

小特集

ソーシャル ブックマークは 進化し続ける!

～ソーシャルブックマーク
研究会の議論から

2



SBMデータを用いた Webコンテンツ推薦

佐々木 祥 ▶ 東京工業大学大学院理工学研究科集積システム専攻

宮田 高道 ▶ 東京工業大学大学院理工学研究科集積システム専攻

情報推薦とは

近年、インターネット上では Consumer Generated Media (CGM) などと呼ばれる、一般ユーザからのコンテンツ生成および、コンテンツ配信が盛んである。これに対して、テレビの視聴率が低下するなど、大衆向けのコンテンツの需要は相対的に低くなってきている。これは、インターネットを利用した情報収集が広く浸透し、その結果として、ユーザの嗜好に多様性が生まれたことを示す1つの事実であると言える。

一方で、ユーザによる情報生成は情報爆発と表現されるほどの大量な情報を生んでおり、ユーザが嗜好するであろう情報は発見されることなく埋没してしまっている。このような情報を救うためには、前述のとおり、ユーザの嗜好は多様化しているため、従来のような視聴・閲覧回数などのランキングによる情報提示はふさわしくなく、各ユーザの嗜好に合わせた情報提示が求められている。

そこで、各ユーザに対する適切な情報提供を可能とする技術として、各種情報推薦技術が注目されている。情報推薦とは、ユーザの嗜好情報(プロフィール)を収集し、それに合致する情報を大量の情報源から機械的に発見し、提示することを指す。

情報推薦においては、何よりもユーザの嗜好情報が重要となるが、これをユーザから収集するのは非常に困難である。たとえば、ユーザからアンケートをとり、嗜好するコンテンツを聞くなどの明示的な収集があるが、ユーザに労力を強いるため、データ数を集めにくい。一方、ブラウザにおける閲覧時間を記録できるようにして、それから嗜好を判定するような暗示的な収集もあるが、嗜好を正確に判定することは困難である。

このように、情報推薦は必要とされながら、いかにしてユーザの嗜好情報を正しく得るか、という課題に対する決定的な解決策は存在しなかったといえる。

ソーシャルブックマークと情報推薦

しかしながら、ソーシャルブックマークの登場によって状況は一変する。ソーシャルブックマークでは、ユーザの嗜好コンテンツ収集データが、ユーザの手によって生成されていく。これはまさにユーザの嗜好情報そのものである。ソーシャルブックマークの登場は、特にデータマイニング系の研究者に大きな影響を与えたと言えるだろう。現に、多くの研究がここ数年でなされている。情報推薦においても同様に、ソーシャルブックマークのデータを用いた研究がなされてきた。

これらの研究の主眼は、タグの情報をいかに用いるかにある。

タグは自由記述のキーワードであり、コンテンツ1つ1つにユーザが能動的に付与する。これをコンテンツ1つに着目すると、多数のユーザによるメタデータ付与と見ることができ、これによってボトムアップにコンテンツの特徴付けができるのではないかと考えることができる。このような考え方はフォークソノミーと呼ばれており、情報推薦を含め、多くの研究はこの視点に基づいて行われてきた。

たとえば丹羽ら¹⁾は、ユーザ全体におけるコンテンツに対するタグの付与傾向と、個々のユーザのタグの利用傾向とをマッチングさせるといって、フォークソノミーに基づいた推薦手法を提案している。

しかしながら、情報推薦に限って言えば、フォークソノミーとの相性はよくないと言わざるを得ない。それは、タグが自由記述であることに起因する。

情報推薦では前述の通り、ユーザの嗜好情報と合致する情報を提示することになるが、この嗜好情報をタグの名称そのものに置き換えてしまうと、合致する情報とは「当該ユーザと同様の記述ルールを用いて記述されたブックマーク」に限定されてしまう。すなわち、たとえまったく同じコンテンツを嗜好(ブックマークおよびタグ

tags	21 う_宇宙	45地理・歴史 46
10 (#`D`)ゴ`カ!!	2 う_歌	食 47ラーメン 48
12(`w`)`キ`ン	8 え_エッシャー	酒 50文化 51漫
22(`D`)`D`)ア`ニョ`コ	17 え_エミュレータ	画、アニメ 51美・
17(`w`)`ヨ`ク`ン	31 え_映画	デザイン 52音楽
36(σ`V`)`カ`ツ!!	37 え_絵	53映画 54作家・本
12(`d`)`ヒ`)`ヒ`)`ヒ`)`	51 え_英語	55()`ト` 56フ
72(`D`)`本`に`こ`の`デ`タ`で`推`薦`で`き`る`の`?	6 お_おまじ	
11(`D`)`	5 お_おまじ	
52(`D`)`カ`ン	172 `w`)`つ`あ`と`で`	4 study/english ましいおたくおっお
8(`A`)`イ`イ	4(`w`)`あ`れ`?	5 study/P2P ならおはきな おめでと
124 `V`)`イ`!!	6(`w`)`う`ん`?	2 study/pdf うございますおわびお
14(`D`)`ア`-	2(`w`)`あ`ら	10 study/SBM フランスざんす おロシ
36(`D`)`ア`-	39(`w`)`お`お	10 study/scale-free アお仕事るんるんお勉
1`do	13(`w`)`な`ー	8 study/scale-out 強 お受験 お年頃 お店
	80(`w`)`ふ`む`ふ`む	11 study/search お役人クオリティ お役
	5(`w`)`売`げ`し`く`同`意`	

図-1 SBM のタグの実データの一例

ング)したユーザ間であったとしても、タグの記述法が違うために情報を共有できないということになる。このような記法の違いはタグのゆらぎと呼ばれる。

この問題を解決するには、ユーザが統制された言語群のみをタグに用いるよう強制してゆらぎをなくすか、類似するタグを同一とみなしてゆらぎを吸収するか、いずれかの方法を取らなければならないだろう。

多くの研究は後者を行おうとするものであるが、タグの実データは図-1に一例を示すとおり千差万別であり、統制がいかに困難で一筋縄でいかないか、分かっていたただけだろうか。

ところで、そもそもこのような努力が必要となったのは、情報推薦という目的に対して、フォークソノミーという、いわば遠回りな手段を利用してしまったのが原因ではないだろうか。情報推薦のためには情報推薦に適した手段を取るべきではないだろうか。

そこで我々は、タグの表面上の表記でしかない、タグの名称は見ないことにし、ユーザが同一のタグを付与してきたコンテンツ集合に着目することにした。フォークソノミーは、コンテンツにどのようなタグが付与されてきたかに着目し、コンテンツの絶対的な意味を抽出しようとするものであるが、我々の手法では、その絶対的な意味を捨てて、ユーザがタグを用いてひとまとめたコンテンツの関係性のみを取り出し、その相対的な関係を用いて推薦を行う。いわば、アンチフォークソノミー的発想による推薦である。アンチフォークソノミーという考えは、文字情報を利用していないので、特殊な記述ルールを用いているユーザに対しても有効であり、また、他言語への拡張も容易である。

また、我々の研究ではこれに加えて、ブックマーク間の類似度合を、統計的信頼性に基づいて算出する手法を提案しているが、詳細は文献2)を参照されたい。

ソーシャルブックマークと研究の今後

研究者の立場から言えば、ソーシャルブックマークは強力な嗜好情報収集方法であるといえる。しかしながら一方で、ソーシャルブックマークの利用率は非常に低いのが現状である。そのため、十分に多様なユーザの嗜好が収集できているとは言えず、それから作り出される情報推薦も限られた機能しか発揮できていないだろう。

この現状を打破するための1つの方法として、情報推薦という付加価値をソーシャルブックマークに与えることがあるのではないかと考えている。情報推薦とソーシャルブックマークそれぞれがサービスの両輪として働き、ユーザによるユーザのための情報提供の場となることが、私の考えるソーシャルブックマークの発展形である。

現在サービス提供中のソーシャルブックマークは、まだまだ発展する余地を残していると感じる。今後のソーシャルブックマークの発展をサービスサイドにも期待するが、新たな情報推薦の可能性を示すことで、研究者サイドからも発展の可能性を示すことができれば幸いである。

参考文献

- 1) Niwa, S., Doi, T. and Honiden, S.: Web Page Recommender System based on Folksonomy Mining, Proceedings of the Third International Conference on Information Technology: New Generations (ITNG'06) (2006).
- 2) 佐々木 祥, 宮田 高道, 稲積 泰宏, 小林 亜樹, 酒井 善則: Social Bookmark におけるコンテンツクラスタ間の類似度を用いた web コンテンツ推薦システム, 情報処理学会論文誌: データベース, Vol.48, No.SIG20, pp.14-27 (Dec. 2007).

(平成 20 年 10 月 28 日受付)

▶▶▶ 佐々木 祥 (学生会員)

brnw@net.ss.titech.ac.jp

2005 年神奈川大学工学部電気電子情報工学科卒業。2007 年東京工業大学大学院理工学研究科集積システム専攻修士課程修了。現在、同大学院理工学研究科博士後期課程在籍中。Web コンテンツ推薦などの研究に従事。

▶▶▶ 宮田 高道 (正会員)

miyata@net.ss.titech.ac.jp

2003 年富山大学大学院理工学研究科博士前期課程修了。2006 年東京工業大学大学院理工学研究科博士後期課程修了。同年、同大学院理工学研究科助手。現在、同助教。画像符号化、画像処理、Web コンテンツ推薦などの研究に従事。博士(工学)。