

J-044

# デジタルテレビジョン放送における効率的な番組検索手法に関する一検討

## A Study of Efficient Method for Searching Digital Television Program

高橋 寛明†      佐藤 利光†      志田 哲郎†      小坂 英明†  
Hiroaki Takahashi    Toshimitsu Sato    Tetsuro Shida    Hideaki Kosaka

### 1. はじめに

日本におけるデジタルテレビジョン放送では放送信号に番組情報が重畳されており、これを利用した電子番組表の表示や、ジャンルによる番組検索ができる。

ジャンルによる番組検索では、番組情報として付与されたジャンルを利用し、ユーザが番組を探す際に、所望する番組の番組名や放送時間を明確に知らない場合でも、番組内容から想定されるジャンルを指定することで所望の番組を探すことができる。

しかしながら、実際の番組情報には似たような意味を持つジャンルが複数存在し、ジャンル間の直交性は保証されておらず、また、各番組にジャンルを付与する条件も明確に規定されていないため、番組へのジャンル付与には恣意的に運用される余地がある。

そのため、従来のジャンル検索では、ユーザの指定したジャンルが、ユーザの所望する番組に付与されていない可能性があり、その場合、ユーザが所望する番組は検索結果に含まれないという問題がある。

そこで、本稿ではジャンル間の関連度を定義し、関連度を利用した番組検索手法を提案する。本手法により、従来のジャンル検索の問題であった、番組の取りこぼしを防止し、ユーザは何度もジャンルを選択することなく効率的に番組検索を行うことができる。

### 2. ジャンル間の関連度を利用した番組検索

#### 2.1 概要

日本の地上デジタルテレビジョン放送において番組に付与可能なジャンルは社団法人電波産業会 (ARIB) による規格“地上デジタルテレビジョン放送運用規定 (ARIB TR-B14)”によって規定されており、大分類・中分類と階層化され、中分類は 104 種類である。また、ジャンルは 1 つの番組に対して最大 3 つまで付与することができる。

前節で述べたように、ジャンルは必ずしも直交性が保証された分類がなされているわけではなく、例えば「旅・釣り・アウトドア」と「歴史・紀行」といったように互いに類似したジャンルが存在する。こうした類似ジャンルの使い分けに関して明確な基準がなく、また、ジャンルを含む番組情報は放送局が個別に付与するものであるため、例えば、図 1 に示すように旅に関する内容の番組であっても、番組 A に付与されているジャンルは「旅・釣り・アウトドア」であったり、番組 B に付与されているジャンルは「歴史・紀行」であったりと、番組によって付与されているジャンルが異なる場合がある。

そのため、ユーザが旅に関する番組を検索する際、従来のジャンル検索では、ユーザは「旅・釣り・アウトドア」もしくは「歴史・紀行」のどちらか一方のみを指定するこ

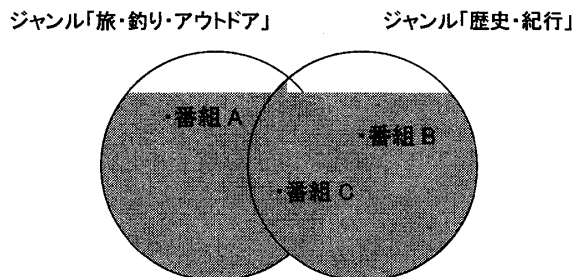


図 1 : 番組に付与されるジャンル

とになり、例えばユーザが「旅・釣り・アウトドア」を選択した場合、「歴史・紀行」しか付与されていない番組 B は検索結果として提示されず、取りこぼしが発生する。

そこで今回提案する手法は、ジャンル間の関連性を関連度として定義し、ユーザが指定したジャンルに対し、関連度が一定以上であるジャンルを検索結果に含めることで、ユーザの所望する番組の取りこぼしを防ぐ。

#### 2.2 関連度の定式化

前節で説明したように、地上デジタルテレビジョン放送では、1 つの番組に対し、複数のジャンルが付与できるため、図 1 における番組 C のように 1 つの番組に複数のジャンルが付与されている場合、それらのジャンルは互いに関連性の強いジャンルであるといえる。そこで、ジャンル間の関連度  $f_{ij}$  を式(1)のように定義する。

$$f_{ij} = \frac{N_{ij}}{N_i} \quad \text{式(1)}$$

式(1)において、 $f_{ij}$  はジャンル  $g_i$  とジャンル  $g_j$  ( $i, j = 1, 2, \dots, n; n$  はジャンルの総数,  $i \neq j$ ) について、ジャンル  $g_i$  に対するジャンル  $g_j$  の関連度であり、ジャンル  $g_i$  が付与された番組の中でジャンル  $g_i$ 、ジャンル  $g_j$  が同時に付与されている番組の割合を表す。ただし、 $N_i$  はジャンル  $g_i$  が付与された番組の総数を表し、 $N_{ij}$  はジャンル  $g_i$ 、 $g_j$  の両方が付与された番組の総数を表す。

例えば、実際の番組情報で「旅・釣り・アウトドア」が付与された番組数が 28 あり、このうち「旅・釣り・アウトドア」と「歴史・紀行」の両方が付与されている番組数は 18 であったとき、 $g_1$  を「旅・釣り・アウトドア」、 $g_2$  を「歴史・紀行」とすると、 $N_1 = 28$ 、 $N_{12} = 18$  であるから、 $f_{12} = 0.6429$  が得られる。

#### 2.3 番組検索手法

前節で定義したジャンル間の関連度を用いた番組検索手法を以下に説明する。本手法では、ジャンルによる番組検

†三菱電機 (株) 先端技術総合研究所

索を行う際に、ユーザが指定したジャンルと一定以上の関連度を有するジャンルを関連性のあるジャンルとして抽出し、ユーザの指定したジャンルと、関連性のあるジャンルのいずれかが付与された番組を抽出し、検索結果として提示する。

番組検索処理の手順を図2に示し、以下に説明する。

- ① ユーザが検索を行うためにジャンル  $g_u$  を指定する。
- ② ユーザが指定したジャンル  $g_u$  と、 $g_u$  以外の全てのジャンルの組み合わせに対して関連度を算出する。すなわち、 $f_{uk}$  および、 $f_{ku}$  ( $k=1,2,\dots,n$ ;  $n$  はジャンルの総数,  $k \neq u$ ) を算出する。
- ③ 閾値  $T$  とし、次式を満たす  $k$  を全て求め、 $g_k$  を  $g_i$  と関連性のあるジャンルとして抽出する。

$$f_{uk} > T \text{ かつ } f_{ku} > T \quad \text{式(2)}$$

- ④ ユーザが指定したジャンル  $g_u$  と、③の手順で求めた、 $g_u$  と関連性のあるジャンル  $g_k$  のうち、いずれかが付与された番組を抽出し、検索結果として通知する。

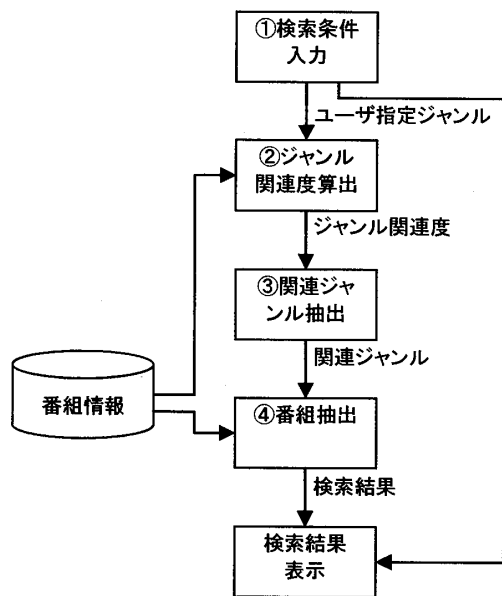


図2：検索処理の流れ

### 3. 番組検索手法の評価

#### 3.1 実行例

本検索手法のプロトタイプシステムをPC上に構築し、2008年10月28日から11月4日までの一週間に京都地区で放送された地上波デジタルテレビジョン放送の番組を対象として、本提案手法を適用した実行例を示す。

まず、取得した番組情報から、各ジャンルについて関連度を求めた。関連度が高い値を示したジャンルの組み合わせと、その関連度の一部を表1に示す。

次に、構築した番組検索システムを用いて、検索を行う動作例を以下に示す。なお、閾値  $T$  は事前実験により、1つのジャンルに対して関連性があるとみなされるジャンルの数が最適となると考えられる値を求め、0.1とした。

表1：ジャンル間の関連度（一部抜粋）

$g_i$	$g_j$	$f_{ij}$	$f_{ji}$
カルチャー・伝統文化	音楽・美術・工芸	0.1818	0.2069
グルメ・料理	暮らし・住まい	0.1806	0.1215
グルメ・料理	料理バラエティ	0.1250	0.6429
海外ロック・ポップス	国内ロック・ポップス	0.8667	0.3611
国内ロック・ポップス	音楽バラエティ	0.1944	0.2500
社会・時事	ドキュメンタリー全般	0.1860	0.1905
歴史・紀行	旅・釣り・アウトドア	0.2069	0.6429

例として、ユーザが、旅に関する番組である「日本の絶景再発見」を検索しようとしたとする。旅に関するジャンルとして「旅・釣り・アウトドア」や「歴史・紀行」が想定されるが、この番組に付与されているジャンルは「歴史・紀行」のみであり図1における番組Bに相当する。

そのため、従来のジャンル検索では、ユーザが「歴史・紀行」を選択した場合は検索結果として提示されるが、「旅・釣り・アウトドア」を選択した場合は提示されない。

一方、本手法によって検索を行う場合、ユーザが「旅・釣り・アウトドア」を指定した際、「旅・釣り・アウトドア」と、「旅・釣り・アウトドア」との関連度が閾値0.1よりも高い「歴史・紀行」のいずれかを含む番組を抽出しユーザに提示することで、ユーザの所望する番組を含む検索結果を得ることができる。

#### 3.2 有効性評価

本検索手法の有効性を確認するために、無作為に選んだ50の番組について、16人の被験者が番組名と番組内容から想定されるジャンルを選択し、従来のジャンル検索の場合に該当の番組が含まれる割合と本検索手法によるジャンル検索の場合に該当の番組が含まれる割合を比較した。

結果、従来のジャンル検索では検索結果に番組が含まれる割合は67.1%であったのに対し、本手法によるジャンル検索では77.5%と上昇し、取りこぼしの発生を減少させることができるため、本検索手法の有効性が確認できた。

以上のように、従来のジャンル検索ではユーザの想定するジャンルが付与されていないために検索結果として得られなかった番組を、本検索手法を適用することにより検索結果として得ることができ、取りこぼしを防止することが可能となる。そのため、ユーザは何度もジャンルを選択することなく効率的に番組検索を行うことができる。

#### 4. まとめ

本稿ではジャンル間の関連度を定義し、これを利用して効率的に番組検索を行う手法を提案した。

提案手法を用いることで、ユーザが想定したジャンルと、実際に番組に付与されているジャンルが異なる場合でも、関連するジャンルを検索結果に含めることで、ユーザの所望する番組の取りこぼしを防止することが可能となる。

また、本稿では、互いに関連性があるジャンルとするためのジャンル間の関連度の閾値を実験的に求めた値に設定したが、番組数などに応じて最適な値を算出することで、より効率的な番組検索が可能となると考えられるため、今後は閾値の算出法について検討していきたい。