

# 法政大学学術機関リポジトリ

HOSEI UNIVERSITY REPOSITORY

PDF issue: 2024-09-03

## ソーシャルタグを教師データとしたコンテンツ推薦システム

HASEGAWA, Naohiro / 長谷川, 直広

---

(出版者 / Publisher)

法政大学大学院情報科学研究科

(雑誌名 / Journal or Publication Title)

法政大学大学院紀要. 情報科学研究科編 / 法政大学大学院紀要. 情報科学研究科編

(巻 / Volume)

9

(開始ページ / Start Page)

131

(終了ページ / End Page)

136

(発行年 / Year)

2014-03

(URL)

<https://doi.org/10.15002/00010534>

# ソーシャルタグを教師データとしたコンテンツ推薦システム Content Recommendation System using Teacher Data created by Social Tag

長谷川 直広

Naohiro Hasegawa

法政大学大学院

情報科学研究科情報科学専攻

E-mail: naohiro.hasegawa.6w@stu.hosei.ac.jp

## Abstract

Social tag as a significant feature in social bookmark service can be used for metadata which describes contents related to tags. The metadata is applied to generate teacher data. Generally, a tag used by a user can somehow reflect the user's interests and preferences. This research is focused on using the relations of tags and their associated contents as learning data for text classification. The resulting text classifiers take a text as input and generate a vector of values from 0 to 1 associated to tags. The text classifiers are applied to a user's visited text documents on SNS for generating the user's profile including a collection of vectors. Based similarity of any two vectors, the system recommends the most relevant contents to the user. In the experiments, 200 tweets are using for generating a user's profile including about 50 tags and the three types of contents, website, web advertising, and books from Amazon.com are recommended to the user. The result shows that the system can generate a user profile reflecting the user's interests to a certain degree if the user is an often Twitter user. For obtaining reasonable recommendation contents, it is important to have sufficient tags in the user profile.

## 1. はじめに

ソーシャルタグとは、あるコンテンツに対して多様なユーザーがそれぞれの尺度で自由に単語を付与したものである。多様なユーザーがコンテンツにタグを付与することで、コンテンツの様々な側面を説明した情報が集約される。こうした仕組みはソーシャルタギングやフォークソノミーと呼ばれている。ソーシャルタギングの特徴として、不特定多数のユーザーがコンテンツにタグを付与することから、コンテンツに対して適切ではないタグが多少含まれていても、付与されたタグを集約することで、コンテンツに精度が高く幅広いタグがコンテンツに付与される。特に、ソーシャルブックマークでは、図1で示すように、ユーザーがブックマークしたウェブサイトと付与されたタグをウェブサイト分類の教師データとして利用できる側面と、ユーザーがブックマークしたウェブサイトをユーザーの趣味趣向を反映したユーザーファイリングの教師データとして有用である側面を持っている。

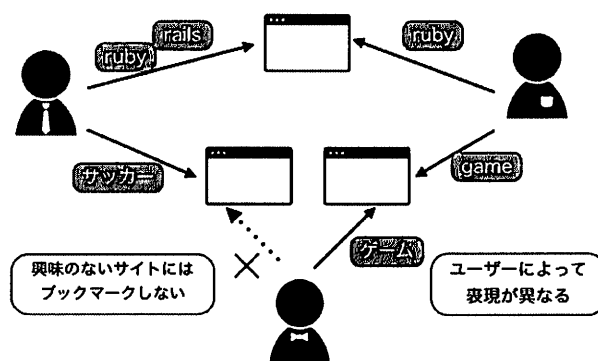


図1 ソーシャルブックマーク

ソーシャルタグを取り入れたサービスとして、画像共有サイトの「Flickr」やソーシャルブックマークを提供する「del.icio.us」が2000年代前半に人気となり、ソーシャルタギングが認知されるようになる。ソーシャルタギングを対象にした研究が注目されるようになった。ソーシャルタギングを対象にした研究として、ソーシャルブックマークにおけるユーザーのブックマーク情報をユーザーの趣味趣向としてユーザープロフィールを生成し、ウェブサイトの推薦を行うシステムがある。しかし、この研究の問題点として、推薦システムを利用するユーザーがソーシャルブックマークを利用している、あるいはSNSでウェブサイトの共有を行っている必要があるという制約があり、システムの対象となるユーザーが限定されてしまい、応用するアプリケーションが限られていた。

本研究では、ソーシャルタギングにおけるタグとタグに関連したコンテンツの文書の関係性から、ソーシャルタギングの特性に応じた教師データを生成する。ユーザープロフィールとコンテンツのプロフィールをそれぞれの関連する文書からタグ分類器を用いてその特徴を抽象的に表現し、ユーザーに推薦するコンテンツをそれぞれのプロフィールの類似度を用いて決定する。これにより、関連研究の問題点である推薦システムを利用するユーザーの制約を排除し、推薦するコンテンツをウェブサイトだけではなく、関連する文書を持った様々なコンテンツに対応する。そして、タグ分類器の分類精度とTwitterのツイートをユーザープロフィールに用いる推薦システムの実験を行い、提案する手法の有用性を示す。

## 2. 関連研究

ソーシャルブックマークを用いてユーザープロフィールリングを行いコンテンツの推薦を行う研究として、Folksonomy マイニングに基づく Web ページ推薦システム [1]がある。ユーザープロフィールと推薦するウェブサイトをそれぞれタグの親和度で表現し、算出された親和度を用いて推薦するウェブサイトを比較している。ユーザープロフィールのタグの親和度は、ユーザーがこれまで様々なウェブサイトに付与してきたタグ群を、ウェブサイトのタグの親和度は、不特定多数が付与したタグ群から算出している。

この研究を実際のサービスに応用する場合に次の二つの問題点が存在する。一つ目が、ソーシャルブックマークを利用しているユーザーが全体から見ると少ないため、サービスの提供を希望するすべてのユーザーには対応できない点である。二つ目が、推薦するウェブサイトは作成されたばかりである、あるいは人目につきにくい場合、十分なタグが付与されていない点である。実際のサービスに利用できるシステムに用いる場合、システムの推薦精度だけではなく、これらの問題を解決する必要がある。

さらに、関連研究として、ソーシャルメディアを利用したセレンディピティな情報推薦 [2]がある。この研究は、[1]の研究をベースに、隠れた興味であるセレンディピティに着目し、推薦結果の新規性や意外性を向上した。特筆すべき点としては、ユーザープロフィールの作成に、ソーシャルブックマークで用いるタグではなく、Twitter や Facebook などの SNS 上で共有したウェブサイトに付与されているタグをユーザーのプロフィールに用いていることである。ソーシャルブックマークを利用しているユーザーは少数であるが、SNS を利用しているユーザーは大多数であるため、アプリケーションへの応用の面で有効な改良である。しかし、SNS におけるウェブサイトの共有は、標準的な利用の仕方ではなく、すべてのユーザーが行っている訳ではなく、ユーザープロフィールの生成において、更なる改良の余地があると考えられる。

## 3. 推薦システム概要

図 2 に従来研究と提案するシステムのユーザープロフィールの生成方法の違いを示す。従来研究では、SNS 上での共有されたウェブサイトなどのユーザーの趣味趣向を表すウェブサイトの集合を用いて、タグ  $T$  との各ウェブサイトの親和度の総和である  $e$  がユーザープロフィールとなる。タグとウェブサイトとの親和度は、ソーシャルブックマークで付与されたウェブサイトのタグ集合における、タグの多重度と希少度からなる重みである。また、提案するシステムでは、ソーシャルブックマークのデータをもとに学習したタグ分類器  $T$  を用いて、タグ  $T$  とユーザーが利用している SNS から取得できる文書との親和度  $e$  をユーザープロフィールに用いる。

システムがユーザーに推薦するコンテンツは、ユーザープロフィールと同様の手法でプロフィールが生成されるため、推薦できるコンテンツはプロフィール生成手法に依存している。そのため、従来の研究では推薦できるコンテンツがソーシャルブックマークでタグが付与され

ているウェブサイトに限定されているが、提案する手法では、コンテンツを説明する文書が存在する様々な種類のコンテンツを推薦できる。

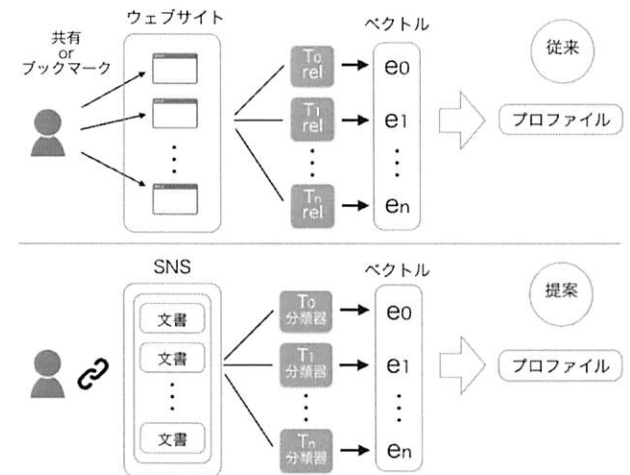


図 2 ユーザープロフィール生成方法の比較

図 3 に推薦システムの流れを示す。はてなブックマーク [3]からタグとタグが付与されたウェブサイト取得し、教師データとする。各タグにタグと与えられた文書の親和度を出力するタグ分類器を作成し、教師データを用いて学習を行う。ユーザーに推薦する候補の各コンテンツに、コンテンツに関連した文書をもとにタグ分類器を用いてプロフィールを生成する。コンテンツに関連する文書は、ウェブサイトなら HTML 文書、商品なら商品の説明文などを利用する。コンテンツのプロフィール作成が完了すれば、推薦システムの準備が完了する。

ユーザーがシステムにアクセスすると、ユーザーが普段利用している SNS やウェブサービスなどから外部向けの API を用いて、ユーザーに関連する文書を収集し、タグ分類器を用いてユーザープロフィールを生成する。ユーザーに推薦するコンテンツはユーザープロフィールと類似度の高いプロフィールを持つものとし、探索したコンテンツをユーザーに提示する。

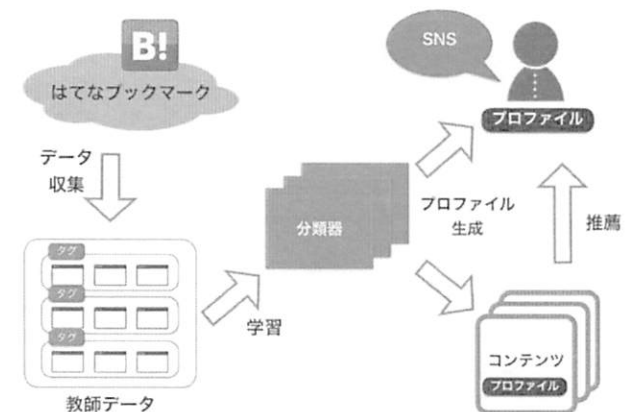


図 3 推薦システムの流れ

### 3.1 教師データ

教師データの素性として、ウェブサイトの HTML 文書に含まれる単語とその出現回数をもとにした重みを用いた。次元の削減を行うために、日本語形態素解析ライブラリである MeCab を用いて、名詞の中でも固有名詞と一般名詞のみに限定した。各単語の重みは、tf-idf 法[4]をベースにした手法により算出した。tf-idf 法において重みは、単語の出現頻度である tf と逆文書頻度である idf から算出される。特に idf が重要な役割を果たしており、多くの文書に出現する一般的な単語ほど重みが小さくなり、特定の文書でのみ出現する単語の重みを大きくする。

はてなブックマークから収集したデータは、タグとタグが付与されているウェブサイトという関係になっていることから、同じジャンルのウェブサイトが、無作為に収集した場合と比べて多い。特定のジャンルにおいて頻繁に使用されるが、他のジャンルでは使われない単語は、文書のタグ分類において重要であるが、文書頻度が高くなり、idf が低くなることで重みが小さくなる。この問題を解決するために、tf-idf 法を改良した式を次に示す。

$$tfidf = tf \cdot idf \quad (1)$$

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (2)$$

$$idf_i = \log \frac{|D|}{df_i} \quad (3)$$

$$df_i = \left| \left\{ d : d \ni t_i \wedge d \notin \{T : T \ni d_j\} \right\} \right| + 1 \quad (4)$$

$n_{i,j}$  は文書  $d_j$  における単語  $t_i$  の出現回数、 $|D|$  は総文書数である。 $df_i$  は通常、単語  $t_i$  が出現する文書数であるが、条件を加え、文書  $d_j$  が属するタグ  $T$  に属していない文書の集合で単語  $t_i$  が出現する文書数とすることで、ジャンル特有の単語の重みを大きくする。例として、総文書数が 10000、ある文書中の単語の出現頻度  $tf$  が 0.1、tf-idf 法による  $df$  が 100、提案する手法による  $df$  が 20 とすると、tfidf 値は、tf-idf 法で約 0.46、提案手法で約 0.69 となり、タグ特有の単語の重みが大きくなる。このことから、学習データがジャンル分けなされているような場合において、提案する手法は tf-idf 法より有効である。

### 3.2 タグ分類器

分類器に用いるタグとして、ソーシャルブックマークでは、ユーザーが自由にタグを入力することから、タグの種類は膨大で、あまり利用されないタグなどを含めた分類器を作成することは、計算時間を考えると現実的ではない。そこで、ユーザープロファイルの生成に適したタグを効率的に収集するために、はてなブックマークのタグ一覧(図 4)を利用する。はてなブックマークのタグ一覧では、タグクラウドと呼ばれるタグのフォントの大きさで、その時々々のタグの利用頻度を視覚的に表している。そのため、システムの計算能力に応じて利用頻度を高いタグに限定することで、プロファイルに用いる分類器の数を調整できる。また、インターネットで流行しているキーワードに応じたユーザープロファイルの生成に効果的なタグの抽出ができる。

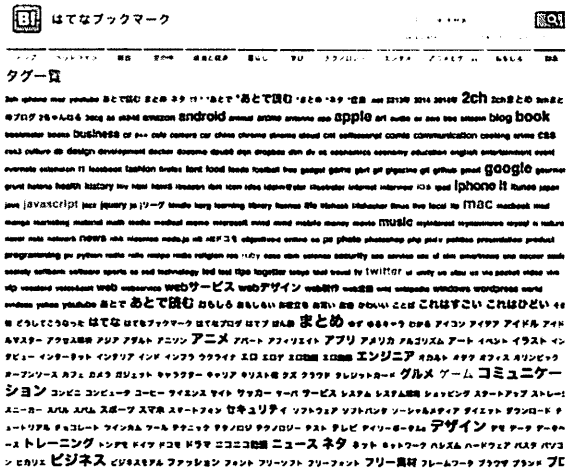


図 4 はてなブックマークのタグ一覧

また、タグ分類器に用いる文書分類手法に SVM を用いる。SVM はあらかじめ学習を行っておくことで高速に分類ができ、文書分類において性能が高いと知られている[5]。また、文書分類は多くの場合、線形分類可能であると実証されている[6]。そして、各タグ分類器の出力は SVM の確率推定手法[7]を用いて、タグと与えられた文書の親和度を 0 から 1 の実数で表す。

### 3.3 推薦

ユーザープロファイルとコンテンツのプロファイルのベクトルの類似度をもとに推薦するコンテンツを決定する。ベクトルの類似度計算手法として、コサイン類似度がある。しかし、SVM を用いたタグ分類器において、0.5 未満の出力は、ネガティブな判定の値であるが、コサイン類似度をそのまま適用すると、通常のリレーティブな値をユーザープロファイルとして用いるため、式(5)を用いて、ベクトルの値  $L_i$  が、0.5 より大きければ 0 から 1 の範囲にスケールした値、それ以外であれば 0 を変更後値  $L'_i$  とする。

$$L'_i = \begin{cases} 2L_i - 1 & (if L_i > 0.5) \\ 0 & (otherwise) \end{cases} \quad (5)$$

また、文書を複数のタグ分類器で多面的に分類するため、ある部分のベクトルの類似度が場合でも、その他の部分の類似度が低ければ、ベクトル全体の類似度は低下する。式(6)を用いて、類似度を計算する二つのベクトルの値  $L1_i$  と  $L2_i$  が互いに 0 より大きい場合のみ類似度の計算に用いることで、ベクトルにおいて特徴的な部分に重点をおいた類似度を得ることができる。

$$L1'_i = \begin{cases} L1_i & (if L2_i > 0) \\ 0 & (otherwise) \end{cases} \\ L2'_i = \begin{cases} L2_i & (if L1_i > 0) \\ 0 & (otherwise) \end{cases} \quad (6)$$

表1 ベクトル修正後の類似度の比較

	なし	(5)	(5) + (6)
(ex1) L1 = [0.1, 0.2, 0.7, 0.1] L2 = [0.1, 0.6, 0.2, 0.1]	0.58257	0	0
(ex2) L1 = [0.1, 0.2, 0.7, 0.1] L2 = [0.1, 0.6, 0.2, 0.1]	0.78842	0.48953	0.78935

類似度が低い二つのベクトルの例を ex1, 類似度が高い二つのベクトルの例を ex2 とし, 二つの例に対して, コサイン類似度の計算のみ行ったものと, (5)の式を用いて正規化してからコサイン類似度を計算したもの, さらに, 正規化と(6)のベクトルの修正を加えてコサイン類似度を計算したものの結果を表1に示す.

類似度が低い ex1 の例において, コサイン類似度のみの値は, 0.58 と比較的高い数値となっているが, 提案する方法では類似度 0 となり期待した結果を得ることができる. また, 類似度が高い ex2 の例では, コサイン類似度のみの値と比較して, 正規化後コサイン類似度を計算した値は, 類似度が高いとは言えない水準にまで低下している. これは, 正規化の処理によってネガティブな値が 0 に置き換えられ, 部分ごとの類似度に影響を与えたためである. (6)の処理を加えることで, 0 である要素を類似度に反映されなくなり, もとのコサイン類似度と同じ水準の類似度を得ることができる. このことから, 提案するベクトルの修正手法は, 複数のタグ分類器によって生成されるベクトルの類似度の計算において有用である.

## 4. 実験

### 4.1 タグ分類器

推薦システムの評価を行う前に, プロファイルの属性として利用しても問題のないタグ分類器を選別するために, 各タグ分類器の精度を検証した. また, タグ分類器には, SVM のライブラリである LIBSVM[8]を用いて, カーネルを線形カーネルとした.

#### 4.1.1 実験データ

はてなブックマークのタグ一覧のタグクラウドで, 各タグの html タグの class に付与されている tagcloudN(N は 0 以上の整数)の N が 1 以上に限定し, 134 つのタグを選択した. また, 各タグの教師データに用いるウェブサイトは, はてなブックマークの検索機能を用いて, タグごとに, 1 ページにつき 40 件, 2 ページ分で最大で 80 件, 合計 10477 件収集した. 検索条件として, ブックマーク数を 50 以上とした. また, 各タグ分類器の教師データポジティブなサンプルとそれ以外のすべてのタグのネガティブなサンプルの比率が 1 対 1.5 となるように, ネガティブなサンプルをランダムに選択した.

#### 4.1.2 評価方法

各タグ分類器の精度の評価には, 3 分割交差検証を用いる. K-分割交差検証では, 学習データを K 分割し, K-1 組のデータを学習に用い, 残りの 1 組を検証に用い, これ

をすべての組み合わせに対して行う. また, 分類器の精度を適合率(precision), 再現率(recall), さらに適合率と再現率の調和平均である f 値を指標とした.

### 4.1.3 実験結果

表2 分類精度が高いタグ一覧

	precision	recall	f
さくらの vps	0.891	0.868	0.880
aws	0.859	0.762	0.807
大阪	0.925	0.714	0.806
グルメ	0.875	0.727	0.794
サッカー	0.931	0.683	0.788
Ruby	0.828	0.725	0.773
料理	0.890	0.671	0.765
wordpress	0.765	0.712	0.765
photoshop	0.843	0.675	0.750
python	0.773	0.725	0.748
株	0.852	0.666	0.748
電子書籍	0.905	0.623	0.738
軍事	0.659	0.810	0.727
機械学習	0.886	0.602	0.717
漫画	0.884	0.582	0.702

表3 分類精度が低いタグ一覧

	precision	recall	f
プログラミング	0.604	0.362	0.453
学習	0.724	0.269	0.392
メディア	0.656	0.280	0.392
ネット	0.567	0.269	0.365
これはすごい	0.606	0.250	0.353
まとめ	0.612	0.237	0.342
ツール	0.542	0.246	0.339
あとで読む	0.558	0.237	0.333
book	0.607	0.217	0.320
amazon	0.615	0.200	0.301
ネタ	0.571	0.202	0.299
お役立ち	0.560	0.175	0.266
技術	0.387	0.151	0.218
web	0.161	0.063	0.090
資料	0.250	0.053	0.087

表2に示す分類精度が高いタグ一覧のタグは, 表3に示す分類精度が低いタグ一覧のタグと比較して, より具体的に専門的な単語となっている. “学習”, “ツール”, “技術”や“web”などの分類精度が低いタグは, 抽象度が高く, 様々なジャンルのウェブサイトが付与されるため, タグを決定づける単語が弱いため精度が低くなっていると考えられる. また, はてなブックマークで特殊な使われ方をする“あとで読む”タグがある. これは, 記事の内容に対する情報ではなく, ユーザーがあとで読みたいと思ったものに対してつけられる. このタグもジャンル関係なく使われるため, 精度が低くなっていると考えられる.

また、専門性が高いと思われる“プログラミング”タグの精度が低い理由として、ruby や python, Java などの“プログラミング”タグをより具体的なタグが存在するためタグを決定づける単語の重みが分散してしまったと考えられる。これらの分類精度が低いタグを使わず、高いタグのみを使うことで、説明力の高いタグのみをユーザープロフィールに用いることができると考えられる。もし、“プログラミング”タグのような抽象度は高いが専門性のある単語をプロフィールに用いたいならば、タグの類似度や共起度をもとに階層的クラスタリングを行い、そのタグの下位のタグの分類結果を補完すれば良い。しかし、ユーザープロフィールをもとにユーザーにコンテンツを推薦することを考えると、ユーザープロフィールの各タグは、より具体的である方が、ユーザー趣味趣向に合致した推薦が可能になるため、プロフィールに抽象的なタグを取り入れる必要はないと考えられる。

## 4.2 推薦システム

提案するシステムの実用性を検証するため、ユーザープロフィールを Twitter アカウントから生成し、プロフィールに応じたコンテンツ推薦するウェブシステムを実装した。プロフィールを生成するタグ分類器は、タグ分類器の実験において  $f$  値が 0.6 以上の精度が比較的高いものに限定した。また、ユーザープロフィールには、Twitter の API を用いてユーザーのツイートを 1 回のリクエストで取得できる最大数である 200 件を 1 つの文書に統合するものを用いる。これは、分類器は文書量が少ないと精度が低くなること、さらに 200 ツイートを 1 ずつ分類すると時間が非常にかかることが理由である。

### 4.2.1 実験データ

推薦するコンテンツとして、ウェブサイト、さらにビジネスへの応用として、ウェブ広告と Amazon.com で販売されている本の 3 種類を用意した。ウェブサイトはタグ分類器の学習に用いた 10477 件を流用した。ウェブ広告は、アフィリエイトサービスの一つである A8.net から 2888 件収集した。広告のプロフィールには、広告のリンク先であるウェブサイトの HTML 文書を用いた。また、本のデータ収集では、A8.net ではデータの母体数が少なくデータの絞り込みを行わなかったが、Amazon.com のデータは膨大ですべて収集することは不可能と判断し、各タグ分類器のモデルのファイルから、重みが大きくタグと関係が強いと思われる単語を上位 5 つ取り出して、その単語を検索ワードとして指定し、合計 4356 件収集した。

### 4.2.2 実験方法

Twitter から取得したツイートをもとに生成したユーザープロフィールに応じて、ウェブサイト、ウェブ広告、本をそれぞれ 10 件ずつ推薦する。ユーザーは、推薦されたコンテンツをそれぞれ、“まったく興味がない”1 から“興味がある”の 5 までの 5 段階で評価してもらい実験を 6 人の学生に対して行い、提案するシステムの実用性を検証した。また、推薦システムの評価画面の様子を図 5 に示す。



図 5 推薦システムの評価画面

## 4.2.3 実験結果

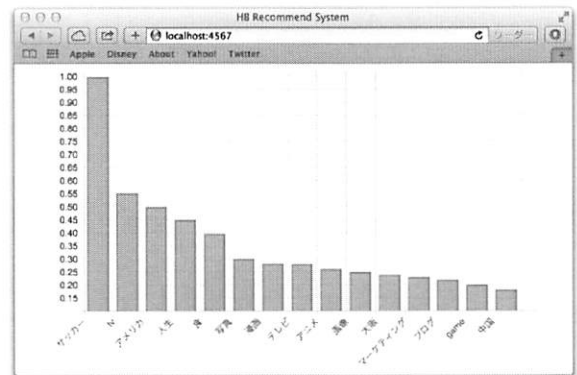


図 6 評価が高かったユーザーのプロフィール

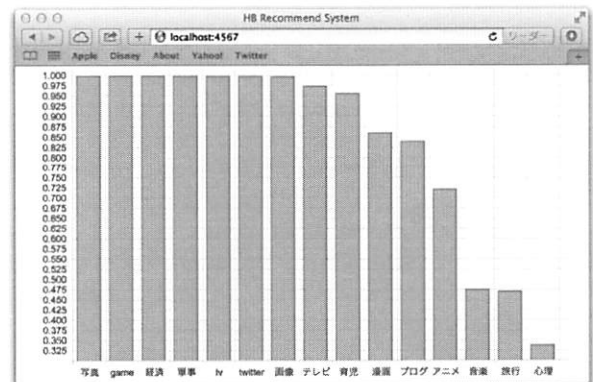


図 7 評価が低かったユーザーのプロフィール

評価が高いユーザーと評価が低いユーザーのプロファイルをそれぞれと図6と図7に示す。評価の高い被験者は、サッカー観戦が趣味でサッカー関連のツイートを主にしていたことから、サッカータグとの親和度が群を抜いて高くなっている。実際に推薦されたウェブサイトや本はサッカー関連のもとなり、高い評価を与えたと考えられる。また、評価の低い被験者は、0.8以上のタグが多数存在する異常なプロファイルが生成されている。原因として、この被験者はTwitterを積極的に利用しておらず、ツイートの大半が友人とのリプライと呼ばれるメッセージのやりとりで、それらは非常に短い文書であると証言している。実際に、文書量の少ない仮の文書からプロファイルを生成した結果この被験者と同じような出力となった。このことから、ユーザープロファイルに用いるには、リプライなどの短い文書ではなく、ある程度の長さの文書であることが望ましいとわかった。

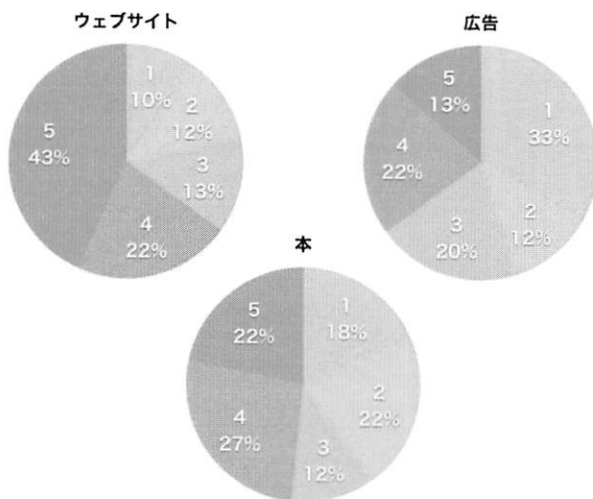


図8 各コンテンツの評価

図8のグラフは、システムが推薦したウェブサイト、広告、本のコンテンツに対しての被験者の回答を集計し示したものである。ウェブサイトでは、5と4の高い評価が多く、広告では最低の1が最も多く、5の評価も少ない。また、本では3の評価が少なく、評価の高い5、4か評価の低い1、2などに回答が分かれた。

広告の推薦が低い評価であった理由として、分類器に用いるタグとA8.netでのジャンルが適合せず、さらに広告の絶対数も少ないことため、プロファイルに適していない広告を推薦せざるを得なかったためと考えられる。今回の実験では、実行時間の短縮のため、分類器に用いるタグの数を少なくしたが、推薦するコンテンツに応じて分類器に用いるタグを柔軟に変更することで、広告の推薦で発生した問題に対応できると思われる。そして、本の推薦では、分類器のモデルから取得した重みの大きい単語を検索ワードにしたことで、ジャンルに適合したコンテンツを収集したが、ウェブサイトのデータと異なり、人気のある本であるかは検索条件に指定しなかった

ため、ユーザーが興味を持ちにくいデータも多数存在し、評価が分かれたと考えられる。

また、ある被験者は本の推薦リストに対して、プロファイルに応じて漫画が推薦されている点は良いが、興味のない少女漫画などが推薦されていたため評価を下げたと述べていた。このことから、分類器に用いるタグが抽象度の高いものが多く、具体的なタグが少なかったと言える。漫画タグを例に挙げると、ギャグ漫画や格闘漫画などのより体的なタグを分類器に用いることで、よりユーザーの趣味趣向に応じた推薦が行える。しかし、分類器の数に応じてプロファイルの生成時間が増加するため、システムの要求する実行時間との兼ね合いで、分類器に用いるタグの種類を調整する必要がある。

## 5. まとめ

本論文では、ソーシャルタギングの特性に応じた、教師データの生成手法とユーザーに推薦するコンテンツの選択手法を提案した。ユーザープロファイルを抽象化することで、様々なコンテンツの推薦を行えるシステムを構築し、システムの有用性を検証した。ユーザープロファイルの生成においてTwitterのツイートを用いたが、リプライなどの短い文書が多い場合を除き、ユーザーの趣味趣向に応じたプロファイルを生成できた。また、コンテンツの推薦に関しても、プロファイルに用いるタグにより具体的なタグを加えることでユーザーの趣味趣向に応じた推薦ができるとわかった。今後の課題として、推薦するコンテンツの特性に応じて、プロファイルに用いるタグを柔軟に変更し、より詳細にユーザーの趣味趣向に応じたシステムが構築できるか検証したい。

## 参考文献

- [1] 丹羽 智史: Folksonomy マイニングに基づく Web ページ推薦システム, 情報処理学会論文誌, vol47 no.5, 2006.
- [2] 福島 良典, 大澤 幸生: ソーシャルメディアを利用したセレンディピティな情報推薦, 人工知能学会全国大会, 2012.
- [3] はてなブックマーク : <http://b.hatena.ne.jp/>
- [4] Gerard Salton and Christopher Buckley. Term-weighting approaches in automatic text retrieval. *Information Processing and Management*, 24(5):513-523, 1988.
- [5] Fabrizio Sebastiani. Machine learning in automated text categorization. *ACM Computing Surveys*, 34(1):1-47, 2002.
- [6] Thorsten Joachims. Text categorization with support vector machines: learning with many relevant features. In Claire Nédellec and Céline Rouveïrol, editors, *Proceedings of ECML-98, 10th European Conference on Machine Learning*, number 1398, pp. 137-142, Chemnitz, DE, 1998. Springer Verlag, Heidelberg, DE.
- [7] Tzu-Kuo Huang. Generalized bradley-terry models and multi-class probability estimates," *The Journal of Machine Learning Research* 7, pp. 85-115, Jan. 2006.
- [8] Chih C. Chang and Chih J. Lin. LIBSVM: a library for support vector machines, 2001