

個人情報保護を考慮した オンラインショッピング支援システムの提案

川口 賢二 加藤 由花 箱崎 勝也

電気通信大学 大学院情報システム学研究科

概要

近年、インターネットが爆発的に普及し、インターネットを用いたサービスが注目を浴びている。その中の1つにインターネットショッピングがあり、その取引額や店舗数が拡大している傾向にある。ユーザにとって有用な情報を提供することが、インターネットショッピングの支援として有効な手段の1つであるが、従来からのシステムにはインターネット上でサービスを行う際に個人情報を求なくてはならないという問題が内在している。しかし、個人情報を保護して欲しいという要望が高い。

そこで本稿では、個人情報を保護した上で、オンラインショッピングを支援できるシステムの提案を行う。

A proposal of an online shopping support system based on explicit privacy protection.

Kenji Kawaguchi Yuka Kato Katuya Hakozaki

Graduate School of Information Systems, University of Electro-Communications

Abstract

The Internet are deployed over a wide area, Internet services have become more remarkable. Online shopping is in one in it, it is in the tendency which the amount of dealings and number of shops have expanded. It is considered to be one of the effective means to give useful information to an online shopping user. However, conventional systems have a problem that a user must register personal information.

This paper proposes an online shopping support system based on explicit privacy protection.

1 はじめに

近年、インターネットが爆発的に普及し、インターネットを用いたサービスが注目を浴びている。その中の1つにインターネットショッピングがあり、その取引額や店舗数が拡大している傾向にある。ところが、現在3万店以上あるといわれるインターネットショッピングサイトのうちほとんどのサイトは、一部の大手サイトもしくは優良サイトと呼ばれている小規模サイトを除き苦戦が続いている状態である。そうした売上が伸びていない大多数の問題点として、インターネットでなくては実現できない特徴をうまく生かしてあげることが挙げられる。それゆえ、インターネットショッピングサイトは、実店舗で与えることのできないサービスをユーザに提供できるように改善する必要がある。

2 インターネットショッピングサイトの改善手法

2.1 従来のサイト改善手法

ショッピングサイトの改善策として、クローラーを効率よくホームページに呼び込めるようにする検索エンジン最適化(SEO)やバナー広告がある。これらはいかに多くの人に見てもらえるかが改善の主な目的である。ところが、アイテムの種類にもよるものの、インターネットショッピングではサイト訪問者は衝動買いをしない。よって、ショッピングサイトが行う改善策として最も優先度を高くしなくてはならないのは、店舗が便利であると認識してもらうことで

ある。それにはユーザにとって検索機能が有用であることや、アイテムについて提示された情報を有用なものとする必要がある。その実現方法の一つとして推薦システムがある。推薦システムはユーザに最適な商品情報をリストの中から提示するシステムである [1, 2]。

2.2 従来からのシステムの問題点

従来からある代表的な推薦システムの例として、Movie Lens [3] や Amazon.com [4] がよく知られている。これらは、Information Filtering や Collaborative Filtering を用いて推薦する機能を実現している。Movie Lens では、ユーザが評価したビデオに基づいて、ユーザが好きそうな映画を推薦するサービスを提供している。また、Amazon.com では、類似ユーザが購買した商品や高く評価した商品に基づいて推薦するサービスを提供している。これらの推薦システムをショッピングサイトへ適用するときに共通した問題点として以下の3点が挙げられる。

1. プライバシー保護が困難である。
2. ユーザの嗜好に対して正確性と潜在性のバランスの取れた推薦ではない。
3. ユーザの嗜好の変化に適応できない。

1つ目の問題はユーザが高品質な推薦とプライバシーの保護の両方の実現を望んでいるためである。高品質な推薦

を行うためには、大量の個人情報をサーバ側においておく必要があるため、この問題が起こる。2つ目は推薦の本質に関わる問題である。推薦はユーザに対して有益な情報を提示することにある。それゆえ、ユーザが知っているような商品情報ばかり提示しては、推薦サービスを利用する意味がない。よく購入しているものが欲しいのか、または意外性のあるアイテムを提示して欲しいのかを推薦処理時に考慮する必要がある。3つ目は、従来から提案されている推薦システムに共通する問題である。従来から提案されている推薦システムは、ユーザの個人情報を大量に得ることで、高品質な推薦を目指しているためである。

上記の問題に対して、我々は、プライバシーを保護した上で、ユーザの嗜好に対する正確性と潜在性を融和した情報を提供できる推薦手法の研究を進めてきた [5]。しかし、評価実験の被験者から、被験者の意識にあるアイテムがあまり推薦されないことに対する指摘があった。従って、本稿で提案するシステムに導入した推薦手法は、ユーザの嗜好特性に対して潜在性を保ちつつ正確性を向上させることを目的としている。

3 システムの提案

本章では、本稿で提案するシステムの詳細について説明する。まず、提案システムに導入した手法を 3.1 節から 3.4 節に示す。その後で、システムの処理の流れを 3.5 節で明らかにする。

3.1 プライバシーの保護

プライバシーを保護した上での情報提供を実現するアプローチの1つとして、エージェントを用いる方法がある [6]。ユーザの PC や携帯端末上のユーザの個人情報を持たせたエージェントに情報を収集させて、店舗などから情報を受け取る場合にはユーザの個人情報を匿名化したプロフィールに基づき商品情報の選別を行う仕組みである。この方法の例として、感性語データベースを用いた手法が挙げられる。

矢野ら [7] は、消費者の商品選択を主観的なイメージ語と客観的なイメージ語を用いてモデル化し、それから商品情報を推薦する手法を提案している。矢野らの手法はユーザの感覚に適した推薦を目的としている。

エージェントを用いた手法は、ユーザの個人情報を匿名化して登録することによりプライバシーの保護を実現しているが、インターネットユーザの持つオンラインショッピングに対する不安感はオンラインショッピング全体に対する漠然としたものである。従って、オンラインショッピングにおけるプライバシーの保護はユーザに個人情報が守られていることを明示しなければならない。

提案システムでは、それらを同時に実現するため、推薦処理をユーザ PC 上でを行い、個人情報を一切サーバに送らないことで解決する。このことによって、ユーザの知らないところで勝手にユーザの個人情報が利用されることがなくなる。ここでの個人情報はユーザの購買特性情報である。その例を図 1 に示す。また、商品リストはテキストファイルで送信する。

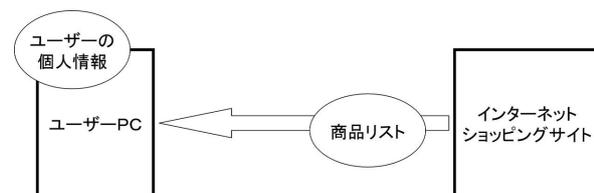


図1 プライバシー保護

3.2 マッチング選択

廣岡ら [8] は、順マッチングと交差マッチングを用いた手法を提案している。順マッチングは購買履歴のあるカテゴリ A 内のユーザ履歴から生成したユーザの履歴キーワードベクトル d とカテゴリ A 内 m 個の商品別特徴キーワードベクトル $a_i (i = 1, 2, \dots, m)$ との比較を行う方法である。順マッチングのイメージを図 2(a) に示す。一方、交差マッチングは購買履歴のあるカテゴリ A 内のユーザ履歴から生成したユーザの履歴キーワードベクトル d と A 以外のカテゴリ B 内 n 個の商品別特徴キーワードベクトル $b_j (j = 1, 2, \dots, n)$ との比較を行う方法である。交差マッチングのイメージを図 2(b) に示す。情報を提供する手段として電子メールを用いたシステムを構築し、評価結果から順マッチングではユーザの意識にあるような嗜好に対する正確性の高い情報を推薦できること、交差マッチングではユーザの意識にないが興味のあるような嗜好に対する潜在性の高い情報を推薦できることが明らかにされている。

しかし、Web 環境で用いた場合、2つの方法を組み合わせる用いることが想定されていないため、サイト訪問者に対し偏った情報を与えることになる。そこで提案システムでは、ユーザの過去の履歴から、ユーザの嗜好に対して正確性の高い情報と潜在性の高い情報を提示する個数を確率的に求めることで、ユーザの嗜好に対して推薦する情報の正確性と潜在性を融和させる。順マッチングの選択確率 P_{order} は、

$$P_{order} = (I_{order} + C_{order}) / Sum \quad (1)$$

として求め、交差マッチングの選択確率 P_{cross} は、

$$P_{cross} = 1 - P_{order} = (I_{cross} + C_{cross}) / Sum \quad (2)$$

として求める。 I_{order} と I_{cross} はそれぞれ正確性と潜在性の初期値である。 C_{order} と C_{cross} はそれぞれ正確性と潜在性の履歴で選択された個数である。 Sum は総数 ($= I_{order} + I_{cross} + C_{order} + C_{cross}$) である。

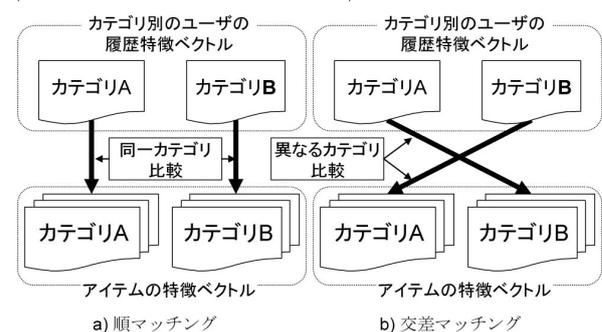


図2 順マッチングと交差マッチング

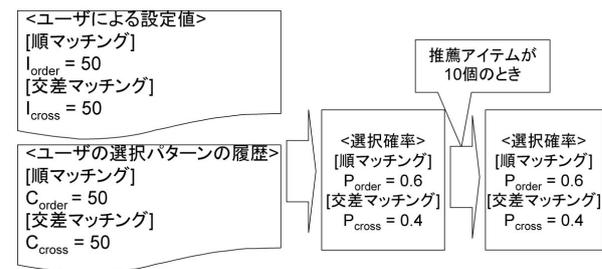


図3 マッチング選択の例

3.5 推薦処理の流れ

推薦処理の流れは、以下の1から8のようになり、例を図7に示す。この処理はすべてユーザPC上で行われる。

1. ユーザのアクションにより、処理の開始が開始する。
2. 3.4節で説明した方法に基づき、GAを用いて特徴ベクトル作成を行う。
3. 2で求めた特徴ベクトルとカテゴリ別のユーザの履歴特徴ベクトルの比較から、ユーザが最も好むと推測されるカテゴリを1個選択する。
4. 3.3節で説明したフィルタリング処理を行い、推薦対象とするカテゴリを決定する。
5. 3.2節で説明したマッチング選択を行い、順マッチングと交差マッチングを適用する個数を求める。
6. 類似度を用いて推薦するアイテムを決定する。
7. 推薦するアイテムをHTMLページを作成する。
8. ユーザが出力されたHTMLページを閲覧する。

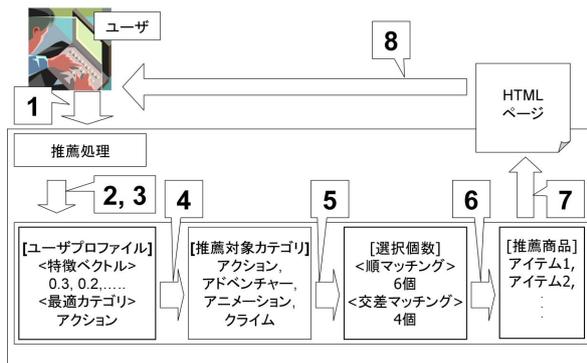


図7 推薦処理の流れ

4 提案システムの実装

4.1 提案システムの構成

提案システムは、クライアントであるユーザPCとサーバである店舗から成り、対象としている商品はビデオである。ユーザの個人情報はサーバ側に置かない。サーバとクライアントの構成をそれぞれ以下に示し、システム全体の構成を図8示す。

(1) クライアントの処理

クライアントの処理は、推薦処理、商品リストのアップデートである。このとき、クライアントは処理の過程でユーザを特徴できる情報をサーバに送らない。

(2) サーバの処理

サーバの処理は、商品リストの保持、クライアントから要求される商品の情報と画像の送信である。

4.2 提案システムを実装した環境

システムの実装は以下のソフトウェア構成で研究室内に構築した。実装したシステムの環境を図9に示す。

<クライアント>

OSとして"WindowsXP", ブラウザとして"Internet Explorer6.x", Javaの実行環境として"JRE1.4.x"を用いる。

<サーバ>

OSとして"RedHatLinux9", Javaの実行環境として"JRE1.4.x", "データベースサーバとして"MySQL4.0.x"を用いる。

提案システムの対象としているアイテムはビデオである。各ビデオのタイトルには1つ以上の内容を示すジャンルが付与されていて、ビデオの特徴ベクトルはジャンルに基づいて生成している。

提案システムの操作は、Java言語で実装したコントローラを用いて行う(図10)。

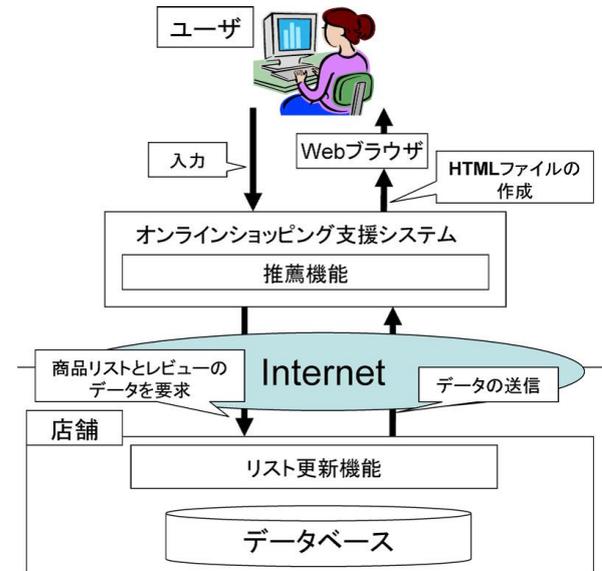


図8 システム構成

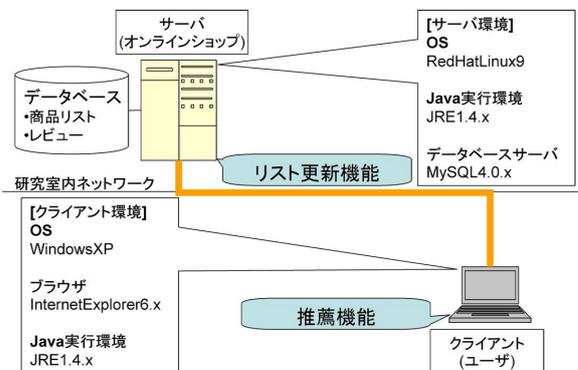


図9 実装した環境

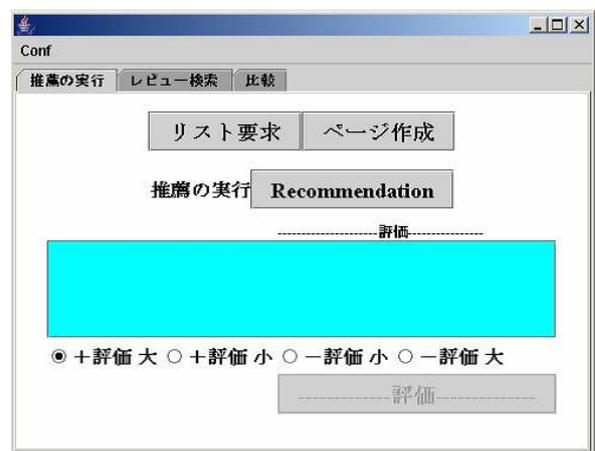


図10 コントローラ

4.3 提案システムの操作例 推薦機能の操作方法を説明する。

(1) 前処理

前処理は、設定フォームを用いて操作を行う。

a. 初期データの入力

初期データは、各カテゴリからランダムに1つのアイテムを提示し、ユーザからの評価を求める。ユーザに要求する評価の方法は、「+評価」と「-評価」の2種類であり、評価の重みはそれぞれ「+1」と「-1」となる。操作手順を以下に示す。操作の例を図11に示す。

1. 「データの初期化」をボタンを押す。
2. 別ウインドウにアイテムが表示される。
3. 評価した値をラジオボタンで選択し、「初期データの評価」ボタンを押す。

b. 正確性と潜在性の初期値

順マッチングと交差マッチングの割合の初期データの入力をユーザに求める。スライダを用いて正確性と潜在性の初期値を選択する。スライダで0から100までの値を選択するが、0のとき正確性が100%となり、100のとき潜在性が100%となる。操作手順を以下に示す。操作の例を図12に示す。

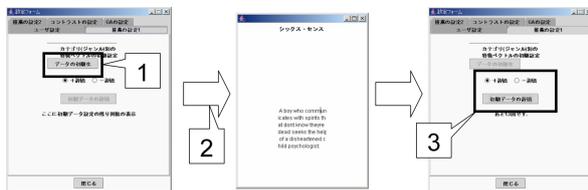


図11 初期データの入力例



図12 正確性と潜在性の初期値の入力例

(2) 推薦処理の実行と推薦された商品の閲覧

推薦処理の実行の手順を以下に示し、操作の例を図13に示す。

1. コントローラ内の推薦の実行タブを選択し、「Recommendation」のボタンを押す。
2. ブラウザの「お気に入り」から「MainPage」を選択する。
3. 「MainPage」の「推薦ページ」をクリックする。
4. 推薦されたアイテムから、閲覧したいアイテムを選択する。
5. アイテムの詳細を表示するページに移動する。

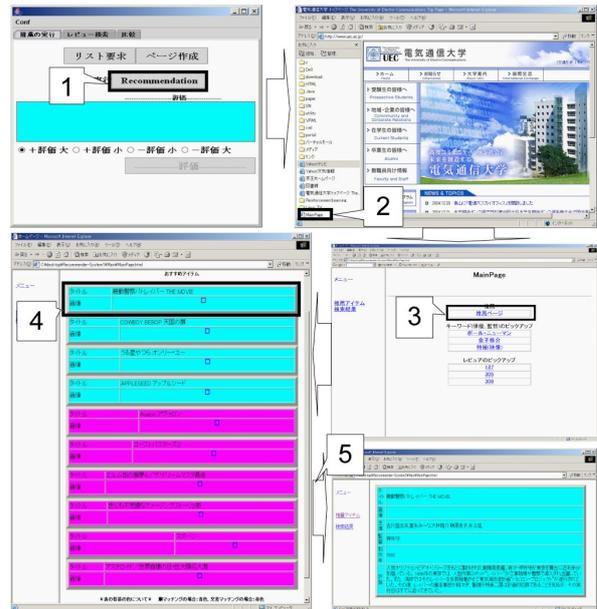


図13 推薦処理の操作例

5 評価

推薦手法の有効性を確認するために行った実験結果について示す。提案手法の効果を明らかにするために、「GAを用いない」、「3回の推薦処理の内1回だけGAを用いる」、「推薦処理の毎回でGAを用いる」の3種類の推薦方法の比較を行う。GAを適用している場合と適用していない場合に分けているのは、GAを用いた場合にユーザの嗜好に対して潜在性に影響を与えずに正確性が向上しているか確認するためである。また、GAを適用する間隔を変化させている理由は、GAを用いることで処理の負荷に対する懸念が考えられるためである。

推薦機能の実験は、10個の推薦結果の提示と10個の推薦結果に対する評価を1セットとし、合計9セット行う。比較のため、3セットごとに初期(1~3)、中期(4~6)、後期(7~9)で主観評価を記入を求めた。主観評価は、3セットにおける正確性と推薦の潜在性について、5段階で行った。被験者は20代男性10人である。正確性の結果を図14に示し、潜在性の結果を図14に示す。また、GAのパラメータは実験的に求めていて、交叉率を0.6、突然変異率を1/64としている。

図14から、GAを用いた場合は、潜在性がGAを用いていない場合との比較結果から、潜在性を保ったまま正確性を向上させていることが分かる。また、GAを用いる頻度であるが、毎回用いるの方が、正確性と潜在性の推移の両方でわずかであるが、評価が高いといえる。

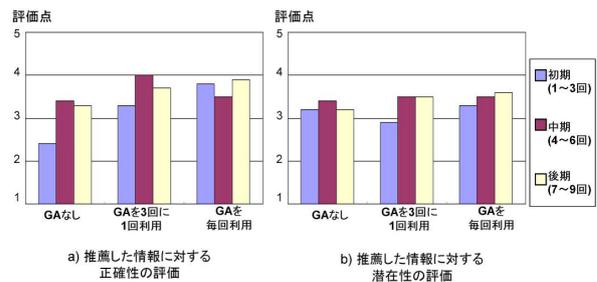


図14 評価結果

6 考察

6.1 推薦処理について

本稿では、インターネットショッピングにおいて、より有効な推薦システムの構築を目指した。それぞれの手法の特徴を以下に述べる。

(1) プライバシー保護

近年、インターネットユーザのプライバシー保護に対する意識が高まっていて、従来の方法がユーザに対して暗黙的であることを指摘した。提案システムではプライバシー保護として、クライアントに処理の大部分を割り与えていて、ユーザに対するサービスを行う過程で、明示的に個人情報を出さないことをユーザに示せる利点がある。

(2) マッチング選択

マッチング選択の方法は、順マッチングを用いて推薦する個数と交差マッチングを用いて推薦する個数をユーザの過去の選択傾向から求め、ユーザの嗜好に対して推薦結果の正確性と潜在性の価値を制御することで、より質の高い情報の提示を目的としている。

(3) フィルタリング処理

従来の方法の問題点として、ユーザに一定期間ごとに登録した情報の更新をする負担を要求すること、内容に基づいて情報をふるい落とすことが困難であることが挙げられる。提案システムに導入したフィルタリング処理は、ジャンルに基づいたカテゴリ別の履歴特徴ベクトルからクラスタリングを行い、最もユーザの嗜好に合ったカテゴリの属するクラスタのみを推薦対象とする方法である。この方法により、上記で挙げた2点の問題点を解決している。

(4) 感性にあうユーザプロフィールの作成

推薦処理は過去のユーザの履歴を元に処理を行うが、履歴の量とユーザの嗜好の変化から、ユーザの嗜好に最適なユーザプロフィールの作成が難しい。そこで、感性に合うユーザプロフィールの作成を行うため、GAを用いて特徴ベクトルを作成する方法を導入した。また、GAを用いることから処理の負荷に対する懸念があるが、1回の処理に掛かるGAの処理時間は4秒程度であり、現在の一般ユーザのPC環境ではそれほど問題にならないと考えられる。

6.2 システムの特性について

提案システムでは、クライアント側から個人情報を出さないシステム構成にすることでプライバシー保護を目指した。このとき、以下に示す2点について考慮する必要がある。

(1) システムの処理量

従来型のシステム構成ではサーバ側のみ負担がかかり、処理量を減らす努力が必要であったが、提案システムでは、クライアントにほとんどの処理を割り与えているため、システムの推薦処理における処理能力がユーザの利用環境に依存することになる。しかし、多くのユーザは、処理能力や記憶容量が潤沢にある現在のPCを最大限使用しては、PCへの負荷は少ないと考えられる。また、近年のブロードバンド化により、プレーンなテキストで書かれた1000タイトルのリストファイルが19.2KBであり、特徴ベクトルのリストファイルが116KBであるから、それらを送信してもあまりネットワークに負担がかからない。そのため、提案手法において、そうした懸念にそれほど注意を払う必要がないと考える。

(2) システムの汎用性

従来型の推薦システムは、登録された個人情報をもとにサーバ上で推薦処理が行われるため、推薦を適用する対象やシステムの利用において汎用性がないと言える。ところが、提案システムでは、推薦処理を実行するアプリケーションをインストールして処理を行うため、ただ1つのユーザプロファイルがあればよい。それゆえ、ユーザ登録を行わなくても、推薦サービスを受けることが可能となる。また、サービスを提供するショッピングサイトは、推薦サービス導入において、推薦の実行を行うアプリケーションの配布を行えばよく、導入コストが軽減されることになる。

7 まとめ

本稿では、ユーザのプライバシーを明示的に保護した上で、ユーザの嗜好特性に対して潜在性の価値を保ちつつ正確性の価値を向上させる手法の提案を行った。さらに、実装したプロトタイプシステムの評価結果から、提案したシステムの有効性を確認した。今後は、運用を見据えて、商品数と商品カテゴリのデータベースを拡張して検証していく予定である。

参考文献

- [1] P. Resnick and H.R. Varian. Recommender Systems. *Communications of the ACM*, Vol. 40, No. 3, pp. 66-72, 1997.
- [2] Saranya Maneeroj, Hideaki Kanai, and Katuya Hakozaiki. An Improved Recommendation Method for Better Filtering Information out of Database. *情報処理学会論文誌:データベース*, Vol. 43, No. SIG5(TOD 14), pp. 66-74, 2002.
- [3] MovieLens. <http://movielens.umn.edu/>.
- [4] Amazon.com. <http://www.amazon.com/>.
- [5] 川口賢二, 加藤由花, 箱崎勝也. インターネットショッピングにおけるユーザの利用状況に基づいた商品情報推薦システム. *情報処理学会 DICO MO シンポジウム*, pp. 481-484, 2004.
- [6] Bradley N. Miller, Joseph A. Konstan, and John Riedl. PocketLens: Toward a Personal Recommender System. *ACM Transactions on Information Systems*, Vol. 22, No. 3, pp. 437-476, July 2004.
- [7] 矢野絵美, 北野有亮, 末吉恵美, 篠原勲, ピンヤボンシニナット, 加藤俊一. 消費者の感性モデルを利用したレコメンデーションシステムの構築. *情報処理学会論文誌:データベース*, Vol. 44, No. SIG8(TOD 18), pp. 46-54, 2003.
- [8] 廣岡康雄, 寺野隆雄, 大塚雄吉. 意外性の高い情報を提供するリコメンダーシステム. *人工知能学会研究報告*, No. SIG-KBS-9904-11, pp. 61-66, 2000.
- [9] 寺野隆雄. Web上の情報推薦システム. *情報処理*, Vol. 44, No. 7, pp. 696-701, 7 2003.
- [10] Zhiwen Yu and Xingshe Zhou. An Adaptive Assistant for Personalized TV. *IEEE Transactions on Consumer Electronics*, Vol. 50, pp. 393-399, 2004.
- [11] 中西泰人. 選考関数を用いた対話型進化システムの制御と評価-遺伝的プログラミングのデザイン支援システムへの応用. *人工知能学会誌*, Vol. 13, No. 5, pp. 36-43, 1998.