

協調フィルタリングを用いた共同研究支援システム

岸本 義晴[†] 阿部 淳也^{††} 堀 幸雄^{†††} 今井 慈郎^{†††}

香川大学工学部[†] 香川大学工学研究科^{††} 香川大学総合情報センター^{†††}

1. はじめに

近年、研究に関する様々な情報が Web 上に公開されるようになったことで、大学や企業などの各研究機関では研究に関する情報や成果を整理し、紹介するようになった。またそれにとともに、Know-Who システムなどの登場により誰が何に詳しいか、どのような研究を行っているのが簡単に調べられるようになった[1]。しかし従来のシステムは各研究者についての情報を検索するものであり、研究者間での交流を支援するものではない。

本研究では、組織内で行われている様々な研究テーマを組み合わせることで新たな共同研究テーマを提示するシステムを提案する。このシステムを用いることで、新たな組織内での研究交流の支援になることが期待できる。研究テーマの組み合わせは協調フィルタリングを用いることで実現する。

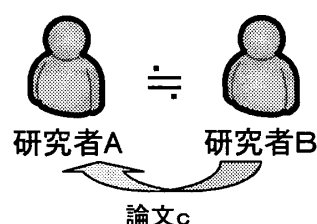
2. 協調フィルタリングの活用

協調フィルタリングとは、ユーザの嗜好を過去の行動という形で記録し、そのユーザと似たような行動を取っているユーザの嗜好情報をもとに、ユーザの嗜好を推測するシステムである。関連研究として、ネットニュースの記事の推薦システムである J. Riedl 氏の GroupLens[2]や、E-mail やインターネット上のニュースのフィルタリングを行う目的で開発された Tapestry がある。

3. システム概要

本システムはユーザが登録されている研究者 A を選択すると、その研究者と似た傾向の研究を行っている研究者 B を探し出しその研究者の論文 c を推薦する機能を持つ。推薦する論文は、

研究者 B が行っていてかつ研究者 A にとって未知の内容の研究となる。研究者 A と研究者 B は行っている研究に似た傾向を持っているので、論文 c は研究者 A が行いたいと思える内容の研究であると言える。



研究者 A と研究者 B が論文の評価値が似ている場合、B が行っている論文は A も好む内容であると推測できる。

図 1 : システムの利用目的

4. 共同研究支援の流れ

4.1 事前処理

(1) 論文取得

組織内の研究者総覧より、研究者の研究キーワードおよび所属学会を抽出し、研究者が所属している学会が発行している学会誌から過去 5 年分の論文を収集する。

(2) 論文キーワード作成

取得した論文とその論文に対応するキーワードを作成する。キーワードは取得した論文の抄録に対して形態素 N-gram で名詞句を抽出し、TF-IDF 値を算出する[3]。このとき上位 10 個をキーワードとする。

(3) 著者キーワード作成

取得した論文の第一著者に対応するキーワードを作成する。キーワードは(2)と同様に論文の抄録に対して形態素 N-gram で名詞句を抽出し、TF-IDF 値を算出して上位 10 個をキーワードとして登録する。

(4) 著者-論文行列作成

(2)および(3)で作成した著者キーワードと論文キーワードから著者-論文行列を作成する。重みは著者キーワードと論文キーワードとの一致数とする。

Support System on Joint Research Based on Collaborative Filtering

Yoshiharu Kishimoto[†], Junya Abe^{††}, Yukio Hori^{†††}, Yoshiro Imai^{†††}

[†] School of Engineering, Kagawa University

^{††} Graduate School of Engineering, Kagawa University

^{†††} Information Technology Center, Kagawa University

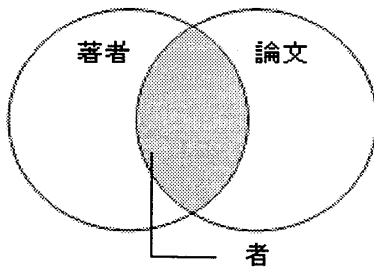


図2：著者-論文行列の評価値

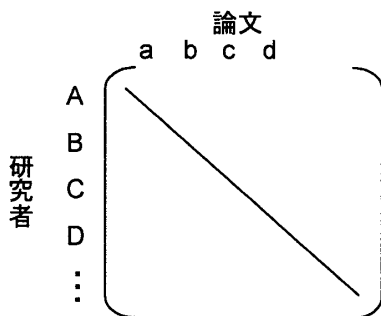


図3：著者-論文行列

ムについて述べた。今後は、システムの設計と実装、ユーザインタフェースの開発ならびに評価の検討が課題となる。

参考文献

[1] KnowWho using RDF, ISWC(Intl. Semantic Web Conference) poster, 2003
 [2] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl. "GroupLens: Open Architecture for Collaborative Filtering of Netnews" In Conference on Computer Supported Cooperative Work, pp. 175-186 (1994)
 [3] D. Goldberg, D. Nichols, B. M. Oki, D. Terry, Using Collaborative Filtering to Weave an Information Tapestry, cacm, Vol. 35, No. 12, pp. 51-60, December 1992
 [4] 山本一晴, 獅々堀正幹, 小泉大地, 北研二 「URLの類似性に着目したWWW空間からの関連語自動収集手法」, 電子情報通信学会 第18回データ工学ワークショップ c3-4

4.2 協調フィルタリング処理

(1) 著者間の類似度算出

4.1 で作成した著者-論文行列から著者の評価値ベクトルの相関係数を著者間の類似度として利用する。評価値ベクトルの共分散を各ベクトルの標準偏差の積で割ることにより相関係数を求める。ここで相関係数を r とすると、 r は次式で求められる。

$$r = \frac{\sum_x (R_{ix} - R_i)(R_{jx} - R_j)}{\sqrt{\sum_x (R_{ix} - R_i)^2} \sqrt{\sum_x (R_{jx} - R_j)^2}} \dots (1)$$

(2) 予測評価値算出

著者-論文行列の重み 0 の要素の評価値を算出する。次式がその計算式となる。

$$R_{ix} = R_i + \frac{\sum (R_{jx} - R_j)r_{ij}}{\sum |r_{ij}|} \dots (2)$$

研究者 i の論文 x に対する評価値 R_{ix} を研究者 i 以外の研究者の評価値に類似度の重みをつけて足し合わせて計算している。

5. おわりに

本稿では、研究者に新たな共同研究テーマを推薦することを目的とする共同研究支援システ