

投稿型サイトにおけるユーザの興味を反映したチャート型ランキングの個人化手法

金子 鷹弥[†] 牛尼 剛聡^{††}

[†] 九州大学大学院芸術工学府 〒 815-8504 福岡県福岡市南区塩原 4-9-1

^{††} 九州大学大学院芸術工学研究院 〒 815-8504 福岡県福岡市南区塩原 4-9-1

E-mail: [†]takaya@gospel.aid.design.kyushu-u.ac.jp, ^{††}ushiana@design.kyushu-u.ac.jp

あらまし 現在, Web 上で多くのユーザが投稿型サイトを利用している. 投稿型サイトでは毎日膨大な量のコンテンツが投稿されるが, それらは視聴するユーザにとっては, 未知の新規コンテンツである. そうした中で, 投稿型サイトにおいて, 未知の新規コンテンツを適切にユーザに提示する手法が重要になっている. 我々は, 多くの投稿型サイトや, その他のサービスにおいて広く利用されている, チャート型ランキングに注目した. チャート型ランキングとは, アクセス数等の普及度に基づいて決定されるランキングのことである. このチャート型ランキングを利用することにより, ユーザへの新規コンテンツの提示を行なう. 本研究では, チャート型ランキングの基準として利用されている, 普及度や流行度と, ユーザの興味の 2 つの情報を利用し, チャート型ランキングの個人化を行う.

キーワード コンテンツ管理, ランキング, 投稿型ウェブサイト

A Method for Personalizing Chart Type Ranking Based on Interests of a User for Content Sharing Site

Takaya KANEKO[†] and Taketoshi USHIAMA^{††}

[†] Graduate School of Design, Kyushu University Shiobaru 4-9-1, Fukuoka-shi Minami-ku, Fukuoka, 815-8504 Japan

^{††} Faculty of Design, Kyushu University Shiobaru 4-9-1, Fukuoka-shi Minami-ku, Fukuoka, 815-8504 Japan

E-mail: [†]takaya@gospel.aid.design.kyushu-u.ac.jp, ^{††}ushiana@design.kyushu-u.ac.jp

1. はじめに

近年, Web 上でのコンテンツの利用形態が大きく変化している. Web が登場してしばらくの間, Web を利用する目的は主に情報取得のためであった. そのため, Web 上に存在するコンテンツの多くは, マスメディアが発信したニュースや記事であったり, またはプロのアーティストが歌う楽曲等, 従来型のマスメディアで流通してきたコンテンツであった. しかし, YouTube [1] をはじめとするコンテンツ投稿型サイトが普及し, ユーザがコンテンツ閲覧者でありながら, 制作者, 投稿者となった. SNS や, Twitter 等のマイクロブログの登場により, こうした傾向はさらに加速し, 現在では, Web 上で発見可能なコンテンツの多くに, ユーザが投稿したコンテンツが含まれている. ユーザが投稿するコンテンツが増加することにより, Web 上に存在する情報は飛躍的に増大している. 例えば, YouTube の 2011 年の総合アクセス数は 1 兆を超え, 1 分間に投稿され

る動画の総再生時間は 60 時間分を超えている [3]. このように, 投稿型サイトには, 一定間隔に大量の新規コンテンツが投稿され, ユーザは既に, 投稿されたコンテンツすべてにアクセスすることが不可能である.

また, 従来流通していたコンテンツの多くは, 専門家が作成したものであったため, ある程度の品質の保証が行われていた. しかし, 投稿型のコンテンツでは, コンテンツを投稿しているユーザの多くは投稿したコンテンツに関する専門家では無いため, コンテンツの品質の保証が無い. 非専門家 (アマチュア) が作成したコンテンツは, 品質が低いものが含まれている一方で, アマチュアのユーザが作成したコンテンツ中にも, 非常に良質なコンテンツが含まれていることがある. その結果, 現在 Web 上に存在するコンテンツ数は膨大になり, それらの品質は, 良質なものから低質なものまで, 玉石混合といった状況にある. そうした中で, ユーザが大量のコンテンツから良質なコンテンツを見つけることが困難になっている.

以上から、本研究では投稿型サイトにおいて、ユーザにとってアクセスする価値のあるコンテンツの効率的な発見が可能なシステムを開発する。

2. 投稿型サイトにおけるコンテンツ提示

2.1 投稿型サイトの特徴

現在、Web上の投稿型サイトに投稿されたコンテンツの利用が一般化している。動画投稿サイトにおいては、2005年頃からYouTubeやgoogle videoといったサービスが開始され、今ではYouTubeの2011年の総合アクセス数は1兆を超えている。また、近年多くのユーザを獲得しているソーシャル・ネットワーキング・サービス(SNS)やTwitterをはじめとしたマイクロブログにおいても、テキスト、画像、動画等、コンテンツの種類は違うものの膨大な数のコンテンツが投稿されている。したがって、SNSやマイクロブログなども、ユーザが投稿するという形式から、投稿型サイトの一種と捉えることもできる。

投稿型サイトの特徴の一つは、そこに投稿されるコンテンツの多くが、UGC(User Generated Content)であるという点である。ユーザは投稿型サイトにおいて、様々なUGCの閲覧と視聴を行う一方で、自身で作成したコンテンツを投稿する場合があり、従来のブロードキャスト型のコンテンツ配信サービスとは異なり、ユーザは閲覧者であると同時に、コンテンツの制作者、投稿者となることができる。現在では、作曲やアニメーション制作等、自身の創作活動の発表の場として、一般的に利用されている。YouTubeやニコニコ動画[2]のような大規模な投稿型サイトでは、一般ユーザがコンテンツを投稿できるため、短期間に大量のコンテンツが投稿される。したがって、個々のユーザが投稿された全てのコンテンツにアクセスして内容を確認することは不可能である。また、Twitter等のマイクロブログにおいては、多くのfollowingを持つユーザのtimelineは非常に高速に流れ、ユーザは多くの情報を見逃してしまう。このように、ユーザが投稿型サイトに投稿されるコンテンツを、ユーザが1つ1つ吟味し、アクセスすることは困難である。

多くの研究では、アクセスする価値のあるコンテンツは、ユーザの嗜好に一致するコンテンツとしている。しかし、アクセスする価値とは、ユーザの嗜好との一致だけではない。日頃アクセスしないようなコンテンツや、意外性のあるコンテンツにアクセスした際、ユーザが新しい興味や知見を得ることがある。本研究では、この、ユーザにとって、新しい興味や知見を得る可能性のある、意外性のあるコンテンツを、発見性の高いコンテンツと呼ぶ。

2.2 代表的なコンテンツ提示手法とその問題点

現在の投稿型サイトで利用されているコンテンツ発見手法は、主に以下の3つの手法に分類できる。

- 検索エンジン
- 情報推薦
- ランキング

代表的なコンテンツアクセス手法である、検索、情報推薦、チャート型ランキングの、発見性の程度を表した概要図を図に示す(図1)。

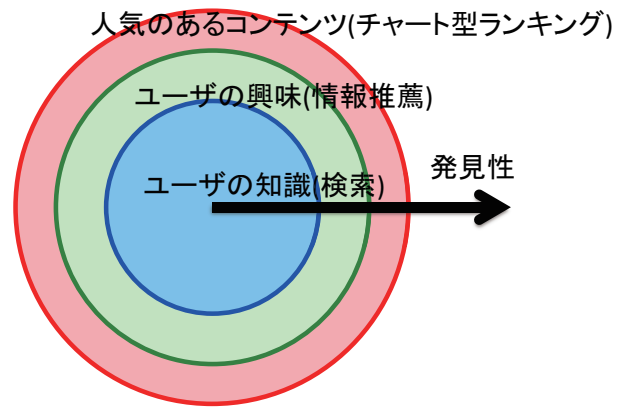


図1 代表的なコンテンツアクセス手法と発見性

検索は、ユーザが検索要求をクエリとして明示的に指定して、クエリに合致するコンテンツ一覧を取得する方法であり、投稿型サイトにおいても広く利用されている。投稿型サイトの多くはソーシャルタギングを導入しており、投稿者や他のユーザが付与したタグに基づいて、1つ1つのタグを検索時のインデックスとすることが可能な場合も多い。検索は、既知のユーザが投稿したコンテンツや、有名な楽曲等へのアクセスに対しては非常に有用である。しかし、前述したように、多量の新規コンテンツが存在する投稿型サイトにおいて、新規コンテンツに対するクエリを生成することは困難である。そのため、「ユーザが知らないが、もし提示すれば興味を持つコンテンツ」の発見にはつながらない。

推薦では、ユーザが明示的にクエリを入力する必要がなく、ユーザの興味に合致すると思われるコンテンツ一覧を提示する。推薦手法には、ユーザのアクセスログに基づいてユーザプロフィールを取得し、ユーザプロフィールに基づいた推薦を行う手法や、特定のコンテンツに対して、そのコンテンツに興味を持つユーザに推薦するコンテンツを提示する等、様々な方式が存在する。

上記の3種類のコンテンツ発見手法の中で、投稿型サイトにおいて、特徴的な手法であると考えられるのが、ランキングである。ランキングでは、一定期間内のユーザのアクセス頻度などに基づいて、コンテンツを順位付け、ユーザに提示する。ニコニコ動画において提供されているチャート型ランキングを図2に示す。

チャート型ランキングとは、検索結果のランキングに対して、YouTubeやニコニコ動画で提供されているランキングのように、アクセス数や販売数等といった基準によって順序づけられたランキングである。Webでチャート型ランキングが利用される対象としては、投稿型サイトのような膨大なコンテンツ数を持つサイトや、ニュースのように、ある一定の間隔で提示すべきコンテンツが変化するサービス等がある。チャート型ランキングは、CDや書籍などの実世界のコンテンツにおいても、「売上ランキング」などとして古くから利用されてきた。評価の基準となっているのはアクセス数や販売数等で、これらは主に流行度や普及度を示しており、ユーザはチャート型ランキングを閲覧することで、流行のコンテンツを知ることが出来る。



図 2 ニコニコ動画におけるチャート型ランキング

このチャート型ランキングの最も大きな特徴は、ユーザからの入力が必要ないことである。また、一定期間ごとに更新され、YouTube のチャート型ランキングは 2 時間毎に更新され、ニコニコ動画のチャート型ランキングは 1 時間毎に更新されるため、一定間隔ごとにチャート型ランキングにアクセスすることにより、新規コンテンツへのアクセスが容易になっている。

チャート型ランキングでは、多くのユーザが多くアクセスしているコンテンツは、そのコンテンツに興味を持つユーザが多いであろうという仮定に基づいている。ランキングでは、推薦同様にユーザは明示的に、取得したいコンテンツの特徴を明示的に指定する必要はない。

チャート型ランキングは、ユーザの明示的な要求を指定すること無しに、有用なコンテンツを発見するための有効な手段を提供する。しかし、現在のランキングでは、ユーザの嗜好が反映されていないため、必ずしもユーザの興味に合致したコンテンツばかりが提示されるわけではない。

この問題点を解決するために、多くの投稿型サイトでは、ジャンル毎のランキングを提供している。図 3 に YouTube において提供されているランキングのジャンルの一覧を示す。現在 YouTube では、YouTube が提供している 14 のカテゴリに分けたランキングと、全体のランキングを含め、合計 15 種類のランキングが提供されている。これは、コンテンツ数が膨大であり、主に上位 100 件を提示するランキングでは、ユーザに価値のあるコンテンツの発見が困難であること、「音楽」等のカテゴリは比較的アクセスされやすく、「政治」や「教育」のように、アクセスされにくいカテゴリがランキングに入りにくいいため、ジャンル毎のランキングが提供されている。

しかし、ジャンル毎のランキングを提示したとしても、ユーザの嗜好が反映されていないため、多数のユーザの興味と一致しない興味を持つユーザに取っては、有効的に機能しない。本研究では、この問題を解決するために、ユーザのアクセス履歴に基づいて、ユーザの嗜好を反映したランキング手法を提案する。

ユーザの入力が無いため「サッカー」等、決まったコンテン

全て	ペットと動物
エンターテインメント	教育
ゲーム	旅行とイベント
コメディ	映画とアニメ
スポーツ	科学と技術
ニュースと政治	自動車と乗り物
ハウツーとスタイル	非営利団体と社会活動
ブログ	

図 3 youtube に於けるランキングのジャンル

ツしか閲覧しないユーザや、流行しているコンテンツに興味がないユーザにとっては、的確なコンテンツ提示ができない。

チャート型ランキングにユーザの興味を加えることにより個人化を行い、新規コンテンツ提示に適したチャート型ランキングを開発する。

3. チャート型ランキングの個人化手法

チャート型ランキングの特徴として、表 1 で比較したように、ユーザの入力が必要ないことから、未知のコンテンツの提示に適していることが挙げられる。しかし多くのチャート型ランキングはユーザの興味や嗜好を反映させておらず、全てのユーザが同じチャート型ランキングを提示されている。つまり、ユーザによってはチャート型ランキングを閲覧しても、満足できるコンテンツを見つけられない場合がある。そのため、チャート型ランキングに対し、ユーザの興味を反映させた個人化を行なうことを考える。ここでは、チャート型ランキングがコンテンツの評価基準としている、普及度や流行度といった基準と、ユーザの興味の 2 つの要素を用いて個人化されたランキングの計算を行なう。

3.1 アプローチ

チャート型ランキングの個人化を行うにあたり、我々はチャート型ランキングの継続的閲覧に注目した。チャート型ランキングの特徴から、ユーザは一定期間毎にチャート型ランキングにアクセスすることが多い。これは、短い間隔で変化する流行を評価の基準としており、そのため、チャート型ランキングを提供しているサービスの多くが、短期間での更新を行なっているからだと考えられる。このような利用形態は、ニュースや株式情報等、情報の鮮度が重要な場合や、真新しく変化するコンテンツを利用する際に利用されており、実際に毎日新聞を読む、といった我々の行動にも見られる特徴である。本研究ではこの一定期間毎の継続的閲覧を利用し、ユーザの興味を取得する。

提案手法の概要図を (図 4) に示す。ユーザは提示されたチャート型ランキングに対し、それぞれのコンテンツにアクセスする。その際、ユーザがアクセスしたコンテンツ集合に対し、ユーザが興味を持ったコンテンツ集合であると判断する。取得されたチャート型ランキング閲覧の履歴により抽出したユーザの嗜好をフィードバックし、次のチャート型ランキングに反映させる。以上の処理を継続することにより、よりユーザの嗜好を反映したチャート型ランキングを出力する。

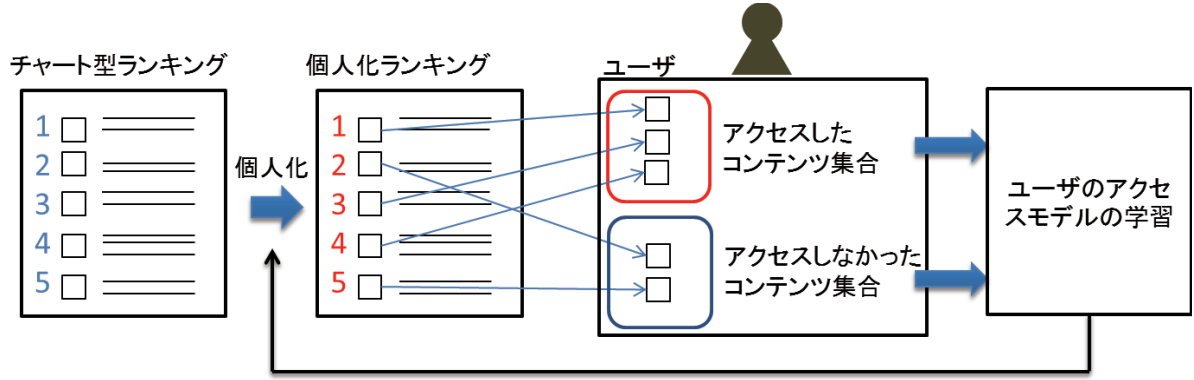


図 4 アクセス履歴に基づくランキングの個人化の概要



図 5 コンテンツのモデル化

3.2 ランキング手法

個人化されたチャート型ランキングを作成するにあたり、確率モデルを用いる。本手法では、個人化ランキングのために、チャート型ランキングの中で、コンテンツ c がユーザに提示された場合に、ユーザがそのコンテンツにアクセス（閲覧）する条件付き確率 $P(\text{アクセス} | c)$ を求める。この確率 $P(\text{アクセス} | c)$ を求める式は、ベイズの定理により以下のように変形できる。

$$P(\text{アクセス} | c) = \frac{P(c | \text{アクセス})P(\text{アクセス})}{p(c)} \quad (1)$$

いま、対象とするコンテンツ集合を C とすると、 C の構成要素全てに対して上記の条件付き確率を計算し、値が高いコンテンツから順にユーザに提示する。この式を計算する際、 $P(\text{アクセス})$ は、全てのコンテンツに於いて共通であるため、ランキングに影響は与えない。我々は時刻 t におけるコンテンツ c^t を、コンテンツ c^t の特徴を表すキーワード集合 $W = \{w_1, \dots, w_n\}$ と、アクセス数に基づいたランキング順位 r の組 $c^t = (W, r)$ としてモデル化する（図 5）。 C の例として、以下のような例を考えることができる。

$$c_1^t = (\{\text{なでしこ, サッカー}\}, 5) \quad (2)$$

$$c_2^t = (\{\text{AKB48}\}, 7) \quad (3)$$

ここで、 c_1^t は、キーワードが「なでしこ」、「サッカー」で、順位が 5 であるコンテンツを表す。一方、 c_2^t は、キーワードが「AKB48」で、順位が 7 であるコンテンツを表している。モデル化した c^t を用い、 $P(c^t | \text{アクセス})$ は以下の式で計算できる。

$$P(c^t | \text{アクセス}) = P(W | \text{アクセス}) \times P(r | \text{アクセス}) \quad (4)$$

この式 4 により、コンテンツの評価値を決定する。以下に、 $P(W | \text{アクセス})$ 、 $P(r | \text{アクセス})$ の計算について説明する。コンテンツのモデル化を行なったため、 $P(c)$ は以下のように計算できる。

$$P(c) = P(W) \times P(r) \quad (5)$$

ここで、式 5 より、 $P(W)$ は以下のように計算できる。

$$P(W) = \prod_{w_i \in W} P(w_i) \times \prod_{w_i \notin W} (1 - P(w_i)) \quad (6)$$

このとき、アクセスしたコンテンツの特徴語集合が W である確率 $P(W | \text{アクセス})$ は以下の式で計算できる。

$$P(w_t | \text{アクセス}) = \prod_{w_i \in w_t} P(w_i | \text{アクセス}) \times \prod_{w_i \notin w_t} (1 - P(w_i | \text{アクセス})) \quad (7)$$

ここで、ユーザ u がアクセスした全てのコンテンツ集合を $AC(u)$ 、単語 w を含むコンテンツ集合を $C(w)$ とするとき、 $P(w | \text{アクセス})$ は以下の式によって求めることができる。

$$P(w | \text{アクセス}) = \frac{|AC(u) \cap C(w)|}{|AC(u)|} \quad (8)$$

図 6 に上記の計算を行うために必要な集合の関係を示す。本研究では、例題として動画コンテンツを対象とし、単語はコンテンツに付与されたタグを利用する。

また、ランクが r であるコンテンツの集合を $RC(r)$ とするとき、アクセスしたコンテンツのランキングが r である確率 $P(r | \text{アクセス})$ は、指数分布に基づく以下の確率分布に従うものとする。

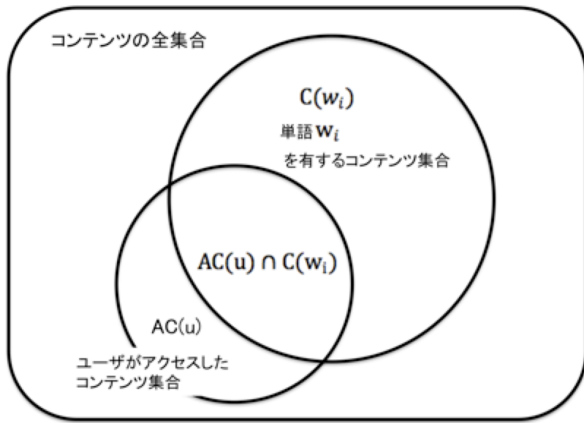


図 6 $P(W| \text{アクセス})$ の算出にあたる各種集合図

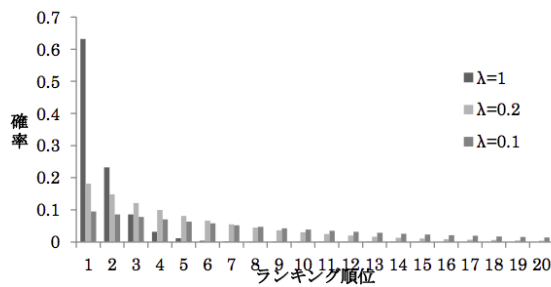


図 7 λ の値に対する $P(r| \text{アクセス})$ がとる値

$$P(r| \text{アクセス}) = \int_{r-1}^r \lambda e^{-\lambda x} dx \quad (9)$$

$$= e^{-\lambda r} + e^{-\lambda(r-1)} \quad (10)$$

$$= (e^{-\lambda} - 1)e^{-\lambda r} \quad (11)$$

このとき λ は標本平均の逆数で求められ、

$$\lambda = \frac{n}{\sum x_i} \quad (12)$$

となる．ここで、 x_i はユーザ u が i 番目にアクセスしたコンテンツのランキングある． λ の値によって、 $P(r | \text{アクセス})$ の値は変化する．

λ の値が 1, 0.2, 0.1 のときの $P(r | \text{アクセス})$ の、ランキング順位における推移を図 7 に示す．

4. プロトタイプ

提案手法に基づくプロトタイプシステムを制作した．開発環境は本プロトタイプシステムシステムでは、ユーザの利用開始時に、ユーザに対して流行に基づくチャート型ランキングを提示する．

今回対象としたチャート型ランキングは、YouTube で公開されているチャート型ランキングである Charts である．YouTube のチャート型ランキングは 2 時間毎に更新されており、本プロトタイプシステムにおいても、同様の更新間隔とした．提示されるコンテンツは YouTube が提供するチャート型ランキングに関しては 96 件、個人化を行なったチャート型ランキングに

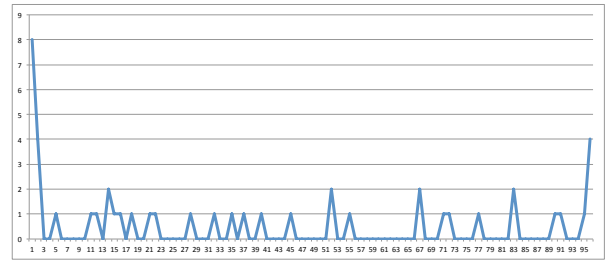


図 8 ユーザ A のランク順位別アクセス数

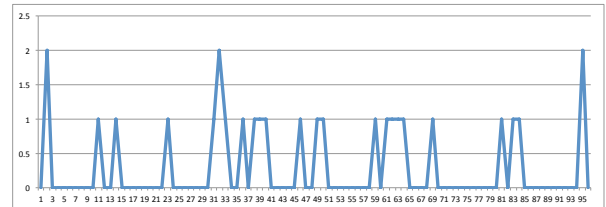


図 9 ユーザ B のランク順位別アクセス数

関しては上位 20 件としており、ユーザはチャート型ランキング中のコンテンツにアクセスし、動画を視聴できる．このプロトタイプを用い、データの収集を行なった．被験者は用意された 2 つのチャート型ランキングにアクセスし、コンテンツの視聴を行なう．ユーザのアクセスログを収集し、個人化したチャート型ランキングと、YouTube で提供されているチャート型ランキングの 2 つを利用して評価を行った．

実験によって取得された 2 名のアクセスデータに基づいて、ユーザがアクセスしたコンテンツのランキングの頻度を図 8 と 9 に示す．なお、プロトタイプの利用開始時期は 2011 年 12 月 20 日であり、A、B とともに同じ時期からプロトタイプを利用している．図 8 より、ユーザ A は 1 位のコンテンツへのアクセス率が非常に高く、流行を重視する傾向にあるユーザであると考えられる．対して、ユーザ B はランク順位に関わらずアクセスしており、流行よりも自身の興味を重視したコンテンツアクセスを行う傾向にあると考えられる．

ユーザ A、B のアクセス回数の高いタグを、表 1 および表 2 に提示する．表 1 より、最もアクセス回数の多いタグが Games であること、またモンスターハンターというタイトルが挙げられていることから、ユーザ A はゲーム関連のタグによくアクセスしていることがわかる．一方、表 2 により、ユーザ B も同様に Games が最も高いアクセス数であるが、他のタグと比較して大きな差が無い．また、News タグや金正日タグから、ユーザ A と比較して News への関心が高いと考えられる．

表 1 ユーザ A のアクセス率の高いタグ

タグ	アクセス回数
Games	12
Entertainment	10
Sports	5
2011	4
モンスターハンター	4

上記の例に示したように、ユーザのアクセスの傾向には多様

表 2 ユーザ B のアクセス率の高いタグ

タグ	アクセス回数
Games	6
News	5
Entertainment	4
mobile	3
金正日	3

性が存在する．また，提示されたコンテンツ数に対し，それぞれのユーザがアクセスするコンテンツ数が大きく異なる．これは，流行だけを取得するユーザは上位 5 件しかチャート型ランキングを閲覧しない等の理由によるものだと考えられる．

次に，YouTube が提供しているチャート型ランキングと，得られたユーザのアクセス履歴に基づいて個人化を行なったチャート型ランキングの 2 つをユーザに提示し，それぞれのチャート型ランキングにおいてユーザがアクセスしたコンテンツの順位を比較した．被験者は 4 名で，ユーザに提示したのはそれぞれのチャート型ランキングにおいて上位 10 件とし，アクセスするコンテンツ数は指定していない．4 名のユーザがアクセスした回数と，それぞれのチャート型ランキングにおける順位を表 3 および表 4 に示す．

順位	アクセス回数
1	1
2	0
3	2
4	1
5	1
6	1
7	0
8	3
9	3
10	2

表 3 YouTube

順位	アクセス回数
1	4
2	2
3	1
4	3
5	0
6	0
7	1
8	2
9	1
10	0

表 4 提案手法による個人化

また，それぞれの結果に対し，以下の式により，提案手法によるランキングおよび従来手法によるランキングの精度を求めた．精度の計算は以下の式によって行う．

$$\text{精度} = \frac{\text{興味を持ったコンテンツ数}}{\text{提示されたコンテンツ数}} \quad (13)$$

上記の計算方法により，それぞれのチャート型ランキングの精度を求めた結果を図 10 に示す．

この結果から，提案手法により個人化されたチャート型ランキングの精度は，1 位から 10 位では YouTube が提供しているチャート型ランキングを上回っていることが分かる．また，我々の提案手法により 1 位に提示されたコンテンツの精度が 1 となっており，的確な個人化が行なわれていると判断することができる．

図 10 では，提示されたそれぞれのチャート型ランキング上位 10 件に対し，ユーザのアクセス回数の平均はそれぞれのチャート型ランキング中で同じであった．つまり，今回の実験におい

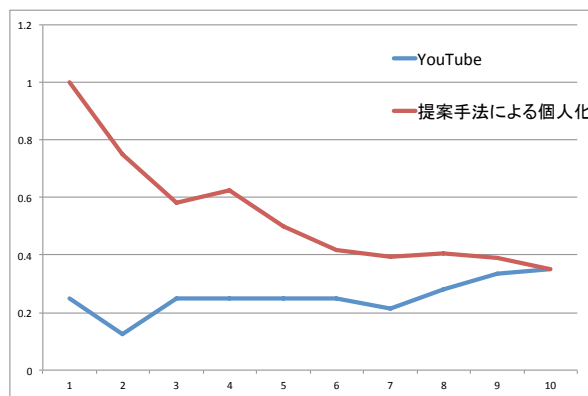


図 10 提案手法と従来手法の精度の比較

て，上位での精度が我々の提案手法が上回っていた理由として，比較的上位のコンテンツのアクセス率が高いためであると考えられる．

今回の実験では流行度，ユーザの興味の 2 つの要素を用いて個人化を行なったが，今後は 2 つの要素がどのように精度に対して影響しているのかを調査する必要がある．具体的に，被験者に対してシステム学習用のテストデータを提示し，興味のみを利用した個人化を行なうことにより，実際に興味情報がどれほど精度に寄与するのか判断する．それに対して，流行度については，YouTube 等が提供するアクセス数に基づくチャート型ランキングは流行度に基づくランク付けを行なっているため，今回実験を行なった結果をもとに判断することができる．しかし，正確なデータの取得を行なうためには，上位 100 件等，提示するコンテンツ数を増やし，さらに多くの被験者からのデータが必要であると考えられる．

5. 関連研究

ランキングの研究は主に検索エンジンが提示する結果を再ランキングするというものが多く，提案手法ようにアクセス数ベースのチャート型ランキングを用いるものは少ない．山本ら [4] は，ランキング結果をユーザの様々なインタラクションにより，自在に変化させることが可能な RerankEverything を提案している．RerankEverything は本研究と同様に，チャート型ランキングをユーザの意思で変化させることが出来るが，操作を加えるのはユーザであり，我々が対象とした投稿型サイトにおける未知の新規コンテンツを発見することは難しい．

小林ら [5] は，仮想の生態系を用いることにより，情報検索を行う Information Fishing を提案している．Information Fishing はクエリを餌，被検索文書を魚とし，ユーザの操作によって自在に変化する検索結果の提示を行うことができる．また，提示されるコンテンツには，ユーザの操作次第で未知のコンテンツが含まれる場合もあり，投稿型サイトにおいて適用することも十分可能である．しかし，クエリを入力する必要があるため，未知のコンテンツの発見に対して，十分な機能を提供しているとは言えない．

湯本ら [6] が提案する Web ページ集合を解とするランキング手法に関しては，対象が検索結果のランキングであるが，ラン

キング結果として集合を提示する点は我々も考慮する必要がある。情報推薦と比較して、チャート型ランキングが主に提示しているコンテンツ数は少なくないが、普及度を利用したチャート型ランキングにおいては、ユーザの嗜好に対する適合率が著しく低い。そのため、コンテンツ集合の提示を行うことにより、ユーザの嗜好に「サッカー」が含まれるのか、または「なでしこ」が含まれるのか等、類似した特徴語からの確かな嗜好を取得することができる可能性がある。

投稿型サイトに投稿されるコンテンツと同様に、ユーザにとって未知のコンテンツであるものとして、ニュース記事が挙げられる。馬ら [10] は、現在の情報フィルタリングでは困難である、ニュース記事のフィルタリングを行った。彼らは過去の配信履歴における時系列的特徴量として新鮮度、流行度、配信頻度を用いた。我々の目標はコンテンツのフィルタリングではないが、チャート型ランキングには類似したコンテンツが同時に上位にランキングされることが多々あり、ユーザの満足度向上や、多様なコンテンツの提示による正確な嗜好の取得のため、同様な考え方を導入を考慮する必要があると考えている。

6. おわりに

本論文では、流行度とユーザの興味を利用したチャート型ランキングの個人化システムの提案を行なった。

チャート型ランキングはアクセス数により評価されているため、コンテンツの流行の指標として利用されている。しかし、多くのチャート型ランキングはユーザの興味情報を利用していないため、閲覧するユーザによっては、適切なコンテンツ提示であるとは言えない場合がある。我々は、チャート型ランキングが持つ流行の指標としての側面を生かしつつ、ユーザの興味を反映した個人化を行なった。また、結果として本手法の有用性を確認することができた。今後、手法に対し、流行・ユーザの興味がそれぞれどれほど寄与しているかの確認が必要である。

本論文で定義した流行度はコンテンツのアクセス数によるランキング順位であったが、ユーザの興味についてタグを利用したことと同様に、流行しているワードを取得し、タグとのマッチングを行なうことにより、ワードの流行度・ユーザの興味との合致度というアプローチを行なうことができる。我々は kizashi [7] や Google Trend [8] からの単語の流行度取得を試みたが、タグとの的確なマッチングには至らず、本論文での実装ができなかった。プロトタイプについて非常に大きな処理時間を要する可能性があるが、ワードの流行度を適切に定義することが出来れば、より良い個人化が可能になると考えられる。

今回は一定期間毎に大量の新規コンテンツが投稿される。投稿型サイトを対象としてチャート型ランキングを利用したコンテンツ提示を行った。その理由は、SNS やマイクロブログ等を含め、ユーザがコンテンツをポストする、という機会が増えていることが最も大きな理由である。携帯端末が進化し、スマートフォンで撮影した写真や動画を、気軽にアップロードするユーザも増え、今後もユーザがコンテンツを投稿する機会が急速に増えていくと考えられる。これらの理由により、Web 上に存在するコンテンツは加速度的に増え続け、ユーザは情報の

取捨選択を迫られている。その際、ユーザのコンテンツ選択の基準となるのが、普及度や流行度といった物であると我々は考えている。本来流行しているコンテンツ提示に利用されているチャート型ランキングを利用し、流行しているコンテンツに対し、ユーザの興味等によるフィルタリングを行なうことにより、タイムリーなコンテンツ提示を行なうことが出来れば、ユーザの満足度は向上する可能性がある。一方で、ユーザがアクセスすることなく、消えていく多くのコンテンツの中にも、良質なコンテンツが存在し、そういったコンテンツを 1 つでもユーザに提示していきたいとも考えている。

タイムリーなコンテンツに対しフィルタをかけ、情報の取得を行なう、というモデルは、Twitter [9] と共通している。Twitter では、チャート型ランキングと同様に、短い期間に膨大な量のタイムリーなコンテンツがポストされている。ユーザはその膨大なコンテンツに対し、follow や list により、自らフィルタリングを行なっている。多くの following を持つユーザの timeline では、利用者の目に触れずに消えていくコンテンツが多数存在するが、ユーザが振り返って見返すことはそれほど多くない。この形態は、コンテンツの価値について、新鮮さ (コンテンツが投稿されてからの経過時間の短さ) が占める割合が大きくなっているということを意味する。

今後、新たなサービスを含め、ユーザが投稿したコンテンツが増加していくことは明らかである。チャート型ランキングをはじめ、Twitter の timeline 式のコンテンツ提示等、可能性を持つシステムは多い。まったく新しい手法の開発も含め、「新鮮な」コンテンツ提示の可能な提示手法を確立することが重要である。

文 献

- [1] YouTube <http://www.youtube.com/>
- [2] ニコニコ動画 <http://www.nicovideo.jp/>
- [3] YouTube 日本版公式ブログ <http://youtube.jpblog.blogspot.com/>
- [4] 山本 岳洋, 中村 聡史, 田中 克己「RerankEverything: ランキング結果閲覧のための 柔軟な再ランキングインタフェース」情報処理学会論文誌 (トランザクション) データベース, Vol.3, No.4 (TOD48), pp.48-64, 2010 年 12 月.
- [5] 小林正朋, 五十嵐健夫「Information Fishing: 即応的な情報検索と持続的な情報提示の統合」第 13 回インタラクティブシステムとソフトウェアに関するワークショップ (WISS 2005) 論文集, 香川, 2005 年 12 月, pp.63-68.
- [6] 湯本 高行, 田中 克己「Web ページ集合を解とするランキング手法」日本データベース学会 letters 5(4), 13-16, 2007-03
- [7] kizashi.jp <http://kizashi.jp/>
- [8] Google Trend <http://www.google.co.jp/trends/>
- [9] Twitter <http://twitter.com>
- [10] 馬強, 角谷和俊, 田中克己: 放送型情報配信システムのための時系列性を考慮した情報フィルタリング, 情報処理学会論文誌: データベース (TOD7), pp.46-57, 2000