検索質問に含まれる単語と適合文書内の単語の距離に着目した適合フィードバックの改善

辻　裕樹† 藤本　典幸† 萩原　兼一†

† 大阪大学 大学院情報科学研究科
〒 560–8531 大阪府豊中市待兼山町 1–3
E-mail: †{h-tuji,fujimoto,hagihara}@ist.osaka-u.ac.jp

あらまし WWW 検索時に，検索結果中の重要な単語に基づいた元の検索質問の修正と，修正した検索質問を用いた再検索を繰り返すことで検索結果を改善していく検索手法の 1 つに適合フィードバックがある．適合フィードバックにおける検索結果中の重要な単語の選択は，検索結果中からユーザが対話的に指定した適合文書集合を基に何らかのアルゴリズムを用いて選択する．本論文では，ユーザが指定した適合文書中の単語の中でも，その周辺に検索質問中に現れる単語が出現する単語は，ユーザが求める情報と関連が強い可能性が高いと考え，そのような単語に大きい重みを与える単語選択手法を提案する．さらに，既存の検索システムの 1 つである Google での検索に提案手法を適用した場合の適合率上昇の比較実験を行った結果，10 件中 8 件の検索課題において検索質問修正後の適合率が既存の wpq アルゴリズムより向上した．

キーワード WWW 検索，適合フィードバック，単語選択，検索質問，周辺スコア

An Improvement of Relevance Feedback Based on the Distance between a Query Term and a Term in Relevant Documents

Hiroki TSUJI†, Noriyuki FUJIMOTO†, and Kenichi HAGIHARA†

† Graduate School of Information Science and Technology, Osaka University
1–3 Machikaneyama-cho, Toyonaka-shi, 560–8531 Japan
E-mail: †{h-tuji,fujimoto,hagihara}@ist.osaka-u.ac.jp

Abstract Relevance feedback is an effective method for iteratively improving user's initial query in Web information retrieval. In relevance feedback, several documents in retrieved documents are selected as relevant documents by a user. Then, terms contained in the relevant documents are used for query modification. In this paper, we propose an improved relevance feedback in which terms close to query terms are given high weights assuming that such terms are highly relevant to the information required by a user. By applying our method to web information retrieval at Google, we achieved higher precision than wpq algorithm for eight of ten tasks.

Key words Web information retrieval，Relevance feedback，Term selection，Query，Around score

1. はじめに

近年 World Wide Web（WWW）では，HTML 文書をはじめとした多くのデータが爆発的に増加しており，膨大な量のデータ中から必要な情報を見つけ出すシステムとして数多くの WWW 検索システムが利用されている．WWW 検索システムでは検索質問として，求める情報に関連するいくつかの単語をユーザが入力する．しかし，求める情報を結果として得ることができるような検索質問を最初に思いつくことはまれであり，そのため検索質問の修正を何度か行い，検索結果を改善していく過程が必要となることがほとんどである．このような検索質問修正を支援する手法の 1 つに適合フィードバック（Relevance Feedback）[1] がある．

適合フィードバックでは，ユーザが検索結果中の各文書について，適合するかしないかという評価をシステムに与える．すると，システムは適合文書集合中から検索質問修正のために有用と思われる単語を選択し，選択した単語に基づいて検索質問を修正して再検索する．検索質問の修正および再検索は，システムが自動的に行う場合と，ユーザがシステムから提示された選択結果の単語を使って行う場合とに分けられる．こうして

ユーザがシステムと対話を繰り返し行い，検索結果を改善する手法が適合フィードバックである．この手法は，検索結果中の適合文書を増やすことを目的とするため，「上りの新幹線の JR 新大阪駅での時刻表が知りたい」といった，1 つの情報を得ることで検索が終了してしまうような検索には不向きである．しかし，「JR 新大阪駅周辺のおいしいと有名なレストランの情報をできるだけ多くほしい」といった，検索結果としてできるだけ多くの情報を必要とするような検索には有効である．

適合フィードバックにおける単語選択手法としては，候補となっている各単語に単語重要度（Term Selection Value）を計算し，単語重要度の大きな単語を選択結果とする手法がよく用いられる．このタイプの手法として wpq アルゴリズム [8] や emim アルゴリズム [6] などがある．これら既存の単語選択手法では，1 つの HTML 文書中に複数のトピックが含まれる場合を考慮していない．ところが，ニュースサイトなど，1 つの HTML 文書中には複数の異なるトピックの情報が存在することがあり，検索結果中の適合文書においても，実際にユーザが求める情報は複数存在する情報の中の一部分であることがほとんどである．したがって既存の単語選択手法を用いると，適合文書中の情報であっても必要のない情報から単語が選択される可能性が生じる．

そこで本研究では，適合文書中においてユーザが求める情報部分を特定する手がかりとして，検索質問中の単語が出現する部分周辺に着目した単語選択の改善手法を提案する．提案手法では，適合文書中において検索質問中の単語が出現する部分に近い単語ほど大きい重みを与える周辺スコアを提案し，wpq アルゴリズムと組み合わせて使用することで，上記の問題点の解決を目指す．

Google [5] をサブ検索システムとして利用する検索支援システムに提案手法を実装して評価したところ，10 件の検索課題のうち，8 件において検索質問修正後の適合率が既存の wpq アルゴリズムより向上した．

2. wpq アルゴリズム

本節では，既存の単語選択手法の 1 つである wpq アルゴリズムについて説明し，その問題点を述べる．

2.1 単語重要度の決定

wpq アルゴリズムでは，適合文書集合中に存在する各単語 t に対して（1）式で表される単語重要度 $wpq(t)$ を求め，その値が大きい単語を選択する．ここで，$w(t)$ は Robertson/Sparck Jones の重み (Robertson/Sparck Jones relevance weight) [7] と呼ばれる，(2）式で表される値である．$w(t)$ の値は，検索結果の文書集合中においてユーザが指定した適合文書集合と，それ以外の文書集合中のそれぞれおける単語 t の出現割合（表 1 参照）により決定される値である．$p(t)$ は，ユーザが検索結果から選択した適合文書集合中に単語 t が存在する確率，$q(t)$ は，それ以外の文書集合中に単語 t が存在する確率であり，それぞれ以下の（3），(4）式で表される．(2），(3），(4）式中の N は検索結果中の全文書数，$n(t)$ は検索結果中において単語 t を含む文書数，R はユーザが検索結果から選択した適合文書数，そして

表 1　検索結果の文書集合の内訳

	適合文書	非適合文書	計
単語 t が現れる	$r(t)$	$n(t)-r(t)$	$n(t)$
単語 t が現れない	$R-r(t)$	$N-n(t)-R+r(t)$	$N-n(t)$
計	R	$N-R$	N

図 1　「イチロー」という検索質問で検索した結果の文書例
（http://macamp.site.ne.jp/shaka/ichiro/index.html）

$r(t)$ はその適合文書集合中において単語 t を含む文書数を表す．ただし $q(t)$ は $p(t)$ と比べて十分小さな値であるので，無視できる [9].

$$wpq(t) = w(t) \cdot (p(t) - q(t)) \tag{1}$$

$$w(t) = \log \frac{\frac{r(t)+0.5}{R-r(t)+0.5}}{\frac{n(t)-r(t)+0.5}{N-n(t)-R+r(t)+0.5}} \tag{2}$$

$$p(t) = \frac{r(t)}{R} \tag{3}$$

$$q(t) = \frac{n(t)-r(t)}{N-R} \tag{4}$$

2.2 単語重要度決定の際の問題点

WWW 検索システムが主な検索対象としている HTML 文書では，複数の異なるトピックの情報が存在することが一般的であり，ニュースサイトは特にこの傾向が顕著な例である．そのため，検索の際においても，ユーザが求める情報は適合文書中の数ある情報の中の一部の情報だけである場合がほとんどである．図 1 に示している HTML 文書は，Google で「イチロー」という単語を検索質問として検索した結果，検索結果リストの最上位に現れた HTML 文書である．この文書は，日本人メジャーリーガーに関する情報を扱った文書であり，イチロー選手に関する情報も存在する．しかし，文書中に存在する全情報に対して，イチロー選手に関する情報は全体の半分にも満たない．残りの情報は，他の日本人メジャーリーガーに関する情報や広告などの情報となっている．

しかし wpq アルゴリズムでは，ある単語 t が適合文書内においてユーザの求める情報部分に出現した場合でも，逆に必要のない情報部分に出現した場合でも，等しく $r(t)$ の値に 1 加算される．つまり，wpq アルゴリズムにおいては，適合文書内に出現するすべての単語がユーザの求める情報と関連があると見

なされている．そのため，上で述べた，適合文書内に存在するユーザの求める情報とそれ以外の情報とが区別されない．その結果として，適合文書内でもユーザの求める情報と関連しない情報中の単語に高い単語重要度が与えられ，単語選択結果として選ばれる可能性が出てくる．このようにして選ばれた，ユーザの求める情報とは関連の無い情報部分中の単語は，検索質問修正に用いても検索結果上位の適合文書数の増加に関して効果が無いばかりか，逆に適合文書数を減少させる原因となる可能性も考えられる．

3. 提案する単語選択手法

本節では，2.2 節で述べた問題点を解決する単語選択手法を提案する．

3.1 提案手法における単語選択方針

2.2 節で述べた問題解決のため，適合文書内の単語を選択する際に，適合文書内に複数存在するであろう情報の区別を行い，ユーザが求める情報中から優先して単語を選択する必要がある．そこで，提案する単語選択手法では適合文書内で出現する単語でも，特にユーザが求める情報部分内に出現する単語に高い重みを与えるスコア (周辺スコア，$Ard(t)$) を導入する．そして，wpq アルゴリズムに周辺スコアを適用した新しい単語重要度として，(5) 式で表される $TSV(t)$ を導入することで単語選択の精度向上を目指す．

$$TSV(t) = Ard(t) \cdot wpq(t) \tag{5}$$

そのためにはまず，適合文書中からユーザが求める情報部分を特定する必要がある．ここで，ユーザに適合文書の指定だけでなく，文書中の適合部分も指定してもらう手法が考えられる．この手法では，適合文書中のユーザの求める情報部分を正確に特定できるが，ユーザにかかる負担を考えると良い方法とはいえない．そこで，システムが自動で適合部分を特定することが望ましいが，特定のためには何らかの手がかりが必要となる．ここで，ユーザが入力した検索質問は，ユーザが求める情報を複数の単語で表現したものであるために，適合部分の特定の際に大きな手がかりとなる，つまり以下の性質 1 が成り立つと考えられる．そこで，性質 1 を考慮し，検索質問中の単語の出現位置に近い単語ほど重みを高くするように周辺スコアを定義する．

［性質 1］ 適合文書内において，検索質問中の単語の出現位置周辺にはユーザが求める情報が存在する可能性が高い．

3.2 周辺スコア

本節では，3.1 節で述べた性質 1 を考慮した周辺スコアの計算方法について述べる．

HTML 文書は図 2 のように HTML タグ（閉じタグは省略）と，HTML タグによって区切られた文章とに分けることができる．以降それぞれタグノード，テキストノードと呼ぶ．また，適合文書中に存在するテキストノードの中でも，検索キーワードが存在するテキストノードをクエリノードと呼ぶ．図 2 のクエリノードの例は，検索質問中の単語が「イチロー」と「ファン」であった場合の例である．ここで，多くのテキストノードに関して次の性質 2 を満たす可能性が高いと考えられる．

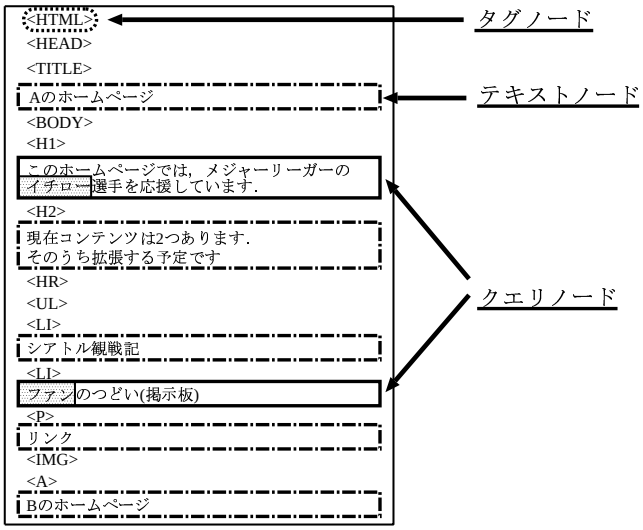

図 2 タグノード，テキストノードおよびクエリノード

［性質 2］ 1 つのテキストノード中には，1 つのトピックに関する情報のみが存在する．

性質 1 および 2 を考慮して，適合文書内の各単語への周辺スコアを，以下の 2 ステップに分けて計算する．

（Step1） 適合文書内の各テキストノードに対して，クエリノードからの距離に応じてスコアを割り当てる．

まず初めに，各テキストノード中の情報が，ユーザの求める情報にどれだけ関連があるかを示すスコアを割り当てる．その際，性質 1 より，検索質問中の単語の近くに存在するテキストノードはユーザが求める情報と関連が高いと考えられるために，検索質問中の単語が含まれるクエリノードとテキストノードとの距離を基にスコアを割り当てる．

（Step2） テキストノードに割り当てられたスコアを基に，適合文書集合中の各単語の周辺スコアを求める．

適合文書集合内の全てのテキストノードに，ユーザが求める情報との関連度に基づくスコアを与えた後，テキストノード内に出現する全ての単語に，そのテキストノードに与えられたスコアを割り当てる．その後，各単語ごとに，割り当てられたスコアの和を，適合文書集合中におけるその単語の出現回数で割った平均のスコアの値を，その単語の周辺スコアとする．

以降，Step1 と Step2 の詳細について，それぞれ 3.2.1 節と 3.2.2 節で述べる．

3.2.1 テキストノードへのスコア割り当て

性質 1 より，クエリノード周辺のテキストノードほど，ユーザが求める情報が存在する可能性が高いと考えられる．また，HTML 文書において，出現する単語と画像や音楽などのマルチメディアデータとの関連の強さは，単語とマルチメディアデータとの距離によって指数的に低下することが，多数の HTML 文書の解析の結果適切であることが分かっている [2]．そこで，クエリノードとテキストノードとの関係について以下の性質 3 を仮定する．

［性質 3］ テキストノードとクエリノードとの関連の強さはク

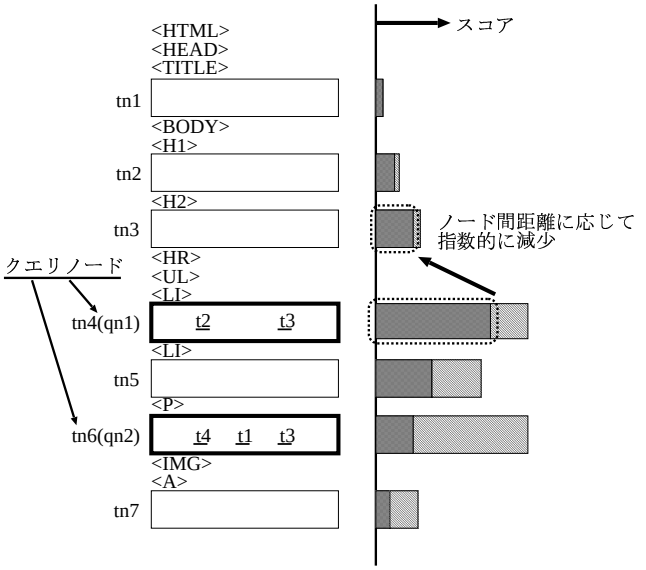

図 3　テキストノードへのスコア付け

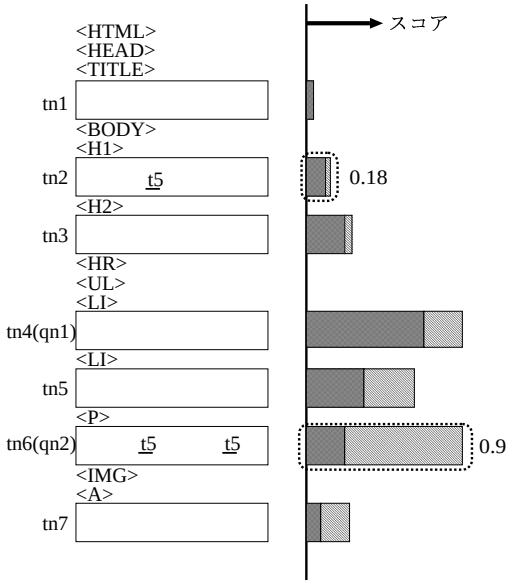

図 4　単語への周辺スコア付け

エリノードからの距離によって指数的に低下する．

性質 1 および 3 より，クエリノード qn がテキストノード tn に与えるスコア $Score(tn, qn)$ を（6）式で定義する．

$$Score(tn, qn) = a(qn) \cdot e^{-2 \cdot d(tn,qn)/d_{max}} \tag{6}$$

$d(tn, qn)$ はテキストノード tn とクエリノード qn のノード間距離である．ここでノード間距離とは，タグノードとテキストノードを含めた場合の，2 つのノード間の距離となる．また，クエリノード qn から距離が一定の値 d_{max} 以上離れたテキストノード tn，つまり $d(tn, qn) > d_{max}$ となるような tn は，クエリノード qn 中の情報と関連がないと考えられるため，そのような tn については，$Score(tn, qn) = 0$ とする．本研究においては実験的な値として $d_{max} = 10$ としている．

同じクエリノードであっても，検索質問として入力した複数の検索質問中の単語のうち，多くの種類のキーワードを含むクエリノードは，よりユーザの求める情報と関連が高いと考えられる．例えば,「イチロー選手のファンサイトを探したい」という検索要求に対して,「イチロー」と「ファンサイト」の 2 つの単語からなる検索質問を入力したとする．その際，適合文書内において,「イチロー」と「ファンサイト」の 2 単語とも出現するクエリノードはユーザの検索要求を満たす情報である可能性が高いが,「ファンサイト」だけが出現するクエリノードは他の選手のファンサイトに関連する情報であることも考えられる．さらに，この場合は「イチロー」や「ファンサイト」という単語出現数の多寡による影響は小さい．つまり，クエリノードがユーザの求める情報に関連するかどうかは，クエリノード内における，検索質問中の単語の単純な出現回数ではなく，クエリノードが含む検索質問中の単語の種類の数に大きく依存すると考えられる．そこで，スコアを上記の性質に対応させるために，クエリノード qn が含んでいる検索質問中の単語の種類の割合を示す値 $a(qn)$ を導入する．

図 3 における，クエリノード qn_1 がテキストノード tn_3 に与えるスコア $Score(tn_3, qn_1)$ を例に説明する．検索質問中の単語が $\{t_1, t_2, t_3, t_4\}$ の 4 種類であり，そのうちクエリノード qn_1 が $\{t_2, t_3\}$ という 2 種類の単語を含んでいた場合，$a(qn_1) = 2/4 = 0.5$ となる．そしてノード間距離は $d(tn_3, qn_1) = 4$ となる．その結果，クエリノード qn_1 がテキストノード tn_3 に与えるスコアは $Score(tn_3, qn_1) = 0.5e^{-2 \cdot 4/10} \simeq 0.2247$ となる．

また，図 3 の場合のように，1 つの適合文書 d 内にはクエリノードが複数存在することが考えられるため，各 qn について求めた $Score(tn, qn)$ の和をテキストノード tn のスコア $Score(tn)$ とする（(7) 式）．

$$Score(tn) = \sum_{qn \in d} Score(tn, qn) \tag{7}$$

3.2.2　単語の周辺スコア決定

適合文書集合 D 中の全てのテキストノード tn にスコアを割り当てたあと，次はそのスコアを基として，D 中に存在する各単語 t に周辺スコアを与えていく．まず，単語 t のテキストノード tn における周辺スコア $Ard(t, tn)$ を（8）式と定義する．(8) 式における $tf(t, tn)$ は，テキストノード tn における単語 t の出現回数である．

$$Ard(t, tn) = tf(t, tn) \cdot Score(tn) \tag{8}$$

そして，単語 t の周辺スコアを求めるために，全てのテキストノードにおける周辺スコア $Ard(t, tn)$ の和を求めるが，単純に和をとると，ユーザの求める情報とは関係のないテキストノードに頻出する単語にも高いスコアが与えられてしまう問題が生じる．この問題を回避するため，全てのテキストノードにおける周辺スコア $Ard(t, tn)$ の和を，D 中における単語 t の出現回数で割った値を，単語 t の周辺スコア $Ard(t)$ とする．

$$Ard(t) = \frac{\sum_{d \in D} \sum_{tn \in d} Ard(t, tn)}{\sum_{d \in D} \sum_{tn \in d} tf(t, tn)} \tag{9}$$

図 4 中の，単語 t_5 における周辺スコアの計算方法を例に説明

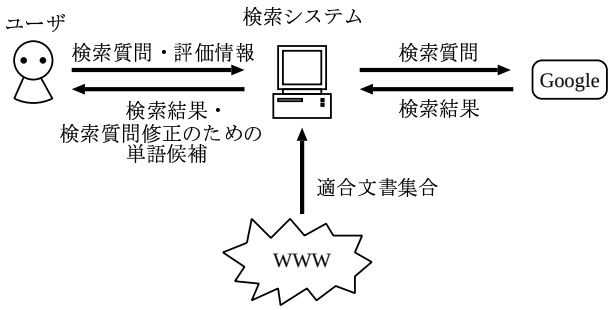

図 5　検索システムの構成

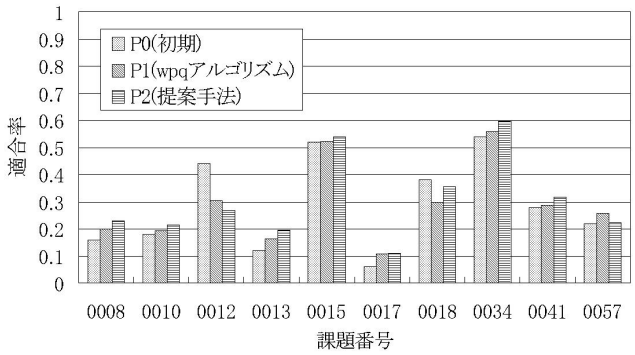

図 6　wpq アルゴリズムおよび提案手法の適合率

する．単語 t_5 はテキストノード tn_2 に 1 回と tn_6に2 回の合わせて 3 回出現している．テキストノード tn_2 と tn_6 に与えられたスコアがそれぞれ $Score(tn_2) = 0.18$，$Score(tn_6) = 0.9$ とすると，それぞれのテキストノードにおける単語 t_5 の周辺スコアは，$Ard(t_5, tn_2) = 1 \cdot 0.18 = 0.18$ 及び $Ard(t_5, tn_6) = 2 \cdot 0.9 = 1.8$ となる．各テキストノードにおける単語 t_5 の周辺スコアから，最終的な単語 t_5 の値は $Ard(t_5) = (0.18 + 1.8)/(1 + 2) = 0.66$ となる

4. 評 価 実 験

4.1 実 験 方 法

評価実験をするにあたって，Google をサブ検索システムとして利用する検索システム（図 5）を構築し，既存の wpq アルゴリズムと提案手法を実装した．システムは既存と提案の各手法でそれぞれ 10 単語を選択し，ユーザに提示する．ユーザは提示された単語を追加することで検索質問を修正する．

一般的に，ある単語選択手法を適合フィードバックに用いた場合，その手法が検索質問修正に有効であるか評価する方法として，テストコレクション [3], [4] と呼ばれる検索システム評価用のデータセットを利用した評価方法がよく用いられる．テストコレクション中の文書データや検索課題を用いて，検索質問修正前と修正後それぞれにおいて，以下に示す検索結果の適合率および再現率の値を求め，比較することによって評価する．

- **適合率**（precision）

適合率は，検索結果中のうち検索質問に適合する文書数の割合を示す値であり，以下の式で表される．この値が高いほど，検索結果中に多く適合文書が含まれていることとなる．

$$\text{適合率} = \frac{\text{検索結果中の適合文書数}}{\text{検索結果中の全文書数}}$$

- **再現率**（recall）

再現率は，検索対象全体の中に存在する適合文書が，検索結果中にどれだけ含まれているかを示す値であり，以下の式で表される．この値が高いほど，検索結果中に，検索対象中に存在する適合文書がもれなく含まれていることとなる．

$$\text{再現率} = \frac{\text{検索結果中の適合文書数}}{\text{検索対象全体中の適合文書数}}$$

しかし，本システムは文書データベースとして Google の持つ文書データベースを利用していることから，テストコレクションが用意している文書集合を本システムの文書データベースとして用いることができない．また，同様の理由から各検索課題において再現率を求める際に，検索対象集合である Google の文書データベース中に存在する適合文書数を求めることも非現実的である．そこで，今回の評価実験においては以下の方針を用いる．

- テストコレクション中の検索課題のみを利用する
- 評価値としては，検索結果上位 50 件の適合率だけを求めて評価する

評価実験で用いる検索課題として，WWW 検索評価用テストコレクションである NTCIR-3WEB [4] 中の検索課題のうち 10 件を用いた．最初にシステムに与える検索質問としては，検索課題中の，もっとも中心的な主題を表わす 1〜3 単語からなる TITLE フィールド中の単語の AND 検索を用いた．検索課題の概要を表 2 に示す.

4.2 適合率の比較

提案手法による単語選択によって，検索質問修正に有効な単語が選択されているか評価するために，検索結果の上位 50 件の適合率が，提案手法に基づく適合フィードバックによる 1 回の検索質問修正でどのように変化するか，以下の手順で求める．

（1） 検索質問をシステムに入力して検索，上位 50 件を得る

（2） 上位 50 件の各文書に対して適合判断を行い，適合率 P_0 を求める．

（3） （2）で求めた上位 50 件の適合文書を基に，wpq アルゴリズムによる選択結果 10 単語からなる単語集合 $\mathbf{T}_1$ を取得

（4） $\mathbf{T}_1$ 中のそれぞれ 1 単語ずつ検索質問に追加して適合率を求め，平均適合率 P_1 を求める

（5） （2）で調べた適合文書を基に，提案する単語選択手法による選択結果 10 単語からなる単語集合 $\mathbf{T}_2$ を取得

（6） （4）と同様にして平均適合率 P_2 を求める

各検索課題において，P_0 と，P_1 および P_2 を比較したグラフを図 6 に示す．実験の結果，10 件中 8 件の検索課題において提案手法が既存の wpq アルゴリズムより高い適合率を示したが，逆に 10 件の検索課題のうち 2 件において，提案手法における検索質問修正後の適合率が修正前の適合率より低下した．

また，wpq アルゴリズムと提案手法による単語選択の結果をより詳しく評価するために，以下の 3 つの適合率を求め，修正前の適合率 P_0 と比較する．(図 7)．

P_3： wpq アルゴリズムと提案手法の両方で選択された単語集

表 2　評価実験で用いた検索課題

課題番号	検索課題内容（DESC フィールド）	検索質問（TITLE フィールド）
0008	サルサを踊れるようになる方法が知りたい	サルサ，学ぶ，方法
0010	観測のために，オーロラの発生する条件を知りたい	オーロラ，条件，観測
0012	各地域でお正月に食べる雑煮に入っている具や味噌などの違いについて調べたい	正月，雑煮，地方
0013	京都の寺や神社について，歴史的背景など，一歩踏み込んだ情報を知りたい	京都，寺，神社
0015	オゾン層の破壊，オゾン層の拡大における人体への影響について知りたい	オゾン層，オゾンホール，人体
0017	野球とベースボールの比較を行った文書を探したい	野球，ベースボール，比較
0018	ロープワークにおける様々な結び方について記述されている文書を探したい	ロープワーク，結び方
0034	キューブリック氏の監督した映画の感想を聞きたい	キューブリック，映画，感想
0041	印象派に属する画家の絵画がどこで見られるかを探したい	印象派，モネ，美術館
0057	亀の寿命について記述された文書が欲しい	亀，寿命

合 $\mathbf{T}_3$ 中の単語をそれぞれ 1 単語ずつ検索質問に追加した場合の上位 50 件の適合率の平均

P_4：　wpq アルゴリズムでのみ選択された単語集合 $\mathbf{T}_4$ 中の単語をそれぞれ 1 単語ずつ検索質問に追加した場合の上位 50 件の適合率の平均

P_5：　提案手法でのみ選択された単語集合 $\mathbf{T}_5$ 中の単語をそれぞれ 1 単語ずつ検索質問に追加した場合の上位 50 件の適合率の平均

wpq アルゴリズムによる単語選択結果を詳細に評価するため，P_3 と P_4 を比較すると，10 件の検索課題のうち 8 件の検索課題において，P_4 の値が P_3 の値より低くなった．これら 8 件の検索課題においては，wpq アルゴリズムで選択された単語集合 $\mathbf{T}_1$ の中でも，wpq アルゴリズムでのみ選択された単語集合 $\mathbf{T}_4$ の単語が精度悪化の原因となっていることがわかる．さらに，P_0 と P_4 を比較しても，10 件の検索課題のうち 6 件の検索課題において P_4 の値が P_0 の値より低くなっている．このことからも，$\mathbf{T}_4$ 中の単語は検索質問修正に多くの場合，有効とはいえない．

そこで，wpq アルゴリズムにおいて検索質問修正に有効でない単語集合 $\mathbf{T}_4$ が選択されているという問題点を，提案手法が改善できているか評価するために，P_4 の値が P_3 の値より低くなっている，つまり単語集合 $\mathbf{T}_4$ が検索質問修正に悪影響を与えている 8 件の検索課題について，P_4 の値と P_5 の値を評価したところ，8 件のうち 7 件の検索課題において P_5 の値が P_4 の値より向上していた．これより提案手法においては wpq アルゴリズムにおける単語選択の際の問題点を解決できていることが確認できる．さらに，提案手法が wpq アルゴリズムの問題解決だけでなく，検索質問修正に有効な単語選択に結びついているか評価するために，さきほどの 7 件の検索課題において P_0 と P_5 を比較する．その結果，7 件のうち 5 件については P_5 の値が P_0 の値より向上していることが確認できた．

4.3　実験結果の考察

本節では，評価実験で用いた 10 件の検索課題のうち，提案手法が有効であった検索課題と，逆に有効でなかった検索課題について詳細な分析を行い，提案手法の有効性に関して考察する．

4.3.1　提案手法が有効であった例

提案手法により単語選択結果が改善された例として，10 件の検索課題の 1 つである，課題 0008 の実験結果の詳細な分析

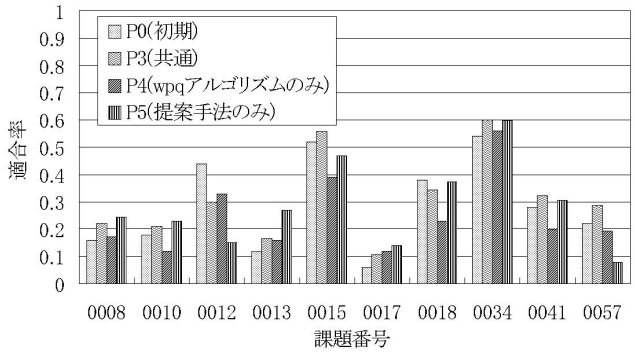

図 7　wpq アルゴリズムまたは提案手法でのみ選択された単語追加時の適合率

を基に，今回の実験結果の原因を考察する．wpq アルゴリズムと提案手法それぞれにおいて選択された 10 単語を表 3 に示す．これらの単語の中でも，wpq アルゴリズムと提案手法のどちらかにしか出現しない単語に着目する．

wpq アルゴリズムだけに出現する単語は，「会員」，「太極拳」，「筋肉」そして「ピアノ」といった，サルサとは関連が弱い単語である．逆に提案手法だけに出現する単語の中には，「基本」，「基礎」，「教室」そして「歌」といった，初心者向けのレッスン情報に関する単語が多い．

これらの単語の違いを，追加後の適合率で比較すると，wpq アルゴリズムにだけ出現する単語の追加による適合率の平均が 0.173 であることに対して，提案手法だけに出現する単語では適合率の平均が 0.245 となる．このことからも，提案手法においては wpq アルゴリズムより適切な単語が選択できていることがわかる．

この課題においては，サルサだけでなくエアロビクス，太極拳そしてフラダンスなど，多種多様な体操やダンスの講習会の案内情報をまとめた HTML 文書が適合文書の大多数を占めた．これらの適合文書において検索要求を満たす情報は，サルサの講習会に関する情報だけであり，同じ文書に存在する他の講習会の情報は必要のない情報である．しかし，wpq アルゴリズムでは同一文書中において，サルサの講習会と他の講習会の情報を等しく扱うため，他の講習会に関する情報中の単語（「太極拳」，「筋肉」など）が選択されてしまった．

提案手法では，検索質問中の単語周辺に着目することによっ

表 3　課題 0008 における単語選択結果

$P_0 = 0.16$			
wpq アルゴリズム		提案手法	
単語	適合率	単語	適合率
体操	0.16	体操	0.16
講師	0.38	講師	0.38
家庭	0.18	参加	0.22
参加	0.22	家庭	0.18
リズム	0.18	リズム	0.18
会員	0.18	基礎	0.26
太極拳	0.23	教室	0.30
筋肉	0.12	歌	0.16
心身	0.20	心身	0.20
ピアノ	0.16	基本	0.26
$P_1 = 0.201$		$P_2 = 0.230$	
$P_3 = 0.220$			
$P_4 = 0.173$		$P_5 = 0.245$	

表 4　課題 0012 における単語選択結果

$P_0 = 0.44$			
wpq アルゴリズム		提案手法	
単語	適合率	単語	適合率
福岡	0.46	福岡	0.46
頭	0.18	歴史	0.18
市	0.34	頭	0.18
歴史	0.18	島根	0.30
つゆ	0.42	県	0.28
いりこ	0.26	津	0.22
島根	0.30	つゆ	0.42
うどん	0.40	市	0.34
津	0.22	青森	0.20
県	0.28	山口	0.10
$P_1 = 0.304$		$P_2 = 0.268$	
$P_3 = 0.298$			
$P_4 = 0.330$		$P_5 = 0.150$	

て，サルサ以外の講習に関する情報を抑えることができており，その結果 wpq アルゴリズムと比べて適切な単語を選択できている．以上の結果より，提案手法の方針であった，1 つの適合文書中に存在する多様な情報の中でも，ユーザの必要とする情報部分に存在する単語の選択ができていることが確認できた．

4.3.2　提案手法が有効でなかった例

10 件の検索課題のうち 2 件においては提案手法の平均適合率 P_2 が wpq アルゴリズムの平均適合率 P_1 よりも低下した．本節では低下した 2 件の検索課題のうちの 1 つである検索課題 0012 の実験結果の詳細な分析を基に，提案手法の問題点について考察する．

表 4 より，検索課題 0012 においては，既存と提案の両手法によって選ばれた単語がともに検索質問修正に有効でなかったことが，修正後の検索質問による適合率により明らかになっている．その原因としては，単語を追加することで検索質問の対象が本来の検索課題の目的とする検索したい対象より狭くなってしまっているたことが挙げられる．例えば，本来は各地方の正月に食べる雑煮の調査を目的として,「正月」,「雑煮」,「地方」という 3 単語を検索質問としたにもかかわらず，適合フィードバックの結果選択された「福岡」という単語を選択することによって「福岡で」正月に食べる雑煮の検索に変化してしまったということである．そのために，検索質問修正によって，検索結果から不適合文書を排除するだけでなく，多くの適合文書も排除してしまった．

さらに，提案手法が wpq アルゴリズムよりも適合率を悪化させた要因としては以下のことが考えられる．

- この検索課題における適合文書は各地方の雑煮の紹介をする HTML 文書であり，適合文書中のほぼすべての情報が検索要求を満たす情報であった．このために，提案手法が用いる周辺スコアのメリットが活かせなかった
- 検索質問の周辺単語を重視する周辺スコアの性質のため，適合文書中のユーザの求める情報内でも，検索質問中の単語からの距離が遠い単語へのスコアが低下する．そのために，検索質問中の単語から遠く離れた場所に出現する重要な単語を見落としてしまう問題が生じる．特に 1 つの HTML 文書に 1 つの情報だけを取り扱う場合は，適合情報が広範囲にわたる傾向があるためにこの問題が起こりやすい．

この要因は，検索課題 0012 と同じく，1 つの適合文書中のほとんどの情報が検索要求を満たす情報であった検索課題 0017 において，wpq アルゴリズムと提案手法との適合率にほとんど差が見られないことからも確認できる．よって，提案手法においては，1 つの適合文書中のほとんどの情報が検索要求を満たす情報である時は，既存手法と同程度かそれ以下の性能しか出ないという問題点が存在することがわかった．

4.3.3　併用手法に関する検討

4.3.1 節における実験結果の分析より，ニュースサイトや各種講習会の案内情報のように，1 つの HTML 文書中に異なる複数の情報が存在する文書が適合文書内に数多く存在する場合には，提案手法は既存の wpq アルゴリズムと比べて，検索結果修正に有効な単語を選択できていた．しかし，逆に 4.3.2 節における実験結果の分析より，適合文書のほとんどが 1 つの情報だけを取り扱う場合は，wpq アルゴリズムと提案手法それぞれの単語選択結果を用いて検索質問修正した場合の適合率の差ががほとんど見られない，もしくは提案手法が wpq アルゴリズムよりも検索質問修正に有効ではない単語を選択してしまう場合も存在した．これらの考察より，wpq アルゴリズムと提案する単語選択手法では適合文書中の適合情報の割合の大小に応じて優劣が逆転することがわかった．

以上の結果より，検索ごとに，適合文書内におけるユーザの求める適合情報の割合に応じて，提案手法と wpq アルゴリズムを使い分けることで，それぞれの単語選択手法を単独で使用するよりも，適合フィードバックによる検索質問修正の効果を上げることが可能であると考えられる．本研究で開発した検索システムにおいては，適合フィードバックを行う際に wpq アルゴリズムと提案手法の選択を行うことが可能であるため，ユーザは検索に応じて単語選択手法を使い分けて検索することが可能である．しかし，適合フィードバックを行うごとに提案手法と

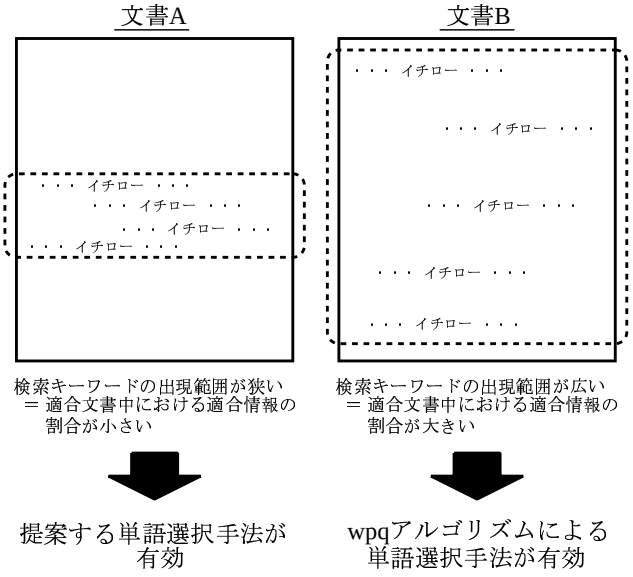

図 8　検索質問中の単語の出現範囲による単語選択手法の使い分け

wpq アルゴリズムのどちらを用いることが有効か判断することはユーザの負担につながることを考えると，適合フィードバックを行う際には，システムが自動で単語選択手法を決定することが望ましい．

そこで，単語選択手法を決定する際に，適合文書内におけるユーザの求める情報の割合を自動で判定する必要がある．この判定は，文書内における検索質問中の単語の出現範囲の分散の程度を求めることで解決できると考えられる．検索質問中の単語が出現する場所が適合文書内のある一定の範囲にのみ限られている場合は，適合文書内におけるユーザの求める情報の割合が少なく，他の情報も多く含まれると考えられるために，提案手法の利用が有効であると考えられる．逆に，検索質問中の単語が適合文書内の広い範囲で出現する場合には，適合文書内のほとんどの情報がユーザの求める情報であると考えられるために，wpq アルゴリズムの利用が有効であると考えられる．

図 8 の例で説明すると,「イチロー」という検索質問中の単語 1 つから成る検索質問で検索した場合，文書 A のように検索質問中の単語がある限られた範囲内に出現する場合，イチロー選手に関する情報は文書 A の中に存在するいくつかの情報の 1 つであると考えられる．逆に文書 B のように検索質問中の単語が全体的に分散して出現する場合，文書 B 中の情報のほぼすべてがイチロー選手に関する情報であると考えられる．

5. ま と め

本研究では，適合フィードバックにおいて既存の単語選択手法では考慮されていなかった，HTML 文書における情報の多様性を考慮した単語選択手法を述べた．提案手法では適合文書内に出現する検索質問中の単語周辺の単語に高いスコアを与える周辺スコアを提案した．評価実験をしたところ，10 件の検索課題のうち，8 件において提案手法は検索質問修正後の適合率が既存の wpq アルゴリズムより向上した．

今後の課題としては，HTML 文書の木構造を意識した周辺スコアの拡張が挙げられる．また，本研究においては提案手法の評価を既存の検索システムの文書データをそのまま用いて行ったが，より厳密な評価のためにテストコレクションの，検索課題だけの利用ではなく文書データの利用とあわせて評価を行いたい．

謝辞　本研究の一部は，文部科学省特定領域研究（課題番号 15020236）の補助による．

文　献

[1] R.Baeza-Yates and B.Ribeiro-Neto, “Modern Information Retrieval”, Addison Wesley Longman, 1999.
[2] M.La Cascia, S.Sethi, and S.Sclaroff, “Combining Textual and Visual Cues for Content-Based Image Retrieval on the World Wide Web”, Proc. of the IEEE Workshop on Content-Based Access of Image and Video Libraries, pp.24–28, 1998.
[3] N.Craswell and D.Hawking, “Overview of the TREC-2002 Web Track”, Proc. of the 11th Text REtrieval Conference (TREC-11), 2002.
[4] K.Eguchi, K.Oyama, E.Ishida, N.Kando, and K.Kuriyama, “Overview of the Web Retrieval Task at the Third NTCIR Workshop”, Proc. of the 3rd NTCIR Workshop on Research in Information Retrieval, Automatic Text Summarization and Question Answering, 2003.
[5] http://www.google.com/
[6] C.J.Van Rijsbergen, “A theoretical basis for the use of co-occurrence data in information retrieval”, Journal of Documentation, 33.2. pp.106–119, 1977.
[7] S.E.Robertson and K.Sparck-Jones, “Relevance weightning of search terms”, Jounal of the American Society of Information Science, 27.3. pp.129-146, 1976.
[8] S.E.Robertson, “On term selection for query expansion”, Journal of Documentation, 46.4. pp.359–364, 1990.
[9] S.E.Robertson, S.Walker, S.Jones, M.Hancock-Beaulieu, and M.Gatford, “Okapi at TREC-3”, Proc. of the Third Text REtrieval Conference (TREC-3), pp.109–126, 1994.