

Folksonomy におけるタグ情報を用いたコンテンツ推薦手法の提案

数原良彦[†] 篠沢佳久[†] 櫻井彰人[†][†]慶應義塾大学大学院理工学研究科

1 はじめに

近年、ウェブの普及に伴うウェブ利用者の増加に従い、ユーザがタグと呼ばれる自由記述のメタデータを付与することでコンテンツを管理する folksonomy と呼ばれる分類手法が注目されている。folksonomy を利用しているサービスとして、ソーシャルブックマーク、画像共有、動画共有などが挙げられる。folksonomy の現状や問題点については様々な指摘があるが、その中のひとつとして、タグ利用の難しさが挙げられる。例えば現状の folksonomy でコンテンツ検索を行うためには、タグをクエリとして検索をする方法が一般的で、異なる意味で同じタグが付与されている場合、ユーザが目的のコンテンツを見つけることが難しい。このように folksonomy による分類は、発見的探索には役立つことがあっても、ユーザの視点に基づく目的指向の検索には向いていないことが指摘されている [1]。

Folksonomy におけるタグ情報を用いたコンテンツ推薦の研究がある。丹羽らは、ユーザの嗜好をユーザとタグの親和度によって表現し、ユーザ間の嗜好類似度に基づくコンテンツ推薦手法を提案している [2]。この手法ではタグ名称が同じものを同一の意味づけをしていると解釈するため、ユーザ間の視点の異なりを考慮していない。我々は、ユーザが付与したタグは、個人の視点に大きく依存するため、タグ名称の一致によるコンテンツ推薦は難しいと考える。佐々木らは、各ユーザが同じタグを付与したコンテンツ群における web コンテンツの共起性に基づく推薦システムを提案している [3]。この手法では、選択されたタグへコンテンツを推薦するためには、他ユーザが同じタグを付与したコンテンツ群（コンテンツクラスター）との類似度を計算する必要がある。このように、我々はタグをユーザにとってのひとつの概念と考え、ユーザが選択した概念に対してコンテンツを推薦する手法を提案する。

本研究では、推薦対象となるユーザを [3] と同様に folksonomy の利用者で、ある視点に基づいて共通の名称でコンテンツにタグを付与しているユーザと想定している。また推薦手法の利用方法として、推薦対象ユーザが保持しているタグの中から推薦を受けたいものを選択し、システムがユーザがそのタグを付与した視点

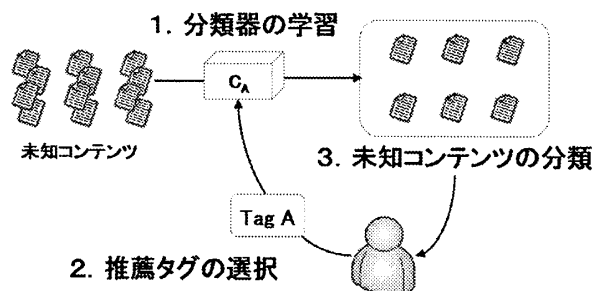


図 1: 提案手法の概要

に基づいてコンテンツが推薦するというシステムを想定している。本稿では推薦対象となるタグのことを推薦タグと呼ぶことにする。ユーザひとりに着目すると、ユーザが持つタグ情報はユーザの概念空間を反映していると解釈することができる。本研究では、ユーザが未知のコンテンツに対してタグが付与されるかを判別する分類器を学習することにより、課題の解決を試みる。

2 提案手法

本研究では、選択された推薦タグに対してコンテンツの推薦を行うため、あるコンテンツへ推薦タグを付与するか判別する分類器を学習し、分類することでユーザへ未知コンテンツの推薦を実現する。

我々は、多数のユーザによって付与されたタグが十分多量に存在すれば、タグ情報がコンテンツの客観的な特徴を表現すると考えた。具体的には、コンテンツに付与されたタグの種類を各次元とし、タグ頻度を各次元の値とする bag-of-words と同様の考え方で特徴を表現する。この際、どのユーザによってタグが付与されたかという情報は考慮しない。

事例の考え方は、推薦対象ユーザによって推薦タグが付与されたコンテンツを正例、それ以外のタグが付与されたコンテンツを負例とする。ただし、本研究ではユーザがコンテンツにタグを付与した時間情報を利用しないため、コンテンツが追加された後にタグが新たに作成された可能性を考慮することができず、厳密には負例とはいえないことに注意する。

提案手法の大まかな流れは以下の通りである (図1)。

1. あらかじめ各ユーザの推薦タグについて分類器の学習を行う。
2. ユーザが推薦タグを選択する。
3. 未知コンテンツに対して分類を行い、推薦を行う。

Web Content Recommendation Based on Folksonomy Tags

Yoshihiko SUHARA[†], Yoshihisa SHINOZAWA[†], Akito SAKURAI[†][†]Graduate School of Science and Technology, Keio University Hiyoshi 3-14-1, Kouhoku-ku, Yokohama, Kanagawa, 223-8522 Japan

{suhara, shino, sakurai}@ae.keio.ac.jp

表 1: 提案手法とベースライン手法の比較

提案手法	マイクロ平均			マクロ平均		
	適合率	再現率	F 値	適合率	再現率	F 値
提案手法	0.10	0.50	0.16	0.15	0.53	0.23
ベースライン	0.21	0.11	0.14	0.12	0.14	0.13

3 評価

3.1 実験

提案手法の評価を行うために評価実験を行った。代表的なソーシャルブックマークサービスである del.icio.us¹から 2007 年 5 月時点における約 16,000 ユーザ、約 870 万コンテンツのデータを取得した。その中から、7 ユーザが保存している合計 718 個のタグについて分類評価を行った。

ベースライン手法として、コンテンツのウェブページのテキストから HTML タグと不要語を除去し、接辞処理を行った bag-of-words を特徴とする手法を用意し、提案手法との比較を行った。

実験には Weka²を使用し、多項モデルの naïve Bayes 分類器 (NaiveBayesMultinomial) を選択した。評価方法は 2-cross validation で行った。そして正例について適合率、再現率、F 値を算出し、ベースライン手法との比較を行った。

3.2 結果

表 1 に実験結果、図 2 に各タグにおける提案手法とベースライン手法による分類結果の F 値の散布図を示す。図 2 は x 軸にベースライン手法、y 軸に提案手法の F 値を取っており、 $y = x$ より上側にある点は提案手法の方が高い F 値を示していることを表している。

表 1 では、マイクロ平均の適合率を除いた全てにおいて提案手法がベースライン手法より高い値を示している。また、いずれの場合においても適合率が低い値を示しているが、これは今回の実験では必然的に負例の割合が多く、誤って正例に分類してしまう false positive の数が多くなるためである。

いずれの手法においても適合率のマクロ平均とマイクロ平均の値が大きく異なっているが、これはユーザごとの分類精度にばらつきがあるためであり、これについても追加実験を行い、検証を行う必要がある。

図 2 において、各タグについての F 値を比較すると、一部のタグを除いて提案手法が高い値を示していることがわかる。

以上の結果より、bag-of-words を用いたベースライン手法では、多数のユーザの視点を判別することが困難であることがわかる。また、ベースライン手法では全ての URL 先のコンテンツを取得しなければならず、コスト面からも提案手法が優れているといえる。

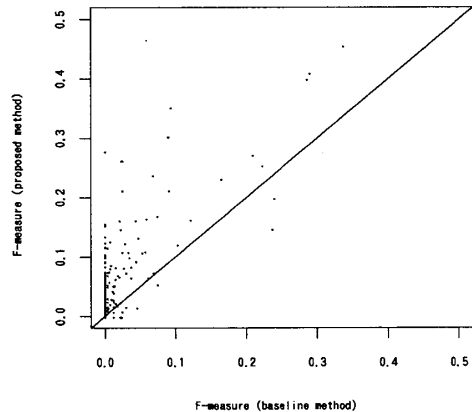


図 2: 各推薦タグにおける F 値の比較

4 おわりに

本稿では、多数のユーザによって付与されたタグをもとに、folksonomy ユーザに対して、付与タグについての分類器を用いることで、未知コンテンツの推薦を行う手法を提案した。評価実験より、ウェブページの単語を特徴とするベースライン手法より高い精度で分類できることを示した。

提案手法では naïve Bayes 分類器のような追加学習が可能な分類器を利用すれば、データ全体に対する再学習をせずに利用することが可能であり、スケーラビリティの面からも有効であるといえる。

コンテンツにタグが付与されていなければ、提案手法を利用することができない。今後の課題としては、タグのないコンテンツについて、タグを補完することで提案手法を適用できるようにする必要がある。

Folksonomy は集合知 (wisdom of crowds) [4] の一形態とも考えられ、従来の分類体系に比べ利点が述べられることが多い。はじめに述べたように現状の folksonomy は個人の視点の集合体でしかなく、まだそれが十分に利用可能な状態には至っていないと考えている。隠れた有益な情報を利用し、個人の視点に基づいた推薦手法を実現したいと考えている。

参考文献

- [1] Voss, J.: Tagging, folksonomy & co - renaissance of manual indexing? In Proceedings of the International Symposium of Information Science, pp.234-254 (2007).
- [2] 丹羽智史, 土肥拓生, 本位田真一: 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, Vol.47, No.5, pp.1382-1392 (2006).
- [3] 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 (TOD36), pp.14-27 (2007)
- [4] Surowiecki, J.: The Wisdom of Crowds, Doubleday (2004).

¹<http://del.icio.us/>

²<http://www.cs.waikato.ac.nz/ml/weka/>