

## 推薦システム—情報過多時代をのりきる

神 島 敏 弘\*

情報過多とは、自身が望む情報があっても、あまりに多くの情報の中に埋もれてしまい、それを見つけることができなくなっている現在の状況のことである。推薦システムは、この情報過多を克服するため、利用者が望む情報を見つけることを補助するものである。本稿ではこの推薦システムの動作原理を中心に述べる。

キーワード：情報過多、推薦システム、内容に基づくフィルタリング、協調フィルタリング、セレンディピティ

### 1. 推薦システム

「推薦システム (recommender system)」は、利用者にとって有用と思われる対象、情報、または商品などを選び出し、それらを利用者の目的に合わせた形で提示するために利用される。「レストランガイド」のような人手を介する方法も広義には含むが、ここでは計算機を用いて自動化されたシステムに限定する。

この推薦システムが必要になった理由について述べよう。第一に、大量の情報が発信されるようになったことがある。情報化技術の進展により、個人・団体が容易かつ低コストで発信できるようになった。第二に、これら大量の情報の蓄積や流通が容易になり、誰もが大量の情報を得ることができるようになったことである。これも計算機の記憶媒体の大規模化や、通信の高速化によるものである。以上のことから、大量に発信された情報を、だれもが大量に取得できる状況が生じた。しかし、欲しい情報が何か分からない(例：統計資料として公開されているがその名前が分からない)とか、探している情報を見つけ出せない(例：類似した資料が大量にあり目的のものが埋もれてしまう)といった理由により、情報を参照できる状態にあるにもかかわらず、それを利用できないという状況が生じた。この状況を「情報過多 (information overload)」という\*1。この状況を打破するため、利用者にとって有用な情報を見つけ出す推薦システムが考案された。

以下、本稿では、この推薦システム、特に、協調フィルタリングと呼ばれるものの実現方法を中心に述べる。

### 2. 内容に基づくフィルタリングと協調フィルタリング

推薦システムには、内容に基づくフィルタリングと協調

フィルタリングの二種類の主な実現方法がある。これらについて、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例に説明する。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するので「内容に基づくフィルタリング (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、面白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。他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)」と呼ばれる。

ここで、「推薦をする」という同じ目的のために、幾つもの方法がある理由について述べておきたい。推薦システムでは、商品情報や他人の嗜好といったデータに基づいて、利用者の嗜好などの未知の事柄を予測する。このような予測問題では、どのような種類のデータでも、他の方法より正確に予測できる万能な手法が存在しないことが知られている。よって、推薦システムが扱うデータの特徴や、システムが用いられる状況を考慮して適切な方法を選択することは重要である。内容に基づくフィルタリングと協調フィルタリングについて、それぞれの長所と短所を以下にまとめる。

内容に基づくフィルタリングの利点としては、cold-start 状況でも適切な推薦が可能なが挙げられる。cold-start 状況 (cold-start condition) とは、システムを使い始めたばかりの利用者へ推薦したり、新しくシステムに登録されたものを推薦対象にする状況のことである。新しい利用者の嗜好はすぐには把握できなかったり、新しいものを知っている利用者は少ないといった理由により、協調フィルタリングでは適切な推薦ができない。それに対し、内容に基づくフィルタリングでは、推薦対象の内容さえ分かれば推薦が可能である。他に、推薦候補として考慮されるものの範囲が広く、少数派の嗜好の利用者でも比較的よい推薦が受けられるといった利点がある。内容に基づくフィルタリングにはある。

一方、協調フィルタリングには、セレンディピティの高

\*かみしま としひろ 産業技術総合研究所  
ホームページ <http://www.kamishima.net/>

(原稿受領 2006.7.21)

\*1 情報爆発 (information explosion) や情報洪水 (information overflow) とも呼ばれる

い推薦ができる利点がある。映画を推薦する場合を例に、セレンディピティについて説明する。推薦には、目新しさ (novelty) が要求される。目新しさには、利用者の嗜好や目的に適合しているという関心と、利用者がまだ知らないものであるという新規性の二つの要素が要求される。例えば、利用者がスピルバーグ監督のファンであり、この利用者にスピルバーグ監督の新作映画を推薦したとする。このとき、利用者はこの映画に関心をもち、まだ知らない作品なので新規性もあるため目新しい推薦であるといえる。セレンディピティ (serendipity) とは、この目新しさに「思いがけなさ」や「意外性」の要素が加わった概念である。例えば、スピルバーグ監督とよく似た作風の新人監督の作品を考える。このとき、作風が似ているため利用者はこの作品に関心をもち、また新規性もある。さらに、利用者はこの新人監督の作風がスピルバーグ監督と似ていることを知らないため、この作品が推薦されることを予見できない、すなわち、意外性がある。よって、この推薦にはセレンディピティがあるといえる。内容に基づくフィルタリングでは、推薦対象の情報を利用するため、どうしても利用者が予見できる範囲の予測になりがちである。一方、協調フィルタリングでは、他人の知識をうまく利用して、セレンディピティのある推薦ができる余地が大きい。その他、協調フィルタリングには、全体的にデータが少ない状況でも相対的に適切な推薦ができるとか、推薦対象の内容のデータの設計や収集の手間がないといった利点もある。

以上のような長所と短所を考慮し、いずれかの方法を選択するか、うまく、二種類の手法を組み合わせた手法を採用する必要がある。例えば、推薦対象の商品が頻繁にほとんど全て入れ替わる電子商取引サイトでは内容に基づくフィルタリング適切であろう。また、映画や音楽のように推薦に娯楽性も求められるような場合には、セレンディピティを重視した協調フィルタリングが有利だろう。

### 3. 協調フィルタリングの手順

前節では、推薦システムの二種類の手法についてまとめた。これらの手法のうち、内容に基づくフィルタリングは、情報検索や情報フィルタリングの技術を用いて実現できる。本稿では、もう一方の協調フィルタリングを中心に紹介する。

最初に、この協調フィルタリングの手順の概略を図1に

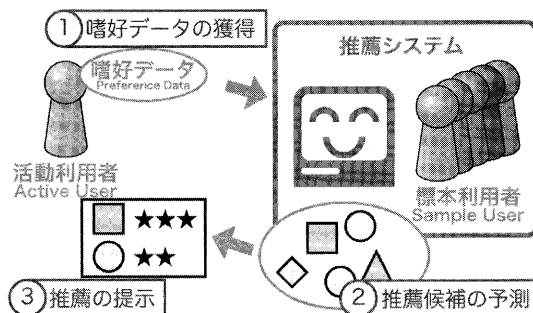


図1 協調フィルタリングの手順

示す。推薦システムを利用して、推薦を受けようとしている人を活動利用者 (active user) と呼ぶ。推薦システムは、利用される前に、いろいろな検索対象について、その好き嫌いを多くの人に尋ねた結果 (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。このとき、図1に示した三段階によって推薦を行う。

1. 嗜好データの獲得: 活動利用者は自身の嗜好データを推薦システムに提示する。
  2. 推薦候補の予測: 獲得した嗜好データと利用者 DB とに基づいて、活動利用者がまだ知らない対象についての、活動利用者の嗜好を予測する。
  3. 推薦の提示: 予測した嗜好を基に、目的に応じた適切な形式で、推薦結果を活動利用者に提示する。
- 以降、これらの各段階について詳細を述べる。

### 4. 嗜好データの獲得

嗜好データの獲得とは、嗜好データ、すなわち、利用者の関心、要求、好みについての情報を、推薦システムが収集することである。適切な推薦のためには、正確で、十分な量の嗜好データを収集する必要がある。

嗜好データを獲得するアプローチには、暗黙的と明示的の二種類がある。明示的な獲得とは、利用者に好き嫌いや、関心のあるなしを質問し、利用者に回答してもらう方法である。例えば、電子商取引サイトには、自分が購入した商品进行评估するページがよく設けられている。この評価は、良くなかった場合には★を一つ、良かった場合は★を五つ付けるなどして行う。こうして得られた★の数によって、評価した商品に対する関心や嗜好の度合いを測ることができる。他に、好きなものから、嫌いなものへ順に複数の対象を並べてもらう順位法を利用する「なんとなく協調フィルタリング」<sup>4)</sup>を筆者は提案している。

もう一方の暗黙的な獲得とは、利用者の行動から、利用者の嗜好や関心を推察することで嗜好データを得る方法である。例えば、レンタルビデオ店で、ある映画のビデオを顧客が借りたとしよう。このとき、関心がなく、嫌いな映画をあえて見るとは考えられないので、この顧客はこのビデオに関心があると推察される。他に、電子商取引サイトの例を挙げよう。商品リストから、ある商品を選んで、その詳細な情報を利用者が閲覧したとする。閲覧したこと自体で、その商品に関心があることが推察され、また、閲覧していた時間を計測することで、その関心の度合いも測ることができるだろう。さらに、推薦システムが幾つかの対象を推薦したとき、推薦に従って購入や利用をした商品と、そうしなかった商品では、相対的に前者の方により関心があると考えられる。このように、利用者の過去の行動を解析することで、暗黙的に嗜好データを獲得する方法が考案されている。

嗜好データの獲得法の長所と短所を表1にまとめた。利

表1 嗜好データ獲得法の長所と短所

	明示的	暗黙的
データ量	×: 少ない	○: 多い
データの正確さ	○: 正確	×: 不正確
利用者の認知	○: 知っている	×: 知らない

利用者の嗜好の予測には統計的な方法が用いられるので、予測を正確にするにはより多くのデータを収集できた方が有利となる。利用者は、質問に答えるといった手間を一般には嫌うため、明示的な獲得では多数のデータの収集は難しい。よって、これらの点では暗黙的な手法が有利である。逆に、暗黙的な獲得では、誤ってクリックしてしまったとか、人に頼まれて購入したなどの理由で、本当は関心がないものも、関心があるとみなされてしまう場合がある。このため、収集されたデータの正確さにおいては明示的な獲得が優れている。最後の利用者の認知とは、利用者が自分の嗜好データをいつ、どのように取得されたかを知っているかどうかということである。システムが提示した推薦は、利用者がその根拠を把握していた方が受け入れられやすい。暗黙的な獲得では、推薦が根拠なくなされたもののように感じられやすいので不利となる。

## 5. 推薦候補の予測

この段階では、前節の方法で獲得した、活動利用者や標本利用者の嗜好データに基づいて、活動利用者の嗜好を予測する。ここでいう予測とは、活動利用者がまだ知らないが、他の標本利用者は知っている対象について、活動利用者の関心の有無や、嗜好の度合いを予測することである。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このような問題は、機械学習や統計的予測の枠組みによって解く。だが、万能な予測手法はありえないため、利用者数や対象数などの利用者 DB 性質や、推薦の利用目的に応じた手法が必要になり、さまざまな手法が開発されている。本節では、これらの手法についてまとめる。

推薦候補の予測手法は**メモリベース法 (memory-based method)\*2**と**モデルベース法 (model-based method)**に分けられる。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なにもせず、ただ利用者 DB を保持している。そして、推薦をするときには、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。モデルとは、「Aさんが好むものは、Bさんも好むことが多い」といったような、利用者と推薦対象の嗜好についての規則性を表したものである。推薦をするときには、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

\*2 事例ベース法 (instance-based method) とも呼ばれる

表2 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	○: 早い
適応性	○: あり	×: なし

いが重要である。

これら二つの手法にも、やはり、長所と短所があるので表2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は、時間がかかるためである。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事前に構築するので、この時間は推薦の早さに影響しない。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、活動利用者に早く推薦することができる。表2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、頻繁に行うことは無理である。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。以上、メモリベース法とモデルベース法について述べた。以後は、これらについて詳細を述べる。

### 5.1 メモリベース法

メモリベース法とは、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。まず、自動化された協調フィルタリングの最初の手法で、単純だが適切な推薦ができる GroupLens の方法<sup>9)</sup>について述べる。

GroupLens は、NetNews の記事の中から、利用者に関心をもつ記事を推薦するシステムとして開発された。だが、現在は NetNews があまり利用されなくなったため、同じ手法を用いた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、GroupLens プロジェクトのホームページ (<http://www.grouplens.org/>) で公開されているので、推薦システムを体感するため利用してみるとよいだろう。レストランを探す場合に、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。人のコミュニティで行われている「口コミ」によるこうした推薦の過程を自動化するのが GroupLens の方法である。この過程を、GroupLens では次の二段階で実現する。

1. 類似度の計算: 利用者 DB 中の各標本利用者との嗜好の類似度を求める。類似度とは、嗜好の傾向がどれくらい似ているかを数値化したものである。
2. 嗜好の予測: 活動利用者が知らないが、いずれかの標本利用者は知っている対象について、活動利用者がど

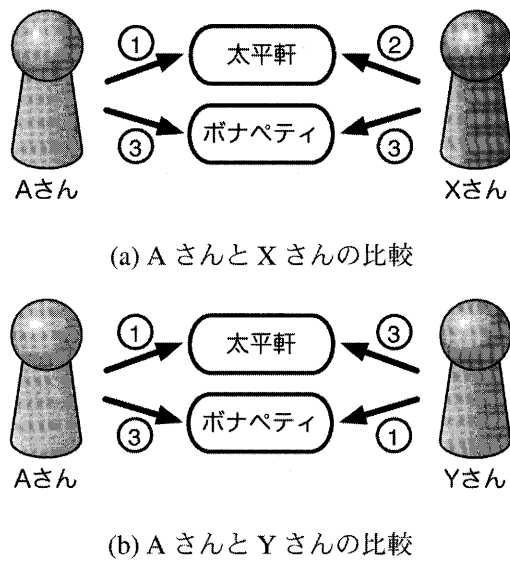


図2 活動利用者と標本利用者の比較

れくらいその対象に関心があるかを予測する。

以下、各段階について詳細に説明する。

類似度の計算について述べる。この類似度は、活動利用者と標本利用者が共に知っている対象についての評価がどれくらい一致しているかによって測る。レストランの推薦システムを例に説明する。まず、図2(a)の活動利用者Aさんと、標本利用者Xさんの類似度を考える。評価は1-3の三段階で行い、二人とも「太平軒」と「ボナペティ」の2軒のレストランは知っているとしよう。Aさんは中華料理は好きではないので太平軒は1と評価しているが、Xさんにとってはどちらともいえないので2と評価している。一方、AさんもXさんもフランス料理は好きなので、ボナペティはどちらも3と評価している。このように、どちらも同じようなものを好み、同じようなものを嫌う傾向がある場合は、二人の嗜好は類似しているとみなすべきである。次に、図2(b)のAさんと、別の標本利用者Yさんの類似性を考える。今度は、AさんとYさんの評価は全く正反対で、二人の嗜好は似ていないとみなすべきである。こうした性質をもつように類似性を数値化するために、GroupLensでは相関係数を用いている。詳細は略すが、図2(a)のように二人の嗜好が似ているときは相関係数は正の値となり、完全に嗜好が一致するなら最大値1をとる。逆に、図2(b)のように正反対なら負の値となり、最小値は-1である。この例では、Xさんとの類似度は正、Yさんとの類似度は負になる。この相関係数を、全ての標本利用者X、Y、Zさんについて計算する。

こうして求めた類似度をもとに、Aさんが知らない店についての評価を予測しよう。X、Y、Zさんは「満腹庵」について知っているが、Aさんは知らないものとする。Xさんの満腹庵への評価に基づく、Aさんの評価の予測値は次のように計算される。

$$[X \text{ との類似度}] \times ([X \text{ の評価}] - [X \text{ の評価の平均}])$$

この式で、Xさんの満腹庵への評価が平均より大きけれ

ば括弧の中は正、平均より低ければ負となる。括弧の中が正、すなわち、Xさんが満腹庵を好きで、かつ、Xさんとの類似度は正、すなわち、AさんとXさんは嗜好が似ているとする。すると、Xさんの嗜好に基づくAさんの嗜好の予測値は正になり、Aさんは満腹庵を好むと予測される。逆に、Xさんの評価が平均より低ければ、Aさんの評価も低くなる。Yさんについても同様の計算をする。今度は、YさんとAさんの類似度は負であることに注意されたい。すると、Yさんの満腹庵への評価が平均より低ければ、Aさんの嗜好は正となる。YさんとAさんの嗜好が正反対なので、Yさんが嫌うものをAさんが好むということが、この式によりうまく表されている。X、Y、Zさんそれぞれの嗜好に基づく、Aさんの満腹庵への評価を予測し、これらの評価値の平均をもって、Aさんの評価の予測値とする。ただし、多くの標本参加者が知っているお店の評価は過大になってしまうため、これを補正する方法が導入されている。以上の手続きを、満腹庵やバスタランドといったAさんが知らない全ての店について行う。

このメモリベース法には幾つか改良や、バリエーションがある。嗜好が類似している標本利用者、すなわち、類似度が大きな利用者のグループを求めておき、そのグループの中の標本利用者の意見だけを利用する。こうすると、利用者DB全体を調べ直さなくて済むため、推薦に要する時間を短縮できる。

## 5.2 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。このモデルの種類や、構築の方法は多種多様なものが用いられている。ここでは、代表的なモデルを簡単に紹介する。

最も多種多様な手法が提案されているのは確率を使ったモデルである。確率は、偶然に生じる事象を扱うために用いられ、ある事象の確からしさを0から1までの値で表現する。例えば、評価を好き/嫌いの二段階で行う、レストランの推薦を考えよう。すると、0と1のちょうど中間である0.5が好き/嫌いの分岐点になる。ここで、「Aさんが、満腹庵を好きな確率は0.2」などといった場合には、この分岐点をかなり下回っているので、あまり満腹庵を好きではないことが表される。協調フィルタリングのモデルでは、この確率の一種である条件付確率が利用されるので、これについて述べる。

条件付確率とは、その名の示すように、条件によって変化する確率である。例えば、ある人がボナペティを好きならば、太平軒を好きな確率は0.1で、ボナペティが嫌いならば、太平軒を好きな確率は0.8といったようなものである。この条件付確率は、利用者DBに登録されている嗜好データの情報を数え上げることで求めることができる。ありとあらゆる推薦対象や条件について、上記のような条件付確率を求めたものを条件付確率分布といい、この条件付確率分布がモデルとなる。

このモデルを使って、活動利用者の嗜好は次のように予

測される。活動利用者 A さんから嗜好データを獲得し、そのデータからボナペティを好きなことが分かったとする。さらに、事前に構築した条件付確率分布によるモデルによれば、ボナペティを好きな人が、太平軒を好む確率は 0.1 であるとする。この二つの事柄から、活動利用者は太平軒が好きではないことが分かる。この条件付確率を表現する方法には、多種多様なものが考案されているが、ベイジアンネットワークや潜在変数を用いるものなどが主に利用されている。

確率を使わないモデルとしては、クラスタリングを利用する方法がある。クラスタリングは、似ているものは同じものに、似ていないものは違うものに、グループ分けする手法である。この手法により、標本利用者を、嗜好の傾向が類似しているグループに分類し、このグループ分けをモデルとして利用する。ここで、利用者間の類似性は、5.1 節のように、いろいろな対象への評価のパターンがどれくらい類似しているかによって評価する。活動利用者への推薦は、活動利用者と各グループとの類似度を調べ、最も似ているグループを見つける。そして、そのグループ中の標本利用者が好む対象は、活動利用者にも好まれると考えて、活動利用者の嗜好を予測する。

最後に、少し傾向の違うモデルを紹介する。今までの手法では、時間について考慮することはなかった。しかし、この時間に伴う変化が推薦に役立つ場合もある。例えば、レンタルビデオ店で、先々週はある映画のエピソード I を、先週はこの映画のエピソード II を借りていった顧客がいたとする。このとき、今週はこの映画のエピソード III を借りることが予測できる。利用者 DB に、標本利用者の時間順の行動や評価の履歴情報を蓄積しておけば、こうした時間に伴う利用者の行動の傾向を表すモデルを構築できるであろう。このモデルを使えば、活動利用者の現在にいたる行動や評価に基づいて、次の行動を予測し、その予測結果を推薦に役立てることができる。この時間に伴う変化を表すモデルには、マルコフモデルと呼ばれる時間変化を表現するものがよく利用されている。

## 6. 推薦の提示

前節の方法で、活動利用者がまだ知らないいろいろな推薦対象について、彼/彼女の評価が予測できた。次に、この予測結果を、活動利用者の目的に適した形式で提示する方法について述べる。文献 3) では、推薦システムを利用する主な目的には「よい対象の発見 (find good items)」と「注釈の添付 (annotation in context)」があると述べている。これらの目的と、それに応じた推薦の提示手法について述べる。

「よい対象の発見」とは、利用者が積極的な動機を持って、情報を見つけるために推薦システムを利用する場合である。例えば、今から食べに行く店を決めるために、レストラン推薦システムを利用する場合などである。この場合には、予測される評価の高いものから順に、推薦対象を整理して利用者に提示する。利用者はこのリストを上位から閲

覧することで、自身の決定に必要な情報を知ることができるだろう。また、この場合には推薦候補の範囲が網羅的であるとか、セレンディピティが高いといった要素より、予測結果がどれだけ正解となるかといった要素を重視した推薦をすべきである。

「注釈の添付」とは、いろいろな推薦対象の閲覧や検索をしているときに、副次的な情報として予測される関心の度合いを参考にしたい場合である。例えば、決まった予定はないが、レストランの紹介 Web サイトなどを閲覧しているときなどである。関心の度合いは、★の数や、アイコン、グラフなどを対象と共に表示することで利用者に提示する。このような情報を参照することで、多数のレストランの中から、利用者にとって関心のあるものを中心に閲覧できるようになるであろう。「よい対象の発見」とは逆に、予測が正解することより、推薦候補の網羅性やセレンディピティを、この場合には重視すべきである。

以上、推薦結果を単純に提示する方法について述べたが、この段階についても、幾つか改良手法が考案されている。推薦システムは、推薦が作為的なものではないと、利用者に信頼されなければ利用されない。例えば、電子商取引サイトなどでは、利用者は不必要に高価なものを薦められていると疑うかもしれない。この点についても研究がなされており、推薦の根拠を簡潔に提示することが効果的であるとされている。例えば、「過去にこういった商品を購入したことがあるので、この商品はどうでしょうか?」とか、「あなたと趣味が似ている人たちの、この映画についての評価を棒グラフで表示しました」といった手法が知られている。

音楽、映画、旅行など娯楽性の高い推薦対象を扱う推薦システムでは、推薦結果に高いセレンディピティが要求される。よって、推薦の提示段階でセレンディピティを向上させる研究もある。推薦対象をその特徴により分類しておく、すでに類似した対象が推薦されていれば、予測評価値が高くてもそれを推薦せず、代わりにより下位の候補を推薦する方法が提案されている。例えば、旅行の推薦システムで、単純に予測評価の高さを重視すると、どれも同様のグルメツアーだけになってしまったりする。この場合には、予測評価値が多少低めの、体験ツアーやエステツアーなど別のカテゴリの旅行プランとリスト一部をあえて交換することで、セレンディピティを向上させる。

## 7. プライバシー保護協調フィルタリング

最後に、今までとは少し異なる問題点に対処するための方法を紹介する。協調フィルタリングでは、標本利用者の嗜好データを収集する。このデータが、商品の購入履歴などであった場合は、注意して管理すべき個人情報となる。現在は、この個人情報の収集に伴う問題には、プライバシー・ポリシーを公開し、それを遵守することを誓約することで、社会的手段によって対処している。これに対し、技術的手段によってこの問題に対処するのがプライバシー保護協調フィルタリング (privacy-preserving collaborative filtering) である。

まず、プライバシー保護協調フィルタリングの基本的なアイデアを述べる。モデルベースの協調フィルタリングでは、標本利用者の嗜好データではなく、それらの規則性を表したモデルさえあれば推薦が可能である。よって、嗜好データそのものを知ることなくモデルが構築できれば、プライバシーの問題は生じない。これを、データを暗号化したまま計算する安全な計算 (secure computation) と呼ばれる技術によって実現する。このアイデアに基づく枠組みを図3に示す。まず、標本参加者は自身の嗜好データを暗号化して推薦システムに渡す(図の上)。ここで、渡されたデータは暗号化されているため、個人情報には漏れない。これらの暗号化嗜好データに、安全な計算の技術を適用して、暗号化された状態のモデルを獲得する(図の中)。最後に暗号化モデルを復号して目的のモデルを得る。ここで、このモデルは標本利用者全般の嗜好の規則性の情報を表しているが、このモデルから個々の標本利用者の嗜好データ自体を復元することはできない。よって、個人情報の漏洩にはならない。活動利用者は、自身の計算機上で、このモデルと自身の嗜好データから、推薦を得ることができる(図の下)。

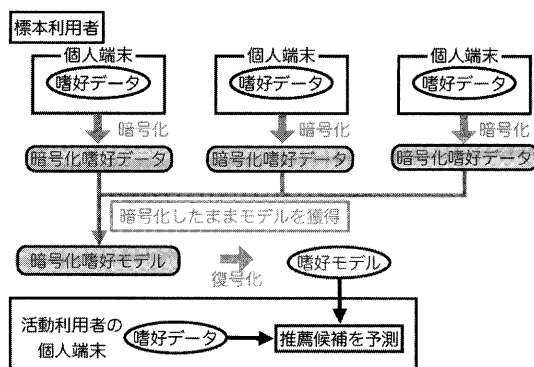


図3 プライバシー保護協調フィルタリング

しかし、現状のプライバシー保護協調フィルタリングでは、社会的手段が補助的に必要となる。安全な計算を実現するには、半数以上の参加者は、個人情報を明かすほど信

頼はできないが、計算の手続きは遵守する程度には信頼できるという semi-honest という前提が必要である。しかし、この前提の技術的手段による保証は難しく、社会的手段によって保証しなければならない。

## 8. まとめ

本稿では、情報過多への解決策としての推薦システムについて述べた。

最後に、推薦システムについての参考文献を挙げておく。推薦システムのパイオニアは GroupLens<sup>5)</sup> というシステムであるが、推薦システムという呼び方は ACM Communications 誌での特集<sup>6)</sup>に始まる。推薦システムの良いチュートリアルとしては 1) がある。また、3) では推薦システムには何が求められ、どういったことを考慮すべきかについて深い考察がなされている。推薦システムのアルゴリズムの比較としては 2) がある。推薦システムについて本格的に知りたい方は、これらの文献が参考になるであろう。

### 参考文献

- 1) J. Ben Schafer, J. A. Konstan, and J. Riedl. E-commerce recommendation applications. *Data Mining and Knowledge Discovery*. Vol.5, p.115-153 (2001)
- 2) J. S. Breese, D. Heckerman, and C. Kadie. Empirical analysis of predictive algorithms for collaborative filtering. *In Uncertainty in Artificial Intelligence 14*, p.43-52 (1998)
- 3) J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl. Evaluating collaborative filtering recommender systems. *ACM Trans. on Information Systems*. Vol.22, No.1, p.5-53 (2004)
- 4) T. Kamishima. Nantonac collaborative filtering: Recommendation based on order responses. *In Proc. of The 9th Int'l Conf. on Knowledge Discovery and Data Mining*. p.583-588 (2003)
- 5) P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl. GroupLens: An open architecture for collaborative filtering of Netnews. *In Proc. of The Conf. on Computer Supported Cooperative Work*. p.175-186 (1994)
- 6) P. Resnick and H. R. Varian. Recommender systems. *Communications of The ACM*. Vol.40, No.3, p.56-58 (1997)

Special feature : Information filtering. Recommender system - Solving the problem of information overload. Toshihiro KAMISHIMA(National Institute of Advanced Industrial Science and Technology (AIST), AIST Tsukuba Central 2, Umezono 1-1-1, Tsukuba, Ibaraki 305-8568 Japan)

Abstracts : We are recently encountering the situation where we cannot find out our target information even if we have an access to it, because it is buried in so enormous amount of irrelevant information. This situation is known as information overload. Recommender systems are developed to solve this problem of information overload. These systems help for users to find out their target information. In this article, we overview the algorithms for making recommendations.

Keywords : information overload / recommender system / content-based filtering / collaborative filtering / serendipity