

# ウェブ上の文書はどう面白いのか？

古川 忠延<sup>1)\*</sup> 松尾 豊<sup>2)3)</sup> 大向 一輝<sup>4)5)</sup> 内山 幸樹<sup>6)</sup> 石塚 満<sup>1)</sup>

1) 東京大学 大学院情報理工学系研究科 2) 産業技術総合研究所 3) スタンフォード大学  
4) 国立情報学研究所 5) 総合研究大学院大学 6) 株式会社ホットリンク  
\*) furukawa@mi.ci.i.u-tokyo.ac.jp, <http://www.miv.t.u-tokyo.ac.jp/mem/tfuru/>

**Abstract:** ウェブコンテンツの増加に対応すべく、これまで多くのコンテンツ推薦手法が提案されてきたが、実際に効果をあげているものは稀である。推薦手法における問題点は、ユーザがどのような「面白さ」を感じているのかを把握できていないからではないかと筆者は考えている。そこで本研究では、ウェブ上の文書について、ユーザにとってどのようなコンテンツがなぜ面白いと感じるのか、その要因を探っていきたい。

## 1. はじめに

近年、mixi<sup>1</sup>やGREE<sup>2</sup>などのSNSや、YouTube<sup>3</sup>のような動画共有サービスの流行に見られるように、ウェブ上のコンテンツは増加・多様化の一途を辿っている。ウェブを閲覧するユーザにとって魅力的なコンテンツの増加は好ましい事態であるが、反面、それを「面白い」と感じるユーザによって発見されないままになってしまいうコンテンツが増加することも危惧される。

こうした状況への解決案として、これまで多くのコンテンツ推薦手法が提案されてきた。多くは内容の類似性や統計的な行動傾向、世間の流行・評判などに基づいて情報を推薦しようというものであり、それらを導入した多くのサービスやアプリケーションが開発されてきた。しかし、対象とするコンテンツやユーザによって精度には難があり、現在効果をあげているものは稀であるというのが実情である。

推薦手法におけるこのような問題の原因は、そもそも「面白い」とはどういうことであるのかを把握できていないからではないかと筆者は考えている。一言に「面白さ」と表現しても、ブログであれば思わずコメントをしたくなる面白さや自分でもその話題で記事として書きたくなる面白さ、動画であれば何度も鑑賞したくなる面白さなど、様々な側面があり、その結果としてユーザのとる行動も異なる。面白いと思われるコンテンツを推薦する際、それがどんな面白さであるのかを考慮する必要があるのではないだろうか。

本研究では、コンテンツとしてウェブ上の文書に対象としたとき、ユーザに対して様々な面白さを持った未知のコンテンツを推薦するためにはどうすればよいのかを考えたい。そのために、ユーザはどんな情報がどのように面白いと感じているのかを分析し、要因が何であるのかを探っていく、

## 2. ウェブサイト閲覧行動

人々がウェブサイトを閲覧する目的としては、主に以下の3つが考えられる。

- 娯楽  
暇潰しとしての話題探し、友人との交流。
- 調査

---

<sup>1</sup> <http://mixi.jp/>

<sup>2</sup> <http://gree.jp/>

<sup>3</sup> <http://www.youtube.com/>

仕事・趣味のための知識獲得や、天気予報や電車乗り換えに関する調べ物。

- オンラインショッピング  
書籍や航空券などの購入。

本研究ではこのうち、「娯楽」に焦点を当て、娯楽目的でウェブを閲覧するユーザに対してコンテンツを推薦するためには、どんな分析をすればよいのかを考えていきたい。「調査」に関しては必要な文書が如何に整理されているのかが重要であるし、「オンラインショッピング」においては扱う対象がウェブコンテンツではないため、ここでは扱わないこととする。（「調査」の際に全く新しい情報を推薦することのニーズも考えられるが。）

一方で、ウェブ上でユーザが閲覧する範囲はどうだろうか。ウェブのサーフィンには、以下の方法が考えられる。

- ブックマーク, RSS リーダ  
ブラウザ機能や単独でのアプリケーションのもの、もしくはオンラインのブックマーク/RSS 登録サービス。
- 検索  
Google, Yahoo のような検索サイトまたは, YouTube や mixi などのサービス内での検索。
- 上記で辿り着くサイトからのリンク

ユーザの能動的な行動で行き着くサイトはこの程度であり、数ホップのリンクを辿ったところで、その範囲には限界がある。この問題を解決するために、これまで多くの推薦アルゴリズムが考案されてきた。

### 3. 既存の推薦手法

推薦システムにおいて用いられる主な手法は以下の2つである。

- 協調フィルタリング  
二人のユーザ A, B の行動履歴が似ているとき, A が興味を示したアイテム(商品やウェブサイトなど)のうち, B がまだ興味を示していないものを, B に推薦する手法。ユーザの興味を, 各アイテムを軸とするベクトル(次元数=アイテム数, 大きさ=各アイテムに対する興味)で表し, ユーザ間の類似度はベクトル同士の内積やコサイン相関を用いて計算する。履歴が少ないとベクトルがスパースになり, 有効な類似度が計算できない問題点がある。
- 内容に基づくフィルタリング  
ユーザがあるアイテム x に興味を持っているとき, x と類似した内容の別のアイテム y を推薦する手法。例えばアイテムが書籍であった場合, 著者が同じである, タイトルに同じ単語が含まれている, などのような書籍が推薦対象となる。全く異なるジャンルのものを提示できないのが欠点である。

両者の欠点を補うよう, 行動履歴が少ないうちは内容に基づくフィルタリングをおこない, ある程度の履歴が溜まると協調フィルタリングに移行する, ハイブリッド型の手法が用いられることも少なくない。その他, オンラインストアの Amazon.co.jp<sup>4</sup>(対象はウェブコンテンツではないが)では, ユーザがこれまで購入・評価した商品の履歴と, 商品間の同時購入されやすさの関係から, ユーザに対して推薦リストが提示する手法を提案している[1]。しかし, 現状で効果を挙げている推薦システムは稀である。

---

<sup>4</sup> <http://www.amazon.co.jp/>

#### 4. 考えていること

コンテンツに対してユーザの持つ印象は様々である。例えば、Yahoo! JAPAN<sup>5</sup>のニュース記事には「みんなの感想」という機能(図1)が付加されており、

考えさせられる・役に立つ・興味深い・誰かに教えたい・びっくりした

の5項目について、ユーザが記事に対する感想を5段階で評価、投票できるようになっている。面白さに関して5種類のベクトルを考慮できると言える。同サイトのスポーツ記事においても

すごい・へえ～・やった！・がっかり・腹立つ～

の5項目のいずれかから、記事に対する印象を投票できる機能が用意されている(図2)。ユーザはアンカーテキストとなっている記事タイトルを読み、それで興味を持ってニュース記事にアクセスしているわけであるが、結果として受ける「面白さ」はこのように様々である。また、ブログサイトの閲覧を考えた場合に、記事を読んだユーザがコメントを残す場合や、自身のブログ上でも同じ話題で書く、もしくは何もしないなど、複数のアクションをとるのも、面白さの違いが関係していることが考えられる。

このような面白さの違いは、既存手法では考慮されていないのではないだろうか。Amazon.co.jpの推薦リストにおいても、アンケート[2]に見られるように、一度だけ気まぐれで購入した商品の影響が大き過ぎるという問題が指摘されている。過去の事例から、ユーザの買いそうな商品や興味を持ちそうな話題を分析するのであれば、学習データにおいて、なぜ商品を購入したのか／話題にどのような面白さを感じたのかを考慮し、その要因を分析する必要があるのではないだろうか。



図1. みんなの感想.



図2. Yahoo! スポーツにおける感想投票.

#### 5. おわりに

本稿では、ユーザがコンテンツに感じる「面白さ」に様々なベクトルがあることを述べ、それが考慮されていないことを既存の推薦手法の問題点として指摘した。今後はウェブ上の文書(主にブログを考えている)に対してどのような面白さがあるのか仮説を立て、どのような要因があるのかを解析していく予定である。解析には[3]で松村らが2ちゃんねるが盛り上がる仕組みを分析する際に用いた、共分散構造分析が参考にできるのではないかと考えている。

<sup>5</sup> <http://www.yahoo.co.jp/>

## 参考文献

- [1] G., B. Smith and J. York: Amazon.com Recommendations: Item-to-Item Collaborative Filtering, IEEE Internet Computing, Vol. 07, No. 1, pp. 76--80, 2003.
- [2] コトノハ - ○×ソーシャル: コトノハ - Amazon の New for You はちゃんと自分の趣味を反映している, <http://kotonoha.cc/no/4369>
- [3] 松村真宏, 三浦麻子, 柴内康文, 大澤幸生, 石塚満: 2ちゃんねるが盛り上がるダイナミズム, 情報処理学会 45 卷 3 号, pp. 1053--1061, 2004.