

SNSを利用した 協調フィルタリングによる番組推薦手法

澤井里枝^{†1} 有安香子^{†1}
藤沢寛^{†1} 金次保明^{†1}

近年、放送予定番組の中からユーザの好みに合った番組を推薦する手法が多数提案されている。従来の手法では、番組の視聴履歴や Web の閲覧履歴などを収集することでユーザの好みを学習し、推薦する番組を判定するものが多かった。しかし、そのような履歴データは、テレビをあまり見ていないユーザや、個人情報を公開しないユーザなどからは収集できない。さらに、本人の履歴を利用する手法では、新しい種類の番組や嗜好の変化への対応が困難という問題があった。そこで本研究では、ソーシャルネットワークサービス (SNS) からユーザどうしの関係や他ユーザの書込みを取得し、協調フィルタリングにより番組推薦する手法を提案する。SNS を利用することで、履歴データや Web への書込みが取得できないユーザも対象とすることが可能となる。また本手法では、単に本人の好みに合った番組を推薦するだけでなく、流行の番組や友人が視聴している番組などユーザが知らなかった番組も推薦する。

A TV Program Recommendation Method Using SNS Based on Collaborative Filtering

RIE SAWAI,^{†1} KYOKO ARIYASU,^{†1}
HIROSHI FUJISAWA^{†1} and YASUAKI KANATSUGU^{†1}

In recent years, a lot of methods for recommending the scheduled TV programs adapting to user's preference have been proposed. Many methods estimate the user's preference by using his collected history data of TV watching or Web browsing, and decide recommend TV programs. However, such history data cannot be collected from users who rarely watch TV, and those who decline to offer personal data. Also, it's difficult for the method using user's history data to correspond to occurrence of new kinds of programs and change of user's preference. We propose a method for collecting users' relation and their descriptions from SNS (Social Network Service), and recommending TV programs based on collaborative filtering. By using SNS, we also target the

users who offer neither history data nor description on Web. Moreover, our method recommend not only programs adapting to users' preference but also users' unknown programs, such as popular programs and those users' friends watching.

1. はじめに

近年、放送のデジタル化や多チャンネル化により多種多様の番組が放送されるようになり、視聴者の番組に対する嗜好や視聴形態は多岐に渡るようになってきた。このため、放送予定番組の中から自分の好みに合った番組を毎日探し出すのは労力のかかる作業であり、そのような番組を自動的に推薦するサービスが求められている¹⁾。従来の番組推薦手法では、文献 2)-4) のようにユーザの番組視聴傾向を学習することでユーザの嗜好を推定し、本人の好みに合った番組を推薦するものが多かった。しかし、個人情報保護の観点から視聴傾向を番組推薦サービスに提供してもらうことは困難である。また、PC 型の受信機であれば視聴履歴蓄積機能を備えた専用のアプリケーションをインストールすることで視聴傾向を学習することは可能であると考えられるが、一般の受信機にそのような機能をもつものは未だ存在せず、家族で受信機を共有している場合は個人の嗜好をそれぞれ区別して推定することができない。

番組の視聴履歴以外にも、文献 5), 6) のように、Web の閲覧履歴や書込みを取得することでユーザの嗜好を学習する手法がある。このような手法ではテレビの視聴履歴よりも学習に必要な情報の取得は容易であるが、日常的に Web を利用しないユーザや、あまり書込みを行わないユーザからは十分な情報を取得できないため、ユーザ本人のプロファイルを作成して番組推薦できない。また、嗜好の変化に対応した推薦を受けるためには、継続的な Web の閲覧や新たな書込みを行ってもらうことが必要となる。

さらに、テレビ視聴や Web 閲覧など本人の行動から好みを学習して推薦する手法では、過去に利用した情報と類似した情報が推薦されがちとなり、新しい種類の番組や本人が出会ったことのない番組は推薦されにくいという問題があった。

そこで本研究では、行動履歴を取得できないユーザや Web にあまり書込みをしないユーザにも番組を推薦できるよう、ソーシャルネットワークサービス (SNS) からユーザどうし

^{†1} NHK 放送技術研究所
Science & Technology Research Laboratories, Japan Broadcasting Corporation

の関係や、代表的なユーザの書込みを取得し、協調フィルタリングにより番組推薦する手法を提案する。SNS とは、参加するユーザが互いにコミュニケーションを取り合うことを目的としたコミュニティ型の Web サービスである。サービスによって備えている機能は異なるが、コミュニケーションを促進するための機能として、自分のプロフィールを公開する機能や、お気に入りのユーザを登録する機能、登録したユーザをたどり新たな人間関係を構築する機能、登録ユーザと直接メッセージをやり取りする機能、同じ趣味をもつユーザどうしで利用できる掲示板の機能、登録ユーザの書込みのみを優先的に閲覧する機能などさまざまな機能をもつ。このようなソーシャルメディアは特に若い世代を中心に広まっており、家族・親戚や友人・知人、世代間の絆を深めるためのツールとなっている⁷⁾。

その中でも、Twitter⁸⁾、Google Buzz⁹⁾、Ameba なう¹⁰⁾、Timelog¹¹⁾ などのミニブログと呼ばれる SNS では、短い文章でリアルタイムにコミュニケーションする場を提供している。ミニブログ型の SNS では、現在いる場所やしていること、感じていることが頻繁に書き込まれており、その中で現在視聴している番組に関して感想を書き込むユーザも多い。本手法では、このようなリアルタイム性の高い SNS から放送中の番組に関する書込みを取得し、SNS 上のユーザの視聴状況を推定することで、推薦する番組を算出する。人と人との繋がりが形成されている SNS からユーザどうしの関係を取得することで、多くのユーザからお気に入り登録されている、すなわち多くのユーザから支持されている代表的なユーザが好んで視聴している番組や、SNS 上で繋がりのある、すなわち共感するユーザが視聴している番組、さらに流行の番組を推薦する。このように、本人の履歴データから直接本人の視聴しそうな番組を推薦するのではなく、SNS 上で関係のあるユーザや代表的なユーザの書込みを利用することで、本研究では履歴データも Web 上での書込みも取得できないユーザや、継続的に履歴データを取得して嗜好の変化を検出できないユーザをも対象とし、普段見ていなかった番組や新しいジャンルの番組を視聴してもらうきっかけを作ることを目指す。

以下、第 2 章で関連研究について述べ、第 3 章で本手法で利用する SNS の特徴をまとめる。次に第 4 章で我々が提案する番組推薦手法を定義し、第 5 章で実験を行う。最後に第 6 章でまとめと今後の課題について述べる。

2. 関連研究

協調フィルタリング手法とは、嗜好が類似したユーザあるいはユーザグループのプロファイルを利用して推薦する手法である^{12),13)}。協調フィルタリングを利用した代表的なシステムとして、ネットニュースの記事を推薦する GroupLens¹⁴⁾ や音楽を推薦する Ringo¹⁵⁾ が

ある。これらのシステムでは、ユーザが利用したアイテムに対して数段階の得点を与える。そして、各アイテムに対する得点の傾向から他ユーザとの類似度を算出し、その類似度と他ユーザが付けた得点から推薦アイテムの得点を予測する。このようにユーザどうしの類似度をもとにした協調フィルタリング手法では、多数のユーザ間のプロファイルを比較して類似したユーザを探す必要があるが、これは非常に計算コストの高い作業である。本研究は、SNS 上でのユーザどうしの関係を用いることで、多くのユーザから類似したユーザを探すことなく、どのユーザの書込みを利用して番組推薦するかを容易に決定できる。

SNS の研究としては、社会学および心理学の観点から SNS におけるコミュニケーションの特性を明らかにする研究があり、本人プロフィールの公開状態とコミュニケーションの積極性との関係を求める研究や、性別・年代別の利用者意識を抽出する研究などがある¹⁶⁾。本研究では、SNS 上での番組に対する反応を利用した番組推薦手法を提案しているが、SNS 利用の積極性とテレビ番組の視聴状況の関係という観点でも今後調査していく必要がある。

3. SNS の特徴

現在運用されている SNS は国内外に多数存在するが、本稿では特に会員数が多く書込みのリアルタイム性の高い SNS として Twitter を利用することとする。Twitter のようにリアルタイム性が高く「いま」に関する書込みが頻繁にされている SNS であれば、他の SNS でも本手法は同様に適用できると考えられるが、お気に入りのユーザ登録や書込みの内容を表すタグなど備えている機能は SNS によって異なるため、個々のシステムに合わせた調整が必要である。

Twitter では、ユーザがお気に入りのユーザを登録することをフォローと呼ぶ。また、発言の内容を表すためのタグをハッシュタグと呼び、例えば「#nhk」と「#」を頭に付けたタグを発言に付与した場合、NHK に関する発言であることを意味する。ハッシュタグは多くのユーザで共有され、番組に関する発言には放送局名や番組名をキーとした複数のハッシュタグを付与することが慣習化している。これと同様の機能は、名称や形式は異なるが他の SNS でも用意されているのが一般的である。

3.1 予備実験

SNS 上での発言の傾向を調べるために、2010 年 1 月 31 日 24 時間分の発言を Twitter から収集し、発言数・発言者数と、流行の番組を知るための尺度として一般的な視聴率を比較した。

図 1 に、1 月 31 日に NHK 総合で放送された各番組に対する視聴率と、各番組の放送

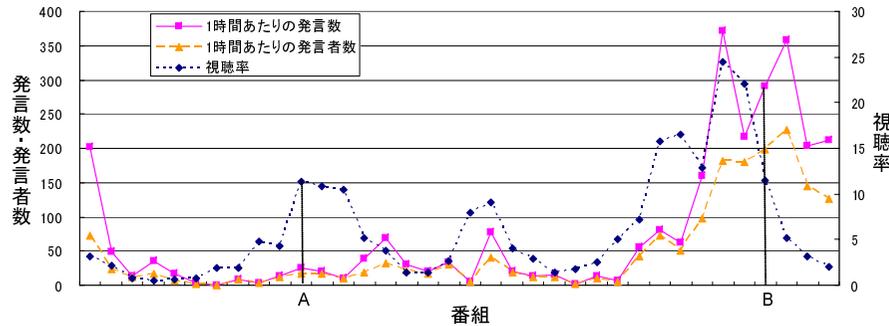


図 1 視聴率と単位時間あたりの発言数，発言者数

時間帯に Twitter 上でハッシュタグ「#nhk」を付与して発言された数，それらの発言者数を示す．横軸には番組を番組開始時刻の早い順に並べた．ただし，放送時間長 10 分以下のニュース番組や告知番組などは除外してある．また，各番組で放送時間長にばらつきがあるため，発言数・発言者数は番組の放送時間長を 1 時間に換算した．図 1 は各番組の放送時間帯に対する発言の推移を示しており，前後に放送された異なる番組に対する発言も含まれる．発言対象となる番組の推定については 4.1.3 節で述べる．

図 1 より，視聴率が高くても，その放送時間帯の発言が多いとは限らないことがわかる．これは，定時ニュースや大河ドラマなどもともと視聴率が高い番組ほど次回放送の番組も視聴する人が多いのに対して，発言数・発言者数は実際に放送された内容に応じて変動したためと考えられる．なお，各番組の視聴率と発言数，視聴率と発言者数の相関係数はそれぞれ 0.49，0.53 であった．

収集した発言の中身を調べると，実際に番組を視聴した上で，興味深い話題を扱っている場合や面白い映像が流れている場合に番組内容に対する発言が多くなる傾向にあった．特に，図 1 の番組 A「おはよう日本」と番組 B「NHK スペシャル」は同じような視聴率 11% 台であるにも関わらず，発言数はそれぞれ 1 時間あたり 24.5，291 と全く異なった．この日の NHK スペシャルは無縁社会という多くの視聴者から反響が得られたテーマを取り扱っており，リツイート（他ユーザの発言を引用して自分のアカウントから発言すること）に自分のコメントを追加することで活発に意見を交わす動きもあった．また，他者の発言を見て「今，何をやっているのか？」と番組に興味を示す発言も見られた．

このように，それまで気付いていなかった番組や関心のなかった番組でも，SNS 上の人

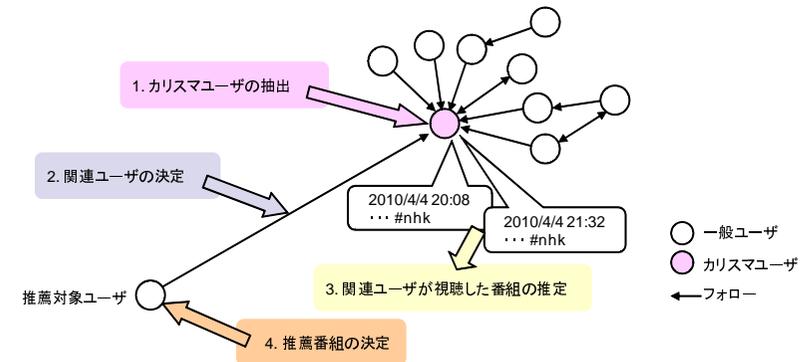


図 2 番組推薦の概要

と人との繋がりを利用し，かつ実際に番組を視聴したユーザの発言を基に番組推薦することで「今面白そうなことをやっているらしい」「さんが見ているなら」と放送番組のリアルタイム視聴に結び付けられる可能性がある．そこで本番組推薦手法では，視聴率ではなく，SNS 上の書込みとユーザ間の関係を利用する．

4. 番組推薦手法

本章では，SNS 上の書込みやユーザどうしの関係から番組を推薦する手法を定義する．番組推薦の基準に応じて以下に 3 つの方法を示す．以下の方法は Twitter に適用することを想定しているが，Twitter 以外の SNS でも同様に本手法を適用できると考えられる．3 つの方法はそれぞれ独立しているため，推薦方法 1~3 の結果を全て推薦する，推薦方法 1 と推薦方法 2 で共通する結果のみ推薦する，など方法の組合せは対象とする SNS やユーザの状況に応じて変更できる．

4.1 推薦方法 1

図 2 に本稿で提案する番組推薦手法の概要を示す．図 2 の丸印がユーザを表し，色により後で定義するカリスマユーザと一般ユーザを区別する．また，矢印によりフォローを表し，矢印元のユーザが矢印先のユーザをフォローしていることを意味する．

図 2 の 1. から 4. が本手法の手順である．以下，推薦対象ユーザに番組推薦するための方法を手順ごとに示す．

4.1.1 カリスマユーザの抽出

発言やフォローなどを積極的に行わないユーザに対しても番組推薦ができるよう、番組推薦の基準となる代表的なユーザをあらかじめ抽出しておく。本研究ではこのようなユーザをカリスマユーザと呼び、以下の2つの条件を全て満たすユーザとする。なお、カリスマユーザでないユーザは一般ユーザと呼ぶ。

- フォローされている数が一定以上
全ユーザのフォローの状況を解析し、多くのユーザからフォローされている、すなわち人気のあるユーザをカリスマユーザの条件とする。
- 一定以上の頻度で番組について発言
番組に関するハッシュタグが発言に付加されている場合、番組について発言したとみなす。全発言を調べ、よく番組に関する発言しており、ある程度テレビ番組に関心をもっているユーザをカリスマユーザの条件とする。なお、発言対象番組の推定については4.1.3節に示す。

一般に、ハッシュタグを付加せずに発言するユーザも多数存在するが、個々の発言が短文であるために、文章から発言の対象を特定するのは困難である。そこで、ここではハッシュタグを付加して発言する習慣をもったユーザのみをカリスマユーザとみなす。

4.1.2 関連ユーザの決定

本手法では、推薦対象ユーザへ推薦する番組を決定するために推薦対象ユーザのユーザプロフィールを作成する。しかし、推薦対象ユーザが積極的に発言しないユーザである場合、本人の発言からユーザプロフィールを直接作成できないため、SNS上で関連のあるユーザやカリスマユーザの発言を基に推薦対象ユーザのユーザプロフィールを作成する。ユーザプロフィールの作成に利用するユーザを本研究では関連ユーザと呼ぶ。関連ユーザの条件を以下に示す。

- 推薦対象ユーザがフォローしている場合
 - フォローしているカリスマユーザ
 - フォローしている一般ユーザのうち、一定以上の頻度で番組について発言しているユーザただし、カリスマユーザの条件である発言頻度よりも小さい頻度を閾値とする。
- 推薦対象ユーザがフォローしていない場合、あるいは前項の条件に一致するユーザが1人もいない場合
 - 全カリスマユーザ

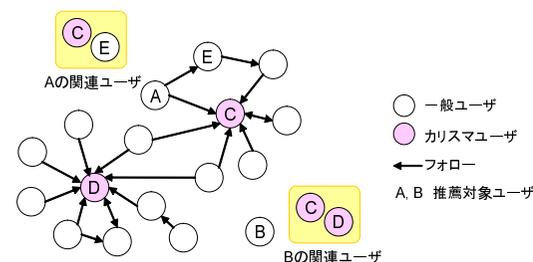


図3 ユーザ間の関係

図3にユーザのフォロー状況と関連ユーザの例を示す。ユーザAとユーザBが推薦対象ユーザであり、ユーザEは一定以上の頻度で番組について発言しているユーザとする。

4.1.3 関連ユーザが視聴した番組の推定

本稿では、NHKに関する発言を扱うために、関連ユーザの発言の中でも特にハッシュタグ「#nhk」の付加された発言を対象とし、以下の手順で関連ユーザの視聴した番組を推定する。

まず第一に、関連ユーザの発言に番組を表すハッシュタグが付加されていれば、ハッシュタグに記載の番組を視聴していたと推定する。例えば「龍馬伝」に関する発言には一般に「#ryomaden」、「#taiga」というハッシュタグが用いられるため、これらのハッシュタグをキーとして視聴番組を推定する。ただし、ユーザが発言を記述し投稿するまでにはある程度のタイムラグが発生するため、放送時間帯の前後の発言についても、明示的なハッシュタグが付加されていればその番組を視聴していたとみなす。

第二に、番組を表すハッシュタグが付加されていない発言に対して、番組名や副題のキーとなる単語を含んでいればその番組を視聴していたと推定する。例えば『NHKスペシャル「無縁社会～“無縁死”3万2千人の衝撃～」』については「NHKスペシャル」、「無縁社会」、「無縁死」をキーとしてあらかじめ登録しておき、これらのキーとのマッチングにより視聴番組を推定する。一つ目の手順と同様に、放送時間帯の前後の発言であっても、キーがマッチすればその番組を視聴していたとみなす。

第三に、上記2つの手順で視聴番組が推定できない発言については、編成情報と照らし合わせ、その発言時刻にNHKで放送されていた番組を視聴していたと推定する。

ここで、ハッシュタグ「#nhk」を付与しているとしても、NHKという法人について発言している場合や過去の番組について発言している場合もあるため、現在放送中の番組につ

いて発言しているとは限らない。そのため、厳密には発言の内容を解析する必要があるが、Twitterのようにリアルタイム性の強いSNSでは、番組を視聴しながら放送に関するハッシュタグを付与して発言されることが多い。そこで本研究では発言時刻とハッシュタグにより番組を視聴しているとみなすこととする。正確な発言内容の特定については今後の課題とする。

4.1.4 推薦番組の決定

前節までに求めた関連ユーザと視聴番組の履歴から推薦番組を決定する手順を以下に示す。まず、各番組について番組のEPG情報を形態素解析し、各語の評価値を算出する。番組 d_i の特徴を表現する番組ベクトルを

$$D_i = \langle w_{i1}, \dots, w_{in} \rangle \quad (1)$$

とする。 n は総単語数である。 w_{ij} は番組 d_i における語 k_j の重みとし、

$$w_{ij} = e_{ij} \cdot idf_j$$

と定義する。 e_{ij} は、番組 d_i の EPG 情報に語 k_j が含まれていれば 1, 含まれていなければ 0 とする。 idf_j は $tf-idf$ 法で求められる idf (Inverse Document Frequency) 値と同様に、全番組の EPG 情報に対して語 k_j が出現する番組の割合から求める、いわゆる語の希少性である。ここで、本手法では、 $tf-idf$ 法で求められる tf (Term Frequency) 値と同様の語の出現頻度は考慮しない。 EPG 情報は長さ上限のある短い文章であり、かつ重要かどうかに関わらず同じフレーズが EPG 情報の番組概要と詳細情報に重複して掲載される場合が多いためである。

次に、関連ユーザの発言から推定した視聴番組をもとに各関連ユーザのユーザプロフィールを作成する。関連ユーザ q_r が視聴した番組の番組ベクトルから、関連ユーザ q_r のプロフィールベクトル Q_r を

$$Q_r = \sum_m b_m \cdot D_m \quad (2)$$

とする。ただし、関連ユーザが一つの番組に対して複数回発言をしても、その番組を 1 回視聴したとみなす。 b_m は番組 d_m の希少性を意味する。本手法では、語の希少性に加え、放送の編成上の特性を考慮して番組の希少性という概念を導入する。番組の希少性の詳細については次節で述べる。

そして、推薦対象ユーザ q_u の関連ユーザのプロファイルベクトルから、 q_u のプロファイルベクトル Q'_u を

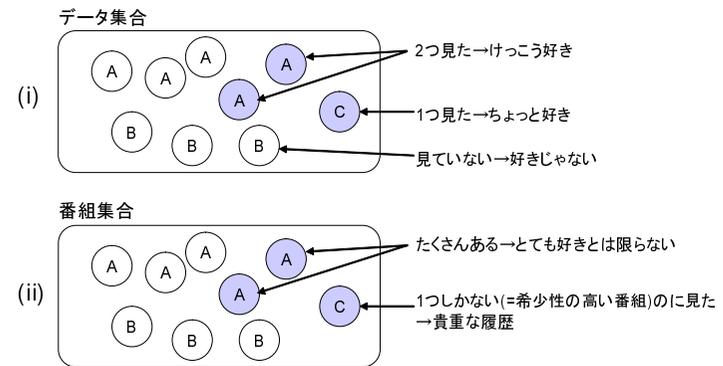


図 4 対象となる集合に応じた嗜好の学習

$$Q'_u = \sum_l a_{ul} \cdot \frac{Q_l}{|Q_l|} = \langle z_{u1}, \dots, z_{un} \rangle \quad (3)$$

とする。ただし、 a_{ul} は推薦対象ユーザ q_u と q_u の関連ユーザ q_l との関連の度合いを表し、 a_{ul} の大きさはフォローの有無に応じて降順に、双方がお互いフォローし合っている場合、片方からのみフォローしている場合、お互いフォローしていない場合、とする。 z_{uj} は推薦対象ユーザ q_u にとっての語 k_j の重みである。

このように、関連ユーザの視聴履歴から各関連ユーザのユーザプロフィールを作成する。そして、推薦対象ユーザの全関連ユーザのユーザプロフィールから作成するユーザプロフィールを推薦対象ユーザのユーザプロフィールとする。

最後に、推薦対象ユーザ q_u と番組 d_i の類似度を

$$Similarity_{ui} = \frac{\sum_k z_{uk} \cdot w_{ik}}{\sqrt{\sum_k z_{uk}^2} \sqrt{\sum_k w_{ik}^2}} \quad (4)$$

とする。この類似度が高い番組を推薦対象ユーザ q_u に推薦する。

4.1.5 番組の希少性

本手法では、視聴履歴からユーザのプロファイルを学習するにあたり、番組の希少性という概念を導入する。図 4 にアイテム集合の違いに応じた履歴の考え方を示す。図 4 の色付きのアイテムが利用履歴のあるアイテムであり、色の付いていないアイテムが利用履歴のないアイテムとする。各アイテムには内容を区別するために記号が振られているが、同じ記号

が振られたものは同じ種類のアイテムであることを意味する。

Webのように大量かつ多種多様なデータ集合の場合、図4の(i)に示すように、多く閲覧された種類のデータほどユーザの興味の度合いは大きく、リストなどで提示されたにも関わらず閲覧しなかったデータにはあまり興味をもたないと推定されることが多かった。しかし、放送という環境では編成が限られているため、図4の(ii)に示すように、ユーザが何度も視聴した番組ほど興味の度合いが大きいとは限らない。例えば、ニュース番組は日々何度も放送されているため、放送されているからという理由のみで何となく視聴している可能性もある。逆に、冬のスポーツに興味があったとしても夏には放送されないため視聴したくても視聴できない。放送環境では、希少性の高い番組ほど、多くの番組からあえて選択して視聴したという貴重な履歴になると考えられる。

そこで、*tf-idf*法の*idf*値(語の希少性)が、全アイテムに対する特定の語を含むアイテムの割合から求めるのと同様に、全番組に対する同じ種類の番組の割合から求める値を番組の希少性と定義する。番組の種類判別には、(社)電波産業会の標準規格 ARIB STD-B10¹⁷⁾で定義されているジャンル大分類(*content_nibble_level1*)とジャンル中分類(*content_nibble_level2*)を利用する。これらはEPG情報に含まれる。ジャンル大分類には「ニュース/報道」や「スポーツ」などがあり、ジャンル中分類には例えば「ニュース/報道」の中に「定時・総合」や「天気」などがある。ジャンル大分類とジャンル中分類の2つ組で番組のジャンルを表すが、1つの番組あたり1~3のジャンルが割り当てられることが多い。本手法では、ジャンルの順序と組合せが全て同じである番組を同じ種類の番組とみなし、番組の割合に反映させる。

4.2 推薦方法2

推薦方法2の手順1~3は推薦方法1の手順1~3と同じとし、手順4でプロフィールベクトルを算出せず、定時編成番組・シリーズ番組に対して、関連ユーザが視聴する頻度を求める。そして、一定以上の頻度で視聴した番組を特定する。

次に、全関連ユーザから一定割合以上の関連ユーザが高頻度で視聴した番組を推薦対象ユーザへ推薦する。

4.3 推薦方法3

推薦方法3は視聴傾向を調べるユーザを関連ユーザに限定せず、現在放送されている番組に対する全ユーザの発言数が一定数以上を超えた場合に、その番組が目玉されていると判断して推薦対象ユーザへ知らせる。

4.4 各推薦方法の特徴

前節までに定義した3つの推薦方法について、それぞれの特徴を以下にまとめる。

● 推薦方法1

推薦方法1は、関連ユーザが視聴した番組のEPG情報をもとにユーザプロフィールを作成するため、番組内容に基づく推薦方法であるといえる。そのため、単発の番組・シリーズ番組に関わらず、特に好きな俳優が出演している回や、興味のあるトピックを扱う回のみ推薦することが可能である。ただし「NHKスペシャル」や「クローズアップ現代」のように放送ごとに内容が異なる番組は、毎回視聴していても内容によっては推薦されない可能性がある。

● 推薦方法2

推薦方法2は、シリーズ番組に対して、視聴したかどうかをもとに推薦番組を決定するため、番組タイトルに基づく推薦方法であるといえる。そのため、定期的に視聴されている番組は必ず推薦できる。しかし、番組タイトルごとに視聴したかどうかに基づいており、番組の内容や出演者などは考慮しない。

● 推薦方法3

推薦方法3は、これまでの視聴履歴に関わらず、現在放送中の番組が目玉されているかどうかをもとに推薦する。推薦方法1と推薦方法2では、番組の開始直前に推薦することが可能であるが、推薦方法3では既に開始している番組について推薦するという点で異なる。そのため、過去の番組に対する反響からはこれから放送する番組が面白いかどうか予測できなくても、実際に視聴したユーザからのリアルタイムな反応から推薦するかどうかを決定できる。

5. 実 験

本章では、前章までに定義した番組推薦手法を実際のデータに適用した場合の効果について評価する。以下、本手法のプロセスである視聴番組の推定、カリスマユーザの抽出に関する評価をそれぞれ5.1節、5.2節で行う。そして、本稿で定義した番組の希少性やユーザプロフィールの効果を確認するために、推薦方法1により算出された番組推薦結果を5.3節に示し、5.4節で本手法の特長について考察する。

5.1 視聴番組の推定

4.1.3節で定義した視聴番組推定方法の精度を評価する。

2010年1月31日にTwitterからハッシュタグ「#nhk」の付与された発言を収集し、本

手法によって推定した視聴番組と実際に入手により分類した発言対象番組を比較した。ただし、Twitter 上の発言は「ああ… #nhk」のように、入手によっても発言対象が判断しにくいものがあるが、放送内容や周囲の発言を参考に類推した。

この結果、収集した 2093 件の発言に対して、88.2%の発言の対象番組を正しく推定できた。一方、0.9%の発言は、NHK の番組に対する発言ではなく、ハッシュタグの羅列や Web ニュースの引用などであった。それ以外の 10.9%の発言に関しては、発言内容をさらに詳しく解析することで視聴履歴を推定できると考えられるが、具体的な方法については今後の課題とする。

5.2 カリスマユーザの抽出

前章で定義したように、本手法では書込みもフォローもしないユーザに対してはカリスマユーザの書込みを基に番組推薦をする。そのため、もしも全てのカリスマユーザの番組視聴傾向が同じであれば、全てのユーザに対して同じ番組しか推薦できなくなってしまう。そこで本節では、カリスマユーザに応じて異なる番組が推薦できるよう、また全カリスマユーザを関連ユーザとするユーザに対しては特定の番組だけでなく多様な番組を推薦できるよう、本手法により視聴傾向の異なるカリスマユーザが偏りなく抽出されていることを確認する。

実験の概要は次のとおりである。実験期間は 2010 年 1 月 25 日から 2 月 19 日とし、Twitter からハッシュタグ「#nhk」の付与された発言を収集したところ、39701 の発言があり、発言したユーザは 9389 人であった。また、カリスマユーザの抽出条件を

- 1000 人以上のユーザからフォローされる
- 発言収集期間内に NHK に関して 50 回以上発言した

としたところ、9 人のカリスマユーザが抽出された。形態素解析には茶筌¹⁸⁾を用いた。

まず、収集した発言から式 (2) により各カリスマユーザのユーザプロフィールを作成し、3 月 4 日から 6 日に NHK 総合で放送される全番組とユーザプロフィールとの類似度を算出した。次に、各日において、任意の 2 カリスマユーザの類似度に対して t 検定を実施した結果、有意水準 5% で有意な差があった。また、カリスマユーザごとに類似度の高い順に番組のランキングを作成したところ、上位 5 番組の順位が一致するカリスマユーザはいなかった。これらの結果から、本手法によりそれぞれ異なる視聴傾向をもったカリスマユーザが抽出できたことを確認した。したがって、カリスマユーザに応じて多様な番組を推薦できるといえる。

5.3 番組推薦結果

本手法による番組の希少性とユーザプロフィールの効果を確認するために、日常的に SNS

表 1 推薦結果

順位	3 月 4 日		3 月 5 日		3 月 6 日	
	番組タイトル	視聴率	番組タイトル	視聴率	番組タイトル	視聴率
1	プラタモリ	10.3	ニュースウオッチ 9	8.8	土曜スタジオパーク	3.4
2	ニュースウオッチ 9	10.5	とくせん ETV	1.5	NHK アーカイブス	7.7
3	世界と出会った日本人	(6)	世界ふれあい街歩き	5.8	LIFE	2.6
4	あなたが主役 50 ボイス	5.7	プライム H	1.4	東京カワイイ TV	2.9
5	おはよう日本	9.5	連続テレビ小説	14.0	おはよう日本	11.3

を積極的に利用しないユーザに対して推薦方法 1 による推薦結果を算出した。前述のように、フォロー機能を利用しないユーザの関連ユーザは、本手法によって抽出する全カリスマユーザである。そこで、5.2 節の条件で作成したカリスマユーザのユーザプロフィールから推薦対象ユーザのユーザプロフィールを作成した。なお、式 (3) の a_{ui} は全てのカリスマユーザで 1 とした。そして、3 月 4 日から 6 日に NHK 総合で放送される番組から推薦番組を算出した結果、各日において評価値が上位の番組を表 1 に示す。参考のため、視聴率もあわせて表記する。ただし、括弧付きの視聴率は、番組としての視聴率データが入手できなかったため、毎分視聴率のデータから概算したものである。

表 1 より、推薦結果の上位には視聴率の高い番組も低い番組も入っていることがわかる。視聴率の低い番組であっても、EPG 情報との類似度が高いカリスマユーザが複数いると評価値は高くなるためである。したがって、これまであまり広く知られていなかった番組でも、カリスマユーザのユーザプロフィールを用いることで、カリスマユーザにとって関心の高い番組であれば推薦できる。また、NHK 総合の編成はニュース番組の割合が比較的多いため、ニュース番組に対する発言が多くなりがちであるが、番組の希少性を考慮することにより、ニュース番組だけに偏らずドキュメント番組やドラマ番組など多様なジャンルの番組を推薦できた。

本稿では特に SNS の発言機能やフォロー機能をあまり利用しないユーザを対象としたが、今後は SNS を積極的に活用しているユーザに対する評価も実施する必要がある。

5.4 本手法の特長

5.1 節の結果より、本手法では広く公開されている SNS 上の書込みから番組の視聴履歴をある程度推定できることが示された。そのため、本手法では個人情報である視聴履歴や Web の閲覧履歴を各ユーザから提供してもらわなくてよい。初期設定情報として従来のようなキーワード登録や個人プロフィール設定なども必要ないため、ユーザは意識的に番組推

薦サービスを受けるための作業をしなくてよい。

本手法では、SNS 上でフォローしているユーザがいれば、そのユーザの書込みから番組推薦することもできるが、フォローしているユーザが全くいなくても、カリスマユーザの書込みをもとに番組推薦できる。すなわち、ユーザは必ずしも SNS の機能を使いこなしていなくてもよい。全カリスマユーザをもとに番組推薦する場合、5.2 節、5.3 節で示したようにこれまで知らなかった番組も含め多様な番組を推薦できるため、新しい番組を視聴するきっかけを作ることができる。逆に、SNS でユーザが知人や友人をフォローしている場合、知人・友人がよく視聴している番組を推薦することが、視聴行動へと結びつきより強いきっかけとなり得る。フォローしているユーザと同じ番組を視聴することにより、その番組に関して後で話題を共有することができる。

また、一般にプロファイルを自動学習する手法は、履歴などを逐次的に反映させていくため、プロファイルが完成するまで時間がかかるという問題があった¹⁹⁾。しかし、本手法は、ユーザが新たなアプリケーションの使用を開始し、それを継続的に使用することで自分のプロファイルを育てていく形態ではない。既に膨大なユーザの発言が蓄積されている SNS のデータを基にしてユーザプロファイルを協調的に作成するため、番組推薦サービスを受けたユーザのこれまでの SNS の使用状況に応じて番組推薦できる。

6. おわりに

本稿では、SNS を用いた協調フィルタリングによる番組推薦手法の提案と実装をした。本手法では、カリスマユーザを定義したことで、積極的にフォローや書込みをしないユーザに関してもユーザプロファイルを作成でき、実験で確認したようにカリスマユーザに応じて多様な番組を推薦できる。これにより、あまり広く知られていない番組であっても、SNS 上で話題になっている番組や、実際に視聴しているユーザから大きな反響がある番組の視聴へと誘導できる。一方、ユーザが知人や関心のあるユーザをフォローしている場合、そのユーザが視聴しそうな番組と一緒に視聴できる。

実験では、経験的にカリスマユーザの抽出条件を設定したが、今後は実用的なパラメータを検討することが課題として挙げられる。また、SNS 上でフォローしているユーザがいないうーザを対象として番組推薦結果を示したが、SNS を積極的に利用しているユーザを対象とする場合も、その妥当性を評価する必要がある。さらに、本手法を利用したサービスの実施の可能性について放送現場と議論を進めていく予定である。

参 考 文 献

- 1) 村上知子: “AV 機器利用者に対する放送コンテンツの推薦,” 情報処理, Vol. 48, No. 9, pp. 984–988 (2007).
- 2) 土屋誠司, 佐竹純二, 近間正樹, 上田博唯, 大倉計美, 蚊野浩, 安田昌司: “TV 番組推薦システムの構築とその有用性の検証,” 情報処理学会研究報告 (HCI), Vol. 2006, No. 3, pp. 95–102 (2006).
- 3) W. Dai and R. Cohen: “Dynamic Personalized TV Recommendation System,” in *Proc. Workshop on Personalization in Future TV(TV’03)*, pp. 12–21 (2003).
- 4) Z. Yu and X. Zhou: “TV3P: An Adaptive Assistant for Personalized TV,” *IEEE Trans. Consumer Electronics*, Vol. 50, No. 1, pp. 393–399 (2004).
- 5) 山口瑶子, 瀬々潤: “Web 閲覧履歴を用いた TV 番組推薦システム,” DEIM2010, A3-2 (2010).
- 6) Y. Nakamura, T. Itou, H. Tezuka, T. Ishihara, and M. Abe: “Personalized TV-Program Recommendations Based on Life Log,” in *Proc. ICCE’10*, pp. 143–144 (2010).
- 7) 総務省: “平成 22 年度 情報通信白書,” <http://www.soumu.go.jp/johotsusintokei/whitepaper/>.
- 8) Twitter: <http://twitter.com/>.
- 9) Google Buzz: <http://www.google.com/buzz>.
- 10) Ameba なう: <http://now.ameba.jp/now/top/>.
- 11) Timelog: <http://timelog.jp/>.
- 12) 澤井里枝: “ブロードバンド時代における情報フィルタリングの動向,” 第 13 回データ工学ワークショップ, ミニサーベイ (2002).
- 13) 土方嘉徳: “嗜好抽出と情報推薦技術,” 情報処理, Vol. 48, No. 9, pp. 957–965 (2007).
- 14) P. Resnick, N. Iacovou, M. Sushak, P. Bergstrom, and J. Riedl: “GroupLens: An Open Architecture for Collaborative Filtering of Netnews,” in *Proc. CSCW’94*, pp. 175–186 (1994).
- 15) U. Shardanand and P. Maes: “Social Information Filtering: Algorithms for Automating “Word of Mouth”,” in *Proc. CHI’95*, pp. 210–217 (1995).
- 16) 大向一輝: “SNS の現在と展望 - コミュニケーションツールから情報流通の基盤へ -,” 情報処理, Vol. 47, No. 9, pp. 993–1000 (2006).
- 17) 社団法人 電波産業会: “標準規格 ARIB STD-B10 4.8 版「デジタル放送に使用する番組配列情報」”(2010).
- 18) 茶釜: <http://chasen-legacy.sourceforge.jp/>.
- 19) A. I. Schein, A. Popescul, L. H. Ungar, and D. M. Pennock: “Methods and Metrics for Cold-Start Recommendations,” in *Proc. SIGIR2002*, pp. 253–260 (2002).