

論文リポジトリに基づく研究支援のための 対話的ユーザモデル構築手法の提案

鈴木 亮詞[†] 工藤 聖広[‡] 辻野 友孝[‡] 清水 堅[‡]
白松 俊[‡] 大園 忠親[‡] 新谷 虎松[‡]

名古屋工業大学 工学部 情報工学科[†] 名古屋工業大学大学院 工学研究科 情報工学専攻[‡]

1 はじめに

本研究では、研究支援のための論文推薦を目的としている。本研究では研究室における研究活動支援のためのシステムである Papits¹が開発されている。Papits ではユーザの閲覧論文、執筆論文からユーザモデルを構築する。構築したユーザモデルと論文との類似度を計算することで推薦を行っている。そのため、研究室に配属されたばかりの学生のようにユーザモデルが構築されていない学生に対する支援が不十分であった。本稿ではそのようなユーザを初心者と呼ぶ。本研究では、初心者に対する推薦機構を新たに構築することにより、ユーザモデルを構築し推薦を行うことを目指す。初心者に対する推薦として、まずは初心者の興味分野の特定をする。アプローチとしては、初心者に論文リポジトリ内のそれぞれの分野の論文を推薦し、どの論文に興味があるかを回答させる。回答に基づき、興味分野内の複数の部分的な分野に関する論文を推薦し、さらに興味分野を詳細化していく。ここでの課題としては、初心者にどのような論文を提示すべきかが挙げられる。例えば、背景知識の乏しい初心者でも読めるような論文を選択するなどの工夫が必要である。

2 ユーザモデルについて

Papits はユーザが閲覧した論文を用いてユーザモデルを構築し、論文推薦を行う。情報推薦にはユーザをプロファイリングすることが重要である [1]。ユーザが閲覧した論文はユーザの興味を知るのに有用である。Papits のユーザモデル構築手法を示す。論文のアブストラクトに含まれる語をノード、同じアブストラクト内の語の共起関係をリンクとしたネットワークを構築する。ユーザが Papits に論文を登録した時にネットワークを更新する。その際、ネットワークに同じノードまたはリンクが存在する場合は重複させない。構築したネットワークはスケールフリー性を持つ [2]。構築したネットワークをユーザモデルとしている。

Papits では研究室に配属されたばかりの学生のようにユーザモデルが構築されていない学生に対する支援が不十分であった。そのため本稿では、システムが初心者に対して質問をし、それに対する初心者の回答からユーザモデルを構築する手法を提案する。論文リポジトリ内の論文からユーザモデル構築木 (図 1) を作成する。ユーザモデル構築木の各ノードは論文集合とその論文集合の代表論文を持っている。質問としてユーザにノードの代表論文を提示する。ユーザの選択した代表論文を含むノードに進んでいき、そのノードの子ノードの代表論文をユーザに提示する。このように論文の提示、ユーザによる論文の選択、ノードの移動を複数回繰り返す。ユーザは質問に対して論文を選択しないことも可能とする。論文を選択しなかった場合、提示する論文を変更する。また、

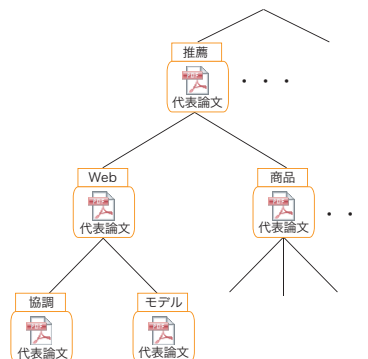


図 1: ユーザモデル構築木

同じノードにおいてユーザが論文を選択しなかった回数がある値を超えた場合に親ノードに戻る。その際に、元のノードに進むきっかけとなった論文をユーザモデルの構築に使用する情報から除外する。得られたフィードバックからユーザモデルを構築する。

初心者が論文を読む際、熟達者に与えられた論文を読むことがある。熟達者にとっては有用な論文であるが、初心者にとっては論文が難解であり自身の興味との関係性を判断することが困難である場合が考えられる。そのため、質問に対する初心者のフィードバックを用いることで初心者に適した論文を提示する。

3 ユーザモデル構築木について

3.1 前処理

本稿では、日本語論文のアブストラクトを対象とする。論文は自然言語で書かれているため、そのままでは扱いにくい。そのため、全論文に対して以下の処理を行う。まず、MeCab²を用いて形態素解析を行う。「名詞、サ変接続」、「名詞、一般」、「名詞、固有名詞、一般」と判断された単語だけを用いる。次に、論文でよく使用される「研究」、「手法」、「提案」などのストップワードの除去を行う。

3.2 作成手法

論文リポジトリ内の類似した論文同士を群平均法により階層的にクラスタリングをする。論文のアブストラクト内の単語の出現頻度を重みとしたベクトルを作成し、それらのユークリッド距離を論文間の類似度としてクラスタリングを行う。階層的なクラスタをユーザモデル構築木とする。木の各ノードは各クラスタの代表論文を持っている。代表論文は最初ランダムに決定する。質問に対してユーザが論文を選択しなかった場合、その論文は初心者に対して適した論文ではないと考えられる。そのため、ユーザが論文を選択しなかった場合に論文の重みを変更する。重みは最初全て同じ値である。重み変更時に重みを α 倍する ($0 < \alpha < 1$)。次回以降はこの重みに従って提示論文が決定される。これにより背景知識の乏しい初心者でも読めるような論文を代表論文とすることが可能となる。

[†]Interactive Construction of User Model for Supporting Study based on Paper Repositories

Ryoji SUZUKI, Kiyohiro KUDO, Tomotaka TSUJINO, Ken SHIMIZU, Shun SHIRAMATSU, Tadachika OZONO, and Toramatsu SHINTANI

Dept. of Computer Science, Nagoya Institute of Technology

¹Papits は PAPER Information Tailor System の略である

²<http://mecab.sourceforge.net/>

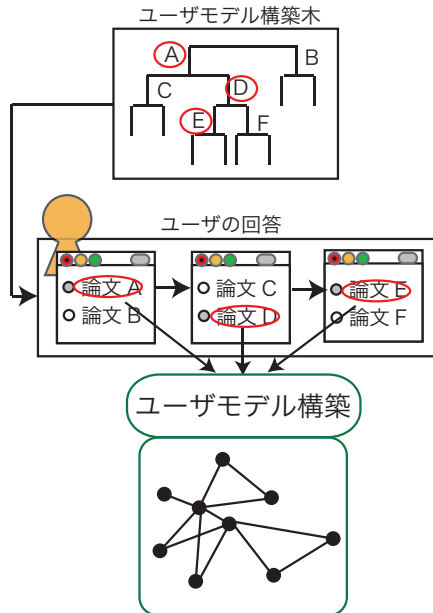


図 2: ユーザモデル構築の概要

3.3 ユーザモデル構築

ユーザモデル構築の概要図を図 2 に示す。Papits と同様に論文のアブストラクト内の語をノード、同アブストラクト内の単語の共起をリンクとしネットワークを構築する。ネットワークの更新はユーザが選択した論文をネットワークに反映させる時に行う。構築したネットワークを初心者の仮のユーザモデルとし、初心者に対する論文推薦に用いる。

3.4 類似度計算

構築したユーザモデルを用いて論文推薦を行う。ユーザモデルと論文の類似度計算にはコサイン尺度を用いる。

$$\text{sim}(U_X, P_Y) = \frac{U_X \cdot P_Y}{\|U_X\| \|P_Y\|} \quad (1)$$

式 (1) における U_X はあるユーザ X のユーザモデルを表すベクトルである。 P_Y はある論文 Y を表す特徴ベクトルである。ベクトル U_X の要素 i はユーザモデルを表すネットワークのノードの持つ重みである。重みは式 (2) で求められる。

$$\text{weight}(U_X, i) = \frac{k_i}{\sum_{j=1}^n k_j} \quad (2)$$

式 (2) の k はノード i が持つリンクの数である。ここではノード i が持つリンク数を語の頻度と見なす。ベクトル P_Y の要素 i は論文 Y のアブストラクト中出现する語の出現確率とする。式 (1) を用いて論文リポジトリ内の論文とユーザモデルとの類似度を計算し、類似度が高い論文をユーザに推薦する。本システムでの推薦のタイミングは、ユーザにより推薦機能が呼び出された時とする。

4 実行例と考察

システム実行例を図 3 に示す。図 3 は初心者がシステムに初めてアクセスした時に行われる。質問として論文のタイトルとその論文のアブストラクトを表示している。表示する論文はシステムの論文リポジトリに存在する論文である。ユーザは選択した論文のラジオボタンにチェックを入れることで論文を選択する。選択後、ユーザが「次へ」ボタンをクリックすることでシステムはユーザが選択した論文を受け取る。

実際の例としてある初心者 A の行動を考える。初心者 A がシステムの質問に対して「プレゼンテーション蓄積検索シ

Papits2.0

- 検索
- プレゼンテーション蓄積検索システムにおける適合度計算の改善
- HTMLテキストの重要文を用いた画像ラベリング手法
- 推薦

タイトル	アブストラクト
プレゼンテーション蓄積検索システムにおける適合度計算の改善	我々は、講義・講義資料と動画をメタデータで統合するプレゼンテーション検索蓄積システム UPRISE を提案してきた。本稿では、UPRISE での検索精度の向上をを目的とし、これまでの適合度計算手法に、シーンのプレゼンテーション単位で出現頻度と、説明を伴わないスライドの出現を考慮する改良を行う。さらに本論文ではそれれに対する実験を行い、有効性を確認する。
HTMLテキストの重要文を用いた画像ラベリング手法	現代社会において「マルチメディア」という言葉をよく耳にする。その言葉の本質にはかなり幅広い内容を含んでいる。しかし、確かであることは我々にとって直感的で分かりやすく、しかも多くの情報をもつ映像、画像などの視覚的な情報がこれからの情報処理の中心になっていくことである。そこで本論文では、Web ページ上の画像と関連テキスト領域中の重要文との関係に着目し、この重要文に基づいて画像へのラベル付けを行うことで、Web 上の画像検索を支援する画像ラベリング手法を提案する。
どれでもない	

次へ

図 3: 実行例

テムにおける適合度計算の改善」、「プレゼンテーション・シナリオに注目したスライド作成支援」、「プレゼンテーションスライド情報の構造化」を順に選択したとする。システムはユーザが選択した 3 本の論文から仮のユーザモデルを作成する。論文リポジトリ内の論文と作成した仮のユーザモデルの類似度計算を行った。類似度計算の結果上位 3 件は順に「トラックバックコミュニティにおける特徴的なブログ記事集合の抽出について」、「繰返し構造に基づいた Web ページの構造化」、「類似画像とキーワードを利用した Web 画像の説明文抽出」であった。ここでは、ユーザモデル作成に用いた論文は除いてある。作成された初心者 A のモデルは「構造」、「プレゼンテーション」、「説明」がハブとなったネットワークになっていた。今回の推薦結果では類似度の計算に出現頻度だけを用いているため、多く出現した語である「構造」、「説明」の影響を強く受けていると考えられる。また、論文のアブストラクト中に含まれる語の数の差の影響も見られる。

5 まとめ

本研究では研究支援のための論文推薦を目的とし、初心者に対する論文推薦のための対話的ユーザモデル構築手法を提案した。初心者はユーザモデル構築のための情報を持っていない。そのため本手法では、システムが初心者に提示した代表論文を初心者が選択し、その回答を用いて仮のユーザモデルを構築した。これにより、ユーザモデル構築のための情報を持たない初心者に対しても論文を推薦することが可能となった。また、初心者の回答をフィードバックすることで背景知識の乏しい初心者でも読めるような論文を選択することが可能となった。

本稿ではクラスタリングやユーザモデル構築に論文のアブストラクトのみを用いた。今後の課題として、推薦の精度をあげるために、論文全文を用いてクラスタリングやユーザモデル構築をする必要がある。

参考文献

- [1] 土方嘉徳: “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”, 人工知能学会論文誌, 669-701, 2004
- [2] 渡邊倫, 伊藤孝行, 大園忠親, 新谷虎松: “研究活動におけるスケールフリーネットワークを用いたユーザモデルの試作”, 日本ソフトウェア科学会第 20 回大会論文集, 163-166, 2003