

③ 実世界のユーザ行動に着目した嗜好抽出・情報推薦

1. AV機器利用者に対する 放送コンテンツの推薦

村上知子

(株)東芝 研究開発センター

多チャンネル時代

BS(Broadcast Satellite)やCS(Communication Satellite)を用いた衛星放送の普及と全国的な地上波デジタル放送の開始により、本格的な多チャンネル時代の幕開けを迎えた。すでに2007年時点で、1日あたりに放送される番組数は、地上波放送で約370件、BS放送で約420件、CS放送では約2,000件に達し、専門チャンネルは今後ますます増加の傾向にある。さらに、2011年には、アナログ放送からデジタル放送への完全な移行が宣言されており、放送のデジタル化、多チャンネル化に伴い多様な放送コンテンツが飛躍的に増大することが容易に予想される。

一方、利用者にとって放送コンテンツを知る術も、従来の新聞のテレビ欄に代表される書誌上の番組表から電子番組表(EPG: Electronic Program Guide)へデジタル化が進んだ。インターネット上のiEPG(internet EPG)やインターネット上からAV機器に番組情報をダウンロードして用いるDEPG(Dynamic EPG)などが代表例である。デジタル化が進んだものの、利用機器上の表示スペースに対する制約と、時間と放送局をそれぞれ縦横軸としたインタフェースは旧態依然であるため、大量の放送コンテンツの中から利用者が自分の嗜好にかなうコンテンツを発見するのは容易な作業ではない。

近年、高度情報化社会を特徴付けるキーワードの1つに「情報洪水(information overflow)」が挙げられる。多種多様で大量の情報が存在するため、利用者がそれらを取捨選択する手間暇ばかりがかかり、必要としている情報を獲得できない状況を表したものである。放送環境の整備に伴う放送コンテンツの多様化・大量化は、AV機器(TV、DVDレコーダ)利用者に対する情報洪水の一事例といえるであろう。本稿では、情報洪水対策の1つとして期待が高まっている、AV機器利用者の嗜好に適した放送コンテンツの推薦技術を紹介する。

放送コンテンツの付加情報

放送波を通じてAV機器に送られる放送コンテンツに関する情報には、EPG情報や字幕情報などが挙げられる。それぞれの情報に関して、詳細を説明しよう。

• EPG情報

EPG情報の内容は、ARIB(Association of Radio Industries and Businesses of Japan: 電波産業会)¹⁾で定められた基準に準拠している。EPG情報から取得可能な情報の一部を以下に紹介する。

- 放送形態: 地上アナログ/デジタル, BS/CS放送の識別
- 番組識別: 放送日でユニークな番組ID
- 放送日時: 放送日, 開始時刻, 放送時間
- 放送チャンネル: 放送局のID
- 番組タイトル: 再放送や新番組のラベルも付与
- 出演者: 声優, 解説, 司会, 監督などの分類ラベルも付与
- 番組ジャンル: 定義された大分類・中分類情報が1番組につき最大3個まで付与
- 番組内容: 概要と詳細内容

これらの情報がすべての番組に付与されていることは珍しく、出演者や番組ジャンル、番組内容が欠落した番組が数多く存在する。

• 字幕情報

テレビの音声(アナウンスやナレーション、ドラマの台詞など)を文字にしたり、BGMやドラマの効果音を記号にしたりして画面に表示したものが字幕情報である。番組中の各シーンの字幕内容とそれらの開始、終了時間などがテキスト形式で取得される。字幕情報は番組の放送と同時に利用者に提供される情報であるため、放送後EPG情報に加えて字幕情報も利用者の嗜好の抽出に利用することが可能である。

放送コンテンツの推薦と課題

AV機器上の推薦システムは、放送コンテンツの付加情報を用いて推薦内容を決定する。その際、推薦にはさまざまな形態が存在するが、放送コンテンツの推薦における共通の課題を以下に挙げる。課題の多くは、放送コンテンツの推薦に特化したものでなく、推薦システム全般に共通の課題でもある²⁾。

• 嗜好の個人差への対応

放送コンテンツに対する嗜好は利用者ごとに異なる。利用者個人の嗜好を捉え、それらをもとに嗜好に適した放送コンテンツを予測することが可能な技術が必須である。

- context への適応

放送コンテンツに対する嗜好は、長期的には利用者個人内においても変化すると考えられる。たとえ嗜好が変化しなくても、利用者の生活スタイルや利用者のおかれる状況 (context) は時間とともに変化することが予想される。そのため、利用者の context に応じた放送コンテンツの推薦が望まれている。

- 嗜好情報の獲得

利用者の嗜好に適した放送コンテンツを推薦するためには、利用者の嗜好情報 (好きや嫌い、好みの度合いなど) を獲得する必要がある。利用者が明示的に嗜好情報を入力することも考えられるが、利用者への負荷が大きく、必ずしも毎回利用者が積極的に評価情報を機器へフィードバックするとは考えにくい。そのため、放送コンテンツに対するアクセス履歴情報から、利用者が実際に好む放送コンテンツを暗黙的に高精度に特定する必要がある。

- 推薦システムの評価

AV 機器における推薦機能の価値は、推薦システムが嗜好に適した放送コンテンツを推薦できたかどうかという観点からだけでなく、実際に利用者の視聴行動を支援できたかどうかという観点からも評価されなければならない。従来、推薦システムは、情報検索の研究分野で有名な適合率 (precision) や再現率 (recall) などの指標を用いて評価されてきたが、推薦システムの提供する情報の新規性 (novelty)、意外性 (serendipity)、有用性 (usefulness) などの評価やそのための指標は情報推薦の研究分野における共通の課題であると認識されている。

- 機器利用開始時の精度向上

多くの利用者は AV 機器を購入した直後に推薦機能を利用することになるが、その時点で視聴・録画履歴情報が利用機器に蓄積される可能性は低い。推薦機能の利用開始時、すなわち、視聴・録画履歴情報が少量の場合、一般的に、推薦システムの予測精度は低い傾向にある。しかし、推薦機能の利用開始時にある程度の予測精度を維持しないと、推薦システム自体の信頼性が低下し、永続的な利用が望めない。推薦機能の利用開始時にも予測精度の高い推薦システムが求められている。

既存の放送コンテンツ推薦システム

米国では、地上波放送、衛星放送、ケーブル放送を含め 20,000 チャンネルにも上るテレビ放送番組の視聴環境が整備されているため、視聴者が番組表の中から自力で好みの番組を検索するのは非常に困難であるという状況がある。このような状況において、視聴者の好む放

送コンテンツを好きな時間に視聴することを可能にする Tivo が高い支持を得ている。Tivo はハードディスクを搭載した専用機器と番組推薦や自動録画機能などのサービスが一体化した形で利用者に提供される。Tivo の番組推薦機能は、全米の Tivo 利用者による録画番組に対する評価に基づく協調フィルタリングで実現されている³⁾。リモートコントローラ上のサムズアップボタン (良い評価) とサムズダウンボタン (悪い評価) の押下回数による段階的な評価情報がサービスセンタに収集、蓄積され、サーバ上の推薦システムがそれらを基に各利用者に対する推薦リストを作成し、各専用機器へ推薦リストを送信する仕組みになっている。プライバシー保護の観点から、視聴履歴情報の利用者 ID は暗号化され、サーバ側で処理する際には利用者の特定ができないような仕組みも考慮されている。

欧州では、日本とほぼ同時期に衛星放送やケーブル放送が開始したが、普及は日本よりやや遅れており、現時点では、欧州の中でも英国が他国に先んじて放送のデジタル化が進んでいるといわれている。米国ほどケーブル放送の利用が盛んでないため、推薦のような高度に利用者の嗜好を扱う機能やサービスは今のところ登場していない。現状では、放送コンテンツプロバイダによるハードディスク録画予約サービスが、利用者の嗜好に関係する最新のサービスのようである。利用者は放送コンテンツプロバイダが提供するハードディスク付き専用機器を購入、月極めの料金を支払うことにより、テレビ番組のシリーズ録画などの自動録画機能を利用することができる。欧州では、従来から放送チャンネル数が少なかったことや文字放送が普及していたため、高画質化よりも多チャンネル化、双方向化に関心の高い利用者が多いといわれている。また、英国やドイツでは、2010 年以降を目処にアナログ放送からデジタル放送への完全な移行を目指しているため、今後欧州においても、利用者の嗜好を捉え視聴や録画に活かすサービスの市場は拡大するものと予想される。

一方、国内に目を向けると、現時点では放送コンテンツ量は米国と欧州の中間に位置する。依然多くの人の視聴・録画の対象が地上波放送番組のみであるという実態ではあるが、来たるべき本格的な多チャンネル時代に向けた番組推薦機能も登場している。東芝の HDD&DVD レコーダ RD シリーズ上のおすすめ機能を一例として紹介しよう。図-1 は、東芝の RD 上のおすすめ機能による推薦番組の表示画面例である。おすすめ番組の日時、チャンネル、番組名とともに、おすすめ度合いを王冠マークで表示している。表示画面上のおすすめ番組を 1 つ選択して、番組解説を表示したり、録画を予約したりできるような仕組みになっている。RD 上のおすすめ機能



図-1 推薦番組の表示画面

には、大きく分類して「みんなからのおすすめ」、「あなたへのおすすめ」と「お楽しみ番組自動録画」の3種類が存在する。「みんなからのおすすめ」は、全国のおすすめサービス利用者の録画予約情報に基づく協調フィルタリングによる推薦機能である。「あなたへのおすすめ」は、利用者の録画履歴情報と EPG 情報を用いたナイーブベイズによる推薦システムに基づく機能である。ナイーブベイズでは、ある番組 s_i に f_1, \dots, f_n の特徴が見られる場合に、ある利用者 u_k が s_i を視聴する確率は以下の式で求められる。

$$P(v_{ki} | f_1, \dots, f_n) = \frac{P(f_1, \dots, f_n | v_{ki}) \cdot P(v_{ki})}{P(f_1, \dots, f_n)}$$

式の $P(f_1, \dots, f_n)$ はそれらの特徴が見られる番組の出現頻度、 $P(v_{ki})$ は利用者 u_k の視聴頻度、 $P(f_1, \dots, f_n | v_{ki})$ は利用者 u_k の視聴番組中に存在するそれらの特徴が見られる番組の割合を示している。「お楽しみ番組自動録画」は、「あなたへのおすすめ」と同様に、ナイーブベイズによる推薦システムに基づく機能である。「お楽しみ番組自動録画」を利用すると、推薦システムが利用者の録画・削除・視聴などの詳細な履歴情報と EPG 情報から抽出した出演者、番組タイトル、番組ジャンルなどの番組を特徴付ける要素をもとに利用者の好む番組を予測し、それらの予測番組を機器が自動的に録画する。「みんなからのおすすめ」と「あなたへのおすすめ」は、利用機器をインターネットに接続した環境下での視聴・録画を前提としているのに対して、「お楽しみ番組自動録画」は利用機器をインターネットに接続しなくても利用可能な機能である。

確率モデルに基づくテレビ番組推薦

本章では、AV 機器における推薦に適した確率モデル

に基づく利用者の嗜好のモデル化とそれに基づく推薦システムの研究事例を紹介する。

【ベイジアンネットによる嗜好モデル】

ベイジアンネット⁵⁾とは、確率変数間の定性的な依存関係をグラフ構造で表現し、確率変数間の定量的な関係を条件付き確率で記述する確率モデルである。嗜好モデルのグラフ構造によって嗜好のメカニズムを表現し、条件付き確率値を計算することによって多様な嗜好の程度を表現することが可能である。Breese らが示したように、クラスタリングや k-nearest neighbor などと比較して、ベイジアンネットで表現された嗜好モデルに基づく予測精度は高いことで知られている⁴⁾。

ベイジアンネットによる嗜好のモデル化の最大の特長は、放送コンテンツの情報が刻々と変化する状況において、嗜好モデルの学習および更新が容易という点である。一度嗜好モデルのグラフ構造が同定されれば、嗜好モデルの学習は、観測データに基づき確率変数値の頻度や確率変数値間の条件付き確率値を計算すればよく、嗜好モデル更新時にも、新規観測データから求められるそれらの値を計算し、既存の嗜好モデルに対して追加・置き換えなどの更新をすればよい。また、ベイジアンネットは欠損データに強いという利点もある。現在の EPG 情報には、出演者情報の欠落や番組内容の未定などによる欠損データが多数存在するため、欠損データに頑強な嗜好モデルの学習方法や視聴の予測方法が求められている。

● 嗜好モデルの作成

EPG 情報と視聴履歴情報を用いて、機器の利用者の嗜好モデルを作成する。嗜好モデルは、確率変数と確率変数間の因果関係を条件付き確率で定義することによって決定するベイジアンネットで記述される。ベイジアン

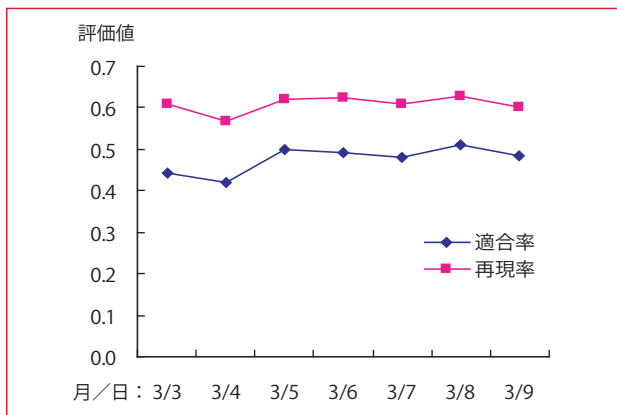


図-2 推薦の精度

ネットの構造は設計者が一意に定めるが、確率変数値の頻度分布に従い確率変数の条件付き確率値を計算する。すなわち、嗜好を説明する要素は利用者共通であるが、各確率変数の条件付き確率値で利用機器ごとに異なる嗜好を表現するモデルを学習する。我が国の放送番組のサイクルは1週間が多く、1週間に1度放送される番組がほとんどであるため、嗜好モデルの学習期間を過去1週間、嗜好モデルの更新は1日1回とした。そのため、過去1週間以前の視聴履歴情報は忘却され学習されない。

● 推薦番組リストの作成

作成された嗜好モデルに基づき、予測日1日の各番組に対する視聴の予測確率値を計算し、それらの値に基づいて番組を降順にソーティングした結果得られる上位の番組を視聴者への推薦番組に決定する。具体的には、各番組の属性値を証拠としてページャネットにおける確率変数を具象化し、各番組に対する視聴の確率値を計算する。嗜好モデルの定期的な更新に伴い、推薦番組リストの作成も1日1回行われ、毎日翌日の推薦番組を決定する。

【番組推薦リスト配信実験】

機器利用者の嗜好モデルをページャネットによって表現し、それに基づいてテレビ放送番組を推薦する実験を行った。実験では、1日あたり約800件の地上波放送とBS放送の番組をDEPGから収集した。曜日の誤差をなくすため、1週間毎日推薦リストを配信し、同時に視聴番組と視聴を逃したが興味のある番組を網羅的に回答してもらうアンケート調査を実施した。図-2は、アンケート調査の回答を正解とした場合の、全被験者46人の平均適合率と平均再現率を1日ごとに示したものである。ここで、利用者にとって好ましい放送番組の集合を F 、推薦されたコンテンツ集合を R とすると、適合率と

再現率は次のように定義される。

$$\text{適合率} = \frac{|F \cap R|}{|R|}$$

$$\text{再現率} = \frac{|F \cap R|}{|F|}$$

「つまらない番組は見たくない」というニーズに対応する指標が適合率、「好きな番組は見逃したくない」というニーズに対応する指標が再現率と考えることができる。両指標は共に視聴生活における基本的なニーズであり重要である。しかし、利用者の視聴時間の伸び率に対して多チャンネル化に伴う放送コンテンツ量の増加率の方がはるかに大きいと予想されるため、限界のある視聴時間内に好きな番組をすべて楽しむことはほぼ不可能になると思われる。したがって、放送コンテンツの推薦においては適合率がより重要な指標であり、今後さらにそのような偏重傾向が強まるであろう。

推薦システムの有用性は、推薦システムが嗜好に適したアイテムを推薦できたかどうかという観点からだけでなく、実際に利用者の意思決定に役立ったかどうかという観点からも評価されなければならない。利用者が日常生活において視聴に割くことのできる時間には限界があるため、推薦システムの有用性は一概に視聴コンテンツ数の変化で評価することはできない。そこで、視聴した番組の質的な変化に注目し、探索視聴適合率を導入した。被験者の視聴番組は、視聴が習慣づいている番組と放送番組を探索して発見し視聴した番組に分類することができる。前者の番組は視聴が固定化している可能性が高いのに対して、後者の番組は流動的である。探索視聴適合率とは、推薦システムによる推薦番組に含まれる探索的に発見し視聴した番組の割合を指す。推薦システムによって利用者に提示された推薦リスト上の番組が、探索的に発見し視聴した番組に多く含まれていれば、推薦システムによる推薦番組が多く視聴されたことになり、推薦システムが視聴番組の決定を支援したという考えに基づいている。図-3に、全被験者の平均探索視聴適合率を、推薦リストを配信した日としなかった日の同じ曜日で比較した結果を示す。すべての曜日で推薦システムの効果が確認され、水、木、金、日曜日に関しては統計的に有意な差も確認できた。

実験ではアンケート調査を実施して利用者の視聴番組情報を収集したが、実際には利用者は主にリモートコントローラを介してAV機器の操作を行う。そのため、視聴する放送コンテンツの選択や切り替えは容易にかつ頻繁に行われ、ザッピング^{☆1}も可能である。利用者が録画した番組は、放送開始前にあらかじめ明示的に録画予約情報がAV機器に送られるため、利用者の好む番組情報として扱うことができる。一方、リアルタイムで放送さ

☆1 テレビを視聴している際にリモートコントローラを操作してチャンネルをしきりに切り替える行為のこと。

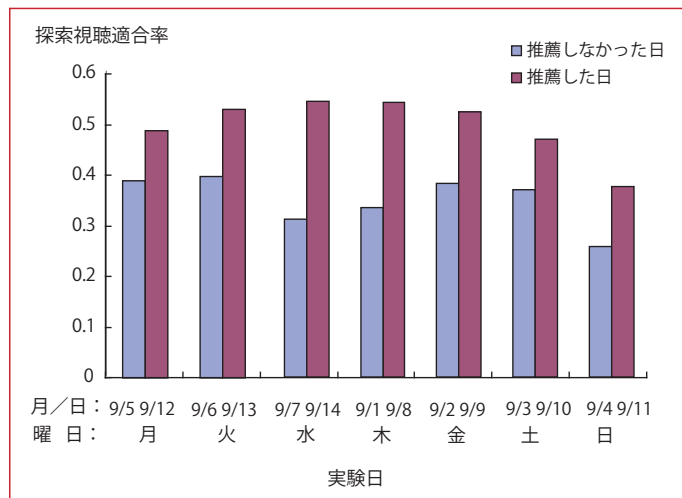


図-3 推薦の精度

れる番組に対しては、利用者による視聴チャンネルの切り替えが多発するため、利用者が実際に視聴し好む番組であるかどうかの判定が困難である。テレビをはじめとするAV機器では、利用者の負荷の観点から、利用者が放送コンテンツに対して明示的に自らの嗜好情報を機器に教えないことを前提としている。このように、機器の操作履歴情報や放送コンテンツ情報を用いて暗黙的に利用者の嗜好情報を特定し、推薦に利用する仕組みは、ユーザビリティの観点から推薦システムの実用化に必須であるといえる。

視聴スタイルに応じた推薦

本章では、利用者の context に適応した放送コンテンツ推薦の一例を紹介しよう。近年、大量のハードディスクを搭載したDVDレコーダなどのように、放送コンテンツを録画し長期間保存できるようなAV機器が急激に普及している。そのようなAV機器を利用すれば、大量の放送コンテンツをハードディスクに録画・保存し、それらのコンテンツの中から選択して視聴するという視聴スタイルが確立されるようになる。そのような場合に、限られた時間ですべてのコンテンツを視聴しきれないという新たな問題が生じてくる。そのため、録画コンテンツの中から、その時点で利用者にはふさわしいコンテンツを提示してくれる機能は有用である。

また、膨大な録画コンテンツの中から必要な録画番組を選択し、さらにその中の見たいシーンだけを選出する作業は、利用者には多大な負荷を与える。そのような将来の課題に対しては、番組シーン推薦により、利用者が録画コンテンツの中から見たいシーンだけを選択して視聴する視聴スタイルの支援を目指した研究がすでに始められている。たとえば、Pic-A-Topic⁶⁾ という番組を話題ごとに分割する技術を用いることによって得られたシーンの中から、利用者の嗜好にかなうシーンを利用者に推

薦する。現在のところ、推薦の実験規模が小さいが、旅行番組を対象にしたシーン推薦実験では、適合率は平均0.40に達している。

推薦技術の今後

EPG情報に始まり画像情報、音声情報、利用者が放送コンテンツに対して付与するメタ情報などの放送コンテンツに付加されるデータは、放送コンテンツ量と同様に増加の一途をたどるのであろう。それらの豊富なマルチメディア情報を利用して放送コンテンツの推薦精度を高める工夫は今後も多く試みられると考える。一般的な情報推薦における課題と同様に、AV機器における推薦機能の価値は、推薦システムが嗜好に適した放送コンテンツを推薦できたかどうかという観点からだけでなく、実際に利用者の視聴行動を支援できたかどうかという観点からも評価されなければならない。本稿では探索視聴適合率を導入したが、このような視聴行動の支援能力の観点から推薦機能を評価するための指標や方法論にも研究の焦点が当てられるであろう。一方で、利用機器上のハードディスク容量の増加に伴い、放送コンテンツをより多く録画しさらに長期間保存できるようなAV機器においては、今後ますます利用者の context に適応した放送コンテンツの推薦が現実になると予想する。

本稿では、AV機器利用者に対する放送コンテンツの推薦技術の動向を紹介した。本稿が、読者に対して、利用者(人間)の嗜好という主観的な対象を扱う研究開発の面白さと難しさの両面を理解する助けとなれば幸いである。

参考文献

- 1) (社)電波産業会: <http://www.arib.or.jp/>
- 2) Herlocker, J. L. et al.: Evaluating Collaborative Filtering Recommender Systems, ACM Transactions on Information Systems, Vol.22, No.1, pp.5-53 (2004).
- 3) Ali, K. and Stam, W. V.: TiVo: Making Show Recommendations Using a Distributed Collaborative Filtering Architecture, Proceeding of the International Conference on Data Mining and Knowledge Discovery, ACM Press, pp.408-413 (2001).
- 4) Breese, J., Herlocker, J. and Kadie, C.: Empirical Analysis of Predictive Algorithms for Collaborative Filtering, Proceeding of the 14th Conference on Uncertainty in Artificial Intelligence, pp.43-52 (1998).
- 5) Pearl, J.: Probabilistic Reasoning in Intelligent Systems, Morgan Kaufmann, CA (1988).
- 6) Sakai, T. et al.: Pic-A-Topic: Gathering Information Efficiently from Recorded TV Shows on Travel, AIRS 2006, Lecture Notes in Computer Science 4182, pp.429-444 (2006).

(平成19年7月20日受付)

村上 知子(正会員)

tomoko.murakami@toshiba.co.jp

1998年慶應義塾大学政策・メディア研究科修士課程修了。同年(株)東芝入社。現在、同社研究開発センターにて、機械学習、データ・テキストマイニング技術の研究開発に従事。博士(政策・メディア)、人工知能学会、行動計量学会各会員。