

## テレビ番組の視聴履歴を用いた番組推薦システムの構築 A Recommender System of TV Programs based on the Viewer's Watching History

井川 一樹<sup>1</sup> 福原 知宏<sup>2</sup> 藤井 秀樹<sup>3</sup> 武田 英明<sup>4</sup>  
Kazuki IKAWA Tomohiro FUKUHARA Hideki FUJII Hideaki TAKEDA

東京大学工学部<sup>1</sup> 東京大学人工物工学研究センター<sup>2,3,4</sup> 国立情報学研究所<sup>4</sup>  
Faculty of Engineering Research into Artifacts, Center for Engineering National Institute of Informatics  
The University of Tokyo The University of Tokyo

### 1. はじめに

電子番組表 (EPG) の普及に伴い、テレビ番組のデータを計算機上で処理できるようになった。テレビ番組の EPG と視聴履歴のデータを組み合わせることで、利用者の嗜好に応じた番組推薦が可能となる。

テレビ視聴者の選局行動については以前より研究がなされている [1] が、本研究ではテレビ番組の EPG と視聴履歴を用いた番組推薦システムを構築し、実際の視聴者に対して番組推薦実験を行ったので、その結果を報告する。

### 2. 番組推薦手法

#### 2.1 使用するデータ

番組推薦の評価実験として、ある多チャンネルテレビ放送の 2009 年 11 月と 12 月の番組表と 907 加入世帯の視聴履歴を利用した。まず 11 月の地上波放送、BS 放送、多チャンネルテレビ放送の 2 週間 (A 期間) の番組の視聴履歴を番組推薦のための解析データとして取り扱い、その後 1 週間分 (B 期間) の多チャンネルテレビ放送の番組に限定し、1 日あたり最大 5 番組を推薦、得られた約 600 世帯のデータを実験に利用した。推薦手法は (キーワード推薦)、(出演者推薦)、(キーワード+出演者推薦)、(時間帯+チャンネル推薦) の 4 種類である。本実験では都合上、番組放送前に推薦を行うのではなく、番組放送後にアンケート調査を行い、システムが推薦した番組について見たか、見なかったか、また興味があったか、なかったかを尋ねた。

#### 2.2 番組表からのキーワード、出演者名抽出

全番組のタイトル・説明文のテキストから形態素解析エンジン MeCab<sup>1</sup> を用いてキーワード及び出演者名を抽出するプログラムを作成した。

キーワードは、各番組の説明文から [人名]・[数] 以外の [名詞] を抽出した後に正規表現を用い、ひらがなのみ・記号のみで構成されるものを除外したものである。その結果、番組表に記載されている全番組から 8,315 個のキーワードが抽出され、当該全キーワードを実験に用いた。抽出したキーワードはデータベースに格納され、各キーワード

とそのキーワードを含む番組とはデータベース中で関係づけられている。

また出演者名は、キーワードと同様に各番組のテキストから [人名] の [名詞] を抽出する。その際、[人名-一般] と判別されるものはそのまま出演者名とし、[人名-姓]・[人名-名] と別々に判別されるものはそれぞれを結合して出演者名と見なした。一方、出演者名には “SMAP” のように人名ともグループ名とも一概に判別しづらい名称も多いため、あらかじめ Wikipedia<sup>2</sup> で人名・グループ名として登録されている項目名から辞書を作成して MeCab に追加辞書として学習させた。その結果 A 期間の全番組表から 4,371 人の出演者名が抽出され、当該出演者全員を実験に用いた。抽出した出演者名はデータベースに格納され、各出演者名とそのキーワードを含む番組とはデータベース中で関係づけられている。

#### 2.3 キーワードと出演者による推薦

各世帯の視聴履歴から A 期間の間に視聴した全番組のキーワード・出演者に対する選好度を算出する。その後、B 期間の間の全番組に対してキーワード・出演者に基づく推薦度を求める。式 1, 2 にキーワード  $k$  による番組選好度  $P_{key}(k)$  ならびに推薦度  $R_{key}(title)$  の算出方法を、式 3, 4 に出演者  $p$  による番組選好度  $P_{person}(p)$  ならびに推薦度  $R_{person}(p)$  の算出方法を示す。式 5 にキーワードと出演者の双方を用いた番組推薦度  $R_{mix}(title)$  の算出方法を示す。

式 1, 2 において  $Watched(x)$  はキーワードまたは出演者  $x$  を含む番組のうち、視聴された番組集合を返す関数を、 $Programs(x)$  はキーワードまたは出演者  $x$  を含む番組集合を返す関数である。また式 2, 5 において  $Keywords(title)$ 、 $Persons(title)$  はそれぞれ番組  $title$  のキーワード集合と出演者集合を返す関数である。式 4, 5 において  $PERSONS$  は全出演者名の集合、 $TITLES$  は全番組の集合である。

#### 2.4 時間帯とチャンネルによる推薦

キーワード・出演者による推薦の優位性を比較するために、時間帯とチャンネルによる確率的な

<sup>1</sup> MeCab: <http://mecab.sourceforge.net/>

<sup>2</sup> Wikipedia: <http://ja.wikipedia.org/>

$$P_{key}(k) = \frac{|Watched(k)|}{|Programs(k)|} \dots (式1)$$

$$R_{key}(title) = \sum_{k \in Keywords(title)} P_{key}(k) \dots (式2)$$

$$P_{person}(p) = \frac{|Watched(p)|}{|Programs(p)|} \dots (式3)$$

$$R_{person}(p) = \sum_{p \in PERSONS} P_{person}(p) \dots (式4)$$

$$R_{mix}(title) = \frac{R_{key}(title)}{\max_{t \in TITLES} R_{key}(t)} + \frac{\sum_{p \in PERSONS(title)} R_{person}(p)}{\max_{p \in PERSONS} R_{person}(p)} \dots (式5)$$

推薦を導入した。各世帯の視聴履歴から A 期間の間に視聴した全番組の時間帯とチャンネルに関する確率分布を求め、その確率に応じて推薦番組を決定した。

### 3. 推薦結果

#### 3.1 番組構成・世帯の視聴状況

解析対象の A 期間の全番組数は 54,990 番組で、そのうち地上波・BS 放送は 25 チャンネルで 14,366 番組、多チャンネルテレビ放送は 69 チャンネルで 40,624 番組であった。推薦期間である B 期間の多チャンネルテレビ放送は 69 チャンネルで、番組数は 11,993 番組であった。平均キーワード選好度は 0.0029、平均出演者選好度は 0.0232 であった。

#### 3.2 推薦方式別回答結果

- (1) キーワードによる推薦(112 世帯に推薦した)
- (2) 出演者による推薦(141 世帯に推薦した)
- (3) キーワードと出演者による推薦(132 世帯に推薦した)
- (4) 時間帯とチャンネルによる推薦(220 世帯に推薦した)

の 4 類型で番組推薦を行い、アンケートの回答結果を比較する。ユーザに対しては「見た・面白かった」「見た・つまらなかった」「見ていない・見たかった」「見ていない・見たくなかった」の 4 段階の評価とした。それぞれのアンケート結果を図 1 に示す。

「見た・面白かった」の結果はキーワード、出演者、キーワードと出演者、時間帯とチャンネルによる推薦類型でそれぞれ 9.9%、3.1%、6.2%、13.9%であった。「見た・つまらなかった」の結果はそれぞれ 0.5%、0.4%、0.4%、1.1%であった。「見ていない・見たかった」の結果はそれぞれ 17.5%、12.1%、16.7%、16.3%であった。「見ていない・見たくなかった」の結果はそれぞれ 72.0%、84.4%、76.6%、68.7%であった。

### 4. 推薦手法の評価

#### 4.1 評価方法

番組推薦の評価方法として、推薦番組の視聴の

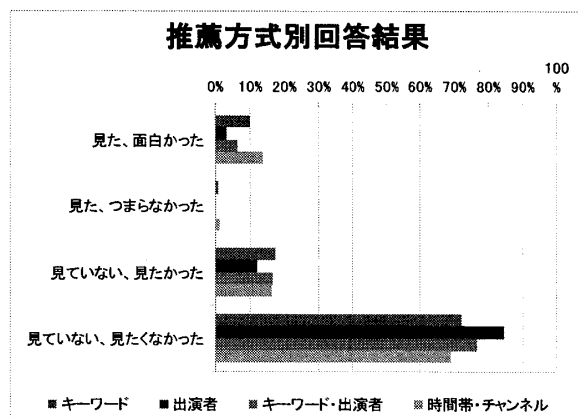


図 1. 推薦方式別回答結果

有無と推薦番組に対する興味の有無の 2 通りの指標を用いた。前者の評価方法では「見た・面白かった」と「見た・つまらなかった」の項目が該当し、後者の評価方法では「見た・面白かった」と「見ていない・見たかった」の項目が該当する。

#### 4.2 推薦類型別比較

推薦番組の視聴の有無に基づく評価では時間帯とチャンネルによる確率的な推薦の結果が最も高かった。その理由としては推薦の手法にユーザのライフスタイルを考慮している点があげられる。つまり、今回実験に参加したユーザが日常的に視聴している番組を推薦する傾向が強いことに起因する。ただ普段から視聴している番組を推薦することが必ずしもユーザの満足度を向上させるとは限らない。一方で推薦番組に対する興味に基づく評価では、キーワードによる推薦の結果が最も高かった。これは視聴履歴から抽出したキーワードによる選好度を採用することにより、ユーザが持っている潜在的な興味を拾い上げることができたためであると考えられる。

### 5. 総括と今後の展望

本論文ではテレビ番組の EPG と視聴履歴を用いた番組推薦手法を提案し、その性能を検証した。予測という観点では時間帯とチャンネルによる推薦手法が有効であり、推薦という観点ではキーワードによる推薦手法が有効であることが分かった。今後の展望としては、チャンネルと時間帯による推薦とキーワードの推薦を複合させ、生活スタイルを反映させたアルゴリズムに変更する。また多チャンネルテレビ放送に限定せず地上波放送と BS 放送も含めた番組推薦と評価も検討する。

#### 参考文献

- [1]黒木修隆, 廣瀬裕二, 鈴木達也, 片岡充照, 沼昌宏, 山本啓輔, “テレビ視聴者の選局行動に基づく番組嗜好度の推定”, 映像情報メディア学会誌 (2006).