

ユーザが抱く印象を用いた個人嗜好の表現 —Folksonomy を利用したユーザの印象の推測方法—

田上 道士^{a)}・山場 久昭^{b)}・高塚 佳代子^{c)}・岡崎 直宣^{d)}・富田 重幸^{d)}

Representation of Human Preference Using Users' Impressions —A Method of Inference of Users' Impressions Using Folksonomy—

Michihito TANOUE, Hisaaki YAMABA, Kayoko TAKATSUKA, Naonobu OKAZAKI,
Shigeyuki TOMITA

Abstract

This paper proposes a method to represent human preference using impressions of items derived from users. Since such impressions are not known explicitly, we infer the impressions from information of tagging in a folksonomy. An inferred impression is called a "concept." In this study, we investigate validity of the concepts obtained by our previous method, and propose the new improved derivation method of concepts based on the results of the investigation. Experiments were carried out in order to evaluate the performance of the improved method through comparison with the previous method, and a recommendation experiment was carried out in order to confirm the usefulness of the representation of human preference using the concepts obtained by the new method.

Keywords: Human preference, Information recommendation, Folksonomy

1. 緒言

近年、ユーザにとって有用な情報を自動的に提示する「推薦システム」の重要性が増している。これは、インターネットの普及や情報化技術の発展によって膨大な情報が発信されようになったことに伴い、それら大量の情報の中からユーザが自分の欲しい情報を見付けることが困難となっているためである。この問題に対処するために、情報推薦に関する数多くの研究がなされており、また、様々な Web サービス (例えば amazon) などで推薦システムが実用化されている。

推薦システムが適切な情報推薦を達成するためには、人の好みの適切な表現 (または把握) が非常に重要である。従来の推薦手法の多くでは、推薦者側があらかじめ定義した属性やカテゴリを用いて人の好みが表示されていたが、多数のユーザのアイテムに対する感覚や認識、印象を反映させることによって、より適切な人の好みの表現が期待できる。

以前の研究¹⁾²⁾で、筆者らは Serendipity を考慮した推薦の実現を目指し、「コンセプト」と呼ぶ考えを導入している。この研究では、ユーザ達がアイテムに対して抱く印象を推測し、その推測結果を用いた推薦手法を提案した。この推測した印象が「コンセプト」である。この推測を行うにあたって、

Folksonomy におけるタグの情報を用いている。

本研究では、コンセプトを用いて人の好みを表現することを目指す。すなわち、推薦を受けるユーザ個人の嗜好構造の把握の実現を目指し、Folksonomy における多数のユーザ達によるアイテムへのタグの付与情報から、その背景にあるであろうユーザ達のアイテムに対する印象をボトムアップに浮かび上がらせる。「ユーザ達がアイテムに対して抱く印象」を推測したものであるコンセプトを用いて表現・把握した人の好みは、アイテムに対してユーザ達が感じている多様な視点が反映されているといえるため、「あらかじめ定義された属性」を用いる場合よりも、より適切であると期待できる。

ただし、このような人の好みの表現を実現するためには、コンセプトが妥当なものである必要がある。そこで本研究では、まず、以前提案した手法で得られたコンセプトの妥当性を調べた。その結果、ほぼ期待するようなコンセプトが得られていることを確認できたが、一方で、いくつかのコンセプトは人の好みを表示するのに不十分な点があった。

そこで、コンセプトの精度を向上させるため、コンセプトの形成方法の改良を行った。そして、改良したコンセプト形成方法の性能を評価するために、以前提案した手法との比較実験を行った。さらに、改良手法で得られたコンセプトを用いて表現した人の好みの妥当性を調べるための推薦実験を行った。

2. 人の好みの表現

本研究では、「ユーザ達がアイテムに対して抱く印象」を用いた人の好みの表現を提案する。具体的には、以前の研究で

^{a)}情報システム工学専攻大学院生

^{b)}情報システム工学科助教

^{c)}教育研究支援技術センター技術専門職員

^{d)}情報システム工学科教授

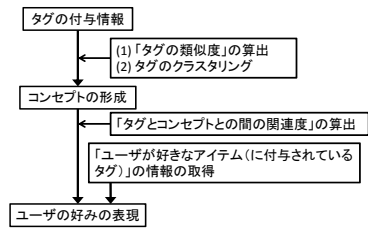


図 1. 本研究における人の好みの表現までの流れ。

Folksonomy におけるタグの情報を利用して導入した「コンセプト」と呼ぶ考え (2.1 節) を活用する。本研究で提案する、人の好みの表現までの全体の流れを図 1 に示す。まず、タグの付与信息を用いて、コンセプトの形成を行う (2.2 節)。そして、得られたコンセプトを用いて、ユーザーが好きなアイテム (に付与されているタグ) に基づき、そのユーザーの好みを表現する (2.3 節)。

2.1 「コンセプト」の概要

2.1.1 コンセプトと呼ぶ考え

従来の推薦手法の多くは、人の好みを表現するにあたって、推薦者側 (例えば、推薦対象となるアイテムに関しての専門家) があらかじめ定義した属性やカテゴリを用いていた。従来の推薦手法は、協調フィルタリングと内容ベースフィルタリングに大別され³⁾⁴⁾⁵⁾、特に後者では、アイテムの内容・特徴に基づいて推薦されるアイテムが選ばれるが、その際にアイテムの特徴は推薦者側があらかじめ定義した属性 (例えば、著者やジャンルなど) を用いて表現されることが多く、このときユーザーの好みもまた、それと同じ属性で表現されている。

しかし、人の好みの表現には、「あらかじめ定義された属性」を用いるよりも、「ユーザー達のアイテムに対する印象」を用いる方がより好ましいといえる。なぜなら、ユーザー達自身を感じている多様な視点が反映されることによって、人の好みより適切な表現が期待できるためである。「人の好みをどのような属性やカテゴリをもって表現するのが最適であるか」は本来は不明確であるため、特定少数の視点でそれを決めてしまうよりも、多数ユーザー達の視点から浮かび上がらせる方がより好ましいと考えられる。

以前の研究で、筆者は Serendipity を考慮した推薦の実現を目指し、ユーザー達がアイテムに対して抱く印象を用いて、アイテムの特徴の表現した。そして、「推薦を受けるユーザーが好きなアイテムに対してユーザー達が抱く印象と、同じ印象がユーザー達に抱かれているアイテム」を推薦する手法を提案した。ここでの「印象」とは、アイテム (の内容や特徴、テーマ、トピックなど) に対してユーザーが感じるもの一つ一つを示す。言い換えると、印象は、ユーザーが感じるアイテムのある一つの特徴、テーマ、トピックなどを指す。例えば、ある本に対して「ミステリー小説だ」「感動するストーリーだ」などが抱かれているとき、それら一つ一つのことを指す。

ただし、実際にユーザーがどのような印象を抱いたかはわからないので、Folksonomy におけるタグの情報を用いて、ユーザー達が抱いた印象の推測を行った。この推測された印象を「コンセプト」と呼ぶ。そして、このコンセプトを用いてアイテムの特徴、及び、ユーザーの好みを表現した。

2.1.2 Folksonomy

Folksonomy とは、従来のトップダウンな方法とは対照的な、ボトムアップ方式の分類法である。従来、ユーザーに提供される情報は、サービス提供者側があらかじめ定義したカテゴリなどに基づいて分類されていた。これに対して Folksonomy では、ユーザー各々が各情報に対して「タグ」を付与することで、情報の分類が行われる。「タグ」とは、その情報の内容や特徴などをもとに、ユーザーが自由に作成したキーワードである。使用できるキーワードに制限はない。また、一つの情報に複数のタグを付与することも許されている。

Folksonomy を利用している Web サービスには例えば、ソーシャルブックマークサービスの「Del.icio.us」や「はてなブックマーク」、動画共有サイトの「ニコニコ動画」、画像投稿コミュニティサイトの「Pixiv」などが挙げられる。

従来の分類法と比較して、Folksonomy は以下のような特徴を持つ。

- ユーザーの感覚や認識構造が分類に反映されやすい。
- タグ付けを行うにあたり、ユーザーは任意のキーワードを入力するだけであり、そこに特別な知識は必要としないため、気軽に行うことができる。

2.1.3 関連研究

情報推薦に Folksonomy におけるタグの情報を用いた研究はいくつか報告されている (例えば⁶⁾⁷⁾⁸⁾⁹⁾¹⁰⁾。Szomszor⁶⁾らは、映画に対するユーザー達の評価値と、その映画に付けられているタグに着目し、それらに基づいて、ユーザーがまだ評価していない映画に対するそのユーザーの評価値を予測する手法を提案した。丹羽ら⁷⁾と Shepitsen ら⁸⁾は、タグの同義性や多義性による問題を避けるために、タグのクラスタリングを導入し、各 Web リソースに付けられたタグが含まれるクラスタに着目した推薦手法を提案した。また、Krestel ら⁹⁾と Said ら¹⁰⁾は、タグとユーザー (或いはアイテム) との間に潜在トピック (隠れ変数) を仮定し、確率モデルを適用した推薦手法を提案した。

筆者の研究では、本来は不明確なユーザー達の印象 (その数も不明) の推測に、タグの情報を用いている。先述したように、Folksonomy による分類にはユーザーの感覚や認識構造が反映されているため、アイテムへのタグ付けの情報から、その背景にあるであろう「ユーザー達がアイテムに対して抱く印象」を推測することは妥当であると期待できる。

2.2 コンセプトの形成

コンセプトの形成は次のように行う (図 2)。「ユーザーはアイテムに対して抱く印象に基づいて使用するタグを選ぶ」ものと仮定する。この仮定から、「ある共通の印象に基づいて使用されているであろうタグ群」を見付け、それらタグから、その背景にある印象を推測する。

- (1) まず、「タグ間の類似度」を定義する。先程の仮定から、ある二つのタグが同じ印象に基づいて使用されているとき、それらタグは共起しやすい (同じアイテムに付与されやすい) と考えられるので、タグ間の共起情報を用いて、この類似度を計算する。

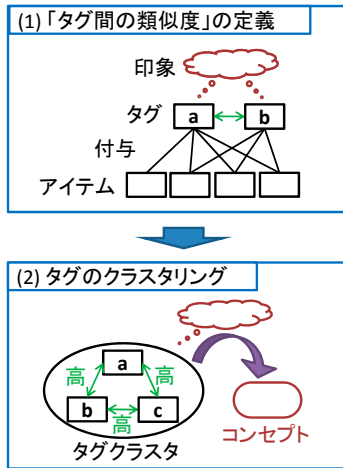


図 2. コンセプトの形成.

(2) 互いに高い類似度を持つタグ同士が集まったクラスタ群を生成する。このようなクラスタ内のタグ群は、ある一つの共通の印象に基づいて使用されているであろうとみなせるので、できたクラスタそれぞれを一つのコンセプトと対応付ける。

以前の研究で提案したコンセプトの形成方法に関して、上記 (1) と (2) それぞれの詳細を、2.2.1 節と 2.2.2 節で説明する。

2.2.1 タグ間の類似度

以前の研究では、「ある二つのタグについて、一方のタグが付けられているアイテムにほぼ常に他方のタグも付けられていれば、それらタグは共通の印象に基づいて使用されている」と仮定し、二つのタグの共起のしやすさに基づいて、タグ間の類似度を定義した。各アイテムに対する二つのタグ (a, b) 間の付与関係は、以下の四つにまとめられる。

- (A, B) 両方のタグが付けられている
- (\bar{A} , \bar{B}) どちらのタグも付けられていない
- (A, \bar{B}) a は付けられているが、b は付けられていない
- (\bar{A} , B) b は付けられているが、a は付けられていない

「(A, B) と (\bar{A} , \bar{B}) が起りやすく、(A, \bar{B}) と (\bar{A} , B) が起りにくい」なら、a, b は共起しやすいといえる。このため以前の研究では、タグ間の類似度として AEMI (Augmented Expected Mutual Information) ¹¹⁾ を採用した。

$$AEMI(a, b) = MI(A, B) + MI(\bar{A}, \bar{B}) - MI(A, \bar{B}) - MI(\bar{A}, B)$$

ここで、A は「アイテムにタグ a が付けられている」という事象を、B は「アイテムにタグ b が付けられている」という事象を示す。 \bar{A} は「アイテムにタグ a が付けられていない」という事象を示す。また、 $MI(A, B)$ は「事象 A と B が相互にどのくらい依存しているかの情報量」であり、以下の式で表される。

$$MI(A, B) = P(A, B) \log \frac{P(A, B)}{P(A)P(B)}$$

$P(A)$ は「事象 A の発生頻度 (タグ a が付けられているアイテムの割合)」を、 $P(A, B)$ は「事象 A と B の同時発生頻度 (タグ a と b が共に付けられているアイテムの割合)」を示す。

2.2.2 タグのクラスタリング

同じ印象に基づいて使用されているであろうタグ同士を集めるために、互いに類似度が高いタグ群からなるクラスタ群を生成させる。そのために、以前の研究では、高い類似度を持つ「タグの組」に着目し、一方のタグが含まれるクラスタには他方のタグも含まれるようなクラスタリング手法を提案した。具体的な手順は以下の通りである。

まず、生成されたクラスタの集合 C_{set} を用意しておく (初期状態では空集合)。次に、全てのタグの組 (t_i, t_j) ($i \neq j$) を生成し、その中から類似度があらかじめ決めておいた閾値 V 以上のタグの組を選出する。そして、そこからタグ間の類似度が高い順にタグの組を取り出して、その二つのタグ (t_i, t_j) について以下の処理を行う。

1. C_{set} に含まれるクラスタのうち、 t_i と t_j の双方をも含むクラスタを除いた、残りのクラスタ群を選び出し、それらクラスタそれぞれ (cl_k) について以下を行う。
 - (i) $t_i \notin cl_k, t_j \in cl_k$ の場合、 t_i と cl_k 内の各タグとの類似度の平均値を求め、それが閾値 V を超えていれば、 t_i を cl_k に追加する。
 - (ii) $t_i \in cl_k, t_j \notin cl_k$ の場合、 t_j と cl_k 内の各タグとの類似度の平均値を求め、それが V を超えていれば、 t_j を cl_k に追加する。
 - (iii) $t_i \notin cl_k, t_j \notin cl_k$ の場合、 t_i と cl_k 内の各タグとの類似度の平均値と、 t_j と cl_k 内の各タグとの類似度の平均値を求め、その平均が V を超えていれば、 t_i と t_j を cl_k に追加する。

ただし、上記 (i)~(iii) のいずれかによってタグを追加した後の cl_k と、同一のタグ構成のクラスタが既に C_{set} 内に存在する場合は、 cl_k は C_{set} から削除する。

2. 1. の処理によって更新された C_{set} の中に、 t_i と t_j の双方をも含むクラスタが一つも存在しなかった場合は、 t_i と t_j からなる新たなクラスタを生成し、 C_{set} に加える。

2.3 コンセプトを用いた人の好みの表現

あるユーザの好みは「得られた各コンセプトへの関心の強さ」で表現する。これは、「そのユーザの好きなアイテムに付けられているタグと、コンセプトとの間の関連強さ」に基づいて算出する (図 3)。具体的には、まず、「タグとコンセプトとの間の関連強さ」を、そのタグと、そのコンセプトに対応するクラスタ内の各タグとの類似度に基づいて算出する。あるアイテムに付けられているタグがあるコンセプトと関連強いとき、そのコンセプトに対応する (クラスタができた背景にある) 印象に基づいてそのアイテムにそのタグは付与されたと期待できる。このため、次に、「アイテムとコンセプトとの間の関連強さ」を、そのアイテムに付けられているタグと、そのコンセプトとの間の関連強さに基づいて算出する。あるアイテムがあるコンセプトと関連強いとき、そのアイテムには、そのコンセプトに対応する印象が抱かれていると期待できる。そして、「あるユーザが好きなアイテムにユーザ達が抱いてい

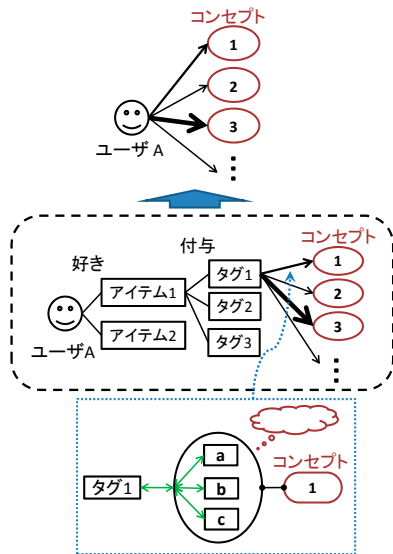


図 3. コンセプトを用いた人の好みの表現.

る印象と、同じ印象がユーザ達によって抱かれている別のアイテムに、そのユーザは関心を持つ」と考えられるので、「ユーザのコンセプトへの関心強さ」を、そのユーザが好きなアイテムと、そのコンセプトとの関連強さに基づいて算出する。

ユーザがどのアイテムが好きかどうかという情報を獲得する方法は、アンケートなどでユーザに直接答えてもらう明示的な方法と、行動履歴（閲覧した Web ページや、購入した商品）などから仮定する暗黙的な方法がある³⁾¹²⁾。

3. 以前の研究で提案した手法で得られたコンセプトの妥当性の分析

前述したような人の好みの表現を実現するためには、コンセプトが妥当なものである必要がある。そこで、以前の研究で提案した手法を用いて実際に得られたコンセプトが妥当なものであるかどうかを調べる。

具体的には、各コンセプトに対応したクラスタ内のタグ群から、ユーザが抱いたであろう印象が連想できれば、妥当なコンセプトが得られていると判断することとする。逆に言えば、クラスタ内のタグ間に関連性が見出せず、ユーザが抱いているであろう印象の連想が困難であるときは、そのコンセプトは妥当でないと判断することとする。また、ユーザの好みの表現に用いる「タグとコンセプトとの間の関連度」にも焦点を当てる。各コンセプトと関連強いタグ群から、ユーザが抱いたであろう印象を連想できれば、算出されたタグとコンセプトとの間の関連度は妥当なものであり、前述した方法による人の好みの表現に用いるものとして妥当なコンセプトが得られていると判断することとする。さらに、本研究の基本的な考え方下では、ユーザが抱いた印象一つに対して一つのコンセプトが対応するので、同じ印象に対応するコンセプトが複数存在することは妥当ではないと考えられる。このため、そのようなことが起きてしまっていないかどうかを調べる。

表 1. 各クラスタ内のタグ群の例.

Cluster	Part of tags
1	IT, プログラミング, 技術書, オライリー, C, C 言語
2	IT, プログラミング, 技術書, オライリー, Java, java
3	it, プログラミング, 技術書, JavaScript, javascript, Javascript
4	IT, プログラミング, 技術書, JavaScript, javascript, Javascript
5	Web, web, ウェブ, デザイン
6	Web, web, ウェブ, WEB

表 2. 各コンセプトと関連強い上位タグ群の例.

Concept	Top tags
1	C, C 言語, プログラミング, Ruby
2	Java, java, programming, Android
3	Javascript, JavaScript, javascript, Java
4	Javascript, javascript, JavaScript, Java
5	design, WEB, PHP, web
6	WEB, PHP, Web, ウェブ

3.1 方法

この分析に用いるデータは「ブックログ (Booklog)」(<http://booklog.jp/>) から収集した。ブックログとは仮想本棚を作成できる Web サービスである。得られた本の数は 6,717 冊、その本に付けられていたタグは 18,922 個だった。今回、コンセプトへのタグの影響を見やすくするために、7 冊以上の本に付けられているタグに絞った (1,115 個)。

このデータから、以前の研究で提案した手法を用いて、コンセプトの形成を行った。そして、以下の (i) と (ii) のタグ群それぞれを見て、ユーザが抱いたであろう印象を連想できるかの判断によって、そのコンセプトの妥当性を調べた。

- (i) 各コンセプトに対応したクラスタ内のタグ群
- (ii) 各コンセプトと関連強い上位タグ群

また、同じ印象に対応するコンセプトが複数存在しないかを調べるために、互いに似たタグ構成のものが複数できていないかを調べた。

3.2 結果

(i) の例を表 1 に、(ii) の例を表 2 に示す。表 2 では、対象のコンセプトとの関連度が高いタグほど左の方に表示している。

被験者は、(i) と (ii) のどちらにおいても、ほとんどのコンセプトが妥当なものであると判定した。例えば、表 1 の Cluster1 と表 2 の Concept1 からは共に、「C 言語」という印象を抱いたと判定できた。

一方で、同じ印象に対応するであろうコンセプトが複数できてしまっていた。例えば、表 2 の Concept3 と 4、5 と 6 のように、同じタグ群と関連強いコンセプトがいくつか存在した。また、表 1 の Cluster1 と 2、3 と 4、5 と 6 のように、その大半が同じタグ群で構成されるクラスタがいくつか存在した。

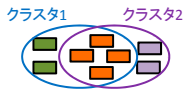


図 4. 互いに似ているクラスタ.

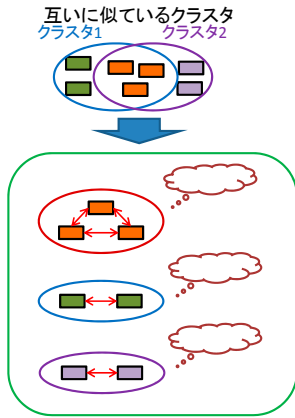


図 5. 目標とするクラスタ.

3.3 考察

今回の分析の中で、二つのクラスタが「互いに似ている」(図 4 に示すように、それら二つのクラスタが多くのタグを共有している) とき、その二つのクラスタの中から取り出した二つのタグの間の類似について、以下の知見が得られた。すなわち、互いに似ている二つのクラスタを選んで、そのクラスタ内のいずれかのタグを二つ選び、その類似度を算出した結果から、以下二つに該当するタグの組は、それらに該当しないタグの組よりも、類似度が比較的高い傾向にあることがわかった。

- (α) 一方のクラスタにしか含まれていないタグの組
(図 4 の緑のタグの組、或いは、紫のタグの組)
- (β) それぞれのクラスタに共に含まれるタグの組
(図 4 のオレンジのタグの組)

例えば、表 1 の Cluster1 と 2 に注目したとき、Cluster1 にしか含まれていないタグの組「C, C 言語」や Cluster2 にしか含まれていないタグの組「Java, java」それぞれの類似度は、(α)(β) に該当しないタグの組 (例えば「プログラミング, C」など) の類似度よりも高かった。

このことから、「互いに“特に”高い類似度を持つようなタグ同士からなるクラスタ群」が生成できれば、「(α) に該当するタグ群」と「(β) に該当するタグ群」とを分けることで、互いに似ているクラスタの生成を避けられると期待できる (図 5)。ただし、「特に高い類似度」の高さの程度は各タグによって異なるため、工夫が必要である。

4. コンセプトの形成方法の改良

コンセプトの精度向上のために改良した、新たなコンセプトの形成方法を提案する。具体的には、以前提案した手法から、(1) タグ間の類似度の定義、(2) タグのクラスタリング手法、及び、(3) タグとコンセプトとの間の関連度の算出方法を改良する。

- (1) 一部のタグの組の類似度が、人が思う類似性とは一致し

ていない場合があった。このため、人が思う類似性と一致するよう、タグ間の類似度を再定義する (4.1 節)。

- (2) 先述したように、互いに似ているクラスタが生成されていたため、これを回避するための工夫を導入したクラスタリング手法を提案する (4.2 節)。
- (3) “あいまいさを持つタグ” (多義的・抽象的な意味で使用されているタグ) による、人の好みの表現への悪影響が懸念されるため、“あいまいさを持つタグ”とコンセプトとの間の関連度が低くなるように、その算出方法を改める (4.3 節)。

4.1 タグ間の類似度

まず、タグ間の類似度は、「ある二つのタグについて、一方のタグが共起しやすいタグ群に、他方のタグも共起しやすいとき、それらタグは共通の印象に基づいて使用されている」と仮定し、二つのタグの共起パターン間の類似度に基づいて定義する。以前の研究では、二つのタグの共起のしやすさ (以下、この尺度を「タグ間の共起度」と呼ぶ) に基づいて、タグ間の類似度を定義した。しかしこの定義では、タグ間の類似度の高さがそれらタグの発生頻度に依存してしまっていた。例えば、同じ意味で用いられているタグ同士であっても、それら各タグの発生頻度が低ければ、そのタグ間の類似度も低くなってしまふ。

各タグの共起パターン (そのタグの、全てのタグそれぞれとの共起度からなるベクトル) 間の類似度を、そのタグ間の類似度とすることで、単にその二つのタグ間の共起度を用いた場合よりも、より多くの情報に基づいてタグ間の類似度を定めているため、人が思う類似性と一致するようになることが期待できる。また、共起の傾向としての類似度を算出することで、各タグの発生頻度による影響を避けることが期待できる。

具体的には、まず、以前の研究と同様の考えに基づき、タグ間の共起度として AEMI を採用する。

次に、この共起度を用いて各タグの共起パターンを表現する。全てのタグ間の共起度 $AEMI(t_i, t_j)$ に関して、 $N \times N$ (N : タグ数) 次元の行列ができるので、その各行ベクトルを、対応するタグ t_i の共起パターン $\vec{v}(t_i)$ とする。

$$\vec{v}(t_i) = [AEMI(t_i, t_j)]_{j=1, 2, \dots, N}$$

ここで、 $AEMI(a, a)$ はタグ a の発生に関するエントロピーである。

$$AEMI(a, a) = -P(A) \log P(A) - P(\bar{A}) - \log P(\bar{A})$$

そして、タグ a と b の共起パターン間の類似度を、 a と b の間の類似度 $sim(a, b)$ とする。本研究ではコサイン類似度を採用する。

$$sim(a, b) = \frac{\vec{v}(a) \cdot \vec{v}(b)}{\|\vec{v}(a)\| \|\vec{v}(b)\|}$$

4.2 タグのクラスタリング

新たなクラスタリング手法も、基本的には、以前提案した方法と同様、「ある共通の印象に基づいて使用されているであろうタグ群」を見付けるために、互いに高い類似度を持つタ

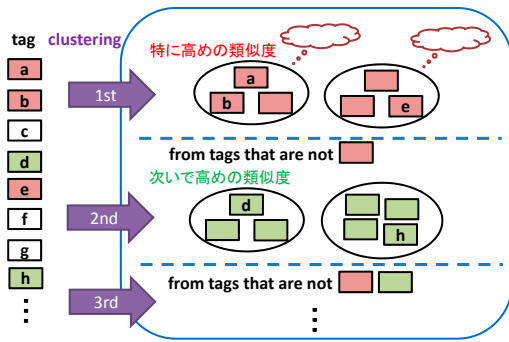


図 6. 多段的なクラスタリング.

グ同士からなるクラスタを生成する。本研究ではいくつかのアイデアを導入した (4.2.1 節～4.2.4 節)。これらアイデアを導入して提案するクラスタリング手法全体のアルゴリズムを 4.2.5 節に示す。

4.2.1 多段クラスタリングの導入

互いに似ているクラスタが生成されることを避けるために、3.3 節で述べたような、以前提案した方法で生成された互いに似ているクラスタの「一方のクラスタにしか含まれていないタグ群」と「それぞれのクラスタに共に含まれるタグ群」とが別々のクラスタに含まれるようになることを目指す (図 5)。このために、「互いに特に高い類似度を持つタグ同士」からなるクラスタを生成したいが、「特に高い類似度」の高さの程度は各タグによって異なる。

そこで、「互いに特に高い類似度を持つタグ同士」からなるクラスタを生成するために、タグ間の類似度の高さに応じた多段的なクラスタリングを導入する。まず、最初のステップでは、「特に高い類似度」として高めの閾値を設定し、その高さ以上で互いに類似し合うタグ群からなるクラスタ群を生成する。このステップでできた各クラスタ内のタグは、ある共通の印象に基づいて使用されているタグとみなせるので、以降のステップではそれらタグを除いたタグ群を対象にクラスタリングを行う。次のステップでは、先程よりも低い閾値を新たに設定し、残りのタグ群 (以前までのステップでできたクラスタには含まれないタグ群) から、前のステップに次いで高めの値で、互いに類似し合うタグ群からなるクラスタ群を生成する。この処理を数回繰り返す (図 6)。

4.2.2 タグクラスタの生成方法の改良

以前提案した方法では、類似するタグの組に着目し、その一方のタグが含まれるクラスタには他方のタグも含まれるようクラスタを生成していたために、本来はそのクラスタ内のタグ群と共通の印象に基づいて使用されるようなタグでなくとも、そのクラスタに含まれてしまう可能性があった。これは、着目しているタグの組の一方のタグが、他方のタグにひかれてしまうために起こり得る問題である。つまり、着目しているタグの組の一方のタグはクラスタ内のタグ群と類似していなくても、他方のタグがそれらタグ群と非常に高い類似度を持つ場合、類似していなかった方のタグもそのクラスタに追加されてしまう可能性が懸念される。

そこで本研究では、あるタグにとって、同じ印象に基づいて使用されているであろうタグを、一つ一つ見付けていく方

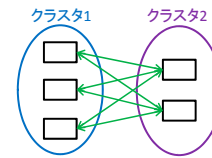


図 7. クラスタ間の類似性.

法をとることで、そのような問題の緩和を目指す。具体的には、まず、類似したタグの組の一つを選び (つまり、あるタグにとって、そのタグと同じ印象に基づいて使用されているであろうタグの一つを選び)、それら二つのタグからなるクラスタを作る。そのクラスタ内のタグ群に最も類似するタグを選んで (つまり、そのクラスタ内のタグ群にとって、それらタグ群と同じ印象に基づいて使用されているであろうと最もいえるタグの一つを見つけ)、追加する。この繰り返しによって、一つのクラスタを拡大していく。

4.2.3 互いに似たクラスタのマージ

多段的なクラスタリングを導入しても、互いに似ているクラスタの生成を完全に抑えることができるとは限らない。このとき生成された互いに似ているクラスタ同士は、(たまたま少し異なるタグ構成のクラスタができてしまっただけで) 同じ印象に基づいて使用されているタグ群からなるクラスタ同士であると考えことにする。

そこで、一つのクラスタが生成されたとき、既にできているクラスタ群の中に、そのクラスタと類似性の高いクラスタがある場合は、それら二つのクラスタをマージする。ここでのクラスタ間の類似性は、一方のクラスタ内の各タグと、他方のクラスタ内の各タグとの間の類似性に依存すると考えられる (図 7)。このため、二つのクラスタ間の類似度の計算には、それらタグ間の類似度を用いる。

4.2.4 順応性のある閾値の導入

4.2.2 節において、クラスタにタグを追加するかどうかを判断する際に使用する閾値は、そのクラスタに十分に類似しているタグが追加されるようにするために、そのクラスタ内のタグ間の類似度に基づいて、その都度計算した値を使用する。固定の閾値は用いない。

また、4.2.3 節において、十分に似ているといえるクラスタ同士をマージするために、このとき用いる閾値もまた、それら各クラスタ内のタグ間の類似度に基づいて、その都度更新する。

4.2.5 アルゴリズム

初めに以下を用意しておく。

- 生成されたクラスタの集合 C_{set} (初期状態では空集合)
- 全てのタグからなるタグの集合 T
- ステップ数 S
- 閾値 V_k ($k = 1, 2, \dots, S$), ($V_k > V_{k+1}$)

タグ間の類似度が閾値 V_k 以上のタグの組 (t_i, t_j) ($t_i, t_j \in T | i \neq j$) からなるリストを生成する。そのリストから昇順にタグの組を選択し、以下の処理を行う。

1. タグ t_i と t_j からなるクラスタ cl を生成する。

2. 以下の処理を繰り返す。

- (i) cl 内のタグ群に最も類似するタグ t_{max} ($t_{max} \in T, t_{max} \notin cl$) を選ぶ。タグ t とクラスタ cl との間の類似度 $sim(t, cl)$ は、 cl 内の各タグと、 t との間の類似度の平均値とする。
- (ii) 以下の閾値 $V(cl)$ を求める。

$$V(cl) = \alpha \times ave_{sim}(cl)$$

ここで、 $ave_{sim}(cl)$ は、 cl に含まれるタグ間の類似度の平均値を示す。また、 α はこの閾値の高さを調節するためのパラメータである。

- (iii) $sim(t_{max}, cl) \geq V(cl)$ のときは、 t_{max} を cl に追加し、(i) に戻る。
 - (iv) (iii) の条件を満たさなければ、この処理を終了する。
3. cl と C_{set} 内の各クラスタ cl_s との間の類似度 $sim(cl, cl_s)$ 、及び、閾値 $V(cl, cl_s)$ を求める。 $sim(cl, cl_s)$ は、 cl 内の各タグと cl_s 内の各タグとの間の類似度の平均値とする。また、 $V(cl, cl_s)$ は $V(cl)$ と $V(cl_s)$ の平均値とする。
4. C_{set} 内に、 cl と類似するクラスタ cl_{sim} が含まれているとき ($sim(cl, cl_{sim}) \geq V(cl, cl_{sim})$)、 cl 内のタグ群 ($\notin cl_{sim}$) を cl_{sim} に追加する。
5. 類似するクラスタがなければ、 cl を C_{set} に追加する。

これら処理によって更新された C_{set} 内のクラスタに含まれるタグ群を、 T から取り除く。これを S ステップ繰り返す。

4.3 タグとコンセプトとの間の関連度

タグとコンセプトとの間の関連度は、そのタグとそのコンセプトに対応するクラスタ (内の各タグ) との間の類似度に基づいて算出する。あるアイテムに付けられているタグが、あるコンセプトに対応するクラスタ内の各タグと類似しているとき、そのコンセプトに対応する印象に基づいてそのタグが付与されていると期待できる。ただし、3.2 節で示したような、多数の同じタグ群と関連強いコンセプト群があるとき、具体的には、タグとの関連強さの傾向が互いに類似したコンセプトが存在するとき、人の好み (或いはアイテムの特徴) を表現する上で、同じ印象に対応するようなコンセプトが複数存在してしまっていることになってしまう。このようなことが起こる原因として、互いに似ているクラスタが生成されてしまっていることと、その他に、“あいまいさを持つタグ” (多義性を持つタグ、或いは、抽象的な意味で使用されているタグ) による影響が考えられる。複数の印象に基づいて使用されているであろう“あいまいさを持つタグ”は、複数のクラスタ (内の各タグ) と類似する可能性が高く、つまり、複数のコンセプトそれぞれとの関連度が高くなるのが十分に起こり得るといえる。すなわち、互いに同じ“あいまいさを持つタグ”群と関連強いコンセプトが複数存在してしまうことが懸念される。そこで、“あいまいさを持つタグ”とコンセプトとの間の関連度が低くなるような、タグとコンセプトとの間の関連度の算出方法を検討する。

具体的には、まず、全てのタグとクラスタの組み合わせ (t_i, cl_j) それぞれについて、その間の類似度 $sim(t_i, cl_j)$ を求める。 $sim(t_i, cl_j)$ は、 t_i と cl_j 内の各タグとの類似度の平均値とする。そうすると、 $sim(t_i, cl_j)$ を要素とする $N \times M$ (N : タグ数、 M : クラスタ数) 次元の行列が表現できる。次に、“あいまいさを持つタグ”による影響を避けるため、各タグに関するベクトルの正規化を行う。基本的には、こうしてできた行列の各要素を、タグと (そのクラスタに対応する) コンセプトの間の関連度 $rel(t_i, co_j)$ とする (式 (1))。すなわち、多数のクラスタと類似するようなタグを“あいまいさを持つタグ”とみなし、そのようなタグとコンセプトとの間の関連度が低くなるようにする。

$$rel(t_i, co_j) = rel(t_i, cl_j) = \frac{sim(t_i, cl_j)}{\sqrt{\sum_k sim(t_i, cl_k)^2}} \quad (1)$$

また、更に“あいまいさを持つタグ”の影響を避けるために、この関連度を拡張する。具体的には、先程得られた行列の各要素 (式 (1) の値) を二乗した値を、該当するタグとコンセプトとの間の関連度として採用する。この二乗した値を採用することにより、“あいまいさを持つタグ”であるほど、そのタグとコンセプトとの関連度がより低くなるようにする。実際には、 $sim(t_i, cl_j)$ が負の値を取ることもあるため、タグとコンセプトの間の関連度は以下とする。

$$rel(t_i, co_j) = \begin{cases} rel(t_i, cl_j)^2 & \text{if } sim(t_i, cl_j) \geq 0.0 \\ -rel(t_i, cl_j)^2 & \text{else} \end{cases} \quad (2)$$

5. 実験

改良した手法 (以下、新手法) の有効性を調べるために、以前の研究で提案した手法 (以下、旧手法) との比較実験を行った。具体的には以下三つに関して、期待した効果が得られていることを確認する実験を行った。

- (1) タグ間の類似度 (5.1 節)
人が思う類似性と一致するようになっているか。
- (2) タグクラスタ (5.2 節)
互いに似ているクラスタの生成が回避できているか。
- (3) タグとコンセプトとの間の関連度 (5.3 節)
“あいまいさを持つタグ”による影響を緩和できているか。

実験に用いたデータは 3. 節のものと同じである。

5.1 タグ間の類似度に関する実験と結果

まず、新手法によるタグ間の類似度の精度向上 (人が思う類似性と一致するようになっているかどうか) を調べる実験を行った。具体的には、タグ間の類似度の再定義によって、旧手法における、タグ間の類似度の高さがタグの発生頻度に依存してしまう問題が改善されていることの確認を行った。

5.1.1 方法

まず、いくつかのタグの組を選び出す。これらタグの組に関して、旧手法を用いた場合と新手法を用いた場合のそれぞれの類似度を求める。そして、各タグの発生頻度によらず、同じ意味で用いられているタグの組の類似度が比較的高い値になっているか否かを調べる。

表 3. タグ間の類似度の結果の例.

A pair of tag	旧手法	新手法
(最大値)	0.1569	0.9983
ミステリー, ミステリ	0.1396	0.9377
C, C 言語	0.0063	0.8837
Java, java	0.0061	0.7341
javascript, Javascript	0.0083	0.9619
プログラミング, 技術書	0.0159	0.6528
プログラミング, C	0.0049	0.4176
プログラミング, Java	0.0058	0.3786
プログラミング, javascript	0.0048	0.4207

5.1.2 結果

各手法での定義によるタグ間の類似度の値の例を表 3 に示す。この表の一番上には、各手法で求めたタグ間の類似度の最大値を示している。

この結果から、新手法は、タグ間の類似度の高さがタグの発生頻度に依存してしまう問題が改善されていることがわかり、タグ間の類似度の精度向上が確認できた。新手法では、同じ意味で用いられているようなタグの組の類似度は、それ以外のタグの組の類似度や最大値と比較して、いずれも高い値となっていた。発生頻度の高いタグの組である「ミステリー, ミステリ」間の類似度だけでなく、比較的発生頻度が低いタグの組である「C, C 言語」や「javascript, Javascript」間の類似度も比較的高い値となっていることから、タグ間の類似度の高さがタグの発生頻度に依存してしまう問題の改善が確認できた。一方で、旧手法の場合は、同じ意味で用いられているタグの組であっても、最大値と比べてその類似度が比較的低いタグの組が多かった。例えば、表 3 に示すように旧手法だと、発生頻度が高いタグの組である「ミステリー, ミステリ」間の類似度は比較的高いが、発生頻度が低いタグの組である「C, C 言語」間の類似度は比較的低い。

5.2 タグクラスタに関する実験と結果

次に、新手法が目指す、互いに似ているクラスタの生成の抑制が実現できているかを調べる実験を行った。

5.2.1 方法

まず、旧手法と新手法それぞれを用いてタグのクラスタリングを行う。そして、得られたクラスタ群から、クラスタの全ての組み合わせを選ぶ。このクラスタの組それぞれについて、それら二つのクラスタに共に含まれるタグの数を求める。共に含まれるタグがない場合は、0 として数える。そして、この平均値（「クラスタの組それぞれが共有するタグの数の総和」／「クラスタの組の数」）を求め、旧手法と新手法それぞれの結果を比較する。

また、新手法で生成された各クラスタ内のタグ群に着目し、3.3 節で目指していたようなクラスタ (図 5) が得られているかを調べる。

5.2.2 結果

平均値の結果 (表 4) から、新手法では互いに似ているクラスタ群の生成を回避できていることが確認できた。旧手法

表 4. 二つのクラスタに共通するタグの数の平均値.

旧手法	新手法
5.1688	0.0019

表 5. 新手法で生成された各クラスタ内のタグ群の例.

Cluster	tags
1	C, C 言語
2	JavaScript, javascript, Javascript
3	Web, web, WEB, ウェブ, インターネット
4	デザイン, design

よりも、新手法の方がクラスタ間で共通するタグの数は少ないことがわかる。

また、新手法で生成された各クラスタ内のタグ群に着目しても、目指していたようなクラスタが得られていることが確認できた。いくつかのクラスタの例を表 5 に示す。例えば表 5 の Cluster1 のように、旧手法で生成された互いに似ている二つのクラスタ (表 1 の Cluster1 と 2) の「一方のクラスタにしか含まれていないタグ群」からなるクラスタが生成されていた。

5.3 タグとコンセプトとの間の関連度に関する実験と結果

新手法によって、4.3 節で述べたような“あいまいさを持つタグ”による影響を緩和できているかどうかを調べる実験を行った。具体的には、タグとの関連強さの傾向が互いに類似した (多数の同じタグ群と関連強い) コンセプトの組が存在しないかどうかを確認した。

5.3.1 方法

旧手法と新手法それぞれを用いて、コンセプトを形成し、タグとコンセプトの全ての組み合わせについてその関連度を算出して、タグとの関連強さの傾向に関するコンセプト間の類似度を求める。このコンセプト間の類似度の計算には、コサイン類似度を採用する。

$$\text{sim}(co_i, co_j) = \frac{\sum_k \{rel(t_k, co_i) \times rel(t_k, co_j)\}}{\sqrt{\sum_k rel(t_k, co_i)^2} \sqrt{\sum_k rel(t_k, co_j)^2}}$$

新手法に関しては、タグとコンセプトとの間の関連度として 4.3 節の式 (1) を用いた場合と、式 (2) を用いた場合とで、それぞれにおけるコンセプト間の類似度を求める。これは、タグとの関連強さの傾向に関するコンセプト間の類似性に対する、式 (1) と式 (2) の影響の違いを調べるためである。すなわち、人の好みを表現する上での、タグとの関連強さの傾向が互いに類似したコンセプトが存在する問題に対して、“あいまいさを持つタグ”であるほどそのタグとコンセプトとの関連度をより低くすることの有効性を調べるためである。

5.3.2 結果

得られた結果を以下のようにまとめた。まず、各手法 (旧手法、式 (1) を用いた新手法、及び式 (2) を用いた新手法) で求めたコンセプト間の類似度の平均値、最大値、及び最小値を求め、これを表 6 に示す。また、各手法で得られた結果に関するヒストグラムをそれぞれ図 8、図 9、及び図 10 に示

表 6. コンセプト間の類似度の結果.

	旧手法	新手法 (式 (1))	新手法 (式 (2))
平均値	0.3779	0.0878	0.0185
最大値	0.9994	0.9967	0.9903
最小値	0.0	-0.4321	-0.0925

す。横軸はコンセプト間の類似度であり、ビン幅は 0.1、縦軸はそのビン幅に対応する類似度を持つコンセプトの組の数の割合である。

これらの結果から、“あいまいさを持つタグ” による影響を緩和できていることが確認できた。コンセプト間の類似度の平均値、最大値はいずれも、式 (2) を用いた新手法、式 (1) を用いた新手法、旧手法の順に小さいため、その順に、タグとの関連強さの傾向が互いに類似したコンセプトができていないことがわかった。また、図 8 をみると以前の方法では、類似度が 0.7 以上のコンセプトの組が全体の 30 % 程度であるのに対して、図 9 と図 10 をみると新手法では、そのような類似したコンセプトの組は少なくなっていることがわかる。これは、新手法では互いに似ているクラスタの生成を回避できていることが一つの大きな要因であると考えられる。さらに、式 (2) を用いた新手法は、式 (1) を用いた場合よりも、互いに類似したコンセプトの組は少なくなっていることから、タグとの関連強さの傾向が互いに類似したコンセプトが存在する問題に対して、“あいまいさを持つタグ” とコンセプトとの関連度をより低くすることは、有効であるといえる。

6. 推薦実験

新手法によって得られるコンセプトを人の好みの表現に適用した場合の改良と有用性を調べるために、推薦実験を行った。旧手法と新手法のそれぞれで得られたコンセプトを用い、次の 6.1 節に示す推薦方法で、3. 節で集めた本の中から被験者の好みに合う本を推薦し、その推薦結果を被験者に評価してもらった。

6.1 推薦方法

ユーザの好みに合う本を見付けるために、得られたコンセプトを用いて、各本について「ユーザの好みに合う度合い」を算出する。そして、この値が高い本を推薦する。

そのために、まず、「本の、各コンセプトとの関連強さ」及び「ユーザの、各コンセプトへの関心強さ」を求める。本とコンセプトとの間の関連度 $rel(b, co)$ は、その本 b に付けられている各タグ t_i とコンセプト co との間の関連度の平均値とする。また、ユーザのコンセプトへの関心度 $int(u, co)$ は、そのユーザ u が好きな本 b_i とコンセプト co との間の関連度の総和とする。

$$rel(b, co) = \frac{\sum_i rel(t_i, co)}{(a \text{ size of tags attached to } b)}$$

$$int(u, co) = \sum_i rel(b_i, co)$$

そして、各本の「ユーザの好みに合う度合い」を、その本の特徴 (各コンセプトとの関連強さ) とそのユーザの好み (各コンセプトへの関心強さ) に基づいて、算出する。本 b の「ユーザ u

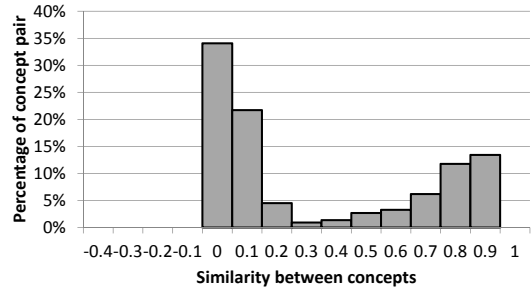


図 8. 旧手法に関するヒストグラム.

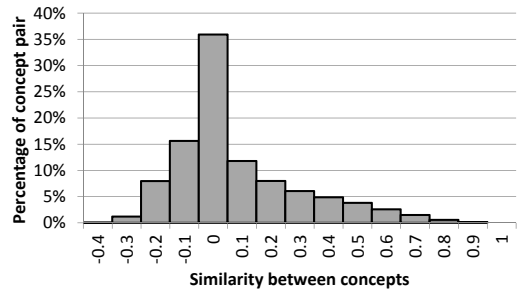


図 9. 式 (1) を用いた新手法に関するヒストグラム.

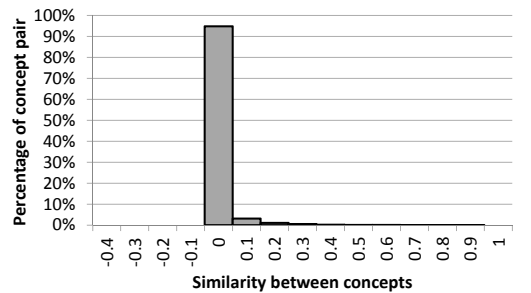


図 10. 式 (2) を用いた新手法に関するヒストグラム.

の好みに合う度合い」の算出方法として以下二つ ($pointA(u, b)$ と $pointB(u, b)$) を考えたが、どちらの方が適切かの結論を出すには至らなかったため、本実験では、双方の方法それぞれを用いた場合の推薦結果を被験者に提示する。

$$pointA(u, b) = \sum_i \{int(u, co_i) \times rel(b, co_i)\}$$

$$pointB(u, b) = \frac{\sum_i \{int(u, co_i) \times rel(b, co_i)\}}{\sqrt{\sum_i int(u, co_i)^2} \sqrt{\sum_i rel(b, co_i)^2}}$$

6.2 実験の内容と評価方法

あらかじめ、3. 節で集めたデータ上の各本について、その本の特徴 (各コンセプトとの関連度) を、旧手法で得られたコンセプトを用いた場合と、新手法で得られたコンセプトを用いた場合とで、それぞれ求めておく。次に、被験者には各自「好きな本」を列挙してもらい、それら本の特徴 (各コンセプトとの関連度) から各被験者の好み (各コンセプトへの関心度) を求める。被験者の「好きな本」が集めたデータ上にない場合は、その本に付けられているタグの情報をブックログから新たに取得することによって、その本の特徴を求める。求めた各被験者の好みに基づいて、6.1 節の推薦方法を用い、以下四種類の推薦結果を各被験者に提示する。

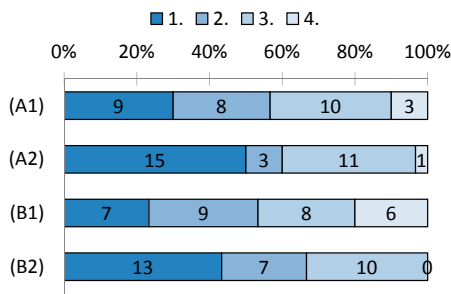


図 11. 推薦実験の評価結果.

- (A1) *pointA* が高い上位 5 冊の本
旧手法で得られたコンセプトを使用
- (A2) *pointA* が高い上位 5 冊の本
新手法で得られたコンセプトを使用
- (B1) *pointB* が高い上位 5 冊の本
旧手法で得られたコンセプトを使用
- (B2) *pointB* が高い上位 5 冊の本
新手法で得られたコンセプトを使用

そして被験者には、推薦された本それぞれに対する興味関心の有無を、以下の四段階で評価してもらう。

1. とても興味関心を持てる
2. 少し興味関心を持てる
3. あまり興味関心を持ってない
4. 全く興味関心を持ってない

6.3 実験結果

この推薦実験を被験者 6 名に対して実施した。(A1)~(B2)それぞれの推薦結果に対して、得られた 30 件 (6 名×5 冊分)の評価結果を図 11 に示す。

この推薦実験の結果から、新手法の改良と有用性が確認できた。推薦された本に対して興味関心があると評価した回答数は、(A1)が 17 件 (56.7%)、(A2)が 18 件 (60.0%)、(B1)が 16 件 (53.3%)、(B2)が 20 件 (66.7%) だった。旧手法で得られたコンセプトを用いた場合よりも、新手法で得られたコンセプトを用いた場合の精度が高かった ((A1)よりも(A2)、(B1)よりも(B2)の精度の方が高い)。また、これらの精度は他の研究と同等の精度である (例えば⁷⁾)。このため、人の好みの表現への、得られたコンセプトの適用は有用であるといえる。すなわち、本研究で目指していた、コンセプトを用いての、ユーザ個人の嗜好構造の把握を実現できているといえる。

7. 結言

本研究では、人の好みを適切に表現するために、ユーザ達がアイテムに対して抱く印象を推測したものである「コンセプト」を用いることに着目した。以前の研究で提案した手法で実際に得られたコンセプトの妥当性を調べ、新たなコンセプトの形成方法を提案した。以前の手法からの改良のために、タグ間の類似度を再定義し、新たなタグのクラスタリング手法を提案し、タグとコンセプトとの間の関連度を拡張した。比

較実験の結果、新たなコンセプトの形成方法の改良と有効性を確認した。また、推薦実験の結果から、新たな方法で形成されるコンセプトを、人の好みの表現に適用することの改良と有用性を確認した。

参考文献

- 1) 田上 道士, 山場 久昭, 高塚 佳代子, 岡崎 直宣, 富田 重幸: Serendipity を考慮した推薦システムの Folksonomy による実現とその評価, 宮崎大学工学部紀要 第 42 号, pp.313-318, 2013.
- 2) H. Yamaba, M. Tanoue, K. Takatsuka, N. Okazaki, S. Tomita: On a serendipity-oriented recommender system based on folksonomy, *Artificial Life and Robotics*, vol.18, pp.89-94, 2013.
- 3) 神嵐 敏弘: 推薦システムのアルゴリズム (2), *人工知能学会誌*, vol.23, no.1, pp.89-103, 2008.
- 4) J.B. Schafer, D. Frankowski, J. Herlocker, S. Sen: Collaborative Filtering Recommender System, *The Adaptive Web*, LNCS4321, pp.291-324, 2007.
- 5) M. J.Pazzani, D. Billsus: Content-Based Recommendation Systems, *The Adaptive Web*, LNCS4321, pp.325-541, 2007.
- 6) M. Szomszor, C. Cattuto, H. Alani, K. O'Hara, A. Baldassarri, V. Loreto, B. D.P.Sevedio: Folksonomies, the Semantic Web, and Movie Recommendation, *Bridging the Gap between Semantic Web and Web2.0*, pp.71-84, 2007.
- 7) 丹羽 智志, 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.
- 8) A. Shepitsen, J. Gemmell, B. Modasher, R. Burke: Personalized Recommendation in Social Tagging Systems using Hierarchical Clustering, In *Proceeding of the 2008 ACM conference on Recommender systems*, pp.259-266, 2008.
- 9) R. Krestel, P. Fankhauser: Tag Recommendation using Probabilistic Topic Models, *ECML PKDD Discovery Challenge 2009*, pp.131-141, 2009.
- 10) A. Said, R. Wetzker, W. Umbrath, L. Hennig: A hybrid PLSA approach for warmer cold start in folksonomy recommendation, *Recommender System & the Social Web*, pp.87-90, 2009.
- 11) P.L. Cahn: A non-invasive learning approach to building web user profiles, *KDD.99 Workshop on Web Usage Analysis and User Profiling*, 1999.
- 12) 土方 義徳: 情報推薦・情報フィルタリングのためのユーザプロファイリング技術, *人工知能学会誌*, Vol.19, No.3, pp.365-372, 2004.