

## グループ嗜好モデルと視聴履歴を利用したコンテンツ検索サーバの試作

河村晃好<sup>†</sup> 黒武者健一<sup>†</sup> 佐藤亮一<sup>†</sup> 芥子育雄<sup>†</sup>

シャープ(株) 情報通信システム開発研究所<sup>†</sup>  
〒639-1186 奈良県大和郡山市美濃庄町 492 番地  
Tel: 0743-55-4049 Fax: 0743-52-1659

E-mail: {kawamura, kuro, sato, kesi}@isl.nara.sharp.co.jp

**あらまし** 膨大な量のデジタルコンテンツが蓄積されるようになり、ユーザが興味を持っている情報を選択的に探し出すのは困難になってきている。ユーザが欲しい情報を得るためには、検索についての専門的な知識や技術も必要になってくる。そのため、ユーザの嗜好を学習し、嗜好に合致した情報だけを推薦するリコメンダシステムが注目されている。本論文では、簡単なアンケートを元にユーザグループを決めた後、グループ嗜好モデルを利用して検索キーワードを拡張するコンテンツの検索方式と視聴履歴による検索とを融合した嗜好類推型検索方式を提案する。さらに、ユーザにお薦めコンテンツを推薦する検索サーバのプロトタイプについて紹介する。

**キーワード:** 推薦, 連想検索, グループ, 嗜好モデル, 履歴, 自己組織化マップ

### **A Prototyping of Retrieval Server System based on the Group Preference Model and the Operation Histories.**

Akiyoshi Kawamura<sup>†</sup> Ken'ichi Kuromusha<sup>†</sup> Ryoichi Sato<sup>†</sup> Ikuo Keshi<sup>†</sup>  
Information & Communication Systems Development Laboratories,  
SHARP Corporation<sup>†</sup>

#### **Abstract**

Recently, it has been getting difficult to find out the information in which a user is interested from many accumulated digital contents. In order for users to acquire the information which they need, the special knowledge and technique about retrieval are also needed. Therefore, there is growing interest in recommender systems which can learn user preferences automatically and recommend only the information corresponding to user interests.

In this paper, we introduce a prototype of retrieval server system which can recommend for users based on a search method combined group preference model from initial profiles provided by users and content-based filtering by extracting user preferences from user's operation histories to improve the recommendation accuracy.

**Keywords:** Recommend, Associative Retrieval, Group, Preference Model, SOM

## 1. はじめに

デジタルコンテンツの蓄積が加速度的に増加しており、ユーザが興味を持っている情報を選択的に探し出すのは困難になってきている。そのため、ユーザの嗜好を学習し、嗜好に合致した情報だけを推薦するリコメンダシステム[1]が注目されている。ユーザの嗜好に沿った情報を推薦する一番簡単な方法は、ユーザにアンケート形式で事前に興味あるキーワードを登録してもらい、登録されたキーワードが含まれる情報を推薦することが考えられる。しかし、単純にこの方法だけを用いれば、登録されたキーワードに関連する情報だけしか推薦できず、ユーザの嗜好が時間の経過や、その他の外的要因によって変わった場合、再度キーワードを登録しなおす必要がある。加えて、一人のユーザが一度に入力できる情報は限られているため、少ない入力情報からユーザの嗜好を抽出するには十分ではないと考えられる。一方、協調フィルタリングを用いたリコメンダシステムも提案されている[2]。協調フィルタリングは、情報に対する索引付けが必要ないという利点があるが、他の利用者の評価に基づいて推薦を行うため、誰からも評価されていない未知の情報は推薦することができないという欠点もある。また、ユーザの購入履歴や行動履歴を使って商品や情報を推薦するシステムも存在しているが履歴がない状況ではユーザの嗜好を知ることはできない。

本論文では、今回試作した検索サーバのプロトタイプについて紹介する。この検索サーバでは、グループ嗜好モデルとユーザの視聴履歴からユーザの嗜好を類推し、連想検索を用いて推薦情報の検索を行う。グループ嗜好モデルを利用することにより、少ないアンケート入力でもユーザの嗜好モデルを類推できると共に、登録されたキーワードには全く関係ないがユーザの興味を誘発するような情報の推薦と、履歴が無い新規ユーザへの推薦も可能となる。さらに、視聴履歴から構築したユーザの嗜好モデルもユーザプロフィールに反映することにより、ユーザの嗜好の変化にも自動で追従した情報の推薦も実現できる。

## 2. 意味ベクトルと連想検索技術

意味ベクトルは、多くの FEATURE との意味的な関係をベクトル表現したものである。 $n$ 個の概念分類を FEATURE とし、各次元が一つの FEATURE に対応した

$n$ 次元ベクトル空間上の一点で、意味を表現するものである。例えばベクトル  $X = (x_1, \dots, x_n)$  の各要素を 2 値で表す場合は式(1)のようになる。

$$x_i = \begin{cases} 1 & (\text{FEATURE}_i \text{と関係あり}) \\ 0 & (\text{FEATURE}_i \text{と関係なし}) \end{cases} \quad (1)$$

例えば、FEATURE として{人間, 悲しい, 芸術, 科学, 興奮, 政治}を採用した場合には、単語「パイロット」の意味ベクトルは(1,0,0,1,1,0)となる。本論文では、このように各 FEATURE を関係あり、なしの2値で表現している。FEATURE との関係性を2値で表現することで、分野に依存しない汎用的な意味ベクトル辞書を構築できると考えている。

意味ベクトルの FEATURE の付与は、検索の対象となるコンテンツの内容を記述したテキストデータ(以下、コンテンツデータ)を用いたブートストラップ学習[3], [4]によって行っている。FEATURE として266種類の概念分類を利用し、人手で作成した2万語の基本単語の意味ベクトルを元に、コンテンツデータに適応させると共に、コンテンツデータ中に出現する基本単語に含まれていない未知語に対する意味ベクトルの生成を自動的に行う。また、意味ベクトルの生成と同時に、各単語の出現頻度による重みも計算する。コンテンツデータに含まれている全レコード数(1コンテンツ=1レコード)を  $N$ 、単語  $j$  を含むレコードの数を  $df_j$  とすると、単語  $j$  の重み  $W_j$  は式(2)のように、コンテンツデータ中の単語の出現確率の逆数によって決める。

$$W_j = \log(N/df_j) \quad (2)$$

次に、意味ベクトル DB について説明する。コンテンツデータを意味ベクトルを用いて検索するために、予めコンテンツデータに含まれる各レコードを意味ベクトルに変換して、意味ベクトル DB ファイルを作成しておく。レコードの意味ベクトルは、レコードから抽出された各単語の意味ベクトルの和として計算される。レコードに含まれる各単語の意味ベクトルを  $T_j$ 、その単語の重みを  $W_j$  とすると、レコードの意味ベクトル  $X$  は、

$$X = k \sum_j W_j T_j \quad (3)$$

と計算される。ただし、 $k$  はレコードのベクトルの大きさを一定にするための係数である。このようにしてコンテンツデータの各レコードに対して意味ベクトルを作成し、

意味ベクトル DB ファイルを生成する。

連想検索は、質問文と検索対象(レコード)との意味ベクトルによる距離と単語含有率による距離の両方を用いて検索結果のランキングを行っている。意味ベクトルによる検索は、検索質問文から式(3)を用いて質問文のベクトル  $X_A$  を計算しておき、意味ベクトル DB 中のそれぞれのベクトル  $X_i$  との内積  $SAV_i$  を計算することで行われる。

$$SAV_i = X_A \cdot X_i \quad (4)$$

この内積値が大きいくほど、質問文との距離が近いレコードであるということになる。また、キーワード検索は、

$$SAW_i = \frac{\sum_{j_i} W_{j_i}}{\sum_{j_A} W_{j_A}} \quad (5)$$

を計算することによって行われる。分母は検索質問文に含まれる単語の重みの和を表し、分子は検索対象 DB テキストのレコード  $i$  に含まれる単語の中で検索質問文にも含まれている単語の重みの和を表している。つまり、検索質問文に含まれている単語を多く含むほど、その中でも重みの大きい単語を多く含むほど  $SAW_i$  の値が大きくなることになる。そして、これらの 2 つの結果を組み合わせて、連想検索の得点は、

$$SA_i = SAV_i + \alpha \cdot SAW_i \quad (6)$$

で計算される。ただし、 $\alpha$  は両者の和の比率を調整する係数である。 $\alpha$  の値が小さいほど、意味ベクトルによる検索の度合いが高くなる。この得点  $SA_i$  の順にレコードをランキングしたものが連想検索の検索結果となる。

### 3. 自己組織化マップを用いたグループ嗜好モデルの構築

ユーザの嗜好モデルは、年代や性別ごとにユーザが持つ嗜好の偏りによってある程度典型的なユーザグループの嗜好モデルが存在すると考えられる。ここでは、嗜好に関するアンケート調査結果から、自己組織化マップを用いて典型的なグループの嗜好モデルを構築する方法について説明する。

#### 3.1 嗜好調査アンケートについて

ユーザグループの嗜好モデルを構築するために、7つの年代層(小学生以下, 高校生以下, 18~20代, ..., 60代以上)の男女別20人ずつの計280人に

対して嗜好調査のアンケートを行った。調査項目は年代, 性別などの基本属性と, 一般的趣味・嗜好を表すキーワード, TV番組のジャンルを表すキーワードの計357項目のキーワードについて, 2週間分のTV番組について行った。評価は, A:とても興味がある, B:興味がある, C:気になるの3段階で興味の度合いをチェックしてもらい, 何もチェックしない場合は全く興味がないと判断した。

#### 3.2 自己組織化マップについて

自己組織化マップ(Self-Organizing Map)[5]はニューラルネットワークの教師なし競合学習モデルであり, 雑然とした情報群の中からいくつかの特徴を見つけ出し, その特徴によってデータを分類し, 通常2次元空間に写像することができる。アルゴリズムを以下に示す。

**Step1.** 入力データをベクトルでパターン化する

$$X_i = (X_{i1}, X_{i2}, \dots, X_{in}) \quad 1 \leq i \leq p$$

**Step2.** 出力層の各ユニットのベクトルパターンを初期化する

$$M_j = (M_{j1}, M_{j2}, \dots, M_{jn}) \quad 1 \leq j \leq q$$

**Step3.** ランダムに入力データ  $X_i$  を選択する

**Step4.** 入力データ  $X_i$  に最も近いパターンを持つ出力ユニット  $U_c$  を探す

$$\|X_i - M_c(t)\| = \min\{\|X_i - M_j(t)\|; j = 1, 2, \dots, q\} \quad (7)$$

$\|\cdot\|$  は距離を表し, ユークリッドノルム等が用いられる

**Step5.**  $U_c$  とその近傍ユニット  $N_c(t)$  を  $X_i$  に近づける

$$M_j(t+1) = M_j(t) + (t)(X_i - M_j(t)) \quad \text{if } j \in N_c(t) \\ = M_j(t) \quad \text{else} \quad (8)$$

$\alpha(t)$  は学習回数  $t$  に関して単調減少関数を用いる。

**Step6.**  $i = i + 1$  ( $i \leq n$ ) として3~5を繰り返す

**Step7.**  $t = t + 1$  ( $t \leq T$ : 総学習回数) として  $N_c(t)$  と  $\alpha$  を小さくしながら3~6を繰り返す

**Step8.** 各入力ベクトルをパターンが一番近いユニットに配置する

この結果, 入力ベクトル空間で近いものが, 2次元空間でも近傍のユニットに射影された写像が完成する。

#### 3.3 グループ嗜好モデルの構築

アンケート調査結果から7つの年代層の男女別計14個の基本グループごとにグループ嗜好モデルを構築し

た。図1はグループの嗜好モデル構築処理のフローを示している。まず、各基本グループについて、アンケートの回答結果から、ユーザが興味を示したキーワードを抽出し、各ユーザの嗜好ベクトルを求める。嗜好ベクトルは266次元の意味ベクトルで表され、式(3)の方法で求められる。このとき、ユーザが強く興味を示したキーワードは単語の重みを大きくして嗜好ベクトルに強く反映されるようにする。

次に、自己組織化マップで嗜好ベクトルを入力データとして、嗜好ベクトルを分類し、グループ嗜好マップを形成する。自己組織化マップの特徴から各ユニットには同じような嗜好の特徴を持った嗜好ベクトルが集まっている。すなわち、それぞれのユニットが似通った嗜好モデルを持ったユーザの集合であり、この集合をグループの嗜好モデルとみなすこともできる。

最後に、嗜好グループの特徴単語として各嗜好グループに帰属するユーザのアンケート調査結果から、357項目のキーワードに対する評価値の平均が高いものN個を選択し嗜好キーワードとした。例えば、20代/女性のある嗜好グループの嗜好キーワードは、「音楽鑑賞、芸能/タレント、ショッピング、カラオケ、温泉、アミューズメント」である。このようにして全14個の基本グループそれぞれについて嗜好グループを形成し、各嗜好グループの特徴を表す嗜好キーワードを集めたグループ嗜好DBを構築した。

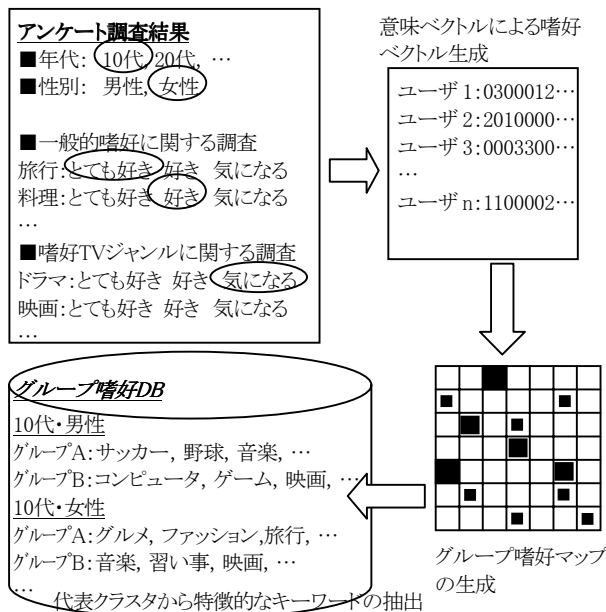


図1 グループ嗜好モデル構築フロー図

#### 4. 視聴履歴によるユーザ嗜好モデルの構築

図2にユーザ嗜好モデル構築フロー図を示す。まず、ユーザが視聴した個々コンテンツ視聴履歴情報をジャンルなどコンテンツの内容を表すような付加情報を使って予め履歴情報を分類し、ユーザの視聴傾向を大まかに分析する。例えば、ジャンルが「映画」である履歴情報が多く溜まっていれば、このユーザは「映画」に興味を持っていることが容易に推測できる。次に、多く視聴されているジャンルの履歴情報を自己組織化マップを用いてさらに細かく嗜好分析を行う。自己組織化マップの学習用の入力ベクトルとしては、各視聴履歴情報を式(3)によって意味ベクトルに変換した履歴ベクトルを用い、履歴の特徴に応じた履歴マップを生成する。意味ベクトルを元にした履歴ベクトルを学習用の入力ベクトルとすることで、意味的に近い内容を持った履歴情報が同一または隣接のユニットに分類される。例えば、ジャンルが「映画」でも「アクション」映画に関するもの、「コメディ」映画に関するものというように内容に応じた分類ができる。履歴が多く分類されているユニットについては、それだけユーザの興味が集中していると考えられる。つまり、この履歴特徴マップは複数の興味を表すユーザの嗜好モデルとみなすこともできる。

意味ベクトルと自己組織化マップを併用することにより、個々の視聴履歴の内容そのものを考慮したより詳細なユーザの嗜好分析が可能となる。最後に、ユーザの嗜好の特徴を強く表した履歴情報をユーザ嗜好情報としてユーザ嗜好DBに記録する。ここでは、多く視聴されたジャンルの履歴マップで、履歴が多く集中しているユニットの履歴情報をユーザの嗜好情報とした。

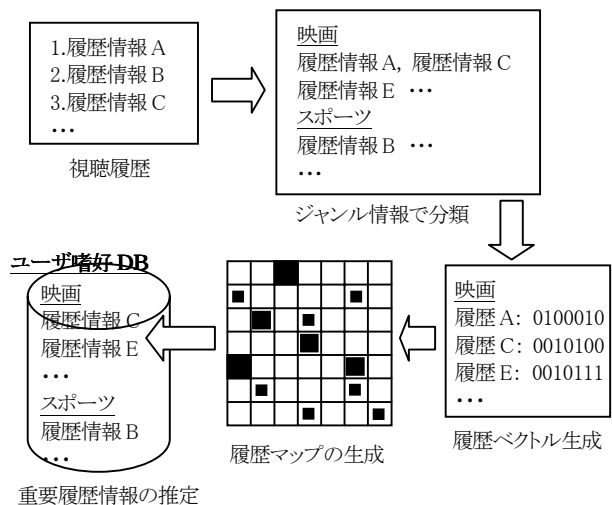


図2 ユーザ嗜好モデル構築フロー図

## 5. お薦めコンテンツの検索

簡単なアンケートを元にユーザの嗜好グループを類推し、グループ嗜好DBを利用して検索キーワードを拡張するお薦めコンテンツの検索と視聴履歴を元にしたユーザ嗜好モデルによる検索を融合した、嗜好類推型コンテンツ検索方式について説明する。

### 5.1 グループ嗜好モデルを使ったユーザの嗜好モデルの類推

ユーザにとってリコメンダシステム側に自分の嗜好情報を入力するのは結構煩わしい作業である。できれば、少ない入力ですべての嗜好情報をシステム側に伝えることができればユーザにとっては有り難いものである。ここでは、グループ嗜好モデルを使って、簡単なアンケート入力だけでユーザ嗜好を類推するための方法を説明する。

まず、ユーザに年代/性別などの基本情報と嗜好に関するいくつかの簡単なアンケートに回答してもらう。アンケートは、「映画」や「スポーツ」など嗜好を表す数個のキーワードを提示して興味の度合いを入力してもらう。例えば、20代/女性のあるユーザがアンケートで以下の2項目のキーワード  $K_u$  について興味があると評価したとする。

$K_u$  = 「旅行/レジャー, コンサート」

このキーワード列  $K_u$  から嗜好ベクトル  $V_u$  を求める。 $V_u$  は式(3)によって生成する。実際には、キーワード個々に対する興味の度合いに応じて単語の重みを大きくすることで、ユーザの嗜好をより反映した嗜好ベクトルが生成できる。この  $V_u$  が簡単なアンケートで得られたユーザの嗜好モデルの1つであると考えられる。

次に、前述のグループ嗜好DBを参照して、ユーザと同じ基本情報に属する嗜好グループを全て抽出する。抽出された嗜好グループの嗜好キーワード列  $K_g$  から、同様に式(3)を用いてグループ嗜好ベクトル  $V_g$  を求める。ユーザの嗜好ベクトル  $V_u$  との内積が一番大きい嗜好ベクトル  $V_g$  を持つ嗜好グループがユーザの嗜好モデルに一番近い嗜好グループと類推できる。

一方、簡単なアンケートでユーザが何も興味を示さなかった場合、すなわち、年代/性別の基本情報だけしか入力されなかった場合、アンケート調査で同じ年代/性別のユーザが一番多く集まった嗜好グループをこのユーザの嗜好グループとすることで、より少ない入力

情報でも嗜好グループの類推の精度を落とさないようにした。簡単なアンケートによって得られたユーザの嗜好モデルはユーザが持つ全嗜好領域の部分空間に過ぎないが、予めアンケート調査によって構築した典型的なグループの嗜好モデルを利用することで、簡単なアンケートだけでは抽出できなかったユーザの他の嗜好領域についての類推が行える。

### 5.2 グループ嗜好モデルを使った拡張検索

簡単なアンケートの回答で得られた嗜好ベクトル  $V_u$  を持つユーザは、ユーザの嗜好グループが興味を示したキーワード  $K_g$  についても興味がある可能性が高いと推測できる。そこで、ユーザがアンケートで興味があると評価したキーワード列  $K_u$  に、ユーザの嗜好グループの嗜好キーワード列  $K_g$  を付加した拡張キーワード検索を行うことで、簡単なアンケートでは類推できなかった、他の嗜好領域についての情報の検索ができる。

例えばユーザの嗜好グループの嗜好キーワードが、 $K_g$  = 「音楽鑑賞, 芸能/タレント, ショッピング, カラオケ, 温泉, アミューズメント」であったとすると、ユーザがアンケートで興味を示したキーワード列  $K_u$  に  $K_g$  を追加したキーワード列を質問文として検索を行う。この例では、「旅行/レジャー, コンサート, 音楽鑑賞, 芸能/タレント, ショッピング, カラオケ, 温泉, アミューズメント」という質問文で連想検索を行う。「旅行/レジャー」「コンサート」に関する情報以外に、「音楽鑑賞」「芸能/タレント」といった情報も推薦できる。このように、アンケートから得られたユーザの嗜好と、ユーザの嗜好に近いと類推されたグループの嗜好とを組み合わせた拡張検索を連想検索を用いて行うことにより、ユーザの嗜好に合致する情報の推薦と、ユーザの興味を誘発するような情報の推薦も行うことができる。

### 5.3 キーワードとコンテンツデータの関連付け

推薦するコンテンツの検索を行う際に、5.2のように一般的なキーワードを質問文として検索を行うより、各キーワードに関連性が高いコンテンツデータを追加した質問文で連想検索を行うことで、ユーザの嗜好をより反映したキーワードと意味ベクトルによる検索ができ、さらに精度が高いコンテンツの推薦ができると思われる。そこで、各キーワードを質問文としてアンケート調査で

用いたテレビ番組の前半1週間分の検索を行い、検索結果が上位の番組でアンケートでの評価が一番高かった1番組のデータをそのキーワードと関連性が高いコンテンツデータとした。これを14個の基本グループごとに357個のキーワードに対して行い、キーワード対応テーブルを作成した。

#### 5.4 グループ嗜好モデルと視聴履歴によるユーザー嗜好モデルを融合したお薦めコンテンツ検索

ユーザーの嗜好の変化にも自動で追従できるような推薦の仕組みを実現するために、視聴履歴の無いユーザーにはグループ嗜好を利用した推薦を行う一方で、視聴履歴が蓄積されているユーザーに対しては視聴履歴に基づくユーザー嗜好モデルも用いることで、より精度の高いお薦めコンテンツの検索を行うような仕組みを実現した。また、グループ嗜好データと視聴履歴によって得られたユーザー嗜好 DB とを予め構築しておき、ユーザーのプロファイルに記録しておくことでリアルタイムにお薦めコンテンツの検索できるようにした。図3はお薦めコンテンツ検索のフローを示している。

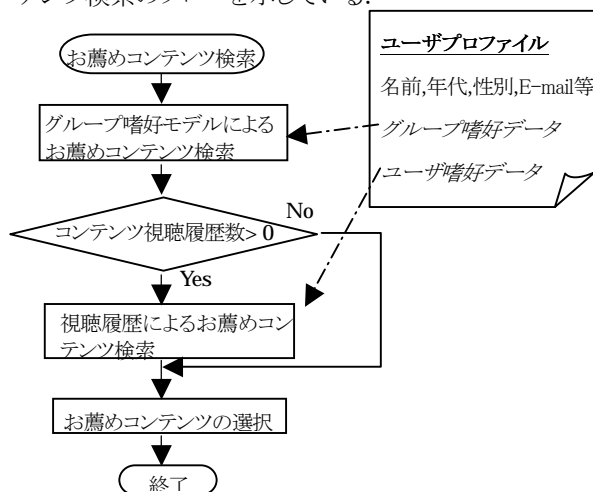


図3 お薦めコンテンツ検索フロー

グループ嗜好モデルによるお薦めコンテンツの検索は前述の通りである。視聴履歴に基づくユーザー嗜好モデルによるお薦めコンテンツ検索は、ユーザー嗜好 DB に記録されているコンテンツ視聴履歴データを連想検索の質問文としてコンテンツの検索を行う。すなわち、よく見られたジャンルでかつ、よく見られた内容に近いコンテンツがお薦めコンテンツとして検索できる。ユーザーに推薦するコンテンツとして、全ての検索結果から、ユーザーの視聴履歴にあるものは除外し、検索の点数が高いものやランキングが上位のものを選択することで、

ユーザーの嗜好により合致した情報だけを推薦できる。

このように、グループ嗜好モデルによるコンテンツ検索は履歴が無いユーザーに対しても推薦が可能となる反面、一般的なキーワードによるコンテンツの検索であるため、ユーザーの嗜好を反映したコンテンツを精度良く検索するのは限界がある。キーワード対応テーブルでキーワードを拡張して検索を行うことである程度解決はできるが、テレビ番組データが基となっているので全てのコンテンツに適応した検索ができるわけではない。

また、視聴履歴に基づいたユーザー嗜好モデルによるコンテンツの推薦は、実際にユーザーが見た視聴履歴情報を用いてコンテンツの検索を行うため、ユーザーの嗜好をより反映したコンテンツの推薦ができる。しかし、視聴履歴が全く無いユーザーに対しては推薦することができないという欠点もある。

このような利点・欠点を持つ2つのコンテンツ推薦方式を融合することで、それぞれの欠点を補完することができる。さらに、履歴が無いかあまり溜まっていない初期段階ではグループ嗜好によるコンテンツ検索を用い、履歴が溜まってくるに従って履歴に基づいたユーザー嗜好によるコンテンツ検索へと徐々にシフトすることによって、使う程に推薦情報の質の向上が期待できる。

#### 5.5 ベンチマーク評価結果

アンケート調査で得られた280人分のテレビ番組データの嗜好結果をベンチマーク(ベンチマーク1)として、年代/性別だけで類推されたグループ嗜好モデルを使った推薦と、簡単なアンケートで類推されたグループ嗜好モデルでの推薦と、視聴履歴によるユーザー嗜好を用いた推薦の3つについて嗜好相関精度を評価した。履歴によるユーザー嗜好を用いた推薦は、各ユーザーについて、アンケート調査で用いた前1週間分のテレビ番組で興味があった番組情報を視聴履歴として、後1週間分の推薦をした場合の嗜好相関を調べた。評価尺度として、以下の2つを用いた。

適合率 = 興味を示した番組数 / 推薦件数

再現率 = 推薦情報内で興味を示した番組数 / アンケートでユーザーが興味を示した全番組数

表1はベンチマーク1での結果を示している。推薦件数は25件、検索対象のコンテンツ総数は844件である。グループ嗜好モデルによる推薦は各キーワードにキーワード対応テーブルで得られたテレビ番組データを拡張

張したものを質問文として検索した結果である。この結果から、グループ嗜好モデルと履歴によるユーザ嗜好モデルを使ってTV番組を推薦すれば、約8~12件興味がある番組が含まれていたことが分かる。年代/性別のみでの推薦でも、約3割の適合率が得られた。

さらに、長期に渡る視聴履歴によるお薦め評価を行うために、別の14人に対してテレビ番組の評価を付けてもらった結果をベンチマーク(ベンチマーク2)として同様の評価を行った。14人には5週間に渡り、毎日その日の詳細な番組内容を読んで評価を付けてもらった。表2は前4週間で興味を持った番組を視聴履歴として5週目の番組を推薦した場合の結果である。表2の結果から、ユーザの興味が良く反映された番組が十分数履歴として残っていれば、適合率約7割の精度でテレビ番組の推薦ができたことが分かる。

表1 ベンチマーク1での嗜好相関精度評価結果

	適合率[%]	再現率[%]
年代/性別のみ	29.2	6.8
簡単なアンケート	32.2	7.4
視聴履歴	46.5	10.8

表2 ベンチマーク2での嗜好相関精度評価結果

	適合率[%]	再現率[%]
視聴履歴	68.1	51.6

## 6. コンテンツ検索サーバについて

図4は検索サーバの概略構成図である。

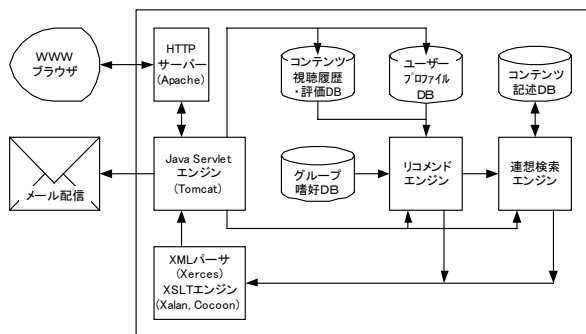


図4 検索サーバの概略構成図

コンテンツやユーザプロフィール及び視聴履歴は全てXML言語で構造的に記述するようにした。また、連想検索技術をXML形式の構造的な記述方式に対応させ、意味検索に加えて、自由な論理演算や数値の大小による検索対象の絞り込みも行えるように拡張した。

リコメンドエンジンは、図3のフローに従ってグループ

嗜好モデルと視聴履歴によるユーザ嗜好モデルを使って、推薦情報の検索を行う。

## 6.1 連想検索エンジンの拡張

図5はXML連想検索エンジンのブロック図である。検索は、検索対象となるXMLデータのどのタグを対象に連想検索を行うのか、ブール検索を行うのか、数値検索を行うのかを指定するために、質問文と検索対象タグを組にした検索質問文XMLデータによって指定する。入力された検索質問文XMLデータがXMLパーサに掛けられ、XMLタグの解析処理を行ってから、それぞれの検索毎の処理を行うことになる。

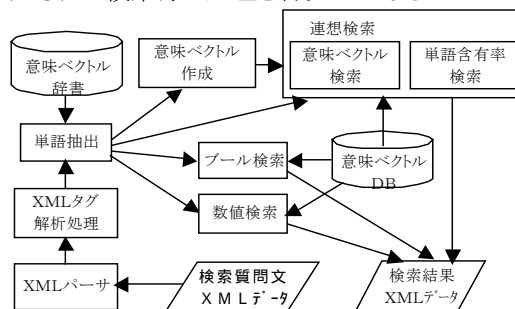


図5 XML連想検索エンジンのブロック図

図6は検索質問文XMLデータの一例である。この検索質問文XMLデータにより、コンテンツ記述データのDateタグ内に20010718以上の数値を含み、かつ、Genreタグ内に「映画」というキーワードを含むものが検索結果の候補となり、さらに、連想検索の質問文として指定されている「派手なアクションもの」による連想検索を行いブール検索・数値検索で絞り込まれた検索結果の候補を意味的に近いものからランキングして出力することができる。お薦めコンテンツの検索は、予めユーザプロフィールに記述されているユーザの嗜好情報を、図6に示した検索質問文XMLデータの連想検索用の質問文タグ部に記述することで、例えば、2001年7月18日以降のコンテンツが検索できる。

```
<RetrievalQuery>
<DatabaseName>MPEG7</DatabaseName>
<RetrievalNumber>10</RetrievalNumber>
<KeywordRetrieval>
<And>
<Keyword match="//Date" pattern="ge">20010718
</Keyword>
<Keyword match="//Genre" pattern="part">映画
</Keyword>
</And>
</KeywordRetrieval>
<ContentsRetrieval>派手なアクションもの
</ContentsRetrieval>
</RetrievalQuery>
```

図6 検索質問文XMLデータの一例

## 6.2 検索サーバの動作例

図7は、ユーザ登録画面例であり、ここで年代/性別、嗜好に関する簡単なアンケートに回答してもらい、プロトタイプでは、アンケート調査で興味が高かった60個のキーワードをアンケートに用いた。



図7 ユーザ登録画面例

また、図8は、コンテンツのお勧め情報を検索する画面例である。ここで「検索」ボタンを押すとリコメンドエンジンが起動し、ユーザプロフィールに記述されているユーザの嗜好情報を元にコンテンツの検索を行いユーザに推薦を行う。この画面で日付、ジャンル、出演者など推薦条件の絞込みを行うための条件も指定できるようになっている。



図8 コンテンツお薦め検索画面例

## 7. おわりに

本論文では、検索サーバのプロトタイプについて紹介した。また、グループ嗜好モデルと視聴履歴を用いたユーザ嗜好モデルを利用した嗜好類推型検索技術を紹介した。少ないアンケート入力でもユーザの嗜好モデルを類推できると共に、登録されたキーワードには全く関係ないがユーザの興味を誘発するような情報の推薦と、履歴が無い新規ユーザへの推薦も可能となった。さらに、コンテンツの検索に視聴履歴を利用したユーザの嗜好モデルも利用することでユーザの嗜好の変化にも自動で追隨した情報の推薦も実現できた。

今後は協調フィルタリングも取り込んで質的により高いコンテンツを優先的に推薦できる仕組みの研究を行うつもりである。

## 謝辞

本研究を進めるにあたり御討論頂いた、情報通信システム開発研究所の坂田安男所長、斗谷充宏チーフに感謝します。なお、本研究の一部は、通信・放送機構の平成12年度通信・放送研究成果展開事業によるものである。

## 参考文献

- [1] Resnick, P., Hal R. Varian: Recommender Systems. In *Communications of the ACM*, vol. 40, no.3, pp. 56-58.(1996).
- [2] Resnick P., Neophytos, I., Miteth, S., Bergstrom, P.,and Riedl, J. GroupLens: An Open Architecture for Collaborative Filtering of Netnews. *CSCW94 Proceedings*, pp. 175-186 (1994).
- [3] 芥子育雄, 乾隆夫, 石倉謙一郎: “大規模データベースからの連想検索,” *信学技報*, AI92-99, pp. 73-80 (1993).
- [4] 芥子育雄, 池内洋, 小淵保司: “意味ベクトルによる百科事典テキストデータベースの構築,” *Proc.Advanced Database System Symposium '93*, ” pp. 227-234 (1993).
- [5] T. Kohonen : “The Self-Organizing Map,” *Proc. the IEEE*, Vol. 78, No. 9, pp. 1464-1480 (1990).