情 報 学 基 礎 51－10
自 然 言 語 処 理 127－10
（1998. 9. 17）

複数マニュアルの自動ハイパーテキスト化における類似度計算手法について

岡村 潤　　田中 俊一　　森 辰則　　中川 裕志
横浜国立大学 工学部 電子情報工学科

複数マニュアルにおける自動ハイパーテキスト化は，電子機器やソフトウェアに添付されるマニュアルが目的別に分冊されることが多い昨今，マニュアルを読むユーザ，マニュアルライターの両視点からみて，重要な技術であるといえる．ユーザが複数マニュアルを読み進めていく場合，一連の操作手続きなどが書かれた文書単位（セグメント）に基づく参照が重要となる．本稿では，セグメント同士間でのハイパーリンクの自動構築について考える．また，その際の類似度計算手法について，語の共起情報と語彙連鎖の利用について説明し，その効果を考察する．

The Method for Similarity Calculation in Automatic Hypertext Generatiton from Related Manuals

Jun Okamura, Shun'ichi Tanaka, Tatsunori Mori and **Hiroshi Nakagawa**
Division of Electrical and Computer Engineering, Faculty of Engineering,
Yokohama National University
{jun@forest,tanashun@forest,mori@forest,nakagawa@naklab}.dnj.ynu.ac.jp

The automatic hypertext generation from related manuals is of use for manual reader and writer, because recently manuals of products become large and often consist of separeted volumes by various purposes. Links between highly relevant segments among related manuals, in terms of operation sequences, are of great help in reading these manuals. In this paper, we propose a method for linking relevant segments between a set of related manuals, and describe new techniques to calculate similarities among segments that show much better performance than the conventional $tf \cdot idf$ method.

1 はじめに

文書間の参照・被参照情報をもつハイパーテキストは，読み手にとっては，参照したい部分を目次や索引から探しだす作業を軽減できるという点で有用である．その反面，その参照・被参照情報の設定は今現在，ほとんど人手によって行なわれており，大規模化・複数化する文書におけるその作業は大きな負担となる．自動ハイパーテキスト生成は，その問題を解決する重要な技術であるといえるだろう．

従来の自動ハイパーリンク生成は，ある機能語やキーワードにあたるアンカー語に注目し，そのアンカー語からテキスト中のある部分への自動リンク生成を目的とするものが多かった．

最近のマニュアルはユーザのレベルや使用目的に合わせて分冊化されることが多くなってきており，そのような関連マニュアルにおいては，ある機能語の説明だけではなく，一連の操作手続きに基づいた文書の参照情報が重要となる．

我々は各種電子機器やソフトウェアに添付されるマニュアルからの自動ハイパーテキスト生成について取り組んでおり，従来の語単位でのリンク生成に対して，マニュアル中の文書小部分（これを以後“セグメント”と呼ぶ）単位でのリンク生成を自動的に行なうシステムをこれまでに提案している [大森 97]．

本システムにより生成されるハイパーテキストの概

念図を図1に示す.

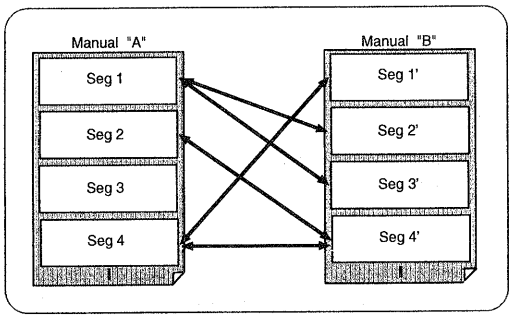

図 1: 本システムにおけるハイパーテキストの概念図

本システムはセグメント中の語を抽出し，$tf \cdot idf$ 法 により重要度を付加し，ベクトル空間モデル に基づくセグメント間の類似度計算を行ない，その類似度の高いものから 自動的にハイパーリンクを生成する.

マニュアルのような文書においては，セグメント間での類似度計算に，セグメント中の語句の他に，文構造などの情報を利用することができる．これが通常の情報検索との相違点であり，この文書構造などの情報を類似度計算にどのように反映させるかという点で，工夫の余地が大きいといえる.

本稿では複数マニュアルにおけるセグメント間での類似度計算に語の共起情報と語彙連鎖を利用することを提案し，またその効果について考える.

2 複数マニュアルのセグメント間の類似度計算

我々の提案するシステムは，マニュアル中のセグメントを一つの単位として考えている.

ここでいう“セグメント”とは，操作手順や語の説明など，意味的にまとまりのある文書小単位のことを想定している．マニュアルにおいては，文書中の「章」「節」といった構造単位で 一つの操作手順が書かれていることが多い．よって，我々はここで本稿でいうセグメントを以下のように定義する.

セグメント：
マニュアル中の「章」や「節」(LaTeX でいう section, subsection など) に相当する文書小単位.

また，システムの対象とするマニュアルは，次のようなものである.

対象となるマニュアル： 同カテゴリのマニュアル群.

これは，例えばあるアプリケーションソフトウェアに添付されるチュートリアルマニュアルとリファレンスマニュアルなどが考えられる．このようなマニュアル群においては，同じ単語は同語義を示し，また同じ概念は同じ表記の語により指し示されることが期待できるため，シソーラスなどの概念辞書を用いる必要がないと期待できる.

2.1 自動ハイパーテキスト生成システム

我々の提案する複数マニュアルの自動ハイパーテキスト生成システムは，図2 に示すように 4 つのサブシステムからなっており，以下のような手法でハイパーテキストを生成する.

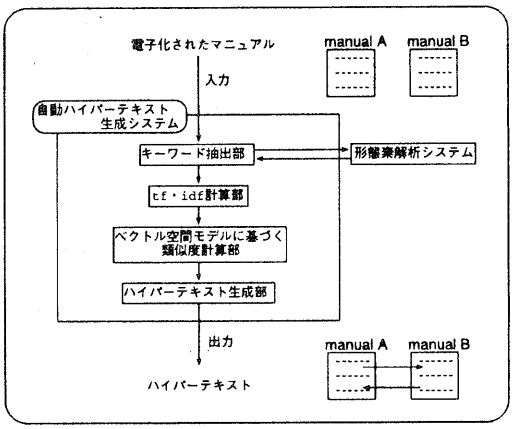

図 2: 自動ハイパーリンク生成システムの構成

1. 対象となるセグメント群から形態素解析によって語を抽出し，それぞれに $tf \cdot idf$ 値によって重要度を与える．tf はセグメント内における語の出現頻度であるが，ここではセグメント内の全形態素数で正規化している.

2. ベクトル空間モデルによってセグメント間の類似度を計算する．類似度計算においては，以下のようなベクトルを生成する.

   - 一つのセグメントに一つのベクトルを対応させる.
   - ベクトルの各次元には各単語が，各成分には対応する単語の重要度，すなわち単語の $tf \cdot idf$ 値が割り当てられる.

3. 計算の結果，セグメント組合せの中で類似度の高いものからハイパーリンクを生成していく.

現在，システムの出力は HTML としている．システムの利用画面の例を図3に示す.

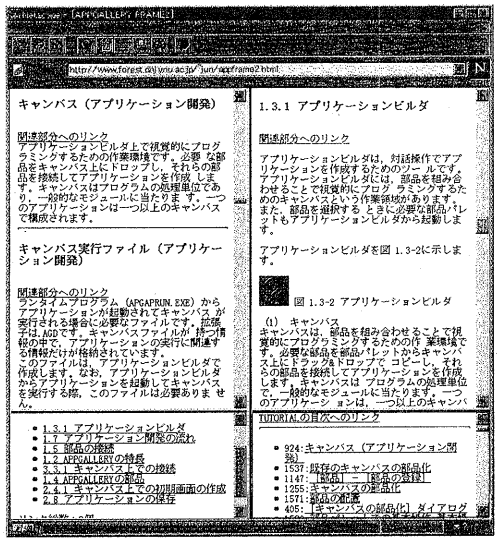

図 3: システムの利用画面

画面を縦分割し，左右フレームに異なる 2 つのマニュアルがそれぞれ表示される．それぞれのフレームは更に上下分割され，下のフレームには現在表示されているセグメントのリンク先が表示されており，いずれかをクリックすることにより，参照先がそれぞれのフレーム上部分に再表示される．その後も同様にリンク先をたどっていくことができる．

2.2 セグメント間類似度計算の補正

2.1で述べたように，本システムにおけるセグメント間類似度計算には，標準的な $tf \cdot idf$ 法 と ベクトル空間法 を用いている．

マニュアルのハイパーテキスト化においては，文章中の一連の操作手順などに注目したセグメント間対応付けが重要となるが，それには1で触れたように，文書の構造情報などを利用して類似度計算に反映することができる．

このことによって，人間がより「適当である」と思えるセグメント対応を類似度計算の結果として得ることができると考えられる．

我々は，類似度計算に反映させるものとして，語の関連に注目した．文書中における語の関連には，その構造によって次の 3 つのレベルが考えられる．

- **文内レベルでの語の関連**
- **セグメント内レベルでの語の関連**
- **セグメント間レベルでの語の関連**

語の関連の範囲がマニュアル中の大きな構造になるにつれて，より大局的な操作や話題の流れを考慮することになるといえる．我々はこの視点から，より高精度なセグメント対応を得るために，

- **格情報・語の共起情報**
- **語彙連鎖 (Lexical Chain)**

という二つをセグメント間の類似度計算に反映させることを考えた．

格情報・語の共起情報は一つの操作に注目しており，これは文内レベルでの語の共起を考慮する手法である．語彙連鎖の利用は，セグメント間に跨る語の出現を類似度計算に反映させる方法であり，実際には隣接するセグメント間における語の関連を考慮していることになる．

次章から，各手法について述べる．

3 共起情報の利用

一般的に，操作の説明は「スイッチをビデオ側に合わせる」のように

名詞 1- 格助詞 1 名詞 2- 格助詞 2 … 動詞

といった操作対象を表す名詞と操作内容を表す動詞で表される．そこで，文中の名詞や動詞の間の関係を利用することで，その文の表す操作 (の一部) に重きを置くことにより，より高精度なセグメント間の対応を取ることができると考えられる．我々は，単語の共起情報によってセグメント内の単語頻度 tf を補正して類似度計算を行う．

3.1 共起情報を単語頻度 tf を補正する方法

情報検索における文書の重要度決定に，検索要求文内で共起している単語対の共起重要度を利用すると，同じ再現率に対する適合率が向上することが報告されている [高木 96]．本稿では文書間のハイパーテキスト化を考えているので，対象となる両方のマニュアルについて，出現する全ての共起単語対についての共起重要度 cw を計算し，類似度計算に反映させることを考える．さらに高木らの方法に加えて格情報を考慮する．図4に本手法のイメージを示す．

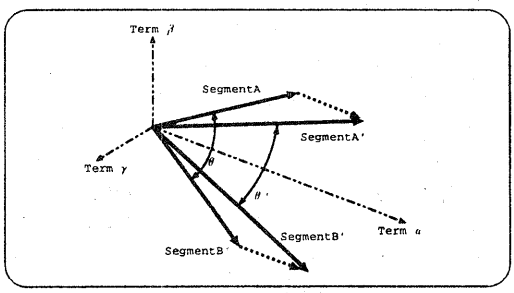

図 4: 共起を類似度計算に影響させるイメージ

共起による効果はベクトル計算をする時に付加する．実際には，共起は単語の対を単位としているので，本手法では操作としての共起があるところに対して影響をあたえていることになる．

この手法では，2 セグメント d_A, d_B 間の類似度計算において，両セグメントに出現している共起単語対について，tf の値を次のように補正する．ある語 t_k がセグメント d_A に f 回出現した場合，新たに $tf'(d_A, t_k)$ を文書内出現頻度として語の重要度を算出する．$tf'(d_A, t_k)$ は以下の式により計算する．

$$
\begin{aligned}
tf'(d_A, t_k) &= tf(d_A, t_k) \\
&+ \sum_{t_c \in T_c(t_k, d_A, d_B)} \sum_{p=1}^{f} cw(d_A, t_k, p, t_c) \\
&+ \sum_{t_c \in T_c(t_k, d_A, d_B)} \sum_{p=1}^{f} cw'(d_A, t_k, p, t_c)
\end{aligned}
$$

ここで，$T_c(t_k, d_A, d_B)$ は d_A, d_B の両セグメントで t_k とある範囲内の位置で共起している単語の集合である．p は，セグメント d_A 内で，ある語 t_k が出現する場所を表しており，セグメント内の全ての出現箇所に対しての cw の和を計算している．この計算を T_c に含まれる全ての単語について行い，tf に加算する値を得る．

また，cw は，共起を調べる単語として名詞のみを考慮した共起重要度であるが，cw' は，名詞とその直後に出現する格助詞を一つの単語と考え，cw と同様に求めたものであり，格助詞と名詞の組に関する共起に着目した共起重要度である．

次に共起重要度 cw の算出法を説明する．cw' についても名詞と格助詞の組を 1 つの単語と見なす以外は算出法は同様である．まず，t_k と t_c における語間の近接出現係数 $\alpha(d_A, t_k, p, t_c)$ と 共起係数 $\beta(t_k, t_c)$ を次のように定義する．

$$
\begin{aligned}
\alpha(d_A, t_k, p, t_c) &= \frac{d(d_A, t_k, p) - dist(d_A, t_k, p, t_c)}{d(d_A, t_k, p)} \\
\beta(t_k, t_c) &= \frac{rtf(t_k, t_c)}{atf(t_k)}
\end{aligned}
$$

$d(d_A, t_k, p)$ はどれくらいの距離までを共起の範囲とするかを表すパラメタである．本稿では 1 つの意味的なまとまりである一文中の単語の共起を見ており，$\alpha(d_A, t_k, p, t_c)$ は文内に共起した単語についてのみ計算する．よって，$d(d_A, t_k, p)$ は注目している動詞句が出現している一文の形態素数である．また，$dist(d_A, t_k, p, t_c)$ は，セグメント d_A で p 回めに出現した t_k について単語数で計算した t_c との距離である．$atf(t_k)$ は注目しているマニュアル内の t_k の出現総数，$rtf(t_k, t_c)$ は一文内に共起している t_k と t_c の出現総数である．

次に，t_k の共起語 t_c の近接出現共起単語の重要度 $\gamma(t_k, t_c)$ を定義する．N は各マニュアル中のセグメント数であり，$df(t_c)$ は t_c の出現する文書数である．

$$
\gamma(t_k, t_c) = log(\frac{N}{df(t_c)})
$$

以上で定義した，近接出現係数 $\alpha(d_A, t_k, p, t_c)$，共起係数 $\beta(t_k, t_c)$，接出現共起単語重要度 $\gamma(t_k, t_c)$ から，セグメント d_A 内の p 番めに出現する語 t_k の共起重要度 cw を次の式で表す．

$$
\begin{aligned}
&cw(d_A, t_k, p, t_c) \\
&= \frac{\alpha(d_A, t_k, p, t_c) \times \beta(t_k, t_c) \times \gamma(t_k, t_c) \times C}{M(d_A)}
\end{aligned}
$$

$M(d_A)$ はセグメント d_A 内の形態素数であり，tf と同様の正規化を行なっている．C は共起重要度正規化係数である．この値は，大きいほど共起重要度が tf にあたえる影響が大きくなる．

我々はこの手法についての実験をすでに行なっており [HTNJ98]，2.1で述べた標準的な $tf \cdot idf$ による類似度計算よりも同じ再現率に対する適合率が向上することを確認している．

4 語彙連鎖の利用

一般にマニュアル内容にも話の流れがあり，$tf \cdot idf$ 法では検出されないような，複数のセグメントにまたがって出現する重要な概念がたびたび登場する．

このような概念をとらえることができれば，セグメント間の対応付けの精度向上に有用であると考えられる．我々は，この効果を類似度計算に反映させるために語彙連鎖の導入を考えた．

4.1 語彙連鎖

語彙連鎖とは，文章中で語彙結束関係にある語のまとまりのことである．一般に言う語彙結束性とは，語の意味的なつながりのことであり，この意味での語彙連鎖は概念辞書上の同一カテゴリ (意味分類) に属するものとして計算される [Gre96]．しかし我々のシステムは同カテゴリのマニュアル群を対象としているため，同じ単語は同語義を示し，また同じ概念は同じ表記の語により指し示されることが期待できるため，シソーラスなどの概念辞書を用いる必要がないと期待できる．そこで，本稿における語彙連鎖を以下のように定義する．

語彙連鎖：

文書中で同じ語が連続して出現している部分．

4.2 語彙連鎖の効果と導入

語彙連鎖を用いると，セグメントを越えて出現する語を捉えることにより，複数のセグメントに渡る重要語を見つけることが可能となる．

本稿における語彙連鎖の概念図を図5に示す．

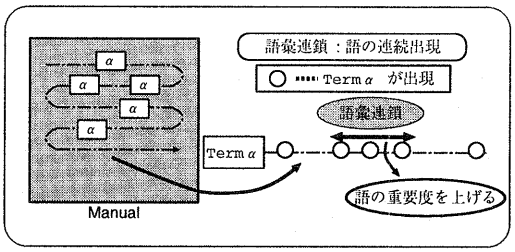

図 5: 語彙連鎖の概念

我々のシステムにおける語彙連鎖は，語彙結束関係を意味的なつながりでなく，同一語の連続出現性とみなしている．よって，語彙連鎖はセグメント軸上に構成されることになる．

ある語が語彙連鎖を形成している場所は，その語についての何らかの説明・操作内容などが記述されていると考えることができる．よって語彙連鎖により，同じ語でも，その語が話題の中心になっているセグメントと，そうでないセグメントによって重要度に差をつけることができる．実際には2.2で述べたように，語彙連鎖を利用することによって，隣接するセグメント間における語の関連というものを考慮していることになる．

4.3 語彙連鎖のパラメタと類似度計算への反映方法

語彙連鎖は類似度計算の補正のために構築されるものである．その視点からは，例えばマニュアル全体に跨って構築されてしまう語彙連鎖や，あまりにも短い語彙連鎖などはあまり意味がないと考えられる．よって，形成される語彙連鎖の連鎖長などに対していくつかのパラメタを設定する必要がある．

我々は語彙連鎖の導入において以下のような3つのパラメタを設定した．

- **連鎖長閾値:** 連鎖の長さに関するパラメタ．ここでは，以下の2つを考える．
  - **長連鎖閾値:** どの長さ“以下”の連鎖を語彙連鎖と見なすか．逆に，どの長さ以上の語の連鎖は語彙連鎖とみなさないか．
  - **短連鎖閾値:** どの長さ“以上”の連鎖を語彙連鎖と見なすか．逆に，どの長さ以下の語の連鎖は語彙連鎖とみなさないか．
- **空白閾値:** 語と語の間の出現距離(時間的距離)，すなわち語の間のギャップ(空白)をどこまで無視し，連鎖とみなすか．

語彙連鎖は，以下のようにして構築される．

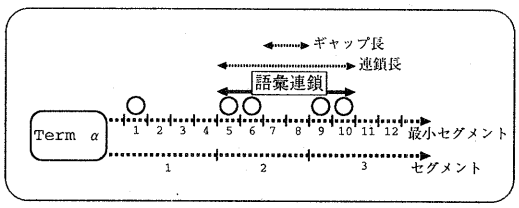

図 6: 語彙連鎖の構築

1. まず，セグメントを，それを構成するより小さな単位に分解する．ここでは，このセグメント中の小単位を「最小セグメント」と呼ぶ．本稿では，最小セグメントを「一文」としている．
   例えば図6においては，セグメント2は4つの単位(5,6,7,8)に分解される．
2. その最小セグメント中に，注目した語(図6の例では“Term α”)が出現するかどうかをチェックする．
   図6中の“○”がこの“Term α”が出現したことを示している．
3. 次に，語の出現の間にギャップがあるとき，そのギャップが空白閾値以下の場合はそれを無視し，語の出現連鎖が続いているものとみなす．
   図6の例では，最小セグメント7,8が無視される．
4. 上の処理の結果できた語の出現連鎖について，長連鎖閾値，短連鎖閾値を当てはめ，長すぎる，もしくは短すぎる出現連鎖は語彙連鎖とみなさなくする．
5. 残った語の出現連鎖を語彙連鎖とみなす．
   図6の例では，最小セグメント5～10において，語彙連鎖が形成されているとする．
6. この処理をセグメント間類似度を計算するために抽出された語すべてについて行なう．語彙連鎖は，それぞれの語に対して独立に構築される．

語彙連鎖をセグメントの類似度計算に反映させるには，語の tf 値を補正を行なうことによって実現した．すなわち，あるセグメント d_i において語彙連鎖を構成している語 t_k の tf 値を以下のように補正する．

$$tf'(d_i, t_k) = tf(d_i, t_k) \times (1 + \delta)$$
$$(\text{ただし} \delta > 0)$$

δについては，平均適合率を最大にするものとして今回は$\delta = 3.5$としている．

この語彙連鎖を使った類似度計算手法についても，我々は実験の結果，2.1で述べた標準的な手法よりも同じ再現率に対する適合率が向上することをすでに確認している [岡村 98]．

また上述のパラメタ群については，ある2つのマニュアル組合せにおける最適パラメタ群が，他のマニュアル組合せについてもある程度有効であることが我々の交差検定による実験で確認されている．

5 共起情報と語彙連鎖の統合

3章と4章において2.1で述べた類似度計算に対する補正方法を述べた．ここでは，この二つの手法を組み合わせることによって得られる効果について考察する．

共起情報と語彙連鎖は異なったプロセスを経て類似度計算を実現しているため，それぞれ異なったマニュアル中の概念・要素に影響を与えていると考えられる．そこで，今回は両手法の組合せを以下のようにして考える．すなわち，あるセグメント組合せ(n, m)について，共起情報による類似度$Sim_{coc}(n, m)$と，語彙連鎖による類似度$Sim_{lex}(n, m)$から以下のようにして新たな類似度$Sim_{int}(n, m)$を得る．

$$Sim_{int}(n, m) = Sim_{coc}(n, m) + \chi \times Sim_{lex}(n, m)$$

χは，共起情報と語彙連鎖のどちらの手法に重きをおくかを決定する結合重みパラメタである．この値が大きいほど語彙連鎖の類似度を重視することになる．実際には，両手法とも類似度はベクトル空間法によって求められるので，両手法を統合した類似度$Sim_{int}(n, m)$は，余弦値の線形和となる．

5.1 各類似度計算手法の評価

ここでは，今まで述べてきたセグメント間の類似度計算手法に対する評価を行なう．3.1,4.3でも述べたように，我々は2.1で述べた標準的な類似度計算手法と格情報・語彙連鎖の各手法の比較・評価については，すでに実験を行なっており [HTNJ98, 岡村 98]，その有効性を確認している．そこで今回は，

1. 共起情報を考慮した手法
2. 語彙連鎖を考慮した手法
3. 共起情報と語彙連鎖を組み合わせた手法

という三つの手法について，類似度計算の結果ランキングされるセグメント対応を評価した．

評価には，情報検索で一般的に利用される再現率(*recall*)，適合率(*precision*)を用いる．

$$\text{再現率}(recall) = \frac{\text{検索された適合対応数}}{\text{全ての適合対応数}}$$

$$\text{適合率}(precision) = \frac{\text{検索された適合対応数}}{\text{検索された対応数}}$$

再現率はある順位までに出現する正解の割合，適合率はノイズの割合をそれぞれ示す．

本システムが対象としている大規模マニュアルについては，我々は既に実験を行なっており，そこである程度のシステムの有効性を確認している [大森 97]．

しかし大規模マニュアルにおいては完全な正解集合を人手で作るのが困難であるため，3，4，5章で述べてきた各手法の効果を考察するには，完全な正解集合を作成することができる小規模のマニュアルで実験を行う必要がある．そこで我々は，同一メーカの3つのビデオのマニュアルA [三菱電a]，B[三菱電b]，C[三菱電c] を用いて実験を行なった．各マニュアルのデータは次の通りである．

マニュアル	A	B	C
セグメント数	32	33	28
大きさ (kbyte)	88	112	102

このマニュアル群のうち，A⇔C ，B⇔Cという組合せで本システムによるハイパーテキスト化を行い，セグメント対応の評価を行なった．各マニュアル組合せにおけるデータは次の通りである．

マニュアル組合せ	A ⇔ C	B ⇔ C
セグメント全組合せ数	896	924
うち正解組合せ数	60	65

適合率・再現率グラフを図7,8に示す．図中に示されたパラメタ群は最適値であり，語彙連鎖のパラメタの単位は“一文”である．

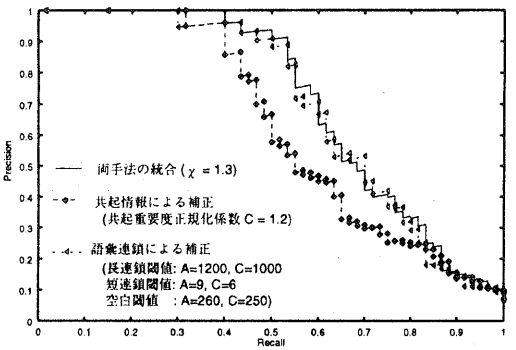

図 7: 適合率・再現率グラフ (A⇔C)

また，図7,8における適合率の11点平均を表1に示す．

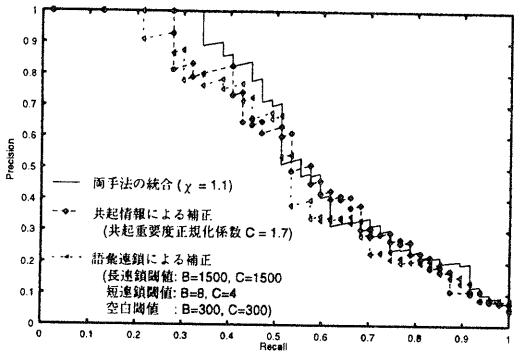

図 8: 適合率・再現率グラフ (B⇔C))

マニュアル組合せ	A ⇔ C	B ⇔ C
共起情報	0.6281	0.5853
語彙連鎖	0.6976	0.5700
両手法の統合	0.7055	0.6208

表 1: 各手法における適合率の 11 点平均

考察

図 7において共起情報と語彙連鎖を用いた手法を比較すると，語彙連鎖の方が共起情報よりも再現率の低～中域において適合率が向上しており，より高精度なセグメント対応を得ていることがわかる．しかし，図 8においてはそれほどの違いが見られず，むしろ再現率の低～中域においては 共起情報を用いた手法の方が良い結果を得ている．つまり対象となるマニュアルに応じて結果が変わることになり，どちらの手法がより優れているとは断言できない．

これと両手法を統合した結果を比べてみると，再現率の低域においては共起情報，語彙連鎖の両手法より適合率が上昇しており，また再現率全域にわたってほぼ両手法の最良値をカバーしていることがわかる．つまり，マニュアルに応じて生ずる両手法の優劣の変動を吸収していることになる．この点において両手法を統合させる最大の有効性が確かめられたといえる．

つぎに，両手法の結合重みパラメタ χ について考察する．χ の最適値は図 7，8に示されている値だが，$\chi = 1$ のとき，統合された類似度は両手法による類似度の単なる和として表される．この χ の基準値と最適値について，適合率の 11 点平均からみた比較を表 2に示す．

適合率の 11 点平均からみた比較では，最適値と基準値，また χ の値が 1 の近傍では性能に大きな違いはなかった．その例として，χ の最適値と基準値における適合率・再現率グラフ (マニュアル組合せは A⇔C) を図 9に示す．

マニュアル組合せ	A ⇔ C	B ⇔ C
χ 最適値 (最適値)	0.7055 (1.3)	0.6208 (1.1)
χ 基準値 ($\chi = 1$)	0.6934	0.6175

表 2: χ 値に注目した適合率の 11 点平均

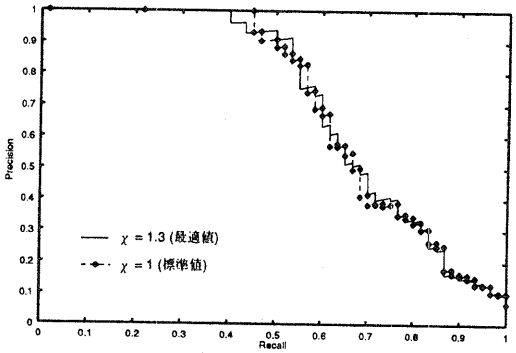

図 9: χ 値に注目した適合率・再現率グラフ (A⇔C)

表 2，図 9より，$\chi = 1$ のとき，すなわち両手法による類似度の単純な和によって，ある程度精度の高いセグメント対応が得られることを示していると考えられる．このことから共起情報と語彙連鎖という，文章構造において対比するレベルでの語の共起を考慮する手法が，お互いに直交する性質のものであると考えられる．

6 おわりに

本稿では，複数マニュアルの自動ハイパーテキスト生成システムについて述べ，その中のセグメント間の類似度計算について，標準的な $tf \cdot idf$ 手法を補正する二つの手法である共起情報と語彙連鎖について述べた．また，その二つの手法を統合する実験を行ない，その有効性を示した．

共起情報と語彙連鎖の統合については，今回は各手法の類似度計算結果を線形結合させるというマクロな形をとったが，ベクトル計算の際に両手法を統合させる方法も考えられる．しかしこれについては，両手法における複合語の最適条件 (共起情報は複合語考慮，語彙連鎖は複合語考慮せずが最適条件) などをどのように組み合わせるかという問題があり，これからの課題である．

我々はマニュアルという限られた形態をとる文書に注目してシステムを構築している．しかし，本稿で述べた 2 つの類似度計算手法は，マニュアルに対してだけではなく，他の電子化文書にも利用できると考えられる．

とくに語彙連鎖のように文書の構造情報に近いものを扱うものは，大規模な電子化文書がネットワーク上

に散在することが多くなると予想されるこれからの時代において，情報検索という点で重要な技術になるといえるだろう．

参考文献

[Gre96] Stephen J. Green. Using lexcal chains to build hypertext links in newspaper articles. In *Proceedings of AAAI Workshop on Knowledge Discovery in Databases, Portland, Oregon*, 1996.

[HTNJ98] H.Nakagawa, T.Mori, N.Omori, and J.OKamura. Hypertext authoring for linking relevant segments of related instruction manuals. In *Proceedings of COLING-ACL '98*, pp. 929–933, 1998.

[大森 97] 大森信行, 蔵方隆宏, 岡村潤, 森辰則, 中川裕志. 情報検索手法を用いた複数文書間の関連箇所抽出 – 電子化マニュアルへの適用 –. 言語処理学会第 3 回年次大会, pp. 257–260, 1997.

[岡村 98] 岡村潤, 大森信行, 山口登志実, 森辰則, 中川裕志. 共起情報を考慮した $tf \cdot idf$ 法に基づく関連文書間の自動ハイパーテキスト化. 言語処理学会第 4 回年次大会, pp. 560–563, 1998.

[高木 96] 高木徹, 木谷強. 単語共起関係を用いた文書重要度付与の検討. 情報学基礎研究会報告 96-FI-41-8, 情報処理学会, 1996.

[三菱電 a] 三菱電機株式会社. 三菱ビデオ HV-FZ62 取扱説明書.

[三菱電 b] 三菱電機株式会社. 三菱ビデオ HV-BZ66 取扱説明書.

[三菱電 c] 三菱電機株式会社. 三菱ビデオ HV-F93 取扱説明書.