

主題グラフ及び関連度情報からの単語重要度付与を用いた情報検索システムの提案

富田 準二, 竹野 浩

NTT 情報通信研究所

{tomita, takeno}@isl.ntt.co.jp

WWW 上での検索システムのように、検索対象が大規模で雑多なものとなり、また、検索キーとして単語列だけではなく文や文書を指定できるようになると、適合度の判定を正確に行なうことは難しくなる。本研究では、グラフモデルという新しい検索モデルを提案する。このモデルでは、検索キーおよびそれぞれの検索対象文書から、単語の重要度をノードの重みとし、単語間の関連度をリンクの重みとした主題グラフを作成し、これら主題グラフ間の類似度に基づき適合度を判定する。また、単なる単語の出現頻度情報ではなく単語間の関連度情報から単語の重要度を決定する手法を提案する。本手法を用いた検索システムを構築し、ベンチマークを用いて検索精度の評価を行なう。

Proposal for a new IR system using subject graph and word's weighting by the relation

Junji Tomita, Hiroshi Takeno

NTT Information and Communication Systems Laboratories

{tomita, takeno}@isl.ntt.co.jp

Precise document ranking is difficult for a search system, especially when documents are gathered from huge and various databases such as WWW, or when it uses sentences as a query phrase. We propose a new search model, Graph-Model. In this model, query phrase and each document is translated into a subject graph regarding every word's weight as its node's weight and relation between words as its link's weight, and the document ranking is evaluated by similarities between these graphs. We also propose a new method to calculate the word's weight using the relation between words not the frequency of a word. And we describe an experimental result using a benchmark data.

1 はじめに

コンピュータ・ネットワークの進歩に伴い、文書検索における検索対象は、従来の図書館の書誌情報といった比較的短くまとまった内容で書かれ、形式の固定されたものから、複数の主題を持ち形式も様々な雑多な文書全文となってきている。また、検索キーとして、従来の単語列や単語のブール演算子結合から、自然言語で書かれた文や、文書そのもの(類似文書検索、関連性フィードバック等)を指定できるようになってきている[6]。

文書検索において検索結果を決定するためには、それぞれの検索対象文書の主題(内容)が検索キーとどの程度似ているのかを表す適合度を判定する必要がある。通常、このような適合度の判定には、検索キーと検索対象文書の主題をある表現に写像し、この表現上で類似度を測定する。多くの検索システムで採用されているベクトル空間モデルでは、検索キーと検索対象文書の主題を、それぞれに使用される単語の重要度の列(主題ベクトル)で表現する。次に、これら主題ベクトル間の類似度を、コサインや内積を用いて測定し、適合度の判定を行なっている。多くの場合、単語の重要度は、文書内の単語の出現頻度(tf法)と文書セット内のその単語の出現する文書数の逆数(idf法)の積によって計算している(tf*idf法)[1]。

ベクトル空間モデルにはそれぞれの主題を単語の重要度だけで表現しているという欠点がある。このため、検索対象の文書が複数の主題を持つ場合、検索精度が低くなってしまう[2]。また、検索キーを文や文書で指定したとしても、これらが持っている単語間の係り受け関係、段落などの特徴を全く利用できていない。さらに、文や文書から単語を抽出する形態素解析という処理の失敗が検索精度の低下を引き起こす。

単語の重要度を求めるのに、文書内の出現頻度(tf法)を利用することの問題点がいくつか指摘されている[3][5]。特に重要なのは、文書の様々な箇所に出現する単語と、ある段落だけに集中的に出現する単語に対して、これらが同じ頻度であれば、等しい重要度を与えてしまうことである。前者はその単語を様々な角度から論じているので、主題を表すのに適当であると考えられるが、後者は主題ではなく副題に関係があるので、主題を表すには適当でない。したがって、後者には前者に

比べて低い重要度を与えるべきである。後者の極端な例として、WWW上の検索において、検索に良く使用される単語を故意に文書の主題とは関係なく文書の先頭に複数回置くことによって、その文書を検索にヒットされ易くするという行為が行なわれている。

このように、ベクトル空間モデルおよび単語の重要度付与(tf法)には、以下の問題がある。

- 検索対象が複数の主題を持つ場合に精度の良い検索ができない。
- 検索キーの特徴を生かした適合度の判定ができない。
- 形態素解析の不完全性への対処ができない。
- 文書内での単語の分布を考慮した単語の重要度付与ができない。

本研究では、これらの問題の解決を目指し、グラフモデルと、単語間の関連度から単語の重要度を決定する手法(td法)を提案する。以下、グラフモデルとtd法の基本的な概念について述べ、実際に本手法を検索システムへ導入した場合の検索精度の評価を示す。

2 グラフモデル

単語同士は、互いに関連し合うことによって、意味を形成している。これは、直観的に明らかである。また、人工知能の分野では概念を表現するのに、「意味ネットワーク」と呼ばれる単語をノードとし、それらのノード間に単語の関連(isa, 対象, 場所など)を表すリンクを付加したグラフを使用している[9]。また、自然言語処理の分野では、辞書を構成する方法や、文書生成の元となるデータに、グラフ(ネットワーク)を使用している[7][8]。

このように、単語間の関連情報を用いれば、文書の主題を、単に単語の重要度で表現する場合と比べて、より正確に表現できると考えられる。また、このように主題を表現した際の適合度を以下のように考える。

「検索キーと適合度の高い文書とは、単に検索キーと同じ単語を使用しているだけでなく、それらの単語同士が検索キー内と同じように関連し合っている文書である。」

そこで、ここでは単語の重要度と単語間の関連情報を同時に表現できるモデルとして、グラフモデルを提案する。グラフモデルにおいては、単語の重要度をノードの重みとし、単語間の関連度をリンクの重みとしたグラフ(主題グラフ)によって、検索キーと検索対象文書の主題を表現する。そして、これら主題グラフ間の一致度に基づき適合度を判定する。

意味ネットワークにおいては、それぞれのリンクにはノードの関係を表すタグが定義されている。これらのタグ情報を利用できれば、より正確な適合度の判定ができると考えられるが、現状の自然言語処理の技術では、これらの関係を自動的に正確に判断することは困難である。そこで、ここでは、単語間の関連を単に関連の強さを表す関連度というスカラー値で表すこととする。

以下、単語間の関連度の計算方法、主題グラフの作成方法、主題グラフ間の類似度測定法について述べる。

2.1 関連度の計算

文書内での単語間の関連度の計算方法として、以下のものが考えられる。

- (a) 任意の2単語が出現する位置が近ければ近い程、これら2単語間の関連度を大きな値とする。
- (b) 同一の文単位内で共起する単語同士には、その共起の度合いに応じた関連度を与える。ここで、文単位とは、それぞれの文、箇条書の項目、タイトルや節の表題などである。
- (c) 単語は文内で係り受けの関係にある。そこで、それぞれの文を係り受け解析することによって関連度を決定する。例えば、直接の係り受けの関係にある単語同士には、大きい関連度を与え、間接的な係り受けの関係にある単語同士には、低い関連度を与える。

これらそれぞれの方法には、以下の特徴がある。

- (a) について 出現する位置が近かったとしてもそれらの単語に関連がない場合は多々ある。例えば、「。」を挟んだ単語同士はいくら出現位置が近くても関連度は小さい。

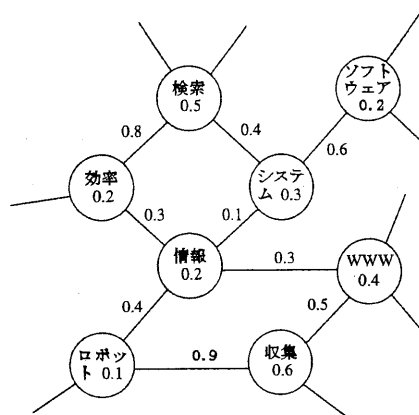


図 1: 主題グラフ

- (b) について 通常、ある単語が他の文単位にある単語と直接係り受けの関係にあるということはない。そのため、文単位内での共起関係を用いれば、係り受け関係にある単語間に関連度を与えることができる。ただし、係り受け関係にない単語同士でも同一文単位内で共起している場合は関連度を与えてしまうという問題がある。
- (c) について 係り受け解析をすれば、(a)、(b)の方法に比べて、正確な関連度を与えることができる。しかし、現状の係り受け解析の技術では、雑多な情報を対象とした場合の精度はあまり高くなく、また、大きな処理コストがかかる。

以上のようにこれらの方法は、関連度の正確さ、処理コスト、検索対象文書の性質によって、使い分ける必要がある。

2.2 主題グラフの作成

2.1節で求めた単語間の関連度をリンクの重みとし、tf*idf法や3節で述べるtd*idf法によって計算される単語の重要度をノードの重みとした主題グラフによって、検索キーおよびそれぞれの検索対象文書の主題を表現する(図1)。

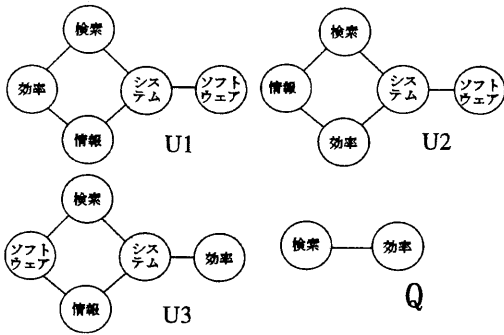


図 2: 適合度判定の例

2.3 主題グラフ間の類似度の測定

主題グラフ間の類似度の測定方法の基本的な考え方は、「同一のノードを多く含み、それらが同じような重要度を持ち、同一のリンクを多く含み、それらが同じような関連度を持つ主題グラフ間には、より高い類似度を付与する。」ということである。ここで、同一のノードとは同じ単語を持つノード、同一のリンクとはリンクの両端のノードに含まれる単語が等しいリンクを表す。

検索キーから作成した主題グラフ Q と検索対象文書から作成した主題グラフ (U_1, U_2, U_3) との類似度の測定の例 (図 2) を用いて、ベクトル空間モデルでは判別不能な適合度の判定をグラフモデルで実現できることを示す。ただし、図 2 において全ての単語の重要度および全ての単語間の関連度は 1 とする。以下、 $f(d, q)$ は、文書 d と質問 q の適合度を表すこととする。

2.3.1 ベクトル空間モデルを用いた適合度の判定

ベクトル空間モデルの場合、適合度は単語の重要度だけによって計算される。そのため単語がどのように関連していたとしても、すべての単語の重要度は 1 なので、

$$f(U_1, Q) = f(U_2, Q) = f(U_3, Q) \quad (1)$$

となる。

2.3.2 グラフモデルを用いた適合度の判定

文書 U_1, U_2, U_3 内の単語の直接の関連を調べると、

- U_1 の「検索」と「効率」の間には、直接の関連 (リンク) がある。
- U_2 および U_3 の「検索」と「効率」の間には、直接の関連 (リンク) がない。

したがって、

$$f(U_1, Q) > f(U_2, Q) = f(U_3, Q) \quad (2)$$

というように「検索」と「効率」の間に関連がある文書 U_1 に高い適合度を与えることができる。しかし、このように単純に直接の関連を用いただけでは、 U_2 と U_3 の間の適合度の区別を付けることができない。

2.3.3 グラフ分割を用いた適合度の判定

U_2 の主題グラフと U_3 の主題グラフを見ると、直観的に考えて、 U_2 の方が U_3 に比べて適合度が高いと考えられる。これは、 U_2 内では「検索」と「効率」は、「情報」と「システム」という 2 つの単語を介して比較的強く関連しているのに対し、 U_3 内では「検索」と「効率」は「システム」という 1 つの単語だけを介してしか関連していないので、これらの単語間の関連度は小さいからである。

そこで、例えば以下の条件のどちらかが成立するように、それぞれの主題グラフを部分グラフに分割する。

- 部分グラフ内の任意の 2 単語 (ノード) 間には直接のリンクが存在する。
- 部分グラフ内の任意の 2 単語 (ノード) 間には長さ 2 以下のパスが 2 本以上存在する。

この方法で作成されたそれぞれの部分グラフは関連を持った単語の集合で構成されているので、それぞれの部分グラフは文書の副題に対応していると考えられる。そして、それぞれの部分グラフ内の任意の 2 単語間には、ある程度の直接の関連があると考えられるので、これらのノード間に弱い重みのリンクを生成する。

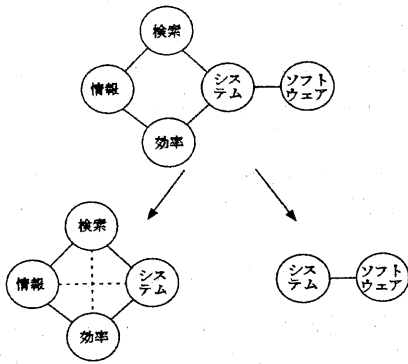


図 3: グラフ分割の例

U_2 を部分グラフに分割した例を図 3 に示す。 U_2 は、グラフ分割されることによって、「情報システムの検索効率」という副題と「ソフトウェアシステム」という 2 つの副題に分けられている。これに対し、 U_3 をグラフ分割すると「情報検索ソフトウェアシステム」と「システム効率」という 2 つの副題に分けられる。つまり、

- U_1 の「検索」と「効率」の間には、直接の関連 (リンク) がある。
- U_2 の「検索」と「効率」の間には、直接の弱い関連 (破線のリンク) がある。
- U_3 の「検索」と「効率」の間には、直接の関連がない。

そのため適合度は

$$f(U_1, Q) > f(U_2, Q) > f(U_3, Q) \quad (3)$$

となる。このようにグラフ分割を利用することによって単語間の間接的な関連を用いた、適合度の判定を行なうことができる。そのため、グラフ分割を利用しなかった場合と比べてより精度の高い適合度の判定が実現できるが、グラフ分割には非常に大きな処理コストがかかるので、グラフ分割の規則、アルゴリズム、適用範囲を検討する必要がある。

3 td*idf 法を用いた重要度の計算

文書内での単語の分布を考慮した、単語の重要度の計算を行なうために、従来の tf 法の代わ

りに、td (term dependency) 法を提案する。td 法を用いた場合、単語 x の重要度は、td 法で計算される重要度 $td(x)$ と、idf 法で計算される重要度 $idf(x)$ の積によって計算する (td*idf 法)。

単語 a の $td(a)$ を、単語 a への関連度の総和

$$\sum_x P(x \rightarrow a) \quad (4)$$

によって計算する。ここで、 $P(b \rightarrow a)$ は、単語 b の単語 a への関連度を表し、 x は文書に含まれるすべての単語を表す。(4) 式から明らかなように、td 法では、多くの単語とより強く関連している単語により大きな重要度を付与することができる。したがって、特定の文脈にしか出現しない単語の td は、様々な文脈に出現する単語の td に比べて小さな値となる。つまり、単語の頻度が同じ場合でも、様々な文脈に出現する、主題と関係がより強いと思われる単語に、より高い重要度を付与することができる。

このように、td 法は、単語の出現分布を考慮した重要度の付与を行なうことができるが、単語の重要度を計算するために単語間の関連度を求める必要があり、tf 法と比べて大きな処理コストがかかる。重要度の正確さ、処理コスト、検索対象文書の性質に応じて、これらの方法を使い分ける必要がある。

4 検索システムへの導入

グラフモデルの考え方をを用いて検索システムを構築した。検索システムの処理の流れを図 4 に示す。ただし、本検索システムでは、単語間の関連度を求めるのに (b) の文単位内共起を用いる方法を採用した。これは、指定範囲内の共起に比べて値が正確であり、また、本検索システムを今後、WWW 上の検索エンジンとして使用することを想定しているので、係り受け解析を利用することは困難であると考えたからである。また、グラフ分割を用いた適合度の判定法は処理コストの制約上採用しなかった。

検索対象文書解析部は、それぞれの文書内の単語間の関連度および単語の重要度を計算し、それぞれの文書の主題グラフを作成する。検索キー解析部は、文書 (または文) で与えられた検索キーから検索キーの主題グラフを作成する。適合度判定部は、検索キーと検索対象文書の主題グラフ

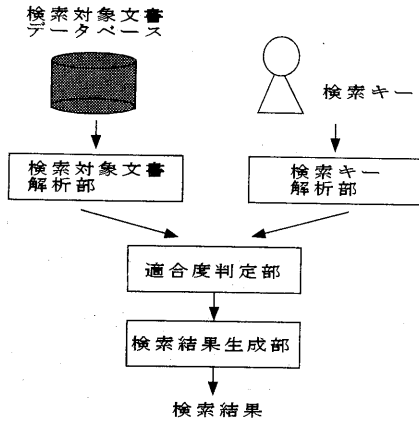


図 4: 検索システムの構成

間の類似度を測定することによって、それぞれの文書の適合度を判定する。検索結果生成部は適合度の降順に検索結果をソートして出力する。

4.1 検索対象文書解析部および検索キー解析部

検索対象文書は HTML で記述されたものを想定している。検索対象文書解析部は、まず、これらの文書を、タグおよび句点を基にして、文単位に分割する。次に、単語 x と単語 y の文書 d 内での出現頻度 $P_d(x)$, $P_d(y)$ と、これらの単語の共起頻度 $P_d(xy)$ を求める。そして、以下の式によって単語 x の単語 y への関連度 $P_d(x \rightarrow y)$ を計算する。

$$P_d(x \rightarrow y) = \frac{P(xy)_d^s}{P(x)_d^t P(y)_d^u} \quad (5)$$

ここで、 s, t, u は、正の定数である。次に関連度の総和によって単語の重要度を計算する (td法)。

$$td_d(x) = \sum_y P_d(y \rightarrow x) \quad (6)$$

そして $td_d(x)$ に idf 法で計算される単語の重要度 $idf(x)$ を掛けた値を単語 x の文書 d 内での重要度 $weight_d(x)$ とする。

このようにして求めた、単語の重み $weight_d(x)$ をノードの重みとし、単語間の関連度 $P_d(x \rightarrow y)$

をリンクの重みとした検索対象文書の主題グラフを作成する。

検索キー解析部は、検索キーが文書の場合、検索対象文書解析部と同様の方法で、主題グラフを作成する。検索キーが文の場合、検索キーに含まれるすべての単語の重みを 1 とし、すべての 2 単語間の関連度を 1 とした主題グラフを作成する。

4.2 適合度判定部および検索結果生成部

適合度判定部は、主題グラフ間の類似度を、以下の方法で計算することによって適合度を判定する。

まず、グラフ q とグラフ d のそれぞれのノードの重みを要素とするベクトルを作成し、これらの内積 f_v を計算する。

$$f_v = 1 + \sum_i v_{qi} * v_{di} \quad (7)$$

ここで、 v_{qi} と v_{di} はそれぞれ、検索キー q 内での単語 i の重要度、検索対象文書 d 内での単語 i の重要度に対応する。ただし、単語 i が存在しない場合は、重要度は 0 とする。また、1 を足しているのは、($f_v \geq 1$) とするためである。

次に、グラフ q とグラフ d のそれぞれのリンクの重みを要素とするベクトルを作成し、これらのベクトルのコサインを計算する。

$$f_r = 1 + \frac{\sum_{i,j} r_{qij} * r_{dij}}{\sqrt{\sum_{i,j} r_{qij}^2} \sqrt{\sum_{i,j} r_{dij}^2}} \quad (8)$$

ここで、 r_{qij} と r_{dij} はそれぞれ、検索キー q 内での単語 i と単語 j の関連度、検索対象文書 d 内での単語 i と単語 j の関連度である。ただし、単語 i と単語 j の間に関連 (リンク) が存在しない場合は、関連度は 0 とする。また、1 を足しているのは、($f_r \geq 1$) とするためである。

以上のように求めた f_v と f_r の積によって、グラフ間の類似度を以下のように求める。

$$\text{類似度} = f_v^p * f_r^{(1-p)} \quad (9)$$

ここで、 $p(0 \leq p \leq 1)$ は定数である。ただし、 $p = 1$ の時、関連度は考慮されなくなり、単語の重要度のベクトルの内積だけによって類似度が計算されるので、ベクトル空間モデルと同様の方法となる。

検索結果生成部は、このようにして求めた適合度の降順に検索対象文書をソートし出力する。

5 評価

グラフモデルを用いた検索システムを構築し、ベンチマーク (BMIR-J2)¹²を用いて検索精度の評価を行なった。

BMIR-J2は、新聞記事5080件、検索要求文60件(標準セット50件+追加セット10件)および、それぞれの検索要求文に対する正解の集合からなる[4]。正解の集合には、A: 検索要求を主題とする記事、B: 検索要求が主題にはなっていないが、検索要求の内容を少しでも記述している記事、C: 全く関連のない記事の3つのランクがある。

検索要求文の標準セットの50件を検索キーとし、ランクA、Bを正解とした。そして、それぞれの検索キーから単語を自動抽出し、それぞれの単語の重要度を1、任意の2単語間の関連度を1とした検索キーの主題グラフを作成して、以下の実験を行なった。

5.1 実験

グラフモデルおよび関連度から重要度を計算するtd法の有効性を示すために、以下の4通りの方法で実験を行なった。ただし、(5)(9)式で用いる定数は、予備実験の結果から $s = 2$, $t = 1$, $u = 1$, $p = 0.5$ とした。

tf-ベクトル 単語の重要度を*tf*法で計算し、関連度を用いないで適合度を判定

td-ベクトル 単語の重要度を*td*法で計算し、関連度を用いないで適合度を判定

tf-グラフ 単語の重要度を*tf*法で計算し、関連度を用いて適合度を判定

td-グラフ 単語の重要度を*td*法で計算し、関連度を用いて適合度を判定

適合度の降順で出力された検索結果の上位 n 件について以下の再現率および適合率を計算した。

$$\text{再現率} = \frac{1}{|Q|} \sum_Q \frac{|S|}{|A|} \quad (10)$$

¹BMIR-J2は(社)情報処理学会・データベースシステム研究会が、新情報処理開発機構との共同作業により、毎日新聞CD-ROM'94データ版を基に構築した情報検索システム評価用テストコレクションである。

²CD-毎日新聞94版を使用。

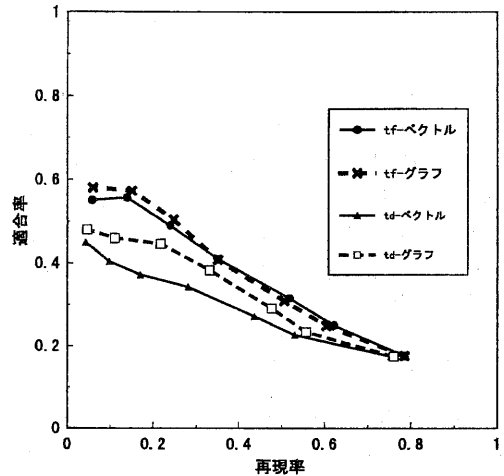


図5: 再現率 - 適合率

$$\text{適合率} = \frac{1}{|Q|} \sum_Q \frac{|S|}{|R|} \quad (11)$$

ただし、 Q は質問集合、 A は正解文書集合、 R は検索された文書集合、 S は検索された正解文書集合、 $|x|$ は、集合 x の要素数を表す。

5.2 結果

$n = 2, 5, 10, 20, 50, 100, 500$ として、実験を行ない、再現率-適合率のグラフを作成した図5。それぞれのグラフのプロットは、左上から順番に検索結果の上位2, 5, 10, 20, 50, 100, 500件の時の再現率と適合率を表す。

5.3 検討および考察

5.3.1 グラフモデルについて

図5を見ると、グラフモデルを用いた場合(*tf*-グラフ、*td*-グラフ)は、そうでない場合(*tf*-ベクトル、*td*-ベクトル)と比べて若干の精度の向上が見られる。*tf*-グラフと*tf*-ベクトルを比べた場合、上位2~10件を対象とした部分(グラフ左上)の精度は向上しているが、それ以外の上位50~500件の部分(グラフ右下)には、精度の向上が見られなかった。このようにグラフモデルは、再現率の向上には寄与しないが、上位数件の適合率の精度向上に寄与していることがわかる。

しかし、実際の検索精度の向上はわずかなものであった。この理由として、検索キーとして与えた検索要求が文書ではなく、比較的短い文(平均3.1単語)であったために検索キーから正確な単語の関連情報が得られなかったこと、検索対象文書が比較的まとまった内容で書かれた新聞記事であったことなどがあげられる。本手法の有効性を実証するためには、文書を検索キーとし、雑多な情報を検索対象とした場合の検索精度の評価を行なうことが必要である。

5.3.2 td法について

td法を用いた場合(*td*-グラフ, *td*-ベクトル)は、そうでない場合(*tf*-グラフ, *tf*-ベクトル)と比べて、検索精度が低下している。この理由として、共起関係から単語の関連度を計算してしまったことが考えられる。前述したように、共起関係から単語の関連度を計算すると、まったく関連のない単語にも同一文単位内で共起すれば高い関連度を与えてしまう。その結果、td法では、これらのノイズの多い関連度の総和から重要度を計算することとなり、正確に重要度が付与できなかつたと考えられる。したがって、td法の有効性を検証するためには、より正確な関連度の計算方法を検討することが重要である。

6 結論および今後の課題

本研究では、適合度の判定に単語間の関連情報を用いることによって、検索精度を向上させる手法(グラフモデル)と、単語間の関連度を用いることによって単語の分布を考慮した重要度が付与できる手法(td法)を提案した。実際に検索システムを構築し、BMIR-J2を用いて本手法の評価を行ない、グラフモデルが、ベクトル空間モデルに比べて、適合率の面でより正確に適合度の判定が行なえることを示した。また、td法に関しては、共起関係から関連度を求めたことが原因で、適合度の判定の精度の向上には繋がらなかった。

今後の課題として、まず、グラフ分割を用いた適合度の判定法の実現がある。しかしながら、td法の場合と同様に関連度を正確に計算することができなければ、グラフ分割を行なっても精度の向上は望めないと考えられるので、より正確に関連度を計算する手法を確立することが先決であ

る。また、グラフモデルのような適合度を向上させる手法は、WWW上の検索エンジンのように検索対象が雑多で、ヒット件数が何百件にもなるような場合により効果的であるので、このような分野に適用し、その有効性を実証したい。さらに、文書検索だけでなく、文書の分類のような文書間の類似度判定に本技術を利用していきたい。

7 謝辞

本研究を行なう上で様々な面でアドバイスを頂いたNTTヒューマンインタフェース研究所の大久保雅且様、NTTソフトウェア研究所の高橋克巳様に感謝致します。

参考文献

- [1] William B. Frakes and Ricardo Baeza-Yates, editors. *Information Retrieval-Data Structures & Algorithms*. Prentice Hall, 1992.
- [2] Marti A. Hearst and Christian Plaunt. Subtopic Structuring for Full-Length Document Access. In *Proceedings of ACM-SIGIR'93*, pages 59-68, 1993.
- [3] 羽田 久一 and 山口 英. キーワードの位置情報を利用した情報検索技術の提案. 情処情報学基礎研究会, 44(3):17-23, 1996.
- [4] 木谷 強ら. 日本語情報検索システム評価用テストコレクションBMIR-J2. 情処データベースシステム研究会, 114(3):15-22, 1998.
- [5] 大澤 幸生, Nels Eric Benson, and 谷内田 正彦. 共起グラフを用いたキーワード抽出. 情処情報学基礎研究会, 43(8):57-64, 1996.
- [6] 五十嵐 幸雄. 「知的」検索技術が製品に、手間、洩れ、無駄を削減. 日系エレクトロニクス, (705):63-70, 1997.
- [7] 長尾 真. 自然言語処理. 岩波書店, 1996.
- [8] 尾崎 正太郎, 黒橋 禎夫, and 長尾 真. 意味ネットワークからの文章生成. 情処自然言語処理研究会, 120(20):133-140, 1997.
- [9] 安西 裕一郎. 認識と学習. 岩波書店, 1989.