ソーシャルブックマークデータを用いた推薦アルゴリズムの提案

齋藤　敬† 富樫　敦‡ 梶　功夫‡
宮城大学大学院　事業構想学研究科　事業構想学専攻†
宮城大学 事業構想学部　デザイン情報学科‡

1 はじめに

従来のキーワードを用いたウェブの検索において，的確なワードが思いつかない場合の対応が課題となっている．本研究ではソーシャルブックマークデータ（以下 SBM）を用いた情報推薦アルゴリズムを提案する．SBM データはフォークソノミとも呼ばれる集合知である．

本手法はタグとユーザのペア集合を標本空間とした推薦モデルを構築し，情報量と組み合わせ，検索における新しい切り口の発見を目指す．本稿は SBM データを用いた推薦アルゴリズムの研究について6節にわたって述べる．

2 ソーシャルブックマーク

本研究では推薦に利用する集合知データセットとして SBM データを用いた．本節では SBM の特徴とタグ付けの問題について述べる．

2.1 特徴

ソーシャルブックマークとは他の利用者に対して公開・共有できるオンラインブックマークサービスである．これまで個人がローカルに溜め込むものだったブックマークを不特定多数で公開・共有することによって，有益な情報源として期待されている．

SBM の主な特徴としてコンテンツのタグ型分類を挙げられる．タグ型分類とは folksonomy ともよばれ，利用者が分類対象に複数の「タグ」を付け加える手法である．さまざまな角度からの分類に対応できるため柔軟な分類を可能にする。

2.2 タグ付けにおける問題点

SBM データにおいて、ユーザのタグ付けにおける多義性，同義性が問題として挙げられる．

多義性(polysemy)とは，同一の単語を異なる意味で用いている状態を指し、ユーザにとって不本意な形でアイテムが分類されてしまう。

同義性(synonymy)とは同じ意の語が複数あることを指し、過剰に種類が増加する原因となる．

Proposal of Recommendation Algorithm based on Social Bookmark Data
†Graduate School of Project Design, Miyagi University
‡Department of Spatial Design and Information Systems, Miyagi University

3 研究のねらいと提案手法

本研究では先行研究を踏まえ，主観的かつ定量的な推薦モデルと評価式の考案に取り組んだ．

SBM データを使った推薦は利用者情報と未知のコンテンツとの距離を推論することに帰結する．本研究は，利用者情報とコンテンツを確率で表した推薦モデルと，お互いの因果関係を情報量で表現する評価式を提案する．本節では基礎集合と推薦モデル，推薦評価式について述べる．

3.1 基礎集合

SBM データ(D)は主にユーザ(A)，タグ(T)、コンテンツ(Q)の三要素で構成されている．

$$A=\{a_1,a_2,\ldots,a_{N_A}\}\quad Q=\{q_1,q_2,\ldots,q_{N_Q}\}$$

$$T=\{t_1,t_2,\ldots,t_{N_T}\}$$

$$D=\{\langle a,q,t\rangle | a\in A,q\in Q,t\in T\}$$

本手法では，利用者情報を推薦プロファイルと呼び，ユーザが収集した任意のコンテンツ群と定義し Qc と表す．また，推薦候補となるコンテンツ群は推薦プロファイルに含まれていないコンテンツ群として¬Qc と表す．

$$Q_C\subseteq Q,\neg Q_C\subset Q$$

3.2 推薦モデルの構築

本研究は推薦プロファイルと推薦候補コンテンツの因果関係を定量的に評価するための推薦モデルを定義した．本項では推薦モデルで用いた標本空間と事象について述べる．

(1) 概念を用いた標本空間の定義

本研究では SBM 上に存在するタグとユーザのペアを，「概念(I)」と定義し，推薦モデルの標本空間として利用する．タグをユーザで細分化することでより厳密なユーザの意図を表現した．

$$I=\{i | i=\langle a,t\rangle,\exists q\in Q,\langle a,t,q\rangle\in D\}$$

(2) コンテンツを表す事象の定義

SBM データよりコンテンツは概念の集合であると言える．よって任意のコンテンツは標本空間 I の事象として定義できる

$$E_q=\{i | i=\langle a,t\rangle,\exists q\in Q,\langle a,q,t\rangle\in D\}$$

(3) 推薦プロファイルを表す事象の定義

定義より，推薦プロファイルは任意のコンテ

Copyright © 2010 Information Processing Society of Japan.
All Rights Reserved.

ンツ群である．よって任意の推薦プロファイルは標本空間Ｉの事象として定義できる．しかし，実際に上記の式を適応すると推薦結果に一部コンテンツの限定的な概念の共起でプロファイルの意図とかけ離れた問題が生じてしまう．本研究では，この問題を解決するために推薦プロファイルの定義について調整を行った．

$$E_C=\bigcup_{q\in Q_C} E_q \qquad E_C=\bigcup_{q',q\in Q_C} E_q\cap E_{q'}$$

3.3 情報量を用いた推薦評価式

推薦プロファイルとコンテンツを推薦モデルの事象として定義した．本項では，この２事象の因果関係を定量的に評価するための推薦評価式と元となったＪ測度について述べる．

(1) Ｊ測度

事象間の因果関係を評価する手法としてITRULE アルゴリズムにおけるＪ測度を挙げる．Ｊ測度は，事象Ｘにおいて条件に事象Ｙが成立したときにおける確率分布の変化をＪ値と呼ばれる情報量で表した値で以下の式で表し、推薦評価式へ採用することを検討した。

$$J(x|y)=p(y)\left(p(x|y)\log\frac{p(x|y)}{p(x)}+(1-p(x|y))\log\frac{(1-p(x|y))}{(1-p(x))}\right)$$

(2) 推薦評価式

Ｊ測度を推薦モデルに適応すると，負の相関の場合も高い値で評価されてしまう．本研究ではＪ値の生成式から負の相関の部分を取り除いた式を次の式で表した．各事象の確率は元の個数を標本空間で割った値である．この式は推薦プロファイルと推薦候補コンテンツとの共起性が高く，かつその共起のみに限定して出現する場合に高い評価をつける．

$$Point(E_q,E_C)=P(E_C)*P(E_q|E_C)\log\frac{P(E_C)*P(E_q|E_C)}{P(E_C)*P(E_q)}$$

4 提案手法の実証実験

本研究では推薦アルゴリズムの有用性を評価するために実証実験を計画している．本項では実験の指標と手順ついて述べる．

4.1 評価指標について

本手法の有用性を確認するために被験者による評価と既存システムとの比較を用いる．

被験者による評価は，開発した検証ソフトウェアを用いて，適合率と推薦によって新たに得られた有益なコンテンツ数の比較を用いる．そして，推薦プロファイルのコンテンツ数と適合率の相関を求めて，一件のコンテンツから新たに有益なコンテンツを発見する尺度を求める．

比較する既存システムとしては，はてなブックマークを用いる．このシステムは Complement Naive Bayes を推薦エンジンに組み込んでおり，本手法の比較対象として適していると考える．

4.2 実験の手順

実験では予め，５０人程度の被験者を予定している．被験者は，開発した検証ソフトウェアから実際にコンテンツの推薦を受け，有益なコンテンツの数を評価してもらう．そのため本研究では図３に示すようなClip treasure を開発した．本項は実験の三つの手順について述べる．

はじめに被験者は，既存の SBM データから興味のあるコンテンツを収集してもらう．次に集めたコンテンツを推薦プロファイルとしてClip treasure からコンテンツの推薦を受けてもらう．最後に推薦された有益なコンテンツを指定してもらいソフトウェア側に記録させる。

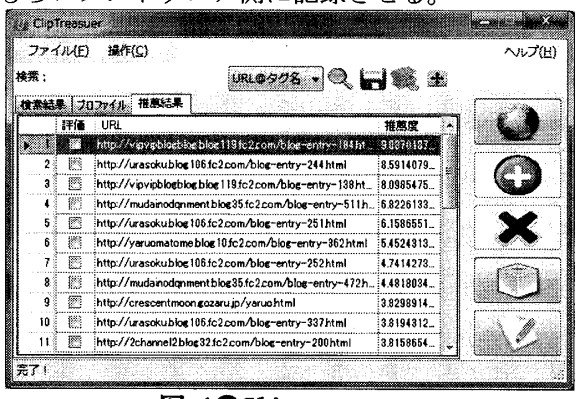

図 1●Clip treasure

5 終わりに

SBM データを用いた推薦モデルと情報量を用いた推薦アルゴリズムについて述べ，煮詰まった検索の状況を打開する新しい切り口を発見する手法であることを説明した．今後はユーザがブックマークしたコンテンツの注目した箇所がコンテンツの主な内容でない場合，推薦結果が意図するものと異なってしまう問題を解決したい．

参考文献

[1] 深見嘉明：ソーシャルブックマークサービスにおけるアノテーション情報の機能分析，人工知能学会全国大会論文集，Vol. . 21，pp. 1G1-4，(2007)

[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，vol20，No20，pp. 14-27(2007)

[3] P. Smyth and R. M. Goodman, an information theoretical approach to rule induction from databases, IEEE Trans. Knowledge and Data Engineering, vol. 4, no. 4, pp. 301-316, Aug. 1992.

Copyright © 2010 Information Processing Society of Japan.
All Rights Reserved.