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

佐々木 祥[†] 宮田高道[†] 稲積泰宏[††]
小林亜樹[†††] 酒井善則[†]

近年急速に普及しているソーシャルブックマークは，ユーザ間でブックマークを共有できるサービスであり，新たな情報収集ツールとして注目されている．ソーシャルブックマークでは，ユーザはwebコンテンツにタグと呼ばれる自由記述のキーワードを付与できるため，既存研究においてタグの名称に着目したwebコンテンツ推薦システムが提案されている．しかしながら，ユーザの嗜好はタグの名称ではなく，タグを表象として関連付けられたwebコンテンツ群（以下，コンテンツクラスタ）として表出するものといえる．そこで本研究では，コンテンツクラスタ間の類似度を仮説検定問題として求め，得られた類似度に基づくwebコンテンツ推薦システムを提案する．また，提案手法の検証実験によって，付与するタグの名称が他のユーザと異なるユーザに対しても有効に推薦することが可能であるなどの有効性を確認することができた．

Web Content Recommendation System Based on Similarities among Contents Cluster of Social Bookmark

Akira Sasaki,[†] Takamichi Miyata,[†] Yasuhiro Inazumi,[††]
Aki Kobayashi[†††] and Yoshinori Sakai[†]

The web-based bookmark management service called social bookmark has recently been in the spotlight and come to be recognized as a new information sharing tool. Social bookmark service allow users to tag keywords to each of their entries. These keywords are called 'tags'. There are some conventional studies of the web content recommendation system based on social bookmark which is focused to actual words of tags. However, the essential information of tags is not tag names, but classification of web contents by tags (we called the result of this classifications as contents cluster). Based on this assumption, we calculate similarities between contents clusters by using hypotheses test. By using calculated similarities, we proposed the web content recommendation system based on these similarities. It has been shown that our proposed method is working well, as the fact that appropriate recommendation can be offered to users, including who tagged different named tags to the same contents.

1. はじめに

近年，blogやweb日記を代表とするユーザ参加型サービスが広く普及したことにともない，WWW上のwebコンテンツは増加傾向にあり，ユーザは所望するwebコンテンツを発見することが困難となってきている．

このような状況を受け，webコンテンツの発見を容易にするwebコンテンツ推薦システムが注目されており，これまで数多くの提案がされてきた[1)~4)]．これらの多くは，大量の情報から有用な情報を選択する情報フィルタリングの一手法である協調フィルタリング[5)~7)]を技術的な核としている．ここで，情報フィルタリングにおける「情報」とは，本や音楽，ウェブページなど，その有用度がユーザごとに異なるもの（アイテム）であり，「有用な情報を選択する」とは，ユーザにとって未知なアイテムの有用度を推定し，有用度の高いアイテムを提示（推薦）することである[8),9)]．

協調フィルタリングとは，広義には，未知のアイテムに対する有用度を，既知のアイテムに対する有用度から類推して推薦を行うものであり，アイテムの内容に依存することなくユーザの嗜好に基づく推薦を可能

† 東京工業大学
Tokyo Institute of Technology
†† 富山大学
University of Toyama
††† メディア教育開発センター
National Institute of Multimedia Education

とするものを指す．本稿では，文献 6) などに示されているような，ユーザなどを単位とした類似度の算出および有用度の類推のアルゴリズムまで含めた狭義の協調フィルタリングのことを単に協調フィルタリングと呼ぶことにする．このときの協調フィルタリングのアルゴリズムは次のとおりである．

(1) ユーザのアイテムに対する有用度（購入した/しなかった，閲覧した/しなかった，5 段階評価など）を収集し，有用度を要素とするユーザ–アイテム行列を作成．
(2) 有用度の行列に基づきユーザ間の類似度を算出．
(3) 類似度に基づき当該ユーザに対するアイテムの有用度を類推・アイテムを推薦．

協調フィルタリングではアイテム数がユーザ数に対して大きすぎるとき，ユーザから収集した有用度を要素とするユーザ–アイテム行列はスパース（疎）となり，類似度が高くなるユーザを得られないため，有効な推薦ができないことが指摘されている[7]．ここで web コンテンツを対象とした協調フィルタリングを考えてみると，アイテムとなる web コンテンツの数は膨大であり，推薦に十分となる量の有用度をユーザから収集することは現実的に不可能である．

このような大量のアイテムデータを対象に有効な推薦を実現するためには，ユーザの嗜好に基づいて類似アイテムを集約することで見た目のアイテム数を減らす，ユーザの嗜好する範囲にアイテムを限定するなど，協調フィルタリングを行う事前処理としてアイテム数を圧縮（以下，アイテム圧縮）する必要がある．しかしながら，ユーザの嗜好は多様であり，事前処理としてアイテム圧縮することは困難である．そのため，ユーザの検索要求に基づいてそのつどアイテム圧縮を行うことができる手法が必要であるといえる．

一方，近年普及しつつあるソーシャルブックマーク（以下，SBM）サービス[10),11)] を利用した web コンテンツ推薦手法が提案されている[12)]．SBM は WWW 上でブックマークを管理および共有できるサービスであり，その利便性により現在急速に利用者数を増やしている．このサービスの最大の特徴として，web コンテンツに「blog」,「面白い」などといった自由記述によるキーワードを付与できることがあげられる．このキーワードはタグと呼ばれる．単一ユーザのブックマークにおいて，タグは web コンテンツを分類する側面と，文字列による注釈という側面とをあわせ持つ．本稿では両者を区別するため，前者の文脈では単に**タグ**，後者では**タグの名称**と使い分ける．

これまでに，同一の文字列の意味の共通性に着目して，ユーザ間のタグの名称の使用履歴の類似性を利用した web コンテンツ推薦[12)] が試みられている．文献 12) では，ユーザ間におけるタグの名称のゆらぎについて考慮しており，類似するタグの名称をクラスタリングすることによってこの問題の解決を図っている．しかし，タグの名称は同一の文字列にもかかわらず，ユーザ間で異なる分類を想定して用いられている場合には，適切な推薦ができない．このとき，あるユーザがあるタグの名称に対して，SBM の分類のために思い描いている概念のことを**タグの名称に対する概念**と呼ぶこととすると，この問題はタグの名称の概念がユーザごとに異なること，すなわちタグの名称の持つ多義性に起因する．

本研究では，タグが持つ「web コンテンツの分類を行う」という側面のみに着目することにより，タグの名称の持つ多義性を原理的に問題としない web コンテンツ推薦システムを提案する．つまり,「あるユーザによって同一のタグを付与された web コンテンツ集合（以下，コンテンツクラスタ）は，当該ユーザの概念において意味的に共通な性質を持つと判断されたものである」という発想に基づき，コンテンツクラスタ間における web コンテンツの共起性のみを利用した推薦システムを提案する．

すなわち本手法では，タグをコンテンツクラスタを表すうえでの表象としてのみ利用し，タグの名称としての類似性はまったく考慮しない．この提案手法は，SBM の特性を利用したコンテンツクラスタベースの（広義の）協調フィルタリングであるといえる．

また，本研究ではコンテンツクラスタの類似性を検定問題として定式化することにより，コンテンツクラスタ間の類似度を統計的な信頼性を考慮した定義として提案する．すなわち,「コンテンツクラスタは web 全体を母集団とする web コンテンツ群から，ユーザがある概念に基づいてサンプリングを行った結果である」と考える．

このとき，2 つのコンテンツクラスタの類似度を，これらのコンテンツクラスタにおけるサンプリングが類似概念で行われたか/異なる概念で行われたかを検定することにより算出する．具体的には，2 つのコンテンツクラスタが類似概念/異なる概念それぞれに基づくときの web コンテンツのサンプリングが一致する確率を仮定することでそれぞれの尤度を算出し，この尤度比を類似度として用いる．

本研究の全体の構成を以下に示す．まず 2 章で関連研究および SBM の概要と一般的な特徴を紹介する．次に 3 章で SBM における web コンテンツ推薦アル

ゴリズムを提案し，4 章において検証実験およびその結果を示す．5 章では本手法の考察を行い，最後に 6 章でまとめる．

2. 関 連 研 究

本章では，関連研究として SBM を用いた web コンテンツ推薦の既存研究および web コンテンツの共起性に基づいた web コンテンツ推薦の既存研究について述べる．関連研究の紹介に先立ち，SBM について説明を行う．

2.1 SBM とは

2.1.1 Social Bookmark（SBM）

SBM とは，従来ローカルで行っていたウェブページのお気に入り登録であるブックマークを WWW 上で行うサービスの総称である．実サービスでは，del.icio.us[10]，はてなブックマーク[11] などが急速に普及している．これら SBM サービスに共通する最大の特徴として，web コンテンツにタグを付与できることがあげられる．**図 1** は，SBM においてユーザが公開しているブックマーク一覧ページのイメージである．このページの web コンテンツ情報には，web コンテンツへのリンクが張られているだけでなく，そのユーザが同じタグを付与した web コンテンツを参照できるページへのリンクや，同じ web コンテンツをブックマークしたユーザを参照できるページへのリンクなども提供されている．これらを参照することによって，自分の興味に合致した情報を探し出すことが可能となる．

2.1.2 Folksonomy

図 2 は，SBM のユーザとタグおよび web コンテンツとの関係モデルである．SBM におけるユーザのタグ付与行動は，図のように，ユーザが生成したタグと web コンテンツとをリンクしていると見なせる．この図においてある特定の web コンテンツに着目すると，ユーザによってタグが複数付与されていることが分かる．このとき，当該 web コンテンツは複数のユーザのタグ付与行動によって意味付けされたと見なすことができる．

このような，ユーザがタグを媒介として，協調的に web コンテンツに意味づけを行うというコンセプトを，一般に Folksonomy と呼ぶ．Folksonomy とは，Folks（みんなの）と Taxonomy（分類学）から作られた造語であり，多数のユーザの自由記述によるタグ付与行動によってボトムアップ的に web コンテンツの分類が行われるというコンセプトを表す[14]．

ウェブディレクトリなどの従来の分類法では，サイト運営者などがトップダウン的に web コンテンツを分類する必要があるため，(1) 分類のための多大なコストを運営者が負わなければならない，(2) 利用者の多様な要求に対応することが困難である，といった問題があった．これに対し，Folksonomy はトップダウンによる分類法に比べて，(1) 利用者に分類におけるコストを分散できる，(2) 利用者の意見に即した分類がなされる，といった利点があるとされている．

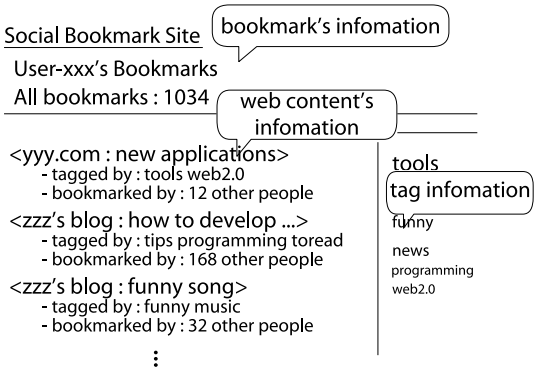


図 1 一般的なソーシャルブックマークサービスの使用イメージ
Fig. 1 Typical schematic of social bookmark service (SBM).

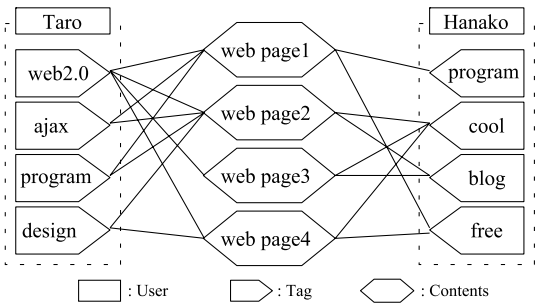


図 2 SBM モデル
Fig. 2 Relationship among users, tags, and contents on SBM.

2.2 SBM を用いた web コンテンツ推薦の既存研究

文献 12) では，Folksonomy の考えに基づいた web コンテンツ推薦システムが提案されている．この手法では，ユーザが SBM サービスにおいて付与してきたタグの履歴と，web コンテンツに付与されてきたタグの履歴とを比較参照し，それぞれが類似すると判定されたとき当該 web コンテンツを推薦する．つまり，利用したタグを媒介とした web コンテンツ推薦システムであるといえる．

しかしながら，ユーザのタグ付与行動を検証すると，タグの名称に基づいた推薦システムは必ずしも有効と

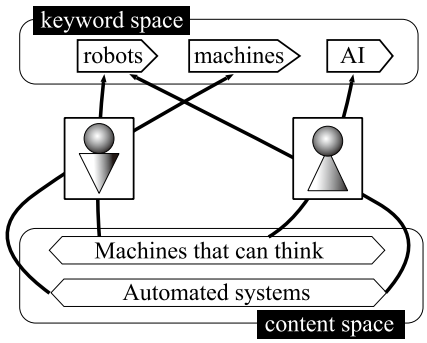


図 3 コンテンツとキーワードとの結び付き
Fig. 3 Relationship between contents and keywords.

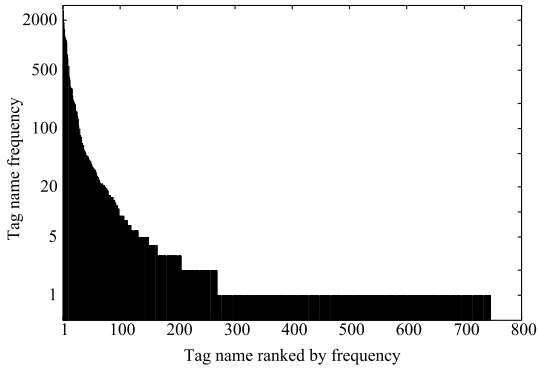


図 4 タグの名称付与行動の頻度分布
Fig. 4 Tag name frequencies sorted by descending order.

はいえないことが分かる．ここで，あるユーザがある web コンテンツ集合に対して同一のタグを付与する行動について考えてみる．この行動は，当該ユーザの概念においてそれら web コンテンツ集合が意味的に共通な性質を持つと考えたために行われたものである．いい換えると，あるユーザによって同一タグが付与された web コンテンツ集合は，当該ユーザの概念において意味的に共通な web コンテンツ集合であると見なせる．

ただし，ユーザは任意のタグの名称を用いるため，ある web コンテンツ集合に対して複数のユーザが共通の概念を持ってタグの付与を行ったとしても，それらのタグの名称がユーザ間で一致するとは限らない．**図 3** は，コンテンツ空間上の web コンテンツを，複数のユーザがキーワード空間上のタグへとマッピングする様子を示したものである．この例のように，各ユーザが「Machines that can think」という共通の概念を持つ web コンテンツに対して，あるユーザはタグ「robots」を付与し，他のユーザはタグ「AI」を付与する可能性がある．

また逆に，意味的に異なる web コンテンツ集合に対して同一のタグの名称が付与されることもある．図 3 に示した例のように，同じタグ「robots」に対して，あるユーザは「Machines that can think」という解釈を行い，他のユーザは「Automated systems」という解釈を行う可能性がある．しかしながら従来の web コンテンツ推薦手法では，キーワード空間にマッピングされた結果に基づいて推薦を行うため，ユーザの web コンテンツに対する概念に基づいていない可能性がある．

また，SBM の実サービスのデータからもタグの名称による推薦の問題点を指摘できる．**図 4** は実際の SBM サービスにおいて，ある特定の web コンテンツに付与されたタグの名称の頻度分布を表したものである．この図から分かるように，多くのユーザがタグの名称に選ぶ上位のタグは少数しかないのに対して，少数のユーザしか選ばない下位のタグの名称は数多く存在している．この下位のタグに着目すると，「toread」（あとで読むための印付け）「web/app」（個人でカテゴリ分け）「**Java**」（特殊な表記）などのような個人の管理手段を含んだタグが多く付与されている．このように，実 SBM サービスでは各ユーザが自分の管理欲求に基づいて利己的にタグを付与していることが分かる．

しかしながら，タグの名称による推薦システムは，そもそも各ユーザが協調的にタグの付与を行うという仮定に基づいており，このようなユーザの利己的なタグ付与行動を考慮していない．そのため，もっぱら上位のタグの情報を利用しており，下位のタグの情報は切り捨ててしまうため，下位のタグを付与しているユーザの情報を有効に活用できない．

一方本研究では，タグの名称によらず，コンテンツクラスタ間の web コンテンツの共起性のみに着目した推薦手法であるため，上記のような特殊な表記のデータを有効に活用することが可能である．

2.3 web コンテンツの共起性のみに基づいた web コンテンツ推薦の既存研究

既存研究においても，web コンテンツ集合間の共起性のみに着目した web コンテンツ推薦システムが提案されている[13]．この手法では，ブックマークのフォルダ構造のみによって定義される web コンテンツ集合間において類似度を算出し，類似度の高いブックマークのフォルダに含まれる web コンテンツ群を推薦するものである．

本研究と既存研究とは，ブックマークのフォルダを対象としている点において類似した研究であるといえるが，文献 13) において類似度の具体的な算出方法

表 1 既存手法との比較による本研究の位置づけ
Table 1 Comparison between proposed method and conventional methods.

	協調フィルタリング	既存研究[12]	既存研究[13]	提案手法
システムクエリ	ユーザ	タグ	ユーザ&カテゴリ	ユーザ&タグ
ユーザ間共起	○	×	○	○
タグ情報の利用	なし	名称	集合	集合
推薦ランキング	○	○	×	○
類似度算出法	共起割合	共起割合	詳細なし	尤度比

については言及されていない．また，文献 13) では類似したブックマークフォルダ，つまり，web コンテンツの集合を発見することが目的となっており，個々の web コンテンツに対して推薦順位を与えるものではない．いい換えれば，この手法では同じ web コンテンツ集合に属している複数の web コンテンツが同じ順位で推薦されてしまい，優先すべき web コンテンツを決定することができない．

一方本研究では，コンテンツクラスタ間の類似度の算出方法を定義し，その類似度を利用して個々の web コンテンツごとに推薦度を算出する方式を提案している．

表 1 に既存研究と提案手法との比較を行い，本研究の位置づけを示した．

3. 提 案 法

本章では提案する web コンテンツ推薦手法について説明する．

3.1 概 要

すでに述べたように，本研究ではタグの名称ではなく，ユーザのタグ付与行動によって関連付けられた web コンテンツ集合，すなわち，コンテンツクラスタに着目する．ある 2 つのコンテンツクラスタのいずれかに含まれる各 web コンテンツが，両方のコンテンツクラスタいずれにも帰属するか，または帰属しないかは，二項分布に従うというモデル化を行う．このモデル化によってコンテンツクラスタ間の類似度を定義し，類似度の総和によって個々の web コンテンツの推薦度を算出する．

ここで，本研究における推薦のターゲットは SBM を利用しているユーザであり，当該ユーザがある嗜好を持って共通の名称でタグを付与していることを想定している．また，想定する推薦システムは，推薦を受けたい SBM ユーザが今まで付与してきたタグの中から興味のあるタグの名称を入力することで，該当するコンテンツクラスタ自体がクエリとして解釈され，そのクエリと適合した web コンテンツが推薦されるというものである．

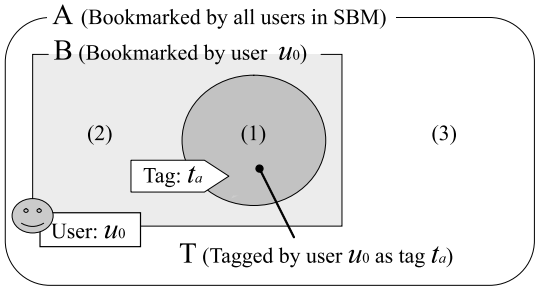


図 5 SBM におけるコンテンツ集合の関係
Fig. 5 SBM modeling about relationship among contents, users and tags.

3.1.1 SBM のモデル化

以下の説明のため，本項ではまず SBM のモデル化を行う．あるユーザのブックマーク行動によって，すべての SBM 内に登録された web コンテンツは「ブックマークされている/ブックマークされていない」のどちらかに分類される．また，ユーザによるタグの付与は，ブックマークしている web コンテンツに対して，タグを表象とする集合へ「帰属させる/帰属させない」という二者択一を再度行っていることと見なせる．つまり，ある特定のユーザ u_0 の，ある特定のタグ t_a に注目すると，SBM に登録されている全コンテンツ集合 A 内の web コンテンツは以下のいずれかに属している（**図 5**）．

(1) ユーザ u_0 にブックマークされており，かつ，タグ t_a が付与されている（$B \cap T$）．
(2) ユーザ u_0 にブックマークされているが，タグ t_a は付与されていない（$B \cap \overline{T}$）．
(3) ユーザ u_0 にブックマークされていない（$\overline{B}$）．

以下本稿では，ユーザ u_0 によってタグ t_a を表象として結び付けられる web コンテンツ群を，u_0 の t_a による**コンテンツクラスタ**と呼ぶこととする．

本研究では，コンテンツクラスタの帰属関係が明示的に示されている集合 (1) および (2) の情報をもとに，(3) にあたる未ブックマークの web コンテンツが当該コンテンツクラスタに帰属する可能性を提案アルゴリズムを用いて推薦度として算出する．

以下では，(1) に含まれる web コンテンツをコン

テンツクラスタに属する web コンテンツと呼び，(2) に含まれる web コンテンツをコンテンツクラスタに属さない web コンテンツと呼ぶことにする．また，(3) に含まれる web コンテンツは，本提案における推薦対象となる web コンテンツである．

提案法のアルゴリズムは以下のステップで構成される．

(1) 二項分布に基づく尤度の算出
(2) 仮説検定に基づくコンテンツクラスタ間の類似度算出
(3) web コンテンツの推薦度算出

3.2 各ステップの説明

3.2.1 二項分布に基づく尤度の算出

web コンテンツの全体集合 A から，任意に複数個の web コンテンツをサンプリングして生成した部分集合をそれぞれ独立に 2 つ考える．これら部分集合を C_1，C_2 とする．ここで，全体集合 A の中から取り出した任意のコンテンツが，これらの和集合 $C_1 \cup C_2$ に含まれる事象を X_1，$C_1 \cup C_2$ の中から取り出した任意の web コンテンツが，積集合 $C_1 \cap C_2$ に含まれる事象を X_2 とおく．このとき，和集合 $C_1 \cup C_2$ から取り出した任意の web コンテンツが，積集合 $C_1 \cap C_2$ にも含まれる確率 p は以下の式で表せる．

$$\begin{aligned} p &= P(X_2|X_1) = P(X_1 \cap X_2)/P(X_1) \\ &= P(X_2)/P(X_1) \end{aligned} \quad (1)$$

これに対し，和集合 $C_1 \cup C_2$ から取り出した任意の web コンテンツが，積集合 $C_1 \cap C_2$ に含まれない確率 q は以下の式で表せる．

$$\begin{aligned} q &= P(\overline{X_2}|X_1) = P(X_1 \cap \overline{X_2})/P(X_1) \\ &= P(X_1) - P(X_1 \cap X_2)/P(X_1) \\ &= 1 - p \end{aligned} \quad (2)$$

また，$X_2|X_1$ と $\overline{X_2}|X_1$ は明らかに排反事象である．

以上により，「和集合 $C_1 \cup C_2$ から任意の web コンテンツを取り出したとき，これが積集合 $C_1 \cap C_2$ に含まれるか含まれないかを確認する」試行はベルヌーイ試行である．以下では，積集合 $C_1 \cap C_2$ に含まれたときを成功，含まれなかったときを失敗と呼ぶこととする．一般に，ベルヌーイ試行を n 回行ったときに k 回成功する尤度 $L(n,k,p)$ は，試行回数 n，成功確率 p の二項分布に従う．よって，尤度は以下の式で表される．

$$L(n,k,p) = {}_n\mathrm{C}_k \; p^k \, (1-p)^{n-k} \quad (3)$$

いま，ユーザ u_0 がある特定のタグ t_{query} についての推薦を受けたいものとする．このとき，u_0 によってブックマークされた web コンテンツ全体の集合を B_s，u_0 によって t_{query} を付与された web コンテンツ全体の集合，すなわち推薦先となるコンテンツクラスタを T_s とし，T_s を**クエリコンテンツクラスタ**と呼ぶこととする．本手法では，ユーザ u_0 とは異なる任意のユーザ u_i の任意のタグ t_j によるコンテンツクラスタが，クエリコンテンツクラスタに対する推薦の度合いを決定するもととなる．そこで u_i によってブックマークされた web コンテンツ全体の集合を B_o，u_i によって t_j を付与された web コンテンツ全体の集合，すなわち推薦元となるコンテンツクラスタを T_o とし，T_o を**推薦元コンテンツクラスタ**と呼ぶこととする．

仮に，ユーザ u_0 がすべての web コンテンツをブックマークしたとすると，すべての web コンテンツはタグ t_a を付与されるか，されないかの 2 つに分類できる．このとき，t_a が付与される web コンテンツの集合を，以下本稿では u_0 の t_a に対応する**理想コンテンツクラスタ**と呼ぶ．いま，ある web コンテンツに対するタグの付与が，他の web コンテンツに依存することなく独立に決定されると仮定する．このとき，2 つの理想コンテンツクラスタの和集合からサンプリングした web コンテンツがその積集合に帰属することを成功と見なすと，このサンプリングは，2 つの理想コンテンツクラスタの共起確率を成功確率としたときのベルヌーイ試行であるといえる．また，ブックマーク集合 B は web コンテンツの全体集合からサンプリングの結果であると見なせる．

以上の議論の結果として，本手法では 2 人のユーザのブックマークの積集合 $B_s \cap B_o$ に含まれる web コンテンツが，双方の理想コンテンツクラスタへの帰属/非帰属の関係を示していると見なす．すなわち，この積集合に含まれる web コンテンツを，上記のベルヌーイ試行を行ったときのサンプリングの結果と見なして利用する．

2 つのコンテンツクラスタ T_s，T_o において，該当する 2 人のユーザのブックマークの積集合 $B_s \cap B_o$ の要素のうち，2 つのコンテンツクラスタの和集合 $T_s \cup T_o$ に含まれる web コンテンツの個数 $|(B_s \cap B_o) \cap (T_s \cup T_o)|$ を試行回数 $n(T_s, T_o)$ とし，積集合 $T_s \cap T_o$ に含まれる web コンテンツの個数 $|T_s \cap T_o|$ を成功数 $k(T_s, T_o)$ とする（**図 6**）．

$$n(T_s, T_o) = |(B_s \cap B_o) \cap (T_s \cup T_o)| \quad (4)$$

$$k(T_s, T_o) = |T_s \cap T_o| \quad (5)$$

以上より，ある 2 つのコンテンツクラスタ T_s，T_o において，$n(T_s, T_o)$ と $k(T_s, T_o)$ が観測されたとき，

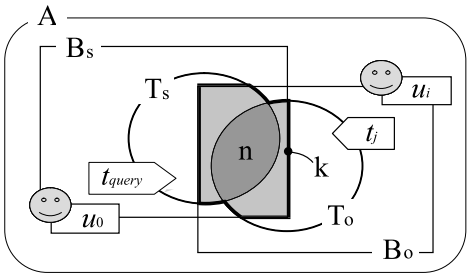


図 6 コンテンツクラスタの比較とコンテンツ推薦
Fig. 6 Content recommendation by comparison of contents clusters.

T_s および T_o と対応する理想コンテンツクラスタ間の共起確率が p であった尤度 $L(n(T_s,T_o),k(T_s,T_o),p)$ は以下の式で求められる．

$$L(n(T_s,T_o),k(T_s,T_o),p) = {}_{n(T_s,T_o)}\mathrm{C}_{k(T_s,T_o)} \times p^{k(T_s,T_o)}(1-p)^{n(T_s,T_o)-k(T_s,T_o)} \quad (6)$$

3.2.2 仮説検定によるコンテンツクラスタ間の類似度算出

このステップでは，前項で得られた尤度を用いて一致度の仮説検定を行う．すなわち，2 つのコンテンツクラスタ間の関係において，以下の 2 つの仮説を設定する．

- 一致度 $p = p_1$：共通の概念に基づく
  共通の概念に基づいて構成されたコンテンツクラスタどうしでは，web コンテンツが高確率で同時に帰属する．
- 一致度 $p = p_0$：異なる概念に基づく
  異なる概念に基づいて構成されたコンテンツクラスタどうしでは，web コンテンツが低確率で同時に帰属する．

ただし，$p_1 > p_0$ であるとする．ここで，2 つのコンテンツクラスタ間の関係が，「共通の概念に基づく」・「異なる概念に基づく」のいずれかに分類されるとすると，観測結果よりそのどちらに近いかを尤度比の大きさによって算出することが可能である．以下に，尤度比を算出する式を示す．

$$\begin{aligned} &sim(T_s,T_o) \\ &= \log \frac{L(n(T_s,T_o),k(T_s,T_o),p_1)}{L(n(T_s,T_o),k(T_s,T_o),p_0)} \\ &= k(T_s,T_o)\log\frac{p_1}{p_0} \\ &\quad +(n(T_s,T_o)-k(T_s,T_o))\log\frac{1-p_1}{1-p_0} \end{aligned} \quad (7)$$

式 (7) において $sim(T_s,T_o)$ は，値が大きいほど共通の概念に基づいていると検定される．本研究では，この対数尤度比 $sim(T_s,T_o)$ をコンテンツクラスタ間の**類似度**として利用し，この類似度が高いコンテンツクラスタから web コンテンツの推薦を行う．すなわち，T_o は T_s に対し，$\overline{T_s} \cap T_o$ に含まれる web コンテンツを類似度 $sim(T_s,T_o)$ によって推薦する．

ここで，本研究では p_0，p_1 の値として，多くの事例で良い結果が得られた数値 $p_0 = 0.1$，$p_1 = 0.6$ を採用した．このパラメータの最適値は，個々の事例ごとに存在するが，事前の検証実験において，上記のパラメータの付近では推薦精度に大きな差異がないことが示されている．

3.2.3 web コンテンツの推薦度算出

このステップでは，クエリコンテンツクラスタ T_s に対する web コンテンツ c（推薦対象は図 6 における $\overline{B_s} \cap T_o$ である）の推薦度 $R(T_s,c)$ を，c を帰属している任意の推薦元コンテンツクラスタ $T_{o_i}(\forall i, c \in T_{o_i}, i = 1,2,3,\cdots,k)$ との類似度 $sim(T_s,T_{o_i})$ の和によって定義する．

$$R(T_s,c) = \sum_{i=1}^{k} sim(T_s,T_{o_i}) \quad (8)$$

ただし，類似度が $sim(T_s,T_{o_i}) < 0$ となる場合においては，前項の式 (7) の仮説検定において p_0 と判定されたと見なし，和に加えないこととする．

以上によって算出された値 $R(T_s,T_c)$ を，T_s に対する web コンテンツ c の**推薦度**と定義する．これに基づき，ランキングや閾値などの処理によって web コンテンツの推薦を行う．

4. 実　　験

del.icio.us [10] から取得したデータを用い提案方式の実証実験を行った．同サービスは，世界中の多くのユーザに利用されている SBM の代表的な存在である．

4.1 実 証 実 験

この実証実験では，2006 年 8 月時点における同サービスに登録しているユーザの情報を収集し，1,000 人分のブックマークデータを抽出した．このデータにおいて，総 web コンテンツ数（ユニーク URL 数）は約 310,000 である．また，各ユーザが使用したタグ数の平均は約 260 個であった．すなわち，このデータセットにおける総コンテンツクラスタ数は約 260,000 であった．本実験では，この約 260,000 件のコンテンツクラスタから被推薦対象であるユーザのコンテンツクラスタを除いたすべてを用いて推薦を行う．

4.1.1 実 験 方 法

本提案アルゴリズムは，あるクエリコンテンツクラスタ T_s においてブックマークされていない web コン

表 2 本実験処理後の web コンテンツの分類
Table 2 Classification of experimental results for calculating recall and precision.

	推薦集合 R	非推薦集合 r
正解集合 X	正解 RX	推薦漏れ rX
不正解集合 x	不正解 Rx	–

テンツの推薦を行うものである．しかしながら，ある web コンテンツが推薦されたとき，それが当該ユーザの所望するものであるかどうかを客観的に判定することは困難であるため，ここではすでに帰属関係が示されている web コンテンツ群，すなわち，当該ユーザのブックマーク B_s に含まれる web コンテンツ群に対して推薦度を算出することで，推薦結果と元のコンテンツクラスタとの web コンテンツの一致度によって推薦精度の検証を行う．

実験方法は以下のとおりである．

(1) 検証するクエリコンテンツクラスタ T_s を選択，T_s に帰属している web コンテンツを正解集合 X，T_s に帰属していない web コンテンツ（$B_s \cap \overline{T_s}$）を不正解集合 x とする．

(2) 他のすべてのユーザによるコンテンツクラスタ T_{o_i}（$i = 1, 2, 3, \cdots$）を対象に，T_s と T_o との間で類似度 $sim(T_s, T_{o_i})$ を算出．

(3) 算出した類似度をもとに，B_s に含まれる web コンテンツの推薦度を算出，上位数件を推薦集合 R，それ未満を非推薦集合 r とする．

(4) 元の T_s および B_s と推薦結果を比較，recall および precision を算出．

ただし，実験後の各 web コンテンツは，**表 2** のいずれかのカテゴリに分けられるので，recall，precision は次の式で与えることとする．

$$\text{recall} = \frac{RX}{RX + rX} \tag{9}$$

$$\text{precision} = \frac{RX}{RX + Rx} \tag{10}$$

また，推薦精度の検証基準として F-measure を用いる．F-measure は次の式で与えられる．

$$\text{F-measure} = \frac{2 \cdot \text{recall} \cdot \text{precision}}{\text{recall} + \text{precision}} \tag{11}$$

今回の実験では，コンテンツクラスタに帰属する web コンテンツ数の大きい順に上位 100 件のコンテンツクラスタをクエリコンテンツクラスタとして実験を行った．これらのコンテンツクラスタにおいて，このコンテンツクラスタを構成した（すなわち，タグを付与した）ユーザのブックマーク数は最大 17,960 件，最小 865 件，平均は 6758.4 件であり，コンテンツクラスタに帰属する web コンテンツ数は最大 1,991 件，最小 477 件，平均は 729.28 件であった．

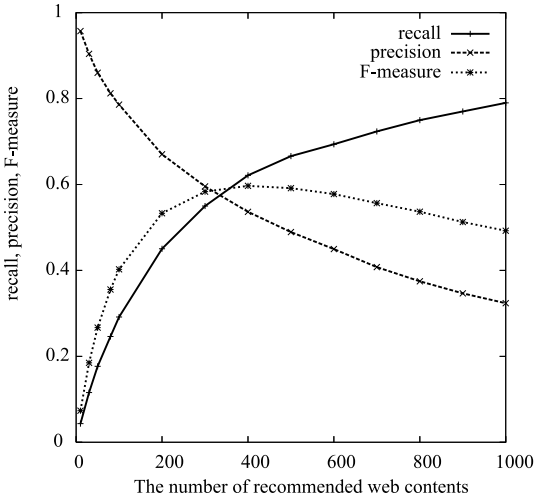


図 7 推薦件数と recall，precision，F-measure の関係
Fig. 7 Recall, precision and F-measure for each number of recommended contents.

上記実験では，推薦対象とあるコンテンツ自体が類似度計算に利用されるため若干の誤差は生じるが，今回の実験で扱ったデータの web コンテンツ数はいずれも十分大きいため誤差は小さいと考えられる．

また，たとえ正解集合に含まれる web コンテンツであったとしても，今回構築したデータベース中において他のだれもブックマークしていないものはそもそも推薦集合にはなりえない．これは一般的な協調フィルタリング手法においても推薦できない web コンテンツであり，本手法における問題ではなく，収集したデータベースの規模に基づく問題である．以上の理由により，以下の実験結果では推薦不可能な web コンテンツとしてこれらを除去して考えることとする．元のコンテンツクラスタ T に対し，推薦不可能な web コンテンツを除去したコンテンツクラスタを T_{clear} と表記することとする．

4.1.2 実験結果

図 7 に，推薦件数を変化させたときの，recall，precision および F-measure の前述の上位 100 件の平均との関係を示す．ただし，横軸の件数分だけ推薦ができないコンテンツクラスタに関しては除外したうえで平均を算出している．

次に，コンテンツクラスタの規模と推薦精度の関係性を検証するため，**図 8** において，コンテンツクラスタに帰属する推薦可能な web コンテンツの数と，各コンテンツクラスタにおいて推薦件数を変更して最大値となるときの F-measure との関係を示す．

表 3 実験結果におけるコンテンツクラスタの詳細
Table 3 Details of contents clusters obtained from the experimental results.

	cluster1	cluster2	cluster3	cluster4	cluster5	cluster6	cluster7	cluster8
ユーザ	user316	user87	user796	user878	user555	user313	user190	user51
$\|B\|$	2006	1632	2242	10078	17965	4782	3925	2701
タグの名称	web	randomlink	music	javascript	art	History	Shopping	funny
$\|T\|$	791	1059	681	845	863	633	547	544
$\|T_{clear}\|$	486	178	52	424	448	209	139	147

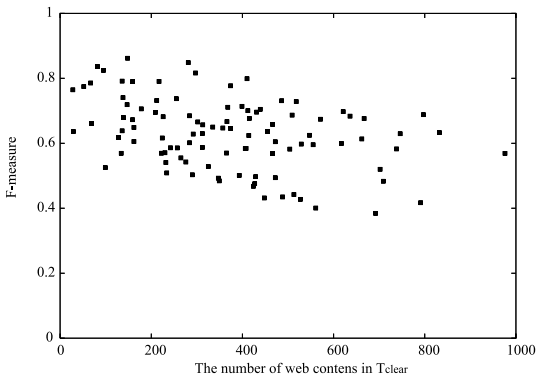


図 8 コンテンツクラスタの規模と F-measure（最大値）の関係
Fig. 8 Relationship between maximum of F-measure and number of contents included in contents cluster.

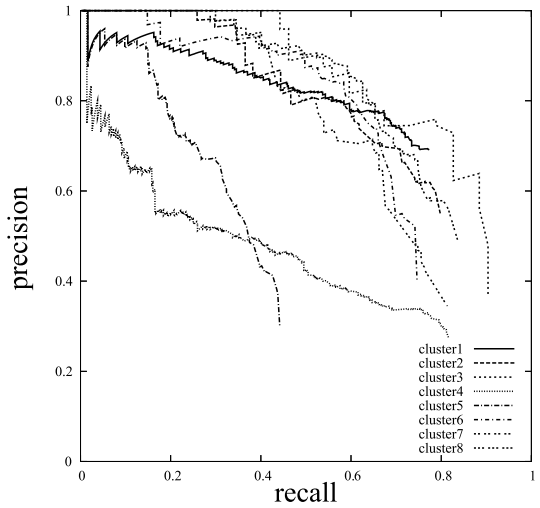


図 9 個別のコンテンツクラスタに対する recall-precision の関係
Fig. 9 Relationship between recall and precision for each contents cluster.

より詳細な考察のため，いくつかのコンテンツクラスタを事例として取り上げ，検証する．**図 9** は，検証したコンテンツクラスタごとに推薦件数を変化させることによって recall と precision を変化させたものである．ただし，各曲線は 1 つのクエリコンテンツクラスタにおける推薦結果であり，**表 3** の cluster1〜8 に対応している．

4.1.3 考　　察

図 7 より，上位 100 件の web コンテンツを推薦したときの平均で precision が 0.79，200 件のとき 0.67 と，web コンテンツ推薦システムの性能として十分な高い値が得られたといえる．また，推薦性能の総合的指標となる F-measure は 400 件付近のとき 0.60 とピークを示し，その後なだらかに下降している．これは，推薦件数を推薦システムとして通常利用されないほど増加させても十分な性能を保つことを示し，本方式の性能の高さを示している．

図 8 からはさほど強い相関性は見られない．よって本手法は，コンテンツクラスタの規模，すなわち，SBM ユーザのタグ付与数によらず，ロバストな性質を示すことが分かる．つまり提案手法は，少量の web コンテンツにしかタグを付与していない SBM ユーザに対しても有用な web コンテンツを推薦できるといえる．

図 9 を見ると，cluster1〜3 および 6〜8 に関しては高い精度が得られていることが分かる．表 3 からも分かるようにこれらのコンテンツクラスタは規模には大きなばらつきがあるものの，いずれの場合も高い精度が得られており，提案手法のロバスト性が確認できる．

また，タグの名称は多岐にわたり，分野の偏りは見られない．良好な結果を示したタグの名称を見ると，ソフトウェア関係と見られる「web」，ユーザの意図がまったく不明な「randomlink」，趣味性が高いと考えられる「music」，「Shopping」，広範な分野を示す「History」，感性的用語とされる「funny」である．この中でも，「randomlink」，「funny」などは人間でも該当する web コンテンツを推測することが難しい．また，「music」，「Shopping」，「History」は，仮に一般的な用語の意味として分類が行われていたとしても，個人の興味範囲は特定の分野に限定されると思われるため，やはりタグの名称からの推測は困難である．

このような事例に対しても有効な推薦が可能となっていることは，タグの名称によらない手法であるからこその結果であろう．これらはクエリコンテンツクラスタと同様の概念で構成された推薦元コンテンツクラスタがデータセット内に存在し，それらから有用な

表 4 比較実験
Table 4 Comparative experiments.

	比較 1（Folksonomy）	比較 2（Jaccard）	提案手法
システムクエリ	タグ	ユーザ&タグ	ユーザ&タグ
ユーザ間共起	×	○	○
類似度算出法	ユーザ数	共起割合	尤度比

web コンテンツが推薦されたため，良い結果が得られたと考えられる．

一方，cluster4 は precision が低く，cluster5 は recall が低くなった．

cluster4 のタグの名称は「javascript」であるが，類似度が高い推薦元コンテンツクラスタのタグの名称には「programming」のような，より広範囲な概念でタグ付けしていたと思われるものが多く見られた．cluster4 については，検証したコンテンツクラスタに比べ，広い概念に基づくコンテンツクラスタがデータセット内に多く存在したため，不正解である web コンテンツが過剰に推薦されて precision が低くなったものと思われる．

また，cluster5 のタグの名称は「art」であるが，推薦元コンテンツクラスタは「webdesign」というタグの名称に見るように，より限定的な概念でタグ付けしていたと思われるものが多く見られた．cluster5 については，検証したコンテンツクラスタに比べ，狭い概念に基づくと解されるコンテンツクラスタがデータセット内に多く存在したため，推薦されるべき正解の web コンテンツが推薦されず，recall が低くなったものと思われる．

これらは，今回用いたデータ内に類似度の高いコンテンツクラスタが存在しなかったことが原因であり，本手法では対応できなかったものと考察される．つまり，今回の実験において構成したデータベース中には同様の概念で構成された推薦元コンテンツクラスタが存在しておらず，有効な推薦が不可能であったことに起因する．

この問題に対しては，協調フィルタリングの事前処理と同様に，適切なアイテム圧縮を行うことによって改善が見込まれると考えられる．また筆者らは，文献 15) においてコンテンツクラスタの論理演算の検討を行っている．これは，推薦元コンテンツクラスタの和集合・差集合などのコンテンツクラスタの合成を行うことで，不足していた広い概念や狭い概念に基づく推薦元コンテンツクラスタの生成を行おうとするものである．これにより上記の問題を解決できる可能性があるが，コンテンツクラスタの論理演算の提案手法への導入は今後の課題である．

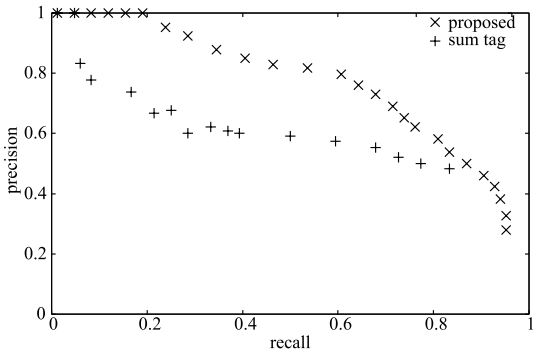


図 10 比較 1：タグの名称を利用する手法と提案手法との比較
Fig. 10 Comparison 1: Proposed method vs. method using tag names.

表 5 比較実験結果におけるコンテンツクラスタの詳細
Table 5 Details of contents clusters in the comparative experiment.

	cluster9
ユーザ	user0
$\|B\|$	2118
タグの名称	ajax
$\|T\|$	117
$\|T_{clear}\|$	84

4.2 比較実験 1：Folksonomy に基づく推薦との比較

本節では提案手法の有効性を検証するため，タグの名称を用いた推薦手法を比較対象として実験を行う．比較手法と提案手法の違いを**表 4** に示す．

4.2.1 実験方法

提案手法では，タグの名称を類似性の評価に利用してしない．一方 Folksonomy では，複数のユーザが付与したタグの名称に従って web コンテンツを分類する．

ここでは，Folksonomy に基づく推薦方法として，多くのユーザが付与したタグの名称に基づく推薦を考える．この推薦方式は以下の手順で行う．

(1) ユーザはクエリとしてタグを入力．
(2) システムはタグが付与された web コンテンツを，多くのユーザが当該タグを付与した順に提示．

4.2.2 実験結果

図 10 において，タグの名称を利用する手法と提案手法との比較結果の一例を示す．今回比較検討に利用したコンテンツクラスタは，**表 5** における cluster9

である．

比較手法よりも本手法が優れた推薦精度が得られていることが分かる．この結果から分かるように，多くのユーザが同一の名称のタグを付与している web コンテンツでも，検証に利用したコンテンツクラスタを構成したユーザにおいてそのタグを付与していないことが多々あった．ちなみに，検証データのタグの名称「ajax」は，web サービスにおける非同期通信技術を指す言葉であり，ユーザ間で「ajax」に対する概念に大きな差が現れにくいと考えられるが，タグの名称に基づく推薦手法よりも提案手法が優れていることは特筆すべきことである．

ここで，このコンテンツクラスタの内容を分析してみると，検証データのユーザは，ajax 技術を用いることで特別な機能を実現している web コンテンツに対して積極的に「ajax」を付与しており，ただ ajax 技術を盛り込んでいるだけの web コンテンツには「ajax」を付与していないことが分かった．つまり，このユーザは他の「ajax」を付与しているユーザと異なる概念に基づいて「ajax」を付与しており，他のユーザに比べて狭い概念を用いているといえる．提案手法では，タグの名称ではなく，コンテンツクラスタに基づいて推薦するため，このような限定的な概念を持つユーザに対しても精度の高い推薦が可能となっていることが確認できた．

4.2.3 考　　察

ここでは，本手法がタグの名称を用いないことに関する考察を行う．

文献 12) の先行研究では，類似したタグの名称をクラスタ化することで利用者ごとの分類表現のゆらぎの影響を軽減している．しかしながら，3 章でも述べたとおり，ユーザの概念は多種多様である．図 10 からも分かるとおり，「ajax」に対する概念もまた多種多様であり，タグの名称「ajax」が等しく付与されている web コンテンツ群であったとしても，当該ユーザの概念と不一致であることは十分ありうる．

また，先行研究ではタグの名称による意見の集約を行っているため，特殊な表記「java/app」，「***java***」のような少数意見は反映されにくく，特に 1 人しか用いていないようなタグの名称の場合は推薦できない．本研究では，タグの名称による比較をいっさい行っておらず，web コンテンツの共起関係のみによって推薦を決定しているため，タグの名称の制限を受けない．

今回の検証では，あるユーザ A のタグ「web2.0」によるコンテンツクラスタと他のコンテンツクラスタとの類似度を算出したところ，他のユーザ B の保持するコンテンツクラスタでは，「web2.0」よりも「ajax」のほうが大きくなるなどの結果が見られた．また，「java/app」のようにタグを階層化して管理しているユーザに対しても，他のユーザが「java」や「app」を付与した web コンテンツを多く推薦していることを確認した．この結果は，タグの名称によらない推薦方式をとることによってのみ達成できることであると考えられる．

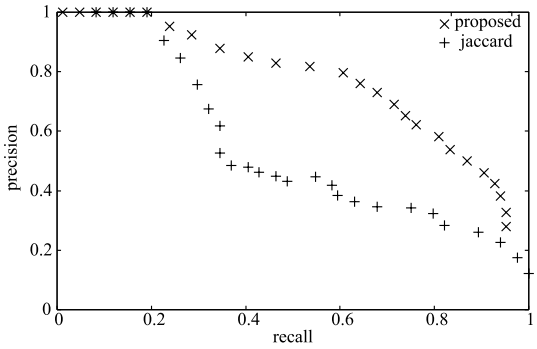


図 11 比較 2：Jaccard 係数による手法と提案手法との比較
Fig. 11 Comparison 2: Proposed method vs. method using Jaccard coefficient.

4.3 比較実験 2：Jaccard 係数との比較

本節では類似度計算において尤度比検定を用いていることの有効性を検証するため，Jaccard 係数を用いた推薦手法を比較対象として実験を行う．

4.3.1 実 験 方 法

協調フィルタリング手法での集合の類似性を比較する尺度として，Jaccard 係数などがある．本手法における類似度と同様の概念を，Jaccard 係数によって定義すると，次の式のようになる．

$$sim_{Jaccard}(s,o) = \frac{k(T_s,T_o)}{n(T_s,T_o)} \tag{12}$$

ただし，$n(s,o)$，$k(s,o)$ は，式 (1)，(2) で定めた値である．

ここで，提案手法の類似度の算出方法を Jaccard 係数に置き換えた手法が考えられる．すなわち，コンテンツクラスタ間の類似度を Jaccard 係数によって求め，各 web コンテンツの推薦度を Jaccard 係数の和によって算出する手法である．これは，一般的な協調フィルタリング手法でユーザ単位であるところを，単にタグ単位に置き換えたものであるといえる．

4.3.2 実 験 結 果

図 11 において，Jaccard 係数による手法と提案手法との比較結果の一例を示す．今回比較検討に利用したコンテンツクラスタは，比較実験 1 同様の表 5 に

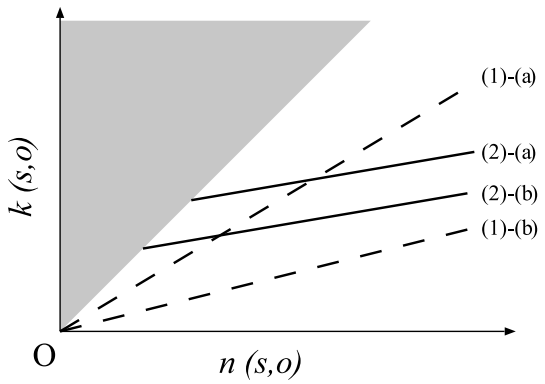


図 12 尤度比と Jaccard 係数との比較
Fig. 12 Numerical comparison between similarity based on hypothesis testing and Jaccard coefficient.

おける cluster9 である．提案手法の推薦精度が，Jaccard 係数に基づく推薦手法と比較して高いことが確認できる．これは，Jaccard 係数に基づく推薦手法では，web コンテンツが偶然一致してしまうコンテンツクラスタとの類似度が結果に大きく影響を与えてしまうことに起因する．

4.3.3 考　　察

ここでは Jaccard 係数よりも尤度比検定を用いたほうが有効であることについて定性的に検証する．

図 12 は，尤度比検定による方式と Jaccard 係数が，それぞれ n 個中 k 個一致するときにおいて，同一の類似度を表す状態を線で表現したものである．(1) の破線は Jaccard 係数，(2) の直線は尤度比による等類似度線であり，それぞれにおいて (a) は類似度 0.6 のとき，(b) は類似度 0.4 のときを示している．

Jaccard 係数は単純な比であるため，n の大小にかかわらず k/n によって類似度が決定される．これは，2 個中 2 個が一致するときと，10 個中 10 個が一致するときとを同一に扱うこととなり，n の小さい推薦元コンテンツクラスタが n の大きい推薦元コンテンツクラスタと同等の影響力を持つことを意味している．また，n の小さい推薦元コンテンツクラスタは n の大きいコンテンツクラスタに比べて数が多い．よって Jaccard 係数による類似度計算では，n の小さい推薦元コンテンツクラスタが悪影響を与え，推薦精度を低下させてしまうと考えられる．

一方，尤度比検定による方式では，n が小さいときはいかなる k に対しても小さい値しかとらず，n が大きいときは k 次第で小さな値から大きな値までをとることが可能となっている．よって，n の小さいコンテンツクラスタにおいては偶発的な web コンテンツの共起による推薦精度への悪影響を除去し，n が大きければは，k に従い類似度を与える計算式であるといえる．

他の Jaccard 係数を用いた既存研究では，閾値処理によってサンプル数が少ないときを除去しているものもある[16]．本方式はこの除去処理を，統計学的なデータの信頼性に基づいて行ったものであるといえる．

5. 考　　察

前章においては，実験の結果から，本手法とタグの名称に基づいた手法との比較，ならびに尤度比検定に基づく類似度計算と Jaccard 係数との比較を行った．ここでは，タグによるコンテンツクラスタを単位とした本手法と従来のユーザを単位とした協調フィルタリングとの比較を行う．

一般にユーザは多種多様な web コンテンツを閲覧しており，学術的な話題から日常的な話題まで様々な分野の web コンテンツを所望する．これらの web コンテンツがすべて SBM に登録されることになれば，研究で使うための技術解説をしているブログ記事，趣味のテニスについてのホームページ，近日計画している旅行先の情報など，これらすべての web コンテンツが単一のユーザによってブックマークされることとなる．

ここで，このユーザ u_1 の各嗜好分野に対応するブックマークの集合をそれぞれ，$B_{u_1}(lab)$，$B_{u_1}(tennis)$，$B_{u_1}(travel)$ と表記することとする．また，異なるユーザ u_2 はテニス，野球，ニュースに興味を持っており，各嗜好分野に対応するブックマーク $B_{u_2}(tennis)$，$B_{u_2}(baseball)$，$B_{u_2}(news)$ をそれぞれブックマークとする．

この 2 人のブックマークを対象に協調フィルタリングを行うと，$B_{u_1}(lab) \cup B_{u_1}(tennis) \cup B_{u_1}(travel)$ と，$B_{u_2}(tennis) \cup B_{u_2}(baseball) \cup B_{u_2}(news)$ との共起関係によって類似度が定義されることとなる．この 2 人のユーザは，テニスという共通した嗜好分野を持つが，嗜好の一致しない分野の web コンテンツがブックマークされていることによって，類似度が低下してしまう．このように，従来法をそのまま用いて SBM 内の web コンテンツ推薦を行ったとしても，これらすべての分野において嗜好が一致するユーザはほとんど存在しないため，良い推薦結果が得られにくい．

一方，提案手法では SBM のタグを用いることで，嗜好分野ごとに類似度を定義することが可能である．SBM ユーザは，後で参照するときの利便性の向上のために，タグを用いてブックマークを管理している．よって，$B_{u_1}(tennis)$ に属する web コンテンツには，

「テニス関連」などといったタグが付与されていることが期待される．その結果，本手法による類似度計算は $B_{u_1}(tennis)$ と $B_{u_2}(tennis)$ のように，嗜好分野を単位とした web コンテンツの共起関係によって定義することが可能となり，従来のユーザ単位よりも高い類似度の web コンテンツ集合を発見することができる．

本手法は，SBM のタグ情報，すなわち，SBM におけるユーザのタグ付与行為を有効に活用した，効率的な web コンテンツ推薦システムであると考えられる．

6. おわりに

本研究では，SBM サービスにおけるユーザのタグ付与行動を，コンテンツクラスタへの帰属関係の意思表明として解釈し，コンテンツクラスタどうしの web コンテンツの共起性に基づいた web コンテンツ推薦システムを提案した．提案手法では，ある 2 つのコンテンツクラスタにおいて，web コンテンツが当該コンテンツクラスタにともに帰属する確率は一致度 p に基づく二項分布に従うという SBM のモデル化を行うことで，信頼性の高いコンテンツクラスタの類似度の算出を可能とした．また，コンテンツクラスタの類似度の算出に仮説検定を導入し，web コンテンツの推薦度を類似度の和とすることで，ユーザのボトムアップ的なタグ付け行動に基づく web コンテンツの推薦を可能とした．実際の SBM サービスから取得したデータに対して提案手法を適用した結果，高い推薦精度が得られることが示された．

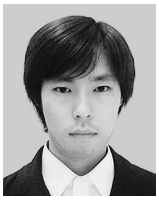
今後の課題としては，概念の大きさに適応した推薦への手法の拡張などが考えられる．

参考文献

1) Li, J. and Zaiane, O.: Combining Usage, Content, and Structure Data to Improve Web Site Recommendation, *Proc. Web KDD-2004 workshop on Web Mining and Web Usage* (2004).
2) Kazienko, P. and Kiewra, M.: Integration of Relational Databases and Web Site Content for Product and Page Recommendation, *International Database Engineering and Applications Symposium* (*IDEAS'04*) (2004).
3) Ishikawa, H., Nakajima, T., Mizuhara, T., Yokoyama, S., Nakayama, J., Ohta, M. and Katayama, K.: An Intelligent Web Recommendation System: A Web Usage Mining Approach, *International Symposium on Methodologies for Intelligent Systems*, pp.342–350 (June 2002).
4) Gunduz, S. and Ozsu, M.T.: A user interest model for web page navigation, *Proc. International Workshop on Data Mining for Actionable Knowledge* (Apr. 2003)
5) Goldberg, D., Nichols, D., Oki, B.M. and Terry, D.: Using collaborative filtering to weave an information tapestry, *Comm. ACM*, Vol.35, No.12, pp.61–70 (1992).
6) Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. and Riedl, J.: GroupLens: An Open Architecture for Collaborative Filtering of Netnews, *Proc. 1994 Computer Supported Cooperative Work Conference*, pp.175–186 (1994).
7) Sarwar, B.M., Karypis, G., Konstan, J.A. and Riedl, J.: Item-based collaborative filtering recommendation algorithms, *Proc. 10th International World Wide Web Conference* (*WWW10*), pp.285–295 (2001).
8) 森田昌宏，速水治夫：情報フィルタリングシステム，情報処理学会，Vol.37, No.8, pp.751–758 (2001).
9) 大杉直樹，門田暁人，森崎修司，松本健一：協調フィルタリングに基づくソフトウェア機能推薦システム，情報処理学会論文誌，Vol.45, No.1, pp.267–278 (2004).
10) del.icio.us. http://del.icio.us/
11) はてなブックマーク．http://b.hatena.ne.jp/
12) Niwa, S., Doi, T. and Honiden, S.: Web Page Recommender System based on Folksonomy Mining, *Proc. 3rd International Conference on Information Technology: New Generations* (*ITNG'06*) (2006).
13) Rucker, J. and Polanco, M.J.: Siteseer: Personalized navigation for the web, *Comm. ACM*, Vol.40, No.3, pp.73–75 (1997).
14) Mathes, A.: Folksonomies — Cooperative Classification and Communication Through Shared Metadata. http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html
15) 佐々木祥，宮田高道，稲積泰宏，小林亜樹，酒井善則：コンテンツクラスタの論理演算を導入したコンテンツ推薦，データ工学ワークショップ (*DEWS2007*) (2007)
16) 松尾　豊，友部博教，橋田浩一，中島秀之，石塚満：Web からの人間関係ネットワークの抽出，人工知能学会論文誌，Vol.20, No.1, pp.46–56 (2005)

(平成 19 年 6 月 20 日受付)
(平成 19 年 10 月 9 日採録)

(担当編集委員　灘本 明代)



佐々木 祥（学生会員）

昭和 55 年生．平成 17 年神奈川大学工学部電気電子情報工学科卒業．平成 19 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．同年より同大学院集積システム専攻博士課程に在学．情報推薦システムに関心を持つ．

宮田 高道（正会員）

昭和 53 年生．平成 13 年富山大学工学部卒業．平成 15 年同大学大学院理工学研究科博士前期課程修了．平成 18 年東京工業大学大学院理工学研究科博士後期課程修了．同年より同大学院集積システム専攻助手．平成 19 年より同助教．画像符号化，画像処理，情報推薦等の研究に従事．博士（工学）．電子情報通信学会，映像情報メディア学会各会員．

稲積 泰宏

昭和 51 年生．平成 10 年富山大学工学部卒業．平成 12 年同大学大学院理工学研究科博士前期課程修了．平成 15 年東京工業大学大学院理工学研究科博士後期課程修了．同年神奈川大学工学部助手．平成 19 年より富山大学大学院理工学研究部（工学）講師．画像符号化，画像処理の研究に従事．博士（工学）．IEEE，電子情報通信学会，映像情報メディア学会，画像電子学会各会員．

小林 亜樹（正会員）

平成 7 年東京工業大学工学部情報工学科卒業．平成 9 年同大学大学院理工学研究科電気・電子工学専攻修士課程修了．平成 12 年同大学院博士課程修了．同年より同大学院集積システム専攻助手．平成 18 年より独立行政法人メディア教育開発センター助教授．平成 19 年より同准教授．画像検索，ネットワーク情報検索，コンテンツ配信の研究に従事．工学博士．電子情報通信学会，映像情報メディア学会，日本データベース学会各会員．

酒井 善則（正会員）

昭和 21 年生．昭和 49 年東京大学大学院工学系研究科電気工学専攻博士課程修了．同年電電公社入社電気通信研究所勤務．ディジタル伝送，マルチメディア通信会議の研究開発に従事．昭和 62 年東京工業大学助教授．平成 2 年より同教授．映像情報伝送，画像情報検索，情報ネットワークの研究に従事．工学博士．平成 13 年電子情報通信学会業績賞受賞．IEEE-COM，CS 各会員．