文では、2 つの平均ベクトルとして

$$\mu_1 = (0.5, 0.5, 0.5, 0.5, 0.5)^T,$$

$$\mu_2 = (-0.5, -0.5, -0.5, -0.5, -0.5)^T,$$

および、分散共分散行列として $\Sigma_{0.04}$ を使用した。

4.2 特徴ベクトル v_j の分布

本論文では、様々な特徴を持つアイテムが偏りなく存在する状況を設定し、各推薦アルゴリズムを解析する。したがって、推薦対象となるアイテムの特徴ベクトルは、一様分布に基づく乱数で生成した。特徴ベクトルを一様分布にすることで、好まれるアイテムとそうでないアイテムの偏りは p_i の分布にのみ依存する。

4.3 検証する推薦アルゴリズム

推薦システムは、格付け行列に基づいて推薦を行う。具体的には、格付け行列の中でユーザの入力により格付けが既知である部分から既知ではない値の推定を行い、推定値が最大のアイテムを推薦する。したがって、推薦アルゴリズムの本質は、まだ格付けされていないアイテムに対する格付けの推定である。本論文では提案手法を使用し、次の4つの推薦アルゴリズムを対象として検証実験を行った。

ランダム推薦 (Random Recommendation)

ランダム推薦では、格付けの推定値として一様乱数を使用する。したがって、すでに入力された格付けには関係なく、まだ格付けされていないアイテムの中からランダムに推薦する。この推薦手法の結果は、他の推薦手法に対する基準として使用する。

人気推薦 (Popular Recommendation)

人気推薦は、格付けのエージェント間の平均値を推定値とし、それが最も高いアイテムを人気アイテムとして推薦する。人気推薦は、各エージェントの嗜好には関係しないため、最も人気のアイテムをまだ格付けしていないすべてのエージェントに推薦する。ただし、格付けの平均値が等しくなった場合は、その中からランダムに決定する。

ユーザ間協調フィルタリング (User-based CF)

ユーザ間 CF は、Resnick ら⁵⁾ によってその基本アルゴリズムが提案された。まず、被推薦者となるエージェント $a \in U$ に対して、嗜好が類似しているエージェント $neighbors$ を求める。その後、 $neighbors$ の過去の格付けを利用して、まだ被推薦者が格付けしていないアイテムに対する格付けの推定を行い、最も推定値が高いアイテムを推薦する。

$neighbors$ の決定には、エージェント a とその他の各エージェント間の類似度 (similarity) を計算し、類似度の高いエージェントから n 人を選択する方法 (Nearest neighbor 法) を用いる。類似度の計算には、2 人のエージェントがともに格付けしたアイテムの格付けのみが使用される。類似度の計算手法としては、ピアソン相関係数⁵⁾ や、コサイン距離¹⁰⁾ を用いる場合が多く、一般的にはピアソン相関係数の方が精度が良いことが知られている¹⁰⁾ ため、本論文でもピアソン相関係数を用いる。エージェント a とエージェント i の類似度 $sim(a, i)$ は、式 (4) で計算される。ただし、 C_i は、エージェント i が格付けしたアイテム集合を示し、 $C_{a,i}$ は、エージェント a とエージェント i がともに格付けしたアイテム集合を示す。また $|C_i|$ は C_i に含まれるアイテムの個数を示す。

$$sim(a, i) = \frac{\sum_{j \in C_{a,i}} (r_{a,j} - \bar{r}_a)(r_{i,j} - \bar{r}_i)}{\sqrt{\sum_{j \in C_{a,i}} (r_{a,j} - \bar{r}_a)^2} \sqrt{\sum_{j \in C_{a,i}} (r_{i,j} - \bar{r}_i)^2}},$$

$$\text{where } \bar{r}_i = \frac{1}{|C_i|} \sum_{a \in C_i} r_{i,a},$$

$neighbors$ に含まれるいずれかのエージェントがすでに格付けをしており、被推薦者 a がまだ格付けしていないアイテム j について、その格付けの予測値 $\hat{r}_{a,j}$ は式 (4) で計算される。

$$\hat{r}_{a,j} = \bar{r}_a + \frac{\sum_{i \in U_j} sim(a, i)(r_{i,j} - \bar{r}_i)}{\sum_{i \in U_j} |sim(a, i)|}, \quad (4)$$

ただし、 \bar{r}_i はアイテム集合 C_i への格付けの平均値、 U_j はコンテンツ j を格付けしたエージェント集合である。したがって、エージェント a には、そのエージェントがまだ格付けしていないアイテムの中で、格付けの推定値 $\hat{r}_{a,j}$ が最大となるアイテムを推薦する。

アイテム間協調フィルタリング (Item-based CF)

アイテム間 CF は、ユーザによる格付けを利用してアイテム間の類似性を計算し推薦する手法である¹⁷⁾。まず、類似アイテムを求めるために、任意の2つのアイテム間で類似度を定義する。一般的に、エージェント間の類似度と同様にピアソン相関係数やコサイン距離が用いられるが、本研究ではピアソン相関係数を用いた。アイテム c とアイテム j の類似度 $sim(c, j)$ は、式 (5) で計算される。

$$sim(c, j) = \frac{\sum_{i \in U_{c,j}} (r_{i,c} - \bar{r}_c)(r_{i,j} - \bar{r}_j)}{\sqrt{\sum_{i \in U_{c,j}} (r_{i,c} - \bar{r}_c)^2} \sqrt{\sum_{i \in U_{c,j}} (r_{i,j} - \bar{r}_j)^2}},$$

$$\text{where } \bar{r}_j = \frac{1}{|U_j|} \sum_{i \in U_j} r_{i,j}, \quad (5)$$

ただし、 $U_{c,j}$ は、アイテム c とアイテム j を両方とも格付けしたエージェントの集合を表

し, \bar{r}_j は, その集合におけるアイテム j の格付けの平均値である. $sim(c, j)$ もエージェント間の類似度同様, 相関係数であることから, その範囲は $[-1, 1]$ である.

被推薦者 i がまだ格付けしていないアイテム c について, アイテム間の類似度を利用した格付けの予測値 $\hat{r}_{i,c}$ は式 (6) で計算される.

$$\hat{r}_{i,c} = \frac{\sum_{j \in C_i} sim(c, j)(r_{i,j})}{\sum_{j \in C_i} |sim(c, j)|}. \quad (6)$$

エージェント i が格付けしていないすべてのアイテムについて格付けの推定値 $\hat{r}_{i,c}$ を計算し, 最も値が大きくなったアイテムを推薦する.

5. 結果と考察

本章では, 提案手法を用いて各推薦アルゴリズムを分析した結果と考察を述べる. 5.1 節では, 格付け行列に何も入力されていない推薦システムを想定し, エージェントとの相互作用を通じて徐々に格付けが入力され, 推薦精度が向上していく過程を分析する. また 5.3 節では, ある程度格付け行列に格付けが入力された推薦システムを想定し, そこを訪れた新規エージェントに対する推薦効果を分析する. 最後に 5.5 節では, ここまでの検証実験で明らかになったユーザ間 CF とアイテム間 CF の特徴をふまえ, 両者の利点を生かした新たな推薦手法について述べる.

5.1 検証実験 1

まず, 格付け行列の要素がまったく存在しない状態から推薦を行う場合を想定し, 実験を行った. 図 2 は, $N_{user} = 500$, $N_{item} = 500$, \mathbf{p}_i の分布を一様分布とし, 4.3 節で述べた各アルゴリズムで推薦した結果である. 縦軸は効用の平均値 (utility average), 横軸はステップ数 t で, 10 回の試行の平均値を示す. ランダム推薦はつねにほぼ一定の値であり, 他の推薦アルゴリズムの基準となる. たとえば, ユーザ間 CF に着目すると, $t = 0$ における効用の平均値は, ランダム推薦と同程度である. これは, 格付け行列に値がまったく存在しない状態では推定ができず, ランダム推薦と等しくなるからである. その後, $t = 30$ 付近で上昇し, $t = 40$ 付近でピークを迎えた後, 高い効用を得られるアイテムがなくなり徐々に低下している.

人気推薦とアイテム間 CF も, $t = 0$ でランダム推薦と等しい値である. しかし, 効用の平均値が増加するステップ数やその最大値は, アルゴリズムによって差が生じており, 推

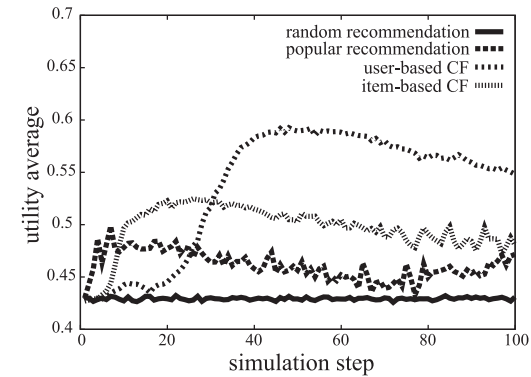


図 2 推薦アルゴリズムによる効用の平均値の推移 (\mathbf{p}_i は一様分布, $N_{user} = 500$, $N_{item} = 500$)
Fig. 2 Changes in utility average obtained using each recommendation algorithm (\mathbf{p}_i is generated based on uniform distribution, $N_{user} = 500$, $N_{item} = 500$).

薦アルゴリズムによって特徴の違いが現れている. このような特徴の差は, N_{user} , N_{item} , および 4.1 節で述べた \mathbf{p}_i の分布を変化させた場合についても観察された. 以下では様々な実験設定における結果を述べるとともに, 5.2 節では得られた結果についてより詳細に考察する.

実験結果を説明するために, 本実験で使用した 2 つの指標について述べる.

まず 1 つ目は, 効用の平均値が立ち上がるまでのステップ数である. 以下「立ち上がりステップ」と呼ぶ. 格付け数が少ない場合, 推薦システムは有効な推薦が実現できない. 立ち上がりステップは, システムが推薦効果を発揮するまでに必要な格付けの数を示す指標である. 本論文では, あるステップの効用の平均値が, 1 ステップ目と比較して 0.05 高くなった時点を, 立ち上がりステップと定義する. ただし, 効用の平均値がそこまで達しない場合は, 立ち上がりステップは計算されない.

2 つ目は「効用のピーク値」である. これは, 推薦効果が立ち上がった後, 効用の平均値がどの程度の値まで達したかを示す指標である. 本論文においては, 立ち上がりステップ $t = r$ から $t = r + 30$ までの全エージェントの効用の平均値を効用のピーク値と定義する. したがって, 立ち上がりステップが計算されない場合は, 効用のピーク値も計算されない.

この指標を用いて, 4.3 節で述べた各アルゴリズムの特徴分析を行った結果を図 3 に示す. 図 3(a) から, ユーザ間 CF およびアイテム間 CF 両方において, N_{user} の増加とともに効用のピーク値が上昇し, また図 3(b) から, N_{user} の増加とともに, 立ち上がりステップが

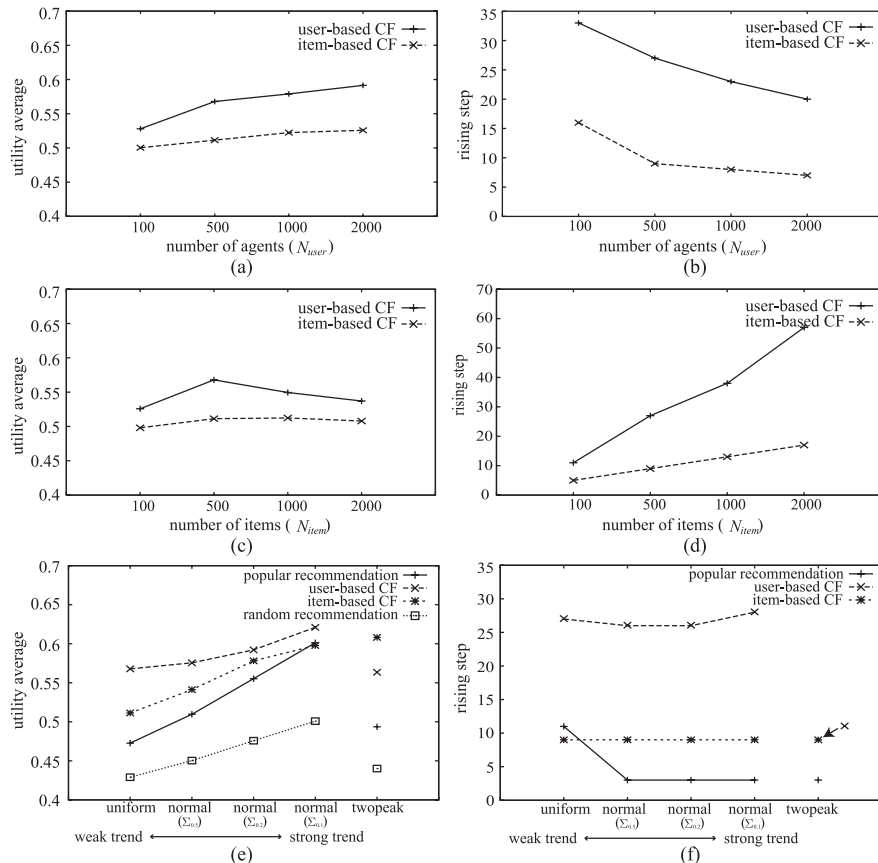


図 3 各推薦アルゴリズムにおける、のピーク値および立ち上がりステップ。ただし、(a), (b) は、 $N_{item} = 500$, p_i を一様分布として固定し、 $N_{user} = \{100, 500, 1000, 2000\}$ での実験結果。(c), (d) は、 $N_{user} = 500$, p_i を一様分布として固定し、 $N_{item} = \{100, 500, 1000, 2000\}$ での実験結果。(e), (f) は、 $N_{user} = 500$, $N_{item} = 500$ で固定し p_i の分布を変化させたときの実験結果

Fig. 3 Peak value of utility average and rising step obtained using each recommendation algorithm. (a) and (b) show the results at $N_{item} = 500$ and p_i is generated based on uniform distribution. (c) and (d) show the results at $N_{user} = 500$ and p_i is generated based on uniform distribution. (e) and (f) show the results at $N_{user} = 500$, $N_{item} = 500$ and p_i is generated based on each distribution.

短縮されることが分かる。同様に N_{item} を変化させた結果が図 3 (c), (d) である。 N_{item} が増加することで、特にユーザ間 CF で、立ち上がりステップが遅くなる。図 3 (e), (f) は、4.1 節で述べた p_i の分布、つまりトレンドの強さと効用のピーク値および立ち上がりステップの関係を示すグラフである。トレンドが強くなるに従って効用のピーク値も向上しているが、基準となるランダム推薦も向上していることに注意が必要である。今回の実験の設定では、トレンドが強くなるにつれて、嗜好ベクトルは空間における中心に集まりやすくなるため、各アイテムとの距離の平均値が小さくなり全体の効用が増加する。この点に注意すると、分布が変化してもユーザ間 CF とアイテム間 CF において、効用のピーク値や立ち上がりステップに大きな違いはない。

次に、図 3 からユーザ間 CF とアイテム間 CF を比較する。いずれのグラフを見ても、効用のピーク値はユーザ間 CF の方が高く、立ち上がりステップはアイテム間 CF の方が早い傾向がある。この差は、2 つのアルゴリズムの特性が異なることを示唆している。次節で、この差についての詳細を述べる。

5.2 考察 1

今回のシミュレーションでは、各エージェントはステップごとに 1 つだけアイテムを推薦され、必ず格付けを行う。したがって、推薦アルゴリズムの特徴は、格付け行列の埋まり方に反映されていると考えられる。そこで、ユーザ間 CF とアイテム間 CF の特徴の違いを分析するため、各ステップにおいて、各アイテムが何人のエージェントに推薦・格付けされているかを測定した。 $N_{user} = 500$, $N_{item} = 500$, p_i を一様分布で与えた場合の結果を図 4 に示す。図 4 の横軸はステップ数 t であり、縦軸は、無作為に選んだ 100 個のアイテムをランダムに並べている。濃淡は各アイテムが、今まで累計で何回エージェントに推薦・格付けされたかを表し、色が薄い場所ほど数多く推薦・格付けされたことを示す。

図 4 (左) は、あるステップにおけるアイテム間の濃淡の差が小さい。つまり、ユーザ間 CF では、アイテムごとの推薦・格付回数について、その差が小さいことを示している。一方、図 4 (右) から、アイテム間 CF では、数個のアイテムが早い段階から白く表示されており、数多くのエージェントに推薦されている。同時に、ほとんど推薦されないアイテムも多数存在していることが読み取れる。また、濃い部分(ほとんど推薦されていない状態)から、急激に薄くなる(多数のエージェントに推薦されたことを示す)アイテムも複数存在している。つまり、アイテム間 CF はその時点で人気と判断したアイテムを積極的に多数のエージェントに推薦していると推測できる。対照的に、ユーザ間 CF の方がより多くのアイテムからエージェントの嗜好に合わせたアイテムを推薦しているといえる。

54 推薦と格付けの相互作用に基づく推薦アルゴリズムの性能評価

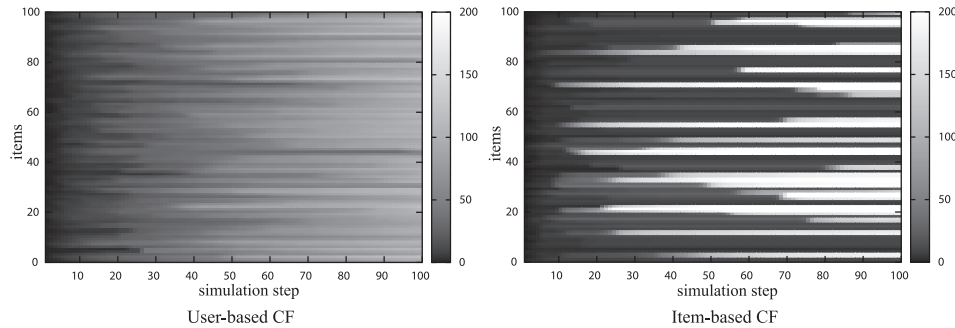


図 4 ユーザ間 CF (左) およびアイテム間 CF (右) において $N_{\text{user}} = 500$, $N_{\text{item}} = 500$, \mathbf{p}_i を一様分布で与えたときの各アイテムに対する推薦・格付け数の変化 (ただし, 図中には無作為に選んだ 100 個のアイテムを示す)

Fig. 4 Changes in the number of ratings to each item at $N_{\text{user}} = 500$, $N_{\text{item}} = 500$, \mathbf{p}_i is generated based on uniform distribution obtained using user-based CF (left) and item-based CF (right). (This figure shows 100 items chosen at random).

このような現象が起きる原因として以下のことが推測される。今回のシミュレーションは、ステップごとに各エージェントが 1 つのアイテムを格付けする。つまり、すべてのエージェントが必ず同じ回数の格付けを行うため、ユーザ間 CF ではエージェント間で格差なく類似度が計算が可能となり、多くのユーザ間の類似度を基に推薦候補を決定できる。ただし、類似度を計算するためにはある程度のステップ数が必要となり、これが立ち上がりステップの遅さの原因とも考えられる。一方でアイテム間 CF では、各アイテムは推薦された回数分の格付けを受けるため、誰にも推薦されないアイテムはまったく格付けが付かないなど、アイテム間で格付けの数に差が生じる。これにより、早い段階で格付けされたアイテムは、他のアイテムとの類似度が計算され、その後も推薦候補となることで、さらに多くの格付けを集めやすくなる。一方で、早い段階にまったく推薦されなかったアイテムは、その後も類似度が計算されず、結果として仮に人気の高いアイテムだったとしても長い間推薦対象から除外されてしまう。これが、立ち上がりステップを早め、効用のピーク値を抑える原因と考えられる。

ユーザ間 CF では効用のピーク値が高くなり、アイテム間 CF では立ち上がりステップ数が早くなる、という傾向は、図 3 (e), (f) に示すように、嗜好ベクトル \mathbf{p}_i を、4.1 節で述べたいずれの分布で生成しても確認された。

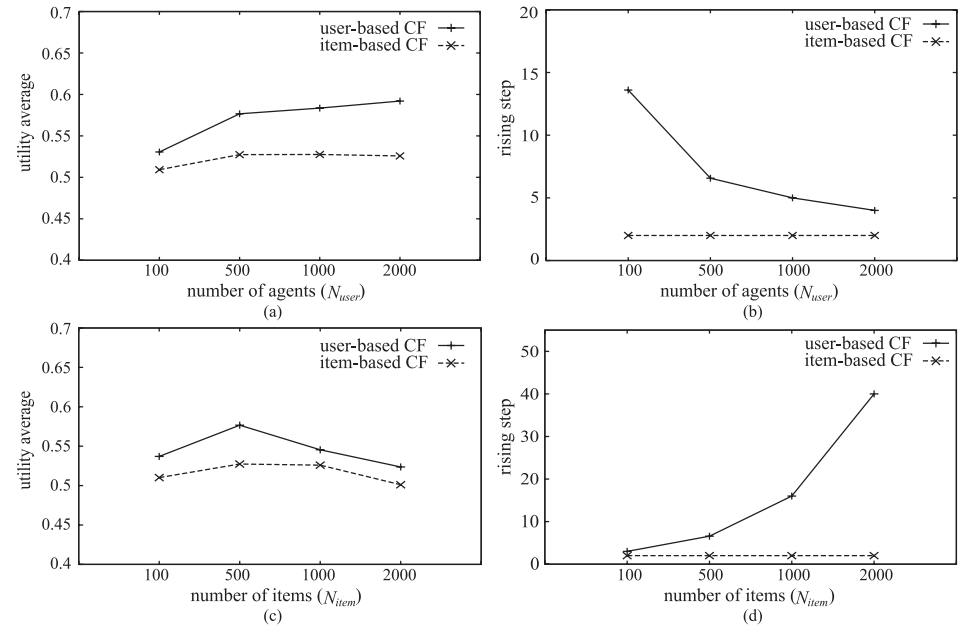


図 5 推薦システムにあらかじめ格付けが入力された状態における新規エージェントに対しての効用のピーク値および立ち上がりステップ。ただし, (a), (b) は $N_{\text{item}} = 500$, \mathbf{p}_i を一様分布として固定し, $N_{\text{user}} = \{100, 500, 1000, 2000\}$ での実験結果。 (c), (d) は $N_{\text{user}} = 500$ と \mathbf{p}_i を一様分布として固定し, $N_{\text{item}} = \{100, 500, 1000, 2000\}$ での実験結果

Fig. 5 Peak value of utility average and rising step of new agent when ratings are set in our recommender system beforehand. (a) and (b) show the results at $N_{\text{item}} = 500$ and \mathbf{p}_i is generated based on uniform distribution. (c) and (d) show the results at $N_{\text{user}} = 500$ and \mathbf{p}_i is generated based on uniform distribution.

5.3 検証実験 2

次に、すでに運用中の推薦システムを想定し、一定の格付けが入力されている状況で、新規エージェントに推薦を行った場合について検証を行った。

まず、あらかじめ半数のエージェントに対して 100 個のアイテムをランダム推薦する。その状況で、残りの半分のまだまったく格付けを行っていないエージェントのみを使用し、ユーザ間 CF とアイテム間 CF のそれぞれのアルゴリズムで推薦を行った。 $N_{\text{item}} = 500$, \mathbf{p}_i は一様分布とし, $N_{\text{user}} = \{100, 500, 1000, 2000\}$ のそれぞれで実験を行った。図 5 は、CF に

て推薦を行ったエージェントに関する効用のピーク値と立ち上がりステップの平均値である。

5.4 考察 2

検証実験 1 と同様、 N_{user} の増加とともに効用のピーク値は上昇し、立ち上がりステップは短くなっている。図 5 (b), (d) によると、アイテム間 CF は 2 ステップ目の推薦から効果を得られるのに対し、ユーザ間 CF は N_{user} が増加しても推薦効果を得るのに 5 ステップ程度を必要としている。この原因として類似度の計算が影響していると考えられる。すでに格付けが入力された状況では、アイテム間類似度は計算可能である。しかし新規エージェントに対してのエージェント間類似度は計算不可能である。今回の実験において類似度はピアソン相関係数を用いているため、エージェント間類似度を計算するためには最低でも 2 つのアイテムを共通して格付けしなければならない。その差が、推薦効果の立ち上がりステップに影響していると考えられる。

5.5 検証実験 3

検証実験 1 よりユーザ間 CF のほうが効用のピーク値が高く、アイテム間 CF の方が立ち上がりステップが早いことが明らかになった。また検証実験 2 より、すでに一定の格付けが入力されている推薦システムにおいて、新規エージェントに推薦を行うことを想定しても、同様の結果となった。図 2 によると、まったく格付けがない状態から推薦を行った場合、アイテム間 CF とユーザ間 CF で効用の平均値が逆転するのは $t = 30$ 付近であることが読み取れる。同様に、検証実験 2 の設定において、ユーザ間 CF がアイテム間 CF の効用の平均値を上回るのは $t = 10$ であることが分かった。

そこで、2 つのアルゴリズムの利点を生かした新たな手法として、アイテム間 CF から途中でユーザ間 CF に切り替えることを考える。図 6 は、まったく格付けがない状態から、すべてのエージェントに対してアイテム間 CF で推薦を開始し、 $t = 30$ でユーザ間 CF に切り替えた場合の結果である。また図 7 は、5.3 節と同様あらかじめ半数のエージェントに 20% のアイテムをランダムに格付けさせておき、残り半数の新規エージェントに対してアイテム間 CF からユーザ間 CF に $t = 10$ で切り替えて推薦した場合の結果である。図 6、図 7 とともに、嗜好ベクトル \mathbf{p}_i を一様分布で与え、 $N_{\text{user}} = 500$ 、 $N_{\text{item}} = 500$ で実験を行った結果である。

5.6 考察 3

図 6 は、推薦アルゴリズムを切り替えた瞬間に効用の平均値がランダム推薦に近い値まで落ち込み、すぐに元よりも高い値まで増加することを示している。図 6 は $t = 30$ で切り替えた結果のみを示しているが、 $t = 10$ から $t = 70$ まで切り替えるタイミングを変化させた

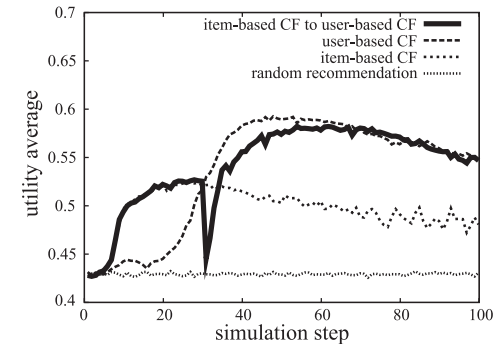


図 6 推薦アルゴリズムをアイテム間 CF からユーザ間 CF に切り替えたときの効用の平均値の推移 (\mathbf{p}_i は一様分布, $N_{\text{user}} = 500$, $N_{\text{item}} = 500$)

Fig. 6 Changes in utility average in the case where recommendation algorithm is switched from item-based CF to user-based CF (\mathbf{p}_i is generated based on uniform distribution, $N_{\text{user}} = 500$, $N_{\text{item}} = 500$).

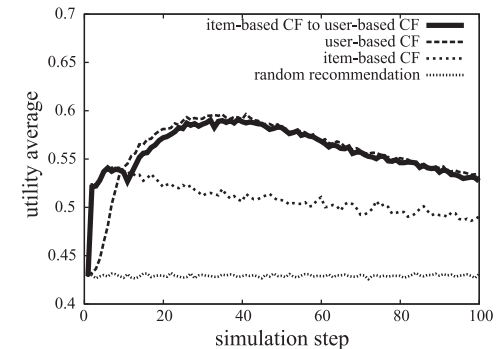


図 7 推薦システムにある程度の格付けが入力された状態において、推薦アルゴリズムをアイテム間 CF からユーザ間 CF に切り替えたときの効用の平均値の推移 (\mathbf{p}_i は一様分布, $N_{\text{user}} = 500$, $N_{\text{item}} = 500$)

Fig. 7 Changes in utility average in the case where recommendation algorithm is switched from item-based CF to user-based CF when several ratings are set in our recommender system beforehand (\mathbf{p}_i is generated based on uniform distribution, $N_{\text{user}} = 500$, $N_{\text{item}} = 500$).

ところ $t = 50$ 付近から切り替えれば効用の平均値の落ち込みが少ないことを確認している。

効用の平均値が落ち込む原因として、図 4 で示した、アイテムごとの格付け回数の差が影響していると考えられる。図 4 (右) によると、 $t = 30$ において、アイテム間 CF では、

数多く格付けされたアイテムや、まったく格付けされないアイテムが多く存在している。このような状況下でユーザ間 CF に切り替えると、数多く格付けされたアイテムを基に、エージェント間類似度は計算できるものの、そのアイテムはほとんどのユーザに推薦済みであるため、新たな推薦候補にはなりにくい。つまり、多くのエージェントに推薦済みのアイテムも、まったく推薦されていないアイテムも、どちらも有効な推薦候補とはならず、一時的にランダム推薦にせざるをえない状況が現れていると推測できる。しかし、エージェント間類似度は測定できているため、いったん落ち込んだ後の立ち上がり早いものと考えられる。

一方で、図 7 では切り替えた瞬間の効用の平均値の落ち込みはわずかである。この結果は、あらかじめランダムに入力した格付けが、有効な推薦候補を選ぶ際に寄与していることを示唆する。すなわち、今回の実験設定において、すでに一定数の格付けが入力されている状態から新規エージェントに対して推薦を行う場合、10 回程度の推薦であればアイテム間 CF を使用し、それ以降の推薦についてはユーザ間 CF による推薦に切り替えたほうが効果的であることが明らかになった。ただし、実システム上で複数のアルゴリズムを切り替える場合、あらかじめ切り替える時期を決定しておくことは、より複雑な状況になることが予測されるため難しく、格付け行列の状態から適切な推薦アルゴリズムを決定するなど、切替え方法の検討が必要である。

6. 結 論

推薦システムは、ユーザの評価を基に推薦するアイテムを決定する一方、その推薦結果によってユーザが閲覧するアイテムが変化し、次にどのアイテムの評価がフィードバックされるかに影響を及ぼす。そのため、推薦システムとユーザは相互作用的であるといえる。

本論文では、これまで考慮されてこなかったこの相互作用に着目し、購入格付けエージェントを主体とする、推薦システムの新たな評価モデルを提案した。また、提案モデルの有効性を検証するために、ユーザ間 CF およびアイテム間 CF を取り上げ、ユーザの評価が相互作用を通して逐次的に入力される状況における特徴分析を行った。

本論文で示した実験結果は、あらかじめいくつかの典型的な状況を設定したうえで、推薦アルゴリズムの基本的な特性を分析した結果である。実世界に近いシミュレーション結果を得るためには、より複雑な状況を示す実データを用いたモデルの拡張と検証が必要である。しかしながら、本論文での検証実験を通じて、アイテム間 CF はより早く効果的な推薦を実現できる一方で、ユーザ間 CF はユーザ効用のピーク値が高く、途中で切り替える手法が有

効であることが明らかになった。

このような実験結果は、従来の評価手法では得られなかった知見であり、提案するモデルが、より効果的な推薦を実現する指針について新たな観点で検証できることを示している。本論文で得た知見を、今後実システム上で有効に適用するためには、推薦アルゴリズムを切り替えるタイミングを予測する技術が必要不可欠である。また、提案モデルを使用することで、単純に推薦アルゴリズムを切り替えるだけでなく、複数のアルゴリズムをより複合的に使用した、新たなアルゴリズムの設計にも貢献できると考えられる。

参 考 文 献

- 1) Resnick, P. and Varian, H.: Recommender Systems, *Comm. ACM*, Vol.40, No.3, pp.56–58 (1997).
- 2) Schafer, J.B., et al.: E-Commerce recommendation applications, *Data Mining and Knowledge Discovery*, Vol.5, No.1-2, pp.115–153 (2001).
- 3) Linden, G., et al.: Amazon.com Recommendations: Item-to-Item Collaborative Filtering, *IEEE Internet Computing*, Vol.7, No.1, pp.76–80 (2003).
- 4) Goldberg, D., et al.: Using Collaborative Filtering to Weave an Information Tapestry, *Comm. ACM*, Vol.35, No.12, pp.61–70 (1992).
- 5) Resnick, P., et al.: GroupLens: An Open Architecture for Collaborative Filtering of Netnews, *Proc. CSCW '94*, ACM, pp.175–186 (1994).
- 6) Herlocker, J.L., et al.: Evaluating Collaborative Filtering Recommender Systems, *ACM Trans. Information Systems*, Vol.22, No.1, pp.5–53 (2004).
- 7) Herlocker, J.L., et al.: An algorithmic framework for performing collaborative filtering, *Proc. 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.230–237 (1999).
- 8) Adomavicius, G. and Tuzhilin, A.: Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions, *IEEE Trans. Knowledge and Data Engineering*, Vol.17, No.6, pp.734–749 (2005).
- 9) 平山巧馬, 小柳 滋: 協調フィルタリングにおける相関係数法の予測性能向上, *電子情報通信学会論文誌*, Vol.J90-D, No.2, pp.223–232 (2007).
- 10) Breese, J.S., et al.: Empirical Analysis of Predictive Algorithms for Collaborative Filtering, *Proc. 14th Conference on UAI'98*, pp.43–52 (1998).
- 11) Ungar, L.H. and Foster, D.P.: Clustering methods for collaborative filtering, *Proc. Workshop on Recommendation System at the 15th National Conf. on Artificial Intelligence* (1998).
- 12) Pennock, D.M., et al.: Collaborative Filtering by Personality Diagnosis: A Hybrid Memory- and Model-Based Approach, *Proc. 16th Conference on UAI'00* (2000).

- 13) 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* (2007).
- 14) Sarwar, B., et al.: Analysis of Recommendation Algorithms for E-Commerce, *Proc. 2nd ACM conference on Electronic commerce*, pp.158-167 (2000).
- 15) Herlocker, J.L., et al.: An Empirical Analysis of Design Choices in Neighborhood-Based Collaborative Filtering Algorithms, *Information Retrieval*, Vol.5, No.4, pp.287-310 (2002).
- 16) 岩田具治, 斉藤和巳, 山田武士: 最大エントロピー法に基づく推薦システムでのユーザ行動のモデル化, 第6回情報科学技術フォーラム (FIT2007) (2007).
- 17) Sarwar, B., et al.: Item-based Collaborative Filtering Recommendation Algorithms, *Proc. WWW'01*, pp.285-295 (2001).

(平成 20 年 4 月 17 日受付)

(平成 20 年 6 月 20 日再受付)

(平成 20 年 7 月 25 日採録)



山下 晃弘 (学生会員)

1983 年生. 2006 年北海道大学工学部情報工学科卒業, 2008 年同大学大学院情報科学研究科修士課程修了, 同年同大学院情報科学研究科博士課程入学, 同年より日本学術振興会特別研究員 (DC1), 現在に至る. 推薦システム, マルチエージェントシステム, 観光情報学に関する研究に興味を持つ. 人工知能学会, 観光情報学会各学生会員.



川村 秀恵 (正会員)

1973 年生. 1996 年北海道大学工学部情報工学科卒業, 2000 年同大学大学院工学研究科博士後期課程期間短縮修了. 同年同大学院工学研究科助手, 2006 年同大学院情報科学研究科助教授, 2007 年同准教授, 現在に至る. 飛行船ロボット, マルチエージェントシステム, 複雑系工学, 観光情報学等の研究に従事. 博士 (工学). 人工知能学会, 日本オペレーションズ・リサーチ学会, 電子情報通信学会, 観光情報学会各会員.



飯塚 博幸

1977 年生. 1999 年北海道大学工学部情報工学科卒業, 2001 年同大学大学院工学研究科博士前期課程修了, 2004 年東京大学大学院総合文化研究科博士後期課程修了. 同年同大学院特任研究員, 2005 年公立ほこだて未来大学特別研究員 (日本学術振興会 PD), 2005~2007 年イギリスサセックス大学客員研究員, 2008 年大阪大学大学院工学研究科特任助教, 同年同大学院助教, 現在に至る. 人工生命, 複雑系, 身体性認知科学, の研究に従事. 博士 (学術) 日本物理学会会員.



大内 東 (正会員)

1945 年生. 1968 年北海道大学工学部応用物理学学科卒業, 1974 年同大学大学院工学研究科博士後期課程修了. 同年同大学工学部助手, 助教授, 1989 年同大学大学院工学研究科教授. 2004 年同大学大学院情報科学研究科教授, 現在に至る. 飛行船ロボット, DNA コンピューティング, マルチエージェントシステム, 医療システム, 観光情報学等の研究に従事. 博士 (工学). 日本オペレーションズ・リサーチ学会, 医療情報学会, 観光情報学会各会員.