自然言語での番組検索における意味フレーム外キーワードを用いた番組ジャンル推定

TV Genre Estimation with Keyword out of Semantic Frame in TV Program Retrieval based on Natural Language Processing

有賀 康顕[*1] Michiaki Ariga
藤井 寛子[*1] Hiroko Fujii
若木 裕美[*1] Hiromi Wakaki
筒井 秀樹[*2] Hideki Tsutsui
鈴木 優[*1] Masaru Suzuki
住田 一男[*1] Kazuo Sumita

[*1] (株)東芝 研究開発センター Corporate R&D Center, Toshiba corporation
[*2] (株)ニューズウォッチ NewsWatch, Inc

In a natural language interface for TV program retrieval, it is important to obtain user's intention so that they can find the adequate program. TV genre is important information to reduce the number of retrieved programs. In this paper, we propose a method that estimates intended genre of TV program from utterances. The method use frequently-appearing Named Entities and words. We verified effectiveness of proposed method.

1. はじめに

近年，大容量 HDD 録再機の普及，多チャンネル化の促進により，多量の映像を録画し，いつでも再生，視聴ができる環境が整ってきた．その一方で，番組数の増加や機能の複雑化にともない，一般ユーザにとって録画予約などの操作が難しくなり，よりわかりやすいインターフェースが求められている．こうした中で我々は，音声認識を利用した自然文での番組検索を実現するためのシステムを開発している．

自然文での番組検索では，少ないやりとりで目的の番組にたどり着くために，発話者の意図を理解して，発話の内容から検索に必要な情報を抽出することが重要である．そのための手法としては，従来，発話の内容から抽出した情報を事前に設計した意味フレームに割り当てることが行われてきた[Konashi 2004]．しかし，実際には全ての意味フレームを予め設計しておくことは難しい．そこで，設計した意味フレームに割り当てられないキーワードからもユーザの意図を理解する何らかの情報を抽出することが必要であると考えられる．

本研究では，意味フレームに割り当てられないキーワードを利用して，発話者が検索したい番組のジャンルを推定する方法を検討する．一般的に，番組を検索する際にジャンルは有効な検索条件であると考えられる．しかし，音声で番組予約を行う場合の実際の発話文は短く，必ずしも番組ジャンルを示す具体的なキーワードは含まれない．また，発話者が想定する番組ジャンルと番組に付与されている番組ジャンルの間にずれがあることも多い．そのため，短い発話文から適切な番組ジャンルを推定することが求められる．本稿では，発話文に含まれる表層語と固有表現クラスから番組ジャンルに関連するキーワードを取得し，発話者の意図する番組ジャンルを推定する手法を提案し，提案手法の精度について評価する．

連絡先：有賀康顕，(株)東芝研究開発センター，神奈川県川崎市幸区小向東芝町 1，044-549-2240，044-520-1308，michiaki.ariga@toshiba.co.jp

2. 番組検索における発話文解析

放送番組に関する情報は，電子番組表(EPG: Electronic Program Guide)により取得できる．EPG には番組名，出演者名，番組ジャンル名，放送時間などの情報が含まれている．これらの情報を番組検索の条件として発話文から抽出するために，発話文に含まれる番組名などの情報を後述する意味フレームに格納する．意味フレームに格納された情報を用いて番組検索を行うが，必ずしも一度で所望の番組にたどり着くとは限らない．少ないやりとりで目的の番組にたどり着くために，ユーザの意図に合った検索条件をできる限り抽出することが重要である．そのため，設計した意味フレームに割り当てられないキーワードであっても，検索対象を絞り込むために利用することが必要である．

本稿では，意味フレーム外のキーワードの利用方法の一つとして，キーワードと番組ジャンルの関連づけを行う．番組ジャンルは，ユーザが多くの番組から所望の番組へと絞込を行うときに，頻繁に利用される情報である．しかし，実際の発話文は「世界遺産の番組が見たい」のように短く，番組ジャンルを表すキーワードが含まれていないことも多い．また，EPG に含まれる番組ジャンルはサブジャンルまで含めると 100 以上あり，発話者が番

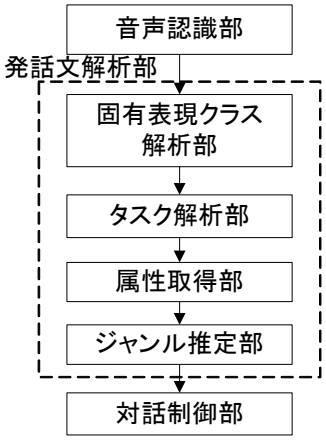

図 1:発話文解析部フロー図

組ジャンルを指定しようとしたときに，どのような番組ジャンルが存在して，ユーザの意図する番組はどの番組ジャンルに含まれているのかを把握するのは困難である．そこで，発話文に含まれているキーワードから，適切な番組ジャンルを推定する．

図 1 に発話文解析部のフロー図を示す．発話文を音声認識した結果として，テキストが発話文解析部に入力される．入力されたテキストを形態素解析した上で固有表現クラスを抽出し，タスク解析及び固有表現クラスの付与されたキーワードの属性を取得する．更に，取得した属性から番組ジャンルを推定し，最終的に対話制御部に受け渡す意味フレームを決定する．

3. 提案手法

本稿で用いる意味フレームについての説明を述べ，提案するジャンル推定手法について説明する．

3.1 意味フレーム

本稿で用いる意味フレームは，発話文からユーザの意図を解析した結果を格納した構造体であり，大きくタスクと属性の 2 つの情報から成る．ここでは，テレビ番組の録画再生についてのタスク及び属性を述べる．

本稿では，テレビ番組の録画再生に必要なタスクとして，「録画」，「再生」，「視聴」，検索結果や番組表を表示する「一覧表示」の 4 つを定義した．

属性は，全タスクに共通な「番組名」，「人物名」，「放送日」，「開始時間」，「終了時間」，「ジャンル」，「放送局」，「期間」，「時間帯」，「時間情報」，「時間長」，「その他」の 12 の属性と，タスクに依存した「回数情報」，「繰り返し情報」の 2 つの属性の計 14 の属性を定義する．属性値は，固有表現クラスが付与されたキーワードを利用して取得する．

「番組名」，「人物名」，「放送日」，「開始時間」，「終了時間」，「ジャンル」，「放送局」は，録画予約に用いられる基本的な属性である．「期間」は「放送日」及び「開始時間」，「終了時間」を絞り込むための“今週”，“来月”などの情報を，「時間帯」は“今晩”，“午前”などの 24 時間以内の時間の情報を表す．「時間情報」はタスクの絞り込みを行うための“過去”，“未来”といった時制に関する情報を表す．「時間長」は再生時の要求として挙がる，番組の放送時間の長さを表す．タスク共通でない「回数情報」は“最終回”などの番組回数の情報，「繰り返し情報」は“毎週”などの繰り返し録画に用いる情報である．

上記どの属性にも含まれないキーワードを，「その他」属性として取得する．番組検索では重要となるキーワードが多岐にわたるため，どの固有表現クラスが重要かをあらかじめ網羅することは難しい．そのため，「その他」属性のキーワードを意味フレーム外キーワードとして，ユーザの意図する情報の取得に利用する．

3.2 表層語を用いたジャンル推定手法

表層語を用いたジャンル推定は，発話文中から EPG のジャンル表記が取得できなかった場合に，発話文のキーワードを用いて番組ジャンルを推定する．なお，本稿では形態素解析には MeCab[工藤 2008]を用い，表層語として人名を除く名詞を利用する．

表層語を用いたジャンル推定の手順は，以下の通りである．まず，事前に EPG データ中の表層語と番組ジャンルの関連性を取得する． EPG の番組ジャンル毎に特徴的な表層語の抽出を行うことで，ある表層語に関連の強い番組ジャンルを推定する．表層語には，TF・ICF(Term Frequency ・Inverse Category Frequency) [Cho 1997]によりジャンルとの関連性を重みづける．これは，一般に索引語の重み付け手法としてよく知られた，TF・IDF(Term Frequency Inverse Document Frequency)を拡張したものである． TF・IDF は単語の出現頻度（TF）と文書頻度の逆数(IDF)との積からなる．TF が単語の網羅性を表し，IDF が単語の文書に対する特定性を表している．そのため，TF・IDF を用いることで，網羅性と特定性が共に高い単語の重みが大きくなる．それに対し，TF・ICF は文書単位ではなく，カテゴリ単位での特定性を算出するため，カテゴリに対する重み付けに有用であるとされている．

番組ジャンル G_i 中の語 t の重み TF・ICF(t ,G_i)は次式のように算出する．

$$TF \cdot ICF(t, G_i) = TF(t, G_i) \cdot ICF(t) = \frac{f_{G_i}}{\sum_{G_i \in G'} f_{G_i}} \cdot \log\left(\frac{|G'|}{|G_t|} + 1\right) \quad (1)$$

この，f_{Gi} は語 t の番組ジャンル G_i での頻度，|G’|は全番組ジャンルの数，$|G_t|$は語 t を含む番組ジャンルの数である．TF・ICF は番組ジャンル毎の語の重要度を表す．

次に，事前に EPG データから番組ジャンル間の共起関係を取得する．EPG データには一番組に対して番組ジャンルが最大 3 つまで付与されており，発話者が把握してない番組ジャンル間の関係も含めるために，番組ジャンル間の共起関係を利用する．本稿では，[松尾 2005]より番組ジャンル間の共起を表す指標として，共起頻度，Jaccard 係数，Simpson 係数,閾値付き Simpson 係数を比較した結果，次式の閾値付きの Simpson 係数を利用する．

$$R_S(G_i, G_j) = \begin{cases} \dfrac{|G_i \cap G_j|}{\min(|G_i|, |G_j|)} & if\ |G_i| > th\ and\ |G_j| > th, \\ 1 & if\quad G_i = G_j \\ 0 & otherwise \end{cases} \quad (2)$$

この，$R_S(G_i,G_j)$は番組ジャンル G_i, G_j 間の関係の強さを表す関数であり，th は閾値である．本稿では th=5 とした．閾値付きの Simpson 係数を用いることによって，単独での出現頻度が非常に少ない番組ジャンルに対して特に高い値が出やすい Simpson 係数の欠点を解消できる．

最終的な語 t の番組ジャンル G_i に対するスコアを次式のように算出する．

$$Scorekey_{G_i}(t) = \sum_{G_K \in G'} TF \cdot ICF(t, G_k) \cdot R_S(G_k, G_i) \quad (3)$$

ただし，$Scorekey_{Gi}(t)$は語 t に対する番組ジャンル G_i の関連度，G’は全ての番組ジャンル集合である．この番組ジャンル関連度を番組ジャンル毎に算出し，値の高いものを推定番組ジャンルとして出力する．

3.3 固有表現クラスを用いたジャンル推定手法

表層語を用いたジャンル推定では人名を除く名詞を全て利用するため，発話文に頻出する口語的表現がある特定のジャンルにのみに現れ，ICF が極端に大きくなってしまうことがある．また，分析する EPG データに含まれない表層語は，ジャンル推定に利用することができない．これらの問題を解決するために，固有表現クラスを用いたジャンル推定を提案する．固有表現クラスを用いたジャンル推定は，発話文の表層から正解ジャンルに含まれるジャンル文字列が取得できなかった発話文に対し，その

他属性として取得されたキーワードの固有表現クラスを基に番組ジャンルを推定する. なお, 本稿では固有表現抽出にはASKMi[市村 2005]の固有表現抽出エンジンを用いた.

固有表現クラスを用いたジャンル推定の手順は, 以下の通りである. まず, 事前に EPG データの固有表現クラス解析を行い, 番組ジャンルと固有表現クラスの共起関係を取得しておく. 番組ジャンルと固有表現クラスとの共起関係を表す指標については, 固有表現クラスの出現頻度の偏りによるノイズの影響が少なかった Jaccard 係数を用いることとした. Jaccard 係数は, 固有表現クラス X と番組ジャンル G_i の単独での出現頻度をそれぞれ $|X|$,$|G_i|$とし, AND,OR をとったときの出現頻度をそれぞれ$|X \cap G_i|$,$|X \cup G_i|$とするとき, 次式となる.

$$R_J(X, G_i) = \frac{|X \cap G_i|}{|X \cup G_i|} \quad (4)$$

この, $R_J(X,G_i)$は固有表現クラス X と番組ジャンル G_i の関係の強さを表す.

発話文から抽出したその他属性に属する固有表現クラス X を取得した後, 式(2)の閾値付き Simpson 係数及び式(4)に示した Jaccard 係数を用いて, 次式のようにジャンル G_i との関連度を算出する.

$$Score_{G_i}(X) = \sum_{G_K \in G'} R_J(X, G_k) \cdot R_S(G_k, G_i) \quad (5)$$

ただし, $Score_{Gi}(X)$は固有表現クラス X に対するジャンル G_i の関連度, G'は全てのジャンル集合である. このジャンル関連度をジャンル毎に算出し, 値の高いものを推定 EPG ジャンルとして出力する.

3.4 表層語と固有表現クラスを組み合わせたジャンル推定手法

前節で提案した固有表現クラスを用いたジャンル推定は, 重要なキーワードをジャンル推定に利用できる反面, 固有表現クラスが取得できない発話文に対応ができない. そこで, 表層語と固有表現クラスを組み合わせたジャンル推定手法を提案する.

まず, 入力された発話文に対し, 固有表現クラスを用いたジャンル推定を行う. そして, 固有表現クラスが取得できない発話文に対しては表層語を用いたジャンル推定を行う. このように, 二つの推定手法を組み合わせた推定結果を用いることで, 重要なキーワードを利用しながらも再現率を高めたジャンル推定が行える.

4. 評価実験

テレビ番組の発話文からのジャンル推定について, 前章で述べた推定手法にてその精度を評価した.

4.1 実験条件

実験に必要な放送番組のデータは, EPG から収集した. 固有表現クラス及び表層語と番組ジャンルの関係の学習は, 2007 年 10 月 11 日から 18 日までの 2 週間の EPG データを基に行った. また, 発話文は上記 2 週間及び 2008 年 7 月 14 日から 21 日までの地上アナログ 7 放送局の EPG データを基に, 被験者のべ 72 人から収集した番組検索のための発話文の内 467 文に正解ジャンルを付与した. その中で今回対象となった発話文は 57 文であった. 正解ジャンルには, [ARIB 2008]にて定義されるジャンル大分類の内, “予備”, “拡張”, “その他”を除く 12 ジャンルとした. また, 正解ジャンルを付与した発話文例を図

発話文:「内藤対亀田のタイトルマッチを予約」

↓

推定ジャンル: “スポーツ”

図 2:ジャンル推定例

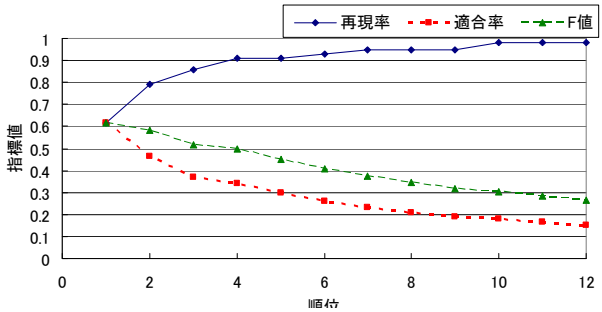

図 3:表層語によるジャンル推定の評価値

2 に示す. このように, 正解ジャンルについては発話文の情報から人手で判断した.

4.2 評価指標

ジャンル推定方法の評価を行うための評価指標について述べる. 推定された番組ジャンルは順位付きで出力されることが想定されるため, 情報検索の精度測定方法を利用する. 第 r 位までの推定結果を用いたタスクに対して平均の再現率を $MRec_r$, タスクに対して平均の適合率を $MPrec_r$, それらの調和平均である F 値を F_rとして次式に示す.

$$M\,\mathrm{Re}c_r = \frac{\sum_{k \in Test} |Cest_r(k) \wedge \mathrm{Re}l_k|}{|T_{est}|} \quad (6)$$

$$Pest_r(k) = \frac{|Cest_r(k) \wedge \mathrm{Re}l_k|}{r} \quad (7)$$

$$M\,\mathrm{Pr}ec_r = \frac{\sum_{k \in Test} Pest_r(k)}{|T_{est}|} \quad (8)$$

$$F_r = \frac{2}{1/M\,\mathrm{Re}c_r + 1/M\,\mathrm{Pr}ec_r} \quad (9)$$

ただし, データセット中のジャンル推定すべき発話文集合をT_{est},発話文集合中の発話文 k に対する r 位までの推定した番組ジャンル集合を $Cest_r(k)$, 発話文 k の番組ジャンル正解集合を Rel_kとする.

複数の検索質問に対する平均の性能評価方法として, 平均適合率の平均(Mean Average Precision: MAP)や平均逆順位(Mean Reciprocal Rank: MRR)が評価指標として用いられている[藤井 2007]. MRR は課題毎に正解が最初に見つかった順位の逆数(Reciprocal Rank: RR)を計算し, 全タスクの RR を平均した値である. MRR はより上位に正解が出現すれば値が高くなり, 正解数が少ない検索の評価に適している. そのため, 本稿では MRR も評価指標として用いる.

4.3 表層語によるジャンル推定実験

表層語を用いたジャンル推定を行った時の, ある順位以上の推定結果を利用した番組ジャンルの再現率, 適合率, F 値を図 3 に示す. 横軸はジャンル推定したものの順位の閾値であり, 縦軸はその順位までの推定結果を用いた場合の各指標値である. 上位 3 位まで利用した時, 再現率 0.86, 適合率 0.37, F 値 0.52

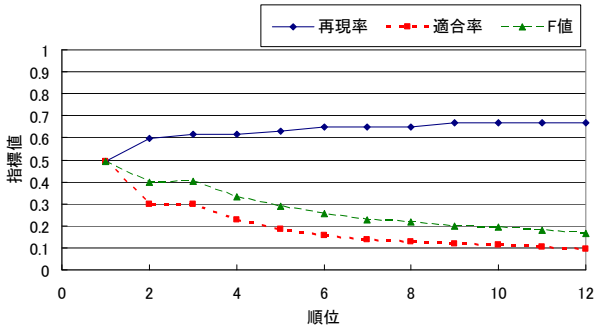

図 4:固有表現クラスによるジャンル推定の評価値

となった. MRR は 0.75 となり上位に正解が分布していることが分かる.

表層語を用いたジャンル推定を行うことで, ジャンルに関連する普通名詞を利用したジャンル推定が行えた. 表層語を用いて適切にジャンル推定できた例としては,「内藤対亀田のタイトルマッチを予約」といった発話文がある. この例では“タイトルマッチ”というキーワードを利用することで, “スポーツ”という番組ジャンルを 1 位に推定できた. 同様に,「サイエンスもの」という発話文に対しても, “サイエンス”というキーワードを利用することで, “ドキュメンタリー／教養”という番組ジャンルを 1 位に推定することができた. 一方,「日ハム戦見せて」という発話文での“日ハム”のようなEPGにあまり含まれていない略称や,「フィギュアのエキシビション」における“フィギュア”といった学習時期に含まれていないキーワードについては, 推定順位が低下する事例や推定できない事例が見られた.

4.4 固有表現クラスによるジャンル推定実験

固有表現クラスを用いたジャンル推定を行った時の, ある順位以上の推定結果を利用した番組ジャンルの再現率, 適合率, F 値を図 4 に示す. 横軸はジャンル推定したものの順位の閾値であり, 縦軸はその順位までの推定結果を用いた場合の各指標値である. 上位 3 位まで利用した時, 再現率 0.61, 適合率 0.30, F 値 0.40 となった.

利用する順位を拡大するときに適合率の低下が大きいのは, 正解とする番組ジャンルの数がたかだか 3 個程度までと少ないため, あまり多くの推定ジャンルを利用すると誤推定が増加してしまうからである. MRR は 0.56 となっており, 比較的上位に正解が分布していることが分かる. しかし, 再現率が全ジャンルを用いても 0.67 と低いのは, 固有表現クラスが付与されない発話文が 57 文中 19 文と多く存在するからである. 一方, 表層語を用いたジャンル推定では得意としなかった,「日ハム戦見せて」や「フィギュアのエキシビション」といった発話文では, “スポーツ”ジャンルを 1 位に推定できた. これは, 表層ではあまり含まれていないキーワードであっても, その固有表現クラスに抽象化することで適切にジャンル推定が行えたと考えられる.

4.5 表層語と固有表現クラスを組み合わせたジャンル推定実験

表層語と固有表現クラスを組み合わせた時の, ある順位以上の推定結果を利用した番組ジャンルの再現率, 適合率, F 値を図 5 に示す. 横軸はジャンル推定したものの順位の閾値であり, 縦軸はその順位までの推定結果を用いた場合の各指標値である. 上位 3 位まで利用した時, 再現率 0.91, 適合率 0.42, F 値 0.58 となった.

また, 1 位に推定された番組ジャンルを用いた適合率は 0.70 と高く, MRR が 0.81 となったことからも, 表層語の再現率の高さ

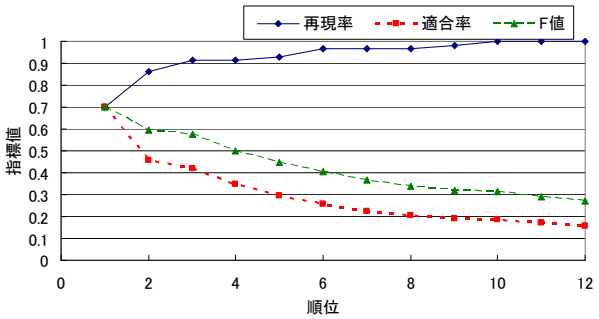

図 5:表層語と固有表現クラスを組み合わせたジャンル推定の評価値

を生かしながら, 固有表現クラスによる重要語を利用したジャンル推定が行えることが分かった. 本推定手法により, 上位の推定結果を用いることで, 適切な番組ジャンルを提示できることが分かった.

5. まとめと今後の課題

本稿では, 音声によって放送番組を検索するシステムにおいて, 予め設計した意味フレームに含まれないキーワードから発話者の意図する番組ジャンルを推定する手法を提案した. 表層語や固有表現クラスと番組ジャンルとの EPG 中での共起関係を事前に取得しておき, これらの情報を組み合わせることで, MRR が 0.81 と上位の推定ジャンルに正解が分布していることが分かった. また, 推定順位 1 位を利用した適合率は 0.70, 上位 3 位までの再現率は 0.91 と高く, 表層語の広い適用範囲を用いつつも, 固有表現クラスで発話者が意図した重要語を利用したジャンル推定が行うことができることが分かった.

今後は, 実際の番組検索による精度の評価を行っていく. また, ジャンル以外の検索条件についても同様に, ユーザの意図する情報を推定する手法を検討したい.

参考文献

[Konashi 2004] Konashi, T., Suzuki, M., Ito, A., and Makino, S.: A spoken dialog system based on automatic grammar generation and template-based weighting for autonomous mobile robots, in Proceedings of INTERSPEECH 2004, pp. 189–192 2004

[松尾 2005] 松尾豊,友部博教,橋田浩一,中島秀之,石塚満: “Web 上の情報からの人間関係ネットワークの抽出”, 人工知能学会論文誌, Vol.20, No. 1E, pp. 46-56, 2005.

[市村 2005] 市村由美,齋藤佳美,酒井哲也,國分智晴,小山誠: “固有表現抽出と回答タイプ体系が質問応答システムの性能に与える影響”, 信学論 D-II, Vol.J88-D-II, No. 6, pp. 1067-1079, 2005.

[Cho 1997] Cho, K. and Kim, J.: Automatic Text Categorization on Hierarchical Category Structure by using ICF(Inverted Category Frequency) Weighting, Proc. KISS Conference, pp. 507-510, 1997.

[工藤 2008] 工藤拓: MeCab : Yet Another Part-of-Speech and Morphological Analyzer, http://mecab.sourceforge.net/

[ARIB 2008] ARIB STD-B10 4.6 版, “デジタル放送に使用する番組配列情報標準規格”

[藤井 2007] 藤井敦:“Web 検索におけるアンカーテキストのモデル化と質問の自動分類”,IC2007,pp87-96,2007