

解説

推薦システムのアルゴリズム (1)

Algorithms for Recommender Systems (1)

神 薦 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.

1. 推薦システム

「推薦システム (recommender system)」とは、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システムである。

最初に、この推薦システムが必要になった背景について述べよう。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。これは、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ためである。第二の理由は、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、誰もが大量に取得できる状況が生じた。しかし、欲しい情報が何かわからない(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^{*1} (information overload)」という。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。

この推薦システムは、広い立場から見れば、情報検索や情報フィルタリング技術の一つとみなせる。そのため、初期の推薦システムは、これらの技術が基盤となっていた。6 章で述べるように、このシステムの実現手法には協調フィルタリングと内容ベースフィルタリングがある。だが、協調フィルタリングという語のほうが推薦システムより古く、1992 年に文献 [Goldberg 92] にて使われた。だが、これは現在のような協調フィルタリングではなく、他人が手動で行った推薦を検索できる協調作

業支援のシステムであった。この過程を自動化した協調フィルタリングが 1994 年の GroupLens [Resnick 94] や Ringo [Shardanand 95] であり、現在の推薦システムの基礎となった。一方の内容ベースフィルタリングの技術は、従来からある情報フィルタリングとみなせ、また、事例ベース推論の応用としても研究されてきた。そのうち、推薦システムとして独自の側面が強くなり、協調フィルタリングに対して内容ベースフィルタリングと呼ばれるようになった。1996 年には、専門のワークショップも開催されるほどに研究が活発化した。1997 年には ACM Communications 誌で特集 [Resnick 97] が掲載され、この種のシステムの呼び名として「Recommender System」が定着した。この頃から、NetPerceptions や Firefly などの企業によってシステムの商業化が始まった。現在では、Web を通じた各種サービスの機能で活用されたり [Das 07, Lawrence 99, Linden 03]、セットトップボックスなどの機器に組み込まれたり [Ali 04] している。その後、物理的な店舗面積に商品数が制限されない電子商取引の発展や、大量の画一的な商品から、少量多品種への消費傾向の変化に伴って、その重要性も広く認識されるようになった。それを象徴する Amazon.com CEO の Jeff Bezos の発言を引用しておこう [Ben Schafer 01].

If I have 3 million customers on theWeb,

I should have 3 million stores on theWeb^{*2}

(Web に 300 万人の顧客がいるなら、

300 万の Web ストアを用意すべきだ)

現在では、推薦システムは多方面で利用され、研究も継続的に行われ、多様な方法が目的に応じて考案されている。これらの推薦システムの設計指針やアルゴリズムについてまとめる。

本稿の構成は以下のとおりである。2 章と 3 章では推薦システムの分類と評価指標について、4 章では全体の三つの実行過程について、5 章では最初の段階である「デ

^{*1} 情報爆発 (information explosion) や情報洪水 (information overflow) ともいう。

^{*2} J. Bezos は、この発言をいくつかの講演で行った。ここでは、300 万人と書いたが、そのときどきの顧客数に応じて、この数字は変えて用いられた。

ータの入力」, 6 ~ 12 章では次の「嗜好の予測」段階, そして 13 章では最後の「推薦の提示」段階を紹介する. 14 章ではその他の問題や視点について, 15 章では参考資料を紹介し, 最後に 16 章では推薦システムの今後の展開について述べる. これらの章を, 1 ~ 5 章, 6 ~ 10 章, および 11 ~ 16 章の全 3 回に分ける. 第 1 回である今回は, 5 章までを掲載する.

1.1 本稿の目的

ここで紹介するアルゴリズムは研究用につくられたシステムで, ほとんど統制環境下で評価されている. よって, そのまま実地環境で用いても, 実用システムとして満足な結果は得られない. それでも, 実用システムの開発者がこれらの研究成果を知ることは, 非常に有意義であると著者は考えている. 本題に移る前に, この点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のデータから, 利用者の嗜好を機械学習の手法を用いて予測する. だが, あらゆるデータに対してもほかより必ず正確に予測できるような手法は存在しないとするノーフリーランチ定理などが機械学習の根幹には存在する [Duda 00]. すなわち, どんな状況にも対応できる万能機械学習アルゴリズムは, 原理的につくれない, いわば永久機関のようなものである. 推薦システムにおいても, 利用者数やアイテム数などの要因で, システムの性能や振舞いに変化することは指摘されている [Ben Schafer 01, Herlocker 04]. よって, 研究システムが実用システムとしてそのまま利用できないのと同様に, 実際にうまく稼働しているある実用システムを別の目的に用いても, やはりうまく稼働しない.

では, 実用システムは毎回, 一からつくり直さなければならぬのであろうか? もちろん, そのようなことはない. 統制環境下での評価によって, 問題をいくつかの独立した要因に分け, 各要因の影響が調査されている. さらに, 各要因の多様な状態について適切な予測ができる手法が研究され, いろいろな知見も蓄積されている. 例えば, 娯楽性の高いものの推薦では, 予測の正しさとともに, 推薦のバリエーションも重要になり [McNee 06a], それを実現する手法が提案されている [Ziegler 05]. 実用システムの開発においては, こうした知見をもとに, 各種の手法を目的に応じて組み合わせて基礎となるシステムをつくり, 細部を実情に合わせてつくり込むことで, 開発に要する労力を低減することが可能である. そのために, 本稿は, 各手法がどういった状況で有利・不利になるかをできるだけ述べるようにこころがけた.

2. 推薦システムの分類と目的

本章では, 推薦の個人化の度合いの分類と, 運用側と利用者側のそれぞれ目的によるシステムの分類について

述べる. これらの目的を考慮して, 推薦システムの設計方針を決めることになる.

その前に, 推薦システムに関連する技術と, それらとの違いについて述べておく. まず, 情報検索における情報フィルタリングがある. これは, 逐次的に入力される情報から, 利用者の関心や興味を記述した利用者プロフィールに適合するものを選別する技術 [徳永 99] で, 推薦システムと良く似ている. だが, 主にテキスト文書を扱い, 利用者プロフィールも主に索引語で示す点や, 必要なものを取り出すより, 不要なものを除外することが主目的である点などが異なる. ほかに, マーケティングの技術とも関連がある. しかし, マーケティングが供給側の視点に立つのに対し, 推薦システムは消費側である. 推薦システムの推薦が利用者に受け入れられるためには, 推薦が実際に客観的なものであり, また, そのことを利用者に示す工夫が必要になる. また, マーケティングでは, レポートなどで全体の傾向などを分析して報告することが必要だが, 推薦システムにはそうしたことは要求されない. また, マーケティングでは, 顧客を客層に分類し, それに応じた対処を行うが, 推薦システムでは, 利用者のグループ化は目的ではない.

2.1 推薦の個人化の度合い

最初に, 推薦の個人化の度合いを 3 段階 [Ben Schafer 01] で示す.

§1 非個人化 (no personalization)

すべての利用者について, 全く同じ推薦をする場合である. 例えば, 編集者による推薦や, 売上げ順位リストなどである. 推薦システムというときは個人化かつ自動化されたものを想定するかもしれないが, こうしたものも広義には含める. 非個人化の推薦の例を図 1 (a) と (b) に示す. 図 1 (a) 中の, (1) は店舗の編集者が手作業で選んだもの, (2) は予約や売上げの順位リストである. システム側だけでなく, 図 1 (b) のように, ほかの利用者が手作業で作成する推薦リストもある.

§2 一時的個人化 (ephemeral personalization)

システムを利用する一つのセッションで同じ入力や振舞いをした利用者には, 同じ推薦をする場合である. 一時的個人化の推薦の例を図 1 (c) に示す. この例では, 利用者がある本を閲覧するという行動をシステムに対してしたとき, その本に関連する情報を示している. 図中の, (1) はこの本と関連が深い本を推薦し, (2) には, 書誌情報や売上げ順位など, (3) ではほかの利用者の評価やコメントなど, この本についての関連情報を提示している.

§3 永続的個人化 (persistent personalization)

たとえ同じ入力や行動をシステムにしている利用者でも, 利用者の個人情報や過去の利用履歴に応じて異なる推薦をする場合である. 例えば, 過去の購入・利用履歴に基づいて推薦したり, 年齢に応じて推薦するアイテムを変

えたりする。永続的個人化の推薦の例である図1(d)では、この利用者の過去の商品への評価に基づいて、関心があるであろう本を予測し、順位付けして提示している。

2.2 推薦システムの運用目的の分類

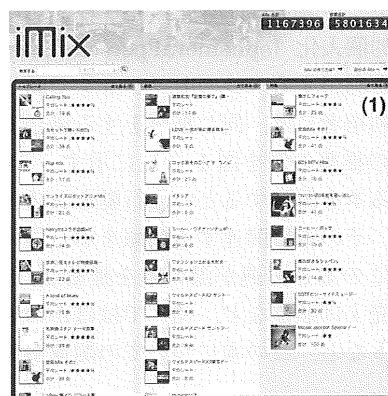
次に、推薦システムを、運用側の目的に基づいて分類する。[Ben Schafer 01]では、自動・手動の推薦システムを次の5種類に分類し、目的に応じて適切に組み合わせて利用すべきと述べている。

§1 概要推薦(broad recommendation)

これは、全体の統計情報や運用者側の編集者を使う場合である。全体の統計情報とは「今週の売上げランキング」や「個々の商品の売上げ順位」といったもので、編集者による推薦には「評論家が推薦する映画」や「特売品リスト」などがある。図1(a)はこうした概要推薦の例である。こうした概要推薦は、システムを利用し始めたばかりか、ごくまれにしか利用しないような利用者を対象とする。これらの利用者が、自身の要求との関連性を見いだして、システムの利用を続けてもらえるように、大まかな情報を提供するのが目的である。よって、



(a) 非個人化(iTunes Store)



(b) 非個人化(iTunes Store)

積極的に探さなくても見えるように、システムにアクセスしたとき最初に見える場所に配置すべき推薦である。加えて、利用者が積極的にシステムに働きかけることは期待できないので、非個人化またはアイテムのカテゴリを示す程度の入力に対応した一時的個人化した推薦をすべきである。

§2 利用者評価(user comments and rating)

これは、運用側が用意したシステム上で、利用者間での推薦をする場合である。例えば、読者の批評文や★の数で示した評価レートを付けられるようにし、それらを



(c) 一時的個人化(Amazon.co.jp*)



(d) 永続的個人化(Amazon.co.jp*)

図1 個人化の度合いの異なる推薦の例

*3 図1(c)は2007/08/04に、図1(d)は2007/07/26にスクリーンショットを取得した。

ほかの利用者が参照したり、平均評価値などの統計情報を見たりできるようにする。図1 (b) のようなほかの利用者による推薦リストや、図1 (c) (3) のようにほかの利用者による評価やコメントなどがこの種の推薦の例である。

一般に、運用側のシステムが第三者的な立場で推薦しているとは、利用者にはあまり思われない。それに対し、ほかの利用者のほうがずっと信用され、その推薦も受け入れられやすい。運用側には、システムに対する利用者の信頼を高めたり、利用頻度を高めたりといった利点がある。こうした推薦は、運用側の関与は少なく、評価値順のアイテムリストなどの非個人化か、閲覧中のアイテムの評価値やコメントなどの単純な一時的個人化が行われる。

§3 通知サービス (notification service)

これは、利用者がシステムを操作していないときに、電子メールなどで推薦を配送する場合である。一つには、過去の購買履歴に基づいて、新規のアイテムの中から推薦リストを生成し、利用者に送付するといった、永続的個人化をするものがある。また、利用者があらかじめ設定した条件、例えばファンである歌手などを設定しておく、その歌手の新譜の案内が届くという一時的個人化をした推薦もある。これらの推薦は、システムの再利用を利用者に促すことを目的とする。

§4 関連アイテム推薦 (item-associated recommendation)

利用者が注目している、例えば電子商取引サイトで、商品閲覧したり、「買い物かご」にすでに商品を入れている状況を想定する。このとき、注目しているアイテムの比較候補 (例：他メーカーの同等品) を示して、購入の判断を助けたり、購入の決断を促す場合や、補足的な商品などを提示して cross-selling (例：ハンバーガーにポテトなどの関連品を薦めて同時に購入させる) を促す場合がこの種の推薦である。例としては、図1 (c) (1) のように閲覧中の商品の関連商品を示すものがある。注目している商品の情報だけにに基づく一時的個人化も、利用者の過去の行動も考慮する永続的個人化も可能である。

§5 緊密な個人化 (deep personalization)

システムが積極的に利用者の情報を能動的に収集したり、過去の行動の情報を蓄積し、それらに基づいて推薦を行う場合である。図1 (d) のような、個人向けの推薦リストなどがこの種の推薦に該当する。利用者がシステムを利用し続けることで蓄積された情報に基づいて、永続的な個人化が可能になる。すると、より適切な推薦ができるようになり、ほかのシステムとの差別化につながり、利用者の長期間にわたるロイヤリティ構築に役立つ。だが、一方で、最も実装に困難が伴い、コストも要する。

2.3 推薦システムの利用目的の分類

利用者が推薦システムを利用する目的の分類を、

[Herlocker 04] に沿って示す。これらの目的に応じて、3・1節で述べる推薦の性質が適切になるように調整する必要がある。

§1 適合アイテム発見 (find good items)

この目的は、利用者が自分の嗜好に適合するものを、何か見つけ出すことである。利用者が積極的な動機をもって、情報を見つけるために推薦システムを利用するときは、こうした目的になる。

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ことが想定される。この場合には、予測される評価の高い比較的少数のものに絞り込んで利用者に提示する。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。

§2 評価閲覧 (annotation in context)

何らかのフォーマットに従ってアイテムが整列されていたり、構造化されている状況を考える。例えば、新商品を発売日が直近になるように整列したり、カテゴリ別に商品进行分类したり、メールや掲示板で記事をスレッド状に表示したりする場合である。このように整列されたアイテムを、利用者が閲覧しているとする。このとき、閲覧の順序を決めたりどの記事を閲覧するかなどを決めるために、手がかりとなる情報があれば便利である。そのような情報として、利用者が関心をもつ度合いをシステムは予測して付随的に提示する。適合アイテム発見とは違って、利用者には必ずしも最終的に意思決定をするような積極的な動機があるとは限らず、また、関心のものを特に選んで表示するといったアイテムの提示フォーマットを変えるような要求もない。

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。レストランは、料理の種別や推薦の度合いは、★の数やアイコン、グラフなどを対象とともに表示することで利用者に提示する。この推薦情報は、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絞り込むのに役立つ。

§3 適合アイテム列挙 (find all good items)

この目的は、利用者が自分の嗜好に適合するものを網羅的に見つけ出すことである。これは裏返しに、適合しないものを排除する目的であるともいえる。例えば、会社の法務部門が関連する特許や判例を検索したり、スパムメールの可能性がないメールだけを閲覧したいといった場合である。

§4 その他の利用目的

意思決定とは関わりのない単なる閲覧といった目的もある。さらに、どの推薦システムを利用するか決めるために推薦の質を見極めようとする利用者などもある。これら目的のうち、主なものとは適合アイテム発見と評価閲覧であり、ほとんどの推薦システムは、これら二つのいずれかの目的を想定している。

そのほか、[Swearingen 01] では、ほかの観点からの

分類を示している。前者のタイプほど、既知のアイテムに近い推薦することが望まれる。

- (1) 備忘録：既知のアイテムを思い出させる。
- (2) 類似品：比較などのため既知のアイテムに類似したものを探す。
- (3) 新規アイテム：自分が確実に好むであろう、未知の新製品を探す。
- (4) 視野を広げる：他のジャンルにも自分の関心を広げる。

3. 推薦システム設計の要素

多種多様な推薦システムのためのアルゴリズムが存在するが、一体、どのアルゴリズムを使えばよいのだろうか？ 1.1節で述べたように、万能アルゴリズムは存在し得ないので、これは、推薦システムを利用する目的や推薦を実行する環境の制約に応じて選択する必要がある。ここでは、そのために考慮すべき要因を大きく二つに分けて述べる。一つは、推薦の性質である。2.3節の利用者の目的などに応じて、推薦は適切な性質を備えるべきである。この性質を測るための規準をいくつか示す。もう一つは、推薦を計算するためのデータや計算機資源の制約である。データや計算機資源は無限にはなく、何らかのトレードオフを考慮しつつ推薦システムは設計する必要がある。

3.1 推薦の性質

推薦システムでは、利用者が好むものを予測して提示する。だが、利用者が何を好むかは、利用者の目的、システムを利用する状況、推薦の候補などによって変化する。よって、提示する推薦の性質を決めるにあたって考慮すべき規準があれば役立つ。これらの規準を以下に示す。なお、推薦の質の評価については、[Herlocker 04]が詳しい。

§1 予測精度

予測精度とは、予測して推薦したアイテムに、実際にどれくらい利用者が関心をもつかという規準である。利用者が関心のないアイテムを推薦しても役に立たないので、予測精度は最も重視すべき規準である [Swearingen 01]。評価はオンラインで行う場合と、オフラインで行う場合がある。オンラインでの評価とは、被験者に実際にシステムを利用させ、推薦が適合したかどうかを調査するものである。一方、事前に被験者から集めた嗜好データと、予測した結果の一致を調べるのがオフラインでの評価である。前者のほうがより実際の運用に近い評価ができるが、この調査のコストは高い。そのため、得られるデータ数が少なく安定した検証が困難であったり、多数の項目での比較が困難になるなどの問題がある。こうした点ではオフラインでの評価のほうが有利になる。

オンラインの評価では、それぞれのアルゴリズムを利

用させて、推薦を利用者が受け入れる比率で、アルゴリズムの相対的な予測精度を評価する。また、二つのアルゴリズムによる推薦リストの上位から交互にアイテムを選んで、一つの推薦リストにまとめる。そして、利用者にそれを提示して、どちらのアルゴリズム由来のアイテムがより頻繁に選ばれるかといったことで評価する方法などがある。

オフラインの評価では、一般の機械学習と同様に、交差確認によって汎化誤差を推定し、その汎化誤差で予測精度を評価する。なお、予測精度の評価については [元田 06] の 5 章に詳しい。ここで、アルゴリズムに超パラメータがある場合の注意を述べておく。超パラメータとは、8.1節の方法の近傍の大きさといった、アルゴリズムで調整すべき変数のことをいう。通常の交差確認では、データを訓練用とテスト用の二つに分けるが、超パラメータがある場合は、厳密には、データを訓練用 (training)、確認用 (validation)、およびテスト用 (test) の三つに分ける必要がある [Bishop 06, 1.1 節, Hastie 01, 7.2 節]。そして、アルゴリズムの学習には訓練用データを、超パラメータの決定は確認用データを、そして最終的な予測精度はテスト用データを用いる。特に、新たに超パラメータを導入したが予測精度の向上がわずかな場合には、こうした厳密な評価実験をしておくことを薦める。

推薦システムでは、次のような尺度がテスト用データに対する予測精度の評価に利用されている。

正解率 (accuracy) 利用者の関心への適合や不適合が、予測結果とテスト用データで一致した割合を示す。評価値を予測するシステムでは 5 段階のうち上位 2 段階のいずれかなら適合とみなしたりする。最も基本的な評価指標である。

精度 (precision) と再現率 (recall) 適合判定されたアイテムのうち実際に適合しているものの割合が精度、すべての適合アイテムのうち適合と判定されたものの割合が再現率である [徳永 99]。適合アイテムを一つ見つければよい適合アイテム発見タスクでは精度を、すべての適合アイテムを見つけたい適合アイテム列挙タスクでは再現率を重視すべきである。情報検索で利用されている、精度や再現率に基づいた F 尺度 (F -measure) や ROC 曲線 (receiver operating characteristic curve) などとも用いられている。

平均絶対誤差 (Mean Absolute Error: MAE) テスト用データの評価値と予測した評価値の差の絶対値のテスト用データ上での平均である [Herlocker 04]。評価閲覧タスクでは、適合か不適合かの判定より、評価値そのものを利用者は見る。よって、このタスクでは、評価値の予測精度のずれを評価するこの指標を重視すべきである。

half-life utility metric と順位相関 (rank correlation)

これらは、推薦するアイテムの並び方の良さを評価する指標である。計算方法の詳細は [Herlocker 04] を参

考にされたい。適合アイテム発見を目的とする場合、推薦システムには意思決定支援の側面が要求される。このときは、評価値自体よりも、評価の大きさを重視すべきである。これらの指標はこうした場面で有用である。

最後に、推薦システムの予測精度の評価についての問題を指摘しておこう。交差確認による評価では、テストに使ったサンプルと今後予測するサンプルは同じ分布から得られることを、通常は仮定している。詳細は5・2節で述べるが、評価されていないアイテムは、利用者に関心を示さなかったものであることが多い。そのため、評価値が欠損するアイテムは、評価が低いアイテムに偏る傾向がある。こうした原因により、実際にシステムが稼働して予測対象となるアイテムの分布とテストに用いられるアイテムの分布は異なり、厳密に予測精度を評価することは難しい。そのため、わずかに予測精度を向上させる試みは実用的には利益がないことが多い。1%なり3%ほどの危険率で統計的に有意な差がなければ、ほぼ同等の予測精度とみなし、他の規準を重視して推薦システムを設計すべきである。

§2 セレンディピティ

利用者に関心をもつであろうアイテムを推薦することは推薦システムの目的であり、上記の予測精度はこのことを定量的に評価する。それに加えて、多くの場合、利用者が知っているアイテムを推薦してもあまり有用ではない。よって、関心があることに加えて、推薦には目新しさ (novelty)、すなわち、わかりきったものではないことが要求される。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持ち、まだ知らない目新しい推薦であり、上記の二つ条件を満たしている。さらに、要求される条件にセレンディピティの高さがある。推薦におけ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、予見のできなさ、または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持ち、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。

しかし、このセレンディピティに伴う感情的な応答を定量的に評価することは難しい。それでも、セレンディピティの一側面ではあるが、それらを定量化する試みもある。グループ向けの推薦には現れないが個人向けの推薦には現れるもの [Herlocker 04] や、単純な予測器では候補にならないが、高度な予測器では候補になるもの [村上 07] などをセレンディピティが高いとみなし、こうした仮定に基づいた定量的評価尺度を提案している。また、

推薦するアイテムの多様性を重視することで、推薦に意外性が生じる可能性を広げ、しいてはセレンディピティを高めようとする研究事例などもある [Ziegler 05]。一般に、利用者が推薦を採用したとき、その結果不満だったときのコストは低い、満足したときの利得は大きい分野では、セレンディピティを重視すべきである。映画や音楽など娯楽に関する推薦では、こうした状況になることが多い。

§3 被覆率

被覆率 (coverage) とは、全アイテムのうち、評価値の予測が可能なアイテムの割合である。評価値の予測が不可能な状況には、以下のようなものがある。協調フィルタリング (7章) では、ほかの利用者の評価値を利用する。すると、誰にも評価されていないアイテムは評価の対象にできない。もう一方の内容ベースフィルタリング (11章) では、アイテムの特徴量が欠損していたり、利用者のプロファイルが未整備である場合には、評価値を予測できない。適合アイテム発見タスクでは、利用者が満足するものが何か見つければよいので被覆率は比較的低くても問題は生じない。評価閲覧が目的なら、評価値のないアイテムが多数あるのは不便なので被覆率は高くあるべきである。適合アイテム列挙タスクでは、推薦すべき対象の見落としは許されない、基本的に被覆率は100%でなければならない。

§4 学習率

嗜好データの増加に伴って予測精度は向上するが、その向上の度合いを学習率 (learning rate) と呼ぶ。これは、実用的な予測精度に達するまでに必要な嗜好データの数で決まる。システム全体を評価する学習率は、嗜好データの総数から計算するが、特定のアイテムや利用者に限定した評価をするための学習率も用いられる。学習率を調整するパラメータをもつアルゴリズムも多いが、学習率を高くしすぎると過学習のため汎化誤差が悪化して、予測精度の向上が不十分なレベルで止まる場合もある。しかし、利用者が評価付けをあまりしない場合や、服飾品など商品のサイクルの早いアイテムでは、過学習の危険性があっても学習率はやや高くすべきである。

§5 推薦の性質に関するトレードオフ

言うまでもなく、正解率やセレンディピティなど上記の規準ですべて良いものが理想的な推薦システムである。しかし、これらの評価規準は、次にあげるようなトレードオフの関係にあり、目的に応じてバランスをとる必要がある。予測精度は、推薦において最も重要な規準だが、これだけでは不十分であることは十分に注意すべきである。[Cosley 03] には、利用者は5段階評価で1段階良く、もしくは悪く改ざんした推薦を見せられると、そのことに利用者は気づき、このような改ざんシステムへの利用者の満足は低いことが報告されている。よって、予測精度は、利用者の満足に影響していることは確かである。だが、計測した予測精度が同じシステムでも、利

利用者の満足には大きなばらつきがあるとも、この文献は報告している。ほとんどの利用者が好むであろう限られたものだけを推薦すれば、予測精度は一般に高くなる。例えば、スーパーマーケットでの買い物で牛乳や卵など、ほとんどの顧客が購入する商品を推薦すると、予測精度の観点からは良い推薦である。だが、当たり前すぎて目新しさはなく、一概に良い推薦とはいえない。新たにシステムを利用し始めた利用者には、システムへの信頼を高めるために、予測精度を重視して確実に好まれるものを推薦するほうが良い。予測精度の評価指標は、評価値を計算できなかったアイテムは無視して計算するのが一般的である。そのため、嗜好データが十分なアイテムだけを推薦対象にすれば予測精度は向上するが被覆率は下がってしまう。また、学習率の向上も過学習などの影響で、予測精度を低下させる場合がある。以上のように複雑なトレードオフの関係があるため、どの指標をどのようなバランスで重視するかは、推薦対象や利用者の目的など多くの要因を考慮して決めなければ、利用者の満足を得られるような推薦システムは設計できない [McNee 06a]。

残念ながら、こうした設計を組織的に行うための研究は始まったばかりである。[McNee 06b]では、Human-Recommender Interaction (HRI) というモデルを提案している。これは、システムと人間のやり取りである推薦ダイアログ (recommendation dialogue)、推薦の傾向を表す推薦器の個性 (recommender personality)、および利用者情報探索タスク (user information seeking task) の三つの点について、それぞれの特徴を記述するための規準を定めている。この規準に基づき、利用者の推薦への要求や推薦アルゴリズムの特徴を記述し、目的に応じて適切な対応付けをする HRI 解析プロセスモデルを提唱している。

3.2 推薦候補の予測に関する制約

推薦候補を予測するために必要な、データや計算機資源は無限ではなく、何らかのトレードオフを考慮しつつ推薦アルゴリズムを選択する必要がある。よって、これらの制約や条件についてまとめる。

§1 嗜好データの制約

嗜好データの最も顕著な特徴は非常に疎 (sparse) なことである。すなわち、非常に多くのアイテムが存在するが、利用者が評価しているのはごく一部で、その他のアイテムへの評価値は欠損している。具体的には、評価値があるのは全体の1～0.001%のオーダーである [Ben Schafer 01]。また、欠損は均一ではなく、Zipfの法則 [徳永 99] のように、被評価数の順に被評価数ごとのアイテム数を整列すると、被評価アイテム数は指数的に減少する現象が見られる [Weigend 03]。こうした疎なデータからの予測は困難である。また、詳しくは5.2節で述べるが、嗜好の評価値は、統制されていない環境で採取され

た心理的な量なので、揺らぎが大きく、評価のたびに変化して不整合を生じる問題もある。その他、利用者数とアイテム数の比率は予測精度に影響する [Herlocker 04] ので、実際の運用状況に合わせてテストをすべきである。最後に嗜好データの更新の問題がある。推薦システムは運用中に随時嗜好データが追加される。また、新たに利用者やアイテムがデータベースに追加されることもある。こうした変化に応じて予測モデルを更新する必要がある。平滑化などを用いた予測技術を使うと、疎なデータでも比較的安定的な予測ができるが、計算量が増えて予測モデルの更新を頻繁に実行できず、これらの変化に対応できなくなるといった問題もある。

§2 その他の制約や条件

データ数が多数であるにもかかわらず、高速な予測が要求されるスケーラビリティは重要な問題である。利用者数は10万～100万、アイテム数は10～100万、利用者当たりの評価数10～1000という大規模なデータにもかかわらず、10～1000の要求に対して、10～100ミリ秒の時間で応答することが要求される [Ben Schafer 01, Linden 03]。このような高いスケーラビリティを達成しつつ、正確に予測することも困難な課題である。

ほかに、推薦を利用する状況の問題もある。例えば、レストランの推薦システムでは、一人で食べに行く場合と、家族で食べに行く場合は異なった推薦をすべきだろう。こうした推薦をする状況や利用者の暗黙的な要求を考慮するかどうかは大きな要因となる。また、利用者がどれくらい詳細な推薦を求めているかも、アルゴリズムを選択するときに考慮すべき事柄である。すなわち、利用者の嗜好に適合か不適合の2段階程度の大まかなものでよいのか、購入のための意思決定のため、いろいろな評価項目について利用者の嗜好への適合度を詳細に要求するのかといった違いである。

4. 推薦システムの実行過程

ここまでは、推薦システムを分類し、設計にあたって考慮すべき事項について述べてきた。ここでは、推薦システムがどのように実現されるかを見ていこう。推薦システムは、図2のように、嗜好データの獲得、嗜好の予測、そして推薦の提示の三つの段階で推薦を行う。これはO-I-Pモデル (Output-Input-Process model) [Konstan 03] とも呼ばれる。以下、これらの各段階の概要について述べる。

4.1 データの入力

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。活動利用者は自身の嗜好データ (preference data) を推薦システムに入力する。嗜好データとは、いろいろなアイテムについて

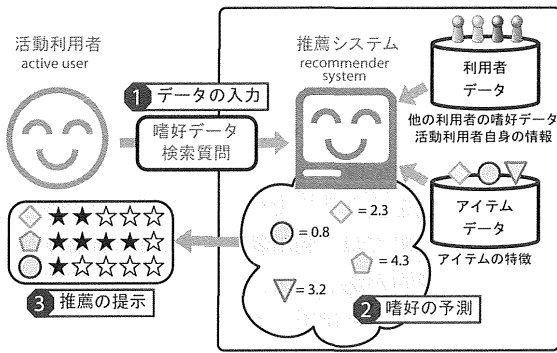


図2 推薦システムの実行過程 (O-I-P モデル)

の関心や好みの度合いを数値化したデータである。この嗜好データの代わりに、関心のあるアイテムについてのより具体的な記述を検索質問 (query) として、活動利用者に入力させるシステムもある。また、推薦には、ほかの利用者の嗜好データ、アイテムの特徴データ、活動利用者自身の情報、推薦の状況や目的なども利用される場合があり、これらの情報を収集する場合もある。この段階については5章で述べる。

4.2 嗜好の予測

活動利用者の嗜好データに加え、収集しておいた利用者のほかの情報やアイテムの情報を利用して、活動利用者がまだ知らないアイテムへの、活動利用者の嗜好を予測する。嗜好の度合いを数値として予測する手法や、単に好きか嫌いかの識別をするだけの手法がある。実現手段としては、機械学習の手法を用いるものや、人手によるルールを用いるものがある。この段階については6～12章で述べる。

4.3 推薦の提示

予測した嗜好に基づいて、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。このために、2.2節や2.3節のいろいろな目的に応じた表示形式の変更や、3.1節の各種指標のバランスを調整するためのアイテムの選別や順位付けの変更などを行う。この段階については13章で述べる。

5. データの入力

ここでは、推薦システムの実行過程の最初の段階である「データの入力」について述べる。この段階では、活動利用者に自身の嗜好データを、推薦システムへ入力させる。嗜好データとは、利用者の各アイテムへの関心や好みの度合いを数量化したものである。システムによっては、この嗜好データの代わりに、活動利用者に検索質問 (批評 (critique) ともいう) を入力させるものもある。この検索質問は、アイテムの特徴についての制約条件を具体的に記述したものである。例えば、レストランの推

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
未評価と不支持の区別	○: 明確	×: 不明確
利用者の認知	○: 認知	×: 不認知

薦システムで、価格帯や、和洋中の別などを具体的に指示するために「価格は6,000円以下で、和食の店」といった形式の検索質問を入力する。こうした検索質問は、情報検索やデータベースのクエリ検索の技術がほぼ転用できるので、ここでは、嗜好データについて述べる。さらに、これら嗜好データや検索質問で表された、活動利用者の嗜好パターン以外のデータも推薦システムは利用する。このようなデータとして、活動利用者以外の利用者の嗜好データ、アイテムの特徴、利用者の年齢や性別などの情報、現在位置などの利用状況を示す情報などがあり、これらについても述べる。なお、嗜好データの収集全般については、[土方 04] にまとめられている。また、[Swearingen 01] は、推薦システムの利用者へのアンケート調査結果に基づいて、入力インタフェースについての設計指針を示している。アイテムについての情報は、利用者が評価するときに見えるようにしておくと、システムへの満足が高まると報告している。

5.1 暗黙的と明示的な嗜好データの獲得

まず、嗜好データを獲得するアプローチは、大きく暗黙的と明示的な2種類に分けられる。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。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、購入したり、閲覧したりしたアイテムには、利用者は関心があるとみなしたりする。

まず、二つの嗜好データの獲得法を比較する。これらの獲得法の長所と短所を表1にまとめた。データ量については、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ほうが有利となる。しかし、質問に答えるといった手間を利用者は嫌うことが多いため、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。

データの正確さについては、暗黙的な獲得では、誤ってクリックしてしまったり、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。

利用者に明示的に評価してもらう場合では、アイテムを利用者が評価したかどうかはもちろん明確である。しかし、暗黙的な評価では、利用者がそのアイテムに対し

て積極的な行動をしなかったことをもって、そのアイテムへの不支持とみなす場合がある。例えば、閲覧しなかったアイテムは好きではないとみなしたとする。このとき、アイテムについて未評価であることと不支持であることの区別ができない。場合によっては、閲覧していないために、利用者が好むアイテムを嫌いだとシステムがみなすこともある。

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ほうが受け入れられやすい。暗黙的な獲得では、嗜好データを意識的に入力していないので、推薦が根拠なくなされたもののように感じられやすく不利である。

5.2 明示的な獲得

アイテムを利用者に提示し、利用者にそのアイテムに対する好みの度合いを答えてもらう明示的な嗜好データの獲得について詳細を述べる。

§1 評価の動機付け

明示的な獲得法では、利用者はアイテムを評価することを面倒だと思うので、暗黙的な方法に比べて多数の嗜好データを集めにくいと述べた。[Swearingen 01]では、利用者は推薦の精度が向上するなど、評価付けによるメリットが明確であれば、ある程度の手間をかけて評価付けをするとの調査結果を報告している。よって、利用者に評価をさせるような動機付けは重要である。

自身への推薦の精度を向上させるということは、利用者にとって主な動機付けとなる。だが、管理者が想定するようなこの動機以外にも、自分の意見の表明をするためやほかの利用者の手助けになるということを動機とする場合 [Herlocker 04] もある。これらの動機は、利用者の評価数の順位の公開などによって喚起することができる。さらに、明示的にインセンティブを与えることも考えられる。[Melamed 07] は、情報検索の結果の順位付けに、ほかの利用者の評価を利用する、一時的個人化の推薦システムを提案している。このシステムでは、市場の考えが導入され、検索結果を閲覧するには、ポイントの支払いが必要である。閲覧する文書の被評価数が多く評価が高いほど多くのポイントを支払う必要がある。一方、ポイントは検索した結果を評価することで獲得でき、被評価数が少なく高い評価をすると獲得ポイント数は増える。すると、検索をするために検索結果の評価をする必要が生じるため、積極的な評価付けが期待できる。

§2 採点法と格付け法

利用者が好みの度合いを答えるには、それを測る尺度が必要になる。好みの度合いを表す尺度として、0～5や-3～+3のような数値尺度を使う採点法 (scoring method) や、上・中・下や適合・不適合などの順序付きカテゴリ尺度を使う格付け法 (rating method) [竹内

89] がよく利用されている。こうした方法は人間の聴覚や味覚などを定量的に計測する官能検査 (sensory test) の分野で研究されてきた [佐藤 85]。採点法や格付け法は、単純な入力フォームを用いて、比較的多数のアイテムに対する嗜好データを得られることが利点である。

これらの方法を使ううえでの注意点をいくつか述べておく。[Cosley 03] では、利用者は評価尺度の目盛りが細かいほうが好む傾向があること、さらに、細かい評価で予測精度が向上することはないが悪くもならないことを報告している。よって、目盛りは細かくに設定することを推奨している。さらに、-3～+3の尺度で、中立の0を抜いた尺度を使うと、中立の評価の多くは弱い肯定的な評価+1に移されること、予測評価値を見せながら評価させると、利用者はそれに「引きずられた」評価をすることも報告している。2.3節で述べた適合アイテム発見を目的とする場合、目的に適合・不適合の2段階でも十分な場合が多いが、評価閲覧タスクでは、どれくらい不要なアイテムを除外したいかは利用者次第なので、より詳細な多段階の尺度を用いるほうが良いだろう。次に、質問の仕方にもいろいろな配慮をすべきである。例えば、採点法では等間隔の尺度を連想させるように、等間隔の目盛りを見せるなどの工夫がある。これらの配慮については [中森 00] を参考にされたい。

§3 評価値の揺らぎや偏り

採点法や格付け法は大量の嗜好データを比較的に容易に得られるので多用されてきた。しかし、当然ながら欠点もある。先に、明示的な獲得は暗黙的な獲得と比べてより正確に利用者の嗜好を評価できると述べた。だが、絶対的には不正確さや揺らぎがある。真の嗜好の度合いは、脳の活動を直接観測するなどすれば将来的には計測できるであろうが、現在のところは厳密には計測できない。そのため、揺らぎがあるかどうかの直接的な証明はできない。よってここでは、採点法や格付け法によって計測した評価値が、真の評価値と乖離している間接的な証拠と、その乖離の原因を示す。

まず、評価値の揺らぎの証拠を示す。官能検査の研究では、たとえ被験者が同じ評価値を与えていても、人によって嗜好の強さが違っていたり、時間が経つと一貫性が保たれなくなる問題があることが経験的に知られていた [Luaces 04]。ソムリエなど訓練された被験者が、同一セッション内、すなわち時間をあけずに連続して評価した場合でなければ、尺度を一定に保つことは難しいとされている。嗜好データについても、一度映画を評価付けしたあと、6週間後にもう一度同じ被験者に同じ評価付けさせると、二つの評価値の間の相関は0.83であったとの報告 [Hill 95] がある。[Cosley 03] でも同様の報告がされている。同一セッション内でも、寿司の嗜好について採点法で尋ねたのち、無関係な質問をいくつかしてから、下記の順位法で再び同じアイテムについて嗜好を質問すると、68.3%の被験者の回答に不一致が観

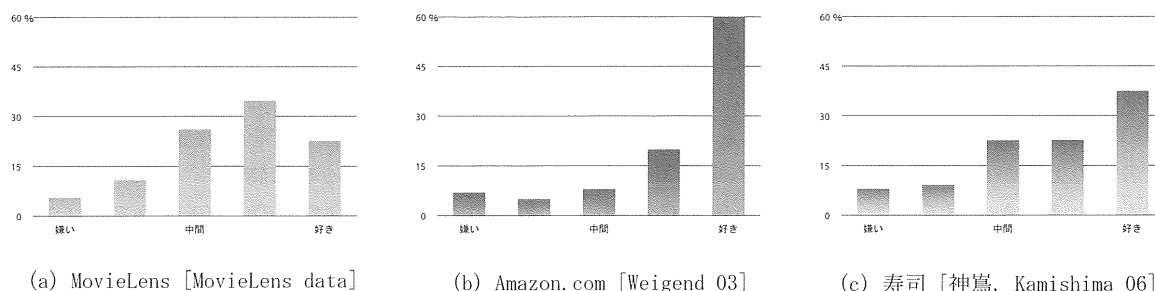


図3 アイテムへの評価値の分布

測された[神島 04]. これらの実験結果は、嗜好データには揺らぎがあることを示している. ほかに、代表的な GroupLens データにおいて、いろいろな工夫をしても、平均絶対誤差 (MAE) を 5 段階尺度で 0.73 の「魔法の壁 (magic barrier)」より小さくできないことから、評価値そのものに揺らぎがあることが示唆されている [Herlocker 04]. 以上のように、絶対的な評価値を使う採点法や格付け法では、被験者は、質問時期の違いなどにより揺らぎが生じるといえるだろう.

次に、評価値の偏りについて述べる. 図 3 に、5 段階の採点法を用いた 3 種類の嗜好データの、評価値の分布を示す. それぞれ、(a) MovieLens の 100 万要素のデータ集合 [MovieLens data], (b) 電子商取引サイト Amazon.com [Weigend 03], (c) 寿司の嗜好調査 [神島, Kamishima 06] での分布である. どのデータでも、好きなほうへ明らかに偏っている. この偏りの原因には、サンプリングと真の嗜好からの乖離の二つが考えられる. サンプリングの偏りの原因として、図 3 (a) や (b) では、利用者が関心がある選択的にアイテムを評価していることや、図 (b) や (c) では、市場の淘汰を受けて人気のあるアイテムのみが評価候補となっていることがあげられる.

このようなサンプリングの偏りは、3・1 節 § 1 で指摘したように、予測誤差の正確な評価を妨げる. 真の嗜好から乖離する理由としては、利用者個人がもつ心理効果の影響が考えられる. 例えば、過剰な酷評は社会的通念的に良くないとの考えをもつ人には、全体的に評価が高くする寛大効果 (leniency effect) が見られ、曖昧な判断や質問では中心のスコアを選びやすい中心化傾向 (central tendency) などが生じる [中島 99]. さらに、質問の仕方による影響も考えられる. 例えば、尺度の一方だけが連続して選ばれるように質問を配列すると偏りが生じる場合がある [中森 00]. しかし、推薦システムでは、推薦結果の提示と嗜好データの収集を兼ねるため、利用者が好むと予測される順にアイテムを並べ評価付けさせることがよく行われる. すると、高い評価値が高頻度で連続してしまう. このように、設計上の制限により、偏りを生じするような質問の仕方をしてしまうという問題もある.

§ 4 順序の利用

そこで、採点法や格付け法以外の調査方法の利用が考えられる. [Herlocker 99] では、利用者の類似度評価

においてスコアの順位関係だけを考慮することや、利用者ごとの平均評価値を 0 に正規化することで予測精度が向上することを報告している. このことは、採点法で得た評価値の絶対的な値ではなく、相対的な大小が重要であることを示唆しているといえるだろう. また、採点法や格付け法で得られる量は、本質的には大小関係にのみ意味がある順序尺度 [Stevens 51, 鷺尾 98] であると指摘されている [中森 00]. そこで、好きなものから嫌いなものへ順に、複数の対象を並べるという順位法を利用する「なんとなく協調フィルタリング」[Kamishima 03, Kamishima 06] を著者は提案した. 少なくとも調査したデータにおいて、順位法の採用で予測精度が向上した. ただし、順位法にも問題点がある. 同時に多数のアイテムを整列するのは難しいので、大量の嗜好データをまとめて得ることは難しい. また、評価は常に相対的で、絶対的な評価は得られない. そのため、相対的に良いものを選ぶような意思決定には役立つが、絶対的な評価が求められる評価閲覧タスクなどには向かない.

官能検査の調査方法としては、取り出したアイテム対のどちらが良いかを指定する一対比較法 (paired comparison) や、いくつかの候補の中から最良のものを指定させる択一法 (method of choice) などもある. これらの方法についての研究は著者はまだ知らず、今後の研究が待たれる.

5・3 暗黙的な獲得

暗黙的な嗜好データの獲得では、アイテムに関連した、利用者の行動に基づいて、そのアイテムについての嗜好を判断する. 利用者があるアイテムを閲覧したり、購入したりすると、これらの行動はそのアイテムへの潜在的な肯定を示していると考えられる. また、購入は閲覧よりもより強い肯定を示すとも考えられる. このような行動による潜在的な嗜好の強弱を Nichols は論じている [Burke 02, Nichols 97]. 強い嗜好を表すものから順に次のような行動をあげている.

- | | | |
|------------------|---------------|---------------------|
| 1. Purchase | 2. Assess | 3. Repeated Use |
| 4. Save/Print | 5. Delete | 6. Refer |
| 7. Reply | 8. Mark | 9. Terminate Search |
| 10. Examine/Read | 11. Consider | |
| 12. Glimpse | 13. Associate | 14. Query |

[Cho 07]では、閲覧、ビデオのプレビュー、購入の3種類の行動それぞれを暗黙的な肯定入力と考え、それぞれの行動別に利用者間の類似性を計算している。

ほかの暗黙的な獲得法として次のようなものがある。推薦リストの上位からA, B, C, …とアイテムを閲覧してCを選んだとき、AやBを見たにもかかわらずCを選択したことから、AよりC、BよりCを好むという相対的嗜好順序を得る方法 [Joachims 02] が提案されている。閲覧するという行為だけでなく、その時間を計測することで、より詳細な情報を得る方法などもある。また、新たに入力装置を導入して、利用者の行動情報を収集し、そこから暗黙的に嗜好データを獲得する試みもある。例えば、マイクで収集した発話内容 [高間 07] や、アイカメラを使って求めた注視領域 [吉高 07] などを利用する試みなどである。

5.4 嗜好データのその他の要因

嗜好データで推薦に影響するその他の要因をあげておく。利用者が初めてシステムを利用するときに、特定のアイテム群について明示的に質問して、嗜好データを集めることが考えられる。すべての利用者が共通に評価しているアイテム群があると、7章の協調フィルタリングでは利用者間の嗜好の類似性を評価しやすくなる利点がある。しかし、音楽のようにその場で少し聞かせて評価できるようなものならよいが、映画などは見たことがないものは評価しにくい。よって、こうした共通アイテム集合を利用できるかどうかは推薦するアイテムの種類に依存する。

評価したときの時間情報も利用できる。服飾品のように流行の影響がある場合には、時間が経った嗜好データはあまり有効ではないだろう。また時間の前後関係に依存する9.4節のような嗜好の予測方法もある。単純な好き・嫌いではなく、Zagatのレストランガイドのように、味・サービス・内装といった項目ごとに分けて嗜好データを収集することもできる。こうしたシステムとしては [Adomavicius 07] がある。

5.5 嗜好データ以外のデータ

嗜好データ以外の、推薦に利用されるデータについてまとめておく。

§1 アイテムの特徴

アイテムを特徴ベクトルで記述したデータで、内容ベースフィルタリングでは必須である。推薦対象がテキストである場合は Bag-of-words モデルで tf-idf 重み [徳永 99] が一般に使われる。この場合多数の特徴量が得られるが、個々の特徴の推薦への寄与は小さい。一方、推薦システムの設計者が意図的に選択した特徴は、多数は得られないが、個々の特徴はより推薦に寄与する。こうした特徴には、映画の場合では監督や制作年、ラーメンであればスープや麺の太さといったものがあげられる。こ

のように、明確な特徴のほかに、アンケート調査などで獲得した印象などを特徴にする場合もある。映画の例では「悲しい」や「楽しい」といった印象語への採点法による評価の平均評価値などを特徴量として利用できる。

§2 デモグラフィックな特徴

利用者の年齢や性別などの、デモグラフィックな情報^{*4}も利用できる [Pazzani 99]。これらの情報は嗜好と関連があると考えられ、マーケティングにおいてもデータベースマーケティングとして利用されてきた [Ben Schafer 01]。デモグラフィックな特徴があれば、新規利用者に対しても推薦が可能になる利点があるが、プライバシー問題の観点から、その収集が困難である問題がある。この問題に対しては、データの利用目的を説明すること [Swearingen 01] や、14.2節のプライバシー協調フィルタリングの導入といった対処がある。

§3 利用状況の特徴

推薦システムを利用する状況の情報である。レストランを利用する場合などでは人数や場所の情報は推薦の制約になる [小野 06, 山本 06]。また、システム側のコンテンツとして、商品の在庫や納期の情報など、推薦時には考慮すべき情報である。

今回は推薦システムの概要とその分類について述べ、その後、推薦システムの実行過程と、その第1段階であるデータ入力について述べた。2008年1月号の次回では、2種類の嗜好の予測手法である内容ベースフィルタリングと協調フィルタリングの特徴について6章で述べ、続く7～10章で協調フィルタリングによる予測アルゴリズムについて述べる。

◇ 参考文献 ◇

- [Adomavicius 07] Adomavicius, G. and Kwon, Y.: New recommendation techniques for multicriteria rating systems, *IEEE Intelligent Systems*, Vol. 22, No. 3, pp. 48-55 (2007)
- [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)
- [Ben Schafer 01] Ben Schafer, J., Konstan, J. A. and Riedl, J.: E-commerce recommendation applications, *Data Mining and Knowledge Discovery*, Vol. 5, pp. 115-153 (2001)
- [Bishop 06] Bishop, C. M.: *Pattern Recognition and Machine Learning*, Springer (2006)
- 邦訳：元田浩ほか訳：パターン認識と機械学習—ベイズ理論による統計的予測 (上下), シュプリンガー・ジャパン (2007, 2008)
- [Burke 02] Burke, R.: Hybrid recommender systems: Survey and experiments, *User-Modeling and User-Adapted Interactions*, Vol. 12, No. 4, pp. 331-370 (2002)
- [Cho 07] Cho, J., Kwon, K. and Park, Y.: Collaborative filtering using dual information sources, *IEEE Intelligent Systems*, Vol.

^{*4} 日本語では個人情報ともいうが、後に述べる秘匿しなければならない情報という意味と区別するため、ここではこのように呼ぶ。

- 22, No. 3, pp. 30-38 (2007)
- [Cosley 03] Cosley, D., Lam, S. K., Albert, I., Konstan, J. A. and Riedl, J.: Is seeing believing? how recommender interfaces affect users' opinions, *Proc. SIGCHI Conf. on Human Factors in Computing Systems*, pp. 585-592 (2003)
- [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)
- [Duda 00] Duda, R. O., Hart, P. E. and Stork, D. G.: *Pattern Classification*, Wiley-Interscience, second edition (2000)
- [Goldberg 92] Goldberg, D., Nichols, D., Oki, B. M. and Terry, D.: Using collaborative filtering to weave an information tapestry, *Commun. ACM*, Vol. 35, No. 12, pp. 61-70 (1992)
- [Hastie 01] Hastie, T., Tibshirani, R. and Friedman, J.: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer (2001)
- [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)
- [Herlocker 04] Herlocker, J. L., Konstan, J. A., Terveen, L. G. and Riedl, J. T.: Evaluating collaborative filtering recommender systems, *ACM Trans. on Information Systems*, Vol. 22, No. 1, pp. 5-53 (2004)
- [土方 04] 土方嘉徳: 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)
- [Joachims 02] Joachims, T.: Optimizing search engines using clickthrough data, *Proc. 8th Int'l Conf. on Knowledge Discovery and Data Mining*, pp. 133-142 (2002)
- [神島] 神島敏弘: 寿司データ: <http://www.kamishima.net/sushi/>
- [Kamishima 03] 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)
- [神島 04] 神島敏弘: なんとなく協調フィルタリングー 順序応答に基づく推薦, 人工知能学会研究会資料, SIG-KBS-A304-37 (2004)
- [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)
- [Konstan 03] Konstan, J. A. and Riedl, J.: Recommender systems: Collaborating in commerce and communities, *Tutorial at ACM CHI2003* (2003)
- [Lawrence 99] Lawrence, S., Lee Giles, C. and Bollacker, K.: Digital libraries and autonomous citation indexing, *IEEE Computer*, Vol. 32, No. 6, pp. 67-71 (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)
- [Luaces04] Luaces, O., Bayón, G. F., Quevedo, J. R., Díez, J., del Coz, J. J. and Bahamonde, A.: Analyzing sensory data using non-linear preference learning with feature subset selection, *Proc. 15th European Conf. on Machine Learning*, pp. 286-297 (2004) [LNAI 3201]
- [McNee06a] McNee, S. M., Riedl, J. and Konstan, J. A.: Accurate is not always good: How accuracy metrics have hurt recommender systems, *Proc. SIGCHI Conf. on Human Factors in Computing Systems*, pp. 1097-1101 (2006)
- [McNee 06b] McNee, S. M., Riedl, J. and Konstan, J. A.: Making recommendations better: An analytic model for human-recommender interaction, *Proc. SIGCHI Conf. on Human Factors in Computing Systems*, pp. 1103-1108 (2006)
- [Melamed 07] Melamed, D., Shapira, B. and Elovici, Y.: MarCol: A market-based recommender system, *IEEE Intelligent Systems*, Vol. 22, No. 3, pp. 74-78 (2007)
- [元田 06] 元田 浩, 津本周作, 山口高平, 沼尾正行: データマイニングの基礎, オーム社 (2006)
- [MovieLens data] MovieLens data, <http://www.grouplens.org/node/12#attachments>
- [村上 07] 村上知子, 森紘一郎, 折原良平: 推薦結果の意外性を評価する指標の提案, 人工知能学会全国大会 (第 21 回) 論文集, 2C5-2 (2007)
- [中島 99] 中島義明 ほか編: 心理学辞典, 有斐閣 (1999)
- [中森 00] 中森義輝: 感性データ解析—感性情報処理のためのファジィ数量分析手法, 森北出版 (2000)
- [Nichols 97] Nichols, D. M.: Implicit rating and filtering, *Proc. 5th DELOS Workshop on Filtering and Collaborative Filtering*, pp. 31-36 (1997)
- [小野 06] 小野智弘, 黒川茂莉, 本村陽一, 麻生英樹: 携帯電話ユーザのための映画嗜好モデル構築と映画推薦システム, 人工知能学会研究会資料, SIG-FPAI-A602-07 (2006)
- [Pazzani 99] Pazzani, M. J.: A framework for collaborative, content-based and demographic filtering, *Artificial Intelligence Review*, Vol. 13, pp. 393-408 (1999)
- [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 Cooperative Work*, pp. 175-186 (1994)
- [Resnick 97] Resnick, P. and Varian, H. R.: Recommender systems, *Commun. ACM*, Vol. 40, No. 3, pp. 56-58 (1997)
- [佐藤 85] 佐藤 信: 統計的官能検査, 日科技連 (1985)
- [Shardanand 95] Shardanand, U. and Maes, P.: Social information filtering: algorithms for automating "Word of Mouth", *Proc. SIGCHI Conf. on Human Factors in Computing Systems*, pp. 210-217 (1995)
- [Stevens 51] Stevens, S. S.: Mathematics, Measurement, and Psychophysics, Stevens, S. S. ed.: *Handbook of Experimental Psychology*, John Wiley & Sons (1951)
- [Swearingen 01] Swearingen, K. and Sinha, R.: Beyond algorithms: An HCI perspective on recommender systems, *ACM SIGIR Workshop on Recommender Systems* (2001)
- [高間 07] 高間康史, 難波広樹, 岩瀬徳宏, 服部俊一, 武藤優樹, 庄司俊寛: テレビ視聴時の情報推薦に基づくヒューマン・ロボットコミュニケーション, 人工知能学会全国大会 (第 21 回) 論文集, 2D5-5 (2007)
- [竹内 89] 竹内 啓 編: 統計学辞典, 東洋経済 (1989)
- [徳永 99] 徳永健伸: 情報検索と言語処理, 言語と計算シリーズ, 第 5 巻, 東京大学出版会 (1999)
- [鷲尾 98] 鷲尾 隆, 元田 浩: 尺度の理論, 日本ファジィ学会誌, Vol. 10, No. 3, pp. 401-413 (1998)
- [Weigend 03] Weigend, A. S.: Analyzing Customer Behavior at Amazon.com, in *Invited Talk at KDD2003* (2003)
- [山本 06] 山本浩司, 片上大輔, 新田克己, 相場 亮, 桑田 仁: 地図上の情報推薦システムにおける投稿情報の信頼度, 人工知能学会論文誌, Vol. 21, No. 3, pp. 276-286 (2006)
- [吉高 07] 吉高淳夫, 脇山孝貴, 平嶋 宗: 絵画を例とした鑑賞活動の共有に基づく情報推薦, 人工知能学会全国大会 (第 21 回) 論文集, 2B4-8 (2007)
- [Ziegler 05] Ziegler, C.-N., McNee, S. M., Konstan, J. A. and Lausen, G.: Improving recommendation lists through topic diversification, *Proc. 14th Int'l Conf. on World Wide Web*, pp. 22-32 (2005)

2007 年 9 月 10 日 受理

著者紹介

神島 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。