

# 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

User Models Using Scale-free Network in Research Activities

渡邊倫<sup>†</sup>, 伊藤孝行<sup>†</sup>, 大園忠親<sup>†</sup>, 新谷虎松<sup>†</sup>

Satoshi Watanabe, Takayuki Ito, Tadachika Ozono and Toramatsu Shintani

<sup>†</sup> 名古屋工業大学大学院工学研究科

Graduate School of Engineering, Nagoya Institute of Technology

{watanabe, itota, ozono, tora}@ics.nitech.ac.jp

長期の行動において、ユーザの関心は常に変化するため、ユーザモデルを動的に修正する必要がある。本論文では、スケールフリーネットワークを用い、ユーザの関心を常に監視し、ユーザモデルを動的に構築することにより、長期の行動で変化するユーザの興味を表現する。スケールフリーネットワークとは、ノードとリンクを持ち、優先的選択に応じて成長する特徴を持ったネットワークである。本論文では、ユーザが所持する論文を用い、語の共起関係のスケールフリーネットワークを構築する。本論文で構築するスケールフリーネットワークは、語をノード、語の共起関係をリンクとする。ここで、ネットワークの成長とはユーザが所持する論文、およびユーザが執筆した論文を順次ネットワークに追加することである。そして、構築された随時変化するネットワークにおけるノードの適応度を計算し、動的ユーザモデルとして利用する。

## 1 はじめに

現在、WWW はさまざまな情報の公開場所であり、研究に関する情報も数多く存在する。また、論文情報を公開しているサイト [1][2] も存在し、研究成果として論文の PDF ファイルを公開する人も多い。そのため、個々のユーザが必要とする情報を効率よく収集するための有効な手法としてユーザモデルが利用されている。しかし、ユーザモデルを構築するにあたって、ユーザの状態は変化するため、ユーザモデルをそれに応じて変更・修正する必要がある。

そこで、本論文ではユーザが過去に使用した全ての語の頻度及び、新近性 [3] を利用する。新近性とは、ユーザが最近使用した語は、使用されやすい性質を持つことである。

本論文では、語の頻度及び、新近性を表現するため、スケールフリーネットワーク [4][5] を用い、ユーザモデルの構築を試みる。スケールフリーネットワークとは、成長 (growth)、優先的選択 (preferential attachment) を特徴として持つネットワークである。さらに適応度 (fitness) を用いる。適応度とは、ネットワーク内に新たなノードが付加される時、ネットワークにおけるノードが持つエッジを獲得する確率である。本論文では、ユーザの知識を、ユーザが所持する論文、および執筆した論文を用いて表現する。ネッ

トワークのノードをユーザが所持する論文に含まれる語、また、エッジを文中における語の共起を用い表現する。

ユーザが所持する論文、および執筆した論文を順次ネットワークに加えることでネットワークのノード及びエッジの数は増加する。適応度をもつネットワークにおいて、ノード及びエッジが増加する際、適応度の高いノードのエッジが増加する。以上から適応度が高いノードがユーザの興味であると考え、ノードの集合をユーザモデルとして提示する。

スケールフリーネットワークは新たなエッジの付加によってネットワークが成長 (growth) し、新たなノードがリンク先として選ぶ確率がノードによって異なる傾向である優先的選択 (preferential attachment) を持つ。スケールフリーネットワークの概念を図 1、及び以下の 2 要素を用いて説明する。

成長とは、スケールフリーネットワークは、図 1 の A のように、エッジを持たない、 $m_0$  個のノード ( $i = 0, 1, \dots, m_0$ ) からスタートする。図の○は新規のノードを表し、●は既に存在するノードを表す。そして、図 1 の  $A \rightarrow B \rightarrow C$  のように、時間ステップごとにノードを付加する。各ステップで付加するノードはすでにネットワークに存在するノードのうち  $m$  個のノードとリンクする。

優先的選択とはステップ  $\tau$  で新たにノード ( $i =$

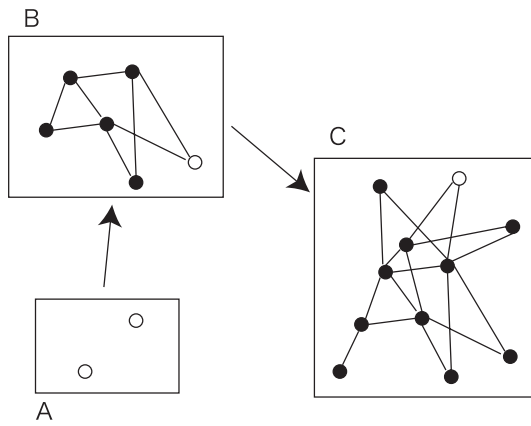


図 1: スケールフリーネットワークの概念

$\tau + m_0$ ) を付加する. この時, すでに存在するノード ( $i = 0, 1, \dots, \tau - 1 + m_0$ ) のいずれかのノードとリンクを持つかを, エッジを多く持つノードとリンクする傾向である, 優先的選択を持たせる. 論文 [5] では, ノード  $i$  とリンクする確率  $\Pi_i$  をノード  $i$  が持つリンク数  $k_i$  に比例するよう定めた.

$$\Pi_i = \Pi(k_i) \equiv \frac{k_i}{\sum_{j=0}^{\tau-1+m_0} k_j} \quad (0 \leq i < \tau + m_0) \quad (1)$$

以上のように生成されたスケールフリーネットワークには次のような特徴がある.  $k$  個のリンクを持つノードの数は, 任意の  $k$  の値に対してべき法則に従い, 多くのエッジを持つノード, すなわちハブが出現する.

上記の特徴を持つネットワークは, 俳優の共演関係, 文献の引用関係, 電力網, Web のハイパーリンク構造などがあげられる [6]. 論文 [4] では, 現実のネットワークを解析し, べき分布を表す式 (2)

$$P(k) \sim k^{-\gamma} \quad (2)$$

における指数  $\gamma$  が, Web のリンクおよび被リンクの関係では,  $\gamma = 2.45 \sim 2.1$ , 映画俳優のネットワークでは,  $\gamma = 2.3$  であることを発見した.

適応度を持つスケールフリーネットワークでは, ノード  $i$  がエッジを獲得する確率を式 (3) で表す [4].

$$\Pi_i = \frac{\eta_i k_i}{\sum_j \eta_j k_j} \quad (3)$$

ここで,  $\eta_i$  はノード  $i$  の適応度を表し, ノード  $i$  がエッジを獲得する確率は, 1 で述べた優先的選択と適応度の積で表せる.

本論文では, スケールフリーネットワークにおける各ノードが持つ適応度によりユーザの興味を抽出する. これまで, ユーザモデルを用いた情報フィルタリングの研究は多く行われている. 論文 [7] は, 提供する情報内容・情報提示方式等をユーザ 1 人 1 人の特性に応じてユーザごとに動的に変更し, ユーザに適応しながら動作する WWW を提案した. 論文 [11] は, 情報システムのユーザの意図や嗜好を推定する能力の必要性を挙げ, 非決定的な挙動を示す人間の不確定性をモデルに表した. 論文 [12] は, 自然言語による対話システムにおいて, ユーザの発話から発話のプランニング過程を推論し, ユーザモデルの更新を行い, 対話への影響を検出する.

本論文の構成を以下に示す. まず, 2 では, 研究活動に必要なとされるユーザモデルを検討し, その構築手法の提案する. 3 では, 具体的な構築方法を述べる. 4 では, 本ユーザモデルを利用した文献推薦手法を示し, 5 では, 本提案手法で作成したユーザモデルの評価実験を行い, 6 で本論文をまとめる.

## 2 研究活動におけるユーザモデル

研究活動におけるユーザモデルを作成する上で, 以下のような条件を満たす必要がある [8].

条件 1 どのようなユーザ情報を対象とし, それをどのように獲得するか.

条件 2 獲得された断片的なユーザ情報をどのように統合し, ユーザモデルとして構築するか.

条件 3 ユーザの状態は変化するため, ユーザモデルをそれに応じて変更・修正する必要がある. どのようにユーザの状態変化を発見し, ユーザモデルを変更修正するか.

本論文で, 条件 1 は, 研究活動を行うユーザを対象とする. 条件 2 は, ユーザの情報を獲得する方法として, ユーザが所持する論文, および執筆した論文を用い, その論文に含まれる語, 及び文中での共起関係をそれぞれネットワークのノードおよびエッジとする. 条件 3 のユーザの状態の変化は, ユーザが論文を新たに取得した時点で, ユーザの知識を表すネットワークにノード及びエッジを付加することで得る.

ネットワークの各ノードが適応度をもつ場合, 式 (3) より, ノードがエッジを獲得する確率は, 優先的選択および適応度の積で表せる. すなわち, ノード

がエッジを多く持たない場合も、高い適応度を持つていればそのノードはエッジを持つ確率は大きくなる。本論文では、各ノードの適応度を計算し、高い適応度をもつノードの集合をユーザの興味と考える。

### 3 ユーザモデル構築法

ここでは、2で述べた、研究活動を行う上でのユーザモデルを構築する手法について述べる。本提案手法におけるノードは、ユーザが所持する、あるいはユーザが執筆する論文に含まれる語である。また、エッジは同一文中における語の共起関係とする。

まず、ユーザが所持する論文、あるいはユーザが執筆した論文の前処理として以下を行う。

1. 不要語 (stop word) のリストを使用する論文から取り除く。本提案手法では、英語文書の検索システム SMART で利用されている不要語リストを用いる [9]。
2. 接辞処理 (stemming) を行う。ここでは、Paice/Husk の手法 [10] を用いる。
3. 語をノードとして付加し、同一文中に共起する語をリンクする。

以上の手順で作成されたノード、およびエッジを、ユーザの知識としてフリースケールネットワークで表したユーザモデルに付加する。ネットワークにノードおよびエッジが既に存在する場合は重複させない。

図2は、本提案手法で作成したユーザモデルを用いて、文献データベースから、ユーザに興味のある文献を比較および選択する概念図である。ユーザモデル生成機構はユーザが所持、及び執筆した論文を基に、上述の接辞処理、および Stemming の前処理を行い、ユーザモデルの構築および修正を行う。構築されたユーザモデルは、比較・選択機能により、収集の対象となる文献データベース中の情報または、そのインデックスと比較される。ユーザが取得および執筆した論文を随時、ユーザモデルに反映させることで、常にユーザの興味がある文献を推薦することが可能となる。

具体例として、"John drinks water. Mary drinks wine. John find a watch. Mary find a picture." という文字列を、前述の前処理を行いネットワークとして表現すると図3となる。

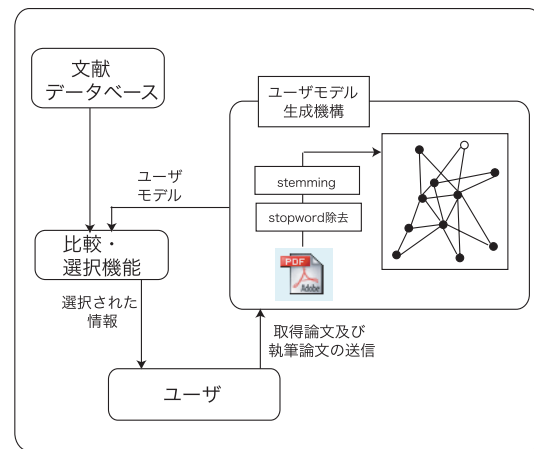


図 2: ユーザモデルの概念

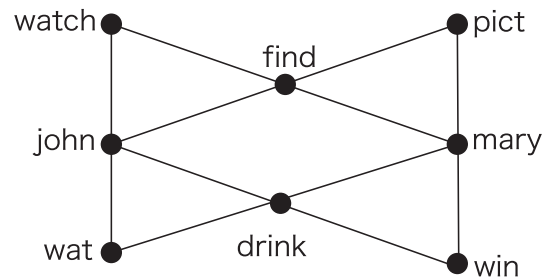


図 3: ネットワーク構成例

### 4 文献推薦への適用

ここでは、本提案手法を利用した文献推薦手法を述べる。本提案手法を利用した文献推薦では、式(3)、および文献中での語の共起関係に着目する。式(3)における、各ノードの適応度  $\eta$  を語の新近性、各ノードが持つエッジ数  $k$  を語の頻度とみなす。また、文献中の語の共起関係を利用することで、語の意味の同一性を同定することが可能である [13]。

以下に本提案手法を用いたユーザモデルと文献の類似度判定関数を示す。ユーザの所持する文献、および執筆した文献から得られたユーザモデルを用い、論文ファイルから得られた語の共起関係を比較し、類似度の高い論文を示すことにより関連性の高い論文を推薦することが可能になる。式(4)はユーザモデル  $N_X$ 、文献  $T_Y$  の類似度を表す関数である。

$$\text{sim}(N_X, T_Y) = \frac{1}{n^2} \sum_{i=1}^n \left( \sum_{j=1}^n \Pi_i \Pi_j \right) \quad (4)$$

ここで、 $n$  は3で述べた前処理を適用した後の文献  $T_Y$  に含まれる語数である。式(4)を用いて計算を

	ベクトル空間	本提案手法
平均評価値	3.1	4.2
平均計算時間 (sec)	1.2	30.5

表 1: 評価実験結果

行うとユーザモデル  $N_X$  と文献  $T_Y$  の類似度を求めることができる。この類似度を推薦する値として使用することが可能である。

## 5 評価実験

### 5.1 評価実験と評価方法

ここでは、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、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を行う。また、実験では、以下の工程で行う。まず、1 か月間に被験者が取得および、執筆した論文を収集する。これを論文を取得もしくは、執筆した順に処理し、3 で述べた、不要語の除去、stemmingを行う。論文に含まれる語をノード、文中における共起関係をエッジとし、ネットワークに付加する。以上のようにユーザモデルを構築する。

本実験では、研究支援システム Papits[14] のデータベースに存在する論文を用いる。式 (4) を用い、Papits のデータベース中に存在する論文と各被験者のユーザモデルの類似度の高い文献を被験者に提示し、被験者 10 人の評価を 5 段階で得る。また、被験者 1 人当たりの文献推薦を実現する処理時間を示す。

### 5.2 実験結果・考察

実験の結果を表 1 に示す。ベクトル空間モデルによる類似度計算手法では平均評価値 3.1、本提案手法による類似度計算手法では平均評価値 4.2 であった。この結果から、本提案手法によるユーザモデル作成手法、及び、文献推薦手法の有用性が確認できる。

しかし、表に示した通り、ベクトル空間モデルを使用した文献推薦に比べ、計算時間が約 25 倍となる。本実験では、被験者の 1 か月分に取得、あるいは執筆した論文を利用したが、さらに長期間の文献データをユーザモデルとして利用する場合、計算時間は増加すると考えられる。

## 6 おわりに

本論文では、スケールフリーネットワークを用いたユーザモデルを提案した。ユーザが随時取得する論文および、執筆した論文をユーザの知識として用い、ノード及び、エッジをネットワークに付加する際の、各ノードが持つ適応度が高いノードの集合をユーザの興味として提示する。

本論文では、研究活動におけるユーザモデルについて、議論を行った。今後は本提案手法が研究活動以外にも利用できるよう、汎用性のあるモデルについて構築を目指す。

### 参考文献

- [1] NEC Research Index  
<http://citeseer.nj.nec.com/cs>
- [2] ACM Portal  
<http://www.acm.org/dl>
- [3] Ramon Ferrer i Cancho and Ricard V. Sole, "The small world of human language," Proceedings of Royal Society of London B, 2261-2265, (2001).
- [4] Reka Albert and Albert-Laszlo Barabasi, "Statistical Mechanics of Complex Networks," arXiv:cond-matt/0106096v1,(2001).
- [5] Albert-Laszlo Barabasi, Reka Albert and Hawoong Jeong, "Mean-field theory for scale-free random networks," Physica A, vol.272, pp.173-187, (1999).
- [6] 新ネットワーク思考-世界の仕組みを読み解く-  
Albert-Laszlo Barabasi 著, 青木薫 訳, NHK 出版, (2002).
- [7] 三浦信幸, 高橋克巳, 島健一, "個人適応型 WWW におけるユーザモデル構築法," 情報処理学会論文誌, vol.39, No.5, pp.1523-1535, (1998).
- [8] 杉本雅則, "情報収集システムにおけるユーザモデリングと適応的インタラクション," 人工知能学会誌 Vol.14, No.1, pp.25-31, (1999).
- [9] 徳永健伸, "言語と計算 5 情報検索と言語処理", 東京大学出版会, (1999).
- [10] Paice, C.D., "Another stemmer", SIGIR Forum, Vol.24, No.3, pp.56-61, (1990).
- [11] 本村陽一, 原功, "確率ネットワークによるユーザモデル構築システム," 平成 12 年度 IPA 未踏ソフトウェア創造事業, (2000).
- [12] 佐川雄二, 大西昇, 杉江昇, "自然言語対話システムにおけるユーザモデルの更新に伴う対話の再プランニングに関する考察," 情報処理学会論文誌 Vol.35, No.6, pp.1042-1049, (1994).
- [13] 後藤将志, 大岡忠親, 新谷虎松, "シソーラスを用いた情報間類似性評価手法について," 第 6 4 回情報処理学会全国大会, (2002).
- [14] 渡邊倫, 大岡忠親, 新谷虎松, "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," 第 6 5 回情報処理学会全国大会, (2003).