

影響伝播に基づく投稿型コンテンツのランキング手法と評価

Design and Evaluation of a Method for Ranking Shared Content Based on Propagation of Influence

金子 鷹弥† 藤村 直美‡ 牛尼 剛聡‡

Takaya Kaneko Naomi Fujimura Taketoshi Ushiyama

1. はじめに

近年、インターネットの普及と Web 上のデジタルコンテンツの増加に伴い、ユーザのコンテンツの利用形態が変化している。インターネットが普及する以前は、ユーザはテレビ放送のようなブロードキャスト型のコンテンツ配信を利用することが多かった。ブロードキャスト型のコンテンツ配信では、同一のコンテンツを同時に複数のユーザが閲覧する。一方、インターネットを利用した Web 上のコンテンツ閲覧では、ユーザが興味のあるコンテンツを見つけ、好きな時に閲覧することが可能となった。

インターネット上に存在する大量のコンテンツから、ユーザが興味のあるコンテンツを見つけるための、代表的なアプローチは検索である。検索を行うためには、ユーザは検索結果となるコンテンツの特徴をクエリとして明示的に指定する。

検索は、ユーザが検索結果となるコンテンツの特徴を想定できる場合には、効果的なアクセス方法である。ブロードバンド放送によるコンテンツアクセスでは、ユーザは受動的にコンテンツを閲覧するために、興味のないコンテンツを閲覧しなければならない場合もある。しかし、検索によるコンテンツアクセスでは、ユーザは主体的にコンテンツを選別するため、興味のないコンテンツを閲覧する必要がない。これは、ユーザの満足度を向上させることにつながる。しかし、検索では効果的なアクセスを提供できない場合も存在する。検索では、未知のコンテンツに対しては、ユーザはそれらの特徴を明示的にクエリとして指定することができないため、ユーザがユーザは自分が興味のないコンテンツを閲覧しなくてよいという利点がある一方で、ユーザの興味を引く、ユーザの未知のコンテンツとの出会いの機会の減少につながっている。

現在、Web 上でのコンテンツ閲覧のための投稿型サイトの利用が一般化している。投稿型サイトには、短期間に膨大な量のコンテンツが投稿されるため、個々のユーザが、投稿された全てのコンテンツを閲覧することは困難である。投稿型サイトに投稿されるコンテンツには、ユーザが未知であるが、ユーザが興味を持つコンテンツの候補が多数存在する。しかし、検索によるアクセスでは、ユーザがそれらのコンテンツと出会う機会を提供することが困難である。

この問題点を解決するために、多くの投稿型サイトではコンテンツを何らかの基準に基づいて順位づけし、上位に順位づけされたコンテンツをユーザに提示している。本研究ではこのような順位づけをランキングと呼ぶ。ランキングによるコンテンツ提示ではユーザはクエリを明示的に指定する必要はない。コンテンツをユーザが興味を持つと思われる順番で提示可能なランキングは、ユーザが未知のコンテンツに出会う方式として有効である。投稿型サイトの利用形態として、定期的・継続的にアクセスし、面白いコ

ンテンツを見ることできないか、ニュースを確認するようにアクセスするような利用形態が考えられる。このような利用形態では、ランキングが有効である。

本研究では、ユーザが投稿型サイトを継続的に利用する際に、ユーザが興味を持つことができるコンテンツの新しい出会いを誘発する手法を開発することを目標とする。そのために、個人化されたランキング手法の開発を目的としている。本論文では、アクセス数等によって与えられたランキングを、コンテンツの質に基づいて再構成する手法を提案する。なお、本研究では、投稿型サイトの例として、ニコニコ動画を利用する。

2. 投稿型サイトにおけるランキングと問題点

従来、投稿型サイトのランキングの基準としてはアクセス数、ブックマーク数などが利用されることが多い。これは、コンテンツの普及度を示す指標である。普及度に基づいたランキングが利用されるのは、他のユーザによるアクセスやブックマーク登録が多く、世間に広く普及しているコンテンツは、ユーザが興味をもつ可能性が高いという仮定に基づいている。

普及度に基づくランキングではいくつかの問題点が存在する。その一つは、普及度が高いコンテンツが、必ずしも質が高いわけではないということである。この問題を解決するために、ランキングによって提示されるコンテンツの質を向上させる必要がある。本研究におけるコンテンツの質とは、そのコンテンツを閲覧することによって、ユーザが新しい発見や興味を持つコンテンツであることと定義する。質の高いコンテンツを発見するために、本研究ではコンテンツ同士の影響関係に注目してコンテンツを評価する。

3. アプローチ

3.1 影響伝播関係

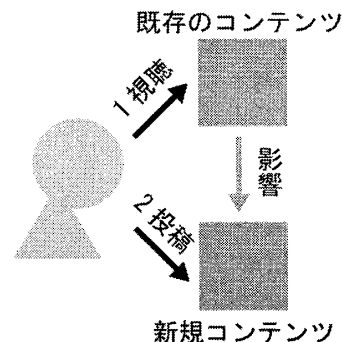


図1 影響伝播関係が出現する様子

本研究では、コンテンツの質を評価するために、コンテンツ間の影響伝播関係を利用する。影響伝播関係とは、あるコンテンツが他のコンテンツの影響を受けて制作された

†九州大学芸術工学府 Graduate School of Design Kyushu University

‡九州大学芸術工学研究院 Faculty of Design Kyushu University

ことを示す関係である。図1に影響伝播関係を有するコンテンツが出現する様子を示している。ユーザは、既存のコンテンツを閲覧し、そのコンテンツから影響を受けた新しいコンテンツを制作し投稿する。ここで投稿された新しいコンテンツに影響を受けて新しいコンテンツが生まれることもある。

本研究で例題として利用するニコニコ動画においては、ユーザがコンテンツを投稿する際、制作時に参考にした動画に対する参照リンクを記載することが一般的に行われている。図2に動画の説明文に存在する参照リンクの例を示す。本研究では、他の動画に対する参照リンクを有する動画に対して、参照先の動画から、影響伝播関係が存在していると考えられる。

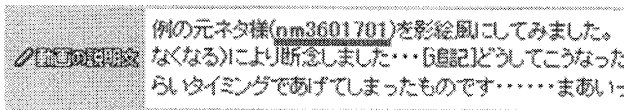


図2 動画説明文における参照リンクの例

3.2 リランキング手法

影響伝播関係によってリランキングを行う手順を以下に示す。

1. アクセス数に基づいたコンテンツの評価値を初期値とする。
2. コンテンツ間の参照関係に基づいて影響伝播関係を抽出する。
3. 影響伝播関係に基づいてコンテンツの評価値を導出する。
4. 再計算された評価値を用いてコンテンツを並び替えてユーザに提示する。

3.3 影響伝播関係に基づく評価値の導出

影響伝播関係に基づいた評価値の導出方法について述べる。コンテンツ c の評価値 $E(c)$ は以下の式によって計算する。

$$E(c) = E_0(c) + \sum_{c_i \in I(c)} \frac{E(c_i)}{L(c_i)}$$

ここで、 $E_0(c)$ は c の評価値の初期値、 $N(c)$ は c に影響を与えたコンテンツの数 (参照リンクの数) を表している。また $I(c)$ はコンテンツ c に影響を与えたコンテンツの集合を表す。図3は評価値の導出の例である。四角形がコンテンツを表し、その中の数字が評価値を表している。

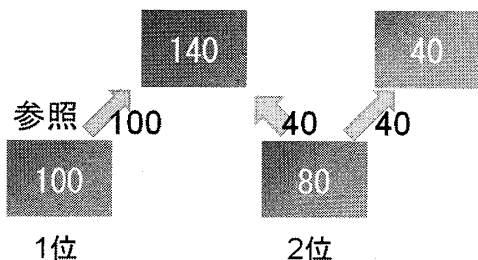


図3 重みづけ手法

4. 評価

提案手法の有効性を検証するために評価実験を行った。評価実験では、プロトタイプシステムを利用して、ニコニコ動画のウィークリーランキングを初期値とし、参照リンクに基づいて抽出された影響伝播関係に基づいてリランキングした。本プロトタイプシステムはPerlとSQLiteを利用して実装し、Windows XP上で動作する。本実験では初期ランキング及びリランキングそれぞれにおける上位10件の動画に対し、被験者にそれぞれのコンテンツの質を10段階で評価してもらった。図4に実験結果を示す。図4の左側の表は初期ランキングの結果とそれぞれの動画に対するユーザによる評価値を示している。同様に右側の表はリランキングの結果とそれぞれの動画に対する評価値を示している。

初期ランキング 平均点4.1

順位	動画	評価値
1	sm8593268	6
2	sm8628149	10
3	sm8781280	6
4	sm8778113	2
5	sm8754291	4
6	sm8312467	1
7	sm8778883	3
8	sm8794366	6
9	sm664723	1
10	sm8790894	2

リランキング 平均点6.3

順位	動画	評価値
1	sm8628149	10
2	nm3601701	6
3	sm2077177	8
4	sm7284045	6
5	nm4595473	5
6	sm9269006	5
7	sm9041517	7
8	sm9015022	3
9	sm9194117	10
10	sm9018966	3

図4 実験結果

今回の実験では、初期ランキングとリランキングの結果に共通して出現する動画が複数存在している。それぞれの上位10件に対するユーザの評価値は、影響伝播関係に基づくリランキングを行うことにより平均点が向上した。また、被験者の感想として、「自分の知らないコンテンツ群を発見できた」というものがあつた。このことから、本実験に於いて提案手法は有効であったと考えることができる。

しかし、今回の実験によって得られたランキング結果は、同一のコンテンツから影響を受けた複数のコンテンツが上位にランキングされていた。この理由として、初期ランキングにおいて非常に高く評価されていたコンテンツから影響を受けたコンテンツが複数存在し、それらがリランキング結果に於いて上位にランク付けされてしまったことが考えられる。本手法の目的は、ユーザの意識していないが興味を持つと思われるコンテンツを提示することであるため、手法の改善が必要である。

5. まとめ

本論文では、投稿型サイトのランキングに登場するコンテンツの質を高めるため、影響伝播関係に基づいてリランキングする手法を提案した。また、ニコニコ動画を利用して単純な実験を行った結果を示した。

今後の進展として、ユーザが定期的にWebを閲覧していることに注目し、ユーザのアクセスを分析することによって、ランキングの個別化を行うことを考えている。ユーザの嗜好を分析し、嗜好に合う可能性のあるコンテンツ群の中から、ユーザがまだアクセスしたことのないコンテンツを推薦することによって、ユーザの興味の活性化を行う。