原著論文

視聴者が主観的に感じる放送番組間の 関連性の強さに関する非線形予測式の最適化

高倉 潤也*, 関根 真弘*, 三原 功雄*, 山内 康晋*, 樋口 靖和**

* 株式会社東芝 研究開発センター, ** 株式会社東芝 ビジュアルプロダクツ社

Optimization of the Non-Linear Formula for Prediction of Subjective Relevance between TV Contents

Jun'ya TAKAKURA*, Masahiro SEKINE*, Isao MIHARA*,
Yasunobu YAMAUCHI* and Yasukazu HIGUCHI**

* Corporate Research & Development Center, Toshiba Corp., 1 Komukai Toshiba-cho, Kawasaki-shi, Kanagawa 212-8592, Japan

** Visual Products Company, Toshiba Corp., 2-9 Suehiro-cho, Oume-shi, Tokyo 198-8710, Japan

Abstract : We have developed a new content-finding navigation system “Roaming Navi”, in which subjective relevance between TV contents is accurately predicted by a complex non-linear formula. In order to optimize the formula, it was necessary, but difficult, to obtain enormous reliable subjective data on the relevance between TV contents. To tackle this problem, we proposed a new probabilistic method to handle subjective data, and the validity of the method was confirmed in empirical questionnaire data. Then, we conducted optimization of the prediction formula combining various optimization methods. As a result, the accuracy of the prediction was significantly improved. Moreover, these methods are broadly applicable in the discipline of Kansei engineering.

Keywords : Subjective evaluation, Probabilistic model, Non-linear formula

1. はじめに

今日、従来からの地上波放送のみならず、BS/CS放送、VOD (Video on Demand)、動画共有サイト、など様々な形態で膨大な量のコンテンツが存在する。しかし、膨大なコンテンツが存在することにより逆に、視聴者にとって興味のあるコンテンツを発見することが困難な状況となっている。

視聴者にとって興味のある番組を効率よく発見することのできる新たな番組検索インターフェースとして、我々は「ローミングナビ」というアプリケーションの開発を行った[1, 2]。ローミングナビの画面イメージを図1に示す。

図1の中心にあるサムネイル画像は、視聴者が選択したある番組（注目番組）を表す。その周囲に同心円状に配置されているサムネイル画像は、注目番組に関連する番組（関連番組）を表す。このとき、関連番組は注目番組との関連性が高いものほど、同心円の中心近くに配置される。視聴者は、ある興味のある番組に関連する別の番組にも興味がある可能性が高いので、本アプリケーションにより興味のある新たな番組を発見することを支援することが可能となる。

このアプリケーションを実現する上で、番組同士の関連性の強さをいかに精度良く算出できるかが重要なポイントとなる。しかし、番組同士が関連しているかどうかは視聴者の主観的な問題である。そこで、我々は視聴者が主観的に感じる番組同士の関連性の強さを予測する式の構築（最適化）を行った。

本論文では、この予測式の最適化を行うために新たに開発した主観評価手法および主観評価結果の定式化手法を紹介し、それぞれの有効性を示す。また、これらの手法を適用し予測式の最適化を行った結果、視聴者が主観的に感じる番組同士の関連性の強さの予測精度が向上したことを示す。

なお、本論文では上記の手法について述べることを目的とするため、ローミングナビ自体の説明は最小限に留める。ローミングナビの詳細に関しては、既報[2]を参照されたい。

2. 従来技術・手法の問題と本研究でのアプローチ

2.1 ローミングナビにおける関連性の強さの算出方法

ある番組に関連する番組を検索する従来手法には、協調フィルタリングという手法が存在する[3]。これは通販サイトなどでよく目にする「この商品を買っている人は、この

図1 ローミングナビの画面イメージ

商品も買っています」という推薦に用いられている技術である。しかし、この手法を番組検索で用いた場合、視聴者が既に視聴(あるいは録画予約)した番組の履歴に依存するため、マイナーな番組や新番組は推薦されにくい。しかし、番組検索インターフェースの目的は、むしろ、こういった番組を視聴者が発見できることである。

そこで、我々は履歴情報には依存せずに、番組に付与されているEPG (Electronic Program Guide) 情報を元に、番組同士の関連性の強さを算出することとした。EPG情報には、番組のタイトルやジャンル、あるいは番組の内容を表す文章など複数の項目が含まれている。まず、関連性の強さを算出したい2つの番組間で、それぞれの項目ごとの一致度合いを表す類似度を算出する。ここでは、タイトル類似度 (S_{title})、ジャンル類似度 (S_{genre})、人物類似度 (S_{person})、キーワード類似度 ($S_{keyword}$)、放送局類似度 ($S_{broadcaster}$)、放送時間類似度 (S_{time})、の6種類の類似度を用いることとした。これらの各類似度は、2つの番組間で各項目の一致度合いが大きいほど値が大きくなる量である。また、人物類似度とキーワード類似度は、EPG情報中の番組内容を表す文章の解析 [4, 5] を行うことにより算出する。たとえば、人物類似度 (S_{person}) は、2つの番組に同じ登場人物が3人いれば $S_{person} = 3$ のように算出する [2]。なお、単位はそれぞれの類似度ごとに異なる。

そして、この類似度をもとに番組全体としての関連性の強さ R_f を算出することとする。すなわち、(1)式で表すように、個々の類似度の関数として算出する。

$$R_f = f(S_{title}, S_{person}, S_{keyword}, S_{genre}, S_{time}, S_{broadcaster}) \quad (1)$$

2.2 主観量を定式化する従来手法の問題点

上記の (1) 式により算出される値は、視聴者が主観的に感じる番組間の関連性の強さとよく一致している必要がある。そのためには、視聴者が主観的に感じる関連性の強さを R_s とすれば、 R_s とよく一致する R_f を算出するように、予測式を定めればよい。

このアプローチにより、関連性の強さを予測する式の精度を高めるためには、主観評価により、視聴者が感じる主観的な番組間の関連性の強さ (R_s) を正確に測定することと、 R_s の値とよく一致する R_f を算出できるように予測式の形式およびパラメータを決定することが重要となる。

ところで、番組の組合せのパターンは、極めて多種多様である。したがって、汎用的に通用する式を構築するためには可能な限り多くの番組の組合せ (例えば数百ペア) について R_s を測定することが望ましいが、数百ペアの番組の組合せについて、関連性の強さを問う設問に回答させる主観評価の実施は回答者の負担が大きくなってしまったため現実的でない。

少ない設問数で効率よく主観量を定式化するために用いられる既存手法としては、コンジョイント分析がある。コンジョイント分析では、直交表に基づき主観評価を行う設問を作成することで、設問数を削減することができる。

我々も、24パターンの番組の組合せを設問として用いたコンジョイント分析を元に、

$$R_f = a_{title} S_{title} + a_{person} S_{person} + a_{keyword} S_{keyword} + a_{genre} S_{genre} + a_{time} S_{time} + a_{broadcaster} S_{broadcaster} \quad (2)$$

という形式の予測式を使用することを試みた。しかし、 R_s の予測に対して十分な精度を得ることができなかった。その原因を検討したところ、(2)式におけるパラメータが不正確であることと、(2)式は単純な線形式であるために多様な番組の組合せにおける関連性の感じ方を再現するには表現能力が不足していること、が主な原因であることが分かった [2]。

一般に、少ないサンプル数から決定したパラメータは精度が低く、パラメータ数が増えるに従い必要となるサンプル数も多くなる。コンジョイント分析に基づいて作成した (2) 式は、24サンプルのみから6パラメータ (a^*) を求めている。これを重回帰式における標準回帰係数の信頼区間を求める式に当てはめると、標準回帰係数の95パーセント信頼区間は0.65~0.71とかなり大きい。仮にこの値を0.10以下に収めようとする、少なくとも196サンプル以上が必要となる [6]。

また、単純な線形式よりも表現能力の高い式 (たとえば非線形で変数間の交互作用を考慮した式) とするには、より多くパラメータを決定することが必要となる。そのためには、さらに多くのサンプル数 (設問) が必要となる。

さらに、決定すべきパラメータの数が多くなった場合、組合せ数が爆発してしまうため、効率的にパラメータを決定することも必要となる。

2.3 本論文での問題解決のアプローチ

前述のとおり、回答者に番組間の関連性の強さを問う設問の数 (サンプル数) を多くすることが望ましい。しかし、たとえば設問の数を数百問に増やすことは回答者の負担が大きくなってしまったため、現実的でない。

従来、感性工学分野においては、コンジョイント分析以外にも、ラフ集合理論 [7] を用いる、あるいはニューラルネットワークの汎化能力を利用して粗なデータの補間を行う [8] など、いかに少ないサンプル数から有用な情報を抽出するかという点に主眼が置かれたアプローチは多い。一方で、いかに多くのサンプル数を得るかという点に関しては、ほとんど考慮されてこなかった。本論文では主観評価において結果の信頼性を確保しつつ、いかに多くのサンプル数を得るかという問題を解決するために、確率モデルを用いた主観評価結果の解析法の提案を行う。

まず、多くのサンプル数を得るために、大量にある設問をいくつかのグループに分割し、また、回答者もいくつかのグループに分割し、分担して回答させることを考えた。例えば、全部で設問が100問あったとしても、設問を10グループに分割し、それぞれを別の回答者グループに回答させれば、1人あたりは10問に回答するだけでよい。

しかし、回答者によって回答の尺度の使い方は異なる。ま

た、回答させる設問の分布や組合せによっても、回答結果にバイアスや歪みが生じることが知られている [9]。したがって、異なる回答者に異なる設問に回答させた結果は同一の尺度上で扱うことができない。

また、回答者に設問を分担して回答させると、それぞれの設問に対して回答する回答者数が少なくなってしまう。たとえば、回答者を100人確保できたとして、10グループに分割して別々の設問に回答させると、各設問に対しては10人からの回答しか得ることができない。そして、少ない回答者からの回答結果は信頼性が低くなってしまう。

そこで我々は、この問題を解決するために各回答者の回答傾向を確率的にモデル化したうえで回答結果を解析することにより、異なる回答者が異なる設問群に回答した結果を同一尺度上で比較可能かつ、信頼性の高い結果を得ることのできる手法を開発した。これにより、従来困難であった大量の主観評価データを取得することを可能とした。

また、関連性の強さを予測する式として、(2) 式のような線形式ではなく、非線形で変数間の交互作用を考慮した (3) 式を用いることにした。(2) 式との違いは、それぞれの類似度に応じた乗数 (β) が乗っていることと、各類似度間の交互作用を表す項があることである。

$$\begin{aligned}
R_j = & \alpha_{\text{title}} S_{\text{title}}^{\beta_{\text{title}}} + \alpha_{\text{person}} S_{\text{person}}^{\beta_{\text{person}}} + \alpha_{\text{keyword}} S_{\text{keyword}}^{\beta_{\text{keyword}}} \\
& + \alpha_{\text{genre}} S_{\text{genre}}^{\beta_{\text{genre}}} + \alpha_{\text{time}} S_{\text{time}}^{\beta_{\text{time}}} + \alpha_{\text{broadcaster}} S_{\text{broadcaster}}^{\beta_{\text{broadcaster}}} \\
& + \alpha_{\text{title} \times \text{person}} S_{\text{title}}^{\beta_{\text{title}}} S_{\text{person}}^{\beta_{\text{person}}} + \alpha_{\text{title} \times \text{keyword}} S_{\text{title}}^{\beta_{\text{title}}} S_{\text{keyword}}^{\beta_{\text{keyword}}} \\
& + \quad \dots \quad + \alpha_{\text{time} \times \text{broadcaster}} S_{\text{time}}^{\beta_{\text{time}}} S_{\text{broadcaster}}^{\beta_{\text{broadcaster}}}
\end{aligned} \quad (3)$$

しかし、(3)式では決定すべきパラメータ数が多くなってしまう。このパラメータを決定するため、重回帰分析に加えて、遺伝的アルゴリズムと実験計画法を組み合わせて用いることにより、効率よく(3)式の最適化を行う手法を開発し、(3)式の最適化を行った。

以下，第3章では主観評価の実施および解析方法について，第4章では式の最適化手法について述べる．

3. 番組間の関連性の強さについての主観評価

番組間の関連性の強さに関する設問は、図2に示すように、2つの番組のEPG情報のペアを呈示し、その2つの番組が関連していると感じる程度を回答してもらう。

EPG情報の番組内容の文章は、最大で数百文字に達するものもある。そのため、2つの番組の内容を把握したうえで関連していると感じる程度を回答する作業は、回答者にとっての負担が特に大きいといえる。無理に多数の設問に回答させようとすれば、回答の信頼性が損なわれることも考えられ、このことから、回答者一人あたりの設問数は可能な限り削減することが望ましい。

第2章でも述べたように、大量の設問を分担して回答させることにより、回答者一人あたりの設問数を削減するためには、異なる回答者が異なる設問群に回答した結果を同一尺度

[illegible]

上記2つの番組が、どの程度関連していると感じますか？最もあてはまる程度をお選びください

非常に関連している ← ○ ○ ○ ○ ○ ○ ○ → 全く関連していない

図2 設問のイメージ

上で比較可能かつ、信頼性の高い結果を得られるようにする
施策が必要である。

3.1 主観評価値を同一尺度で扱うための従来手法

異なる実験における回答結果を、同一の尺度で扱うようにするための手法として、たとえば、心理物理学の分野ではマスター尺度法 [10] という手法が用いられる。マスター尺度法では、規準となる評価値が定まっている対象を回答者に提示し、規準となる評価値と各回答者の回答値のずれを補正する。これにより、尺度の使い方が異なる回答者の回答を同一尺度で扱うことができる。ただし、この手法を適用するためには、各回答者の回答値の補正量を正確に求める必要がある。心理物理学で行われるのは、音の大きさなど比較的単純な判断であるため、一人当たり数百回の評定を行わせることも一般的である。そのため、1つの評価対象に対して、独立に複数回の回答を行わせることで各回答者の回答値を信頼性の高いものとし、補正量も正確に求めることができる。

一方、今回実施しようとする評価は前述の通り数百回の回答を行うことは困難である。したがって、同一の設問に対して十分多い回数回答させることは困難であり、各回答者個々の回答結果の信頼性は決して高くない。その結果、各回答者の回答に対する補正量を正確に求めることができず、マスター尺度法のような手法は適用することができない。

一般的に考えられる対策としては、十分多い回答者数の回答結果の平均値を求めることにより、信頼性の高い結果を得ることである。しかし、前述の通り回答者をグループに分割しているため、各設問について十分多い回答者からの回答を得ることも困難である。

3.2 提案する主観評価についての確率モデル

上述の問題を解決するためには、少ない回答者の信頼性の低い回答結果に対して、尺度の補正を行い、かつ、信頼性の高い結果を得ることのできる施策が必要となる。我々は各回答者の回答傾向を確率的にモデル化することにより、異なる回答者の回答結果を同一尺度上で扱うことができ、かつ信頼

性の高いデータを得られる手法を開発した。以下では、その詳細を説明する。

まず、従来行われる多数の回答者の回答結果の平均値を用いることにより、結果の信頼性が上がるということはどういうことかを説明する。設問 i に対する回答者 j の回答値が x_{ij} ($i=1, 2, \dots, N, j=1, 2, \dots, M$)であったとする。 x_{ij} が平均値 μ_i 、標準偏差 σ_i の正規分布に従うと仮定すれば、 x_{ij} からの真の平均値 μ_i の最尤推定量は、解析的に求めることができ、 x_{ij} の標本平均 $\bar{x}_i$ となる。

$$\hat{\mu}_i = \bar{x}_i = \frac{1}{M} \sum_{j=1}^M x_{ij} \quad (4)$$

すなわち、標本平均を求めることが、真の平均値を推定するための最善の方法である。このとき、 $\hat{\mu}_i$ は、平均 μ_i 、分散 σ_i^2/M の正規分布に従う。すなわち、 M が大きいほど標本平均は真の平均に近い値となる確率が高くなる。逆に言えば、標本平均値と真の平均値の誤差を小さくするには M （主観評価の場合は回答者数）を大きくするしか方法がない。

しかし、これは、あくまで個々の x_{ij} が同じ程度 μ_i からばらついていることを期待した場合における話である。仮に、 $x_{ij}=\mu_i$ である x_{ij} があった場合、十分多い M を用意するまでもなく、上記の $x_{ij}=\mu_i$ である x_{ij} の1サンプルのみから真の平均値を知ることができる[11]。主観評価に置き換えて考えてみれば、真の平均値と完全に一致する回答を行う回答者が1人いれば、その回答者の回答結果のみから真の平均値を知ることができる。

もちろん、真の平均値と完全に一致する回答を行う回答者は存在しないが、どの程度真の平均値に近い回答を行うかは回答者ごとに異なる。したがって、各回答者がどの程度真の平均値に近い回答を行うかを考慮すれば、単純に平均値を求める場合と比較して、より信頼性の高い（真の平均値に近い）結果が得られると考えた。

そこで、各回答者の回答値がどの程度平均に近いかを確率的に表すことを考える。設問 i に対する回答者 j の回答値を x_{ij} とおく。また、設問 i に対する真の平均値を μ_i とおき、回答値は間隔尺度以上の尺度であると仮定する。その上で、回答値 x_{ij} は、真の平均値からの線形変換に誤差項が加わったものと仮定すれば、たとえば以下のようにモデル化できる。

$$x_{ij} = a_j \mu_i + b_j + \varepsilon, \quad \varepsilon \sim N(0, \sigma_j^2) \quad (5)$$

すると、真の平均値が μ_i であったとき、回答者 i の回答値が x_{ij} となる確率は、以下の(6)式により表すことができる。

$$P(x_{ij} | \mu_i) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left[-\frac{(a_j \mu_i + b_j - x_{ij})^2}{2\sigma_j^2}\right] \quad (6)$$

このとき、 μ_i の最尤推定量は、以下の(7)式に(6)式を代入することで求めることができる。

$$\hat{\mu}_i = \arg \max_{\mu_i} P(\mu_i) \prod_{j=1}^M P(x_{ij} | \mu_i) \quad (7)$$

この解は、数値計算により求めることができる。実用上は、対数を取ったものを最大化の方が容易であり、また $P(\mu_i)$ に関して強い仮定がなければ以下の式で求めてもよい。

$$\hat{\mu}_i = \arg \max_{\mu_i} \sum_{j=1}^M \log P(x_{ij} | \mu_i) \quad (8)$$

ところで、(8)式による推定値は、 $a_j=1, b_j=0, \sigma_j=\sigma$ （全ての回答者の回答傾向が全く同一である）と仮定すれば、(4)式による推定値と一致する。すなわち、この手法の特徴（(4)式との違い）は回答者ごとの回答傾向が同一では無い、と仮定することである。

たとえば、以下の2人の回答者1 ($a_1=1, b_1=0, \sigma_1=1$)と回答者2 ($a_2=1, b_2=0, \sigma_2=3$)がいたとする。ある設問に対する回答者1の回答値が「2」、回答者2の回答値が「5」だったとする。(4)式による μ の推定値は、 $(2+5)/2=3.5$ であるが、(6)式を(7)式に代入して求める μ の推定値は2.3となる(図3)。

σ_j の小さな回答者の回答値（の線形変換）は、 σ_j の値が大きな回答者の回答（の線形変換）と比較して、真の平均値に近い確率が高い。上記の例では回答者1の回答値の方が回答者2の回答値に比べて信頼性が高いことを考慮できるため、確率モデルを用いることにより、単純な回答平均値よりも信頼性の高い結果を得ることができる（推定値は回答者1の回答値に近くなる）。

ただし、 σ_j の小ささは、あくまで真の平均値を求めるという目的において、各回答者の回答がどの程度役に立つかを表すものである。 σ_j の小さな回答者の回答が正しく、 σ_j の大きな回答者の回答が間違っている、というわけではない点には注意が必要である。

さて、上記のような確率モデルを用いるためには、各回答者の特性を表すパラメータを推定する必要がある。これは、あらかじめ真の平均値が分かっている設問に対する回答値から求めることができる。真の平均値 μ_i が既知である設問($i=1, 2, \dots, N$)に対する回答者 j の回答値が x_{ij} であったとき、 a_j, b_j, σ_j は、一般的な回帰分析と同様に、

$$E_j^2 = \sum_{i=1}^N \{\mu_i - (a_j x_{ij} + b_j)\}^2 \quad (9)$$

図3 確率モデルの振る舞い

が最小となるように a_j , b_j を求めればよく, σ_j は、そのときの E_j の値から、

$$\sigma_j = \sqrt{\frac{E_j^2}{N-2}} \quad (10)$$

として求めることができる。

なお、信頼性が一定でない複数の結果に対して確率モデルを用いて対処するという考え方は、例えば学力テストの採点等で用いられる項目応答理論 [12] などでも用いられている。本論文における回答者ごとの回答傾向を確率的にモデル化し、各設問に対する結果を確率的に推定するという手法は、項目応答理論において、設問を回答者、回答者を設問におきかえたもの、として解釈することもできる [13]。

3.3 提案手法の有効性の検証

以上の手法の有効性を検証するため、以下の実験を行った。まず、番組同士の関連性の強さを7段階で回答する設問を合計24問、20～64歳の男女1,000名（年齢・性別による層化二段階抽出）に回答させた。設問の呈示と回答の入力はWebアンケート形式で実施した。

回答者数が十分多いため、この回答結果の平均値は真の平均値とみなすことができる。そこで、この1,000名の回答結果の平均値を真の平均値とする。なお、この際1,000名の回答者は全て同一の設問24問に回答しているため、設問の組み合わせに起因する回答値のバイアス [9] が生じていない。設問の組み合わせに起因する回答値のバイアスに対する有効性も検討するため、各回答者について [0, 1] の一様乱数を生成し、回答値に付加した。続いて、24問の設問を12問ずつの2つに分割する（設問グループAと設問グループB）。そして、設問グループAについての真の平均値と、設問グループAの設問に対する各回答者の回答値から、各回答者の特性を表すパラメータ (a_j , b_j , σ_j) を求める。

そのパラメータを用いて、設問群Bに対する真の平均値を(8)式により推定し、真の平均値との誤差平方和を求める。比較対象として、(4)式による確率モデルを用いない回答平均値も求めた。ブートストラップ抽出により1,000人の回答者から任意の人数を抽出し、設問グループBに対する平均値の推定を行う手順を1,000回繰り返し、誤差平方和の95パーセンタイル値を求めた。結果を図4に示す。

図4に示す通り、(4)式による推定値（回答平均値）も(8)式による推定値（確率推定値）も、回答者数が多くなるにつれて真の平均値からの誤差は小さくなっている。しかし、(8)式による推定値の方が、(4)式による推定値に比べて、全域で誤差は小さい。

この差は、① 回答者数が十分多いときに収束する値の差、② 回答者数を大きくしていくときの収束する速度の差、の2つに分けられる。①については、異なる回答者が、異なる設問群に回答した場合に生じる回答値のバイアスを、(8)式を用いることにより軽減できることを示している。また、②については、回答者数が少ないときであっても、(8)式を

用いることにより(4)式に比べて信頼性の高い結果を得られることを示している。たとえば、約30人の回答者のデータから約1,000人相当の精度が得られている。

このことから、各回答者の回答傾向を確率的にモデル化することにより、大量の設問を回答者グループに分担して回答させても、同一尺度上で扱うことのできる信頼性の高いデータを得られることが分かった。

3.4 大量の設問に対する主観評価実験の実施

上記の手法を用いて、実際に大量の設問に対する主観評価を実施した。設問の割り当ては、図5に示す通りであり、回答者グループは各30名、設問グループは設問グループ1が8問、設問グループ2から21は各20問である。それぞれの回答者グループは、図5の色の付いている設問に回答する。

回答者は20歳から64歳までの男女合計600名（年齢・性別による層化二段階抽出）であり、設問の呈示および回答の入力はWebアンケート形式で実施した。

ここで、設問グループ1に対しては、全ての回答者(600名)が回答する。回答者数が十分多いため、設問グループ1に対する標本平均値は真の平均値とみなすことができる。この値を元に各回答者の特性を表すパラメータを求め、そのパラメータを用いて、設問グループ2から21についての真の平均値を(8)式を用いて推定した。

図4 真の平均値の推定誤差

	設問 グループ1 (8問)	設問 グループ2 (20問)	設問 グループ3 (20問)	...	設問 グループ21 (20問)
回答者 グループ1 (30名)					
回答者 グループ2 (30名)					
⋮	⋮			⋮	
回答者 グループ20 (30名)					

図5 設問の割り当て

これにより、回答者1人あたりは28問に回答するだけで、合計では400問もの設問に対して、主観的に感じる番組間の関連性の強さのデータを収集することができた。なお、28問の設問は、20~30分程度あれば回答できる分量であり、回答者に対する負担は十分許容できる範囲であったと考えられる。

4. 関連性の強さの予測式の最適化

第3章では、放送番組同士の関連性の強さに関するデータ(R_s)を取得した。あとは、この R_s に近い値を算出するように式を最適化すればよい。番組の組*i*についての R_s および R_f をそれぞれ $R_s(i)$ ($i=1, \dots, N$)、予測式により算出される値を $R_f(i)$ とすれば、

$$\sum_{i=1}^N \{R_f(i) - R_s(i)\}^2 \rightarrow \min \quad (11)$$

となるような予測式とすればよい。第2章でも述べた通り、ここでは(3)式により R_f を算出する。

4.1 予測式の最適化における問題

(3)式において決定すべきパラメータは、各項にかかる係数 α と、各類似度のべき乗 β である。また、使用する項を多くすれば式を最適化するためのデータに対する当てはまりは良くなるが、式の汎化能力は低下してしまう。そのため、一般性の高い式とするためには、変数選択を行い使用する項を絞り込む必要がある。

いま、(3)式で使用可能な項は21項あるため、使用する項の組合せは2,097,152 (2^{21})通りである。また β に関しては、取りうる値をそれぞれ0.5, 1.0, 2.0の3種類に絞っても729 (3^6)通りある。一方、使用する項と β が定まれば、重み α は重回帰分析により一意に求めることができる。すると、最適な式の候補は、 $2,097,152 \times 729 = 15,288,223,808$ (約150億)通りあり、総当たりを行うことは計算時間の上で現実的でなく、効率的に式の最適化を行うことが必要となる。そこで、重回帰分析による係数の決定に加えて、遺伝的アルゴリズムと実験計画法を組み合わせることにより、最適化を実施した。

4.2 遺伝的アルゴリズムによる変数選択

効率的に重回帰式の変数選択を行う手法には、たとえばステップワイズ法がある。しかし、ステップワイズ法は、局所的な最適解に収束してしまう可能性がある。そこで、変数選択には遺伝的アルゴリズム(GA)を用いた[14]。

GAにより変数選択を行う場合、遺伝型の各ビットを各変数に対応させ、1であればその変数を採用する、0であればその変数を採用しない、とすればよい。その上で遺伝型により採用すると決定された変数を用いて重回帰分析を行い、その結果に対してAIC(赤池の情報量規準)を算出する。AICの値は、予測式の残差の分布に正規分布を仮定すれば、

以下の式で与えられる。

$$AIC = N \log \bar{E} + 2P \quad (12)$$

ここで、 $\bar{E}$ は予測残差の平均、 P は式に含まれるパラメータの数(採用する項の数)、 N は、式の当てはめを行うデータの数であり、いまの場合、合計の設問数(400)である。

AICの値は、式の汎化能力を表す指標であり、汎化能力が高いほどその値は小さくなる。ここでは、野村ら[15]の方法にならない、ある世代におけるAICの最悪値と各個体のAICの差を、その個体の適応度としてGAを適用した。

まず、この手法が交互作用を含む重回帰式における変数選択に対して有効であるかの検討を行った。予め全探索を行い、最適変数の組み合わせが既知である問題に対して、上記のGAによる最適化を行い、最適解に到達するまでの世代数を測定した。各世代での個体数は128、突然変異率は0.02、とし交叉は1点交叉とした。測定は、第1世代の遺伝型の初期値を変えて5回行った。その結果を表1に示す。

表1に示すとおり、平均で178世代の探索で最適解に到達できた。これは、全探索に比べて約1.08%のみを探索するだけで大域的最適解が求まっていることになる。GAを用いた変数選択が、交互作用を含む重回帰分析に対して有効であり、効率的に大域的な最適解を求めることができるといえる。

以上により、GAを用いて、効率的に変数選択を行うことが可能であることが分かったので、あとは各変数に対するべき乗数 β を求めればよい。この探索範囲は729(3^6)であるが、これは実験計画法により絞り込むこととした。

まずL18直交表を用いて、各 β の水準を決定し、それぞれの水準の組合せにおいて、前述のGAを用いた変数選択を実施する。そうして得られた各最適解のAICに対して分散分析を行い、主効果のあった β を最適な水準に固定するそして残りの β に対して再び直交表により水準を割り当てる。この手順を繰り返すことで、最適な各 β の水準を求め、予測式を最適化していく。

予測式の最適化を行う手順を表すフローチャートを図6に示す。以上の手順により最適化を行ったことにより、全15,288,223,808通りの非線形予測式の候補のうち、2,304,000通り(全探索に対する割合は0.151%)の候補に対する探索を行うだけで、効率的に最適化を行うことができた。

表1 遺伝的アルゴリズムの有効性検討結果

試行	最適解に到達するまでの世代数	探索した解候補の数	全探索に対する割合
1	49	6272	0.29%
2	43	5504	0.26%
3	434	55552	2.65%
4	207	26496	1.26%
5	155	19840	0.94%
平均	178	22733	1.08%

図6 非線形予測式の最適化手順

5. 予測式精度の評価

第3章で説明した主観評価法により取得した主観データに対し、第4章で説明した予測式最適化法を適用することで、番組間の関連性の強さを予測する式を求めた。この式および式の最適化手法の有効性を検証するため、以下の実験を行った。

5.1 関連性の予測式精度の指標

ローミングナビでは、図1のように注目番組に関連する番組を配置する。このとき、関連性の強い番組は、関連性の弱い番組よりも上位に表示される必要がある。つまり、関連性の強さの順位関係が重要となる。そこで、主観的な関連性の強さの回答値と、予測式により算出した関連性の強さの値の順位相関係数をもとにして、妥当性を検証することにした。また、本論文で提案した手法の有効性を示すため、従来のコンジョイント分析に基づいて導出した線形の予測式との性能の比較も実施した。

20歳から64歳までの男女600名（年齢・性別による層化二段階抽出）に、図2に示すような番組間の関連性の強さを7段階で回答する設問24問に回答してもらった。設問は、実際の地上波デジタル放送およびBS/CS放送1週間分のEPG情報から、ジャンルに偏りがないように選択した。

設問の呈示および回答の入力は、Webアンケートにより実施した。なお、式の汎用性の評価であることを保障するために、式の最適化のために実施した主観評価に用いたものとは異なる番組のペアに対して、異なる回答者が回答している。また、回答の一貫性を確認するために、2週間の間隔をあけて、同一の回答者に同一の設問に回答してもらった。

各回答者の回答値（1回目）と予測式による算出結果の値のスピアマンの順位相関係数を ρ_{sf} 、1回目の回答値と2回目の回答値のスピアマンの順位相関係数を ρ_{ss} とし、各回答者の主観的な番組間の関連性の強さと予測式により算出された関連性の強さの一致度合いの指標を $\rho_{adj} = \rho_{sf} / \rho_{ss}$ とした。 ρ_{adj} の値は、一致度合いが高いほど大きくなり、予測式による値と回答した値が無相関であれば0、各回答者が自分自身で回答した値と同程度に一致していれば1となる。 ρ_{sf} の値をそのまま指標として用いない理由は、回答が一貫していない回答者の結果を補正するためである。

5.2 予測精度の評価結果

予測式として、400問の設問の結果に基づき作成した非線形予測式を用いた場合と、24問のコンジョイント分析に基づいて作成した線形予測式を用いた場合の ρ_{adj} のヒストグラムを、それぞれ図7に示す。ただし、 ρ_{ss} が有意な正相関でない回答者は除外している。

図7を見比べて明らかのように、非線形予測式の方が、線形予測式の場合に比べて ρ_{adj} の分布が大きい方に寄っている。線形予測式では平均値0.716（標準偏差0.177）、であり、一方、非線形予測式では、平均値0.864（標準偏差0.171）であった。

図7 線形予測式と非線形予測式の性能比較

両者に有意差があるか検定を行ったところ、非線形予測式による ρ_{adj} の値の方が、線形予測式による ρ_{adj} の値と比較して、統計的に有意に高かった（ウィルコクソンの符号付き順位和検定、 $p<0.01$ ）。すなわち、最適化を行った非線形予測式により、より多くの回答者（視聴者）にとっての主観的に感じる番組間の関連性の強さを、より精度よく予測できることが示された。

ところで、(11)式では、主観的に感じる番組ペア i についての関連性の強さを $R_s(i)$ としたが、Inter-Individual Variationを考慮すると、視聴者 j ($j=1, 2, \dots, M$) についての関連性の強さ $R_s(i, j)$ と表すことができる（ここでは、尺度の使い方は統一されているとする）。上記の結果のように、より多くの視聴者にとっての主観的に感じる番組間の関連性の強さをより精度よく予測できたということは、それぞれの番組のペア i に関して以下の(13)式の値はより小さくなっていると考えられる。

$$E_i^2 = \sum_{j=1}^M \{R_f(i) - R_s(i, j)\}^2 \quad (13)$$

(13)式を最小とする $R_f(i)$ を求めるために、(13)式を $R_f(i)$ について偏微分し、その値を0とおくと、

$$R_f(i) = \frac{1}{M} \sum_{j=1}^M R_s(i, j) \quad (14)$$

が得られる。つまり、より多くの視聴者にとっての主観的に感じる番組間の関連性の強さを、より精度よく予測するためには、 $R_f(i)$ の値を $R_s(i, j)$ の全視聴者 j についての平均値と一致させればよいことが分かる。非線形式による予測が、線形式による予測よりも性能が良かったということは、非線形式による予測値がより正確な平均値に近い値を算出できていることを示している。

以上のことから、本論文で示した手法が、主観量に関してより正確な平均値を求めること、および、その定式化を行うために有効であったとすることができる。

5.3 予測式最適化の「ローミングナビ」に対する効果

最後に、本論文において行った予測式最適化の結果が「ローミングナビ」というアプリケーションに対して与える効果およびローミングナビにおける関連性の強さの予測方法の妥当性についても述べておく。

ローミングナビを搭載している「CELL REGZA」の機能である地上波デジタル放送8チャンネル26時間分の録画機能[1]により録画した全番組を対象として、番組検索を行うことを想定したシミュレーションの結果では、 ρ_{adj} の値が0.716（線形予測式の平均値）から0.864（非線形予測式の平均値）に向上した場合、視聴者が注目番組に対して最も関連性が高いと感じる番組が図1の画面内に表示される確率（最初に表示される画面から画面をスクロールせずにその番組にアクセスできる確率）は38.3%から58.0%に向上する。また、画面を上下左右の各方向にスクロールすることによ

り、最も関連性が高いと感じる番組にアクセスできる確率は89.2%から98.1%に向上する。このように、予測式の最適化を行ったことにより、ローミングナビにおいて視聴者にとってより関連性の高いと感じる番組に対するアクセス機会の増大を実現することができた。

また、ローミングナビにおいては、(1)式のようにEPG情報中の限られた情報のみを用いて番組間の関連性の強さを予測している。しかし、前述のとおり最適化した予測式を用いることにより、最も関連性が高いと感じる番組に98.1%という高い確率でアクセスすることが可能となっていることから、(1)式による関連性の強さの予測は、実用上十分妥当な方法であるといえる。

6. おわりに

本論文では、番組間の関連性の強さを題材として、主観評価データを取得する手法および、その定式化手法について述べた。特に、各回答者の回答傾向を確率モデルにより表すことにより、大量の設問を異なる回答者に分担させても同一の尺度上で扱うことのできる信頼性の高いデータを取得できる点に特徴がある。

2.3節でも述べたが、これまで感性工学の分野ではいかに少ないサンプル数から有用な情報を抽出するか、に主眼が置かれていた。本論文は、それらとは異なり、いかに多くのサンプル数を集めるか、という点に着目したアプローチである。

しかし、従来の少ないサンプルから有用な情報を抽出する、というアプローチと、本論文における多くのサンプル数を集める、というアプローチは相反するものではない。たとえば、本論文で示した手法で多くのサンプル数を集めた上で、ラフ集合理論などを用いた解析を行えば、少ないサンプル数に対して解析を行った場合に比べて、より有用で信頼性の高い情報を抽出することができると考えられる。したがって、本論文で示した手法は、従来手法の能力を拡張させるための手法にもなり得る。

ところで、本論文では正確な平均値を推定することを目的として、回答者の回答傾向を表す確率モデルを求めた。しかし、たとえばデータの分布が多峰性となる場合など、データの性質によっては平均値を求めること自体に意味のないこともある。したがって、本手法の適用に限った話ではないが平均値を求めること自体の正当性は確認する必要がある。

一方で、各回答者の回答がどの程度平均的であるかを確率モデルで表したのと同様に、各回答者の回答がどの程度特殊であるかを確率モデルで表すことも可能である（(9)式および(10)式で、 a_j , b_j , σ_j を求める際に用いる μ_i を、ある特定のユーザー層の嗜好を表す量に置き換えれば良い）。すなわち、従来、回答者のプロフィール（年齢・性別など）に従って回答者の分類をして行っていた回答の集計を、回答結果そのものの傾向に従って集計を行うことが可能となる。このよ

うに、本論文で報告した手法は、主観量からの情報抽出や定式化のために広く適用可能な手法である。

なお、本論文では、回答傾向に関して、単純な線形のモデルを仮定したが、異なる形式の確率モデル（たとえば項目応答理論における多段階反応モデルに相当するモデル）なども有効であると考えられる。また、設問数や回答者数の決定法の指針の整備や、より多くの例に対して本手法の有効性を検証することも必要である。

(注) 本文に掲載の商品・機能等の名称は、商標として使用している場合があります。

参 考 文 献

- [1] CELL REGZA 録画機能紹介ページ,
<http://www.toshiba.co.jp/regza/lineup/55x1/recording.html>
- [2] 三原功雄, 関根真弘, 樋口靖和, 高倉潤也, 鈴木優, 山内康晋: テレビ番組間の関連性に基づくコンテンツ検索システム「ローミングナビ」の開発, 情報処理学会研究報告, Vol.2010-HCI-137, No.8, 2010.
- [3] Linden, G., Smith B., York, J.: Amazon.com Recommendations: Item-to-Item Collaborative Filtering, IEEE Internet Computing, Vol.70, No.1, pp.76-80, 2003.
- [4] 市村由美, 齋藤佳美, 酒井哲也, 國分智晴, 小山誠: 質問応答と、日本語固有表現抽出および固有表現体系の関係についての考察, 情報処理学会研究報告, Vol.2004-NL-161, No.3, 2004.
- [5] 山崎智弘, 真鍋俊彦, 川村隆浩: 話題抽出エージェントを用いた番組検索システムの実装, コンピュータソフトウェア, Vol.25, No.4, pp.41-51, 2008.
- [6] Kelly, K. and Maxwell, S.E.: Sample Size for Multiple Regression: Obtaining Regression Coefficients That Are Accurate, Not Simply Significant, Psychological Methods, Vol.8, No.3, pp.305-321, 2003.
- [7] 森典彦: ラフ集合と感性工学, 日本ファジィ学会誌, Vol.13, No.6, pp.600-607, 2001.
- [8] 堀田創, 萩原将文: 文字デザインを例とした感性情報からのファジールール自動抽出法, 電子情報通信学会論文誌 D, Vol.89, No.10, pp.2361-2369, 2006.
- [9] Laming, D.: The Measurement of Sensation, Oxford University Press, 1997.
- [10] Burgland, M. B.: Quality Assurance in Environmental Psychophysics, In Ratio Scaling of Psychological Magnitude: In Honor of the Memory of S. S. Stevens, pp.142-160, Lawrence Erlbaum Associates, 1991.
- [11] 藤子・F・不二雄: 並平家の一泊, SFファンタジア 風刺編, 学習研究社, 1977 (藤子・F・不二雄 異色短編集「パラレル同窓会」収録, 小学館文庫, 1995).
- [12] 豊田秀樹: 項目反応理論 入門編 - テストと測定の科学 -, 朝倉書店, 2002.
- [13] 高倉潤也, 関根真弘, 三原功雄, 山内康晋: 回答者ごとの回答傾向を考慮した主観評価評定値の解析 - 項目応答理論の枠組みを用いた回答傾向のモデル化 -, 日本行動計量学会第38回大会抄録集, pp.200-201, 2010.
- [14] 伊庭斉志: 進化論的計算手法, オーム社, 2005.
- [15] 野村博義, 若見昇: 遺伝アルゴリズムによるファジー推論規則の決定法, 電子情報通信学会論文誌 A, Vol.77, No.8, pp.1241-1249, 1994.

高倉 潤也 (正会員)

2008年九州大学大学院芸術工学府芸術工学専攻修士課程修了。同年、(株)東芝に入社。以来、ヒューマンインターフェースや情報視覚化および、その評価手法に関する研究開発に従事。日本生理人類学会、日本行動計量学会、電子情報通信学会会員。

関根 真弘 (非会員)

2002年東北大学大学院情報科学研究科システム情報科学専攻修士課程修了。同年、(株)東芝に入社。現在、コンピュータグラフィックス、画像処理、ヒューマンインターフェースに関する研究開発に従事。映像情報メディア学会会員。

三原 功雄 (非会員)

1997年東京工業大学大学院情報理工学研究科計算工学専攻修士課程修了。同年、(株)東芝に入社。ヒューマンインターフェース、画像解析、入力デバイス、コンピュータグラフィックスに関する研究開発に従事。ACM、電子情報通信学会、人工知能学会会員。

山内 康晋 (非会員)

1994年 東京大学大学院工学系研究科電子工学専攻修士課程修了。同年、(株)東芝に入社。2004-2006年 スタンフォード大学客員研究員。コンピュータグラフィックス、3次元映像処理、ヒューマンインターフェースに関する研究開発に従事。ACM、情報処理学会会員。

樋口 靖和 (非会員)

2005年東京農工大学大学院工学研究科情報工学専攻修士課程修了。同年、(株)東芝に入社。現在、デジタルTV向けヒューマンインターフェースの研究開発に従事。