

# 研究支援システム Papits における 文献メタ情報を用いた論文推薦機構の試作

渡邊倫 大園忠親 新谷虎松

名古屋工業大学 知能情報システム学科

e-mail: {watanabe, ozono, tora}@ics.nitech.ac.jp

## 1 はじめに

本論文では、研究支援システム Papits<sup>1</sup>における論文推薦機構について述べる。論文推薦において、ユーザプロファイルの作成に関わるユーザの負担を軽減する必要がある。従来ブラウザのブックマークを利用した手法 [1] などが提案されている。本論文では、ユーザが所持する論文とその論文に含まれるメタ情報からユーザプロファイルを作成することで、ユーザの負担なくユーザプロファイルを作成する手法と、本ユーザプロファイルを用いた論文推薦機構を提案する。本提案手法では、ユーザの興味と、論文の重要度を組み合わせて用いることによりユーザプロファイルを作成する。本論文では、本ユーザプロファイルに基づく類似度計算アルゴリズムを示す。さらに、従来の類似度計算手法との比較実験を行い、本手法と Papits における論文推薦機構の有用性を示す。

## 2 論文推薦機構

本論文では、ユーザが所持する論文の Abstract と、その参照先の論文の Abstract を用いて、ユーザプロファイルを作成する。ユーザの興味は、ユーザが所持する複数の論文から得られる共起シソーラスとしてモデル化する。また、論文の重要度は、WWW 上で研究文献に関する情報を公開している ResearchIndex<sup>2</sup>におけるメタ情報を利用して計算することができる。本提案手法では、メタ情報として参考文献を用いる。論文の参考文献リストは、その論文と関連のある論文リストと考えることができる。本提案手法では、論文のメタ情報である参考文献を用いる。ユーザの所持する論文  $X$  が参照する  $Y$  の対を、ユーザプロファイルとして利用する。 $X$  に対する  $Y$  は一般的に複数存在する ( $Y_1, Y_2, \dots$  とする) 論文の対を作成するため、 $Y_1, Y_2, \dots$  から 1つ選択する必要がある。選択方法として、 $X$  と  $Y_1, Y_2, \dots$  の類似度を比較し、類似度が最大の  $Y$  を選択する。そ

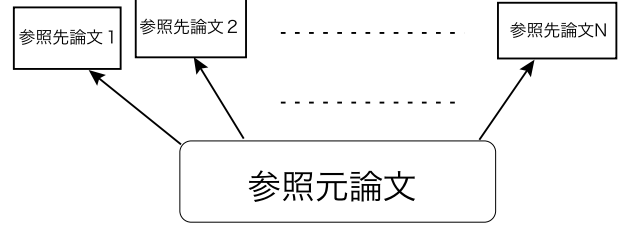


図 1: 参照の概念

の概念を図 1 に示す。この論文の組み合わせを、各ユーザのプロファイルとして用いる。参考文献リストから、ユーザが所持する論文に最も類似度が高い論文を選択するのは、ユーザが所持する論文に、より類似した論文を推薦するためである。このユーザプロファイルと各論文プロファイルを次に提案する類似度計算アルゴリズムにより比較することで、ユーザに論文の推薦を行う。

## 3 類似度計算アルゴリズム

本提案手法では、ユーザが所持する論文ファイル  $p$  が与えられると、その論文の参照先の論文の集合  $\{q_1, q_2, \dots, q_n\}$  とのそれぞれの類似度を計算し、もっとも類似度が高い論文  $q_p$  を選択する。そして  $p$  と  $q_p$  の組み合わせを用い、ユーザプロファイル  $U_p$  を作成する。ここで次のように、情報源  $T_x, T_y$  の類似度を測定する式を定義する。

$$\text{sim}(T_x, T_y) = \frac{1}{n} \cdot \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}} \quad (1)$$

ここで、 $x_i, y_i$  は情報源  $T_x, T_y$  における語の共起関係に基づく、マッピングマトリクスの  $i$  行目の要素である。このマッピングマトリクスは情報源  $T_x, T_y$  に出現する語の総数  $n$  からなる  $n$  正方行列である。

$$q_p = \omega(\max_{i=1}^n \text{sim}(p, q_i)) \quad (2)$$

式 (2) は、ユーザが所持する論文ファイル  $p$  とそのファイルの参照先である論文の集合  $\{q_1, q_2, \dots, q_n\}$  の中でもっとも類似度が高い論文  $q_p$  を選び出す式である。

$$U_p = p * q_p \quad (3)$$

<sup>1</sup>On a paper recommendation mechanism using the reference information in the research support system Papits  
Satoshi WATANABE, Tadachika OZONO, Toramatsu SHINTANI

Dept. of Intelligence and Computer Science, Nagoya Institute of Technology, Gokiso, Showa-ku, Nagoya 466-8555 JAPAN

<sup>2</sup>Papits は PAPER Information Tailor System の略である

<sup>3</sup><http://citeseer.nj.nec.com/cs>

式 (3) はユーザが所持する論文  $p$  とその参照先の論文  $q_p$  を組み合わせて作成したユーザプロフィール  $U_p$  を示す式である。  $x * y$  は、情報源  $x$ ,  $y$  を結合したものである。

$$\text{similarity}(P_p, U_p) = \text{sim}(P_p, U_p) \quad (4)$$

本提案手法における類似度  $\text{similarity}(P_p, U_p)$  は式 (4) のように表すことができる。

## 4 評価

### 4.1 実験手法と評価方法

本章では、本提案手法の有効性を確認するために行った実験方法と評価方法について述べる。本論文では、評価実験を通じて、以下のことを検証する。

#### ベクトル空間モデルとの比較

従来のベクトル空間モデルに基づく手法に対する、提案した類似度計算手法の有効性の確認

#### 文献メタ情報を用いない時との比較

文献メタ情報を用いず、共起シソーラスのみを用いる手法 [2] に対する、提案した類似度計算手法の確認

### 4.2 比較実験

本実験では、実験対象者（実験者）が所持している、論文 3 本を利用する。これは、本実験が汎用性のあるものであることを確認するため、実験者おのおのが所持する論文ファイルを実験に使用することにした。これらの論文を参照元のファイルとし、参照先の論文でもっとも類似度が大きい論文とを組み合わせ、ユーザプロフィールを作成する。さらに ResearchIndex に登録されている論文を利用する。研究支援システム Papits のデータベース内に存在する ResearchIndex のメタ情報を含むデータをランダムで 60 本抽出する。そして、これらランダムで抽出された論文おのおのの論文プロフィールを作成する。このようにして作成されたユーザプロフィールと、論文プロフィールを類似度計算により比較することで、論文の推薦を行うことができる。実験では、各実験者の所持する論文の abstract の組み合わせから作成されたユーザプロフィール 2 つから、残りの 1 つのユーザプロフィールを導き出せるかどうかを測定する。すなわち 3-fold cross validation（交差検定）を行う。交差検定における値は、N-best 候補適合率を用いる。3-fold cross validation においてユーザプロフィールとして用いない 1 つのユーザプロフィールと Papits からランダムで抽出した論文データ 60 本の各 abstract から作成された論文プロフィールとを類似度比較し、その値を順位付けする。そのときの順位

	(i)	(ii)	(iii)
実験者 1	0.207936508	0.672619048	0.676767677
実験者 2	0.1	0.674242424	0.672619048
実験者 3	0.364721485	0.089631881	0.107638889
実験者 4	0.024957004	0.128500452	0.566666667
平均	0.174403749	0.391248451	0.50592307

表 1: 各手法の比較実験の結果

の逆数を値として用いる。その 3 回の検定の平均値をとりベクトル空間モデル、共起シソーラス（論文のメタ情報を利用しない）各手法との比較検討を行うことで、本提案手法の有効性を確認する。

### 4.3 実験結果と評価

本提案手法の実験結果を表に示す。表は (i) ベクトル空間モデルに基づく手法、(ii) 共起シソーラスを用いた手法、そして (iii) 本提案手法である論文メタ情報を用いた手法、それぞれを用いたときの値が表示されている。実験結果を見ると、実験対象者によってベクトル空間モデルによる手法が適する者や、共起シソーラスを用いた手法が適する者もあった。しかしながら、全体の平均を見ると、ベクトル空間モデルに基づく手法、共起シソーラスに基づく手法に比べ、本手法ではユーザプロフィールの作成における精度の向上が見られた。研究支援システム Papits において、ユーザが所持する論文をユーザプロフィールとして用いる点において、有用であると思われる。

## 5 おわりに

本論文では、**論文メタ情報を用いた類似度計算手法を提案**し、アルゴリズムの説明と、評価実験について述べた。本提案手法では、**参照元の論文の Abstract と参照先論文の Abstract** の組み合わせからユーザモデルを構築する。本提案手法では、従来の手法である**ベクトル空間モデル**の基づく手法や、**共起シソーラス**に基づく手法に比べ、有用なユーザプロフィールの作成手法であり、Papits における論文推薦機構の有用性が確認できた。

## 参考文献

- [1] 濱崎雅弘, 武田英明, 河野恭之, 木戸出正継, “ブックマークを用いた人の繋がりの発見手法”, 人工知能学会全国大会 (第 15 回) 論文集, (2001).
- [2] 渡邊倫, 大園忠親, 新谷虎松: “共起シソーラスに基づく類似度比較を用いた情報推薦機構の試作”, 平成 14 年度電気関係学会東海支部連合大会講演論文集, pp259,(2002).