DEIM Forum 2011 F9-4

学術論文閲覧支援システムのための関連論文推薦

鉢木　稔浩† 　太田　学† 　高須　淳宏††

† 岡山大学大学院自然科学研究科　〒 700–8530 岡山県岡山市北区津島中 3–1–1
†† 国立情報学研究所　〒 101–8430 東京都千代田区一ツ橋 2–1–2
E-mail: †{hachiki,ohta}@de.cs.okayama-u.ac.jp, ††takasu@nii.ac.jp

あらまし　我々は，論文から専門用語を抽出し解説等の有用なページへのリンクを提供する，学術論文閲覧支援システムを開発している．本稿では閲覧論文中の専門用語を用いて，その関連論文を推薦する手法を提案する．具体的には，論文中の各専門用語で検索される論文集合とそれらに出現する専門用語集合の二部グラフを生成する．この二部グラフのリンク解析に HITS アルゴリズムを利用し，推薦する関連論文を得る．実験ではベクトル空間モデルに基づく論文推薦手法等と比較し，提案手法の有効性を評価した．

キーワード　閲覧支援，情報推薦，専門用語抽出，HITS アルゴリズム

Related Paper Recommendation for a Browsing Support System of Research Papers

Toshihiro HACHIKI†, Manabu OHTA†, and Atsuhiro TAKASU††

† Graduate School of Natural Science and Technology, Okayama University
Tsushima-naka 3–1–1, Kita-ku, Okayama-shi, Okayama, 700–8530 Japan
†† National Institute of Informatics
Hitotsubashi 2–1–2, Chiyoda-ku, Tokyo, 101–8430 Japan
E-mail: †{hachiki,ohta}@de.cs.okayama-u.ac.jp, ††takasu@nii.ac.jp

Abstract　We developed a browsing support system for research papers, which extracts technical terms from a paper and presents links to useful pages such as those explaining the terms. In this paper, we propose a method to further use the extracted technical terms to recommend users papers related to the original paper. Concretely, the proposed method generates a bipartite graph consisting of papers retrieved by the technical terms, which we call related papers, and technical terms appearing in the related papers. It then ranks the related papers using the HITS algorithm for analyzing the bipartite graph and recommends users top-ranked papers. We evaluated the effectiveness of the proposed method compared with other recommendation methods in experiment.

Key words　browsing support, information recommendation, technical term extraction, HITS algorithm

1. はじめに

現在の電子図書館の多くは Web 経由でアクセス可能なものが多いが，その文書の閲覧においてオンラインであるメリットが十分に生かされているとはいい難い．例えば，検索のサービスは一つの電子図書館内で閉じていることが多く，検索できるデータベースが限定される．そこで我々は，電子図書館と Web を連係させることでより充実したサービスが提供できると考え，学術論文中の専門用語に対して解説等の有用なページへのリンクを提供することで論文の閲覧を支援するシステムを提案した [1].

本稿は，論文閲覧時にその論文の関連論文を推薦することで，更なる論文閲覧支援を行うことを目的とする．本稿では学術論文の OCR テキストを利用し，閲覧論文中の専門用語から関連論文を検索して推薦するシステムを提案する．そこでまず論文文書画像を OCR 処理して得たテキストから専門用語を抽出する．次に各専門用語で検索した論文集合と，その論文集合に出現する専門用語集合を求める．本研究では，これらの論文とそれに出現する専門用語間にリンクを生成し，この二部グラフに Web ページのリンク解析で用いられる HITS アルゴリズム [2] を適用することで，論文をランク付けして推薦する．

本稿は 2 節で関連研究，3 節で提案する論文推薦システムについて述べる．さらに 4 節で実験と評価について述べ，5 節でまとめと今後の課題を述べる．

2. 関 連 研 究

2.1 専門用語の抽出

Web 上にある一般的な文書から自動的に専門用語を抽出する手法は，多く提案されている．例えば久光らは語の話題性や分野特定性を求める指標を提案した [3]．これはある語と文書内共起する語の集合の偏りから，語の重要度を測るものである．また湯本らは出現頻度と連接頻度を用いた専門用語抽出を行った [4]．単語に連接する語，つまり単語バイグラムの出現頻度からその単語のスコアを求める．複合語の場合，複合語を構成する単語のスコアの平均をとる．これに単語または複合語自身の出現頻度も考慮して専門用語を抽出する方法を提案した．本稿の提案手法においても専門用語を抽出するが，出現頻度情報のみを用いてスコア付けする．

2.2 関連コンテンツの推薦

ある文書に対して，関連の深い Web ページや文書などのコンテンツを推薦する研究が多く存在する．矢島らはブックマークを用いて類似ブックマークおよび類似ユーザを推薦した [5]．ブックマークから抽出した特徴語空間におけるコサイン類似度から類似ブックマークを，またブックマーク集合における特徴語の偏りから類似ユーザを推薦した．近藤らは与えられた文書から重要語を抽出し，それらを外部 API に入力することでブログや動画などの関連コンテンツの推薦を行った [6]．またユーザの Web 閲覧履歴からキーワードを抽出し，関連するクエリを推薦する手法を提案した [7]．これは Wikipedia 記事の見出し語をキーワードとして，Wikipedia 内のリンク構造に改良した HITS アルゴリズムを適用し，キーワードの重要度を算出し，関連クエリとして提示するものである．

関連論文の推薦の研究では，Sugiyama らがユーザの研究についての興味を用いて学術論文を推薦するシステムを提案している [8]．これはユーザが過去に発表した論文集合とそれらの引用・被引用文献からユーザの興味を示すプロファイルを構築し，これと類似度の高い論文を推薦するものである．また難波らは論文の参照関係をリンクとした多言語データベースを構築し，HITS アルゴリズムを用いてサーベイ論文の自動検出を行った [9]．彼らはその際，検出対象であるサーベイ論文の特徴を考慮して HITS アルゴリズムを改良している．本稿ではこのようなユーザのプロファイルは用いず，論文の表題と概要から抽出した専門用語を利用して関連の深い論文を推薦する．

3. 論文推薦システム

3.1 概　　要

本節では閲覧している学術論文中に出現する専門用語集合を用いて，関連する論文をユーザに推薦する手法について述べる．本手法では，関連論文を推薦する際に HITS アルゴリズムを用いる．HITS アルゴリズムは元々Web 上のリンク関係を解析し，コミュニティを抽出するものである．しかし本手法では，論文とそれに出現する専門用語間にリンク関係を定義し，論文集合と専門用語集合という異なる二つの集合間のリンク構造を解析する．具体的には閲覧論文中の専門用語集合から得た論文集合とそれらに出現する専門用語集合の関係を解析する．

- 閲覧論文中の専門用語で検索して得た関連論文集合中で，出現回数の多い専門用語は閲覧論文と関係が深い．
- 上記のような出現回数の多い専門用語が多く出現する論文は閲覧論文と関係が深い．

図 1　閲覧論文と関連論文の関係に関する仮定
Fig. 1　Assumption on the relationship between a target paper and its related papers.

本稿では閲覧論文の関連論文を推薦するにあたり図 1 の仮定をおいた．これらの仮定から論文推薦に HITS アルゴリズムが有効であると考え，提案手法で用いた．

提案手法による関連論文の推薦は以下のように行う．

(1) 閲覧している論文 p_{target} の表題と概要から 3.2 節の手順で専門用語集合 T を抽出する．

(2) 各専門用語 $t_i \in T (i = 1, \ldots, K)$ をクエリとして，CiNii が提供する API [10] を利用し，t_i に関連する論文 $p_{ij} \in P_i (j = 1, \ldots, N)$ を最大 N 件取得する．この際，CiNii から取得する論文は被引用件数が多い順とする．またこのとき既に別の専門用語で検索され取得している論文は取得せず，ユニークな論文を最大 N 件取得する．ここで $P = \bigcup_{i=1}^{K} P_i$ とする．

(3) 各論文 p_{ij} の表題と概要から (1) と同様の方法で専門用語集合 T_{ij}^a を抽出する．ここで $T_i^a = \bigcup_{j=1}^{N} T_{ij}^a$，$T^a = \bigcup_{i=1}^{K} T_i^a$ とする．

(4) (2) で取得した論文集合 P から (3) で取得した専門用語集合 T^a へリンクをはった二部グラフを生成する．この二部グラフに HITS アルゴリズムを適用し，得られた authority 値と hub 値によって推薦する論文を決定する．適用方法については 3.4 節で詳しく述べる．

3.2 専門用語の抽出

学術論文を形態素解析器 Sen [11] を用いて形態素解析し，以下のルールに従い専門用語の候補となる特徴語を抽出する．

（1） 品詞情報が以下のものを特徴語として抽出する．
- 名詞
- カタカナ，漢字，英数字のそれぞれのみで構成される未知語

（2） (1) で抽出した語が連続する場合は連結して一つの特徴語とする．

（3） 不要語を除去する．不要語はひらがなや数字のみで構成される語，1 文字の語，除外語リストの語とした．除外語リストは「あらまし」のように論文の書式により必ず出現する語や，「2 種類」のように専門用語として不適当と考えられる語を含む．

一方 OCR テキストには，文字の認識誤りが含まれる．そこで特徴語に含まれる認識誤りを修正するために，Yahoo!ウェブ検索 [12] を利用する．Yahoo!ウェブ検索では綴りを誤った語で検索すると，「…ではありませんか？」のようにシステムが正しいと推測した語を提示してくれるサービスがある．これを利用

し，検索質問の語を提示された語に修正する．例えば「スケーラピリティ」が特徴語として抽出される．この語を検索質問として Yahoo!で検索をすると「スケーラビリティではありませんか？」と正しい語が提示されるので,「スケーラピリティ」を「スケーラビリティ」に変換する．しかし正しい特徴語で検索しても，別の語が提示されることがある．そこで誤りを含む特徴語の場合通常検索結果がかなり少ないので，検索結果の総数が 1000 件を超えないときにのみこの修正を行う．

次に抽出した特徴語を TF・IDF 法により重み付けする．本手法では特徴語 t_i の重要度 $tfidf_i$ を以下のように定義する．

$$tfidf_i = tf_i * log\frac{num}{df_i} \tag{1}$$

ここで特徴語 t_i が抽出された論文文書中におけるその t_i の出現頻度を tf_i とする．出現文書頻度 df_i は，論文検索サイト CiNii において t_i で検索したときの検索結果数，全文書数は CiNii における総収録件数を用いる．本手法で利用した時点でこの総収録件数は 13,206,916 件であった．

論文から抽出したすべての特徴語をこのスコアに基づいてランク付けする．このランキングの上位 K 件の特徴語を専門用語として抽出する．

3.3 HITS アルゴリズム

Kleinberg が提案した HITS アルゴリズムは，Page らが提案した PageRank [13] と並び，代表的な Web ページランキング手法であり，Web 上で形成されているコミュニティの発見などに応用されている．HITS アルゴリズムは「より重要なノードとの関係の方が，そうでないノードとの関係よりも，重要度への寄与が大きい」という考えに基づき，Web ページの内容とは独立して，ページ間のリンク関係を解析することで authority と hub という 2 タイプのコミュニティを抽出するものである．Web 解析における HITS アルゴリズムの authority とは特定のトピックに関する情報が豊富にある Web ページであり，一方 hub とは authority としての価値が高いページへのリンクが豊富にある Web ページのことである．

authority と hub は，以下の 2 式で定義される Web ページ p に与えられた authority 値 a_p と hub 値 h_p を値が収束するまで反復計算することで求められる．ここで Web ページ p から q へリンクされていることを，$p \to q$ と表している．これらの結果に基づいて Web ページをランク付けする．有用な authority と hub は相互に強化すると Kleinberg は主張している．

$$a_p = \sum_{q, q \to p} h_q \tag{2}$$

$$h_p = \sum_{q, p \to q} a_q \tag{3}$$

良い authority であるページは互いにリンクを張ろうとせず，比較的無名なページである hub を通じてつながっていると Kleinberg は述べている．つまり図 2 のように，両者の関係は「よい authority は複数の良質の hub によってリンクされ，また良質の hub は複数のよい authority にリンクをはっている」と再帰的に定義されている．

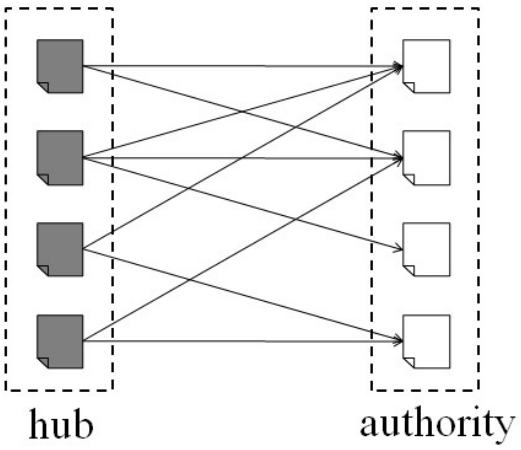

図 2　authority と hub の関係
Fig. 2　The relationship between the authority and the hub.

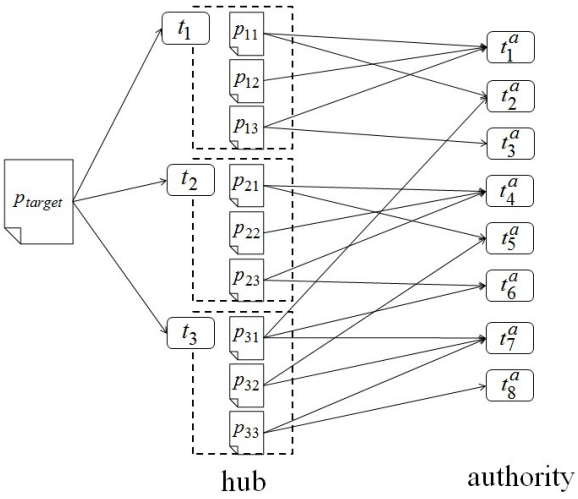

図 3　HITS アルゴリズムの論文推薦への適用
Fig. 3　Application of HITS algorithm to paper recommendation.

3.4 論文推薦への応用

提案手法では HITS アルゴリズムを関連論文の推薦に利用するため，論文集合と専門用語集合について以下のように定めた．

- 論文からその論文に出現する専門用語へリンクをはる．
- 推薦候補論文を hub，専門用語を authority とする．

この二部グラフを HITS アルゴリズムでリンク解析すると，図 1 の仮定に基づく有用な論文推薦ができると考えられる．また本来の HITS アルゴリズムでは各ページがそれぞれ入リンクと出リンクをもち，それに応じた authority 値と hub 値をもっている．しかしこの二部グラフでは論文から専門用語へのリンクしかもたないので，論文を hub，専門用語を authority を定めた．このようにして，論文集合から専門用語集合へのリンクをもった二部グラフを生成し，HITS アルゴリズムを適用した例を図 3 に示す．

この二部グラフにおいて，式 (2)，(3) を用いて a_p と h_p を反復計算により求める．また authority 値と hub 値は毎回それぞれ式 (4)，(6) で正規化し，反復計算は前回の authority 値ま

たは hub 値との差分が十分小さくなるまで繰り返す．

$$\sum_p a_p^2 = 1 \tag{4}$$

$$\sum_p h_p^2 = 1 \tag{5}$$

提案手法では論文から専門用語へリンクをはると定めたため，反復計算により hub 値でランク付けされた論文と authority 値でランク付けされた専門用語が得られる．最後的に，この hub 値のランキング上位の論文を推薦論文としてユーザに提示する．

なお推薦論文のランク付けの際，スコアが同じ場合には，以下の基準を (1)，(2) の順で評価して論文をランク付けする．

（1） CiNii から取得した際のランク

（2） 取得時に用いた専門用語の重要度

4. 評価実験

4.1 実験概要

提案手法により推薦した論文の上位 10 件について適合性の判定を行った．評価に用いる論文は電子情報通信学会論文誌 Vol.J83-D-I〜Vol.J88-D-I から無作為に選んだ 10 件とする．また本手法での HITS アルゴリズムは，抽出する専門用語数が多ければ多いほど，その論文の hub 値が高くなる性質をもつため，各論文から抽出する専門用語数の上限を定める．本手法では論文の表題，概要のみから専門用語を抽出しているが，予備実験の結果その中に含まれる平均専門用語数は一論文あたり約 10.4 語となった．よって実験では，一論文から抽出する専門用語を最大 10 語とした．

さらに提案手法の有効性を調べるために，ベクトル空間モデルを利用した手法（以下，VS モデル）および図 1 の仮定に基づく HITS アルゴリズムを適用しない手法（以下，ベースライン）を用いて同様の実験をした．これら二手法については 4.2 節で詳しく述べる．

また推薦論文の適合率の判定には，閲覧している論文との関連の強さに応じて rigid 判定と relaxed 判定の二つを用いた．rigid 判定の基準は閲覧論文の目的と同じ内容の論文であることとし，relaxed 判定は rigid 判定の基準を満たす，または閲覧論文と同じ技術や概念を利用している論文を適合とした．例えば閲覧論文が「Support Vector Machine を用いた文書分類」という内容であれば，文書分類を行う論文は rigid 判定で適合，Support Vector Machine を利用しているが，文書分類は行っていない論文は relaxed 判定で適合となる．上記のような判定基準に基づいて評価する．なお適合率はランク付けされた推薦論文の上位 k 件の適合率（p @ k）で算出する．

実験では，CiNii から取得する論文の最大件数 N を 5 件，10 件，30 件，50 件と変えて，VS モデル，ベースライン，提案手法の三手法を評価した．

4.2 比較対象とする論文推薦手法

提案手法と比較する手法は，以下の二つである．

- VS モデル

これは情報検索で広く用いられているベクトル空間モデルを利用した手法である．VS モデルは，閲覧論文と関連論文集合 P 中の各論文 p_{ij} の文書ベクトルを，それらの論文に出現する専門用語のベクトルで表現する．このベクトルを用いて文書間の類似度をコサイン尺度で計算し，閲覧論文との類似度が大きい論文を推薦論文とする．

なおコサイン尺度は二つのベクトル $\mathbf{x}$，$\mathbf{y}$ を用いて以下の式で求められる．

$$\cos\theta = \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} \tag{6}$$

- ベースライン

これは，図 1 の仮定を直接推薦論文のランキングに反映する手法である．まず閲覧論文から取得した関連論文集合中の専門用語を，文書頻度によりスコア付けする．次に各論文に現れる各専門用語のスコアの合計をその論文のスコアとし，論文をランク付けし，このランキングが上位の論文を推薦する．この手法は提案手法とは異なり，HITS アルゴリズムの反復計算をせずに専門用語と論文のスコアを算出する．

表 1　平均関連論文数

Table 1　The average number of related papers.

取得論文数/語 (N)	5	10	30	50
最大取得論文数	50	100	300	500
平均関連論文数	37.9	69.3	178.7	269.0

表 2　専門用語 t^a の平均入次数

Table 2　The average indegrees of a technical term t^a.

取得論文数/語 (N)	5	10	30	50
平均専門用語数	346.0	615.3	1514.0	2227.7
平均入次数	1.09	1.12	1.17	1.19

4.3 実験結果

まず実験において，提案手法で CiNii から取得した関連論文数の平均を表 1 に示す．最大取得論文数は専門用語数 10 語 × 取得論文数 N を示している．平均関連論文数がこれを下回っているのは，抽出した専門用語が 10 語に満たない論文，また CiNii から取得した論文数が N に満たない専門用語が存在するためである．

また提案手法では，推薦した論文が 10 件未満の場合があった．これは多くの論文の hub 値が正規化により 0 となったからである．このような論文は，他の論文と出現する専門用語に重なりがないかもしくは極めて小さい．

次に論文集合と専門用語集合の二部グラフにおける，専門用語 t^a の平均入次数を表 2 に示す．ここで平均専門用語数は，実験データ 10 件の論文の一論文あたりの専門用語集合 T の大きさで，平均入次数は，専門用語 t^a のそれである．入次数が多いほどその専門用語が多くの論文に出現することを表し，このような専門用語が authority 値と hub 値に影響を与える．表 2 より取得論文数が多くなれば平均専門用語数は増加するが，平均入次数の増加は緩やかである．本手法では入次数が多い専門用語の authority 値が高くなることから，入次数が 2 以上の専門用語，つまり複数の論文に出現する語のみの平均入次数と，

表 3　専門用語 t^a の平均入次数（≧ 2）
Table 3　The average indegrees of a technical term t^a (≧ 2).

取得論文数/語 (N)	5	10	30	50
平均専門用語数	20.0	39.9	124.3	206.4
平均入次数	2.46	2.74	3.00	3.09
専門用語集合に占める割合 (%)	6.05	6.53	8.32	9.21

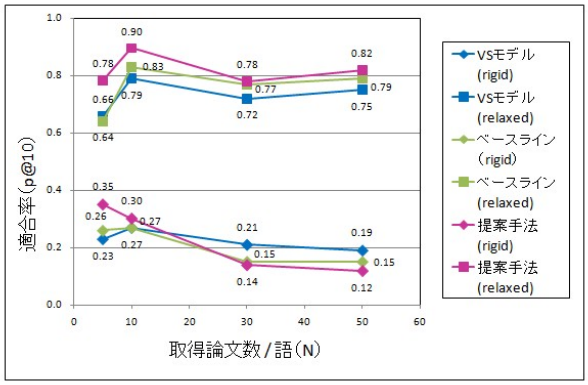

図 4　適合率の比較
Fig. 4　Comparison of precision

専門用語集合中での入次数が 2 以上の専門用語の割合を調べた（表 3）．表 3 を見ると，専門用語集合に占める入次数が 2 以上の専門用語の割合が約 6.1%から約 9.2%であることから，9 割以上の専門用語がその論文にしか出現しない語であり，重なりをもつ一部の専門用語が提案手法の論文推薦に影響を与えていることが分かる．

次に VS モデルとベースライン，提案手法について，各専門用語による取得論文数を 5，10，30，50 件としたときの推薦論文の上位 10 件の適合率を図 4 に示す．図 4 より，提案手法の推薦論文の適合率は，rigid 判定で最も良い $N = 5$ の場合で 0.35，relaxed 判定で最も良い $N = 10$ の場合で 0.90 であった．また rigid 判定では，取得論文数が多くなるほど適合率が下がり，relaxed 判定では取得論文数が 10 件より多くても，少なくても低いという結果となった．他の二手法と比較すると，relaxed 判定では取得論文数によらず最も良い結果を示している．しかし rigid 判定では取得論文数を多くすると，他の手法よりも悪くなっている．

4.4　考　　察

VS モデルを用いた手法は，取得論文数が $N = 5$ のように少ない場合には類似度が高い論文が少なく，類似度が小さい論文を推薦してしまい，適合率が低かった．しかし取得論文数が多くなれば類似度の高い論文も増え，推薦論文の適合率が向上した．ところが N が 30 件，50 件の時よりも 10 件の時の rigid 判定の適合率が良いことから，関連の深い論文は取得論文が 10 件程度で得られることが分かる．30 件，50 件としたとき適合率が下がる理由の一つは，類似度が同じ論文をランク付けする優先順位が適切でない場合があるためと考えられる．一方，提案手法では取得論文数が少なくても関連性の高い論文を推薦することができている．提案手法で取得論文数が多くなると適合率が下がる主な原因は，authority 値が高い語に不適切な語が含まれるようになるためである．例えば閲覧論文が「英文の冠詞誤り検出」という内容の場合,「冠詞誤り」などの語の authority 値が高くなることが期待されるが，実際には「ユーザ」や「入力文」のような語の authority 値が高くなり，それにより適合率が低下した．このように取得論文数が多いと，論文の主題ではないが専門用語として抽出される語が，論文推薦に悪影響を与えることがある．この「冠詞誤り」という語は，式 (1) で計算する専門用語としての重要度は大きかった．しかし提案手法ではこの専門用語の重要度は論文推薦では考慮せず，抽出した専門用語はすべて同等に扱っている．適合率向上のためこの重要度に応じて取得論文集合に重みを付けるなどの方法が考えられる．また専門用語の抽出手法を改善し，より専門性，特定性の高い語を抽出できるようにする必要がある．

次に提案手法とベースラインを比較すると，多くの場合に提案手法の方が適合率が高い．ベースラインは単に専門用語の文書頻度により専門用語のスコアを決定し，それに基づいて論文のスコアを決定する．しかし提案手法では文書頻度に加えて，リンク構造を考慮に入れており，これが論文推薦に一定の効果があったと考えられる．

さらに詳しく提案手法とベースラインを比較するために，推薦した論文の上位にどれだけ関連論文が存在するかについて，実験結果をより上位の推薦論文の適合率，具体的には p@1，p@3，p@5，p@7，p@9 によって評価した．この結果を図 5〜図 8 に示す．提案手法では推薦論文のランクによる適合率の差は $N = 10$ のときを除いてあまり見られなかった．一方でベースラインでは rigid 判定の p@1，p@3 の適合率がその他に比べて高い傾向がある．この傾向が顕著に見られるのが専門用語あたりの取得論文数が 5 件の場合である（図 5）．これは推薦論文の上位に適合論文が集中していることを示している．よって数件の適合論文を推薦する場合にはベースラインの方が有効であるが，より多くの論文を推薦する場合，提案手法の方が適していると言える．

図 4 から三手法に共通して言えるのは，rigid 判定と relaxed 判定の適合率にかなりの差が見られることである．本実験の適合率の rigid 判定では，閲覧論文の目的と同じ内容の論文を適合としている．しかしこの判定に適合する論文自体が十分に取得できていないことがあった．そのため本実験はいずれの手法でも rigid 判定の適合率が低くなったと考えられる．提案手法は 3.1 節で述べたように，各専門用語単独で検索して CiNii から論文集合を取得する．これを複数の専門用語の AND 検索にすることで，より閲覧論文に関連の深い論文集合が取得できれば，rigid 判定での適合率の向上が期待できる．

5.　ま　と　め

本稿では学術論文閲覧支援システムのための関連論文推薦の手法を提案した．提案手法では閲覧論文から取得した各専門用語で検索した論文集合と，それらに出現する専門用語集合との間にリンク関係を定義し，二部グラフを生成した．この二部グラフに HITS アルゴリズムを適用し，得られた hub 値に基づい

図 5　提案手法とベースラインの比較（N=5）
Fig. 5　Comparison of the proposed method with the baseline one (N=5).

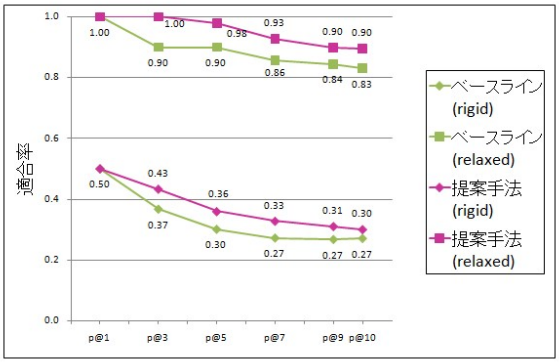

図 6　提案手法とベースラインの比較（N=10）
Fig. 6　Comparison of the proposed method with the baseline one (N=10).

図 7　提案手法とベースラインの比較（N=30）
Fig. 7　Comparison of the proposed method with the baseline one (N=30).

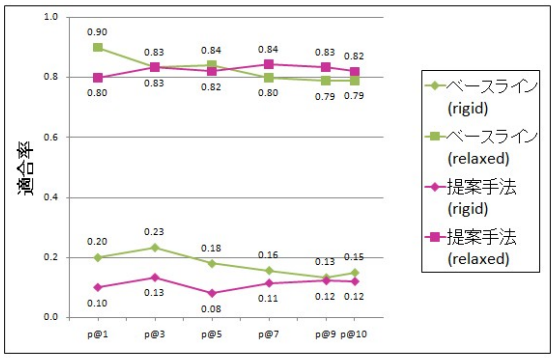

図 8　提案手法とベースラインの比較（N=50）
Fig. 8　Comparison of the proposed method with the baseline one (N=50).

て関連論文をランク付けして論文推薦を行った．また推薦論文の適合性を評価し，VS モデルによる論文推薦等と比較した結果，比較的少ない関連論文集合から提案手法は高い精度で適切な論文を推薦することができた．

本稿の実験では論文の表題と概要のみを推薦に利用しているが，論文の引用・被引用情報を利用することも論文推薦には有効であると考えられる．また閲覧論文中の専門用語の重要度を考慮し，HITS アルゴリズムの反復計算式 (2)，(3) で重みを付与することで，論文推薦精度のさらなる向上が期待できる．今後の展望としては，閲覧支援の対象を論文の表題や概要だけでなく，論文全体に拡張することが挙げられる．

文　献

[1] 鉢木稔浩, 太田学, 高須淳宏. Web 資源を利用した学術論文閲覧支援システム. 情報処理学会研究報告, Vol. 2009-DBS-149, No. 14, pp.1–6, 2009.

[2] J.M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, Vol. 46, No. 5, pp. 604–632, 1999.

[3] 久光徹, 丹羽芳樹, 辻井潤一. タームの representativeness を測る. 情報処理学会研究報告. 自然言語処理研究会報告, Vol. 99, No. 73, pp. 115–122, 1999.

[4] 湯本紘彰, 森辰則, 中川裕志. 出現頻度と連接頻度に基づく専門用語抽出. 情報処理学会研究報告. 自然言語処理研究会報告, Vol. 2001, No. 86, pp. 27–45, 2003.

[5] 矢島健太郎, 井上潮. ソーシャルブックマークにおける文書解析を利用した類似文書および類似ユーザの推薦方法の提案. 電子情報通信学会 第 18 回データ工学ワークショップ論文集, C9-3, 2007.

[6] 近藤光正, 中辻真, 田中明通, 内山匡. 重要語抽出を用いた外部 API からの関連コンテンツ推薦. *The 24th Annual Conference of the Japanese Society for Artificial Intelligence*, 1D2-1, 2010.

[7] 近藤光正, 森田哲之, 田中明通, 内山匡. HITS に基づく Wikipedia ランキングアルゴリズムとユーザ履歴を用いた個人適応型クエリ推薦. 電子情報通信学会 第 19 回データ工学ワークショップ論文集, B2-4, 2008.

[8] Kazunari Sugiyama and Kan Min-Yen. Scholarly paper recommendation via user's recent research interests. *In Proc. of 10th annual joint conference on Digital libraries*, pp. 29–38, 2010.

[9] 難波英嗣, 奥村学. 多言語論文データベースを用いたサーベイ論文検出：サーベイ論文自動作成の実現に向けて. 電子情報通信学会技術研究報告. NLC, 言語理解とコミュニケーション, Vol. 102, No. 199, pp. 35–41, 2002.

[10] CiNii API, http://ci.nii.ac.jp/opensearch/search

[11] Sen Project, http://ultimania.org/sen/

[12] Yahoo!ウェブ検索, http://serch.yahoo.co.jp/

[13] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. *Technical report, Stanford Digital Library Technologies Project*, 1998.