

ユーザの潜在的興味抽出とコミュニティの粒度推定に基づく情報推薦システム

志甫谷 匠†

中島 伸介‡

角谷 和俊†

†兵庫県立大学環境人間学部

‡京都産業大学コンピュータ理工学部

1 はじめに

近年、インターネットやコンピュータの発達により、世の中の情報が爆発的に増加している。この際、ユーザにとって必要な情報を提供する手法の1つとして、情報推薦システムの活用が有効である。しかしながら、従来技術では、ユーザが興味を持つトピックを認識することが出来ても、その興味の範囲を正しく認識した推薦を行うことが困難である。そこで、本研究では、推薦情報を適切なコミュニティに推薦するためのコミュニティ粒度推定と被推薦ユーザの潜在的な興味範囲の推定に基づく情報推薦の手法を提案する。そして、コミュニティの粒度が分散しやすい就職情報の推薦を対象にした研究を行う。

2 本研究の概要と関連研究

2.1 本研究の概要

本研究は、特に就職情報の推薦に焦点を当てている。就職情報の推薦とは、就職情報サイトで行われている就職活動生を対象とした業種・企業情報推薦のことである。これらの推薦には、ユーザのプロフィール情報や趣向情報を利用した類似ユーザの志望状況を基にして行われている。

この際、被推薦者に対して適切な粒度の情報を推薦することが重要である。例えば、同じ自動車業界を志望しているユーザであっても、その中には、カーオーディオに興味を持っているユーザや、開発部門に興味があるユーザ、さらにその中でも F1 の開発に興味があるユーザなど、1つの業界の中にも様々な粒度のコミュニティが存在し、推薦内容を一様にすることは適切ではない。そのため、被推薦者にとって適切なコミュニティの粒度を推定することが重要である。

このように、被推薦者の潜在的な興味を抽出することにより、適切なコミュニティの粒度の推定に基づいた情報推薦システムの提案を行う。なお、対象となる利用ユーザは、就職活動を行っている学生であり、就

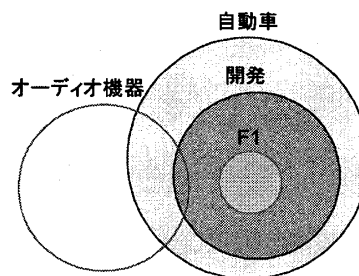


図1: 粒度の違い

職情報サイトへ登録を行っていることを前提とする。

以下、本システムの入出力の流れを説明する。まず、入力情報は、利用ユーザが興味を持っているキーワードや、志望業種名である。システムは、入力されたキーワードや業種と関連の強いニュース記事を複数出力する。出力されたニュース記事には、それぞれ関連の強い業種3つずつ記されており、さらに記事内容を示すタグ候補がキーワード単位で記されている。利用者は、出力された記事の中で、気になるニュースを閲覧することが出来、閲覧した記事に対してふさわしいと思うタグをタグ候補中から選択する。この際、タグとして選択されたキーワードを利用者の興味を示す語として、利用者毎にタグリストへ保存しておく。そして、タグリスト内に含まれる語の状況から、利用者の対象グループ粒度を推定し、最終的に推薦する企業名を決定する。

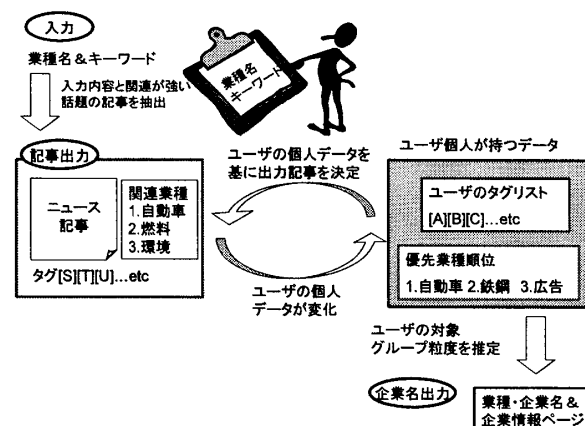


図2: 入出力の流れ

Recommender System Based upon Extracting User Potential Interest and Target Community Identification

†Takumi Shihoya ‡Shinsuke Nakajima †Kazutoshi Sumiya

†School of Human Science and Environment, University of Hyogo

‡Faculty of Computer Science and Engineering, Kyoto Sangyo University

2.2 関連研究

特定コミュニティに限定したトピックの抽出を行っている研究として、[2][3][4]などが挙げられる。中島ら [2] は、熟知度を表すような語をマニア候補語とし、マニア候補語を含む記事は熟知度が高いとし、マニア指向ランキング方式を提案している。関口ら [3] らは、同じ興味を持つ人々の共通の話題を表す“話題語”を定義し、話題語を抽出することで、特定の興味を持つコミュニティの発見を行う手法を提案している。

3 ユーザの潜在的粒度推定と対象コミュニティ粒度推定

3.1 ニュース記事へのタグ付与

ニュース記事へのタグ付与は、利用ユーザの観点を抽出するために行う。例えば、新型自動車の発売というニュース記事であっても、閲覧する人によって、自動車の開発に関連する話題か、ただ単に自動車業界全体の話題と捉えるのかは異なる。そのような観点の違いから、利用者の持つ潜在的な興味を抽出し、さらに対象コミュニティの粒度を推定する。

ユーザへ出力するニュース記事は、1 週間毎に 100 件、[6]などで取り上げられたものを取得しておく。本研究では、各ニュース記事に対して tf-idf 値に基づく候補タグをあらかじめ付与しておき、利用ユーザによって候補タグを選択してもらう。選択された候補タグは、ユーザ毎にタグリストへ保管される。保管されたタグの内容により、利用ユーザの興味と関連が強い記事をその都度出力する。

ニュース記事の出力とタグ付与という一連の流れを繰り返すことにより、タグリスト内に含まれるキーワードから、利用者の興味の範囲を限定出来る。

3.2 ユーザの興味の表現方法とタグの分類

最終的な推薦を行う際には、タグリスト内のキーワードを分類し、ユーザの興味を表現し、具体的な企業名と対応させる。タグを業種関連タグ、職種関連タグ、その他の 3 種類に分類する。そして、業種+職種+その他でユーザの興味の範囲が表現される。

3.3 タグ解析に基づく意外な推薦コンテンツの発見

本研究では、利用者にとって関連性が高い企業名の発見だけでなく、特定コミュニティから支持されている企業名の発見や、利用ユーザの視野外の企業の発見も目的とする。このような検知を行うことで、“世間一般では関心が低いが、特定業種を志望する学生の関心が高い企業の発見”や、“ユーザ自身は発見出来ていな

いが、ユーザの興味の範囲内に存在する企業”を発見することも可能になる。

そこで、利用ユーザごとに保管しているタグの内容を比較し、類似するユーザを探索する。この際、利用ユーザにとって想定していない業種であるが、自分の興味の範囲内に存在する企業を発見するために、業種関連タグを省いた状態でのタグ比較を行い、異なる業種を志望していても、職種やその他の興味が類似するユーザを発見する。

4 おわりに

本稿では、ユーザにとってふさわしい対象コミュニティ粒度を推定に基づいた情報推薦システムの提案を行った。今後システムの有効性を検証する実験を行うと共に、就職情報の推薦においてその時々で話題になっている事柄と特定企業名とのマッチングを可能にするための手法を検討したい。さらに、結婚式場などのブライダル情報や、音楽情報、書籍情報など、異なる分野への適用を検討する。

謝辞

本研究の一部は、科研費補助金若手研究 (B) (課題番号: 20700089) による。また、本研究を遂行するにあたり、解析対象となる CGM データを (株) プログウォッチャーよりご提供いただいた。ここに記して謝意を表します。

参考文献

- [1] 志甫谷匠, 中島伸介, 角谷和俊, トレンド分析および対象コミュニティ推定に基づく情報推薦システム第 149 回データベースシステム研究発表会, 2009 年 11 月
- [2] 中島伸介, 稲垣陽一, 草野奉章, プログァーの熟知度に基づいたブログランキング方式の提案 DEWS2008 B10-6:
- [3] 関口裕一郎, 川島晴美, 奥田英範, 奥雅, ブログ発信者の特徴を利用した話題抽出手法 DBSJ Letters Vol.5 No.1:
- [4] 高木充, 森康真, 田村慶一, 黒木進, 北上始, ブログユーザ空間からの頻出なコミュニティ抽出法 DEWS2007 D3-7:
- [5] クチコミポータル: SHOOTI. <http://shooti.jp/>
- [6] Yahoo!ニュース. <http://headlines.yahoo.co.jp/>