

多様性を持つユーザの動的グルーピング

牧野 浩之[†], 廣安 知之^{††}, 三木 光範[‡], 横内 久猛^{††},

[†] 同志社大学大学院 ^{††} 同志社大学生命医科学部 [‡] 同志社大学理工学部

本提案では、ユーザが簡単な質問にオンライン上で回答し作成できる自己紹介（プロフィール）サービスをもとにしたユーザ同士のマッチングシステムを提案する。ユーザが回答した質問からシステムがグループ分けを行い、ユーザの嗜好とグループの時間的变化を捉える。質問項目はカテゴリ別に分類されており、一定時間ごとにどのようなカテゴリにユーザが興味を持っているかを考慮してグループを形成する。本提案システムの特長は、時間軸でユーザが所属するグループの変化を捉える点である。本提案を実証するため、プロフサイトを構築し、類似の変化をした他のユーザを興味に近いユーザとしてマッチングを行った。将来的には、モバイルデバイスでの利用を想定し、異業種交流会などのイベントにおいて、初対面でもお互いの興味を予め把握でき、コミュニケーションの促進に役立つシステムを目指す。

Development of Dynamic Grouping System for Diverse Users

Hiroyuki MAKINO[†], Tomoyuki HIROYASU^{††}, Mitsunori MIKI[‡] and Hisatake YOKOUCHI^{††}

[†] Graduate School of Engineering, Doshisha University

^{††} Department of Life and Medical Sciences, Doshisha University

[‡] Department of Science and Engineering, Doshisha University

In this paper, we propose users matching system based on self introduction (profile) service which users answer the quick question online. The system makes groups from the questions users have answered, and it captures the user's favor and time change in the group to which the user belongs. The question items are classified according to the category, and the system determines the group in consideration of what category the users are interested in. The feature of this proposal system is to catch the change in the group to which the user belongs by the time axis. To prove this proposal, we made the online service which matches other users who had similar change as a near user. In the near future, we assume the system is used with a mobile device. Thus users can figure out an interest with each other whether they meet for the first time in the event such as the pan-industry social events. Furthermore, we aspire for making the useful system for the promotion of communication.

1 はじめに

SNSやソーシャルブックマーク、ブログなど様々なオンラインサービスの登場により、ユーザがオンライン上に情報を蓄積するようになった。近年では情報爆発と言われるほど情報が蓄積され続けており、経済産業省が推進している情報大航海プロジェクト¹のように、蓄積された情報を解析し活用する動きが活発になってきている。新たなビジネスやイノベーションの創出には、必要な情報をどのように組み合わせ活用していくかが重要である。オンライン上には様々な興味や嗜好を持つユーザがサービスを利用し情報が蓄積されているため、そのような情報を活用し趣味の合う人や、ユーザ自身と同じようなことに興味がある人を見つけた

いというニーズもある。

そこで本研究では、ユーザによって蓄積される情報を解析し、ユーザの特徴を抽出するとともに、特徴に応じたグループ分けを行い、ユーザ同士のマッチングを行う。蓄積される情報源として、オンライン上でユーザが簡単な質問に答えることで、自己紹介ページを作成できるプロフィール（プロフ）サービスをベースに、グルーピングを行う。ユーザが所属するグループはシステムによって自動的に決定され、答える質問に応じてグループが変化する。本システムでは、ある一定時間ごとにユーザがどのようなグループに属しているかを捉えて、属するグループの変化によってマッチングを行う。

マッチング結果は現実社会での利用を想定している。例えば、異業種交流会においては、初対面の人同士が出会うケースが多いが、実際に会話をし

¹ <http://www.igvpj.jp/>

てみないとお互いの興味や求めていることが分からない上、時間が限られているため多くの人とコミュニケーションを取ることが難しい。そこで、参加者には予めプロフィールを作成してもらうことで、その場に居合わせる参加者同士でのマッチングを行うことも考えられる。本提案システムはモバイル環境での利用も視野に入れて、人脈形成やコミュニケーションを促進するサービスへの応用を目指している。

2 関連研究

関連研究として、膨大な情報からユーザの嗜好に合う情報を提供する手法に、情報推薦システムに関する研究が盛んに行われている。時間帯や天気、同伴者、予算など、状況に応じて変化するユーザの嗜好を適切にモデル化する手法として「ユーザの状況に応じて適切な情報を提供する状況依存型情報推薦システム¹⁾」や、各ドキュメントに対する特徴語ベクトルを抽出し、ユーザのプロファイルを一定時間ごとに作成することでその変化を捉える「Identifying and tracking changing interests²⁾」などが挙げられる。

3 プロフマッチングシステム

本研究では、グループ分けの例として、プロフィール作成サービス（プロフサービス）を対象にしたユーザ同士のマッチングを行う。結果として、ユーザがどのような属性を持っているかを示すとともに、ユーザに似た興味や嗜好を持つユーザを提示する。ユーザの興味は、時間が経つごとに変化する場合が考えられる。そのため、本システムでは一定期間ごとにユーザがどのようなグループに所属するかを計算することで興味の変化を捉えられるようにする。

3.1 プロフサービス

プロフサービスとは、ユーザが任意の簡単な質問に回答していくことで自分の自己紹介ページ（プロフィール）を作成できるサービスである。代表的なものに、前略プロフィール²⁾やアバウトミー³⁾などが挙げられる。アバウトミーでは7万個以上の質問が用意されている。本研究では、図1に示すようなプロフサービスを構築し、グループ分けによりユーザをマッチングする手法を実装する。

3.2 システムの利用方法

ユーザは以下の流れで本システムを利用することで、興味や嗜好の合うユーザが提示される。

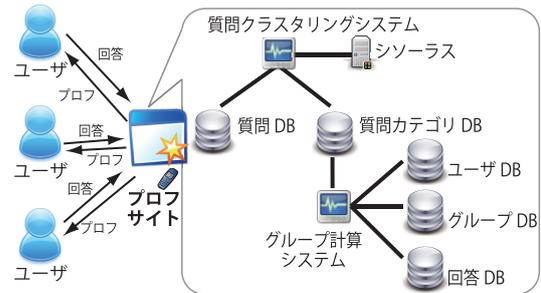


図 1: システムの構成

1. ログイン

本システムでは OpenID⁴⁾に対応しており、mixi や Yahoo!, はてな, livedoor などのアカウントを入力することでログインが行える。

2. 質問の作成

ユーザが任意に質問項目を作成できる。回答項目は Yes か No の選択型のほかに、ユーザが任意に回答の選択肢を作成したり、ユーザが自由に文章を入力して回答できる自由入力型の項目を作成することができる。

3. 質問の回答

ユーザが任意に回答したい質問項目を見つけて回答する。状況に応じてシステムがユーザに質問項目を推奨する。

4. リコメンデーション提示

ユーザの回答履歴をもとにグループ分けが行われ、興味や嗜好の合うユーザが提示される。

3.3 動的グルーピング

ユーザが質問項目に回答していくと、ユーザがどのような質問カテゴリに属しているかが変化することがある。ユーザは複数の質問カテゴリに所属することが考えられるため、所属しているカテゴリと回答数に応じてグループ分けを行う。本研究では、一定時間ごとにユーザがどのようなグループに属しているかの変化を考慮に入れたグルーピングを動的グルーピングと定義した。動的グルーピングの流れを以下に示す。

1. 質問項目を形態素解析し、単語（名詞）を抽出する。
2. シソーラスを用いて質問項目同士の類似度を計算する。
3. 類似度をもとに質問項目のクラスタリングを行い、カテゴリ分けを行う。

²⁾ <http://pr.cgiboy.com/>

³⁾ <http://aboutme.jp/>

⁴⁾ <http://openid.ne.jp/>

4. ユーザの回答をもとに、どのようなカテゴリにユーザが属しているか計算する。
5. ユーザが所属しているカテゴリとユーザ同士の類似度に基づいてグループを決定する。

ユーザの回答状況に応じて、ユーザが属しているカテゴリが変化するため、グループ構成が動的に形成される。

4 提案手法

本提案ではシソーラスを用いてプロフの質問項目全体を体系化し、ユーザがどのような質問に回答しているかを捉えて動的なグルーピングを行う。

4.1 シソーラスの利用

シソーラスとは、単語の上位/下位関係、部分/全体関係、同義関係、類義関係などによって単語を分類し、体系づけた辞書である。主に用いられるシソーラスとして、2,715のカテゴリからなる最大12段の木構造で表される、日本語語彙体系 (ALT) や、202,797のカテゴリからなる最大16段のグラフ構造で表されるEDR電子化辞書 (EDR) が挙げられる。本研究では、日本語語彙大系を用いた。

4.2 質問のカテゴリ分け

本研究では、質問項目を MeCab⁵を用いて形態素解析し、単語 (キーワード) に分解した上で、シソーラスを用いて単語間の類似度を計算した。文書間の類似度計算手法には、tf-idf や N-gram、編集距離がよく用いられているが、質問文のように比較的短文の類似度を計算するには、シソーラスを用いて単語間の階層距離計算を行う手法^{3) 4)}が有効である。

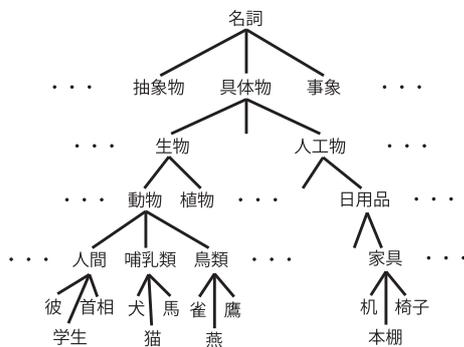


図 2: シソーラスの例

図 2 に示すように、シソーラスは単語が階層的に分類されているため、階層の違いを距離に見立てて類似度を計算することができる。また、異なる

語であっても意味的に似た単語であれば類似度が高くなるよう考慮することができる。2単語の類似度 (階層距離) は以下のように計算できる。

$$sim(q, d) = \frac{2c_{qd}}{(d_q + 1) + (d_d + 1)}$$

ここで、 d_q と d_d は見出し語 q, d が属する意味属性の深さである。また、 c_{qd} は見出し語 q が属する意味属性と見出し語 d が属する意味属性の共通の上位属性の中で最も下の階層に位置するカテゴリの深さである。

質問間の類似度計算 (スコアリング) は、ある質問のキーワードに対する、他の質問のキーワード全ての類似度を計算する。計算した後、類似度が高い順にソートを行い、上位数個を類似度の高い質問として保持する。質問に複数のキーワードがある場合においても同様に、それぞれのキーワードの類似度を計算し、類似度の高い順に保持する。

図 3 に示すように、スコアリングを行った後に、クラスタリングを行い、質問項目を体系化する。クラスタリングにはクラスタ数が自動的に決定される Newman 法⁷⁾を用いた。

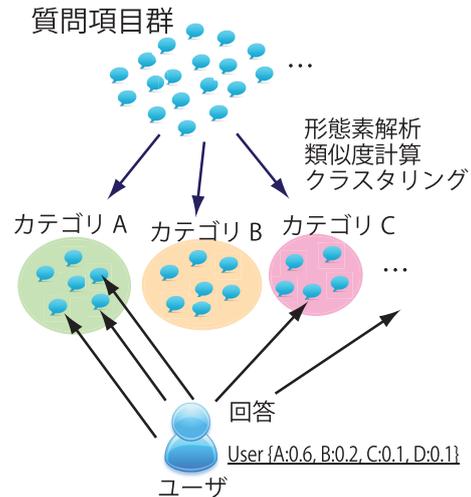


図 3: 質問の体系化

4.3 ユーザのグルーピング

ユーザの回答している質問項目がどのようなクラスタに属しているかで、ユーザがどのような質問群に興味を示しているかを特定し、ユーザ同士の類似度をもとにマッチングを行う。図 4 に示すようにユーザの所属するグループとその割合をベクトル化してユーザ同士の類似度を計算する。ユーザの属する質問群を考慮に入れることにより、あるカテゴリに対する興味の度合いや、同一カテゴ

⁵ <http://mecab.sourceforge.net/>

り内において他ユーザとの比較を行うことが可能になる。

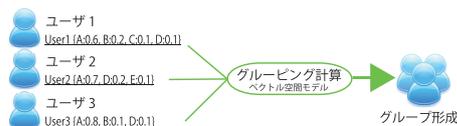


図 4: グルーピングの例

4.4 結果

システムの一連の流れを検証するため、テストを行った。テストでは 100 個の質問項目を任意に作成し、それぞれの質問で MeCab を用いて形態素解析を行い、キーワードを抽出した。次にキーワードが抽出できた質問に対して、シソーラスを用いて関連度計算を行った。これにより、84 個の質問間でシソーラスを用いた関連度計算ができた。個々の質問において、関連度が 0.6 以上となったものを関連度が高い順に最大 5 個抽出し、該当する質問に関連が高い質問とした。続いて、この結果に Newman 法を用いてクラスタリングを行った。その結果、49 個のクラスタが形成された。クラスタの最大サイズは 9、最小サイズは 1 となった。図 5 に生成された主なクラスタを示す。各数字は質問項目ごとの ID に対応しており、関連のある質問同士がクラスタに分類されていることが分かった。

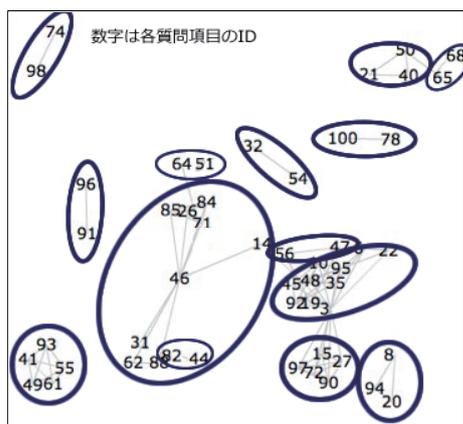


図 5: 質問項目のクラスタリング

最後に、ユーザの回答履歴から、どのクラスタに対して回答しているかをカウントし、ユーザの所属するグループを決定した。この処理を一定時間ごとに行い、ユーザの回答履歴の変化による所属グループの変化を記録する。

5 今後の課題

今後の課題として、現在は、ユーザがどのような質問に回答したかをもとにグルーピングを行っているが、同じグループ内の他のユーザが回答した選択肢も考慮に入れたマッチングを行う。また、異業種交流会といったイベントの場で利用できるよう、モバイルデバイスでの利用を考慮に入れたシステムの構築を行っていく。

6 おわりに

本研究では、ユーザが入力するプロフの情報をもとに、ユーザの興味や嗜好を捉える。提案手法を検証するため、プロフサイトを構築した。プロフの質問項目はシソーラスを用いて分類されており、ユーザがどのような分野に興味を持っているかを同定した。ユーザが属する質問カテゴリとその回答数に応じてグルーピングを行い、他のユーザのマッチングを行った。また、ユーザの興味は時間が経つごとに変化するものとして、一定時間ごとの興味の変化も考慮に入れてグルーピングを行う。

参考文献

- 1) 奥健太 中島伸介 宮崎純 植村俊亮: 状況依存型ユーザ嗜好モデリングに基づく Context-Aware 情報推薦システム, 情報処理学会論文誌, Vol.48 pp. 162-176, 2007.
- 2) Barry Crabtree Stuart J. Soltysiak: Identifying and tracking changing interests, International Journal on Digital Libraries, Vol.2 No. 1 pp.38-53, 1998.
- 3) 大橋敬久: シソーラスを用いた意思決定支援のための文書の抽出, 平成 18 年度卒業論文, 日本大学工学部情報工学科, 2007.
- 4) 川島貴広 石川勉: 言葉の意味の類似性判別に関するシソーラスと概念ベースの性能評価, 人工知能学会論文誌 20 巻 5 号, pp.326-336, 2005.
- 5) 藤井敦 徳永健伸 田中穂積: シソーラスと統計情報を統合した単語の類似度計算について, 電子情報通信学会技術研究報告, Vol.97(199) pp.53-58, 1997.
- 6) 古川忠延 松尾豊 内山幸樹 石塚満: 汎用的な Web 推薦アルゴリズムの提案, 第 70 回情報処理学会全国大会講演論文集, 3ZK-2, 2008.
- 7) M.E.J.Newman: Fast algorithm for detecting community structure in networks, Physical Review E 69,066133, 2004.