

# ソーシャルブックマークと Wikipedia カテゴリを用いた ブックマーク推薦システム

吉田 拓実<sup>†</sup> 井上 潮<sup>‡</sup>

<sup>†</sup> <sup>‡</sup> 東京電機大学大学院 工学研究科 〒120-8551 東京都足立区千住旭町5番

E-mail: <sup>†</sup> 12kmc36@ms.dendai.ac.jp <sup>‡</sup> inoue@ms.dendai.ac.jp

**あらまし** ユーザが Web ページを検索する際、検索クエリの試行錯誤が必要となる。この問題の解決のためにユーザの興味・関心に合った Web ページを提示する様々な推薦システムが研究、開発されてきた。そのなかでも、Web 上でブックマークを共有するソーシャルブックマークのタグ情報を用いてブックマークを推薦する研究が盛んに行われてきた。しかし、これらの推薦システムでは、2つの問題が存在した。1つめは、ベクトル空間や確率モデルに基づき類似するユーザの検索を行うため、処理速度が遅いという問題である。2つめは、ソーシャルブックマークのタグの表記の揺れ問題である。本稿では、ソーシャルブックマークのタグ情報と Wikipedia のカテゴリ情報を用いて上記の2つの問題に対応した推薦手法を提案し、提案手法を実装したシステムと評価結果を報告する。

**キーワード** 情報推薦, ソーシャルブックマーク, Wikipedia

## 1. はじめに

ブログや SNS, Web 上の情報をまとめるキュレーションサービスなどの出現により、ユーザが自由に情報を発信することができるサービスは増加しつつある。そのため、インターネット上で入手できる情報量は増加傾向にあり「情報爆発」[1]が問題になっている。ユーザが何か情報を得ようとする際に Web ページ検索エンジンを使用するが、「情報爆発」では情報量が多く情報の絞り込みのため検索クエリの試行錯誤が必要となり効率が悪くなっている。

検索クエリの試行錯誤の解消に関しては、閲覧履歴から関連情報の検索・自動提示を行う研究[2]や、ソーシャルブックマークを用いて情報を提示する研究[3]などがある。ソーシャルブックマークとは、各ユーザが Web ページをブックマークとして追加し、他のユーザと共有することができる Web サービスの事である。また、各ブックマークには、ユーザが自由に選択したタグと呼ばれるキーワードを付加することができる。ソーシャルブックマークを用いた推薦の研究では、ユーザに推薦するブックマークは、タグ付けの傾向に基づくことが多い。

しかし、ソーシャルブックマークを用いた推薦では2つの問題が存在する。1つめは、処理速度の問題である。既存の推薦システムのほとんどは、ベクトル空間や確率モデルに基づいて対象ユーザと類似するユーザを検索する。これらのモデルでは、ユーザとブックマーク間の類似度を計算するために膨大な数のタグを分析する必要がある。そのため、処理速度が遅いという問題がある。2つめは、ソーシャルブックマークタグの表記の揺れ問題である。これは、同じタグが異なる意味を示す、または異なるタグが同じ意味を示すとい

う問題である。そのため、同じような嗜好を持つユーザだとしてもタグの表記揺れ問題のため、推薦の対象外となってしまうことがある。

我々はこれらの問題を解決する、ソーシャルブックマークと Wikipedia を用いた情報推薦手法について提案した[4]。この手法の特徴は、各ユーザの重要なタグのみを分析するため高速化が期待できることと、Wikipedia を用いてソーシャルブックマークタグの表記の揺れ問題に対処したことである。

本稿では、ソーシャルブックマークと Wikipedia の調査によって提案手法の論理的な根拠を与える。さらに、十分な量の実験データを用いて既存手法との定量的な比較を行ったことである。

以下、本稿の構成を示す。第2章では関連研究としてソーシャルブックマークを用いた情報推薦の研究と Wikipedia を用いた単語の拡張に関する研究について述べる。第3章では、本研究の前提システムであるはてなブックマークと Wikipedia カテゴリに関して調査結果を述べる。第4章では、本研究のアプローチを述べ、第5章では、ブックマーク推薦手法を述べる。第6章では、実験と評価について、第7章ではまとめを述べる。

## 2. 関連研究

### 2.1. ソーシャルブックマークを用いた情報推薦

百田ら[5]はソーシャルブックマークを用いて、Cos類似度というユーザ間の類似度を求める手法を用い、類似度が高いユーザが所持しているブックマークを推薦する手法を提案している。また、 $\alpha$ ブックマーカと呼ばれる新しい情報をすぐ入手している利用者に着目しそのようなユーザが所持しているブックマークを推薦する手法を評価した。その結果、 $\alpha$ ブックマーカを利用することでトピックを限定した時の情報発見について良い結果を得た。

石橋ら[6]は、ソーシャルブックマークからタグのグループを作成し、そのグループによってユーザの嗜好推定を行い、類似ユーザを発見し Web ページを推薦する手法を提案している。ユーザが新規にブックマークを登録する際のタグと URL の関係の推定をベイズ推定によって行う。しかし、推薦される URL は類似度が近いユーザ1名のブックマークしている URL のリストであるため、ユーザの嗜好と異なるページが推薦される場合がある。

### 2.2. Wikipedia を用いた単語の拡張

中山ら[7]は、Wikipedia を用いてシソーラス(単語の上位/下位関係、部分/全体関係、同義関係、類義関係などによって単語を分類した辞書)を構築する手法を提案した。Wikipedia の記事同士がハイパーリングで互いに参照していることを活かし、同義語、多義語を抽出することによってシソーラス辞書の構築を行う手法を提案した。その結果、関連度の高い語を抽出しユーザ評価と一致するシソーラス辞書の構築に成功した。

この研究により Wikipedia によって、ある単語に関して関連する語句を取得することは有用であると考えられる。

## 3. 前提システム

### 3.1. はてなブックマーク

本研究で用いるはてなブックマーク[8]について述べる。はてなブックマークは、日本国内で一番人気のあるソーシャルブックマークサービスであり、2013年12月17日時点で、179,540,202ブックマーク(64,968,666種類の Web ページを含む)を有する。

図1はソーシャルブックマークサービス上でのユーザ、タグ、ブックマークの関係である。ユーザは Web ページをブックマークする際に複数のタグを使用することができ、また同一のタグを他の Web ページをブックマークする際に再利用することができる。

しかし、タグを自由につけることができるためタグの表記の揺れ問題が生じる。タグの表記の揺れ問題とは、「同じタグが異なる意味を示す」、または「異なる

タグが同じ意味を示す」という問題である。例えば図2のように USER A が“Java”のタグを使用していて USER B が“Java”，USER C が“プログラミング言語”，“オラクル”，“サン・マイクロシステムズ”とタグ付けをしたと仮定する。また，USER A～C は全員“Java”に興味があると仮定する。USER A と USER B は同一のタグを使用しているため互いにブックマークを参照することは可能である。しかし，USER C は“Java”というタグを付けていないため“Java”に興味があるのにかかわらず USER A のブックマークを参照することができなくなる。これがタグの表記の揺れ問題である。

次に、はてなブックマークの全体のタグの使用傾向について述べる。表1は、本研究で収集したはてなブックマークタグの上位10件である。上位10件に含まれるタグの内“ネタ”や“まとめ”は記事の性質を示し，“あとで読む”はユーザのタスクを示している。このようなタグを除くと“iPhone”や“javascript”のような技術用語がタグとして多く使われることがわかる。

さらに、ユーザごとのタグの使用傾向もはてなブックマーク全体と同じ傾向を示すかを調査した。図3に無作為に取得したユーザ2名のタグのパレートチャートを示す。ユーザ A は31種類のタグを使用し、タグの累積使用回数は48回である。そのうち使用回数上位5種のタグが2回以上使用され、残りの26種のタグは1回のみ使用したタグであった。また、使用しているタグも“ios 7”や“iPhone”など最新の技術用語が多かった。ユーザ B は95種類のタグを使用し、タグの累積使用回数は205回であった。そのうち使用回数上位37種のタグが2回以上使用され残りの58種は1回しか使用されていなかった。ユーザ B もユーザ A と同じく“プログラミング”や“Java”のような技術用語をタグとして使用していることが分かった。

以上の結果からはてなブックマークのユーザは技術的用語をタグとして多く使用することが分かる。また、使用したタグの内、複数回使用されているタグは少ないため使用回数の上位のタグを使用すれば処理速度の高速化が見込めることが分かる。

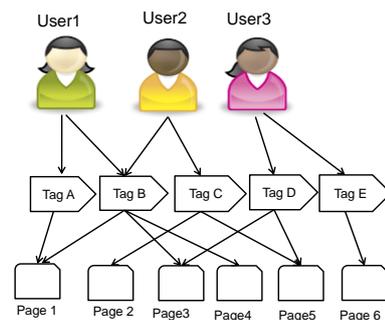


図1 ソーシャルブックマークでのタグ付け

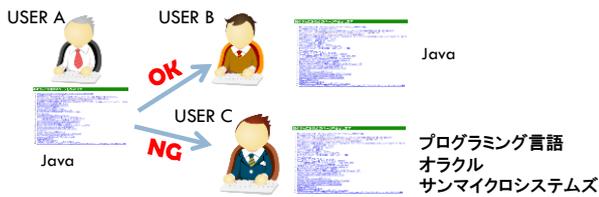
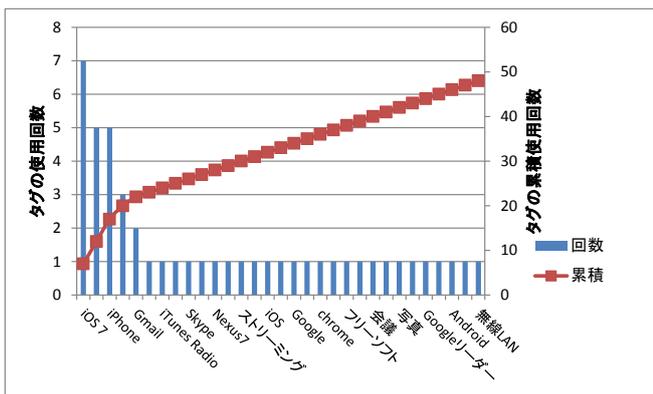


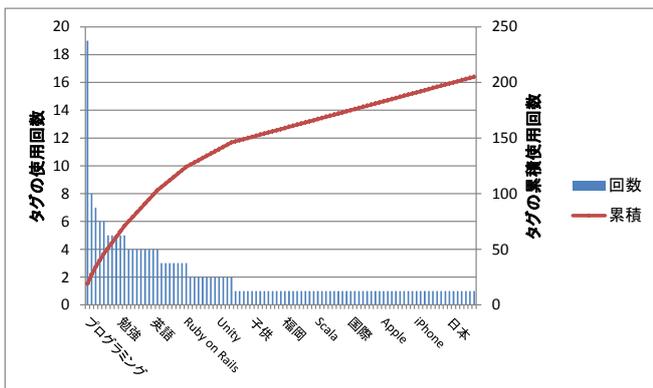
図 2 タグの表記の揺れ問題

表 1 はてなブックマークにおける上位 10 件のタグ

タグ	使用回数(回)	使用人数(人)
ネタ	41832	5177
あとで読む	36550	4242
まとめ	32367	5018
社会	24850	3115
2ch	23723	2572
web サービス	22924	5162
iphone	22657	5036
web	19316	3710
javascript	19138	3871
web デザイン	18676	3562



(a) ユーザ A の使用傾向



(b) ユーザ B の使用傾向

図 3 あるユーザのタグに関するパレートチャート

### 3.2. Wikipedia カテゴリ

Wikipedia のカテゴリ [9] は、分類とキーワードを示すものである。Wikipedia カテゴリはいくつかのサブカテゴリを持ち親子関係が存在する。この分類機能により、カテゴリは階層構造を持つことになる。

図 4 は“Java”を例としたカテゴリ構造の一部である。この図において、各長方形はカテゴリを表し、各矢印はカテゴリ間の親/子関係を示している。例えば、“Java”は“オブジェクト指向プログラミング言語”、“プログラミング言語”、“サン・マイクロシステムズ”、“オラクル”の子カテゴリとなっている。同様に“サン・マイクロシステムズ”は“コンピュータ会社”の子カテゴリである。図のように、“Java”と“サン・マイクロシステムズ”はカテゴリ間に強い関係を持っている。一方で“Java”と“コンピュータ会社”のように直接的なカテゴリ構造にない場合には関係性が薄くなる。

次に日本語版 Wikipedia に存在する用語数について述べる。2012 年 6 月 6 日のダンプデータでは Wikipedia には 810,499 記事、103,886 カテゴリが存在する。これは日本語 Wordnet の 57,238 語より多い結果となっている。また、日本語 Wordnet では、ソーシャルブックマークのタグによく使用されている専門用語や、流行語のほとんどを所持していない。一方 Wikipedia では、だれでも編集ができ、多人数で作成されるため、最新の流行語や専門用語が含まれている。

以上の結果から、Wikipedia を使用することによってソーシャルブックマークのタグデータの関連情報を取得することが可能なため、タグの表記の揺れ問題の改善が期待できる。

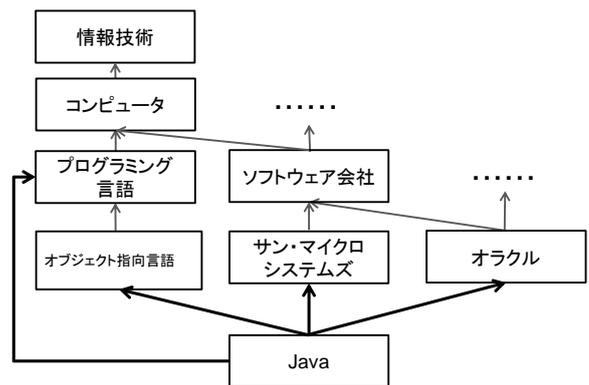


図 4 Wikipedia カテゴリの階層構造

## 4. 本研究のアプローチ

本研究では、検索時のクエリの試行錯誤の煩わしさの解消のためにソーシャルブックマークを用いたブックマーク推薦とシステムを開発する。

本研究において、推薦をしてもらいたいと依頼しているユーザをクライアントユーザ  $X$  とする。クライアントユーザ  $X$  が所持しているタグデータを用いてブックマークの推薦を行うことで、情報検索時の効率の悪さを解決する。ソーシャルブックマークを使用するうえで、生じるタグの種類数が多いという問題点に対しては、3.1.で述べたようにクライアントユーザ  $X$  が使用しているタグの中で重要であるのは使用回数が多い上位何件かのタグであることを利用し、クライアントユーザ  $X$  がよく使用している上位  $k$  件のタグに着目することで解決を図る。また、タグの表記の揺れ問題に関しては、3.2.で述べたように Wikipedia カテゴリにはソーシャルブックマークで使用されているタグの関連情報を取得できるため、Wikipedia カテゴリ情報を用い、タグの関連語を取得することで関連語を含んだブックマークを推薦することで解決する。

## 5. ブックマーク推薦手法

ブックマーク推薦システムのブロック図を図5に示す。本推薦システムは、類似ユーザ抽出部(TK法)、ブックマーク抽出・提示部からなる。

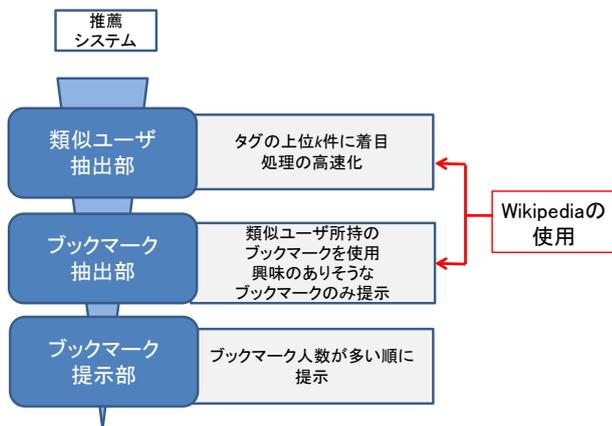


図5 ブックマーク推薦システムブロック図

### 5.1. 原則

ブックマークの推薦において以下の原則に基づいた処理を行う。

- クライアントユーザの使用回数が多い上位  $k$  個のタグに着目する。
- Wikipedia のカテゴリで直接階層にあるタグに関連する用語を新しいタグとして追加する。前者は、類似ユーザやブックマークを検索するタグ

の数を減少させる。後者はタグの表記の揺れ問題を緩和することができる。

### 5.2. TK法

本研究の類似ユーザ抽出における提案手法であるTK法を図6に示す。3.1.で述べたように、ユーザの使用しているタグのほとんどは使用回数が1回である。そのため、ユーザの嗜好を反映しているとはいえない。一方使用回数が多いタグに関しては、ユーザが多く興味関心を持っていると考えられるため、ユーザの嗜好を反映しているといえる。そのため、TK法では、クライアントユーザ  $X$  が使用しているタグのうち、使用回数が多い上位  $k$  件のタグに着目する。はじめに、上位  $k$  件のタグをタグセット ( $T$ ) として取得する。その後、嗜好が類似しているユーザを抽出するために  $T$  をすべてタグ付しているユーザを類似ユーザ ( $U$ ) として抽出する。このように行うことでユーザが深く興味を持っているタグに関して嗜好が一致するユーザを抽出することができる。

### 5.3. ブックマークの抽出・提示処理

ブックマークの抽出・提示の処理方法について図7に示す。ブックマークの抽出・提示ではユーザ  $X$  に推薦するブックマークセット ( $B$ ) を作成する。類似ユーザが所持しているブックマークは、クライアントユーザ  $X$  にとって役に立つという考えの元に、ブックマークセット ( $B$ ) を、5.2.で抽出した類似ユーザ ( $U$ ) が所持しているブックマークの中からタグセット ( $T$ ) のタグが付けられているブックマークを抽出する。これにより、クライアントユーザ  $X$  に類似しているユーザが所持しているブックマークの中からクライアントユーザ  $X$  にとって興味のあるブックマークのみが取得できる。

その後、ブックマーク人数が多いブックマークほど役に立つという考えのもとに、ブックマーク人数順にソートしクライアントユーザ  $X$  に提示する。

### 5.4. Wikipedia法

Wikipedia法は、TK法とブックマークの抽出処理においてWikipediaのカテゴリ情報を使用する手法である。TK法において、類似ユーザ ( $U$ ) が閾値より少ない場合、Wikipediaカテゴリを用いて上位1件のタグの関連キーワードとして取得する。そして、関連キーワードのいずれかと、上位1件を除いた残りの  $k-1$  件のタグすべてをつけているユーザを類似ユーザ ( $U$ ) として取得する。

ブックマークの抽出処理においては、Wikipediaカテゴリを用いてユーザの上位  $m$  件のタグの関連キーワードを取得しタグセット ( $T$ ) を拡張する ( $ET$ )。そして、ユーザ ( $U$ ) の所持しているブックマークの内、拡張タグ ( $ET$ ) のタグが付けられているブックマークを抽出し、多い順に提示する。Wikipediaカテゴリを用いた、

タグセットの拡張例を図 8 に示す。

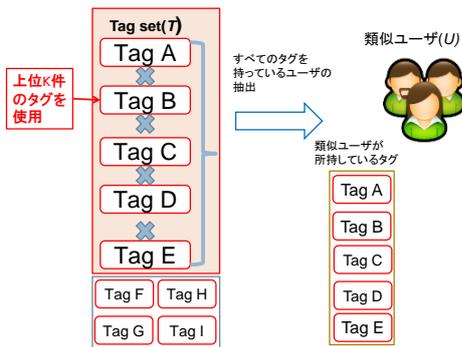


図 6 TK 法

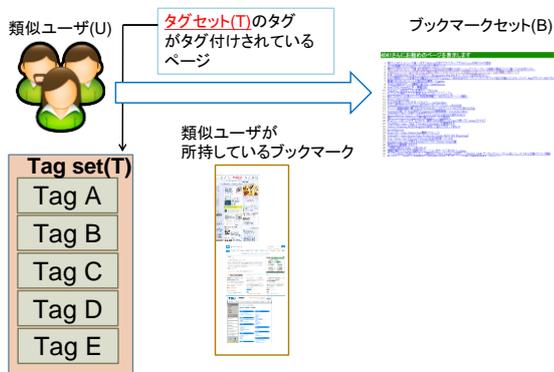


図 7 ブックマーク抽出・提示処理

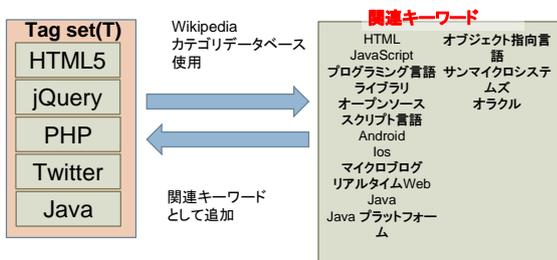


図 8 タグセットの拡張例

## 6. 評価

### 6.1. 実験システム

評価を行うため、実験システムを構築した。実験システムの構成を図 9 に示す。本システムはソーシャルブックマークのデータを収集するクローラとブックマークの推薦・表示・評価を行うブックマーク推薦部、評価を行う評価部からなる。本実験では、国内でよく用いられているはてなブックマークのデータを用いた。使用したデータは、クローラを用いて収集し MySQL データベースに格納したものを利用した。また、Wikipedia のカテゴリ情報は提供されている 2012 年 5 月 4 日版のダンプファイルを用いた。

データの収集に関しては、ソーシャルブックマーク

のデータを収集するクローラを構築した。クローラは、はてなブックエントリー情報取得 API を使用し、ある Web ページをブックマークしているユーザの情報とタグ情報を取得する。また、はてなブックマークページのユーザ個人の RSS の解析によってユーザのブックマークしているページのタイトル、URL、タグを取得する。

実験データとして 2012 年 5 月 19 日から 2013 年 11 月 19 日まで収集したデータを用いた。表 2 に収集したデータを示す。

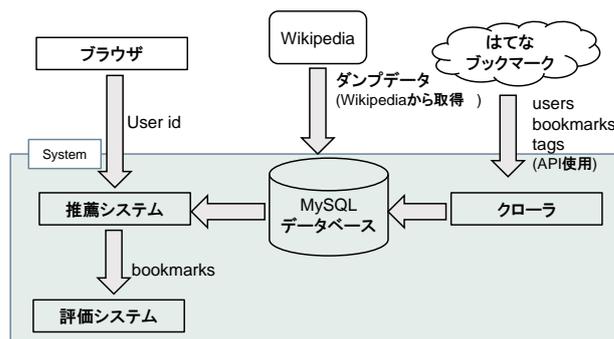


図 9 実験システム

表 2 実験データ

タグをつけているユーザ数(人)	19,659
総ブックマーク数(重複ふくむ)	2,400,355
総タグ数(重複ふくむ)	2,532,753

### 6.2. ブックマーク推薦評価

ブックマーク推薦評価では 2 つの評価を行った。

1 つ目は処理時間の比較である。一般的な手法である Cos 類似度を用いた類似ユーザ抽出手法と本研究の提案手法である TK 法の処理時間の比較を行った。

2 つ目はブックマーク数の比較である。ユーザ抽出法による違いと Wikipedia カテゴリの有用性を検証するために、推薦されたブックマーク数と適合ブックマーク数の評価を行った。

#### 6.2.1. Cos 類似度を用いたユーザ抽出法 (CS 法)

本研究での提案手法との比較のため Cos 類似度を用いて類似ユーザを抽出する手法 (CS 法) について述べる。Cos 類似度は、図 10 のようなベクトル  $a, b$  のなす角  $\theta$  を求める。三角関数の Cos と同様に、1 に近ければ類似しており、0 に近ければ類似性がないことになる。ここで  $\cos \theta$  は

$$\cos \theta = \frac{(a,b)}{|a||b|} \quad \dots(1)$$

( $a, b$ ): 内積  $|a||b|$ : 絶対値  
で表すことができる。なお本手法では、ベクトル  $a, b$

はユーザのすべてのタグの使用回数である。例えば、ユーザ A:(PHP2,Java3,C0) ユーザ B:(PHP5,Java2,C4) がこのようなタグ付け傾向を示すとき、これらのユーザの類似度を計算すると、

$$(a, b) = (2 \times 5 + 3 \times 2 + 0 \times 4) = 16$$

$$|a| = \sqrt{2^2 + 3^2} = \sqrt{13}$$

$$|b| = \sqrt{5^2 + 2^2 + 4^2} = \sqrt{45}$$

$$\text{Cos } \theta = \frac{16}{\sqrt{13} \times \sqrt{45}} = 0.66$$

となる。

CS 法では、図 11 に示す通り Cos 類似度を用いてクライアントユーザ X のタグ全件とその他のシステムを利用しているユーザのタグ全件を用いて類似度を計算し、上位 n 人のユーザを類似ユーザ(U)として抽出する。

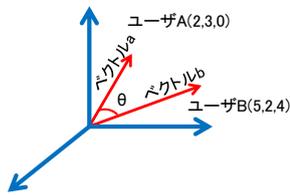


図 10 Cos 類似度のベクトル

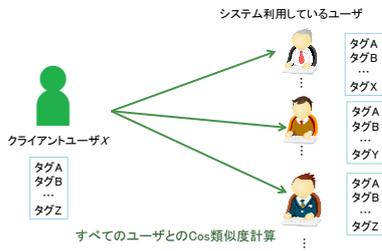


図 11 CS 法

### 6.2.2. 評価パラメータ

今回の評価にあたり、パラメータとして上位 k=5 件のタグを取得し、類似ユーザ x=10 人以下の場合上位 m=1 件を Wikipedia のカテゴリ情報を用いて拡張し、拡張したタグをつけているユーザを追加取得した。また、CS 法では上位 n=50 人のユーザを類似ユーザと使用するため TK 法の類似ユーザの上限も 50 人とした。サンプルユーザとしては 2013 年 10 月 9 日時点のはてなブックマークで多く用いられているタグであった”Web サービス”のタグを付けているユーザ 5 人を無作為に抽出した。

### 6.2.3. 処理時間比較

本手法の有用性を検証するため CS 法と TK 法を使用した場合の処理時間の評価を行った。処理時間は、サンプルユーザの ID を入力して実行をしてから、類似ユーザが表示されるまでの時間である。処理時間の平

均結果を表 3 に示す。

TK 法では類似ユーザを抽出するために平均 7.37 秒かかった。一方 CS では、16888.90 秒(4.69 時間)と TK 法の 2292 倍時間がかかる結果となった。図 12 に今回使用したサンプルユーザ 1 のタグ使用回数のパレートチャートを示す。サンプルユーザ 1 の使用していたタグの種類数は 127 である。その中で使用回数が 2 回以上であったタグの種類数は 49 種類であり残りの 78 種類のタグの使用回数は 1 回であった。よく使用するタグの種類が少なく、あまり使わないタグの数が多いという分布が見られる。すなわち、127 個のタグの内重要なタグは一部で他はあまり意味がない情報となっている。他のユーザを調査したところどのユーザも同じような傾向を示した。そのため、CS 法では、すべてのタグを対象としているため計算量が膨大となってしまいこのような結果が得られたと考えられる。

表 3 処理時間平均

ユーザ No.	TK 法 処理時間(秒)	CS 法 処理時間 (秒)
1	16.19	15911.82
2	2.91	17410.16
3	4.85	17131.68
4	6.42	17077.70
5	6.38	16913.14
平均	7.37	16888.90

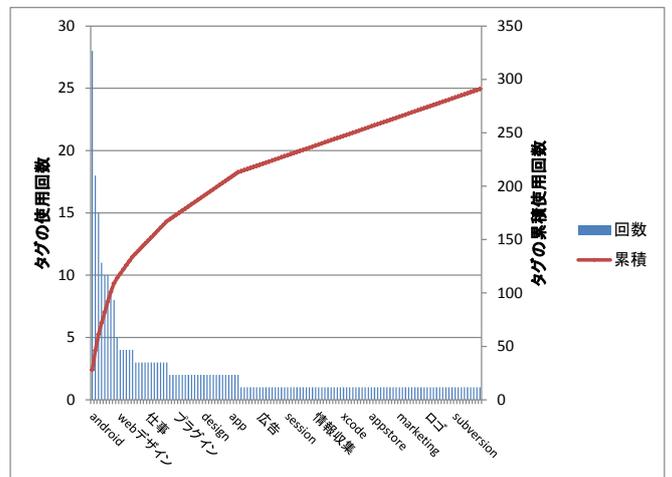


図 12 ユーザ No.1 のパレートチャート

### 6.2.4. ブックマーク数の比較

類似ユーザ抽出法による違いと Wikipedia カテゴリの有用性を検証するために、CS 法、TK 法で抽出した類似ユーザから適合したブックマークが推薦できたかどうか評価を行った。また、評価を行う際に Wikipedia カテゴリを用いることでどれだけブックマーク数が増加するかの評価を行った。抽出されたブックマークが適合しているかどうかは以下の基準で行った。

- I. 抽出したブックマークのタイトルにユーザの上位  $k(=5)$  件のタグが含まれる.
- II. 抽出したブックマークの上位  $k(=5)$  件のタグにユーザの上位  $k(=5)$  件のタグが含まれる.
- III. はてなブックマークのシステムが推薦しているはてなおすすめタグが含まれている.

結果を表 4 に示す. TK 法, CS 法, 共に Wikipedia カテゴリを用いることで適合率は下がるがブックマーク数は増加している. この結果より, Wikipedia カテゴリにより, タグの表記の揺れ問題が解決され, 推薦されたブックマーク数が増加したと考えられる. また, TK 法, CS 法の推薦されたブックマーク数を比較すると, TK 法のブックマーク数が多いため, TK 法の方が良い結果が出ていると言える. これは, CS 法ではタグの表記の揺れがあるため単純に類似度を計算してもユーザの嗜好が結果に適応されていないからである.

表 4 ブックマーク推薦結果

		TK 法		CS 法	
		Wikipedia カテゴリ情報			
		なし	あり	なし	あり
Web ページ数平均	適合	105.8	153.2	44.2	60.2
	不適合	3.2	12.8	1.2	2.2
適合率(%)		96.9	93.8	95.6	95.1

## 7. まとめ

本稿では, 初めにソーシャルブックマークのタグと Wikipedia に対して調査を行い, 調査結果に基づいた 2 つの手法について提案し有用性を検証した.

1 つめは, ユーザにブックマークを推薦する高速な手法である TK 法である. TK 法は, ソーシャルブックマークユーザが使用しているタグの内, ユーザにとって重要であるのは使用回数が上位のタグであるという調査結果を基にしている. そのため, ユーザが使用しているタグの内重要であると思われる上位  $k$  件のタグに着目することで高速化が可能になった. 一般的な手法である CS 法との処理時間の比較を行った結果, 提案手法(TK 法)の方が CS 法より 2292 倍処理時間が早いという結果を得た.

2 つ目は, Wikipedia を用いたソーシャルブックマークの表記の揺れ問題への対処である Wikipedia 法である. Wikipedia 法では, 推薦に使用したタグの関連キーワードを Wikipedia 内で直接的な親子関係にあるカテゴリを使用することで取得することで表記の揺れ問題への対処が可能になった. 結果として, TK 法, CS 法ともに ブックマーク数の増加が確認できた.

以上述べた TK 法と Wikipedia 法を組み合わせること

により, 多数の有用なブックマークを高速に推薦することが可能になった. 今後は, ユーザが推薦されたブックマークをチェックする負担を軽減するため, Web ページからの重要文の自動抽出方法について検討を進めていく.

## 参考文献

- [1] 柴山悦哉, 鳥澤健太郎, 田浦健次郎, 河野健二, “特集: 情報爆発時代におけるわくわくする IT の創出を目指して”, 情報処理, vol.49, no.8, 2008.
- [2] 吉田大我, 中村聡史, 田中克己, “ブラウジングと検索の融合: 閲覧履歴からの関連情報の検索・自動提示にもとづくウェブ閲覧”, 日本データベース学会論文誌, vol.7, No.1, pp.133-138, 2008.
- [3] Ziyu Guan, Can Wang, Jiajun Bu, Chun Chen, Kun Yang, Deng Cai, Xiaofei He, “Document Recommendation in Social Tagging Services”, WWW’10 Proceedings of the 19<sup>th</sup> international conference on World wide web, pp.391-400, 2010
- [4] 吉田拓実, 井上潮 “ソーシャルブックマークと Wikipedia を用いた Web ページ推薦と重要文抽出手法”, DEIM2013.
- [5] 百田信, 伊東栄典 “ソーシャルブックマークに基づく情報発見”, DEWS2008.
- [6] 石橋智幸, 顧優輝, 脇屋達, 真部雄介, 菅原研次, “ソーシャルブックマークを用いた Web 推薦システムの開発”, 電子情報通信学会技術研究報告. AI, 人工知能と知識処理, vol. 109(211), pp. 7-12, 2009.
- [7] 中山浩太郎, 原隆浩, 西尾章治朗, “Wikipedia マイニングによるシソーラス辞書の構築手法”, 情報通信学会論文誌, vol. 47, pp. 2917-2918, 2006.
- [8] “はてなブックマーク”, <http://b.hatena.ne.jp/>
- [9] “Wikipedia: カテゴリの方針”, <http://ja.wikipedia.org/wiki/Wikipedia:カテゴリの方針>