

# 8U-06 履歴データベースを用いたTV番組選択支援エージェントの改良について

磯貝 邦昭\* 中島 太郎\* 渡辺 尚\* 樽口 秀昭\*\*

\* 静岡大学 情報学部 \*\* ヤマハ株式会社

## 1 はじめに

昨年末にBSデジタル放送が開始され、今後は地上波のデジタル化も行われる予定である。放送電波のデジタル化により、デジタル圧縮による多チャンネル化が生じる[1]。多チャンネル化によってユーザは多くの番組を提供され、その選択肢が広がるが、その一方で、膨大な情報の中から自分の見たい番組を選択するための労力も増大する。そのためにユーザは各ジャンルに点数をつけてエージェントに好みを教えたり[2]、あるいはユーザの手を煩わせず、視聴時間によって番組やその出演者などの評価を決めたり[3]する方法が検討されている。

TV番組選択支援エージェント(Testa)[4]は、ユーザがとった行動を履歴として記録し、それをもとにユーザが番組検索のために入力したキーワードの意味を推測して、それを置換や補完して番組表の検索を行い、ユーザが見たい番組を探すのを助ける。本稿では、このシステムの問題点とその改良について述べる。

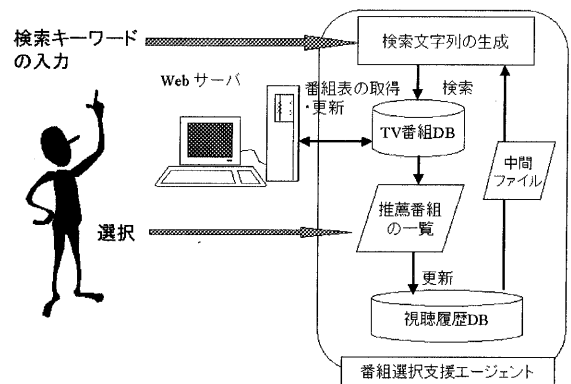


図1: TV番組選択支援エージェント(Testa)の概観

## 2 TV番組選択支援エージェント(Testa)

### 2.1 Testaの概要

図1にこのシステムの概観図を示す。ユーザは見たい番組をキーワードを入力して探す。Testaは視聴履歴データベースに、

- ユーザが番組検索用に入力したキーワード
- ユーザが選択した番組の情報とその推薦順位、その番組評価値
- ユーザが選択しなかった番組のうち、選択した番組より上位で推薦した番組の情報とその推薦順位、番組評価値

を記録している。ここで、番組の情報は、放送曜日、放送開始時間、放送局、ジャンル名、番組タイトル、番組内容(出演者名等)といった、Webサーバから得られる情報である。

ユーザの入力キーワードを視聴履歴データベースの情報により置換、補完して検索文字列を生成し、それを用いて番組表を検索する。そして、番組評価値の高いものから順位付けしてその番組の情報をユーザに提示する。ユーザはTestaの提示した番組の中から目的の番組を選択すると、その情報は視聴履歴データベースに反映され、更新される。

### 2.2 Testaの問題点

Testaは生成した検索文字列により番組表を検索する際、文字列の一致度を調べている。そのため、

An Improvement of TV Programs Selection Support Agent with History Database

\*Kuniaki Isogai, Taro Nakajima, Takashi Watanabe and

\*\*Hideaki Taruguchi

\*Faculty of Information Shizuoka University, \*\*YAMAHA Corporation

- 文字列中の助詞等の、番組内容に関して意味を持たない文字とも一致が起るので、検索結果に悪影響を与える
- ひらがなとカタカナ、漢字とその読みがなとの間の一致がとれない

という問題が生じる。これらの問題には、3.1で述べるように、形態素解析と読みがなを用いて対処する。

もう一つの問題は、ある番組の持つ情報について、番組タイトルのようにほとんど変更されないものと、番組の放送内容のように頻繁に変更されるものを特に区別せずに扱っていることである。なぜなら、履歴を用いて検索をする場合、変更の少ない情報の方が変更の多い情報よりも重視されるため、放送内容が推薦結果に反映されにくくなってしまふからだ。この問題には、3.2で述べるように、番組情報を分割することで対処する。

## 3 Testaの改良

### 3.1 検索方法の改良

番組表の番組内容の中から助詞等を除去するために、番組内容の文字列に対して形態素解析を行なった。形態素解析には茶筌[5]を使用した。そして、文字列の中から名詞のみを取り出して他の品詞のものは削除し、検索結果に影響を与えないようにした。

次に、取り出した名詞には全てカタカナの読みがなを付随させた。そして、番組内容の検索には単語単位で比較を行ない、ひらがなとカタカナ、漢字とカナを比較する場合のみ、読みがなを使用するようにした。単語の比較を全て読みがなで行わないのは、同音意義語の漢字の誤った一致を防ぐためである。

## 3.2 番組情報の分割

番組情報は複数の項目により構成されているが、これは変更の度合により以下のように区分できる。

- ある番組について、あまり変更のない項目  
{ 放送曜日、放送開始時間、放送局、ジャンル名、番組タイトル }
- ある番組について、頻繁に変更される項目  
{ 番組内容 (出演者名、放送内容) }

以後、前者を番組属性、後者を番組内容と呼ぶ。

### 3.2.1 番組評価値の改良

検索を行なう際は、番組表の番組情報と検索文字列を番組属性と番組内容とに分割し、別々に検索を行なう。

検索を別々に行なった結果、2種類の番組評価値が求められる。ある番組  $n$  に対し、番組属性について求められた番組評価値を  $P_A(n)$ 、番組内容について求められた番組評価値を  $P_C(n)$  とする。

番組属性に基づいている  $P_A(n)$  は、過去にユーザが視聴した事のある番組ほど大きくなる特性がある。一方、番組内容に基づいている  $P_C(n)$  は、視聴した事のある番組のものが大きくなるとは限らないが、ユーザの好む出演者や放送内容を含む番組について大きくなるという傾向がある。

これらの特性を活かすために、ユーザに推薦する番組を順位付けする際に用いる番組評価値  $P(n)$  は、それぞれの番組評価値を検索結果中の最大値で正規化し、個々の番組に於いて大きい方の値を用いる。

$$P(n) = \max \left[ \frac{P_A(n)}{\max\{V_i P_A(i)\}}, \frac{P_C(n)}{\max\{V_j P_C(j)\}} \right]$$

## 4 比較実験

### 4.1 実験方法

改良前と改良後のシステムの性能の変化を比較するための実験を行った。実験は以下のような条件で行った。

- 3人の被験者から3週間分の視聴履歴を収集
- 最初の2週間分の視聴履歴を視聴履歴データベースの初期データとし、それらは全て1位で推薦され、選択されたものと仮定
- 初期データの番組評価値、ユーザの入力する検索キーワードの評価値は全て定数とする
- 2つのシステムに同一の検索キーワードを10種類、25個入力し、最後の週に視聴した番組の推薦された順位を比較する

### 4.2 実験結果

最後の週に視聴した番組の推薦順位の平均順位を比較した場合は、改良前のシステムの平均順位が約4.8位であり、改良後のものが約2.6位となった。また、最後の週に視聴した番組が1位で推薦された確率は、改良前のものが48%、改良後が72%であった。

しかし、改良後のシステムでは順位付けの方式により、複数の番組が番組評価値  $P(n) = 1.0$  となって1位

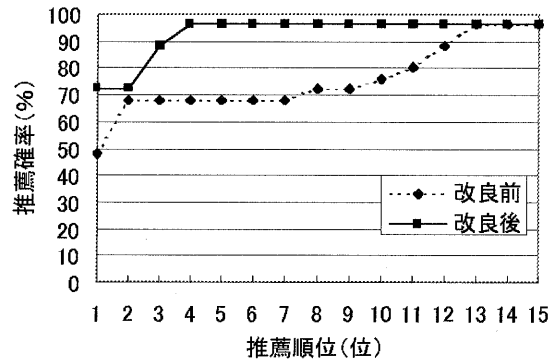


図 2: 視聴した番組がある順位以内に推薦される確率

に推薦される場合が多かった (全体の 88%)。最後の週に視聴した番組が 5 位以内に推薦された確率は、改良前で 68%、改良後は 96% となった (図 2)。ユーザの視聴する番組が上位で推薦される確率の高いほど性能が良いとすれば、改良後の方が性能が向上したと言える。

## 5 おわりに

TV 番組選択支援エージェント (Testa) は、多くの番組の中からユーザが希望する番組を推測し、推薦するシステムである。本稿では Testa の問題点を指摘し、その改良を行なった。その結果、システムの性能が向上した事を比較実験により示した。

今後は、複数の番組が 1 位に推薦されないよう、番組評価値の計算方法について改良を行なう。

## 参考文献

- [1] 金 淵培 “エージェント技術の放送への応用” 映像情報メディア学会誌 Vol.52, No.4, pp.447-451, 1998.
- [2] 村崎 康博, 柴田 正啓, 金 淵培, 山田 一郎 “エージェントテレビのための自動番組選択手法の検討” 1998 年映像情報メディア学会冬季大会, p83, 1998.
- [3] 矢川 雄一, 鶴飼 ひろみ, 田中 哲雄, 柴田 正啓, 金 淵培 “個人の嗜好に合ったテレビ番組を自動編成するエージェントの検討” 信学技報 AI98-55, pp9-16, 1998.
- [4] 中島 太郎, 渡辺 尚, 樽口 秀昭 “視聴履歴データベースを用いた TV 番組選択支援エージェントの評価について” 情報処理学会第 61 回全国大会講演論文集 (3), 3-377, 2000.
- [5] 松本 祐治, 北内 啓, 山下 達雄, 平野 善隆, 今一 修, 今村 友昭 “日本語形態素解析システム 茶釜 Ver.1.0 使用説明書” <<http://chasen.aist-nara.ac.jp>>, 1997.