

D-001

推薦番組の選択行動のパターン化に関する検討

Extraction of User Behavior Patterns to Choose TV Programs

古宮 弘智† 住吉 英樹† 後藤 淳† 佐野 雅規† 藤井 真人†

Hironori FURUMIYA Hideki SUMIYOSHI Jun GOTO Masanori SANO and Mahito FUJII

1. まえがき

我々は、現在視聴中の番組に対して付与されたメタデータを検索キーとして、システムが関連する番組を自動的に検索および推薦を行い、視聴者の多様な興味を満たす「コンテンツ推薦テレビ」[1]を提案している。

NHK オンデマンドでは 2010 年 6 月より、関連番組の推薦、提示サービスを行っている。このサービスでは一つの番組を選択すると、それに関連する番組が複数提示、推薦される。視聴者は提示された番組の中から、番組候補を選択できる。そして、この操作を繰り返すことで見たい番組を探索できる。この関連番組のリストの作成に、コンテンツ推薦テレビの関連番組検索技術を応用した実験システムを利用している。

今回我々は、よりの確かな関連番組の検索と推薦を行うため、NOD のパソコン向けサービスにおいてこの関連番組の選択行動に関する操作履歴を収集し、選択した番組候補からその次に選択した番組候補の遷移パターンを分析した。

関連番組のリストを作成する推薦システムは、番組内容を説明した文章(番組概要文)の類似に基づく関連番組検索、同一の出演者の番組、同じジャンルの番組などを組み合わせる関連番組リストを作成している。収集した履歴情報を確認すると、番組概要文の類似検索によって提示された関連番組のうち約 61%の番組が、類似度の 1 位以外の番組である事がわかった。そこで、類似検索で提示された各番組をジャンルによるカテゴリーに抽象化し、カテゴリーの遷移パターンの分析を試みた。その結果、特定の番組ジャンルでは、他のジャンルへの遷移に特徴があるなど推薦の最適化に有効と思われる行動があることが分かった。

本稿では第 2 章で現在の関連番組に関する推薦システムの概要について述べ、第 3 章で操作履歴の収集方法および番組候補の遷移パターンなどの実験結果・考察について述べる。第 4 章で試作システムの概要、第 5 章で本稿のまとめについて述べる。

2. 関連番組を推薦するシステムの動作概要

2.1 システム構成

システム概念図を図 1 に示す。システムは、NOD の表示に必要な情報を配信する「配信用メタデータ管理システム」、各コンテンツに対応するメタデータを蓄積する「メタデータサーバ」、メタデータを利用して関連するコンテンツを検索する「検索システム」で構成される。検索システムでは、内容の類似検索[2]の他、出演者やジャンルによる検索を行う。番組推薦装置では、推薦ポリシー(番組のジャンル毎に内容の類似検索、出演者検索、ジャンル

検索の割合を定義)に従い、推薦番組を検索システムから取得する。推薦番組の表示画面を図 2 に示す。

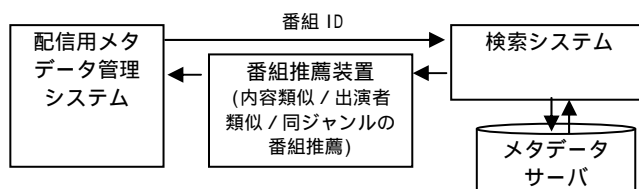


図 1 関連番組に関する推薦システムの概念図



図 2 NOD の関連番組の表示画面(赤枠)

3. クリックログの分析

3.1 推薦番組の選択行動のパターン化

NOD の関連番組(図 2)の画面から、以下の期間中に番組選択に関する行動データを収集した。

実験データの収集期間: 2010 年 9 月 ~ 2011 年 5 月

対象データ数(クリック数): 31743 サンプル

データの内容: 選択された番組 ID, その次に選択された関連番組 ID, クリック数

(番組を特定できない、見逃し見放題パックを含むクリック数を除外とした)

番組 ID から番組情報メタデータを参照し、選択行動のパターン化を試みた。一例として、番組ジャンルのパターン化を以下に示す。今回の実験では、NOD の担当者が番組メタデータの登録時に付与する番組のジャンルコードを利用した。各番組に対して、番組ジャンルコードは、最大 3 個付与されており、きめ細やかなカテゴリーの分類を行うため最大 3 個のジャンルを一つの特徴として扱った。これにより実際の番組に登録されている番組ジャンルコード

† NHK 放送技術研究所, NHK Science and Technology Research Laboratories

は、367種類となる。例えば、「NHKスペシャル 魔性の難問 リーマン予想・天才たちの闘い」では、「サイエンス/科学、ドキュメンタリー/社会、ドキュメンタリー/ヒューマン」となる。

3.2 分析と考察

(1) 類似度の有効性

収集した関連番組のクリックログのサンプル 31743のうち、概要文の類似検索に関する 14388 サンプルを抽出した。そのサンプルのうち、類似度 1 位以外のサンプルが約 61%(8737 サンプル)であった。これは、必ずしも番組概要文の類似度が視聴者の興味を的確に反映しているとは言えないことを示している。

(2) ジャンルによる遷移の偏り

上記の結果を受け、概要文の類似検索に関するサンプルについて、番組ジャンルによる遷移傾向を調査した。

最初に選択された番組のジャンルコード、その次に選択された番組毎に遷移数を集計し、割合を求めた。多くの場合は同一ジャンルが選択される傾向が強いが、ジャンルによっては異なるジャンルが選択される傾向の強いものもある。ここでは代表として「こころとからだ/宗教、ドキュメンタリー/社会、人物/著名者」の番組選択後の行動パターンについて表 1 に、「ドキュメンタリー/社会、歴史/歴史」の番組選択後の行動パターンについて表 2 に示す。

表 1 「こころとからだ/宗教、ドキュメンタリー/社会、人物/著名者」の番組選択後の行動パターン 母数:389

次に選択した番組候補	割合(%)
ドキュメンタリー/文化、自然/紀行、歴史/歴史	93.3
こころとからだ/宗教、ドキュメンタリー/社会、人物/著名者	2.3
ドキュメンタリー/文化、歴史/歴史、自然/紀行	1.8
人物/著名者、ドキュメンタリー/社会、ドキュメンタリー/教育	0.8
ドキュメンタリー/文化、歴史/歴史	0.5
こころとからだ/宗教	0.5
自然/気候	0.5
芸術/文学・文芸	0.3

表 2 「ドキュメンタリー/ビジネス、ドキュメンタリー/環境」の番組選択後の行動パターン(一部抜粋) 母数:74

次に選択した番組候補	割合(%)
ドキュメンタリー/社会、ドキュメンタリー/時事問題、ドキュメンタリー/海外	44.6
ドキュメンタリー/ビジネス、ドキュメンタリー/環境	21.6
ドキュメンタリー/ビジネス、ドキュメンタリー/海外	12.1
ドキュメンタリー/ビジネス、サイエンス/科学	9.5
サイエンス/科学、ライフスタイル/生涯教育	8.1
ドキュメンタリー/ビジネス	4.1

これらの表から以下のことがわかる。

- ・表 1 「こころとからだ/宗教、ドキュメンタリー/社会、人物/著名者」の場合、そのほとんどが次の番組ジャンルとして「ドキュメンタリー/文化、自然/紀行、歴史/歴史」を選択しており、推薦リスト作成時に当該ジャンルの推薦を優先すべきことがわかる。

- ・表 2 「ドキュメンタリー/ビジネス、ドキュメンタリー/環境」の場合、表 1 のような極端な偏りはないものの、同様のテーマの社会や時事問題、海外に関する内容の番組ジャンルに興味を持ち、番組を選択されていることがわかる。

このようにジャンルによっては、同一ジャンルではないジャンルが多く選択される傾向にある事がわかった。

4. リランキングの評価プログラム

今回の分析結果をもとに、以下の手順に従い、選択行動パターンデータを用いたリランキングプログラムの試作を行っている。概念図を図 3 に示す。このプログラムは、関連番組リスト作成システムで動作する。

- ・各番組 ID をキーとして、推薦番組の候補となる関連番組 ID リストを取得する。
- ・推薦番組の候補となる番組 ID が属するジャンルコードを番組情報 DB より取得する。
- ・番組ジャンルのパターン化テーブルで、遷移の割合によって推薦番組のリランキングを行い、この情報を推薦番組の候補リストとする。

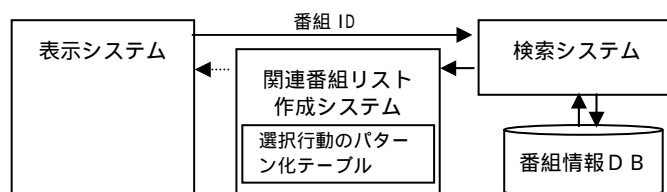


図 3 試作システムの概念図

5. まとめ

本稿では、NOD のパソコン向けサービスにおいてこの推薦番組の選択行動に関する操作履歴を収集し、選択した番組候補からその次に選択した番組候補の遷移パターンを分析した。分析の結果、特定のジャンルでは内容の類似検索に加えて、番組ジャンルによる選択行動のパターン化、推薦順位のリランキングが必要であることが示唆された。これらのパターンは多数のサンプルの収集により明確になってきたことも多く、引き続きデータの収集と分析を継続する。

今後はジャンル以外のカテゴリーについても分析を進めるとともに、番組内容に適した番組の検索、推薦手法の検討およびその評価を行っていきたい。

参考文献

- [1] 藤井他：“情報検索を使う新しい視聴スタイル CurioView”，映像情報メディア学会年次大会，6-5(2009)
- [2] J. Goto, H. Sumiyoshi, M. Miyazaki, H. Tanaka, M. Shibata, A. Aizawa：“Relevant TV Program Retrieval using Broadcast Summaries”, Proc. ACM Int. Conf. on Intelligent User Interfaces, pp411-412 (Feb.2010)