

Twitter を利用した番組視聴傾向によるユーザ分類手法

澤井 里枝[†] 中川 俊夫[†][†]NHK 放送技術研究所

1. はじめに

近年、さまざまなソーシャルネットワーキングサービス (SNS) が利用されるようになり、番組を視聴しながら感想を書き込み、他のユーザと共有するという視聴スタイルが広まってきた。筆者らは、特にリアルタイム性の高い Twitter 上の書き込みやユーザ間の繋がりを利用してユーザプロフィールを作成し、番組推薦をするための研究を行ってきた^[1]。これまでの手法では、1 人のユーザに対して、ユーザの嗜好を表したプロフィールを 1 つ作成し、新しい書き込みなどが増えるごとにプロフィールも累積的に更新してきた。そのため、ユーザに提供されるのは、プロフィールに適合した 1 パターンの推薦結果であった。しかし、ユーザによっては、平日の朝には短時間でニュースをチェックし、休日の夜にはバラエティやドラマをじっくり楽しむなど、その時の状況に応じて視聴したい番組が異なる場合も多い。

そこで本研究では、Twitter 上の書き込みから、番組に関するツイートの時系列的な特徴を解析し、その特徴から得られる番組視聴傾向に応じてユーザを分類する手法を提案する。毎週決まった時間に視聴するユーザ、時間は決まっていないが好きな番組だけ視聴するユーザ、といったようにユーザを分類することにより、分類グループの特性に合わせて、適切なタイミングで必要な番組推薦や情報提供が可能となる。

2. 関連研究

ユーザの状況に適応した推薦技術は従来からいくつも提案されてきた。例えば、奥らは時刻、予算、同伴者などのコンテキストに応じて飲食店を推薦する手法を提案した^[2]。また、倉島らは位置情報や過去の行動履歴などから次の行動を予測して推薦する手法を提案した^[3]。

表 1 に推薦技術で考慮される一般的なコンテキスト情報を示す。コンテキスト情報はその特性から 4 つに大別できる。推薦するアイテムや目的に応じてこれらの情報を組み合わせてユー

“A Method of Classifying Users by TV Watching Habit Using Twitter”

Rie Sawai[†], Toshio Nakagawa[†]

[†]Science & Technology Research Laboratories,
Japan Broadcasting Corporation

ザプロフィールは作成されるが、情報が増えるに従って推薦のためのコストは増大する。また、上の項目ほどプロフィールを作成する際に重要な情報であるが、個人情報の観点から入手が困難な情報である。本研究では、一般に公開されている SNS の情報のみを用いることで、計算コストや個人情報入手の問題を緩和しつつも、ユーザの状況に応じたきめ細かなサービスの実現を目指す。

表 1. 一般的なコンテキスト情報

ユーザ	嗜好、注目の対象・方向、予算、友人、行動履歴など
環境	場所、交通、気象、照明、騒音、近くの人など
端末	性能、ディスプレイ、ネットワーク、通信コストなど
時間	日時、曜日、月、季節など

3. ユーザ分類手法

本章では提案手法の処理手順について述べる。

まず、Twitter から解析対象とするツイートを収集し、文献[1]に示した方法によりツイートの時間とキーワードからツイートが話題の対象としている番組を推定する。

次に、ツイートの時間と対象番組により、以下の 2 つの観点でユーザの視聴傾向を分析する。

i) 時間起因の視聴傾向

生活スタイルによって「毎日同じ時間に視聴する」、「毎週土曜日に視聴する」などの時間に起因した視聴傾向を以下の方法で分析する。

視聴傾向を分析する対象日の集合を D とする。ある日 $d \in D$ において、1 日を 24 時間に分割した時間帯 $t_x \in \{t_0, \dots, t_{23}\}$ ごとのユーザ u のツイート数を a_x とし、時間起因によるツイートのベクトルを $A_{u,d} = \langle (t_0, a_0), \dots, (t_{23}, a_{23}) \rangle$ と定義する。任意の 2 ベクトル $A_{u,d_i \in D}$, $A_{u,d_j \in D}$ の内積により異なる日どうしの類似度を算出し、各値が高いほど視聴時間帯が固定されているとみなす。

ii) 番組起因の視聴傾向

「好きなドラマを毎回見る」、「シリーズのドキュメンタリーを毎回見る」などの番組に起因した視聴傾向を以下の方法で分析する。

全番組の集合を $P = \{p_1, \dots, p_n\}$ とする。ユーザ u の番組 $p_y \in P$ に対するツイート数を b_y とし、番組起因によるツイートのベクトルを

$B_u = \langle (p_1, b_1), \dots, (p_n, b_n) \rangle$ と定義する。また、あるシリーズ番組 $S \subset P$ のベクトルを $P_S = \langle (p_1, e_1), \dots, (p_n, e_n) \rangle$ と定義する。ただし、 e_y は、 $p_y \in S$ ならば1、 $p_y \notin S$ ならば0とする。 B_u と例えば大河ドラマ S' の $P_{S'}$ の内積値が高いほど大河ドラマに注目して視聴しているとみなす。

最後に、i)、ii)で算出した内積値によりユーザを分類する。

4. 実験

2013年7月1ヵ月分のツイートを収集し、「#nhk」が付加されたツイートのうち、10回以上ツイートした4793ユーザのツイートを対象として視聴傾向の分析を行った。

横軸を実験対象日全組合せに関する前章 i) の内積値の平均値、縦軸を ii) の内積値の平均値として、ユーザの分布を図1に示す。また、横軸を同じ曜日全組合せに関しての i) の内積値の平均値とした場合のユーザ分布を図2に示す。

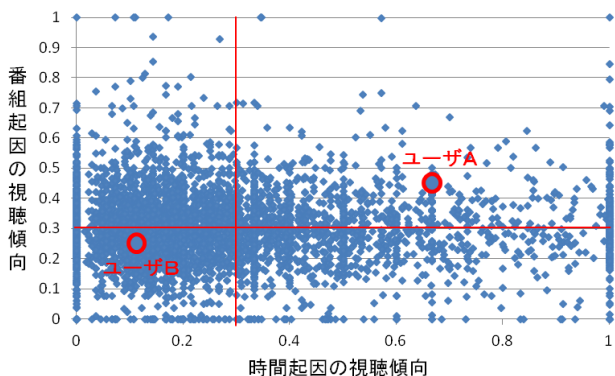


図1. ユーザの分布(横軸: 全日程)

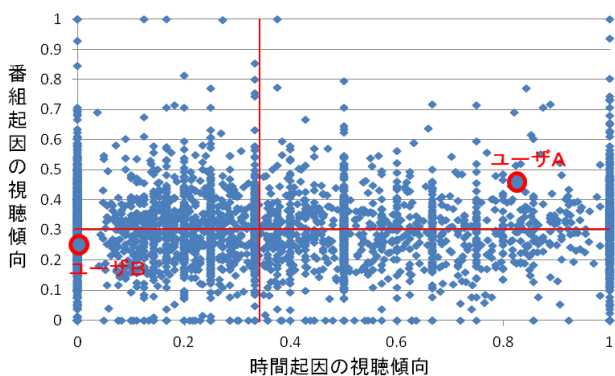


図2. ユーザの分布(横軸: 曜日ごと)

図1と図2を比較すると、図2の方がユーザの分布は左右に広がり、視聴傾向の特徴がより顕著に表れる結果となった。これは、曜日に関わらず視聴時間が決まっているユーザだけでなく、曜日によって視聴時間が異なるユーザも存在したためと考えられる。例えば図中のユーザA

は「大河ドラマ」、「クローズアップ現代」といった夜の番組を集中して視聴したユーザであったため、全日程での視聴時間の類似度もある程度高かったが、曜日ごとの方が類似度はより高くなった。一方、ユーザBは「おはよう日本」、「ニュースウオッチ9」といった番組を不定期に視聴するユーザであったため、全日程での視聴時間の類似度は低く、曜日による規則性は見られなかった。

このようなユーザの視聴傾向を考慮することにより、適切なタイミングで必要な内容の情報提供を可能とする。図中の赤線は全ユーザの平均値を表す。これらの平均値を閾値としてユーザを分類する場合、表2のような方針で情報提供することで、各グループの生活スタイルに合った情報提示ができる。

表2. ユーザ分類と情報提示方針の一例

時間番組	平均値未満	平均値以上(全日程・曜日ごとのうち、値の高いもの)
平均値以上	視聴したい番組が決まっているユーザ ⇒ 時間帯に関係なく、視聴番組の関連番組を推薦	毎日/毎週、同じ時間帯に同じ番組を規則正しく視聴するユーザ ⇒ 視聴番組を深掘できる関連情報やVODなどを提示
平均値未満	視聴する時間も番組も決まっていないユーザ ⇒ SNS上の友人が視聴している番組など、他の手法で導出した番組の情報を提示	生活リズムで視聴時間は決まっているが、気分によって視聴番組は異なるユーザ ⇒ 視聴時間帯に、さまざまな番組を推薦

ここで、本稿ではまず「#nhk」というハッシュタグのツイートに焦点を絞って実験を行ったが、ユーザを分類する閾値の最適値は、タグの付いていないツイートも含め全てのツイートを考慮した上で決定する必要があるため今後の課題とする。

5. おわりに

本稿では、Twitter上の書込みからユーザの番組視聴傾向を解析し、視聴傾向に応じてユーザを分類する手法を提案した。毎日・曜日ごとの視聴時間、視聴番組といった特徴を複合的に解析することで、生活スタイルに合った番組推薦や情報提供を可能とする。今後はユーザの全ツイートを対象としてより詳細な特徴を分析することで、本手法のさらなる改良を目指す。

参考文献

- [1] 澤井ら: “SNSを利用した協調フィルタリングによる番組推薦手法”, 情処研報, vol.2010-DBS-151, no.43 (2010).
- [2] 奥ら: “情報推薦におけるユーザの価値判断基準モデルに基づくコンテキスト依存型ランキング方式”, 情処論文誌データベース, vol.2, No.1, pp.57-80 (2009).
- [3] 倉島ら: “行動範囲と興味の同時推定モデルによる地域情報推薦”, 情処論文誌データベース, vol.6, no.2, pp.30-41 (2013).