

〔公 開〕

TR-C-0119

連想記憶を用いた異質性を含む  
情報の抽出手法の検討

西本 一志  
Kazushi NISHIMOTO

1 9 9 5 4 . 1 8

A T R 通信システム研究所

# 連想記憶を用いた異質性を含む情報の抽出手法の検討

西本一志

ATR通信システム研究所

知能処理研究室

## 1 はじめに

本テクニカルレポートでは、仮想門外漢システムを構成するための一要素としての、異質性と関連性を併せ持つような情報抽出手法について説明する。このような性質の情報には、発散的思考過程において人の固定観念を打開し、発想を刺激する作用があると考えられる[川喜田67]。

第2章では、関連性があると同時に異質性がある情報を人がどのようにして獲得しているか、そしてそのような情報によってどうやって固定観念が打開されるかを、簡単な例を用いて考察する。第3章では第2章での考察に基づき、異なった視点からの情報抽出を実現するための「門外漢モデル」を提案する。第4章では門外漢モデルに基づき作成した実験システムの基本的構成を説明する。第5章では固定観念について考察するとともに情報の関連性と異質性についても詳細に検討を加え、情報提供型の発散的思考支援ツールの有効性評価方法を提案する。第6章では実験システムを用いた三つの実験を行い、この結果に提案した評価方法を適用し、門外漢モデルに基づく本システムの有効性の評価や動作の検証を行なう。第7章では、今のところ実装されていないが今後必要となると思われるいくつかの機能について、その実現方法を説明する。

## 2 視点の変更による固定観念の打開

この章では、例を用いて異なる視点から異質な情報がいかにして得られるか、またそのような情報がいかにして人の固定観念の枠を打開するかを説明する。

月の模様を見てそれが何に見えるかについては、いろいろな国、あるいは文化圏によって様々に異なる。例えば日本では月の模様は兎の餅搗きであるが、アメリカでは男の人の顔、中国では女性（妖精）と兎、モロッコではモロッコの先王が馬に騎乗しているように見えるそうである。月の模様それ自体は誰が見ても同じものであるはずなのに、このように異なったものと解釈されるのは、おそらく以下のような処理過程を経ているからであろう：

- (1) 与えられた全体のパターンから、なんらかの特徴的な部分を切り出す。
- (2) 切り出した部分を、自分の持つ知識を用いて連想的に拡張する。
- (3) 拡張されたパターンと類似した既知のパターンを取り出す。

このような処理過程において、第1の段階で異なった部分を切り出し、さらにこれを第2の段階で異なった知識によって連想拡張することによって同じパターンからまるで違う情報が取り出されるものと考えられる。

我々日本人のほとんどは、月の模様は兎の餅搗きという固定観念ができあがっている。そして月の模様を見てそれを兎の餅搗きと解釈するための固定的な処理過程が完成している。つまり、常にその模様から同じ部分を切り出し、同じように拡張して兎のイメージと重ね合わせる処理を行なってしまう。したがって、たとえばアメリカ人が「月の模様は男の顔だ」と言っても、初めはなぜそう見えるのか、どこをどう見ればそう見えるのかがなかなか理解できず混乱するであろう。しかし、具体的に月の模様のどの部分がどのような男の人の顔に見えるのかをちょっと説明してもらえば、たいていの場合その見方を理解し、月の模様に男性の顔を重ね合わせることができよう。この時、同じ全体パターンから別の部分パターンを切り出し、その部分パターンを従来と異なるやり方で連想拡張するという、「視点の変更」が行なわれていると考えられる。そして、ひとたびこのようにして自分が従来から持っていた「兎の餅搗き」を自動的に導き出してしまふ固定的処理過程を打破すれば、おそらく説明を受けなくても月の模様を「妖精と兎」あるいは「騎乗したモロッコ先王」と見ることは以前より容易に実現できるようになっているであろう。

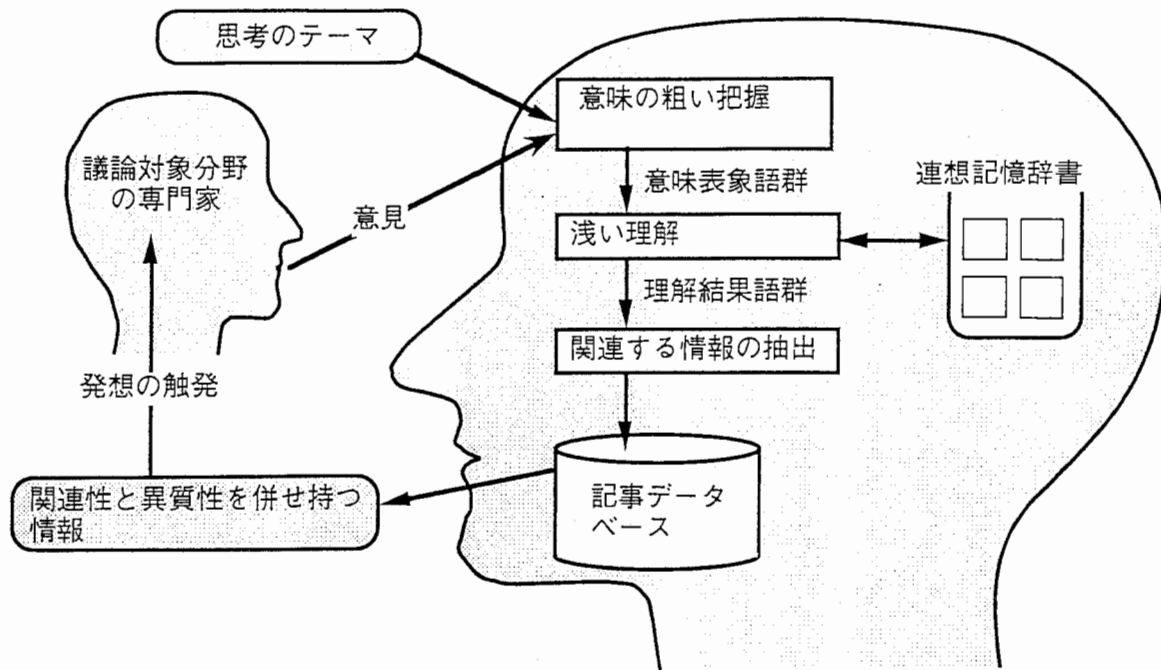


図3-1 門外漢モデル

つまりこの例では、「月の模様は男の顔に見える」という日本人にとっては異質な情報によって、日本人が持っていた「月の模様は兎の餅搗き」という固定観念が打開されたことになる。そこで本研究では、このような異質性を含む情報の抽出手法について検討する。

### 3 門外漢モデル

本研究における、関連があると同時に異質性のある情報の抽出は、基本的に前章で述べたような部分パターンの連想拡張によって実現する。

図3-1に門外漢モデルを示す。本モデルは以下の3つの処理段階で構成される。

- (1) 意味の粗い把握：他の会議参加者が発言する内容を表面的に把握する処理過程
- (2) 浅い理解：把握した発言内容を、自分の手持ちの知識の範囲で浅く理解する過程
- (3) 関連する情報の抽出：浅い理解の結果に基づき、自分の手持ちの情報の中から関連があると思われる情報を抽出する過程

これらの3つの過程は、前章で説明した月の表面の模様から兎の餅搗きのイメージなどを取り出すための3つの処理過程のそれぞれにほぼ対応する。ただし、月の模様のような、本来無意味な情報とは異なり、人の発言には必ず意味が含まれる。このため、本モデルの処理では元になる発言の意味を考慮するようにしている点で先に示した月の模様の場合の処理と異なっている。

#### 3.1 意味の粗い把握過程

この過程の目的は、まず他の会議参加者の発言の意味内容をおおまかに把握することにある。

この過程の処理は、入力された発言（テキスト）から、その発言の内容をおよそ示すであろう語群を取り出すことである。つまりここでは、文（章）から取り出されたそのような語群が、たとえ文（章）の形を成していなくても元の文（章）の意味をほぼ保持していると仮定している。これは、語群に含まれる各語が持つ意味の多様性が、同時に現れる語によって相互に抑制され、結果的に全体としてある程度の具体的な（曖昧性の低い）意味を示すからである。このような語群を以下では意味表象語群と呼ぶ。

### 3.2 浅い理解過程

この過程の目的は、前の「意味の粗い把握過程」で把握した発言の意味を、ある別の視点から理解することにある。

この過程の処理は、意味表象語群を、なんらかの連想辞書を用いて連想的に拡張することである。つまりここでは、ある語から、それに関連する語（群）を連想することをひとつの理解であるとみなしている。たとえば、「りんご」というものを説明しようとするとき、しばしば「りんご」と関連する種々の語（たとえば、赤い、果物、青森、ジャム、ニュートン、マッキントッシュ、など）を列挙する。つまりこれらの語との相互作用（曖昧性の抑制作用）によって「りんご」という言葉の意味が明確化される。このような理由から、ある語から、それに関連する語を想起することをひとつの理解とみなしている。

このようにして意味表象語群に含まれる各語をそれぞれ連想拡張すると、拡張された語の相互作用によって強く連想される語が現れる。例えば、ある意味表象語群に「りんご」と同時に「惑星」のような単語が現れていたとすれば、この語からやはり「ニュートン」という語が連想される可能性がある。このように複数の語から同じ語が連想される場合、その語は強く連想されることになる。こうして意味表象語群に含まれる各語から連想される語全ての相互作用によって最終的に強く連想されるいくつかの語によって構成される語群が得られる。この語群が示す意味が、この浅い理解過程による意味表象語群の理解結果である。この浅い理解過程で得られる理解結果の語群を理解結果語群と以下では呼ぶ。

浅い理解の過程における「理解」は、以上のような処理で得られる。ここで容易に想像できるように、この理解結果はどのような連想辞書を使用するかによって変化する。この作用が視点の変更に対応する。連想辞書を特定分野に依存した構成にすれば、理解の結果の方向性のある程度コントロールできる。例えば法律に関わる分野の連想辞書を使用すれば得られる理解結果語群は当然法律色のかかったものとなるであろう。このように、連想辞書の分野を必要に応じて特定分野のものとすることによって、本モデルは単なる門外漢ではなく、異分野の専門家として使用することが可能となろう。さらに、複数分野の連想辞書を組み合わせて使用することによって、複数の専門分野を持つような仮想会議参加者を構成できる可能性もある。これについては、6.3節で詳細に述べる。

### 3.3 関連する情報の抽出過程

この過程の目的は、前の「浅い理解過程」で得た理解の結果に基づき、記事データベースからもっとも関連のありそうな記事を取り出すことである。

この過程の処理は、理解結果語群と記事データベースに含まれる各記事との類似度を計算し、もっとも高い類似度を持つ記事を提供候補の第1とすることである。記事の分野は、必ずしも使用する連想辞書の分野に一致している必要はない。連想辞書と記事データベースを異なる分野のものとすることによって、なんらかのおもしろい効果が得られる可能性もある。ただし、あまり分野が異なると（例えば連想辞書を日本の古典文学によるもの、記事データベースを量子物理学関係のものを準備すると）、理解結果語群に含まれるほとんどの語が記事データベース中の記事に含まれず、結果的に抽出される記事の理解結果語群との類似度が非常に低くなる可能性が高い。この場合入力された発言と抽出された記事との関連性がほとんどなくなって、結果的に発想を触発する効果に乏しいものとなると予想される。従って、連想辞書と極端に分野の異なる記事データベースは使用しない方が良いでしょう。

## 4 実験システムの基本的な構成

前章で述べた門外漢モデルに基づき、実験システムを構成した。図4-1にソフトウェアの基本的な構成と処理の流れを示す。システムは形態素解析モジュール、連想記憶モジュール、データベースマネージャおよび記事データベースで構成される。各モジュールがそれぞれ順に門外漢モデルの各処理過程にほぼ対応する。

### 4.1 処理の概要

本システムを仮想門外漢として使用する前に、まず連想記憶辞書ならびに記事データベースを構築し

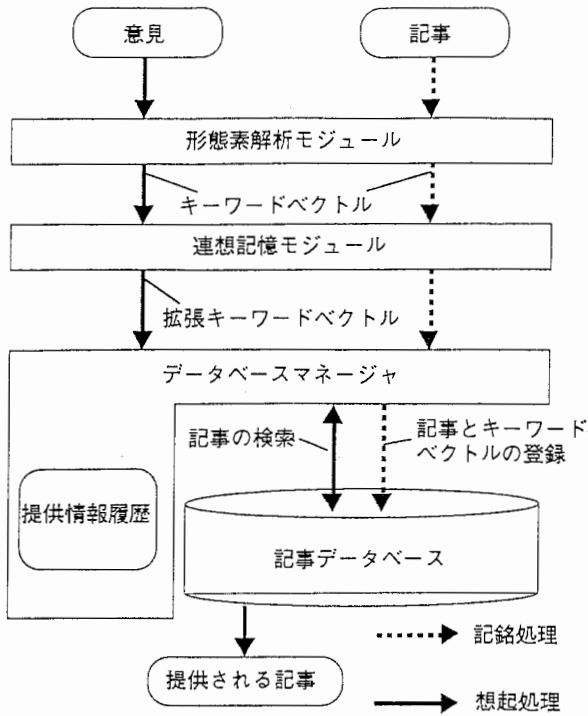


図4-1 実験システムのソフトウェアの基本的構成

データベースマネージャは、この入力された拡張キーワードベクトルと、記事データベースに登録されている各記事に付加されているキーワードベクトルとの類似度を計算する。その結果基本的にもっとも高い類似度を持った記事を抽出し、これを利用者に提供する。

## 4.2 各モジュールの詳細[西本93a]

### 4.2.1 形態素解析モジュール

このモジュールでは、入力されたテキストに対して形態素解析を行う。すなわちまず単語を切り分け（日本語の場合）、ついで文法情報および品詞辞書を用いて各語の品詞を決定する。さらに語尾変化などを行っている語についてはその語の原形に戻す。また、必要ならば同義語辞書を準備し、同義語に関してはそれらを代表するどれかひとつの語に書き直す。この解析の結果、語とその品詞との対応リストを生成する。表4-1に生成されたリストの例を示す。表中、記号!はフィールドの区切りを示す。第1フィールドは単語の正規形、第2フィールドは品詞、第3フィールドはその語のそのテキスト中における出現回数、第4フィールドはその語のそのテキスト中における最初の出現位置である。第5フィールドにSUBJと記述されている語は、文の主語だったことを示す。品詞は、nが名詞、vが動詞、unkが未知語、などである。

ついで、このリストから指定された特定の語だけを抜きだしたサブセットリストを生成する。特定の語としては、たとえば名詞と品詞同定不能語だけを採用することが考えられる。

このサブセットリストからキーワードベクトルが生成される。ある記事（あるいは発言） $A_j$ のキーワードベクトル $K_j$ は以下のようにして生成する。

なければならない。図4-1中、破線で示される処理の流れがこの処理に当たる。まずなんらかの記事を本システムに入力する。入力された記事は形態素解析モジュールで形態素解析され、キーワードベクトルが生成される。このキーワードベクトルが門外漢モデルの意味表象語群に対応する。ついで、このキーワードベクトルは連想記憶モジュールに入力され、連想記憶処理によって連想辞書が生成される。また同時に、必要であれば入力された記事とその記事のキーワードベクトルをセットにしてデータベースモジュールに入力し、記事データベースを構築する。

ついで、本システムを仮想門外漢として使用する場合の処理の流れを説明する。図4-1中、実線で示した処理の流れがこの処理に当たる。この場合、システムに対する入力とは他の会議参加者の発言である。発言はやはり構文解析モジュールで構文解析され、キーワードベクトルに変換される。このキーワードベクトルが連想記憶モジュールに入力され、なんらかの連想辞書を使用して連想的に想起を行なう。この結果連想的に拡張されたキーワードベクトルが得られる。この拡張キーワードベクトルが門外漢モデルの理解結果語群に相当する。拡張キーワードベクトルはデータベースマネージャに入力される。データベースマネージャは、この入力された拡張キーワードベクトルと、記事データベースに登録されている各記事に付加されているキーワードベクトルとの類似度を計算する。その結果基本的にもっとも高い類似度を持った記事を抽出し、これを利用者に提供する。

Interfax!unk!2!23
Jalilov!unk!1!98
Member!n!2!34
Monday!n!1!12
Moscow!n!1!1
OPON!unk!4!36
Ragimov!unk!1!105
Ramil!unk!1!30
Reuter!unk!1!2
after!prep!2!13
against!prep!1!80
agency!n!1!27!SUBJ
and!conj!3!50
announce!v!1!83
arrested!unk!1!88
as!conj!2!32
back!n!1!64
base!n!3!54
bel!v!8!56

表4-1 語と品詞の対応リスト例

$$K_j = (\delta_1, \delta_2, \delta_3, \dots, \delta_i, \dots, \delta_m)$$

$$\delta_i = \begin{cases} 1 & (w_i \in A_j) \\ 0 & (w_i \notin A_j) \end{cases}$$

ここに、 $m$ はシステムがこれまでに処理した語の種類数の総数、 $w_i$ はそのうちの $i$ 番目の単語である。つまり各単語にはユニークなID番号が付与され、キーワードベクトルの各要素はその位置が単語のIDに対応し、その要素の値は元の記事 $A_j$ 中に各要素に対応するIDの単語 $w_i$ が含まれているかいないかを示す。このID番号は単語とID番号を対応付けたリストファイルを生成して管理する。連想辞書作成時、記録する記事数が増えるにつれて、当然処理した単語の種類も次第に増加する。つまり $i$ の上限 $m$ は可変であり、ID番号管理リストファイルには順次新しい単語が追加されていく。これにともない、キーワードベクトルもモディファイされ、常にすべてのベクトルの次元を同じに保つようにする。たとえばある記事 $A_j$ を記録した際に新出単語が2つ含まれていた場合、それまでに生成した $K_1$ から $K_{j-1}$ の全てのキーワードベクトルには末尾に値0の要素が2つ追加されることになる。

なお、サブセットリストに含まれる語は元の記事（発言）に現れた順に抜き出され、その数は通常ある一定数以下に制限される。

#### 4.2.2 連想記憶モジュール

連想記憶の手法は、基本的にアソシアトロン[中野72]の手法に準ずる。まずアソシアトロンにおける基本的な連想記憶・連想想起の方法について説明する。

$n$ 個のベクトル $K_j(j=1 \sim n)$ は以下のようにして記録される。

$$M = \sum_{j=1}^n K_j K_j'$$

ここに、 $K_j$ は縦ベクトルであり、演算子'は転置を表わす。 $K_j$ の要素の値は-1/0/1の3値のいずれかである。また $M$ は連想記憶行列であり、これが連想記憶辞書に相当する。なお、当然すべてのベクトル $K_j$ の次元は同じでなければならない。このようにして連想記憶行列 $M$ を生成することによって連想記憶が行なわれる。

アソシアトロンではこのように、本来記録するベクトルは-1/0/1の3値論理をとり、-1の要素同士、1の要素同士がそれぞれ共起的關係を持つことになる。また-1の要素と1の要素は相互に負の共起的關係を持つ。0の要素は、他の-1/1の要素との関連性が不明であることを示している。

しかし本システムでは前節で述べたようにキーワードベクトルは1/0の2値論理で構成している。これはある記事（意見）中に共起した語相互には関連があるとみなし値を1にセットするが、それ以外の要素との関連については不明であるのですべて0としているのである。明確に、ある語群と別のある語群とに負の共起性（絶対同時には現れ得ないという強い排他性）がある場合には一方の語群を1、もう一方を-1で表現したベクトルを構成することによって連想記憶行列に負の共起性を反映させることも可能であるが、一般にそのような負の共起性を持つ語群というものは想定し難い。したがって本システムでは特に利用者の意図的な指定が無い限り、キーワードベクトルは2値論理で構成する。

次に連想想起について説明する。アソシアトロンでの想起は以下のようにして行なわれる。

$$R = \phi_\theta \left\{ \phi_0(M)Q \right\}$$

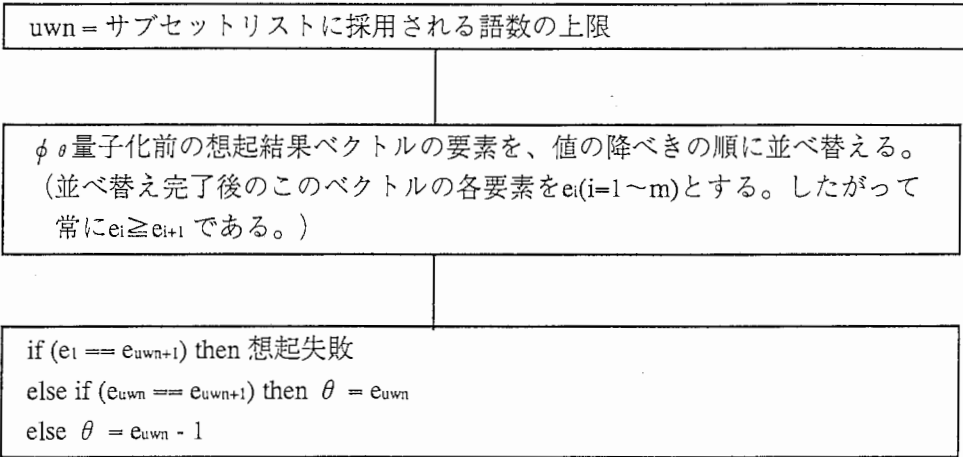
ここに、 $M$ は連想記憶行列、 $Q$ は想起の元になるベクトルであり、本システムでは他の会議参加者の発言から生成されたキーワードベクトルにあたる。 $R$ は想起の結果得られた拡張キーワードベクトルであり、本システムでは理解結果語群に相当する。また、 $\phi_0$ 、 $\phi_\theta$ は以下の量子化関数である。

$$\phi_0(x) = \begin{cases} 1; & x > 0 \\ 0; & x = 0 \\ -1; & x < 0 \end{cases} \quad \phi_\theta(x) = \begin{cases} 1; & x > \theta \\ 0; & -\theta \leq x \leq \theta \\ -1; & x < -\theta \end{cases}$$

これらの量子化関数はベクトルおよび行列にも適用可能であり、その場合は各要素がこれらの関数に従って量子化されるものとする。なお、本来のアソシエーションはすでに述べたように3値論理なので量子化関数も3値論理になっているが、本システムでは2値論理であるため、実際には負の領域に関する量子化は行なわれない。

量子化関数  $\phi_\theta$  の量子化の敷居値  $\theta$  の値は、基本的に想起結果のベクトル  $R$  の要素のうち、値が1となる要素の個数が、形態素解析モジュールでサブセットリストを生成する際に採用する語の個数の上限に一致するように、想起毎に決定する。実際には以下のようなアルゴリズムで  $\theta$  の値を決定している。

\*\*\*\*\*



\*\*\*\*\*

このようなアルゴリズムで  $\theta$  の値を決定することによって、想起結果ベクトル  $R$  のうち、値が1にセットされる要素の個数は常に  $uwn$  個以下となる。この制限を加えておかないと、次のデータベースマネージャにおける類似度の計算がうまくいなくなる。

上記のアルゴリズムで想起失敗は、たとえばベクトル  $Q$  に含まれる値1の要素が非常に少ない場合に発生する。極端な例として、値1の要素が1つしかなく、しかもその要素に対応する語の出現頻度が非常に高い場合、記録したいいずれかの記事でその語と共起したことがあるすべての語に対応する要素が全部値1となり、簡単に  $uwn$  の個数を越えてしまう。

4.2.3 データベースマネージャと記事データベース

データベースマネージャは想起結果ベクトル  $R$  と、記事データベースに格納されている各記事のキーワードベクトル  $K_j (j=1 \sim n)$  との類似度を求める。類似度  $r_j$  は以下の式によって得られる。

$$r_j = \frac{K_j' \cdot R'}{m_R} \times \frac{K_j' \cdot R'}{m_{K_j}}$$

ここに、 $m_R$  は想起結果ベクトル  $R$  のうち値1がセットされている要素の個数、 $m_{K_j}$  は記事  $A_j$  のキーワードベクトル  $K_j$  のうち値1がセットされている要素の個数である。演算子  $\cdot$  はベクトルの内積を示す。この結果、類似度の高いものが提供される情報の候補となる。

データベースマネージャは利用者ごとの提供情報履歴を持つ。つまり過去に記事データベースからの記事を利用者に提供したことがあるかのリストを保持している。データベースマネージャはこのリストを参照し、利用者に対して過去に一度も提供したことがない記事のうち最も高い類似度を持つ記事を提供する。

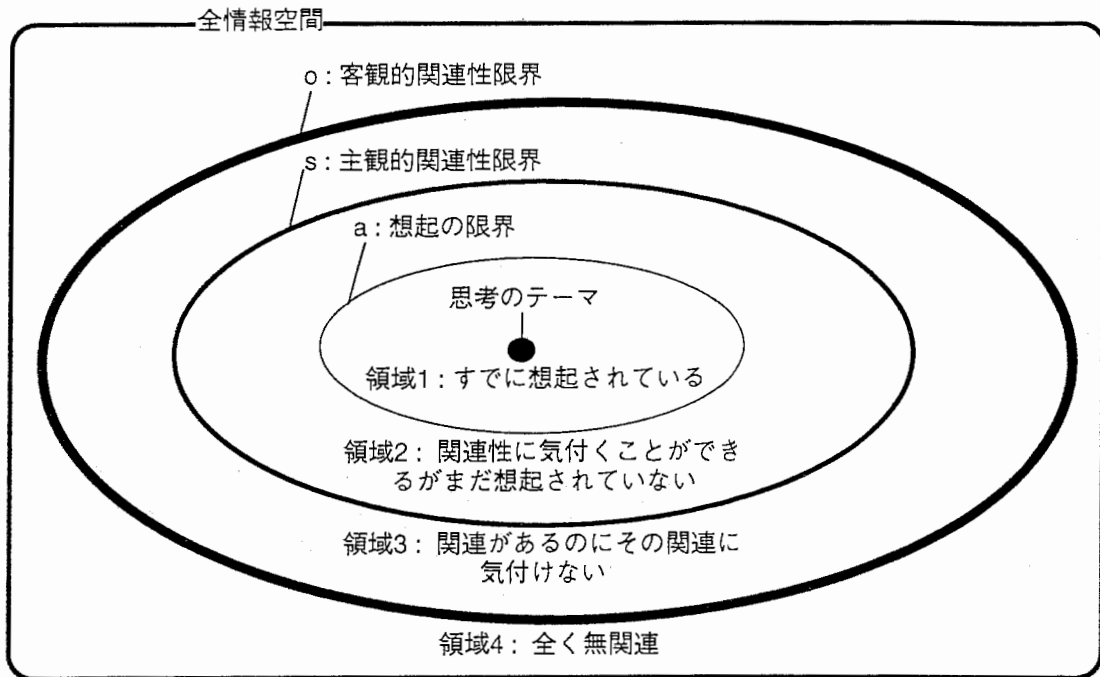


図5-1 ある思考のテーマのもとでの全情報空間の客観的／主観的観点からの分類

## 5 評価手法の考察[西本95]

本章では、発散的思考支援システムのひとつの評価手法について考察する。

### 5.1 人の発散的思考を刺激する情報の種類に関する考察

ある思考のテーマが与えられたときに、全情報空間をどのように分類するかを、図5-1に示す。

まず全空間は二つの領域に分割される。思考のテーマからみて、外側の領域は「客観的に全く無関連な情報」の領域であり、内側の領域は「客観的になんらかの関連がある情報」の領域である。図5-1の線oがこの両領域の境界である。すなわち、領域4に含まれる情報は、思考のテーマとなんの関連もないものである。

境界線oの内側の領域は、人の主観によってさらに二つの領域に分割される。外側の領域は「主観的に無関連な情報」の領域であり、内側の領域は「主観的に関連がある」情報の領域である。線sがこの両領域の境界である。すなわち、境界線oと境界線sとに挟まれる領域3に含まれる情報は、客観的には関連があるにもかかわらず、その人はその関連をなんらかの理由で認識できないようなものである。

境界線sの内側の領域は、さらに二つの領域に分割される。境界線aの内側の領域1は人が思考のテーマとの関連を認識できるし、思考のテーマが与えられた時からすでにそのような情報が想起されているような情報の領域である。境界線sとaとで挟まれる領域2に含まれる情報も人にとっては思考のテーマとの関連性を認識できる情報である。ただし、思考のテーマが与えられた段階ではまだそのような情報は想起されておらず、提供されて初めてそのような情報にも関連があったことに気づく種類のものである。

このように情報空間を分割した際、人の発散的思考を刺激する情報は、領域2と3に含まれるものであると考えられよう。

領域2に含まれる情報は「関連を見落としていた情報」である。すなわち、人はその情報を与えられればその関連性に即座に気づく。しかし、与えられるまではそのような情報、そのような関連性に気づいていない。そして与えられた瞬間に「なるほど、たしかにそんな関連もあった」と驚きとともに気づくのである。したがって、この領域に含まれる情報は非常に直截的に人の制限された視野（固定観念）を打開する効果があると期待できる。

また、領域3に含まれる情報は、その情報が与えられても依然人はその関連性には気づくことができな



い。したがって領域2に含まれる情報のような直截的、即時的固定観念打開効果はないであろう。しかし、やはりそれらの情報も人の気づかない関連性を含んでいるのである。そこで、人が「何か関連があるのか？」と考え思考・調査し、その関連性を見いだすことができたなら、やはりそのような情報も人の固定観念を打開する効果があるだろう。

一方、領域4に含まれる情報は思考のテーマと完全になんの関連もない。このような情報が人の思考に意味のある作用を及ぼすとは考えられない。領域1に含まれる情報は、直截的な関連があるので少なくとも思考を行なう上で必要な基礎的情報であることは間違いない。したがって領域4の情報とは異なって思考に意味のある作用を及ぼすものである。しかし、この領域の情報のもつ関連性はすでに人は気づいている。言うなれば、その人の固定観念の枠内にある情報なのである。このような情報が固定観念の打開に有効な作用を及ぼすとは、やはり考え難い。

以上の考察から、固定観念の枠とは境界線aないしsで表わされ、これを打開するためには領域2ないし3に含まれるような情報を抽出し提供することが有効な手段であると結論できる。

## 5.2 ある情報の発散的思考への支援寄与度の評価手法の提案

ある情報が、前節で示した分類のどの領域に含まれるものであるかを評価するために、以下に説明するような主観的評価方法を提案する。

発散的思考支援システムが提供する情報を被験者に提供し、その情報と、最初に与えた思考のテーマとに対して、以下のふたつの視点から主観的に評価してもらう。

- (1) 関連性：提示された情報が、思考のテーマとどの程度関連していたか。
- (2) 異質性：与えられた思考のテーマの下で、提示されたような情報が提供されることをどの程度予測できなかったか。

直観的に言えば、関連性と異質性とは裏腹な性質である。したがって、さまざまな情報に関して上記のような評価を加えた場合、一般にはその評価結果はほとんど図5-2に示す直線1に沿って分布し、一部は直線1の右上の領域に分布すると予想される。直線1の左下の、関連性も低くかつ異質性も低い（関連が弱いのに、そんな情報が提供されることが期待できた）ものというのは想定し難いので、この領域に分布する情報は少ないであろう。

そこで、直線1付近より右上の領域を図5-2に示すようにA～Dの四領域に分割し、これを前節で説明した図5-1と比較してみる。

図5-1の領域1に含まれる情報は、思考のテーマとの関連があり、そのような関連をその利用者が認識可能であるのみならず、思考のテーマが与えられた段階ですでにはほぼそのような情報を利用者が想起しているような情報である。したがってこの領域に含まれる情報に対しては、被験者は関連性が高く、異質性が低いと判断するであろう。よって、図5-1の領域1は図5-2の領域Aに対応する。

図5-1の領域2に含まれる情報は、思考のテーマとの関連があり、その様な関連をやはり利用者は認識可能である。しかし、思考のテーマが与えられた段階ではまだその関連性に思いが至っていない。したがって、そのような情報が提供された際に被験者はその関連に気づくので関連性を高く評価するが、一方そのような情報が提供されることを予測していなかったため、異質性も同時に高く評価することになる。よって、図5-1の領域2は図5-2の領域Bに対応する。

図5-1の領域3に含まれる情報は、思考のテーマとの関連性があるのだが、その関連に利用者は気づくことができない。したがって、そのような情報が提

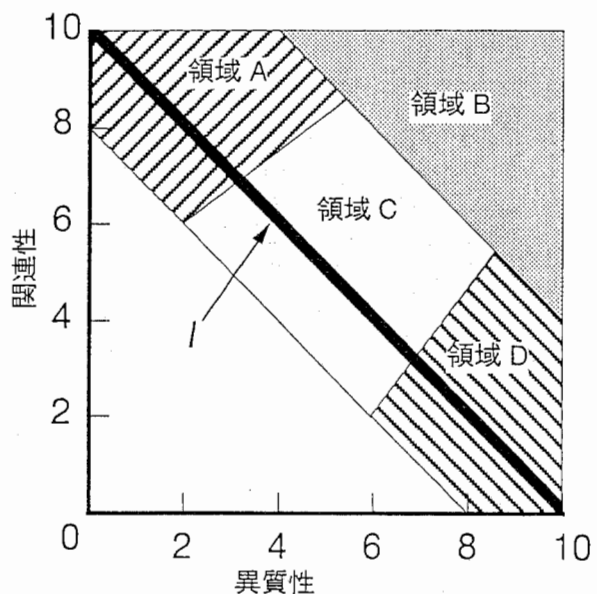


図5-2 主観的評価による記事の予想分布

供されても利用者は関連がわからないので関連性を低く評価するし、当然そのような情報が提供されることを期待してもいないから異質性を高く評価する結果となる。よって図5-1の領域3は図5-2の領域CからDに対応する。

図5-1の領域4に含まれる情報は、思考のテーマとなんの関連もないものである。したがって被験者はそのような情報に関連性を見いだすことは通常ないし、当然そのような情報が提供されることを期待することもない。つまり関連性はきわめて低く、異質性はきわめて高く評価するであろう。よって図5-1の領域4は図5-2の領域Dに対応する。

そこで、ある発散的思考支援システムが提供する情報を、思考のテーマとなんらかの（客観的）関連性があるものだけにすれば、当然図5-1の領域4に含まれる情報を除外することができる。つまりこの場合図5-2の領域Dに含まれる情報は、図5-1の領域3に含まれるものとみなすことができる。

このような方法で主観評価実験を行い、対象システムが提供する情報が図5-2の領域B,C,Dに分布すれば、そのシステムは発散的思考を刺激する効果のある情報を提供する機能があるものとみなせよう。領域Bに多く分布すればそのシステムは非常に即時的な効果をもつものと考えられる。また、領域C,Dに多く分布すれば、そのシステムは即時的な効果には乏しいが、その提供する一見あまり関係が無さそうな情報をじっくり考えたりあるいは調査したりすることによって、今まで知らなかったような関連に気づかせてくれる効果を持つであろう。

## 6 実験

### 6.1 基礎実験と評価[西本95]

まず門外漢モデルに基づく実験システムに対して、5章で述べた評価手法に基づき主観評価実験を行なう。本節で述べる実験は、門外漢モデルの中核部分である「浅い理解」の過程が、システムの抽出する情報の種類に与える影響を調べることを目的としている。

#### 6.1.1 実験手順と結果

この実験では、連想辞書と記事データベースの構築のために、CLARINET newsの記事を使用した。CLARINETとは、インターネット経由で配布される、米国の商用ネットワークニュースである。この記事を使用した理由は、商用の記事であるので品質が高い（つまりミススペルなどが少ない）こと、通常のネットワークニュースの記事に多くみられるような他記事からの引用がなく、各記事が個々に意味的に完結していること、である。実験で使用した記事の数は1465であった。各記事から生成されるサブセットリストには名詞と品詞同定不能であった語のみを採用した。ひとつのサブセットリストに採用される語の数は最大50とした。全記事処理して得られたサブセットリストに含まれた語の種類は全部で12637であった。すなわち、

$n = 1465$

$uwn = 50$

$m = 12637$

である。

実験では、まず以下に示す、思考のテーマ・利用者の興味・意見をシステムに入力する。

思考のテーマ：Computer market

利用者の興味：The future of Apple Macintosh

意見：I think Windows will be in place of the Macintosh, even if the Macintosh's performance is raised dramatically by the adoption of the Power PC chip.

なお、本節で紹介する実験では上記の思考のテーマ・利用者の興味・意見の区別は意味がなく、これら全部を一つのテキストとして単にシステムに入力するだけである。これらの区別は次節で述べる実験で使用される。

上記入力に対して、後で述べる二種類のアルゴリズムを使用して記事を抽出し、これを被験者に与えて前章で述べた手法に基づいて主観評価を行なってもらう。記事は一つのアルゴリズムあたり6つ、合計

12記事を抽出した。被験者は8人であり、各被験者に対して一つのアルゴリズムあたり3つずつ、計6つの記事を与え、それらについて関連性と異質性の面から評価してもらう。どの記事がどのようなアルゴリズムによって抽出されたかは被験者は知らされていない。記事はハードコピーとして被験者に渡される。

使用したアルゴリズムは以下の二種類である。

- (1) 門外漢モデルに基づくアルゴリズム
- (2) 直接類似度計算アルゴリズム

第二のアルゴリズムは、門外漢モデルから浅い理解過程を削除したものである。つまり第二のアルゴリズムでは意見から得られる意味表彰語群そのものと、各記事の意味表彰語群とから直接に類似度を求めることになる。このアルゴリズムによって抽出される記事を第一のアルゴリズムによって抽出される記事と比較することによって、抽出記事に異質性を取り込むために最も重要な過程と思われる「浅い理解過程」の効果がわかるであろう。

このようなアルゴリズムを用いて上記入力から記事を抽出した。アルゴリズム(1)と(2)とで同一の記事が抽出される可能性も当然あるが、今回の実験では同一記事は全く抽出されなかった。

被験者は提供された記事を読み、その内容と入力テキストとを比べて、その関連性と異質性をそれぞれ0から10の値で評価する。図6-1に評価の結果を示す。図中、同じ記号で示されるものは同じ記事に対する4人の被験者による評価結果である。ただし、例えば図6-1(1)の○で示される記事と、図6-1(2)の○で示される記事とは異なる記事である。

### 6.1.2 考察

図6-1(2)の結果を見ると、記事の評価結果はおおよそ図5-2の直線1にそって分布し、特に領域Cに多く分布していることがわかる。つまり、第二のアルゴリズムを使用しても、図5-1の領域3に含まれる情報を抽出できることがわかる。一方、第一のアルゴリズムの結果を見ると、多くの情報が図5-2の領域Dに分布している。したがって第一のアルゴリズムでも図5-1の領域3に含まれる情報が得られているが、第二のアルゴリズムに比べてその情報はより異質性が高いものとなっている。これは浅い理解によって視点をずらすという目的にかなった結果と言える。

第二のアルゴリズムでは、抽出される記事は入力された意見などに含まれていた語だけを使用して得られたものである。そして当然高い類似度を持つ記事にはそれらの多くが含まれている。したがって、高い類似度を持つ記事にはその関連性が非常に直接的な形で現れることになる。そして類似度が低下するにつれて、次第に関連性が低下すると同時に異質性が増していく。この結果、第二のアルゴリズムを

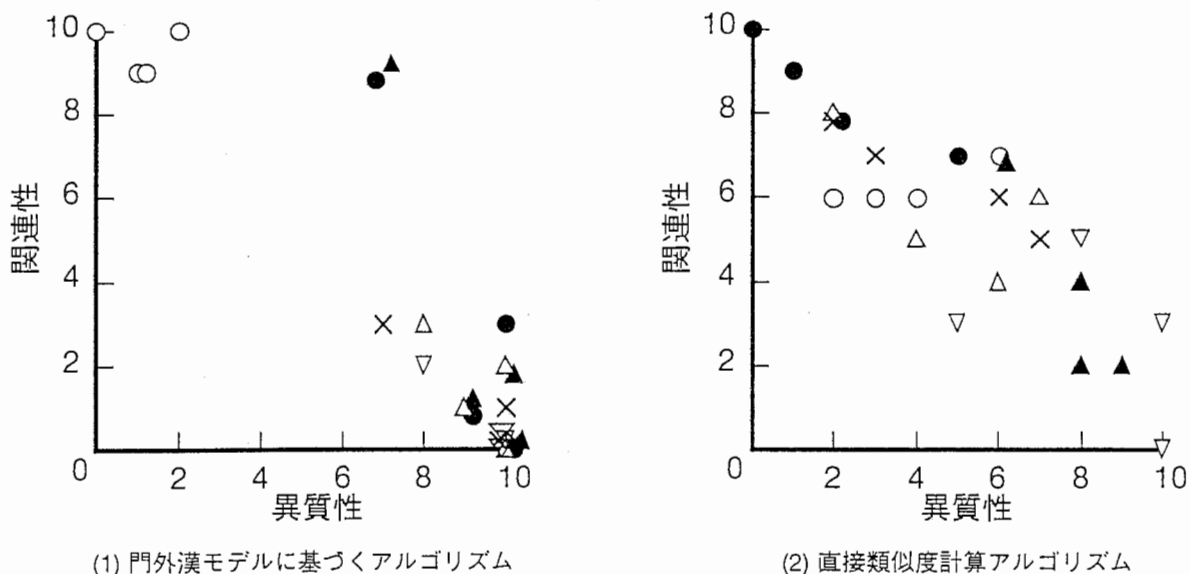


図6-1 門外漢モデルに基づく基本的構成のシステムによる「浅い理解過程」の影響の評価

用いると抽出された記事は図5-2の直線1に沿って分布するようになるものと思われる。よって図5-2の領域Bに含まれるような記事はこのアルゴリズムでは得ることが難しい。

一方、第一のアルゴリズムでは、記事は入力された意見に含まれていた語のみならず、それらの語から連想的に得られた語も用いて抽出される。したがって記事の関連性は間接的でやや曖昧なものとなるであろう。このため被験者は第一のアルゴリズムで得られる記事の関連性を見いだすことが第二のアルゴリズムの場合よりも難しくなる。したがって第一のアルゴリズムで得られる評価結果は、第二のアルゴリズムで得られるものよりも関連性が低く異質性が高いものとなる。ただし、すでに述べたようにこのプロトタイプシステムはなんらかの関連がある情報のみを抽出するように設計されているので、このような関連性が低く異質性の高い情報にも実際にはなんらかの関連性が含まれている。したがってこの図6-1(1)に見られるように評価結果が領域Dに集まっても、その情報は図5-1の領域3に含まれるものであって、領域4に含まれるものではない。ただその関連が間接的で隠れたものであるため、被験者には認識しがたいだけである。

もしある被験者がその情報が提供されたときにその隠れた関連に気づいたとすれば、その被験者はその情報が関連性が高いと評価すると同時に、そのような情報の関連があることにそれまで気づいていなかったもので異質性も高く評価するであろう。それゆえ、第一のアルゴリズムを使用すると図5-2の領域B、すなわち図5-1の領域2に含まれる情報をも抽出可能であると予想される。実際に、ある被験者によって●と▲で表示される二つの記事が図5-2の領域Bに含まれるものとして評価されている。つまりこの被験者は、その情報が提供されるまで見過ごしていたような隠れた関連性に、その情報が提供されたときに気づいたわけである。図6-1(1)に見られるように、同じ記事を他の被験者は全員図5-2領域Dに含まれるものと評価しているが、これらの被験者もこの二つの記事についてよく考え、そしてその隠れた関連性に気づくことができれば、速効性はないもののやはり発想には有益に作用するであろう。

以上の結果から、以下の結論を得た。

- (1) 門外漢モデルに基づく情報抽出手法によって発散的思考に有益な情報が得られる。
- (2) 「浅い理解」の作用によって、思考のテーマと関連性が低く異質性が高いような情報（図5-2領域C,D）だけでなく、関連性も異質性も同時に高いような情報（図5-2領域B）も得られることがわかった。

## 6.2 キーワードフィルタによる異質性の取り込み[西本93b]

前節ではもっとも単純な門外漢モデルの実装による実験を行なったが、本節ではさらに抽出する情報の異質性をコントロールする目的でキーワードフィルタをシステムに追加し、これを用いた被験者実験を行なう。

### 6.2.1 実験

本節で示す実験でも、使用した記事データは6.1節で述べた基礎実験で使用したものと同一である。キーワードベクトルや連想記憶行列の作り方も全く同様であるので詳細は割愛する。

なお、本実験における実験システムの構成は6.1節で用いた基本的な構成と若干異なり、図6-2に示すようにキーワードフィルタが追加され、これにともないキーワードベクトルの連想記憶モジュールに対するフィードバック経路も追加されている。キーワードフィルタは、以下に示す二種類を使用した。

#### (1) 差分フィルタ

連想記憶モジュールに入力された検索キーワードベクトルと、それから想起された想起ベクトルとの差

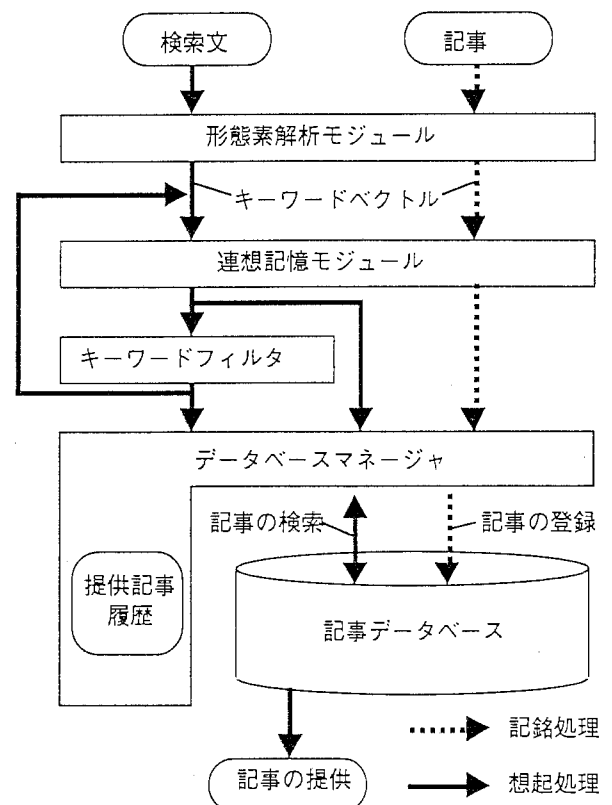


図6-2 キーワードフィルタを含む実験システムのソフトウェア構成

分から得られる差分ベクトルを求めるフィルタ。最初、被験者が入力したテキストから得られた検索キーワードベクトルを $Q_0$ とすると、提供済みキーワードベクトル $P_i$ を、

$$P_0 = Q_0$$

$$P_i = Q_i + P_{i+1} \quad (1 \leq i)$$

と定義する。ここで+はベクトルの要素毎の論理和を示す。この時、 $i+1$ 番目の検索キーワードベクトル $Q_{i+1}$ を次の式で与える。

$$Q_{i+1} = K_i * \{!(P_i)\} + T$$

ここに $K_i$ は $Q_i$ による検索で得られた記事のキーワードベクトル、 $T$ は最初に被験者が設定した思考のテーマと興味だけから生成したキーワードベクトルである。また、演算子 $*$ はベクトルの要素毎の論理積、 $!$ はベクトルの要素毎の否定である。この $Q_{i+1}$ を用いて連想想起・検索を行い $i+1$ 番目の記事を抽出する。

つまりこのフィルタでは、過去に検索キーワードベクトルに使用した単語は、思考のテーマならびに興味として設定された単語以外は二度と使用されない。このフィルタリングによって、ある記事から、その記事に現れた新しい単語をトリガとして新たな話題へと抽出される記事が早く変化していくことを期待している。

## (2) 弱連想フィルタ

想起されたキーワードから、強く想起された部分を捨てて、弱く想起された部分を混ぜるフィルタである。すなわち、通常のアソシエトロンにおける $\phi_{\theta}$ 量子化関数を、以下の様な二つの数居値を用いた量子化関数 $\phi_{\theta_w}$ に置き換える。

$$\phi_{\theta_w}(x) = \begin{cases} 0; & x \geq \theta_U \\ 1; & \theta_U > x > \theta_L \\ 0; & \theta_L \geq x \geq -\theta_L \\ -1; & -\theta_L > x > -\theta_U \\ 0; & x \leq -\theta_U \end{cases}$$

この数居値 $\theta_L$ 、 $\theta_U$ の値を変化させることによって弱さの程度が可変となる。

以下、実験の詳細を述べる。今回の応用実験では5名の被験者にまず個々に任意の思考のテーマ・興味・テーマに対する意見を入力してもらい、これから以下で説明する5種類7通りのアルゴリズムによって各アルゴリズム毎に6記事ずつ抽出し、これらの記事に対して5章で述べた手法に基づいた主観評価を行なった。

使用した記事抽出アルゴリズムは以下の通りである。

### (1) 単純連想想起

6.1節の実験での第一のアルゴリズムと同じ。

### (2) 連鎖連想想起

差分フィルタを使用したアルゴリズム。したがって、第一の記事は(1)の単純連想想起で得られるものと同じであるが、第二の記事以後は差分ベクトルを使用して獲得された記事となる。

### (3) 弱連想想起

弱連想フィルタを使用したアルゴリズム。以下の3通りの $\theta_L$ ・ $\theta_U$ の組み合わせを使用した。

(a) 強く想起された要素3つをリセットし、弱く想起された要素3つをセット

(b) 強く想起された要素6つをリセットし、弱く想起された要素6つをセット

(c) 強く想起された要素12個をリセットし、弱く想起された要素12個をセット

### (4) ランダム抽出

入力された意見などに関連なく、記事データベース中の全記事からまったくランダムに（ただし重複はないように）6つの記事を取り出すアルゴリズム。比較評価用。

### (5) 直接相互関連

6.1節の実験における第二のアルゴリズムと全く同じ。検索ベクトルの連想拡張を行なわない。

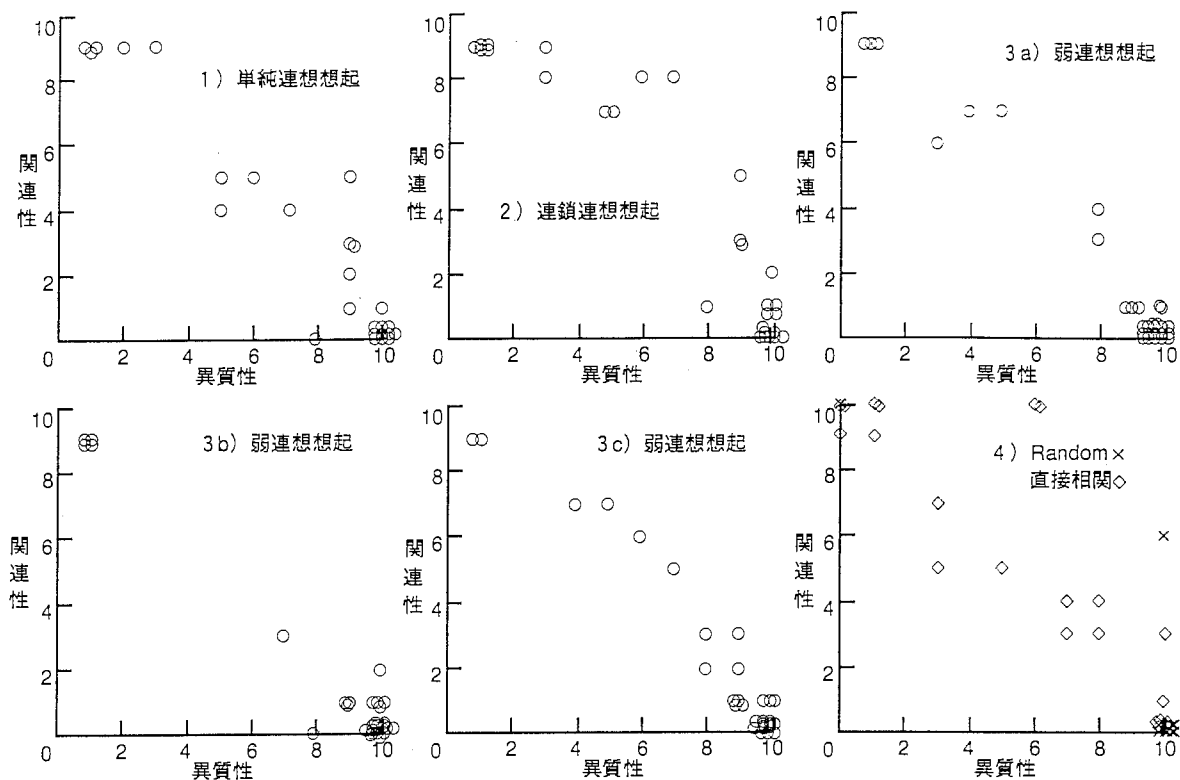


図6-3 各種キーワードフィルタを使用した場合のアルゴリズム別評価結果

各アルゴリズムにおける記事抽出が終わるごとに記事提供履歴を初期化する。したがって、同じ記事が異なるアルゴリズムによって抽出されることはありうる。

このようにして記事を被験者に提供して主観評価を行なってもらう。なお、各記事がどのようなアルゴリズムによって抽出されたかは被験者には知らされない。また、本実験では英文記事を大量に読んで評価せねばならないので、被験者には外国人2名を含む特に英語力のある者を選んだ。

各被験者が設定したテーマ・興味・意見を表6-1に示す。また主観評価の結果を各アルゴリズム毎にまとめたものを図6-3に、各被験者毎にまとめたものを図6-4に示す。

## 6.2.2 考察

まず図6-3の、アルゴリズム毎の結果を見る。4)のグラフに示したランダム抽出の結果を見ると、二つの記事を除いてすべて関連性0、異質性10と評価されている。ランダム抽出の場合、抽出される記事に力した意見などの関連は全く考慮されていないのでこの結果は当然であり、おそらくこれらの記事はすべて本当に無関連な、図5-1の領域4に含まれるものである。関連性を含む二つの記事は偶然に抽出されたものであり、うち一つは関連性10、異質性0で発散的思考支援としての寄与は期待できず、唯一関連性6、異質性10と評価された記事のみが発散的思考支援に有効に作用するものと思われる。このように、ランダム抽出は発散的思考支援のための手段としては効率が悪いことが示された。

その他の結果についてみると、いずれの結果も基本的には図5-2の直線1に沿った分布を示している。このうち、グラフ4)の直接相関の結果を見ると、やはり関連性10の評価を与えられた記事も多く、6.1の結果同様連想想起を用いた場合よりも関連性が直接的に見えるものが多く抽出されていることがわかる。

次に連想想起を用いた1)~3c)の結果を見る。

まず、3a)~3c)の弱連想を用いた場合がやはり1)、2)の結果に比べて図5-2の領域Dに分布する記事が多く、予期通り弱連想フィルタの作用によって関連度が下がっていることがわかる。ただし、本来3a)、3b)、3c)の順に関連度が下がると予想していたのだが、実際には3b)の場合が最も領域Dに多く分布する結果となった。この原因はよくわからない。

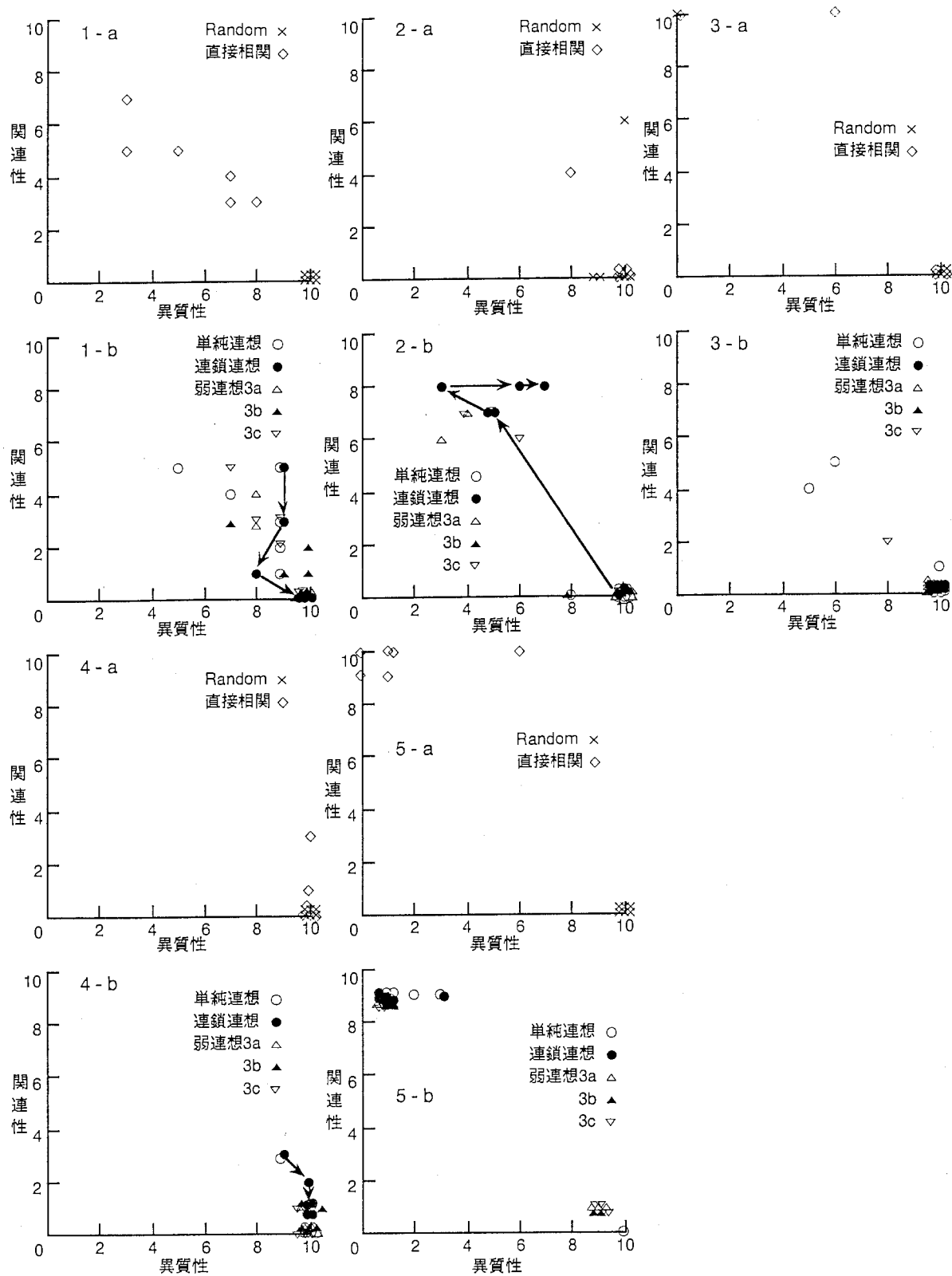


図6-4 各種キーワードフィルタを使用した場合の被験者別評価結果

1)の単純連想の場合図5-2の領域CからDにかけて分布しており、基本的な分布傾向は6.1節での実験の第一のアルゴリズムの場合と同じである。ただし今回は領域Bに含まれるような記事は得られなかった。  
 2)の連鎖連想の結果は今回の実験ではもっともよい分布傾向を示しており、全般に図5-2領域B方向に膨らんだ分布となっている。

被験者1	テーマ	Personal Computer
	興味	Network System
	意見	In near future, All PC's should be connected via network.
被験者2	テーマ	Recreation
	興味	idea, quiz, prize, money
	意見	The best recreation is stimulating my brain and to make me rich.
被験者3	テーマ	World Politics
	興味	NAFTA, Latin America
	意見	The U.S. has always thought that Latin America is its own back yard. This stance has caused that many countries south of the border keep resentful.
被験者4	テーマ	Outdoor life
	興味	kayak, canoe, log cabin, camping, tour
	意見	kayaking tour is very light and exciting.
被験者5	テーマ	Basketball
	興味	Phoenix Suns Charles Barkley win championship fans
	意見	I think that the Phoenix Suns will win the NBA championship this year. I think that most of the fans in Phoenix like Charles Barkley and the enix Suns.

表6-1 各被験者の設定したテーマ／興味／意見

かも分野的にも技術的な内容のものから経済関係のものまで比較的幅広く分布していた。そこで評価結果を見ると、まずランダム検索ではこのように比較的多量の関連記事が含まれている場合でも全く無関連の記事のみが抽出されておりその効率の悪さが明確に現れている。直接相関と単純連想の評価結果はほぼ6.1節の評価結果と同様の傾向を示し、直接相関では図5-2の直線Iに沿ってほぼ領域Cに分布しているのに対し、単純連想ではより異質性が高まり領域Dに主に分布し、しかも異質性が9と高いわりには関連性も5と比較的高いような記事も獲得できている。また、連鎖連想の結果を見ると、最初の記事が最も関連性が高く、以後順次関連性の低い方向に評価が移動しているのがわかる。これは差分フィルタによる異質性の導入の効果と思われる。これらの結果から、記事データベース中にまんべんなく分布するような語を多く含むテーマ・意見などが設定された場合にはこのシステムは期待する機能を発揮すると言える。ただし、弱連想については、単純連想の場合よりはやや異質性の高まる傾向があるものの、明確に単純連想よりも異質性が高まっているとは言い難い結果となっている。これは、記事がまんべんなく分布する結果、多少弱く想起される語も依然比較的強い関連性のあるものばかりであり、12個程度語の入れ替えを行っても全体的な傾向は変化しないためではないかと思われる。

次に被験者2の結果2-a/bについて見る。被験者2の設定したテーマや意見に含まれる語は記事データベース中のほとんどの記事に含まれていないものであった。特に意味的に重要と思われるrecreationやquiz,brainなどの語がほとんどの記事に含まれていなかった。この様な場合は直接相互相関では当然関連のある記事は得ることができず、しかも連想想起を用いた場合でも単純連想ではmoneyなどの語が支配的な作用をし、娯楽からは大きく外れた分野の記事ばかりが抽出される結果となった。この結果単純連想の抽出結果は、たしかに「金」という関連はあるものの被験者の思いとはまるで無関連と評価される結果となった。

ところが、連鎖連想や弱連想では逆に関連性を含んだ記事が抽出されている。連鎖連想では最初の記事は単純連想と同じ記事が抽出されるので関連性0、異質性10のものだが、そこからグラフに見られるように一気に関連性7、異質性5へ移動し、以後比較的高い関連性を保ちつついったん異質性が下がった後次第に異質性が高い方向に動いている。これは最初の記事で導入された新しい語が、テーマと興味に設

以上の結果から、ランダム検索以外の手法はすべて発散的思考を支援できる可能性があること、弱連想フィルタを使用することで単純連想より異質性が高い情報の抽出が可能であること、差分フィルタを用いることによって関連性と異質性が同時に高いような情報の抽出の可能性があることが示された。

なお連鎖連想の場合は他のアルゴリズムと異なり、前に抽出された記事の内容によって次に抽出される記事の内容が影響を受ける。つまり、連鎖連想の場合には単に分布を見るだけでなく、抽出記事に対する評価が順にどのように変化していったかを見る必要がある。そこで、この点も含めて、以下では図6-4の被験者毎の評価結果について考察する。

図6-4の1-a/bの結果が、本システムの最も一般的な情報抽出傾向を示していると思われる。この被験者1の設定したテーマや意見はパーソナルコンピュータやネットワークにかかわるものであり、記事データベース作成に使用した記事群にはこのような語を含む記事が多くあり、し



定されていた語とうまく作用しあってテーマとの関連性を持つ記事を抽出したものと思われる。また、弱連想によっても関連性7、異質性5付近の記事が抽出されている。つまり、単純連想ではmoneyのようなかなり多数の記事に含まれる語のが想起に強く影響し、recreationなどの出現頻度の低い語によって想起される語が最終的な想起に採用されることを妨害しているのだが、弱連想によって強く想起された部分を削除することによってこのような出現頻度の低い語によって想起される語が取り込まれ、テーマとの関連性が導かれたのであろう。つまりこの被験者2の場合のように設定されたテーマなどに含まれる重要な語の記事データベース中における出現頻度が低く、あまりテーマとの関連が密接ではないが出現頻度の高い語が想起を支配してしまうような場合には、差分フィルタや弱連想フィルタなどの、本来異質性を導入するために取り入れたメカニズムが、逆に関連性を導く機能を持つものとなりうることがわかった。したがって、設定されたテーマや興味などに含まれる語に対して事前に出現頻度を調査して、それが全体の平均的レベルから見て非常に低いような場合には差分フィルタや弱連想フィルタを用いて関連性を導入することができるようになる。ただし、差分フィルタによる関連の導入は、抽出された記事の内容に依存するので偶然に頼っている面がある。よって、このような場合には弱連想フィルタを使用した方が良いと考えられる。

ある語の出現頻度に相当する情報は連想記憶行列上に保持されている。まず単純にある語 $w_i$ が、記録したいくつの記事中で使用されたかについては、連想記憶行列Mの要素 $a_{ij}$ の値がそのままこれに対応する。またアソシアトロン的手法では単なる出現頻度の他に、いかに多くの他の要素と共起したことがあるかも影響する。たとえば語 $w_i$ が他のいくつの要素と共起しているかは、i行中の要素で0でない要素が幾つあるかを数えればよい。このような情報を用いれば、入力されたあるテーマや意見の場合に適切なフィルタをシステムが自動的に選び、効率よく情報の抽出を行えるようになるであろう。

しかし、被験者2の場合よりも、設定したテーマや意見に含まれる（特にそれらの内容と直接的に関連のある）語が記事データベース中の記事にさらにわずかしが含まれていない場合には当然上記の機能もうまく作用しなくなる。被験者4の結果がその場合に当たる。被験者4の設定したテーマなどのうち、比較的多く記事に含まれていた語はlifeであるが、この語だけでは被験者4の意図とは間接的関連しかない。被験者4にとって重要なのはOutdoorやkayak, canoeなどの語であったが、outdoorは形容詞であるため採用されず、またkayak, canoeは記事中にいっさい含まれていなかった。このため、どんなアルゴリズムを使用しても抽出結果はほぼlifeからの連想に依存したものとなり、かろうじてtourに関連するような記事に被験者はわずかな関連性を見いだしてはいるが、全般には被験者の意図からは大きく外れたものとなり、ほとんどの記事が関連性0・異質性10と評価される結果となった。このような場合にはなんらかのシソーラス辞書、あるいは国語辞典などから構成したいわゆる「常識」にあたるような連想辞書などを用いて、入力されたテーマや意見に対して前処理を行なう必要があるだろう。

被験者5の場合は、NBAの試合結果記事はかなり多数記事データベース中に含まれていた。そして、それらの記事は、被験者の設定したテーマや意見に含まれる、それらの内容に直接かかわる語を多数含んでいた。このため5-aの結果に見られるように、直接相関はもちろんのこと、単純相関においてもいわゆるデータベース検索を行なったかのように関連性が高く異質性が低い記事しか得られなかった。つまりこのような状況では連想辞書による拡張を行なってもそれらの記事の範囲を逸脱したような方向への拡張が困難なのである。この結果、連鎖連想を行なっても、毎回得られる記事がNBAの結果の記事となってしまうためやはり異質性の高い方向への脱出ができない。このような場合には弱連想が有効であり、単純連想・連鎖連想では得られなかった異質性を弱連想によって獲得することができていることが5-bの結果からわかる。ここで関連性1・異質性9付近に集まっている記事はいずれもホッケーリーグの試合結果記事であった。つまり、弱連想によってbasketballやNBAなどの強い影響力を持つ語の束縛を脱し、championshipや、あるいはbasketballなどの語から連想されるsportsあるいは直接的にhockeyなどの語によってこれらの記事が得られたと考えられる。

ただし、ここで主観評価の評価基準に関する問題点が一つ表面化している。5-bの結果を見ると、評価が完全に二極に分離している。おそらくこの被験者は設定したバスケットボールというテーマに対してかなり強いこだわりを持っていたと思われる。このため、バスケットボール関連記事以外の記事が提示されるとそれはすべてほとんど無関連なものと評価したのであろう。この結果、スポーツあるいは試合結果という点では関連の認められるホッケーの記事を関連性1・異質性9評価しているが、人によっては

これらをもう少し関連性が高く異質性が低いものと評価する可能性も十分に考えられる。つまり設定したテーマや意見に対するこだわりの強さ次第でこの評価が変わる。

被験者3の場合も同様の状況が見られる。被験者3の評価結果も基本的には直接関連で関連性が高く、単純連想、連鎖連想・弱連想と異質性が高くなっている。しかし、被験者1や2と比べれば全体に評価は二極分解している。この被験者の場合NAFTAに対するこだわりが強く、直接NAFTAに関連する記事には高い関連性を与えているが(3-aの直接関連やランダムで得られている)、3-bの結果に見られるように、NAFTAから少しでも外れた内容の記事に対する関連性の評価は一気に低くなり、テーマで設定された world politics の範囲に入るが、NAFTAやLatin Americaと関連の無い記事に対する評価は関連性0・異質性10となっている。

今回の実験では被験者に先入観を与えたくなかったため、「発散的思考」などの条件を予め与えずに実験を行なった。この結果、被験者1,2のように関連性を柔軟に考えた例もあれば、被験者3,5のように関連性をかなり狭い範囲の意味でとらえた例も出た。一般に発散的思考を行なう場合には関連性は柔軟に考えた方が良い。したがって、今後この手法に基づく主観評価実験を行なう場合には、「あなたはいまブレインストーミングを行なっています。この時このテーマのもとである人がこんな情報を提供しました。」というような条件付けを行なうべきなのかもしれない。このように、主観評価の手法についてはさらに実験条件などを考慮する必要がある。

以上の実験結果ならびに考察から、以下の結論を得た。

- (1-1) システム利用者の入力した意見などに含まれる語が記事データベース中の記事に比較的多量に、しかも一部の記事に集中することなく多くの記事に分散して含まれる状況では単純連想想起で発散的思考支援に有効な記事を獲得できる。
- (1-2) (1-1)の状況では、差分フィルタは期待どおり異質性を取り込むのに有効に作用する。
- (2-1) システムの利用者の入力した意見などに含まれる語が記事データベース中の記事にあまり含まれていない場合には直接関連・単純連想のいずれによっても発散的思考に効果的な記事を獲得することは難しい。
- (2-2) (2-1)の状況では、差分フィルタや弱連想フィルタが逆に関連性を取り込む作用を持ち、この結果これらのフィルタによって発散的思考に効果的な記事を獲得することが可能となる。
- (3-1) (2)の状況よりもさらにシステム利用者が入力した意見などに含まれる語が記事データベース中の記事に含まれていないような場合には、差分フィルタや弱連想フィルタも有効に作用しない。
- (3-2) (3-1)の状況では、外部から与えたシソーラス、あるいは別に準備した連想辞書などを使用して、前段階の連想拡張処理を行なう必要がある。
- (4-1) (2)や(3)とは逆に、システム利用者が入力した意見に含まれる語が非常に多く一部の記事に集中して含まれる場合は、直接関連はもちろん、単純相関でも、また差分フィルタを使用しても関連性が非常に高い記事しか得られず、この場合も発散的思考支援に効果的な記事が得難い。
- (4-2) (4-1)の状況では、弱連想フィルタが所期の期待通りに異質性を取り込むのに有効に作用し、この結果発散的思考に効果的な記事を得ることが可能となる。
- (5) 以上のような、利用者の入力情報とシステムの連想記憶辞書との関連の状態は、連想記憶行列の出現頻度・リンク数を調べることで知ることができるので、状況に応じて適切なフィルタ・連想想起手法を使用して常に発散的思考支援に効果的な記事を得ることができるシステムの構成が可能である。
- (6) 主観評価手法については、被験者に与える実験条件にもう少し「発散的思考過程」を意識させる制約条件を付け加える必要があるかもしれない。

### 6.3 異分野専門家としての応用実験[西本94]

本節では、複数の連想辞書を組み合わせて使用し、本システムを単純な門外漢ではなく、異分野の専門家として応用することを試みる。

#### 6.3.1 実験

実験では二種類の連想辞書を準備した。第一の連想辞書は「一般連想辞書」である。これは前の二つの実験で使用した連想辞書と同じものである。第二の連想辞書は、「専門連想辞書」である。今回はこ

## 自由連想

## 状況想定連想

製品責任 大企業、PL法、品質管理	製品責任 不法行為
医用情報処理システム 大病院、白衣、受付、診察料金、カルテ	医用情報処理システム プライバシー

図6-5 専門連想辞書構築のためのアンケート例

の専門分野として、法律・政治分野を設定することにし、京都大学法学部の教職員・学生あわせて30名の皆様のご協力を得て、アンケートを行なってその結果から専門連想辞書を構築した。

アンケートはこちらから与えた語（以下これをseed語と呼ぶ）に対して、

- (1) 全く自由に連想して思いつく語（句）を列挙してもらう
  - (2) 「あなたはコンピュータの専門家数名と、コンピュータと法律の関係に関する議論をしています。あなたには、法律に詳しい人としての立場からの意見を期待されています。」という状況を想定してもらってさらに思いつく語を列挙してもらう
- という二段階の連想を行なってもらった。図6-5にアンケートの例を示す。図中、下線を引いた語がseed語、左覧が第一の自由連想の想起結果、右覧が第二の条件付連想での想起結果である。seed語には以下のような語を選んだ。
- (a) 以下で述べる実験に用いる、システムに入力するテーマ・意見などからシステムが抽出する語  
例：Windows, Macintosh, PowerPCなど
  - (b) JISCT科学技術用語シソーラスから抽出した、コンピュータやその関連技術と関係する種々の階層の語  
例：計算機利用システム、知識ベースシステム、意思決定支援システム
  - (c) JISCT科学技術用語シソーラスから抽出した、法律と関係する種々の階層の語  
例：法規、情報政策、知的所有権

このようにして全部で114語をseed語として選びだし、このうちから一人38語を選んで与えた。

得られた回答に対し、まずseed語および連想語をすべて英訳する。ついで、各回答者毎に一つのseed語とそのseed語に対する回答をそれぞれ個々に一つの記事とし、これをすでに前節までの実験で述べたのと同様の方法で連想記録して専門連想辞書を構築した。

また、実験で使用した記事データベースは「一般連想辞書」構築に使用した記事のみによって構成されたもの（つまり6.1.6.2の実験で用いたものと全く同じもの）である。

このようにして構築した二種類の連想辞書および記事データベースを用いて想起実験を行なう。想起実験では、システムに入力するテーマ・興味・意見は6.1の実験で使用したものと全く同じであるが、ここに再掲する。

思考のテーマ：Computer market  
 利用者の興味：The future of Apple Macintosh  
 意見：I think Windows will be in place of the Macintosh, even if the Macintosh's performance is raised dramatically by the adoption of the Power PC chip.

このテキストを構文解析し、検索ベクトルQを6.1節の実験と全く同様の方法で生成する。このベクトルを用いて、以下の四種類のアルゴリズムを用いて想起を行なった。図6-6に各アルゴリズムの概要を示す。

- (1) 検索ベクトルQと専門連想辞書Jを用いて想起を行い、想起ベクトル $R_i$ を得る。この $R_i$ を用いて記事データベースを検索する。（図6-6(1)）
- (2) 検索ベクトルQと専門連想辞書Jを用いて第一段目の想起を行い、想起ベクトル $R_i$ を得る。この $R_i$

をそのまま第二段目の検索ベクトル $Q'$ とし、一般連想辞書 $U$ を用いて想起を行い、想起ベクトル $R_2$ を得る。この $R_2$ を用いて記事データベースを検索する。(図6-6(2))

(3) 検索ベクトル $Q$ と専門連想辞書 $J$ を用いて第一段目の想起を行い、想起ベクトル $R_1$ を得る。この $R_1$ と $Q$ を要素毎に論理和をとって得られる合成ベクトル $Q_{merge}$ を第二段目の検索ベクトルとし、一般連想辞書 $U$ を用いて想起を行い、想起ベクトル $R_3$ を得る。この $R_3$ を用いて記事データベースを検索する。(図6-6(3))

(4) 検索ベクトル $Q$ と一般連想辞書 $U$ を用いて想起を行ない、想起ベクトル $R_4$ を得る。この $R_4$ を用いて記事データベースを検索する。(図6-6(4))つまりこのアルゴリズムは6.1節の第一のアルゴリズムと同じである。

検索ベクトル $Q$ 、および想起ベクトル $R_1 \sim R_4$ で値1がセットされた要素に対応する語を図6-7に示す。また各アルゴリズムで得られた記事の内容の要約を図6-8に示す。図6-8の括弧内の数字は想起ベクトルと記事ベクトルの相互相関の値である。

### 6.3.2 考察

第1のアルゴリズムでは、与えられた検索文のうちから設定された専門の範囲にある語に関しては連想を幅広く行う一方、範囲外の語に関しては連想はほとんど行わず、専門連想辞書にまったく含まれない語に関しては当然なんの連想も行わない。この結果想起されるベクトル $R_1$ は専門範囲に関しては充実し、専門外に関しては非常に貧弱な構成となる。図6-7(1)の語群を見ればわかるように、もともと検索文に多く含まれていたコンピュータ用語のうち、PowerPCやchipなどの多少専門的な語が消滅し、一方でディスプレイやPCと言った、コンピュータにあまり詳しくない人がコンピュータと言われてとりあえず思いつくような語が増えている。これに対し、検索ベクトル $Q$ にはmarket程度しかなかった社会的分野の語が、想起ベクトル $R_1$ ではcompany, trade, firm, law, corporationと増強されている。このような $R_1$ をそのまま用いて各記事ベクトルとの相互相関をとれば、設定された専門分野方向の記事が上位に現われるであろう。

第2のアルゴリズムでは、 $R_1$ を一般連想辞書によって再想起するため、第1のアルゴリズムよりは専門的色彩は弱くなると思われる。

第3のアルゴリズムでは、検索文に含まれていた語のうち、専門辞書による連想で失われた専門外の語が補充され、これを用いて一般連想辞書で再想起している。このため、第2のアルゴリズムよりもさらに専門的色彩は弱く、一方で検索文との直接的な関連性がより高い記事が上位に現われてくるだろう。

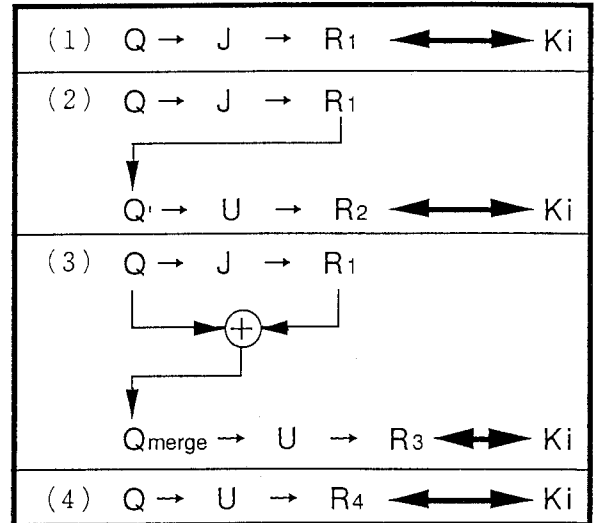


図6-6 複数連想辞書の組み合わせアルゴリズム

(0)検索ベクトル $Q$ think, window, place, Macintosh, performance, dramatically, adoption, PowerPC, chip, computer, market, future, apple
(1)第1のアルゴリズムで想起されたベクトル $R_1$ company, trade, firm, information, computer, law, apple, Macintosh, display, corporation, protection, convenience, window, PC, picture, Panasonic, CRT
(2)第2のアルゴリズムで想起されたベクトル $R_2$ world, all, three, two, company, year, week, month, California, business, time, can, May, product, while, john, make, Thursday, line, program, United States, software, Nov, NB, window
(3)第3のアルゴリズムで想起されたベクトル $R_3$ world, first, all, three, two, one, Texas, company, million, system, Washington, market, year, technology, use, week, president, month, California, business, plan, board, time, can, development, May, number, report, product, while, john, make, Thursday, line, program, United States, computer, chip, need, office, software, people, NOV, NB, window
(4)第4のアルゴリズムで想起されたベクトル $R_4$ world, all, three, one, company, million, system, Inc, market, year, technology, use, week, president, run, business, plan, board, time, number, product, make, Thursday, United States, computer, price, saying, office, software, support, NOV, NB

図6-7 検索ベクトルおよび各アルゴリズムによって想起されたキーワードベクトル

<p>(1) 第1のアルゴリズムによって抽出された記事</p> <p>1-1 (0.0576) 過去に倒産し、bankruptcy法の適用を受けたCompuadd Computer社が、環境保護局のガイドラインに沿った省エネルギー型コンピュータの出荷を開始すると発表。同社は製品のコンピュータ全てにMicroSoft Windows Ver.3.11を登載する。</p> <p>1-2 (0.0423) Compuadd Computer社はbankruptcy法の適用下から脱し、PCメーカーのZeos Internationalまたはその他の会社と合併する。</p> <p>1-3 (0.0423) Computer Technicaはパソコン用カラーTVスクリーンコンバータを開発した。これは通常のTVをコンピュータディスプレイとしても使用できるようにするもの。</p> <p>1-4 (0.0294) IDG World Expo Japanは、来年のMacworldは過去最高規模のものになると発表</p> <p>1-5 (0.0294) Microsoftは、Motorolaと協力して、Windows NTとPowerPCアーキテクチャを結合するためのポート開発を開始すると発表。</p>
<p>(2) 第2のアルゴリズムによって抽出された記事</p> <p>2-1 (0.0967) National SemiconductorとNovellはNECと提携すると発表。日本により有用な情報ハイウェイを構築し、日本にtelecommutingをもたらすことを目指す。</p> <p>2-2 (0.0967) AT&amp;T Paradyneは、Las VegasのFall COMDEXで予定していたVoicespanに関する発表を延期した。Voicespanとは、1本の電話回線を2つの仮想回線に分割し、音声と同時に画像やデータをやりとりする技術。</p> <p>2-3 (0.0967) 1-1と同じ</p> <p>2-4 (0.0799) 1-5と同じ</p> <p>2-5 (0.0799) ビデオ教材の「C++の世界」の続編、「C++の世界を超えて」の紹介。</p>
<p>(3) 第3のアルゴリズムによって抽出された記事</p> <p>3-1 (0.1137) すでに政府の書類やホワイトハウスのプレスリリースなどはインターネット経由でアクセス可能であり、多くの人はその利用法を知らないが、インターネットのマニア達はその方法を知って利用している。</p> <p>3-2 (0.1137) 国際的な株式市場の急落傾向にもかかわらず、雇用は予想外の上昇を示している。</p> <p>3-3 (0.1137) 1-5と同じ</p> <p>3-4 (0.1137) National SemiconductorとNECは日本のコンピュータネットワーク市場の拡大のための提携を結んだことを発表。</p> <p>3-5 (0.1000) 2-1と同じ</p>
<p>(4) 第4のアルゴリズムによって抽出された記事</p> <p>4-1 (0.1406) 1-5と同じ</p> <p>4-2 (0.1406) Novell Inc.はRaymond J. Noorda氏の後任のネットワークソフトベンダーのチーフエグゼクティブオフィサーを探していることを確認。</p> <p>4-3 (0.1056) ゴア副大統領は、コンピュータデータストレージシステムを生産しているハイテク企業Storage TechnologyでNAFTAへの協力を呼びかけた。</p> <p>4-4 (0.1056) 2-1と同じ</p> <p>4-5 (0.1056) ネットワーク製品メーカーのCISCOが発表した1993年第1四半期のネットワーク製品、特にインターネットワーキング製品の売上と収益の金額。</p>

図6-8 各アルゴリズムで抽出された上位5記事の内容の要約

第4のアルゴリズムは従来から試みているものと同じであり、ここでは専門連想辞書は使用されないため、基本的にもっとも検索文の内容との直接的な関連性が高い記事が上位に現われると思われる。

つまり、各アルゴリズムによって抽出される記事は、第1のアルゴリズムでもっとも法律的色彩が濃く、第2、第3、第4と次第に法律的色彩が弱まって行き、逆に検索文との直接的関連性は第1のアルゴリズムでもっとも弱く、第2、第3、第4と次第に直接的関連性が強くなると期待される。

図6-8を見ると、どのアルゴリズムによって抽出された記事も、3-2の記事を除いて、すべてコンピュータにかかわるものとなっている。3-2だけがコンピュータと関係なく株式市場の話題となっているが、抽出された記事はすべて基本的にcomputer marketというテーマに沿ったものになっていると言える。これは、図6-7(1)に見られるように、第1のアルゴリズムにおいてもコンピュータ関係のキーワードは完全には消滅しないためである。

ただし、検索文との直接的な関連性の強さはアルゴリズムによって変化する。図6-8に示した1-5の記事は、すべてのアルゴリズムで抽出されている。この内容はここに示した他のどの記事よりも直接的に検索文の内容と関連していると言える。この結果、第4、第3のアルゴリズムではこの記事は最も強い相関で抽出されているが、第2のアルゴリズムでは2位の相関、第1のアルゴリズムでは3位の相関となっている。このように、第1、第2のアルゴリズムでは抽出される記事と検索文との直接的な関連は弱まっている。

一方、第1のアルゴリズムでは1-1および1-2の記事に見られるように、法律との関連が強い記事が上位

に現われる。これら2つの記事のうち、1-1の記事は第2のアルゴリズムでも1位の相関で抽出されているが、1-2は上位から消えている。さらに第3のアルゴリズムになると1-1の記事も上位から消える。このように、抽出される記事の、設定した専門分野との関連性も、予想通り第1のアルゴリズムで最も強く、第2、第3と次第に弱まっ行く。

以上の結果から、複数の連想辞書を用意し、これらを組み合わせて使うことによって抽出される情報の分野の粗い制御が可能となると言えよう。

次に、想起されたベクトル $R_2$ 、 $R_3$ 、 $R_4$ の3つを比べてみる。すると、 $R_2$ 、 $R_4$ には現われない語が $R_3$ には現われていることがわかる。図6-7(3)に下線付で示したfirst, Texas, Washington, development, report, chip, need, peopleがそれである。

$Q_{merge}$ は、 $Q$ と $R_1$ との論理和である。したがって第2のアルゴリズムと第4のアルゴリズムとを単純に組み合わせれば、一般連想辞書による想起の段階に投入された語は第3のアルゴリズムと同じになる。しかし、 $Q_{merge}$ によって想起されたベクトル $R_3$ は $R_2$ と $R_4$ との論理和にはなっておらず、 $R_2$ と $R_4$ のいずれにも現われない語を含むものとなる。これらは、専門連想辞書を用いた拡張によって追加された語と、元の検索ベクトルに含まれた語との組み合わせによって初めて強く連想されて現れてきたものである。

つまり、このような語は、第2のアルゴリズムでは消滅してしまったコンピュータに関するやや専門的な領域の知識と、第4のアルゴリズムでは存在しなかった法律的な分野に関する専門的な知識の両方にまたがるような知識を持つものでなければ想起できないような語であると言える。人間に例えれば、第3のアルゴリズムは法律にも詳しいと同時にコンピュータにもそれなりに詳しい者にあたると言える。

図6-8(3)の検索結果を見ると、第3のアルゴリズムで得られた検索結果の内、もっとも強く想起された記事のうち3つ(3-1, 3-2, 3-4)が他のアルゴリズムの検索では上位に現われないものである。特に3-1の記事の内容は、法律的な分野とコンピュータの両分野にまたがりかつ意外性を感じさせるものになっていると思われる。

このように、複数の連想辞書を使用することによって複数分野に詳しいような門外漢を構成でき、しかも分野Aだけに精通する専門家と分野Bだけに精通する専門家が二人集まっても得難いような、A、B両分野にまたがる専門知識を持つ人ならではの情報を取り出せる可能性が示唆された。

### 6.3.3 連想辞書の組み合わせ方式に関する考察

$k$ 個の連想記憶行列 $M_i(i=1\sim k)$ を組み合わせる場合、本来のアソシエトロンの手法の延長で考えれば、 $M_{merge}$ は以下の式で得られる。

$$M_{merge} = \sum_{i=1}^k M_i$$

本節の実験でも当初この方法にしたがって専門連想辞書と一般連想辞書を組み合わせてみた。ところが、組み合わせた連想辞書を用いて抽出される記事は一般連想辞書を用いて得られるものとまったく同一になってしまい、今回の実験では結局この方法は採用できなかった。

このような結果となった原因は、二つの連想辞書の規模の差にある。つまり一般連想辞書を構成するために使用した記事は専門連想辞書を構成するために使用した記事よりもほとんどの場合ははるかに長大であった。また一般連想辞書の方がはるかに多数の記事を記録して構成されていた。このため、一般連想辞書には専門連想辞書より多数のリンク情報が構成されていたため、ここに専門連想辞書を単純に重ねてもそれによって一般連想辞書の持っていたリンクの数から増えるリンク数はごくわずかで、結局それらは全体の想起にたいしてほとんど影響せず、想起の結果は一般連想辞書だけを用いた場合とほとんど差がなくなってしまう。

そこで今回の実験では各連想辞書を分離したまま段階的に連想を行うことによってこの問題を回避した。しかし、前節の実験でみたように、段階的な連想を行うと、たとえば第2のアルゴリズムにみられたように、もともと検索文に含まれていたキーワードの一部が第1段階の連想で消失してしまい、第2段階の連想にそれらの語が反映されない。第3のアルゴリズムではそれをわずかにおぎなっているが、やはり、連想辞書そのものを組み合わせて使用することによって検索文に含まれていた全ての語を同時に反

映しつづ、複数の連想辞書にまたがった連想を行うことによるクロスオーバー効果を実現できることが望ましい。

そこで、各連想記憶辞書がどの程度の規模の構成を持つかをあらかじめ確認し、その規模に応じて重みをかけて行列を合成する必要がある。アソシアトロン型の自己相関行列による連想記憶では、この「規模」は二つの観点から評価する必要がある。

一つは、記録ベクトル数による評価である。つまり、ある連想記憶行列がいくつのベクトルを記録することによって構成されたかである。これは単純に記録時に連想記憶行列Aにいままで記録したベクトル数 $N_A$ を行列Aと併せて保持しておくのが最も単純であるが、あるいは以下のような方法でもおよその評価は可能である。 $\phi$ 量子化前の連想記憶行列の要素 $a_{ij}$ は、語 $w_i$ と $w_j$ とがこれまでにいくつの記事の中で同時に出現したかを示す値を持っている。つまり共起頻度である。特に $a_{ii}$ は $w_i$ がいくつの記事に含まれていたかを示す。よって $a_{ii}$ の最大値は連想記憶行列を構成したベクトルにおいてもっとも出現頻度の高かった要素がいくつのベクトルに出現したかを示すものであるから、この値はおよそその行列の $N_A$ の指標となろう。したがって、 $a_{ii}$ の最大値を調べて比べれば、各行列がどの程度のベクトルを記録して構成されたかを相対的に比較することが可能となる。

もう一つは、形成リンク数による評価である。つまり、その行列にどの程度の数のリンクが形成されているかの評価である。これは、行列A中で $a_{ij}=0$ である要素の数 $N_{A(a_{ij}=0)}$ を数えれば良い。行列Aが $m$ 次元であれば、行列A中に形成されているリンクの総数 $L_A$ は $m \times m - N_{A(a_{ij}=0)}$ である。ただし、アソシアトロンの連想記憶行列は必ず対称行列となるので、 $a_{ij}=0$ である要素の数のカウントは行列全体にわたって行う必要はない。また、本研究の場合、一般に処理した語 $w_i$ のうち $i$ の値の小さな語は出現頻度が高い。出現頻度の低い語は当然形成するリンク数は少ないので、結局全リンク数は出現頻度の高い語によってほぼ決定される。つまり本研究の場合、 $i=1 \sim m$ について調査する必要はなく、 $i=1 \sim s$  ( $s < m$ ) についてのリンク数 $L_{As}$ を調査すれば十分である。つまり、 $L_{As}$ は以下の式で得られる。

$$L_{As} = \sum_{i=1}^s \sum_{j=i+1}^s \delta_{ij}$$

ただし、

$$\delta_{ij} = \begin{cases} 1; & \text{if } a_{ij} \neq 0 \\ 0; & \text{if } a_{ij} = 0 \end{cases}$$

以上の二つの観点から得た規模の評価値 $N_A$ （あるいは $a_{ii}$ の最大値）と $L_A$ （ないし $L_{As}$ ）などを用いて、以下のような重みづけをした行列の合成を行えば、先に示したアソシアトロンの手法にしたがった単なる行列の合成による問題は解消できるであろう。

$$M_{\text{merge}} = \frac{\alpha}{N_A L_{As}} M_A + \frac{\beta}{N_B L_{Bs}} M_B + \dots$$

ここで $\alpha, \beta, \dots$ のパラメータは利用者が必要に応じて行列の合成の重みを変えるためのパラメータであり、特に指定がなければ $\alpha, \beta, \dots = 1$ である。

## 7 そのほかの検討項目

本節では、実際にはまだシステムへ実装はしていないが有用性が期待でき、今後実装する可能性のあるいくつかの機能について説明する。

### 7.1 連想記憶行列の白痴化防止対策

本研究のように1/0（または-1/0）の二値論理で記録するベクトルを構成する場合、あるいは三値論理であっても1の出現確率と-1の出現確率とに比較的大きな偏りがあるような場合には、大量の情報を記録し続けるとやがてほとんど全ての要素間にリンクが構成されてしまう可能性がある。この状況になると、どの要素を入力しても他の全ての要素が同じ強さで想起され、結局何も記録していないのと同じ状態に陥る可能性がある。これを連想記憶の白痴化と呼んでいる。この現象の発生を避けるために、従来の想起の式を以下の式に書き換える。

$$R = \phi_{\theta} \{ \phi_{\theta_1}(M)Q \}$$

つまり、従来の $\phi_0$ 量子化関数を、 $\phi_{\theta_1}$ 量子化関数に置き換えただけである。 $\phi_{\theta_1}$ 量子化関数は $\phi_0$ 量子化関数と全く同じ形の量子化関数である。量子化の敷居値 $\theta_1$ の値は固定でも良いが、一般には記録ベクトル数 $n$ の関数とするのが望ましいであろう。 $n$ が小さいうちは連想記憶の白痴化は生じないので $\theta_1=0$ とすれば良いが（この時 $\phi_{\theta_1}$ 関数は $\phi_0$ 関数と同じ）、 $n$ が増加しリンク数 $L$ がある一定割合を越えたらその時の $n$ の値に応じた $\theta_1$ の値を設定する。単純な例としては、

$$\theta_1 = n \bmod 10$$

のような関数が考えられる。ここに、 $\bmod$ は $n$ を10で割った商を与える演算子である。このように連想記憶行列の量子化を変更することによって、記録のし過ぎによる連想記憶の白痴化を防止できる。

## 7.2 記録時期による連想記憶行列の分割による、想起への時刻情報の反映

アソシアトロンをはじめ、多くの連想記憶では、記録した情報がいつ記録されたかに関する情報が無い。しかし、現実の人間だと比較的最近覚えた情報は記録した情報の量としては少なくとも、古い情報よりも想起されやすい場合がある。たとえば、相撲についての新聞記事から連想記憶行列を構成する場合を考える。この時、おそらく今のところ先代の貴の花に関する記事の方が今の貴の花に関する記事より多く存在するので、「貴の花」で想起を行なうと、まず先代の貴の花に関する情報が強く想起されるであろう。しかし、先代の貴の花について詳しい人であっても、今「貴の花」という単語を聞けば、まず二代目の貴の花について思い出す人が大半であろう。

そこで、このような「最近記録した情報が（記録量が少なくとも）まず強く想起される」状況を作るために、記録した時期に依存して連想記憶行列を分割して構成しておき、必要に応じて前節で述べた合成手法によって任意の強度で新しい連想記憶行列と古い連想記憶行列とを合成して使用する。

具体的な手法は以下の通りである。まず入力された記録ベクトルは新記憶行列 $M_{new}$ に記録される。同時に、この記録ベクトルは時刻タグつきで記録ベクトルデータベースに登録される。

システムはこの記録ベクトルデータベースを監視し、記録以後一定の期間を経過したベクトルに関する記憶を $M_{new}$ から $M_{old}$ へ移す作業を行なう。 $M_{new}$ から $K_{aging}$ を抜く（すなわち忘却する）処理は以下の式によって行なう。

$$M_{aging} = M_{new} - K_{aging} K_{aging}^T$$

ここに、 $M_{aging}$ は忘却処理後の $M_{new}$ 、 $K_{aging}$ は縦ベクトル、演算子 $^T$ はベクトルの転置を示す。ついで、 $K_{aging}$ を $M_{old}$ に記録する。記録の仕方は従来のアソシアトロンの方法と同じ。この移動の処理が終了したら、 $K_{aging}$ を記録ベクトルデータベースから削除する。

このような手法によって、 $M_{new}$ にはつねに最近の一定期間に入力された情報のみが記録されていることになる。必要に応じてこの $M_{new}$ と $M_{old}$ とを必要な重みで組み合わせて使用することによって新しい情報を優先的に、あるいは古い情報を優先的に使用することが可能となる。

## 7.3 共起関係が不明な要素間への弱いリンクの設定

アソシアトロンでは、記録ベクトル中で一度も共起したことがない要素 $w_i$ 、 $w_j$ の間には直接的なリンクは構成されない。つまり $a_{ij} = 0$ である。しかし、特に記録ベクトル数が比較的少ない状況、あるいは $w_i$ または $w_j$ の一方がかなり出現頻度の低いような要素である場合には本来 $w_i$ と $w_j$ の間には直接的な関連があるかもしれないのに、 $w_i$ から $w_j$ を直接取り出すことができない。そこで、以下のような方法で共起性が不明な要素間にも弱いリンクを構成することによって、記録ベクトルでは一度も共起したことがない要素を直接取り出せるようにする。

まず、記録ベクトルを従来のアソシアトロンの(1, 0, -1)の三値論理から、(a, b, -c, -d)の四値論理に変更する。ただし、 $a, b, c, d > 0$ 、 $a > b > -c > -d$ である。つまり、従来のアソシアトロンの場合で、値1をとる要素が値 $a$ を、値-1をとる要素が値 $-d$ を、値0をとる要素が値 $b$ ないし $-c$ をとる。 $b$ と $-c$ のいずれをとるかの決定手段は種々考えられるが、ここではとくに限定しない。このようにして記録ベクトルを構成して、従来のアソシアトロンと同様の記録を行なう。

例えば、 $a=d=3, b=c=1$ とした場合を考えると、従来(100101)で表わされていた記録ベクトルはたとえ



ば(3 1 1 3 1 3)となる。このベクトルの自己相関行列は、

$$\begin{pmatrix} 9 & 3 & 3 & 9 & 3 & 9 \\ 3 & 1 & 1 & 3 & 1 & 3 \\ 3 & 1 & 1 & 3 & 1 & 3 \\ 9 & 3 & 3 & 9 & 3 & 9 \\ 3 & 1 & 1 & 3 & 1 & 3 \\ 9 & 3 & 3 & 9 & 3 & 9 \end{pmatrix}$$

となる。つまり、共起している要素間には強いリンク(9)が構成され、出現している要素と出現していない要素の間にはやや強いリンク(3)が、また出現していない要素同士の間には弱いリンク(1)が構成される。この結果すべての要素間になんらかの強さのリンクが構成されるので、そのまま量子化せずに使用しても良いし、あるいは $\phi$ の量子化や、6.2節の実験で示した弱連想フィルタのような量子化を用いることによって適当な強さのリンクを取り出して使用することが可能となる。このような弱いリンク情報は、連想の飛躍などに利用できると思われる。

#### 7.4 マスク行列による特定要素群間の直接的関連の削除

利用状況によっては、ある要素(群)Gと別のある要素(群)Hとの間に意図的に直接的関連を設定したくない場合がある。特に7.3で説明したような方法で、明示的に共起性を与えなかった要素間にも関連が設定されてしまうような場合にこのような要求が発生することが考えられる。このような場合に、マスク行列を構成することによって指定の要素(群)間の直接的関連を切断することができる。

マスク行列は次のようにして生成する。たとえば、

$$G: (e1, e3), H: (e5)$$

という二つの要素群間の関連を切る場合には、一方の要素群を1、もう一方の要素群を-1で表現したベクトルを作成する。この例の場合は、

$$(1 0 1 0 -1 0)$$

である(次元mは6とする)。次にこのベクトルの自己相関行列を求める。

$$\begin{pmatrix} 1 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

この自己相関行列のうち、値が-1の要素を値0に、そのほかを値1に書き換える。すなわち、

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

となる。これがマスク行列 $M_{\text{mask}}$ である。ある連想記憶行列Mのうち、上記の要素群GおよびHの関連を切断したい場合には、Mに対して、この $M_{\text{mask}}$ の要素毎の積をとればよい。

さらに、要素群VとWとの間の関係も切断した場合には上記と同様にVとWについてのマスク行列を求め、このマスク行列とGとHについてのマスク行列との要素毎の積を求めて得た行列がこの二つの関係を併せた合成マスク行列となる。

なお、ここで注意しなければならないのは、たとえば上記のように要素群GとHに対するマスク行列を使用している連想記憶行列にマスクをかけた場合であっても、要素群Gに含まれる要素をもつ検索ベクトルから要素群Hに含まれる要素をもつベクトルが想起される場合はありうる。これは、Hに含まれる要素とGに含まれない要素との共起関係によるものである。つまり、ここでのマスク行列は設定した要素間の直接的な想起への寄与を下げるだけの役割を果たすものであり、検索ベクトルにGに含まれる要素がある場合に、想起ベクトルに絶対Hに含まれる要素が現われないようにするものではない。

## 8 おわりに

以上、発散的思考支援を意図した、連想記憶を用いた情報検索による関連性と異質性とを併せ持つような情報の抽出手段について説明した。関連性と異質性のある情報の抽出メカニズムとして、意味の粗い把握過程、浅い理解過程、関連性する情報の抽出過程の3過程からなる「門外漢モデル」を提案し、これに基づく実験システムを構築した。また、思考のテーマを中心とした客観的／主観的基準による情報空間の分類を行い、これに基づいてある情報がどのように人の発想（発散的思考）に寄与するかを評価する手法を提案した。実験システムを用いて3種類の実験を行い、以下の結論を得た。

- (1) 門外漢モデルによって異質性と関連性を併せ持つ情報の抽出が可能となる。
- (2) システムの持つ連想記憶行列の内容と、利用者が入力した検索文の内容とを比較し、必要に応じて差分フィルタあるいは弱連想フィルタを用いることによって異質性や関連性を積極的に検索結果に導入することが可能となる。
- (3) 分野の異なる情報から構成した連想記憶行列を複数準備し、これを適宜組み合わせで使用することによって抽出される情報の分野を粗く制御することが可能となる。さらに、組み合わせ方によっては、個々の単独の分野の専門家が集まったのでは得られないような、複数分野にまたがる知識をもつものならではの情報を抽出することが可能となる。

さらに、未実装ながら今後必要となってくると思われる四つの機能について説明を加えた。

今後の予定であるが、ここまでで提案した評価手法は、あくまで異質性と関連性とを併せ持つ情報が発散的思考の触発に有効であるという前提の下に、ある情報がどの程度異質性と関連性とを併せ持っているかを評価するものであった。したがって、異質性も関連性も高い評価を与えられた情報が得られたとしても、現段階ではそれが本当に発散的思考に有効に作用するかどうかは推測の域を出ない（ただし、経験的にはその有効性は認められている）。そこで、次段階の評価として、本プロトタイプシステムが提供する情報が本当に発散的思考を効果的に支援するものとなっているかどうかを評価する必要がある。このため、まずその評価手法を提案し、続いて実際に評価実験を行う予定である。

また、現在は本システムが処理可能な情報はテキスト情報に限られているが、画像や音声などのマルチメディア情報を取り込むことも試みたい。ただし、現在のシステムでは、それぞれに意味のあるシンボルの集合として扱えないような情報を処理することはできない。したがって、画像や音声などの情報を処理可能にするためには、それらに意味的に対応したシンボル情報を付加することが必須である。単純に言えばキーワードづけであるが、現在の技術レベルでは画像や音声に全自動にキーワードを付与することはきわめて困難である。したがって、これらの情報に対するキーワードづけは実質人手に頼らざるを得ないが、作業量も膨大であるし、その作業を行う特定個人の視点の制約を受けたキーワードづけとなることもあまり好ましくない問題である。しかし一方で画像や音声情報にはテキスト情報とは異なるタイプの発想の触発効果を期待できるので、なんらかの手段で画像／音声データを本システムで処理可能として被験者実験を試みたい。

## 参考文献

- [川喜田67] 川喜田二郎：発想法 創造性開発のために、中公新書 (1962)
- [西本93a] 西本一志、望月研二、宮里勉、岸野文郎：発散的思考支援 -意外性のある情報の提供手法に関する検討-、人工知能学会研究会資料SIG-HICG-9301-1, pp.1-8, 1993
- [中野72] 中野馨：アソシアトロン -連想記憶のモデルと知的情報処理-、昭晃堂, 1972.
- [西本95] K. Nishimoto, S. Abe, T. Miyasato, and F. Kishino: "A System Supporting The Human Divergent Thinking Process by Provision of Relevant And Heterogeneous Pieces of Information Based on An Outsider Model", proc. of The Eighth International Conference on Industrial & Engineering Applications of Artificial Intelligence and Expert Systems, 1995.
- [西本93b] 西本一志、宮里勉、岸野文郎：連想記憶を用いた情報提供による発散的グループ思考支援の試み、国藤研究室主催 第2回「発想支援ツール」シンポジウム講演論文集、1993
- [西本94] 西本一志、安部伸治、宮里勉、岸野文郎：連想記憶を用いた発散的思考支援システムにおける提供情報の分野制御の試み、計測自動制御学会第15回システム工学部会研究会資料「発想支援技術」、pp. 17-24, 1994.