

解 説

推薦システムのアルゴリズム (2)

Algorithms for Recommender Systems (2)

神 嶋 敏 弘
Toshihiro Kamishima

産業技術総合研究所
National Institute of Advanced Industrial Science and Technology (AIST).
mail@kamishima.net, http://www.kamishima.net/

Keywords: recommender system, collaborative filtering, content-based filtering, serendipity, privacy, shilling, groupLens.

この連載では全3回で、利用者が好むであろうものを予測して提示する推薦システムのアルゴリズムについて解説する。第2回である今回は、6章～10章を掲載する。

6. 嗜好の予測

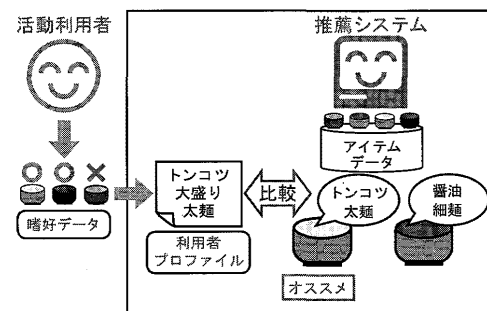
嗜好の予測とは、活動利用者の嗜好データや、アイテムの特徴を用いて、活動利用者の各アイテムへの関心や好みの度合いを予測することである。

嗜好の予測段階の実現方法は大きく二つに分類される。レンタルビデオ店で、顧客が見たい映画を推薦する場合を考えてみよう。一つは、ファンである監督、好みのジャンルを利用者に尋ねてその条件に合ったものを選ぶ方法である。これは、検索対象の内容を考慮して推薦をするので**内容ベースフィルタリング (content-based filtering)**と呼ばれる。もう一つは、映画の趣味が似ている知り合いに、おもしろかった映画を教えてもらう「口コミ」の過程を自動化する方法である。ほかの人との協調的な作業によって推薦対象を決めるため、この推薦手法は**協調フィルタリング (collaborative filtering)**や社会的フィルタリング (**social filtering**) と呼ばれている。

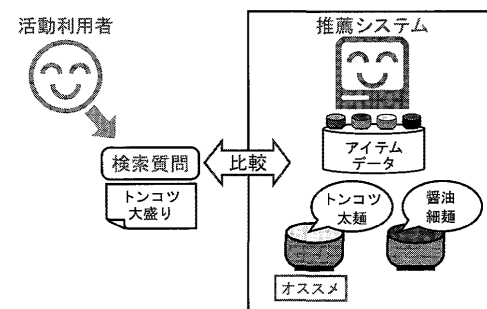
現在では、どちらの方法にもいろいろな派生型が提案されているが、純粋な形では図4のように予測する。内容ベースフィルタリング (図4 (a) と (b)) では、アイテムの性質と利用者の嗜好パターンを比較して、利用者が好むと判断したものを推薦する。このアイテムの性質は特徴ベクトルによって記述される。特徴ベクトルとは、アイテムのいろいろな側面の性質を表す特徴を集めて、ベクトルの形にしたものである。各特徴は、事前に定めた定義域中の値をとることで、そのアイテムの性質を表現する。ラーメンの例を示そう。このとき、スープの種類、麺の太さ、価格といった特徴からなるベクトルでラーメンを表現する。スープの種類という特徴は、トンコツ、醤油、塩のような定義域の値の一つをとり、価格という特徴では自然数がその定義域となる。そして、ある特定のラーメン『トンちゃん』があるとする

(スープの種類=トンコツ, 麺の太さ=細麺, …, 価格=650円)

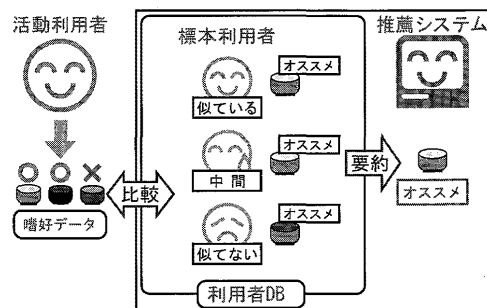
といったベクトルで表現される。こうした情報をいろいろなアイテムについて収集したものをアイテムデータと呼んでおく。一方、利用者の嗜好パターンは利用者プロフィール (user profile) によって表す。この利用者プロフィールを間接指定するシステムと、直接指定するもの



(a) 内容ベースフィルタリング (間接指定型)



(b) 内容ベースフィルタリング (直接指定型)



(c) 協調フィルタリング

図4 内容ベースフィルタリングと協調フィルタリング

がある。間接指定型 (図 4 (a)) では、利用者のいろいろなアイテムに対する嗜好データ、すなわち、好き嫌いの度合いを定量化したデータを集める。この嗜好データと、アイテムデータに基づいて、その利用者が好むアイテムの特徴のパターンを機械学習の手法でモデル化し、利用者プロフィールとする。直接指定型 (図 4 (b)) では、利用者が明示的に、自身が好むアイテムの特徴を表した検索質問 (query) (批評 (critique) と同じ) を入力する。一般的な検索質問は「スープの種類は醤油で、価格は 500 円以下」といった、特徴に対する制約の形式だが、自然言語文などを扱えるものもある。この検索質問はそのまま利用者プロフィールとして用いられる。内容ベースフィルタリングでは、アイテムデータ中のアイテムの特徴ベクトルと、利用者プロフィールとを比較し、プロフィールに最も近い特徴ベクトルをもつアイテムを利用者が好むものと判断して推薦をする。

もう一つの協調フィルタリング (図 4 (c)) では、アイテムの性質は全く考慮しない。その代わりに、システムが利用される前に、多くの利用者の、いろいろなアイテムに対する嗜好データをデータベースに集積している。このデータベースを利用者データベース (user database) (「利用者 DB」と略す) と呼び、この利用者 DB に嗜好データを登録している利用者を標本利用者 (sample user) と呼ぶ。さらに、今までの嗜好パターン、すなわち、どのアイテムを好み、どのアイテムを嫌うのかという傾向が類似している利用者は、これからも同じアイテムを好み、同じアイテムを嫌うであろうという仮定を導入する。この仮定のもと、活動利用者と、嗜好パターンが類似している標本利用者を見つけ、これらの標本利用者が好むものを活動利用者に推薦する。

なお、内容ベースフィルタリングは、いろいろな拡張が行われているので、厳密に定義するのは難しい。[Burke 02] では、デモグラフィックな特徴 (5.5 節) などを用いず、アイテムの特徴のみを用いた、間接指定型の手法のみを内容ベースフィルタリングと定義している。そして、上記の直接指定型にあたる知識ベース型や、デモグラフィックな特徴を使うものも、別の種類として細かく分類している。しかし、細分化しても多種多様な手法を厳密に分類するのは実際には難しいので、本稿では、広義に捉えて、アイテムやデモグラフィックな特徴を利用するような方法はすべて内容ベースフィルタリングとして扱う。一方、これらの特徴を用いず、標本利用者の嗜好データのみを用いる方法を協調フィルタリングとしておく。

7 章～10 章では協調フィルタリングについて、11 章では内容ベースフィルタリングの各種手法を紹介する。12 章の嗜好の予測段階のまとめでは、これら二つの手法を組み合わせるハイブリッド法について述べたのち、アルゴリズムの選択の指針について述べる。これらの話題に移る前に、協調フィルタリングと内容ベースフィル

タリングの長所と短所をまとめておく。

6.1 内容ベースと協調フィルタリングの比較

協調フィルタリングと内容ベースフィルタリングの長所と短所を [Balabanović 97, Burke 02] などに基づき表 2 にまとめた。以下、表中の各項目について詳しく述べる。

表 2 協調フィルタリングと内容ベースフィルタリングの比較

	協調	内容ベース
セレンディピティ	○	×
ドメイン知識	○	×
cold-start問題	×	△
利用者数	×	○
被覆率	×	○
類似アイテム	×	○
少数派の利用者	×	○

§1 セレンディピティ

3.1 節で述べたセレンディピティについては、定量的な評価が確立していないので断言はできないが、協調フィルタリングが有利といわれている。内容ベースフィルタリングでは、利用者自身が知っているアイテムの特徴に、推薦対象が制限されてしまうことが多い。例えば、映画の場合であれば、過去に見たのと同じジャンルや監督の作品が推薦される。さらに、類似した内容のアイテムを推薦し、それを利用者が受け入れることで、利用者プロフィールの偏りが一層強化される現象も生じる。それに対して、協調フィルタリングでは、自身が知らないジャンルや監督でも、ほかの標本利用者の知識を通じて知ることができる場合がある。そうしたときには意外性のある、すなわち、セレンディピティがある推薦ができるとされている。

§2 ドメイン知識

協調フィルタリングの最も重要な長所は、アイテムの特徴、すなわち、アイテムのドメイン知識を全く必要としないことである。アイテムの特徴ベクトルの設計に伴う困難には次のようなことがある。

- アイテムの特徴についてのデータを集める手間やコストが必要になる。仕様についてのデータベースが整備されている商品分野は書籍、CD などに限られている。たとえ整備されていても、食品の成分表などのように推薦という目的にはあまり役立たないものもある。それ以外の分野ではデータベースの構築・更新コストが必要になる。
- アイテムのある性質を表す特徴がないために、適切な推薦ができないことがある。例えば、映画の推薦で、自分がファンであるカメラマンの撮る映画を見る利用者がいたとしても、映画の特徴ベクトルに『カメラマン』の特徴がなければ、内容ベースでは適切な推薦ができない。
- どの特徴を採用するかということが、利用者の判断に影響を与える。推薦アイテムの決定に利用された

特徴は、利用者の意思決定でより重要視され、そうでない特徴は無視されるようになる副作用を生じることがある [Häubl 01].

- 違う分野のアイテムを推薦することが困難である。例えば、ある映画が好きでも、内容ベース法では、そのサントラの CD を推薦することは難しい。なぜなら、映画と CD は、異なる特徴ベクトルで表現されているためである。

§ 3 cold-start 問題

cold-start 問題 (start-up 問題や ramp-up 問題とも呼ばれる) は、2 種類に分けられる。一つは、新たにシステムを利用し始めた利用者 (first rater や early rater という) に対して適切な推薦をする難しさであり、もう一つは、推薦対象として新たにシステムに導入されたアイテムを推薦する難しさである。前者については、協調フィルタリングでは、ほかの標本利用者との類似性を判定できないため良い推薦をするのは難しい。間接指定型の内容ベースフィルタリングでも、利用者プロフィールが不完全になるためやはり難しい。加えて、利用者本人の評価値しか利用できないので、問題がさらに深刻になることがある。だが、直接指定型の内容ベース法では、自身の利用者プロフィールを直接記述するためこうした問題は生じない。よって、観光地の案内端末での推薦など、同一利用者の継続的な利用があまりない状況では、直接指定型の内容ベース法を採用すべきである。一方、継続的な利用がなされるならば、嗜好データの蓄積に伴って予測精度が向上する協調フィルタリングや間接指定型の内容ベースフィルタリングがよいであろう。

一方、後者の新規アイテムに対する問題については、内容ベースフィルタリングが優れている。内容ベースフィルタリングでは、利用者プロフィールがあれば、新規のアイテムでも、その特徴ベクトルを手がかりに、全く問題なく推薦ができる。しかし、協調フィルタリングでは、標本利用者の評価を利用するので、まだ誰も評価していない新規のアイテムは、推薦を予測する手がかりがなく、そのアイテムに対する活動利用者の好みを判断できない。以上のことから、商品が頻繁に入れ換わるような場合は、内容ベースフィルタリングが有利である。

§ 4 利用者数

内容ベース法の場合は、たとえシステムの利用者が一人であっても推薦は可能である。一方、協調フィルタリングは、ほかの利用者の意見を参照するので、利用者がある程度いなければ実行できない。また、利用者がほかにいない状況では、適切な推薦もできないため、新たな利用者の増加も難しい。そのため、電子商取引サイトでのポイントの配布といったインセンティブや、暗黙的な方法で自動収集するといった手段で、十分な規模の嗜好データ量を維持する必要がある。

§ 5 被覆率

3.1 節で述べた被覆率は、適合アイテム列挙タスク

(2.3 節) では特に重視すべき評価規準である。協調フィルタリングでは、まだ誰も評価していないアイテムは推薦の対象にできないため、被覆率を 100% にはできない。一方、内容ベースでは、アイテムの特徴を手がかりに、すべてのアイテムについて推薦すべきかどうかの判定をすることができる。よって、特徴に欠損値がなければ、被覆率は 100% となる。

§ 6 類似アイテム

ドメイン知識の項では、アイテムの特徴を使うことの短所を述べたが、アイテムの特徴を使うことの長所もある。アイテムの性質を無視する協調フィルタリングでは、同じ商品のサイズや色の違うもの、また、同じ目的の競合商品などは全く異なるものとして扱われる。例えば、利用者 A と B はともに、同じサッカーチームのファンであったとしよう。だが、A はこのチームのロゴの入ったマグカップを、B はロゴ入りのタオルを買ったとする。すると、協調フィルタリングでは、チームロゴという特性を明示的に考慮できないので、そのことを推薦に反映できない。一方、内容ベースフィルタリングでは、利用者 A が過去にマグカップとタオルを購入していれば、アイテムの特徴から、そのサッカーチームのロゴ入りのほかのアイテムを推薦することも可能である。

また、推薦されたアイテムを利用者が拒否した場合、そのサイズや色が違うだけの類似アイテムを推薦されてしまう場合も、協調フィルタリングでは生じる。例えば、ある商品を却下したすぐ後で、その商品の色違いを推薦されることがよくある。これを回避するには、どのアイテムとどのアイテムを同じとみなすかが重要になる。すなわち、JAN コードなどの商品 ID が異なっても、協調フィルタリングで扱う場合には、色違いなどの類似アイテムは同じアイテムとして扱うようにする必要がある。しかし、どのアイテムを同じとみなすかは、アイテムのドメインに依存した難しい問題である。例えば、服飾などでは色の違いは重視されるだろうが、ティッシュペーパーなどは色違いでも同じアイテムとみなしてよいだろう。

§ 7 少数派の利用者

協調フィルタリングでは、非常に少数派の嗜好パターンをもつ利用者は、類似した嗜好パターンをもつ利用者を全く、またはごく少数しか見つけられないので、適切な推薦を受けられない場合がある。例えば、ほとんど無名なタレントだが、利用者と出身地が同じであるのでファンである人がいたとしよう。こうした嗜好をもつ人は非常にまれであろう。さらに、そうした人が同じシステムを利用していることはさらにまれである。すると、協調フィルタリングでは、類似した嗜好の人がいなければ嗜好を予測できないので、こうした観点からの推薦は難しい。一方、内容ベースフィルタリングでは、タレントの出身地情報を用いて適切な推薦をすることも可能である。

7. 嗜好予測：協調フィルタリング

二つの嗜好予測手法のうち、まず協調フィルタリングについて述べる。以後用いる記号表記を表 3 にまとめた。

最初に、協調フィルタリングによる嗜好予測について述べる。ここでいう予測とは、活動利用者はまだ知らないが、ほかの標本利用者は知っているアイテムについて、活動利用者の関心の有無や、評価値を推定することである。このように、与えられたデータの中から、規則性を見つけ出し、その規則性に基づいて予測する問題は、機械学習や統計的予測の枠組みによって解く。だが、1・1 節で述べたように、万能な予測手法は理論的にあり得ない。よって、3 章で述べた、利用者数やアイテム数などのデータの特性、望ましい推薦が備えるべき性質を考慮してモデルやアルゴリズムを選択しなくてはならない。もしこれらが適切でなければ、データがいくらあっても予測精度が向上することはないので、この選択は重要である。これらの選択に関連した、モデルやアルゴリズムの特徴に留意して述べるので、参考にされたい。

7.1 メモリベース法とモデルベース法

個々のアルゴリズムについて述べる前に、協調フィルタリング手法の分類と、それぞれのグループの長所と短所を述べる。推薦候補の予測手法はメモリベース法 (memory-based method)^{*1} とモデルベース法 (model-based method) に分けられる [Breese 98]。メモリベース法では、推薦システムが利用される以前には何もせず、ただ利用者 DB を保持している。そして、推薦をするときに、利用者 DB 中の嗜好データそのものと、活動利用者の嗜好データを併せて予測をする。もう一方のモデルベース法では、推薦システムが利用される以前に、あらかじめモデルを構築する。このモデルは、「佐藤さんが好むものは、鈴木さんも好むことが多い」といった、利用者とアイテムの嗜好についての規則性を表す。推薦をす

表 3 本稿の主な記号表記一覧

a	: 活動利用者を示す添字
n	: 総利用者数
m	: 総アイテム数
\mathcal{U}	: 利用者の定義域 $\{1, \dots, n\}$
\mathcal{U}_j	: アイテム j を評価した利用者の集合
\mathcal{I}	: アイテムの定義域 $\{1, \dots, m\}$
\mathcal{I}_i	: 利用者 i が評価したアイテムの集合
\mathcal{R}	: 評価値の定義域 (そのほかに欠損値もある)
r_{ij}	: 利用者 i のアイテム j への評価値
S	: r_{ij} を要素とする行列
\bar{r}_i	: \mathcal{I}_i 中のアイテムへの評価値の平均
\bar{r}_j	: \mathcal{U}_j 中の利用者アイテム j への評価値の平均
x	: 利用者 i を示す確率変数
y	: アイテム j を示す確率変数
r	: 評価値を示す (確率) 変数
$\langle Y^{(u)} \rangle$: アイテムの時系列

*1 事例ベース法 (instance-based method) ともいう。

表 4 メモリベース法とモデルベース法の比較

	メモリベース法	モデルベース法
推薦時間	×: 遅い	○: 速い
適応性	○: あり	×: なし

るときに、利用者 DB は用いずに、このモデルと活動利用者の嗜好データとに基づいて予測する。すなわち、事前にモデルを構築するかどうかという違いが重要である。

これら二つの手法の大まかな長所と短所を表 4 にまとめておく。推薦時間についてだが、メモリベース法は一般に遅い。これは、利用者 DB には多数の標本利用者や推薦対象が登録されており、これら多数の項目を推薦の度に調べ直すのに時間がかかるためである。一方のモデルベース法では、夜間などシステムが利用されない間や、毎月や毎週など定期的にモデルをあらかじめ構築・更新しておく。すると、この時間は推薦の時間からは除外できる。また、モデルの規模は、利用者 DB のそれと比べて小さいので、活動利用者に速く推薦することができる。

表 4 の適応性とは、標本利用者や推薦対象が、削除されたり追加されたりしても、適切な推薦ができるかということである。このような削除や追加が行われると、モデルベース法では、利用者や推薦対象間の規則性に变化が生じるため、モデルを再構築する必要が生じる。だが、モデルの構築には時間がかかるので、頻繁に行うことは難しい。そのため、適応性に関してはモデルベース法は不利になる。一方、メモリベース法では、モデルの構築は行わないためこのような問題は生じない。

8 章にてメモリベース法の、9 章にてモデルベース法の手法を個別に紹介し、10 章でこれらのハイブリッド手法を述べる。

8. 協調フィルタリング：メモリベース法

メモリベース法 (memory-based method) とは、利用者 DB を直接利用して、活動利用者の嗜好を推定する方法である。機械学習の観点からは k 近傍法 (k -nearest neighbor method) であるといえる。この方法は、大きく利用者間型 (user-user type) とアイテム間型 (item-item type) とに分類できる。どちらも、類似性を計算する点では共通しているが、利用者間型は、活動利用者と嗜好パターンが似ている利用者をまず見つけ、彼らが好むものを推薦する。一方、アイテム間型では、活動利用者が好むアイテムと類似したアイテムを推薦する。これらの手法を順に紹介する。

8.1 利用者間型メモリベース法

利用者間型メモリベース法の代表的な手法である GroupLens の方法 [Resnick 94] について述べる。これは、協調フィルタリングの手順を自動化した先駆的

研究である。シンプルな手法だが、その予測精度は高く、多くの改良研究もなされている。GroupLensは、当初はNetNewsの記事を推薦するシステムであったが、現在は映画の推薦システムMovieLensとなっている。この実験システムは、プロジェクトのホームページ[MovieLens]で公開されているので、推薦システムを体験するため利用してみてもよいだろう。

レストランを探す場合のことを考えてみよう。このとき、自分と食べ物の嗜好が似ている何人かの人に尋ねてみて、彼らの意見をもとにどの店で食事をするか決めたりすることがあるだろう。GroupLensの方法は、こうした人のコミュニティでの「口コミ」による推薦の過程を、次の2段階で実現する。

- (1) 類似度の計算：利用者DB中の各標本利用者と活動利用者の嗜好の類似度を求める。類似度とは、嗜好パターンがどれくらい似ているかを定量化したものである。
- (2) 嗜好の予測：活動利用者は知らないアイテムについて、それらのアイテムへの標本利用者の好みと、その標本利用者との間の類似度に基づいて、活動利用者がどれくらいそのアイテムを好むかを予測する。

これらの段階について詳細に説明する前に、記号を定義する。 n 人の全利用者の集合を $\mathcal{X}=\{1, \dots, n\}$ 、 m 種類の全アイテムの集合を $\mathcal{Y}=\{1, \dots, m\}$ とする。評価値行列 S は利用者 $i \in \mathcal{X}$ の、アイテム $j \in \mathcal{Y}$ への評価値 s_{ij} を要素とする行列である。 s_{ij} は、評価済みなら評価値の定義域 \mathcal{R} のいずれかの値をとり、未評価なら欠損値「*」をとる。例えば、5段階のスコアを用いた採点法(5・2節参照)で獲得した嗜好データであれば、評価値の定義域は $\mathcal{R}=\{1, \dots, 5\}$ となる。活動利用者を添字 a で表す、すなわち、 s_{aj} は活動利用者のアイテム j への評価値である。また、利用者 i と活動利用者が評価済みのアイテムの集合を、それぞれ $\mathcal{Y}_i=\{j \text{ s.t. } s_{ij} \neq *\}$ と \mathcal{Y}_a で表す。活動利用者と利用者 i の類似度は、共通に評価しているアイテムについてのPearson相関で測る。

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

ただし、 \mathcal{Y}_{ai} は二人が共通に評価したアイテムの集合、すなわち、 $\mathcal{Y}_{ai} = \mathcal{Y}_a \cap \mathcal{Y}_i$ で、 $\bar{s}_i' = \sum_{j \in \mathcal{Y}_{ai}} s_{ij} / |\mathcal{Y}_{ai}|$ である。なお、 $|\mathcal{Y}_{ai}| \leq 1$ 、すなわち、活動利用者と標本利用者が共通に評価したアイテムが一つ以下ならば、Pearson相関は計算できないので $\rho_{ai}=0$ とする。アイテム $j \notin \mathcal{Y}_a$ の評価値は、式(1)の類似度で重み付けした、各標本利用者のアイテム j への評価値の加重平均で予測する。

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

ただし、 \mathcal{X}_j はアイテム j を評価済みの利用者の集合で、 $\bar{s}_i = \sum_{j \in \mathcal{Y}_i} s_{ij} / |\mathcal{Y}_i|$ である。なお、 $a \in \mathcal{X}_j$ なる状況では、 s_{aj} の推定が不要になるので想定しなくてよい。

式(2)の第1項は、第2項が中間的な評価で0をとるので、それを補正するバイアス項である。また、第2項の分子は上記の加重平均であり、分母は評価している利用者が多い、すなわち $|\mathcal{X}_j|$ が大きいと加重平均が大きくなりやすい問題に対するスケーリング項である。また、[Resnick 94]では、式(2)の第2項のように、活動利用者と標本利用者がともに評価したアイテム \mathcal{Y}_{ai} 上での評価値の平均 \bar{s}_j' を用いている。だが、現在のGroupLensプロジェクトでは \bar{s}_i を用いているとのことであり*2、また、著者の実験でも、どちらを用いても有意な差は見られなかった。加えて、 \bar{s}_i のほうが、活動利用者が決まる前に事前に計算できる利点があるので、実際に利用するときには \bar{s}_i を用いるほうがよいだろう。

ここで簡単な例をあげよう。表5はどんぶり専門店の評価値行列 S の例である。4人の利用者は各行に対応し、4種類のアイテム(=どんぶり)は列に対応する。評価値は3段階の採点法、すなわち、 $\mathcal{R}=\{1, 2, 3\}$ である。この表は、利用者3の佐藤のアイテム2の牛井の評価値は1で、嫌いであることを示している。具体的に、活動利用者を田中、すなわち $a=2$ とし、2: 田中の1: 親子丼への推定評価値 $\hat{s}_{2,1}$ を求めてみよう。

まず、式(1)の相関係数を求める。親子丼を評価済みの利用者、すなわち \mathcal{X}_1 に含まれる各利用者と活動利用者の間の相関係数を求める必要がある。すると、1: 山田、3: 佐藤、4: 鈴木の3人とも親子丼を評価済みなので $\mathcal{X}_1=\{1, 3, 4\}$ の各利用者との相関係数を求める。2: 田中と1: 山田の相関 $\rho_{2,1}$ は、共通に評価しているアイテムが2: 牛井だけで、1個以下なので $\rho_{2,1}=0$ となる。次に、2: 田中と3: 佐藤の間の相関係数を計算する。この二人が共通に評価しているアイテムは2: 牛井と3: 海鮮丼なので $\mathcal{Y}_{2,3}=\{2, 3\}$ となる。すると、これらのアイテムについての $\mathcal{Y}_{2,3}$ 上の平均評価値はそれぞれ

$$\bar{s}_2' = \frac{\sum_{k=2,3} s_{2,k}}{2} = \frac{1+3}{2} = 2$$

$$\bar{s}_3' = \frac{\sum_{k=2,3} s_{3,k}}{2} = \frac{1+3}{2} = 2$$

で、相関係数は次式になる。

$$\rho_{2,3} = \frac{\sum_{k=2,3} (s_{2,k} - \bar{s}_2') (s_{3,k} - \bar{s}_3')}{\sqrt{\sum_{k=2,3} (s_{2,k} - \bar{s}_2')^2} \sqrt{\sum_{k=2,3} (s_{3,k} - \bar{s}_3')^2}}$$

表5 評価値行列 S の例

	1: 親子丼	2: 牛井	3: 海鮮丼	4: カツ丼
1: 山田	1	3	*	3
2: 田中	*	1	3	*
3: 佐藤	2	1	3	1
4: 鈴木	1	3	2	*

*2 J. Riedl のメールより。

$$= \frac{(1-2)(1-2)+(3-2)(3-2)}{\sqrt{(1-2)^2+(3-2)^2} \sqrt{(1-2)^2+(3-2)^2}} \\ = 1$$

同様に計算すると、2: 田中と 4: 鈴木の相関は $\rho_{2,4} = -1$ となる。

次に推定評価値を計算する。まず、2: 田中の全評価済みアイテム上の平均評価値を求める。

$$\bar{s}_2 = \frac{\sum_{k=2,3} s_{2,k}}{2} = \frac{1+3}{2} = 2$$

最後に、これまで計算した値を式 (2) に代入すると

$$\hat{s}_{2,1} = \bar{s}_2 + \frac{\sum_{i=1,3,4} \rho_{2,i}(s_{i,1} - \bar{s}_i)}{\sum_{i=1,3,4} |\rho_{2,i}|} \\ = 2 + \frac{0(1-3)+1(2-2)+(-1)(1-5/2)}{|0|+|1|+|-1|} \\ = 2.75$$

によって、2: 田中の 1: 親子丼への推定評価値は 2.75 と計算できる。この値は、最大値 3 にかなり近く、2: 田中は 1: 親子丼が好きであると予測される。

§ 1 利用者間型メモリベース法の改良

GroupLens の方法にはいろいろな改良が試みられている。[Herlocker 99] では、いろいろな改良を実験的に検証している。すべての計算の前に、評価値 s_{ij} から利用者 i の平均評価値 \bar{s}_i を引いて正規化しておくことで予測精度は向上する。これは、5・2 節で述べたように、計測された評価値揺らぎや偏りがある。肯定的でも否定的でもない評価値を 0 に正規化することで、こうした不整合が緩和されるためであろうと著者は考える。利用者の近傍を使う改良もある。ここでの近傍とは、式 (1) の相関が大きな、すなわち活動利用者と類似した嗜好をもつ利用者の集合のことである。式 (2) の推定評価値は、アイテム j を評価済みのすべての利用者の評価に基づいているが、これを事前に計算した活動利用者の近傍のみに基づいて計算する。実験によれば、近傍利用者数がある程度以上になると、それ以降は増やしても予測精度は向上しない。よって、近傍利用者だけに計算を限定することで、計算を高速化でき、データベースアクセスも抑制できる。ただし、モデルベース法のモデルほど頻繁にする必要はないが、近傍は定期的に更新しなくてはならないので、純粋なメモリベース法の利点は部分的には失われる。また、新規の参加者については近傍を新たに計算する必要が生じ、脱退者がほかの利用者の近傍利用者であれば予測精度の低下を招く。以上の、二つの改良は、ほとんどのデータに対して有効なので、適用しておくほうがよいだろう。その他、この [Herlocker 99] では、データが疎である問題に対処するため、活動利用者と標準利用者が共通に評価しているアイテムの数に応じて重み付けしたり、式 (1) の Pearson 相関の代わりに順位相関を使ったりすることは予測精度の向上に役立ったと報告している。

[Breese 98] にも、いくつかの改良が示されている。

その一つに、評価値の数が少ない場合に有効なデフォルト投票 (default voting) がある。利用者間の類似度は、共通に評価しているアイテム \mathcal{Q}_{ai} への評価に基づいて計算する。しかし、そうしたアイテムが少ないか、全くない場合には、適切に類似度を評価できない。例えば、利用者 1 と 2 の評価済みアイテム集合がそれぞれ $\mathcal{Q}_1 = \{1, 2, 3\}$ と $\mathcal{Q}_2 = \{1, 4, 5\}$ とする。このとき、これら二人の類似度は共通して評価しているアイテム 1 に対する評価値 $s_{1,1}$ と $s_{2,1}$ の二つの値だけに依存する。このとき類似度は、共通に評価したアイテム数 $|\mathcal{Q}_1 \cap \mathcal{Q}_2|$ は一つだけなので 0 となってしまう。そのため、これらの利用者間の類似度は全く推薦の精度向上に寄与しない。そこで、利用者 1 の $s_{1,2}$, $s_{1,3}$ と、利用者 2 の $s_{2,4}$, $s_{2,5}$ の評価値を活用する。例えば、 $s_{1,2}$ の評価値は、利用者 2 の対応する値 $s_{2,2}$ があれば相関の計算に利用できる。そこで、アイテム 2 の中立的な評価値、すなわち、全利用者のアイテム 2 への評価値の平均 $\bar{s}_j = \sum_{i \in \mathcal{U}} s_{ij} / |\mathcal{U}_j|$ をデフォルト投票値とし、 $s_{2,2}$ の値をこのデフォルト投票値で置き換える。ほかのアイテムについても同様の置換えをすると、アイテム集合 $\mathcal{Q}_1 \cup \mathcal{Q}_2 = \{1, 2, 3, 4, 5\}$ 上でより適切に利用者間の類似度を評価できるようになる。著者の実験でも、一人当たりの評価値数が少ない状況でかなりの効果が確認された。また、 s_{ij} の欠損を補うときに、アイテムの平均評価値 \bar{s}_j の代わりに、利用者 i の全アイテム上の平均評価値 $\bar{s}_i = \sum_{j \in \mathcal{Q}_i} s_{ij} / |\mathcal{Q}_i|$ も利用する実験も行ったが、あまり良い結果は得られなかった。この [Breese 98] では、最低や最高など両端に近い評価値をより重視する改良や、評価している利用者が少ないアイテムへの評価値を重視する改良などを提案している。

3・2 節で述べたように推薦システムが扱うデータは非常に疎である。機械学習では、このような場合には、次元縮約の前処理を実行することが多い。代表的な方法に主成分分析^{*3} [Bishop 06, Bishop 08, 石井 98] がある。GroupLens の方法では、各利用者は m 個のアイテムへの評価値を要素とするベクトルで表される。これをデータの分散に関する情報をできるだけ失わないように、 $K < m$ の K 要素のベクトルで各利用者表現するのが主成分分析である。この K 要素に情報を圧縮したベクトルで、利用者間の類似度を求める方法もよく用いられている [Canny 02a, Zhang 06]。ただし、主成分分析は欠損値があると実行できないので、上記のデフォルト投票値や EM アルゴリズムなどによって適当に評価値行列 S を補完する必要がある。

8・2 アイテム間型メモリベース法

GroupLens などの利用者間型では、評価値行列 S の行ベクトル、すなわちある利用者のいろいろなアイテム

^{*3} 自然言語処理分野では latent semantic analysis (LSA) [Baeza-Yates 99] と呼ばれる。

への嗜好を表すベクトルの類似度に基づいて、ほかの利用者の評価値を予測した。一方、アイテム間型メモリベース法では、アイテムベクトル、すなわち、 S の列ベクトルの類似度を用いる。これは、いろいろな人に同じような評価を受けるアイテムは似ていると考え、関心があるアイテムの類似アイテムに利用者は関心をもつという仮定に基づいている。簡単な方法としては、アイテムベクトルのコサイン [Linden 03] や、単純な共起性 [Das 07], Pearson 相関などでアイテム間の類似度を測り、利用者が閲覧中や、買い物かごに入れている、もしくは、直近の利用履歴にあるアイテムと類似しているアイテムを推薦する。これらの方法によって、協調フィルタリングの枠組みで、一時的個人化 (2.1 節) をした推薦ができる。

GroupLens の方法と同様に加重平均を使う方法としては [Sarwar 01] がある。この方法では、活動利用者のアイテム j への推定評価値を次式で求める。

$$\hat{s}_{aj} = \frac{\sum_{k \in \mathcal{I}_a} s_{ak} \rho'_{kj}}{\sum_{k \in \mathcal{I}_a} |\rho'_{kj}|} \quad (3)$$

ただし、 ρ'_{kj} はアイテム k と j の類似度である。この方法は [Ali 04] でも利用され、アイテム間の類似度行列と活動利用者自身の評価値があれば利用者 DB を参照することなく、ローカルマシンだけで推薦を計算できる。よって、ある程度のプライバシー保護 (14.2 節) も実現できる。GroupLens の方法と同様に、 \mathcal{I}_a に含まれるアイテムすべてではなく、 ρ'_{kj} がしきい値以上の近傍に計算を限定すると、予測精度を下げることなく、式 (3) の計算を高速化できる。だが、近傍アイテムが重要なので、公開中の映画・コンサートなどのように推薦対象のアイテムが頻繁に入れ換わるような状況^{*4}では、頻繁な近傍の更新が必要になってしまう。この方法は、評価しているアイテム数が少ない利用者には推薦が早くできる。利用者間型より予測精度は良いとの報告もある。しかし、理論的根拠はないが、実験的には特定のアイテムに推薦が偏る傾向が強く [McNee 06], セレンディピティについてはアイテム間型は不利であるといわれている。

8.3 メモリベース法に関するその他の研究

5.2 節でも述べように、利用者の評価値には一貫性が低い問題がある。そこで、McLaughlin ら [McLaughlin 04] は、投票された評価値を隣接した評価値に配分する信念分配 (belief distribution) を利用した GroupLens 法の拡張を示した。利用者が指定した評価値に最も大きな重みを与えるとともに、その周囲の評価値にも小さな重みの評価値を与える。例えば、評価値 4 を利用者が指定したとき、評価値 4 には 0.6 の重みで、それに隣接する評価値 3 や 5 には 0.2 の重みがあると考え。すなわち、ファジィ理論でのメンバシップ関数のようなもので

ある。しかし、この重みの分配の割合は予測精度に大きく影響するが、この割合は調整は試行錯誤によって行わなければならない問題がある。

5.2 節では、利用者から嗜好データを得るために、採点法や格付け法ではなく順位法を用いるなんとなく協調フィルタリング [Kamishima 03b, Kamishima 06] について述べた。ここでは、順位法で得た嗜好順序を使って推薦する方法について述べる。手法は非常に単純で、嗜好順序中のアイテムの順位をそのまま評価値とする。例えば、アイテム 1, 2, 3, 4 を好きなものから順に $3 > 2 > 4 > 1$ と並べた場合、アイテム 4 の順位は 2 であり、 $s_{i4} = 2$ として扱う。このとき、 s_{ij} が小さいほど好きなことを表すので、式 (2) の \hat{s}_{aj} が小さいものから順に推薦する。また、順位法では評価は常に相対的なもので、予測評価値も相対的な好みしか示さないことに注意してはならない。利用者ごとに整列したアイテム数が異なる場合には、嗜好順序の長さが一定ではない。この場合は、嗜好順序の長さを l とし、この嗜好順序中のアイテム j の順位を rank_j としたとき、 $s_{ij} = \text{rank}_j (l+1) / (m+1)$ とする。これは、嗜好順序に含まれる l 個の対象が、 m 個の全対象から一様にランダムに選ばれたとしたときの、アイテム j の順位の期待値である。著者の実験 [Kamishima 06] では、採点法で得た評価値を正規化するなどしても、順位法による嗜好データに基づく予測順序の精度のほうが良かった。

GroupLens の方法では、利用者間の類似度は式 (1) の Pearson 相関で測っている。実験的にはこの類似度がかなり良い予測精度が得られているが、理論的な根拠は弱い。そこで、さらに予測精度を向上させるため、経験損失を最小にするようにこの類似度関数を学習する方法 [Bell 07] もある。これにより予測精度は向上するが、7.1 節で述べたモデルベース法と同じ適応性の問題を生じてしまう。

純粋に理論的な観点から、Pennock ら [Pennock 00b] は、望ましい協調フィルタリングが備えるべき公理的性質について論じた。これは、社会全体での意思決定を論じる社会選択の研究で著名な Arrow の不可能性定理 [Arrow 63] のような考え方である。四つの公理的性質として、全定義域と最小機能 (universal domain and minimal functionality), 全員一致 (Pareto 性, unanimity), 無関係な候補からの独立性 (independence of irrelevant alternatives), スケール不変性 (scale invariance) をあげている。これらの、性質を満たす予測手法は最近傍法のみであることを示している。実際の推薦システムで、これらの公理的性質が厳密に満たされなければならないわけではないが、こうした研究は推薦システムの手法の選択指針について参考になるであろう。

*4 item churn ともいう。

9. 協調フィルタリング：モデルベース法

モデルベース法では、活動利用者に推薦をする前にモデルを構築する。1.1 節で述べたように、万能なモデルは存在しないので、多様なモデルが目的に合わせて提案されている。これらを順次紹介する。

9.1 クラスタモデル

嗜好パターンが類似している利用者が好むものを推薦するという手順を、直観的に実装したのがクラスタモデルである [Breese 98, Kamishima 03b]。クラスタとは対象の集合を分割した部分集合で、同じクラスタ内の対象は互いに似ているが、違うクラスタでは似ていないという条件を満たすものである。こうしたクラスタを得る手法をクラスタリングという [神畠 03a, 宮本 99, 齋藤 06]。この手法を用いて、嗜好パターンが類似している利用者のクラスタに、標本利用者の集合を分割する。ここで、利用者間の類似度は、8 章の式 (1) の Pearson 相関係数のように、いろいろなアイテムへの評価値がどれくらい類似しているかによって測る。活動利用者への推薦は、活動利用者と各クラスタとの類似度を調べ、最も似ているクラスタを見つける。そして、そのクラスタ中の標本利用者の平均評価値が高いアイテムから順に活動利用者へ推薦する。この方法には、利用者集合とアイテム集合を同時に分割する共クラスタリング (co-clustering) を使う改良 [George 05] などがある。

このモデルでは、利用者 DB を何個のクラスタに分けたかによって、推薦の質が大きく変わる。クラスタ数を小さく設定すると、おおまかであり個人化されていない推薦がなされる。そのため、cold-start 問題に対して比較的強い。しかし、推薦パターンの種類はたかだかクラスタ数に制限されるので、多数の嗜好データを集めても、ある程度以上に個人化された推薦はできない。一方、クラスタ数を多くすると、cold-start 問題に対して弱くなるが、推薦の個人化の度合いは高まる。また、クラスタ数が多すぎると安定したクラスタを求めるのが難しくなる問題もある。よって、目的に合わせてクラスタ数を調整する必要がある。

この方法には、実現が直観的なことに加え、モデルの構築も比較的高速であり、単純なので実装も容易であるといった利点がある。推薦時も、各クラスタと活動利用者の類似性を調べるだけなので、計算量はクラスタ数だけに比例し、高速である。欠点としては灰色の羊 (gray sheep) 問題 [Burke 02] がある。例えば、映画の推薦を考えた場合、特定のジャンルのものだけを鑑賞する利用者はむしろ少なく、サスペンスとホラーなど複数のジャンルの映画を見るであろう。クラスタモデルでは、利用者を特定のクラスタに分けてしまうため、こうした白でも黒でもない灰色の羊の利用者に適切な推薦ができない。

9.2 関数モデル

利用者が好きなものほど大きな値をとる効用関数 (utility function) や、評価値そのものを予測する関数を用いるモデルを、本稿ではまとめて関数モデルと呼ぶ。そして、これらのモデルの獲得を、回帰問題、クラス分類問題、および順序回帰問題に帰着させて解く。これらの手法を順次紹介する。

§1 回帰問題へ帰着させる方法

最初に、回帰問題に帰着させる手法から紹介する。最も簡単な線形関数の場合を考える。8.2 節の式 (3) は、詳細を無視すれば次のような線形モデルとみなせる。

$$\hat{s}_{aj} = \sum_k w_{jk} s_{ak} \quad (4)$$

8.2 節では、パラメータ w_{jk} はアイテム j と k の嗜好パターンの類似性を、相関係数など経験的に選んだ類似度で決めた。しかし、与えられた評価値の集合を訓練事例として、機械学習の手法を適用すれば、もっと予測精度の高い関数を獲得できるのではないかと。この考えに従い、これらのパラメータを決める問題を回帰 (regression) (当てはめ (fitting) ともいう) [Bishop 06, Bishop 08] という機械学習や統計的予測の問題に帰着させて解く。

まず、回帰モデルを、評価値行列 S 行列の分解として捉えてみよう。与えられた行列 S の未評価アイテムの部分は欠損しているが、欠損していない完全な評価値行列 S^* を考え、この行列を次のように分解する。

$$S^{*T} \approx Y^T X \quad (5)$$

ここで、 X は列数が n で、 Y は列数が m 、 X と Y の行数は等しく K である。 X の第 i 列ベクトル \mathbf{x}_i は利用者 i を、 Y の第 j 列ベクトル \mathbf{y}_j はアイテム j を表している。すると、利用者 i へのアイテム j への評価値は $s_{ij}^* \approx \mathbf{y}_j^T \mathbf{x}_i$ のような線形回帰モデルで表される。 \mathbf{x}_i の要素を説明変数と、 \mathbf{y}_j の要素をパラメータとみなせば、式 (4) と同等のモデルであり、逆に \mathbf{y}_j を説明変数にすれば、式 (2) と類似したモデルとも解釈できる。ここで、もし K が $\max\{m, n\}$ なら、近似ではなく厳密に $S^{*T} = Y^T X$ となるように分解できる。だがこれでは、観測データを書き写しただけで、嗜好パターンを要約したモデルとはいえない。また、実際に観測できる評価値行列は S^* ではなく、欠損のある S である。そこで、 K を $K \ll m, n$ に固定し、 S^{*T} との期待的な損失を最小化するように、 S^T を $Y^T X$ と分解することで、モデルを獲得する。

こうした分解をするには、次元縮約 (dimension reduction) や (非負) 行列分解 ((non-negative) matrix factorization) の手法を用いる。こうした手法の一つ [Canny 02b] を述べる。この方法では、残差 $S^T - Y^T X$ の要素が正規分布に従うとモデル化し、観測された評価値の生成確率が高くなるように X と Y を計算する。分解ができれば、任意の利用者の任意アイテムに対する推定評価値 $\hat{s}_{ij} = \mathbf{y}_j^T \mathbf{x}_i$ が計算できる。しかし、実際の S には欠損値があるので、[Canny 02b] では、これらの欠損値を潜在変数とみ

なして EM アルゴリズム [Dempster 77] を適用して解いている。欠損値に対処する方法は、潜在変数以外にも、全員が必ず評価するアイテム集合を利用する、欠損値を平均値などで補完する、欠損した要素は無視して損失を計算する [Bell 07, Weimer 08] といった方法もある。回帰モデルは単純なので、演算操作が複雑になるプライバシー保護協調フィルタリング (14.2 節) などへの適用や、カーネルを使った非線形回帰を導入して予測の向上を図るといった拡張が可能である。

次に [Aggarwal 99] の Horting という方法を紹介する。式 (4) のように全体を一つの線形モデルで表すと、大まかすぎて予測精度が低くなる場合がある。そこで、嗜好が類似している利用者の間で局所的に線形モデルを構築する。利用者 a について、共通に評価したアイテムが十分に多い利用者 i を見つける。さらに、これらの利用者 i の評価値 s_{ij} から $\hat{s}_{aj} = v_{ai}s_{ij} + b_{ai}$ の線形関数で、利用者 a のスコアを高精度で予測できる利用者を選び出す。このとき、利用者 i は利用者 a を予測可能という。なお、 v_{ai} や b_{ai} は、両者が共通に評価しているアイテムの評価値から計算可能である。次に、利用者をノードとし、 i から a への予測可能性を逆向きの a から i への有向辺で示すグラフを生成しておく。このグラフを用いて、利用者 a のアイテム j への評価値を推定する。利用者 a から直接リンクしている利用者で、アイテム j を評価している利用者の集合を \mathcal{X}_j' とする。この集合が空でなければアイテム j への予測評価値は、 \mathcal{X}_j' 中の利用者それぞれによる予測評価値の平均とする。

$$\hat{s}_{aj} = \frac{1}{|\mathcal{X}_j'|} \sum_{i \in \mathcal{X}_j'} v_{ai} s_{ij} + b_{ai}$$

もし、 \mathcal{X}_j' が空なら、リンクを 2 段たどり、アイテム j を評価している利用者集合を求め \mathcal{X}_j'' とする。この集合が空でなければ、次式で予測評価値を求める。

$$\hat{s}_{aj} = \frac{1}{|\mathcal{X}_j''|} \sum_{i \in \mathcal{X}_j''} v_{ai} (v_{ii} s_{ij} + b_{ii}) + b_{ai}$$

ただし、 i' は、各 i について i と a を中継する利用者である。さらに \mathcal{X}_j'' も空ならば、3 段以上のリンクを考慮する。

§2 クラス分類問題へ帰着させる方法

回帰と並ぶ代表的な機械学習の枠組みであるクラス分類 (classification) も利用できる。アイテム j への評価値 s_{aj} は、 \mathcal{A} 中の値の一つをとるので、クラスと考えることができる。また、分類するアイテム j の特徴量には、このアイテムへのほかの利用者による評価値や、活動利用者のほかのアイテムへの評価値が利用できる。あとは、これらの特徴量で表されたアイテムが分類されるべきクラス、すなわち評価値を予測するモデルをクラス分類の学習手法によって獲得すればよい。

このような方法として、逐次型二項関係学習法 (Cross-G-Learn-Relation) [Nakamura 98, 山西 03] を紹介する。これは、Winnow [Littlestone 88] のような手法で、各ク

ラスごとに、対象の分類されやすさを表す関数を獲得し、その関数の値が最大になるクラスに対象を分類する。利用者 i のアイテム j の評価値を予測するとき、逐次型二項関係学習法では、アイテム j へのほかの利用者の評価と、利用者 i のその他のアイテムへの評価を入力とした次の関数を用いる^{*5}。

$$\hat{s}_{ij} = \arg \max_{r \in \mathcal{A}} \left(\sum_{i': s_{ij'}=r} v_{ii'} + \sum_{j': s_{ij'}=r} w_{jj'} \right) \quad (6)$$

利用者 i が仮にアイテム j の評価値を r としたとき、第 1 項ではアイテム j についての評価が同じ r である利用者 i' について、利用者間重み $v_{ii'}$ の総和を求める。同様に、第 2 項では利用者 i が同じ r と評価をしているアイテム j' について、アイテム間重み $w_{jj'}$ の総和を求める。すなわち、類似利用者と類似アイテムの評価値の両方が考慮されている。重み $v_{ii'}$ と $w_{jj'}$ はオンライン学習 [丸岡 99] の枠組みで求める。これは、行列 S の形式で評価値がまとめて与えられてから学習するのではなく、誰かが何かアイテムを評価して、 s_{ij} が一つ与えられるごとに重みをより適切なものに更新する方法である。もう少し詳しく述べると、最初にすべての重み $v_{ii'}$ と $w_{jj'}$ を初期化する。その後、 s_{ij} が観測されるたびに、利用者 i 以外でアイテム j を評価済みのすべての利用者 i' について、 $s_{ij} = s_{ij'}$ なら重み $v_{ii'}$ を増やし、そうでないなら減らす。さらに、利用者 i が評価している j 以外のすべてのアイテム j' について、評価値が一致すればやはり重み $w_{jj'}$ を増やし、そうでなければ減らす。このように評価値が与えられるごとに重みを更新する。この方法はオンライン学習なので、モデルベースだが、アイテムや利用者が追加されても対応できる特徴がある。ただし、アイテムや利用者の削除には対応できない。

§3 順序回帰問題へ帰着させる方法

5.2 節では、採点法や格付け法による評価値は、本来は順序付きカテゴリとして扱うべきものであることを述べた。すなわち、評価値が 2 のアイテムは、4 のアイテムの半分程度好きなのではなく、評価値 3 のアイテムより、4 のアイテムのほうがより好きであるということだけを示している。こうした順序付きカテゴリ値を予測する問題は順序回帰と呼ばれる。この順序回帰を、ブースティング (boosting) [Freund 96, フロインド 99] の枠組みで扱う RankBoost [Freund 98, Freund 03] で推薦をする研究などがある。後に、この順序回帰を用いた、内容ベースと協調フィルタリングのハイブリッド法を 12.1 節で紹介する。

9.3 確率モデル

モデル化に用いた関数が確率分布として解釈できるものを本稿では特に確率モデルと呼ぶ。これら確率モデル

*5 原論文では評価の揺らぎを許す工夫がさらになされているが、ここでは簡潔に記す。

を、大きく履歴条件型と共起型の二つに分け、順に説明する。

§1 履歴条件型

履歴条件型の確率モデルでは、活動利用者のアイテム j への評価値を次の条件付き期待値 [Breese 98] で予測する。

$$\begin{aligned}\hat{s}_{aj} &= E[s_{aj} | s_{ak} \text{ s.t. } k \in \mathcal{G}_a] \\ &= \sum_{r \in \mathcal{R}} r \Pr[s_{aj} = r | s_{ak} \text{ s.t. } k \in \mathcal{G}_a]\end{aligned}\quad (7)$$

これは、活動利用者の過去の嗜好データが与えられたときの、活動利用者のアイテム j への評価値の条件付き期待値である。実際には、式 (7) 中の確率分布は未知なので、推定した関数を利用する。式中ではほかの標本利用者の嗜好データが参照されていないが、確率分布の推定の過程で、これらの嗜好データを利用するため、協調フィルタリングによる推薦といえる。また、適合/不適合の 2 値の嗜好データ、すなわち $\mathcal{R} = \{0, 1\}$ の場合、式 (7) は単なる条件付き確率分布となる。

$$\hat{s}_{aj} = \Pr[s_{aj} = 1 | s_{ak} \text{ s.t. } k \in \mathcal{G}_a] \quad (8)$$

このように簡潔になるので、予測評価値を提示する評価閲覧よりも、適合か不適合かの判定ができればよい適合アイテム発見タスクに適す。この履歴条件型モデルでは、条件付き確率 (8) の計算のため、 $s_{ij} = 1$ と $s_{ij} = 0$ の両方の場合の訓練事例が必要である。しかし、5 章で述べたように、暗黙的な評価では、購入や閲覧などの行動がなかったことで、否定的な評価とみなす。すると、未評価と否定的評価は混同され、 $s_{ij} = 0$ となる事例にはノイズが多くなる。そのため、 $s_{ij} = 0$ となる事例が必要な、この履歴条件型ではこうした暗黙的な評価では不利になる。

式 (7) 中の条件付き確率は、 \mathcal{G}_a に依存するため、各利用者について個々にモデルを獲得する必要性が生じ、実用的ではない。そこで実際には、利用者には依存しない、次の全アイテムへの評価の同時分布を求める。

$$\Pr[s_k \text{ s.t. } k \in \mathcal{G}] \quad (9)$$

そして、式 (7) 中の条件付き分布は、ベイズ則と不要な変数を周辺化で消去することで計算する。だが、各 s_k は $|\mathcal{R}|$ 個の値をとることが可能で $|\mathcal{G}| = m$ なので、この分布の表形式の飽和モデルのパラメータ数は $|\mathcal{R}|^m - 1$ となる。しかし、データは疎で m は大きいので、単純に評価値の各組合せの頻度を数え上げて、この飽和モデルのパラメータを推定することは現実的ではない。そこで、このような場合の一般的な対策を導入し、変数 s_k の間に部分的な条件付き独立性を仮定し学習すべきパラメータ数を減らす。このような条件付き独立性を導入した確率モデルは一般にグラフィカルモデルと呼ばれる。その一つであるベイジアンネット [繁樹 06] は [Breese 98] で利用されている。ベイジアンネット全般の問題として独立性の構造をうまく設計する難しさがあるが、うまく設計できれば学習も、推薦も効率的に実行できる。また、式 (7) の条件付き確率を直接内部構造としてもつ確率モデルである dependency ネットを使う [Heckerman

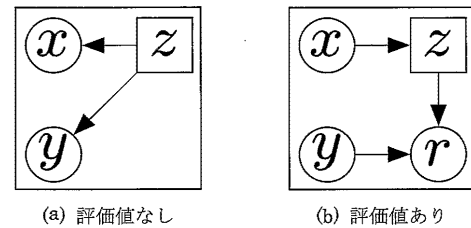


図 5 共起型確率モデル [Hofmann 99b]

00] などもある。このモデルでは、同時確率から条件付き確率へ変換が不要なので、実行時の推薦は高速に実行できるが、事前のモデル学習の計算量は多い。

§2 共起型

もう一方の共起型の確率モデルについて述べる。このモデルでは、ある利用者が、あるアイテムを評価したことを、その利用者が i であるという事象と、そのアイテムが j であるという事象が共起していると捉える。そして、この共起確率をモデル化する。このモデル化には、自然言語処理で文書と語の共起確率を表すために考案された probabilistic latent semantic analysis (pLSA) [Hofmann 99a] を利用する。この確率モデルは非常に柔軟で、純粋な協調フィルタリングだけでなく、評価の揺らぎ、アイテムの特徴、コンテキスト、デモグラフィックな要因を導入するといった拡張が可能である。

本節では、利用者の嗜好データのみに基づく協調フィルタリングを対象としたモデル [Hofmann 99b, Hofmann 04] を紹介する。最初に、利用者とアイテムの共起関係だけを使うモデルについて述べる。利用者とアイテムをそれぞれ多値の確率変数 $x \in \mathcal{X}$ と $y \in \mathcal{Y}$ で表す。嗜好データを、利用者 i がアイテム j を購入するという行為で暗黙的に獲得する場合を考える。これは、 $x = i$ という事象と、 $y = j$ という事象が共起していることに相当する。また、 $\mathcal{R} = \{0, 1\}$ であるとき、 $s_{ij} = 1$ となる対 (i, j) が観測されることもとも等価である。履歴条件型の確率モデルでは $s_{ij} = 0$ となる事例も必要であったが、共起型では $s_{ij} = 1$ となる事例だけでモデルが学習できる。このため、未評価と不支持が区別できず $s_{ij} = 0$ の事例にはノイズが多くなる暗黙的な評価を採用した場合に、この共起型確率モデルは便利である。

さらに、このモデルで重要な役割を果たす潜在変数 z を導入する。これも多値変数で $\mathcal{Z} = \{1, \dots, K\}$ 中の値をとり、潜在的な嗜好のパターンを表す。これはクラスタモデルと類似しているため、灰色の羊問題 (9.1 節) が生じるように思える。だが、 K 個のパターンのうちどれか一つに割り当ててのではなく、それらが混ざり合った状態を考慮するので、この問題は生じない。そして、この潜在変数を用いて、 x と y の同時確率を、次式で表すのが pLSA モデルである。

$$\Pr[x, y] = \sum_{z \in \mathcal{Z}} \Pr[x|z] \Pr[y|z] \Pr[z] \quad (10)$$

このモデルでは、図 5 (a) のように、 z が与えられたと

きに x と y が条件付き独立であることを仮定して、モデルのパラメータの総数を減らしている。また、 $\Pr[x|z]$, $\Pr[y|z]$, $\Pr[z]$ はそれぞれ多項分布に従う。モデルのパラメータは $\theta = (\{\Pr[x|z]\}, \{\Pr[y|z]\}, \{\Pr[z]\})$ であるが、これらを最尤推定で求める。すなわち、いろいろな利用者がいろいろなアイテムを評価した、 N 個の共起データ $\mathcal{D} = \{(x_k, y_k)\}_{k=1}^N$ に対する次の対数尤度を最大にするように求める。

$$\mathcal{L}(\mathcal{D}; \theta) = \sum_{(i,j) \in \mathcal{D}} \ln \Pr[x=i, y=j; \theta]$$

潜在変数 z があるため、この最尤推定は EM アルゴリズム [Bishop 08, Dempster 77] によって行う。具体的には次の二つのステップを交互に反復する。一つ目のステップでは、パラメータ $\{\Pr[z|x]\}, \{\Pr[y|z]\}, \{\Pr[z]\}$ が与えられたときの、潜在変数の分布を求める。

$$\Pr[z|x, y] = \frac{\Pr[z] \Pr[x|z] \Pr[y|z]}{\sum_z \Pr[z] \Pr[x|z] \Pr[y|z]} \quad (11)$$

二つ目のステップでは前ステップで求めた分布 $\Pr[z|x, y]$ を用いてパラメータを更新する。

$$\Pr[x|z] = \frac{\sum_y n(x, y) \Pr[z|x, y]}{\sum_{x', y} n(x', y) \Pr[z|x', y]} \quad (12)$$

$$\Pr[y|z] = \frac{\sum_x n(x, y) \Pr[z|x, y]}{\sum_{x, y'} n(x, y') \Pr[z|x, y']} \quad (13)$$

$$\Pr[z] = \frac{\sum_{x, y} n(x, y) \Pr[z|x, y]}{N} \quad (14)$$

ただし、 $n(x, y)$ は、 x と y が共起している \mathcal{D} 中の対の数である。以上の手続きを収束するまで反復すると、パラメータが計算できる。そして、次式の $\Pr[y|x]$ を求めておく

$$\Pr[y|x] = \frac{\sum_z \Pr[x|z] \Pr[y|z] \Pr[z]}{\sum_{y', z} \Pr[x|z] \Pr[y'|z] \Pr[z]} \quad (15)$$

利用者 i へのアイテム j の適合性は $\Pr[y=j|x=i]$ の大きさで測る。利用者 i に対しては、この適合性を最大化する次のアイテム y^* を推薦すればよい。

$$j^* = \arg \max_{j \in \mathcal{Y}} \Pr[y=j|x=i] \quad (16)$$

次に、利用者 x とアイテム y に、評価値を表す確率変数 r も加えた拡張を考える。この r は、 \mathcal{D} 中の値をとる多値変数である。このモデル化では $x=i, y=j$, および $r=s$ の共起確率 $\Pr[x, y, r]$ を考える。そして、利用者 $x=i$ のアイテム $y=j$ の推定評価値 \hat{s}_{ij} は最頻値

$$\hat{s}_{ij} = \arg \max_s \Pr[r=s|x=i, y=j]$$

や期待値

$$\hat{s}_{ij} = \sum_{s \in R} s \Pr[r=s|x=i, y=j]$$

で計算する。なお、 $\Pr[r|x, y]$ は $\Pr[x, y, r]$ から計算できる。図 5 (b) は $\Pr[x, y, r]$ のモデル化の一例である。このモデルでは、利用者 x は z で表されるグループに分類され、そのグループ z とアイテム y に依存して評価値 r が決まると解釈できる。モデルを決めれば、あとは対数尤度を最大化するパラメータを EM アルゴリズムで求めればよい。[Hofmann 99b] ではほかに 4 種類のモデルをあげている。どのモデルが適切かだが、一般には 3.1 節で述べた交差確認法で、予測精度を最大にするようなものを選ぶ。

この共起型確率モデルは自由に設計できる余地が多く、いろいろな拡張が可能である。5.2 節では、利用者の評価値に一貫性はなく揺らぎがあると述べた。こうした揺らぎを考慮する 2 種類の方法を [Jin 03] では提案している。図 6 (a) の decoupled は、まず揺らぎのない真の評価値はわからないので、これを潜在変数 z_{pref} で表す。この真の評価値は、同じ嗜好の利用者のグループを示す z_p と、類似したアイテムのグループを表す z_y に依存して決まる。さらに、評価値がどのように揺らぐのかというパターンに利用者は分けられ、そのパターンを潜在変数 z_r で表す。観測される評価値 r は真の嗜好 z_{pref} と揺らぎのパターン z_r に依存して決まる。推薦は、予測した真の評価値 z_{pref} によって行う。もう一つの preferred ordering モデル (図 6 (b)) は、評価値を順序尺度と考える方法である。同じ利用者 x が二つのアイテム y と y' のそれぞれに r と r' の評価値を与えた場合を考える。このとき、 r と r' の順序関係だけを取り出す関数 $I(r, r')$ を導入し、この関数 I が、利用者とアイテムのグループを表す潜在変数 z_x と z_y に依存するとモデル化する。このようにモデルを決めれば、あとは EM アルゴリズムでパラメータを学習できる。これらのモデルでは揺らぎを扱うことができるが、パラメータの総数は増えるので、それに応じた十分な嗜好データが必要になる。

ほかの拡張としては、pLSA を、ベイズの枠組みに拡張した latent Dirichlet allocation (LDA) [Blei 03] もある。これは、多項分布のパラメータの事前分布として共役な Dirichlet 分布を用いるものである。これにより、利用者のモデルが訓練事例集合に過学習したものとなることがなくなり、新規利用者への予測性能が向上する。しかし、変分ベイズや MCMC を用いた近似計算が必要になり、pLSA より計算量は増大する。この LDA と類似したモデルとして [Marlin 04] などもある。

9.4 時系列モデル

時間に伴う変化を嗜好の予測に利用する方法を紹介する。例えば、レンタルビデオ店で、先々週はあるドラマ

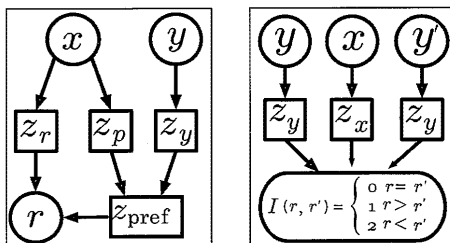


図 6 評価値の揺らぎを考慮したモデル [Jin 03]

の第 1 話を、先週はこのドラマの第 2 話を借りていった顧客がいたとする。このとき、今週はこのドラマの第 3 話を借りるだろうことは容易に予測できる。5・4 節で述べたように、標本利用者の時間順の行動や評価の履歴情報を、利用者 DB に蓄積している場合がある。この情報から、利用者の行動の時間的推移を表すモデルを構築し、活動利用者の現在に至る行動や評価とこのモデルから、嗜好を予測する手法を紹介する。

§ 1 maxent モデル

この購入履歴に基づくモデルを利用した [Pavlov 03] の方法を紹介します。利用者が t 回目購入したアイテムを $y^{(t)}$ で表すと、購入履歴の系列は $\langle Y^{(t)} \rangle = \langle y^{(1)}, y^{(2)}, \dots, y^{(n)} \rangle$ となる。例えば、上記のドラマの例であれば $y^{(1)} \sim y^{(3)}$ は、それぞれドラマの第 1 ～ 3 回に相当する。この購入履歴が与えられたときの、次に購入するアイテムの条件付き確率 $\Pr[y^{(t+1)} | \langle Y^{(t)} \rangle]$ を求める。そして、この確率を最大にするアイテムを推薦する手法を考える。これは、標本利用者の購入履歴を数え上げれば原理的には計算可能である。しかし、3・2 節で述べたようにデータが疎であるので、至るところで確率が 0 となり、実際にはこの方法では計算できない。そこで、自然言語処理の言語モデルでよく利用される素性関数を用いた maxent モデルを導入する。

$$\Pr[y^{(t+1)} | \langle Y^{(t)} \rangle] = \frac{1}{Z} \exp \left[\sum_{k=1}^K \lambda_k f_k(y^{(t+1)}, \langle Y^{(t)} \rangle) \right] \quad (17)$$

ただし、 Z は正規化定数、 λ_k は重みパラメータである。ここで、素性関数 $f_k(y^{(t+1)}, \langle Y^{(t)} \rangle)$ が重要になる。素性関数は購入履歴 $\langle Y^{(t)} \rangle$ にある特徴があるとき、次回に購入するアイテムがある特定のものになるなら 1 をとり、それ以外では 0 となるような関数である。例えば、上記のドラマの場合、購入履歴にドラマの第 2 回が含まれていて、次回購入がドラマの第 3 回なら 1 になる関数である。このような素性関数を trigger と呼ぶ。この trigger には、購入履歴中にアイテム j があるときに、次回にアイテム j' を購入する確率 $\Pr[y^{(t+1)} = j' | j \in \langle Y^{(t)} \rangle]$ と $\Pr[y^{(t+1)} = j']$ との差が大きなものを選ぶ。こうして素性関数を選択すれば、式 (17) のモデルのパラメータ

λ_k は最大エントロピー原理 [北 99] に基づいて推定できる。素性関数には、trigger 以外にも、いろいろなものが利用できる。次に購入するアイテムが、 $s \in 0, 1, 2, \dots$ 個前に購入したアイテムに依存して決まる gap マルコフモデルを素性として利用する方法 [岩田 07]。利用者自身のデモグラフィックな特徴を考慮するのも有効であろう。系列パターンマイニング [Han 06, Pei 01] などと組み合わせれば、trigger よりも複雑なパターンも素性関数として利用できるだろう。

§ 2 マルコフ決定過程モデル

前項のモデルでは、過去の履歴に基づいて、次に利用者が選択する確率が最も高いアイテムを推薦する。だが、推薦がなければそのアイテムを買ったかもしれないが、推薦の影響によってそのアイテムを選ばないという場合もある。加えて、1 年間全体で見て、選択したアイテムの価格の総和を最大化するとか、選択したアイテムへの利用者の評価の総和を最大化するような推薦をしたいとする。これらの目的を達成するため、マルコフ決定過程を使う手法が提案されている [Shani 02, Shani 05]。

マルコフ決定過程の前にマルコフ過程について述べる。この過程では、次に選択するアイテム $y^{(t+1)}$ は、直前の K 個のアイテム系列 $\langle Y^{(t)} \rangle = \langle y^{(t-K+1)}, \dots, y^{(t)} \rangle$ に依存する。このアイテム系列で状態を表し、状態の遷移確率は $\text{tr}_{\text{MC}}(\langle Y^{(t)} \rangle, \langle Y^{(t+1)} \rangle)$ と記す。この確率は、十分なデータがあれば、利用者の購買履歴を数え上げれば計算できる。しかし、データは疎なので、[Shani 02, Shani 05] では skipping や、アイテムのクラスタリングなどの手法を用いて、近似的に求めている。次にマルコフ決定過程を図 7 を用いて説明する。図中①は現在の状態 $\langle Y^{(t)} \rangle = \langle y^{(t-1)} = A, y^{(t)} = B \rangle$ で、この時刻 t にアイテム D を購入して、図中④の次の状態 $\langle Y^{(t+1)} \rangle = \langle y^{(t)} = B, y^{(t+1)} = D \rangle$ に移る。ここで、単に $\langle Y^{(t)} \rangle$ のみに依存して $\langle Y^{(t+1)} \rangle$ が決まる (図の点線矢印) ならマルコフ過程である。一方、マルコフ決定過程では、状態 $\langle Y^{(t)} \rangle$ で行う推薦という行動 (action) を明示的に考慮する。これにより、推薦という行動によって、利用者の選択が変わることをモデル化できる。この時刻 t での行動 $A^{(t)}$ は図中②にあたり、アイテム C を推薦した。利用者は $\langle Y^{(t)} \rangle$ と $A^{(t)}$ に依存して、アイテム $y^{(t+1)}$ を選択し、次の状態 $\langle Y^{(t+1)} \rangle$ に遷移する。図の例では、推薦を受け入れず利用者は③でアイテム D を購入したので、遷移した状態は④の $\langle Y^{(t+1)} \rangle = \langle B, D \rangle$ となった。この状態遷移を $\langle Y^{(t)} \rangle, A^{(t)}, \langle Y^{(t+1)} \rangle$ と、この遷移確率を $\text{tr}_{\text{MDP}}(\langle Y^{(t)} \rangle, A^{(t)}, \langle Y^{(t+1)} \rangle)$ と記す。また、状態 $\langle Y^{(t)} \rangle$ で行動 $A^{(t)}$ を決定する手続きを政策 (policy) と呼ぶ。さらに、図の③にてアイテム D を選択したが、これに依存して報酬 (reward) が発生する。この報酬は、アイテムへの代金や、利用者のアイテムへの評価値などである。報酬の導入により、個別の推薦ごとの効用を最大化するのではなく、複数の推薦を含む長期間の効用の最大化を考慮でき

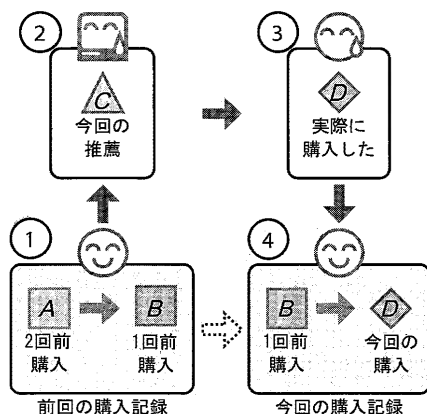


図 7 マルコフ決定過程によるモデル

ようになる。この報酬を最大化するような政策を決める学習問題は強化学習 [Sutton 98, Sutton 00] と呼ばれる。既存の強化学習の方法を用いて、適切な政策が獲得できれば、推薦は政策に従うことで実行可能になる。ただし、データが疎である問題に対処するため、マルコフ決定過程の遷移確率 $\text{tr}_{\text{MDP}}(\langle Y^{(t)} \rangle, A^{(t)}, \langle Y^{(t+1)} \rangle)$ を、マルコフ過程の遷移確率 $\text{tr}_{\text{MC}}(\langle Y^{(t)} \rangle, \langle Y^{(t+1)} \rangle)$ によって初期化するなどのヒューリスティックを用いている。

§3 サブスクリプションサービス

時系列を扱う手法の最後にサブスクリプションと呼ばれるサービスを対象とした手法 [Iwata 06] を紹介する。サブスクリプションとは音楽配信などのサービスの販売形式の一つで、契約すると解約するまでの期間、任意の曲を任意の回数試聴できる。そのため、個々の購入での満足ではなく、サービス全体への顧客満足度を向上させて、解約を防ぐことが目標になる。上記の強化学習の枠組みでは、各購入行動ごとに時間 t が経過する。一方、サブスクリプションでは、各行動が何回目であるかを示す t に加え、実時間とともに推移する契約期間 T もモデル化する必要がある。

利用者 x に t 回目にアイテム $y^{(t)}$ を推薦したとき、現在の契約期間中に解約しない事象 e が生じる確率を最大化するようなアイテムを推薦する。

$$\hat{y}^{(t)} = \arg \max_{y^{(t)} \in \mathcal{Y}} \Pr[e|x, y^{(t)}]$$

推薦の結果、実際に購入したアイテムを $y'^{(t)}$ で表し、この確率を次のように分解する。

$$\Pr[e|x, y^{(t)}] = \sum_{y'^{(t)} \in \mathcal{Y}} \Pr[e|x, y'^{(t)}] \Pr[y'^{(t)}|x, y^{(t)}]$$

利用者 x がアイテム $y'^{(t)}$ を選択したとき解約しない確率が $\Pr[e|x, y'^{(t)}]$ で、 $y^{(t)}$ を推薦したときに $y'^{(t)}$ を選択する確率が $\Pr[y'^{(t)}|x, y^{(t)}]$ である。前者の確率は、ハザード (hazard) 関数 $h(T|\langle Y^{(t)} \rangle)$ で表す。これは購入履歴が $\langle Y^{(t)} \rangle = \langle y^{(1)}, \dots, y^{(t)} \rangle$ である利用者のうち、期間 T まで契約を継続し、かつ、この期間で契約をやめる割合を示す。この関数は生存時間分析の Cox 比例ハザードモデルでモデル化する。 $\langle Y^{(t-1)} \rangle$ のあとに $y'^{(t)}$ を購入した系列を $\langle Y^{(t-1)}, y'^{(t)} \rangle$ で表すと $\Pr[e|x, y'^{(t)}]$ は、適当な仮定のもと、次式となる。

$$\Pr[e|x, y'^{(t)}] = \frac{h(T|\langle Y^{(t-1)} \rangle)}{h(T|\langle Y^{(t-1)} \rangle) + h(T|\langle Y^{(t-1)}, y'^{(t)} \rangle)}$$

一方の $\Pr[y'^{(t)}|x, y^{(t)}]$ は、アイテムを推薦することにより選択確率が定数倍されると仮定して、利用者 x の選択確率 $\Pr[y'^{(t)}|x]$ から求める。この $\Pr[y'^{(t)}|x]$ は、購入履歴に基づく素性関数を用いた maxent モデルでモデル化する。このようにモデル化すれば、あとは購入履歴からパラメータを学習すればよい。

10. 協調フィルタリング：ハイブリッド

協調フィルタリングはメモリとモデルベースとに分類できると述べたが、これらの中間的な手法も存在する。

personality diagnosis 法 [Pennock 00a] は、メモリベースのように利用者間の類似度を用いる。そして、この類似度を重みとした、各利用者の評価値の重み付き平均を推定評価値とする。これだけであれば、メモリベース法だが、類似度計算にモデルベースの要素がある。各利用者個人モデルを作成し、活動利用者と標準利用者の個人モデルが一致する確率を類似度としている。

クラスタリングをしたあとに、メモリベース法を適用する方法は、メモリベース法に、モデルベースの要素を加えた手法と解釈できる。クラスタリングする対象が、利用者集合のもの [Kamishima 03b, Xue 05] と、アイテム集合のもの [O'Connor 99] とがある。前者の方法では、アイテムへの嗜好パターンが類似した利用者のクラスタを生成する。活動利用者と嗜好パターンが近いクラスタを見つけ、そのクラスタ内の利用者を対象に利用者間型メモリベース法を適用する。利用者間型メモリベース法で近傍利用者を利用した場合と同様に、推薦の高速化と、場合によっては予測精度の向上が見込めるが、適応性の面では不利になる。後者のアイテム集合をクラスタリングする方法では、8.2 節のように、 S の列ベクトルの類似性に基づいてクラスタをつくる。その後、各クラスタごとに個別にメモリベース手法を適用する。これは、特定のアイテムカテゴリー内に推薦対象を限定すると、そのアイテム群に特に関心のある利用者に対する予測精度が向上するとの考えに基づくが、実験ではその有効性は確認できなかったと報告している。

今回は推薦システムの実行過程の二つ目である「嗜好の予測」段階の全般的な話題と、協調フィルタリングによる各種の予測アルゴリズムを紹介した。2008 年 3 月号の次回では、「嗜好の予測」段階の残りとして、最後の「推薦の提示」段階について述べる。その後、その他の問題や視点、参考資料、推薦システムの今後の展開についてまとめる。

◇ 参考文献 ◇

- [Aggarwal 99] Aggarwal, C. C., Wolf, J. L., Wu, K.-L. and Yu, P. S.: Horting hatches an egg: A new graph-theoretic approach to collaborative filtering, *Proc. 5th Int'l Conf. on Knowledge Discovery and Data Mining*, pp. 201-212 (1999)
- [Ali04] Ali, K. and van Stam, W.: TiVo: Making Show recommendations using a distributed collaborative filtering Architecture, *Proc. 10th Int'l Conf. on Knowledge Discovery and Data Mining*, pp. 394-401 (2004)
- [Arrow 63] Arrow, K. J.: *Social Choice and Individual Values*, Yale University Press, second edition (1963)
- [Baeza-Yates 99] Baeza-Yates, R. A. and Ribeiro-Neto, B.: *Modern*

- Information Retrieval*, Addison Wesley (1999)
- [Balabanović 97] Balabanović, M. and Shoham, Y.: Fab: Content-based, collaborative recommendation, *Commun. ACM*, Vol. 40, No. 3, pp. 66-72 (1997)
- [Bell 07] Bell, R. M., Koren, Y. and Volinsky, C.: Modeling relationships at multiple scales to improve accuracy of large recommender systems, *Proc. 13th Int'l Conf. on Knowledge Discovery and Data Mining*, pp. 95-104 (2007)
- [Bishop 06] Bishop, C. M.: *Pattern Recognition and Machine Learning*, Springer (2006)
- [Bishop 08] Bishop, C. M., 元田 浩ほか訳: パターン認識と機械学習 (上下), ベイズ理論による統計的予測, シュブリンガー・ジャパン (2007-2008)
- [Blei 03] Blei, D. M., Ng, A. Y. and Jordan, M. I.: Latent Dirichlet Allocation, *Journal of Machine Learning Research*, Vol. 3, pp. 993-1022 (2003)
- [Breese 98] Breese, J. S., Heckerman, D. and Kadie, C.: Empirical analysis of predictive algorithms for collaborative filtering, *Uncertainty in Artificial Intelligence*, Vol. 14, pp. 43-52 (1998)
- [Burke 02] Burke, R.: Hybrid recommender systems: survey and experiments, *User-Modeling and User-Adapted Interactions*, Vol. 12, No. 4, pp. 331-370 (2002)
- [Canny 02a] Canny, J.: Collaborative filtering with privacy, *Proc. 2002 IEEE Symp. on Security and Privacy*, pp. 45-57 (2002)
- [Canny 02b] Canny, J.: Collaborative Filtering with Privacy via Factor Analysis, *Proc. 25th Annual ACM SIGIR Conf. on Research and Development in Information Retrieval*, pp. 238-245 (2002)
- [Das 07] Das, A., Datar, M., Garg, A. and Rajaram, S.: Google news personalization: Scalable online collaborative filtering, *Proc. 16th Int'l Conf. on World Wide Web*, pp. 271-280 (2007)
- [Dempster 77] Dempster, A. P., Laird, N. M. and Rubin, D. B.: Maximum likelihood from incomplete data via the em algorithm, *Journal of The Royal Statistical Society (B)*, Vol. 39, No. 1, pp. 1-38 (1977)
- [Freund 96] Freund, Y. and Schapire, R. E.: Experiments with a new boosting algorithm, *Proc. 13th Int'l Conf. on Machine Learning*, pp. 148-156 (1996)
- [Freund 98] Freund, Y., Iyer, R., Schapire, R. E. and Singer, Y.: An efficient boosting algorithm for combining preferences, *Proc. 15th Int'l Conf. on Machine Learning*, pp. 170-178 (1998)
- [フロインド 99] フロインド Y., シャピリ R., 阿部直樹: ブースティング入門, 人工知能学会誌, Vol. 14, No. 5, pp. 771-780 (1999)
- [Freund 03] Freund, Y., Iyer, R., Schapire, R. E. and Singer, Y.: An efficient boosting algorithm for combining preferences, *Journal of Machine Learning Research*, Vol. 4, pp. 933-969 (2003)
- [George 05] George, T. and Merugu, S.: A scalable collaborative filtering framework based on co-clustering, *Proc. 5th IEEE Int'l Conf. on Data Mining*, pp. 625-628 (2005)
- [Han 06] Han, J. and Kamber, M.: *Data Mining: Concepts and Techniques*, Morgan Kaufmann, second edition (2006)
- [Häubl 01] Häubl, G. and Murray, K. B.: Recommending or Persuading? the impact of a shopping agent's algorithm on user behavior, *ACM Conf. on Electronic Commerce*, pp. 163-170 (2001)
- [Heckerman 00] Heckerman, D., Chickering, D. M., Meek, C., Rounthwaite, R. and Kardia, C.: Dependency networks for inference, collaborative filtering, and data visualization, *Journal of Machine Learning Research*, Vol. 1, pp. 49-75 (2000)
- [Herlocker 99] Herlocker, J. L., Konstan, J. A., Borchers, A. and Riedl, J.: An algorithmic framework for performing collaborative filtering, *Proc. 22nd Annual ACM SIGIR Conf. on Research and Development in Information Retrieval*, pp. 230-237 (1999)
- [Hofmann 99a] Hofmann, T.: Probabilistic latent semantic analysis, *Uncertainty in Artificial Intelligence*, Vol. 15, pp. 289-296 (1999)
- [Hofmann 99b] Hofmann, T. and Puzicha, J.: Latent class models for collaborative filtering, *Proc. 16th Int'l Joint Conf. on Artificial Intelligence*, pp. 688-693 (1999)
- [Hofmann 04] Hofmann, T.: Latent semantic models for collaborative filtering, *ACM Trans. on Information Systems*, Vol. 22, No. 1, pp. 89-115 (2004)
- [石井 98] 石井健一郎, 上田修功, 前田英作, 村瀬 洋: わかりやすいパターン認識, オーム社 (1998)
- [Iwata 06] Iwata, T., Saito, K. and Yamada, T.: Recommendation method for extending subscription periods, *Proc. 12th Int'l Conf. on Knowledge Discovery and Data Mining*, pp. 574-579 (2006)
- [岩田 07] 岩田具治, 山田武士, 上田修功: 購買順序を考慮した協調フィルタリング, 信学技報, AI 2007.3 (2007)
- [Jin 03] Jin, R., Si, L. and Zhai, C.-X.: Preference-based graphic models for collaborative filtering, *Uncertainty in Artificial Intelligence*, Vol. 19, pp. 329-336 (2003)
- [神鷹 03a] 神鷹敏弘: データマイニング分野のクラスタリング手法 (1) — クラスタリングを使ってみよう! —, 人工知能学会誌, Vol. 18, No. 1, pp. 59-65 (2003)
- [Kamishima 03b] Kamishima, T.: Nantonac collaborative filtering: Recommendation based on order responses, *Proc. 9th Int'l Conf. on Knowledge Discovery and Data Mining*, pp. 583-588 (2003)
- [Kamishima 06] Kamishima, T. and Akaho, S.: Nantonac collaborative filtering - Recommendation based on multiple order responses, *Proc. Int'l Workshop on Data-Mining and Statistical Science*, pp. 117-124 (2006)
- [北 99] 北 研二: 確率的言語モデル, 言語と計算, 第 4 巻, 東京大学出版会 (1999)
- [Linden 03] Linden, G., Smith, B. and York, J.: Amazon.com recommendations: item-to-item collaborative filtering, *IEEE Internet Computing*, Vol. 7, No. 1 (2003)
- [Littlestone 88] Littlestone, N.: Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm, *Machine Learning*, Vol. 2, pp. 285-318 (1988)
- [Marlin 04] Marlin, B. and Zemel, R. S.: The multiple multiplicative factor model for collaborative filtering, *Proc. 21st Int'l Conf. on Machine Learning*, pp. 576-583 (2004)
- [丸岡 99] 丸岡 章, 瀧本英二: オンライン予測, 人工知能学会誌, Vol. 14, No. 5, pp. 763-770 (1999)
- [McLaughlin 04] McLaughlin, M. R. and Herlocker, J. L.: A collaborative filtering algorithm and evaluation metric that accurately model the user experience, *Proc. 27th Annual ACM SIGIR Conf. on Research and Development in Information Retrieval*, pp. 329-336 (2004)
- [McNee 06] McNee, S. M., Riedl, J. and Konstan, J. A.: Accurate is not always good: How accuracies have hurt recommender systems, *Proc. SIGCHI Conf. on Human Factors in Computing Systems*, pp. 1097-1101 (2006)
- [宮本 99] 宮本定明: クラスタ分析入門. ファジィクラスタリングの理論と応用, 森北出版 (1999)
- [MovieLens] MovieLens, <http://movielens.org/>
- [Nakamura 98] Nakamura, A. and Abe, N.: Collaborative filtering using weighted majority prediction algorithms, *Proc. 15th Int'l Conf. on Machine Learning*, pp. 395-403 (1998)
- [O'Connor 99] O'Connor, M. and Herlocker, J.: Clustering items for collaborative filtering, *ACM SIGIR Workshop on Recommender Systems: Algorithms and Evaluation* (1999)
- [Pavlov 03] Pavlov, D. Y. and Pennock, D. M.: A maximum entropy approach to collaborative filtering in dynamic, sparse, high-dimensional domains, *Advances in Neural Information Processing Systems*, Vol. 15, pp. 1465-1472 (2003)
- [Pei 01] Pei, J., Han, J., Mortazavi-Asl, B., Pinto, H., Chen, Q., Dayal, U. and Hsu, M.-C.: PrefixSpan: Mining sequential patterns efficiently by prefix-projected pattern growth, *Proc. 17th Int'l Conf. on Data Engineering*, pp. 215-224 (2001)
- [Pennock 00a] Pennock, D. M., Horvitz, E., Lawrence, S. and Lee Giles, C.: Collaborative filtering by personality diagnosis: A hybrid memory-and model-based approach, *Uncertainty in Artificial Intelligence*, Vol. 16, pp. 473-480 (2000)
- [Pennock 00b] Pennock, D. M., Horvitz, E. and Lee Giles, C.: Social choice theory and recommender systems: Analysis of the axiomatic foundations of collaborative filtering, *Proc. 17th National Conf. on Artificial Intelligence*, pp. 729-734 (2000)

- [Resnick 94] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. and Riedl, J.: GroupLens: An open architecture for collaborative filtering of netnews, *Proc. Conf. on Computer Supported Co-operative Work*, pp. 175-186 (1994)
- [齋藤 06] 齋藤堯幸, 宿久 洋: 関連性データの解析法, 多次元尺度構成法とクラスター分析法, 共立出版 (2006)
- [Sarwar 01] Sarwar, B., Karypis, G., Konstan, J. and Riedl, J.: Item-based collaborative filtering recommendation algorithms, *Proc. 10th Int'l Conf. on World Wide Web*, pp. 285-295 (2001)
- [Shani 02] Shani, G., Brafman, R. I. and Heckerman, D.: An MDP-based recommender system, *Uncertainty in Artificial Intelligence*, Vol. 18, pp. 453-460 (2002)
- [Shani 05] Shani, G., Heckerman, D. and Brafman, R. I.: An MDP-based recommender system, *Journal of Machine Learning Research*, Vol. 6, pp. 1265-1295 (2005)
- [繁樹 06] 繁樹算男, 本村陽一, 植野真臣: ペイジアンネットワーク概説, 培風館 (2006)
- [Sutton 98] Sutton, R. S. and Barto, A. G.: *Reinforcement Learning: An Introduction*, MIT Press (1998)
- [Sutton 00] Sutton, R. S. and Barto, A. G.: 強化学習, 森北出版 (2000), (三上貞芳, 皆川雅章 訳)
- [Weimer 08] Weimer, M., Karatzoglou, A., Le, Q. and Smola, A.: COFI RANK. maximum margin matrixfactorization for collaborative ranking, *Advances in Neural Information Processing Systems*, Vol. 20 (2008)
- [Xue 05] Xue, G.-R., Lin, C., Yang, Q., Xi, W.-S., Zeng, H.-J., Yu, Y. and Chen, Z.: Scalable collaborative filtering using cluster-based smoothing, *Proc. 28th Annual ACM SIGIR Conf. on Research and Development in Information Retrieval*, pp.

114-121 (2005)

- [山西 03] 山西健司: データとテキストのマイニング, 言語と心理の統計—ことばと行動の確率モデルによる分析, 統計科学のフロンティア, 第10巻, 第4章, pp. 179-242, 岩波書店 (2003)
- [Zhang 06] Zhang, S., Ouyang, Y., Ford, J. and Makedon, F.: Analysis of a low-dimensional linear model under recommendation attacks, *Proc. 29th Annual ACM SIGIR Conf. on Research and Development in Information Retrieval*, pp. 517-524 (2006)

2007年11月14日 受理

「推薦システムのアルゴリズム (1)」正誤表

5・2節 § 3 第2段落: 「他に, 代表的な GroupLens データにおいて, …」→「他に, 代表的な映画評価データにおいて, …」

著 者 紹 介



神嵐 敏弘 (正会員)

1968年生まれ。1992年京都大学情報工学科卒業。1994年同大学院修士課程修了。同年電子技術総合研究所入所。2001年博士 (情報学)。同年電子技術総合研究所は産業技術総合研究所へ再編。機械学習とその応用の研究に従事。AAAI, ACM, 情報処理学会各会員。