

動的なカテゴリ定義を利用した個人化サービスの実現

岡本 道也[†] 山下 剛士[‡] 鎌原 淳三[§] 下條 真司[◇] 宮原 秀夫[†]

[†] 大阪大学大学院 基礎工学研究科
〒 560-8531 大阪府豊中市待兼山町 1-3
m-okamot@ics.es.osaka-u.ac.jp
miyahara@ics.es.osaka-u.ac.jp

[‡] 西日本電信電話株式会社 ビジネス推進部 [§] 神戸商船大学 情報処理センター
t.yamashita@west.ntt.co.jp kamahara@cc.kshosen.ac.jp

[◇] 大阪大学 サイバーメディアセンター
shimojo@cmc.osaka-u.ac.jp

あらまし

近年、インターネットの普及により、多くの情報に多くのユーザがアクセスできるようになった。今後さらにインターネット上の情報量及びユーザ数が増加していくことが予想され、膨大な情報のフィルタリングを行い、ユーザにとって有益な情報だけを推薦し提供する技術が必要とされている。しかし、既存のフィルタリング手法を用いた推薦システムではコンテンツプロバイダがコンテンツのカテゴリ情報を静的に定義し、それらを全てのユーザに対して共通に用いているため、適切な個人化サービスを行うことができない。

そこで、本稿ではコンテンツプロバイダが静的に定義するカテゴリ情報を動的に変化させる手法の提案を行う。動的なカテゴリ定義を実現するために、ユーザが過去に残したアクションログの解析を行い、その結果からコンテンツ間の類似度を導出し、ユーザ毎に異なる類似コンテンツ集合を導出する。この類似コンテンツ集合に属するコンテンツに対してカテゴリを再定義することにより、動的なカテゴリ定義を実現する。

さらにこの動的カテゴリを導入した「VOD コンテンツ推薦システム」を実装し、実証実験を行った結果から本システム及び提案手法の有効性を示す。

キーワード: 動的カテゴリ, 個人化サービス, コミュニティ, セッション, 推薦システム

Realizing a Personalized Service with Dynamic Category

Michiya Okamoto Tsuyoshi Yamashita Junzou Kamahara
Shinji Shimojo Hideo Miyahara

Abstract

Recently, users can receive a large amount of information easily through network because of the spread of the Internet. In the future, The amount of information and the number of users in the Internet will increase further, and the technologies will be needed which can filter much information and provide users only the information which interests them. But the existing filtering technologies cannot supply approximate personalized services. That is because the services for all users in a system share the static categories which the content providers defined.

In this paper, we propose the method to change the categories of contents dynamically. In our proposed method, the system analyzes the access logs which the users made in the past, calculates the similarities between every two contents, and derives the sets of the similar contents. Then, the system redefines the categories of the contents in the sets of the similar contents.

We implemented “the Recommendation System for VoD Contents” with the idea of dynamic category and examined the effectiveness of our proposed method in the demonstrative experiments.

key words: Dynamic Category, Personalized Service, Community, Session, Recommendation System

1 はじめに

近年、インターネットは急速に普及しており、インターネット上の情報源も急速に膨大化している。ユーザ数も依然として増加傾向にあり、その利用目的も多様化している。またユーザの利用目的の多様化に伴い、インターネット上におけるマルチメディアコンテンツも多様化を見せており、VOD システムなどの新たなインターネットサービスモデルも確立されている。

その結果、ユーザは膨大な情報源におけるニュース、音楽、動画などのマルチメディアコンテンツの中から自分が必要としている情報やコンテンツを確実に獲得するために多大な労力を要するようになってきている。そこでインターネット上での情報検索でユーザにかかる負担を軽減するような技術が求められている。実際に検索サイトやポータルサイトなどのユーザの情報獲得の手助けとなるようなサイトが開設されており、ユーザの嗜好情報に合った情報やコンテンツをユーザに提供する技術による推薦システムなどが実装されている [1-4]。

これらのコンテンツ推薦システムが用いている技術として情報フィルタリングがある。情報フィルタリングを用いると膨大な情報源から情報やコンテンツをユーザに提示する前に情報とユーザとの関連性を計算機が計算し、その関連性を基にユーザが興味を持つような情報だけを抽出しユーザに伝達することができる。この情報フィルタリングを実現するために主に用いられる手法としては content-based filtering と collaborative filtering の 2 つがある [5]。

content-based filtering ではユーザの嗜好情報(ユーザプロファイル)とコンテンツの属性情報や情報に対するキーワードなどとのマッチングを行い、推薦するコンテンツや情報を導出する [6, 7]。collaborative filtering では同様の嗜好を持ったユーザを発見し、その類似ユーザが高い評価を下した情報やコンテンツを推薦する [8]。

我々の研究グループでも情報フィルタリングを取り入れた推薦システムの開発を行ってきており、特に content-based filtering による推薦システムとして「スポーツダイジェスト自動生成システム」[9, 10] や「おすすめ献立配信システム」[11, 12] を実装し実証実験を行った。

この content-based filtering と collaborative filtering は現在最も一般的に採用されている既存のフィルタリング手法であるが、幾つかの問題点も指摘されている。content-based filtering の問題点として、システムを利用しているユーザが複数存在したとしても、ユーザプロファイルとコンテンツの属性情報から個々のユーザに対して推薦するコンテンツを導出するプロセスは独立に行われており、ユーザの興味がユーザプロファイルに反映されていない場合には、ユーザの興味と合致しているコンテンツでもユーザが興味を持っているとは判断されず、推薦することができない。またユーザプロファイルは通常、いくつかの特徴量から成り立っており、ユーザプロファイルによってユーザの興味や関心を忠実に表現するのは不可能であるという問題もある。またコンテンツの属性情報はコンテンツプロバイダによって予め静的に定義され、ユーザ毎

に動的に変化するわけではないので、コンテンツプロバイダとユーザとの間にコンテンツに対する認識に相違が生じる場合にはユーザにとって適切な属性定義のもとで推薦システムが運用されるとは限らない。そういった場合、コンテンツプロバイダが推薦したいコンテンツとユーザが視聴したいコンテンツに食い違いが生じてしまう。

一方、collaborative filtering においても、新たにコンテンツがシステムに追加されても、そのコンテンツがユーザに対して推薦されにくいという問題点がある。さらに類似ユーザがそのコンテンツを視聴しなければ、そのコンテンツがユーザの興味や関心に合致したものであったとしてもユーザに推薦されない場合がある。また特殊な興味や関心を持つユーザにとっては、類似ユーザの発見が困難であり、類似したユーザがいなかった場合にはコンテンツの推薦を行うことができず、推薦システムにおける個人化サービスを実現することができない。

近年、こういった問題を解決するために、content-based filtering と collaborative filtering を組み合わせたハイブリッドなアプローチを行う研究も行われている [5, 13]。本研究においても content-based filtering と collaborative filtering を組み合わせることにより情報推薦を行うが、これら既存の手法にカテゴリ定義の動的化を加えることにより、システムが推薦に利用するカテゴリ定義とユーザのコンテンツに対する認識との相違をなくす手法の提案を行う。また本手法を用いることにより実現できる動的なカテゴリ定義を導入した推薦システムを構築し、これまでの content-based filtering による推薦システムよりも効果的な個人化サービスの実現を目指す。

本稿では、2章で動的カテゴリを実現する手法の提案を行い、3章で提案手法を用いて実装した「VOD コンテンツ推薦システム」の説明を行う。4章では「VOD コンテンツ推薦システム」を用いて行った実証実験について説明し、システムと提案手法の評価を行い、5章で本研究をまとめる。

2 動的なカテゴリ定義の実現

content-based filtering や collaborative filtering などのフィルタリング手法を用いた多くの既存の推薦システムではコンテンツプロバイダによってコンテンツの属性情報やカテゴリ情報が静的に定義され、それらの情報に基づいて推薦コンテンツが決定される。しかしコンテンツプロバイダとユーザとの間にコンテンツに対する認識において相違が生じた場合、コンテンツプロバイダが推薦したいコンテンツとユーザが推薦してほしいコンテンツが一致しない場合がある。

例えば、対象ユーザが「野球」に関するコンテンツを必要としている場合を考える。既存の一般的な推薦システムではコンテンツプロバイダによって「野球」という属性情報を与えられたコンテンツから推薦するコンテンツリストを導出する。つまりコンテンツプロバイダによって「野球」に関するコンテンツであると判断されたコンテンツだけが推薦の対象となる。しかし、コンテンツに対する認識において対象ユーザとコンテンツプロバイダの間に相違が生じる場合は対

象ユーザが本当に必要とするコンテンツが提供されない場合が考えられる。つまり対象ユーザが視聴しても「野球」に関係すると認識しないコンテンツが推薦されたり、対象ユーザにとって「野球」に関係すると考えられるコンテンツが推薦されない場合が生じてしまう(図1)。

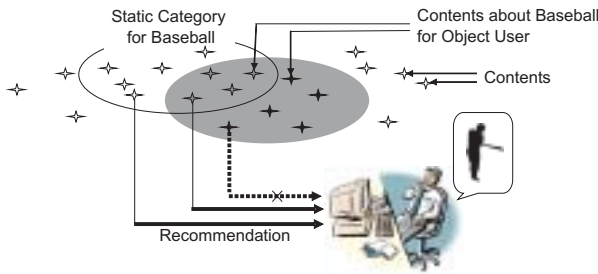


図 1: 静的カテゴリだけに依存した推薦

本研究では従来の推薦システムではコンテンツプロバイダによって静的に定義されていたコンテンツの属性情報としてのカテゴリ情報(以降本稿では静的カテゴリと呼ぶ)に注目し、そのカテゴリ情報をユーザ毎に、また動的に変化させる各コンテンツに新たなカテゴリ(以降本稿では動的カテゴリと呼ぶ)を割り当てる手法の提案を行う。本手法では、ユーザの興味や関心を表す属性情報や嗜好情報から類似ユーザ群としてのコミュニティを形成する。さらにコミュニティに属する各ユーザが過去にサーバに残したアクセスログより互いに何らかの関連を持つコンテンツから形成される類似コンテンツ集合の導出を行う。こうして導出される類似コンテンツ集合に属する各コンテンツに対し、共通のカテゴリを与え直すことにより、新たなコンテンツ分類を行う(図2)。

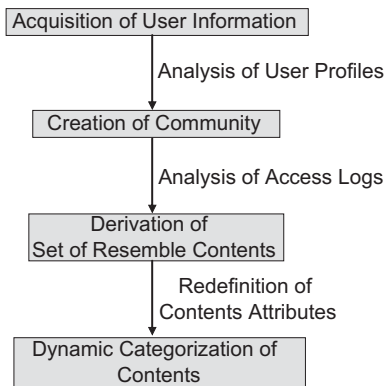


図 2: 動的なカテゴリ定義の流れ

類似コンテンツ集合はユーザ毎に導出することができ、また動的に変化するもので、本研究で提案する手法を用いることにより、コンテンツのカテゴリ定義の個人化及び動的化が実現できる。この動的に定義されるカテゴリを導入することで、静的に定義されるカテゴリだけによる content-based filtering や collaborative filtering を用いた従来の推薦システムよりも効果的なコンテンツ推薦システムを実現できる。

2.1 コミュニティの形成

本手法では互いに類似したユーザのグループとしてコミュニティという概念を導入する。コミュニティに属するユーザ同士のサイト上におけるアクションも類似すると考えることができ、類似ユーザの情報も用いた効果的な推薦プロセスを提供できる。これは従来の collaborative filtering でも用いられる考え方である。

システムはユーザの特徴である属性情報を入力し、ユーザプロフィールとして用いる。ユーザの特徴を入力するためには、ユーザがサイトを初めて利用した際に初期アンケートを実施してユーザの属性情報を得る方式とユーザが過去にそのサイトで行ったアクションのログ情報を用いてユーザの属性情報を導出する方式がある。

システムは初期アンケートやアクションログから得られる全ユーザのプロフィール情報を用いて、現在システムを利用しているユーザ u_{obj} (以降本稿では対象ユーザと表す。)とそれ以外の各ユーザ u_k との類似度 $R_{usr}(u_{obj}, u_k)$ を求める。ユーザ間の類似度は対象ユーザのユーザプロフィール $Pr(u_{obj})$ とその他のユーザのユーザプロフィール $Pr(u_k)$ を比較する関数 f_{res} を用いて計算される。

$$R_{usr}(u_{obj}, u_k) = f_{res}(Pr(u_{obj}), Pr(u_k))$$

こうして計算されるユーザ間の類似度により、対象ユーザを中心としたユーザのマッピングを行う。さらにこのユーザ間の類似度に対する閾値 K_{sh} を定め、閾値以上の類似度を持つユーザを対象ユーザの類似ユーザとして扱う。対象ユーザと属性の類似した類似ユーザのグループをコミュニティ $M(u_{obj})$ として定義する(図3)。

$$M(u_{obj}) = \{u_k | R_{usr}(u_{obj}, u_k) > K_{sh}, k = 1, \dots, obj - 1, obj + 1, \dots, n\}$$

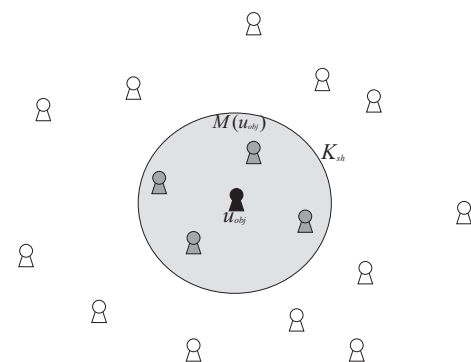


図 3: 対象ユーザとの類似度によるユーザマッピング

コミュニティに属する各ユーザは対象ユーザと興味や関心においてなんらかの関連性があると考えられ、コミュニティに属するユーザに関する情報は対象ユーザに個人化サービスを提供する際に用いられる情報として有効である。

2.2 類似コンテンツ集合の導出

対象ユーザのコミュニティを定義した後、互いに何らかの関連性を持つコンテンツの集合である類似コンテンツ集合を求める。

類似コンテンツ集合を導出するために対象ユーザのコミュニティに属する各ユーザに関する情報を用いる。ユーザに関する情報としては、ユーザが過去にサーバに残したアクセスログを用いる。本研究ではアクセスログにセッションという概念を導入する [14, 15]。ユーザは情報入手を行うためにサービスを行うサイトにログインし、複数のコンテンツ (c_l, c_m, c_n, c_o) にアクセスし、ログアウトすると考えられ、その1回の操作をセッション s_j として扱う。つまり各セッションはアクセス中に視聴したコンテンツ集合であり、視聴した順序を保持したものとして扱う。

$$s_j = \{c_l, c_m, c_n, c_o\}$$

さらに対象ユーザのコミュニティに属するユーザが過去に残したセッションを集約しセッション集合 V を導出する。このセッション集合 V を解析の対象とする。これはコミュニティに属するユーザの情報は対象ユーザに対して個人化サービスを提供する際に有効な情報として機能すると考えられるからである。このセッション集合 V を解析することにより任意の2コンテンツ c_l, c_m 間の類似度を計算する。コンテンツ間の類似度 $R_{cnt}(c_l, c_m)$ は、

- 同一セッション内において2つのコンテンツが共起する確率
- 共起した場合における2つのコンテンツが連続してアクセスされる確率

に注目した式で計算される (図4)。これは互いに類似したコンテンツはユーザがシステムを利用する際に同一セッション内で視聴される傾向があると考えられるからである。さらに、そのセッション内で連続して視聴される2つのコンテンツ間にはさらに有効な関連性があると考えられるからである。

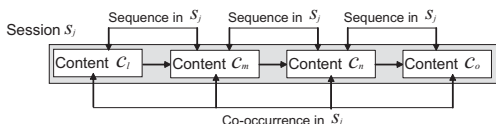


図4: コンテンツの共起性と連続性

以上のことを考慮することにより、セッション集合 V における2つのコンテンツ c_l, c_m 間の類似度 $R_{cnt}(V, c_l, c_m)$ に対する計算式に対して以下の式を定義することができる。

$$R_{cnt}(V, c_l, c_m) = \alpha \frac{N_{co_app}(V, c_l, c_m)}{N_{app}(V, c_l)} + \beta \frac{N_{co_app}(V, c_l, c_m)}{N_{app}(V, c_m)}$$

$$+ \gamma \frac{N_{sqc}(V, c_l, c_m)}{N_{co_app}(V, c_l, c_m)}$$

$$\alpha + \beta + \gamma = 1$$

計算式ではセッション集合 V においてコンテンツ c_l が出現したセッション数を表す $N_{app}(V, c_l)$ 、セッション集合 V において2つのコンテンツ c_l, c_m が共に出現したセッション数を表す $N_{co_app}(V, c_l, c_m)$ 及び2つのコンテンツが連続してアクセスされたセッション数を表す $N_{sqc}(V, c_l, c_m)$ が用いられる。

実際にシステムに適用する際には α, β, γ に対して適切な定数を決定する。この計算式を用いて任意の2コンテンツ間の類似度を計算し、その類似度を基にコンテンツのマッピングを行う。さらに閾値 C_{sh} によるクラスタリングを行い、類似コンテンツ集合 L_i を導出する (図5)。

$$R_{cnt}(c_l, c_m) > C_{sh}$$

$$c_l, c_m \in L_i$$

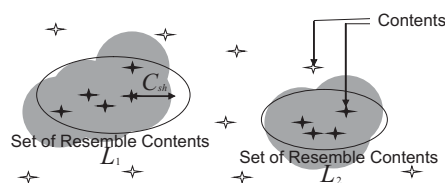


図5: コンテンツ間の類似度によるコンテンツマッピング

提供される全コンテンツの中で、最大の類似度を持つ2コンテンツを選択し、それらを含む類似コンテンツ集合 L_i を形成する。さらに類似コンテンツ集合に属するコンテンツ c_l とその他のコンテンツ c_m との類似度を調べ、閾値 C_{sh} 以上の値を持つものがあれば、コンテンツ c_m を類似コンテンツ集合 L_i に含める。この操作を新たに類似コンテンツ集合に加えるコンテンツが存在しなくなるまで継続する。

本手法では以上のように類似コンテンツ集合を導出する。類似コンテンツ集合はコンテンツ間の類似度を基に導出されるので、同一の類似コンテンツ集合に属する2コンテンツ間にはなんらかの関連性があると考えられる。

2.3 カテゴリ導出

本節では2.2節において説明した類似コンテンツ集合を用いてカテゴリ定義の動的化および個人化を行う手法について説明する。静的カテゴリは、コンテンツプロバイダにとって互いに類似した複数のコンテンツから構成されるように定義される。一方、類似コンテンツ集合は対象ユーザにとって互いに何らかの関連性があると考えられる複数のコンテンツから構成される。

本手法では静的カテゴリを類似コンテンツ集合を用いて動的に変化させ、コンテンツプロバイダだけによって静的に定義された静的カテゴリからユーザの興

味や関心を反映した動的カテゴリを導出することができる。

類似コンテンツ集合は様々な静的カテゴリを持ったコンテンツから構成される。この類似コンテンツ集合のコンテンツが持つ静的カテゴリの中で最も多数を占めるカテゴリを類似コンテンツ集合を構成する全てのコンテンツのカテゴリとして与え直す (図 6)。

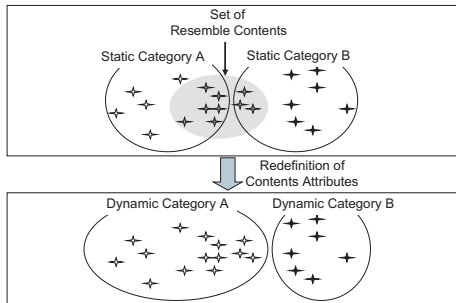


図 6: 類似コンテンツ集合に対するカテゴリの再定義

例えばコンテンツ c_1, c_2, c_3, c_4 から構成される類似コンテンツ集合 L_g を考える。さらにコンテンツ c_1, c_2, c_5 は静的カテゴリ N_a を持ち、コンテンツ c_3, c_6, c_7, c_8 は静的カテゴリ N_b を持ち、コンテンツ c_4, c_9, c_{10}, c_{11} は静的カテゴリ N_c を持つとする。この場合、類似コンテンツ集合 L_g に属するコンテンツで静的カテゴリ N_a を持つものが最大多数を占めることになる。そこで、類似コンテンツ集合 L_{g1} に属するすべてのコンテンツに対して新たにカテゴリ N_a を与え直す。これにより、コンテンツ c_1, c_2, c_3, c_4, c_5 から構成されるカテゴリ N'_a 、コンテンツ c_6, c_7, c_8 から構成されるカテゴリ N'_b 、コンテンツ c_9, c_{10}, c_{11} から構成されるカテゴリ N'_c が新たに定義されることになる。このように定義されるカテゴリを動的カテゴリとして扱う。

$$\begin{aligned}
 L_g &= \{c_1, c_2, c_3, c_4\} \\
 N_a &= \{c_1, c_2, c_5\} \\
 N_b &= \{c_3, c_6, c_7, c_8\} \\
 N_c &= \{c_4, c_9, c_{10}, c_{11}\} \\
 &\downarrow \\
 N'_a &= \{c_1, c_2, c_3, c_4, c_5\} \\
 N'_b &= \{c_6, c_7, c_8\} \\
 N'_c &= \{c_9, c_{10}, c_{11}\}
 \end{aligned}$$

このように静的カテゴリを類似コンテンツ集合を用いて変化させ、動的カテゴリを導出することができる。本章において説明した手法を用いて導出した動的カテゴリはコンテンツプロバイダおよび対象ユーザの両方のコンテンツに対する認識を反映したものであるといえ、この動的カテゴリを用いることで、より効果的な個人化サービスを推薦システムにおいて実現することができる。

3 VOD コンテンツ推薦システムへの適用

本章では 2 章において説明した手法による動的カテゴリを導入した「VOD コンテンツ推薦システム」(図 7) の説明を行う。この「VOD コンテンツ推薦システム」では介護やエステに関する映像コンテンツ約 60 個を扱っている。また約 60 名のユーザが実証実験に参加することを想定しシステムを実装しており、類似度に対する閾値、コミュニティに属するユーザの最大値や計算式における重みなどはそれらを考慮して設定した。



図 7: VOD コンテンツ推薦システム

3.1 システムの概要

「VOD コンテンツ推薦システム」では各ユーザのユーザプロフィールや過去のアクションログを用いて推薦リストの導出を行う。本節では「VOD コンテンツ推薦システム」のシステム全体の流れを説明する。システム全体の流れを図 8 に示す。

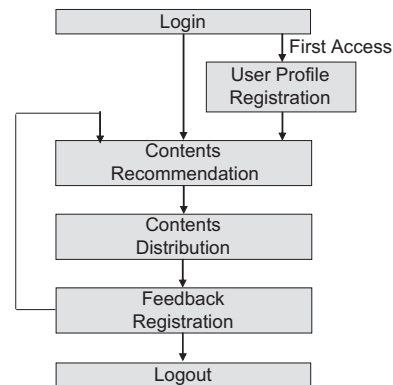


図 8: システム全体の流れ

ユーザプロフィールの情報取得は Web ページ上におけるユーザに対するアンケートによって行う。対象ユーザが「VOD コンテンツ推薦システム」を初めて利用するためにサイトにアクセスするとシステムは

対象ユーザがシステムをこれまでに利用した回数を計算する．対象ユーザがこれまでに利用したことがないということがわかると初期アンケート画面の提示が行われ、対象ユーザは初期アンケートの画面にしたがって、ユーザプロフィールの情報を登録することができる．初期アンケート画面によってユーザプロフィールを登録した後、対象ユーザに対して推薦するコンテンツリストが提示される．

一方、対象ユーザのシステムを利用した回数が2回以上であった場合には初期アンケートを行わず、推薦するコンテンツリストを提示する．

対象ユーザはその推薦されたコンテンツリストの中から視聴したいコンテンツを選択することができる．コンテンツを選択すると映像コンテンツを視聴することができ、視聴した後、視聴したコンテンツに対するアンケートが実施される．このアンケート結果は、「VOD コンテンツ推薦システム」に対するフィードバックとして、今回動的カテゴリを導入して実装したシステムに対する評価を行うために用いる．以降、推薦するコンテンツリストの提示、コンテンツの視聴、視聴したコンテンツに対するアンケートの実施が繰り返されることになる．

コンテンツの視聴が終わるとユーザはログアウトすることができ、視聴した順番を保持したコンテンツ集合からなるセッションをシステムはユーザ毎に用意されるログファイルに記録する．

3.2 コンテンツのカテゴリ定義

本節では2章において説明した動的カテゴリを「VOD コンテンツ推薦システム」に対して導入する方法について説明する．

本システムで扱う各コンテンツはそれぞれの映像や内容に対して特徴を持っている．今回実装した「VOD コンテンツ推薦システム」では各コンテンツの特徴を表す属性情報としてコンテンツの種類を表すキーワードを用いており、この情報をカテゴリとして扱う．各コンテンツに対して定義される静的カテゴリはコンテンツプロバイダが予め用意したカテゴリの候補からコンテンツプロバイダが最も適切であると考えたカテゴリを選択して用いる．本システムではコンテンツプロバイダは静的カテゴリとして各コンテンツに対して3つのカテゴリ Category_1, Category_2, Category_3 をそれぞれの候補から決定する．さらにシステムはユーザやコミュニティのプロファイルによって類似コンテンツ集合を導出する．その類似コンテンツ集合に属するコンテンツに対する新たなカテゴリ定義を行い、コンテンツプロバイダの静的なカテゴリ定義に対して変化を加え、動的カテゴリを決定することができる(図9)．

3.2.1 コンテンツの静的カテゴリ

「VOD コンテンツ推薦システム」では各コンテンツはコンテンツを表現する情報として表1に示すようなエントリ情報を持つように定められる．これらのエントリ情報はコンテンツプロバイダが定める予め定義するものである．

表1に示した Category_1, Category_2, Category_3

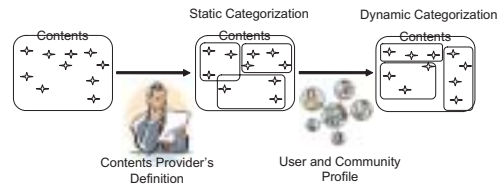


図9: 提案手法におけるコンテンツのカテゴリ定義

は静的カテゴリであり、それぞれに対して表2に示すような9個ずつの候補を予め定めた．この静的カテゴリの候補は、Category_1にはコンテンツを料理の栄養分や効果および番組生成におけるキーワードに基づいて分類することによって決定した．Category_2に関しては料理の調理法やスキンケアに対するアプローチおよび番組の連続性における位置づけに基づいた分類により決定し、Category_3は料理の国籍やエステや介護の番組における目的におけるキーワードに基づいた分類により決定した．

コンテンツプロバイダは各コンテンツの Category_1, Category_2, Category_3 に対して候補の中からコンテンツプロバイダにとって最も適切なものを選択し、静的カテゴリとして定義する．

3.2.2 ユーザ属性とコミュニティ

本システムはユーザは自分の特徴を表すユーザプロフィールをシステムにWebページ上でのアンケートを用いて登録する必要があり、そのエントリ情報を表3に示す．

アンケートにおいて登録される情報は性別、年齢の他に介護や健康に対する関心度などがある．これは今回扱うコンテンツが介護やエステに関するコンテンツが中心となっているためであり、扱うコンテンツに関連するユーザ属性を設定することで、より効果的なコミュニティの形成を行うことができる．

このように入手したユーザプロフィールとしてユーザのモデリングに用いる．さらに、対象ユーザとその他のユーザとのユーザプロフィールのマッチングを行うことにより、対象ユーザとその他のユーザ間の類似度を計算する．ユーザプロフィールは各ユーザの年齢や性別、介護に対する意識の3段階の評価などから構成され、それらの値の差を基にユーザ間の類似度が計算される．ユーザ間の類似度は0から1までの値がユーザプロフィールのマッチングによる計算によって与えられ、ユーザプロフィール間に共通点が多く見られれば高い類似度が与えられる．

さらにユーザ間の類似度に対して閾値を定める．本システムではユーザ間の類似度に対する閾値を0.5に定めた．閾値0.5を越える対象ユーザとの類似度を持つユーザは対象ユーザのコミュニティに属することになる．しかし閾値0.5以上の対象ユーザとの類似度を持つユーザが数多く存在する場合、コミュニティに属するユーザの数とユーザ全体の数が不適切な割合になってしまうことが考えられる．そこで、本システムでは実証実験の参加人数を考慮してコミュニティに属するユーザ数の最大値を5と設定して実装した．対象ユーザとの類似度として閾値0.5以上の値を持つユー

エントリー名	エントリーの説明	エントリー情報の例
Content_Id	コンテンツID	27
Title	コンテンツタイトル	テーマ別健康料理集ビタミン A を多くとる料理「えびと野菜の中華風揚げ漬け」
Local_Path	映像アーカイブの URL	http://www.civa.jp/ipv6/mvx/RM027.mvx
Category_1	静的カテゴリ 1	栄養を考慮した料理
Category_2	静的カテゴリ 2	スキンケアを考慮した料理
Category_3	静的カテゴリ 3	中華料理

表 1: コンテンツのエントリー情報

Category_1	Category_2	Category_3
栄養を考慮した料理	スキンケアを考慮した料理	西洋料理
療養に適した料理	なま物	中華料理
ダイエットに適した料理	ごはん物	日本料理 (メイン)
カロリーの高い料理	煮物	日本料理 (サイド)
社会生活にまつわる話	焼き物	エスニック料理
ホームヘルプに関する話	汁物	その他の料理
教養にまつわる話	導入説明	コミュニケーションの方法
美容にまつわる話	詳細説明	リラックスする運動
ダイエットを目的とした運動	スキンケアの方法	マッサージの方法

表 2: カテゴリ属性の候補

ザの数が 5 を越える場合には高い類似度を持つユーザから順番にコミュニティに属することになる。

3.2.3 類似コンテンツ集合

対象ユーザのコミュニティを導出した後、対象ユーザのコミュニティに属するユーザが過去に残したセッションの集約を行う。この集約されたセッションを 2.2 節で説明した手法を用いて解析し、任意の 2 コンテンツ間の類似度を計算する。コンテンツ間の類似度の計算式のモデルを 2.2 節において示したが、本システムでは α, β に 0.4, γ に 0.2 を適用することにより以下の計算式を定義し、セッション集合 V におけるコンテンツ間 c_l, c_m の類似度計算を行う。

$$\begin{aligned}
 R_{cnt}(V, c_l, c_m) &= 0.4 \frac{N_{co_app}(V, c_l, c_m)}{N_{app}(V, c_l)} \\
 &+ 0.4 \frac{N_{co_app}(V, c_l, c_m)}{N_{app}(V, c_m)} \\
 &+ 0.2 \frac{N_{sqc}(V, c_l, c_m)}{N_{co_app}(V, c_l, c_m)}
 \end{aligned}$$

こうして求まるコンテンツ間の類似度に基づいてコンテンツのマッピングを行い、さらに 2.2 節において説明した手法によるクラスタリングを行う。この場合コンテンツ間の類似の閾値として 0.7 を用いて実装している。このクラスタリングによって導出されるコンテンツのクラスタを類似コンテンツ集合として扱う。しかし類似コンテンツ集合に関してもコンテンツに全体に対する類似コンテンツ集合の規模が大きすぎると類似コンテンツ集合が本システムにおいて適切に機能しないと考えられる。そこで類似コンテンツ集合の規模が大きくなりすぎないように類似コンテンツ集合を構成するコンテンツ数の最大値を 8 に設定した。これ

により、1 つの動的カテゴリを持つコンテンツ数を適切な数におさえることができる。

3.2.4 コンテンツの動的カテゴリ

本システムでは 3.2.1 節で説明したように、各コンテンツに対して 3 つの静的カテゴリ $Category_1, Category_2, Category_3$ が定められている。

そこで類似コンテンツ集合に属するコンテンツを持つ $Category_1$ のうちで最も多数を占めるもの j_a とすると、類似コンテンツ集合に属する全てのコンテンツの動的カテゴリ $Category_1'$ に j_a を割り当てる。類似コンテンツ集合に属さないコンテンツに対しては $Category_1$ で用いられていたカテゴリを $Category_1'$ に割り当てる。

この操作を動的カテゴリ $Category_2', Category_3'$ に対しても行い、コンテンツに与えるカテゴリの動的化を実現する。なお、最も多くを占める静的カテゴリが複数に及ぶ場合は、そのような静的カテゴリの候補からランダムに選択し、定義する動的カテゴリを 1 つに定める。

3.3 推薦リストの導出

本システムではユーザに対して推薦する映像コンテンツのリストを提示するが、その映像コンテンツのリストを導出する過程において個人化が実現されている。本節では個人化された推薦する映像コンテンツリストの導出手法について説明する。

ログインした直後の推薦リストは対象ユーザのコミュニティに属するユーザが過去に多く視聴したコンテンツから構成されている。対象ユーザのコミュニティは対象ユーザと興味や関心が類似するようなユーザから構成されるように設計されているので、対象

エン트리名	エントリの説明	エン트리情報の例
userId	ユーザ ID	00001
Nick_Name	ニックネーム	m-okamot
Sex	性別	male
Birth_Day	生年月日	19770803
Diet	ダイエットした経験の有無 (3 段階)	0
Cook	料理をする頻度 (4 段階)	2
Health	健康に対する意識 (3 段階)	0
Care	介護に対する意識 (3 段階)	1

表 3: ユーザプロフィールのエン트리情報

ユーザがログインした直後の推薦リストは対象ユーザと興味や関心が類似したユーザが多く見たコンテンツが含まれることになる。各ユーザのコミュニティは互いに異なるので、導出される映像コンテンツリストもユーザ毎に異なったものが提示される。

一方、その後の推薦する映像コンテンツリストは対象ユーザが直前に選択したコンテンツと同じカテゴリを持つものから選択して構成する。その際に用いるコンテンツのカテゴリは動的カテゴリである。つまり対象ユーザが直前に視聴したコンテンツ c_1 の動的カテゴリ $Category_1'$, $Category_2'$, $Category_3'$ がそれぞれ j_a , j_d , j_f である場合、 $Category_1'$ として j_a が定義されているコンテンツや $Category_2'$ として j_d が定義されているコンテンツや $Category_3'$ として j_f が定義されているコンテンツから推薦する映像コンテンツリストを導出する。

本稿でここまで述べてきたように、動的カテゴリの定義はユーザ毎に異なる。つまりユーザ毎に異なる個人化されたカテゴリ定義の環境下において推薦する映像コンテンツリストが導出されることになり、このように導出される映像コンテンツリストにおいても個人化サービスが実現されているといえる。

3.4 映像配信とフィードバック

個人化された推薦する映像コンテンツのリストを導出して対象ユーザに提示した後、対象ユーザは推薦されたコンテンツリストの中から自分の視聴したい映像コンテンツを選択する。対象ユーザが視聴したいコンテンツを選択すると、システムは映像コンテンツをネットワークを通じてストリーミング配信する (図 10)。

システムは映像コンテンツのストリーミング配信を終えた後、対象ユーザに対して視聴したコンテンツに対するアンケートを実施し、対象ユーザは視聴したコンテンツに対するフィードバックを入力することになる。さらに本システムではシステムが提示したコンテンツリストおよび対象ユーザが視聴したコンテンツの情報を対象ユーザがサイト内で行ったアクションとして記録する。

本研究ではこれらの情報を用いることにより、「VOD コンテンツ推薦システム」や提案手法の評価を行った。



図 10: 「VOD コンテンツ推薦システム」における映像コンテンツ配信 (映像提供: NTT 西日本)

4 実証実験と評価

本章では実装した「VOD コンテンツ推薦システム」を用いて行った実証実験について説明し、さらに実証実験において得られた結果を用いることにより、システムおよび提案手法の評価を行う。

4.1 実証実験

実証実験では情報家電インターネット技術研究開発事業「情報家電 IPv6 映像配信実験」において本研究で開発した「VOD コンテンツ推薦システム」を NTT 西日本との共同実験として 43 日間にわたって運用した。「情報家電 IPv6 映像配信実験」では学外のユーザ約 50 名と研究室内のユーザ 14 名を対象とした映像をストリーミング配信するサービスを IPv6 のネットワーク上でっており、アプリケーションサーバ、データベースサーバ及びストリーミングサーバは NTT 西日本に置かれている (図 11)。本システムではエステや介護に関する 60 個の VOD コンテンツをモニタに対して推薦し、モニタは VOD コンテンツを視聴した後にその番組に対するフィードバックを入力することができる。

4.2 評価

本研究ではシステムを利用したユーザから視聴したコンテンツに対するフィードバック結果を用いて動的カテゴリを導入して実装したシステムの評価を行う。また、ユーザがシステムを利用する際のサイト上で

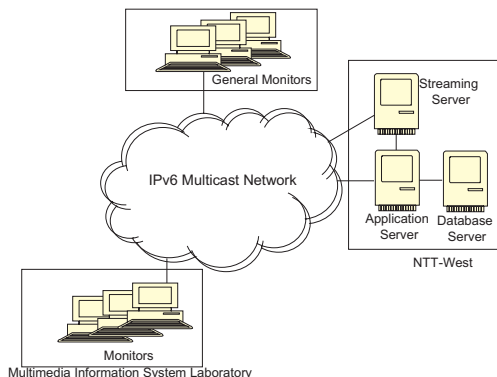


図 11: 情報家電 IPv6 映像配信実験ネットワーク構成

行ったアクションのログを得ることができ、その結果を用いて提案手法の有効性を評価する。

4.2.1 システムの評価

この 43 日間にわたる実証実験を行った結果、モニタ約 60 名のうち 27 名が本システムを利用した。この 27 名のユーザが本システムに平均 2.93 回のアクセスを行い、平均 8.5 個のコンテンツを視聴した。このような実験環境において、ユーザに対して、視聴したコンテンツに対する 3 つのアンケートを行った。

まず、提供されたコンテンツがユーザにとって興味深いものであったかどうかを示す「あなたはこの番組を面白いと思いませんか?」という質問を行った結果、39%のユーザが提供されたコンテンツに対して「面白かった」と回答した。一方、「普通だった」と回答したユーザが最大の 46%を占め、残りの 15%のユーザは「つまらなかった」と回答した。

次に「この番組はあなたに役立ちましたか?」というアンケートを実施した結果、40%のユーザが提供されたコンテンツに対して「役だった」と回答した。一方、42%のユーザが「普通だった」と回答し、残りの 18%のユーザは「無意味だった」と回答した。

最後に「この番組はあなたの求めている情報と関連するものでしたか?」というアンケートを行った結果、対象ユーザからの回答の 46% が「とても関連があった」であった。つまり、本システムによって対象ユーザに提示されたコンテンツのおよそ半数がユーザが求めている情報と関連するものであった。また「少し関連があった」という回答も総回答数の 39%を占めており、全体の 85%が対象ユーザの求めている情報と関連するものであったことがわかる。このことから、本研究で実装した「VOD コンテンツ推薦システム」がユーザにとって必要な情報を提示するという点において、機能していることがわかる。

ここまでユーザが視聴したコンテンツがユーザに対して与えた印象に関するフィードバック結果を示してきたが、コンテンツのユーザにとっての面白さ、有用性、関連性に関するいづれの結果においてシステムへの肯定的な評価が否定的な評価を上まわっていることが分かった。これらの結果から本システムがユーザにとって有効なシステムであったということがいえる。

4.2.2 提案手法の評価

次に、本研究で提案する動的カテゴリ導出手法が、「VOD コンテンツ推薦システム」において機能したかどうかを考察し、提案手法の評価を行う。対象ユーザが視聴したコンテンツには動的カテゴリを導入することによって推薦コンテンツリストに加えられたコンテンツと動的カテゴリを導入しなくても推薦されるコンテンツの 2 種類が存在する。そこで動的カテゴリを導入することによって推薦されたコンテンツが対象ユーザに視聴される確率と動的カテゴリを導入しなくても推薦されるコンテンツが対象ユーザに視聴される確率との比較を行うことにより、本研究で提案する動的カテゴリが「VOD コンテンツ推薦システム」において機能したかを示す。

本実証実験では 27 名のユーザが「VOD コンテンツ推薦システム」を利用し、のべ 229 個のエッセイや介護のコンテンツを視聴した。その過程においてシステムはのべ 1618 個のコンテンツをユーザに提示した。またユーザが視聴したコンテンツのべ 229 個のうち 79 個のコンテンツは対象ユーザがログインした直後に視聴したコンテンツであり、これらのコンテンツはカテゴリの概念によって推薦されるものではなく、対象ユーザのコミュニティに属するユーザが過去に多く視聴したコンテンツから推薦されるものである。よって考察の対象をユーザが視聴したのべ 229 個のコンテンツのうちログインした直後に視聴した 79 個のコンテンツを除いたのべ 150 個のコンテンツとする。またこのべ 150 個のコンテンツが視聴された際に、対象ユーザに推薦されたコンテンツの数はのべ 1182 個であった。

これらの動的カテゴリによるコンテンツと動的カテゴリによらないコンテンツの割合を図 12 に示す。

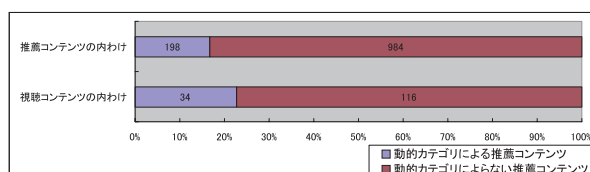


図 12: 「VOD コンテンツ推薦システム」における動的カテゴリの機能

図 12 からわかるようにユーザに視聴されたのべ 150 個のコンテンツが推薦された際に導出された推薦コンテンツリストを構成するコンテンツの総数はのべ 1182 個であり、このうちののべ 198 個が動的カテゴリを導入することによって推薦コンテンツリストに加えられたものであり、残りののべ 984 個は動的カテゴリを導入しなくても推薦コンテンツリストに加えられたものであった。一方、実際に対象ユーザに視聴されたのべ 150 個のコンテンツのうち、動的カテゴリを導入することによって推薦コンテンツリストに加えられたコンテンツがのべ 34 個であり、これは 23%を占めることになる。一方、動的カテゴリを導入しない場合においても推薦コンテンツリストに加えられるコンテンツがのべ 116 個存在し、これは残りの 77%を占めることになる。

図 12 に示した結果における動的カテゴリによる推薦コンテンツがユーザに選択され、視聴される確率が 17.2% であることがわかる。一方、動的カテゴリによらない推薦コンテンツがユーザに選択され、視聴される確率は 11.8% である。つまり動的カテゴリによって推薦されるコンテンツの方が動的カテゴリによらずに推薦されるコンテンツよりもユーザにとって有効なコンテンツである確率が高いことがわかる。

このことは本研究において提案する動的カテゴリを導入して推薦システムを構築するほうが、コンテンツプロバイダが定義する静的カテゴリだけを用いて推薦コンテンツを決定する従来の推薦手法よりもユーザにとって有効なコンテンツを推薦できることを示している。よって本稿において説明した動的カテゴリの導出手法を用いて動的カテゴリを導入した推薦システムは従来の手法を用いて構築する推薦システムよりもユーザの要求に応じた推薦コンテンツリストを導出できるといえる。

5 終わりに

現在、運用されている推薦システムにおけるフィルタリング手法として既存の content-based filtering や collaborative filtering が一般的に用いられている。しかし、それらのシステムではコンテンツプロバイダが各コンテンツに対して定義する属性情報を用いることが多い。このコンテンツプロバイダが定義する属性情報はユーザによって変化するわけではなく、コンテンツプロバイダのコンテンツに対する認識に従った共通のものが全ユーザに対して用いられることになる。そのため、コンテンツプロバイダが推薦したいコンテンツとユーザが推薦してほしいコンテンツに食い違いが生じてしまう。

そこで本研究では、コンテンツプロバイダが定義する静的カテゴリの他に動的カテゴリという概念を導入する手法を提案した。この動的カテゴリを導入することにより、ユーザの求める個人化されたカテゴリ定義の環境下での推薦を実現することができる。その結果、これまでに実現されていた以上に効果的な個人化サービスが可能となる。

また本研究では、動的カテゴリを導入した推薦システムとして「VOD コンテンツ推薦システム」を実装し、情報家電 IPv6 映像配信実証実験において約 60 名のユーザを対象に「VOD コンテンツ推薦システム」を用いたコンテンツ配信サービスを行った。その実証実験の結果、本システムと提案手法の有効性を示すことができた。

今後の課題としてはさらに提案手法の有効性についての考察を深めるためにシステムが導出した動的なカテゴリ定義と各ユーザのコンテンツに対する認識に基づいたカテゴリ定義とを比較することが必要である。また「VOD コンテンツ推薦システム」では直前に視聴されたコンテンツと共通の動的カテゴリを持つコンテンツの中からランダムにコンテンツを選択して推薦コンテンツリストを導出したが、これらのコンテンツに対してユーザプロファイルを用いた優先度計算の概念を導入することでさらに精度の高い推薦システムが実現できると考えられる。今後、これらの課題に対し

でも取り組んでいくことにより、Yahoo! などのような大規模なポータルサイトに対しても実用できると考える。

謝辞

本研究の一部は、日本学術振興会未来開拓学術研究推進事業における研究プロジェクト「マルチメディア・コンテンツの高度処理の研究」(Project No. JSPS-RFTF97P00501)、及び、通信・放送機構 情報家電インターネット技術研究開発事業「情報家電 IPv6 映像配信実験」によるものである。

参考文献

- [1] N. J. Belkin, "Helping People Find What They don't Know," *Communication of the ACM*, vol. 43, pp. 58-61, March 2000.
- [2] U. Manber, A. Patel, and J. Robison, "Experience with Personalization on YAHOO!," *Communication of the ACM*, vol. 43, pp. 35-40, August 2000.
- [3] B. Smyth and P. Cotter, "A Personalized Television Listings Service," *Communication of the ACM*, vol. 43, pp. 107-111, August 2000.
- [4] H. Hirsh, C. Bash, and B. D. Davison, "Learning to Personalize," *Communication of the ACM*, vol. 43, pp. 102-106, August 2000.
- [5] M. Balabanovic and Y. Shoham, "Content-based, Collaborative Recommendation," *Communication of the ACM*, vol. 40, pp. 66-72, March 1997.
- [6] F. Pachet, P. Roy, and D. Cazaly, "A Combinatorial Approach to Content-based Music Selection," *IEEE Multimedia*, vol. 7, pp. 44-51, January-March 2000.
- [7] M. Perkowitz and O. Etzioni, "Adaptive Web Sites : Conceptual Cluster Mining," in *Proceedings of the 16th International Joint Conference on Artificial Intelligence*, pp. 264-269, July 1999.
- [8] J. A. Konstan, B. N. Miller, D. Maltz, J. L. Herlocker, L. R. Gordon, and J. Riedl, "Applying Collaborative Filtering to Usenet News," *Communication of the ACM*, vol. 40, pp. 77-87, March 1997.
- [9] 岡本 道也, 植田 和憲, 鎌原 淳三, 下條 真司, 宮原 秀夫, "教師付学習を利用したストーリー性を持つシナリオテンプレートによるスポーツダイジェスト自動生成機構," 第 11 回データ工学ワークショップ (DEWS2000) 論文集, March 2000.
- [10] M. Okamoto, K. Ueda, J. Kamahara, S. Shimojo, and H. Miyahara, "An Architecture of Personalized Sports Digest System with Scenario Templates," in *Proceedings of 7th International Conference on Database Systems for Advanced Applications (DASFAA2001)*, pp. 170-171, April 2001.
- [11] 岡本 道也, 鎌原 淳三, 下條 真司, 宮原 秀夫, "動的生成されたシナリオを用いた個人化コンテンツの自動生成," 第 12 回データ工学ワークショップ (DEWS2001) 論文集, March 2001.
- [12] M. Okamoto, J. Kamahara, S. Shimojo, and H. Miyahara, "Automatic Production of Personalized Contents with Dynamic Scenario," in *Proceedings of 2001 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM'01)*, vol. 1, pp. 91-94, August 2001.
- [13] G. Polcicova, R. Slovak, and P. Navrat, "Combining Content-Based and Collaborative Filtering," in *Proceedings of ADBIS-DASFAA Symposium 2000*, pp. 118-127, September 2000.
- [14] E. de Vos and E. J. H. Kerckhoffs, "The Web Tracking System IMU," in *Proceedings of 6th Euromedia Conference 2001*, pp. 54-63, April 2001.
- [15] B. Mobasher, R. Cooley, and J. Srivastava, "Automatic Personalization Based on Web Usage Mining," *Communication of the ACM*, vol. 43, pp. 142-151, March 2000.