

1C-1

記号の空間的表現について

田村 淳

日本電気(株) C & C 情報研究所

1. はじめに

概念間の近さを利用して、情報の柔軟な検索を実現することが要請されており、概念語の静的な空間配置によって概念間相互の近さを表現することが提案されている[堀86]。しかし、対象とする概念領域を変更した場合には、それに応じて空間配置を動的に変化させる必要がある。そこで、本論文では、相互作用による空間形成モデルを提案する。

2. 記号とボタン

一般に概念間の関係には、①論理的关系と②連想的関係の2種類がある。①は記号処理に適しており、そこでは基本的には記号間の正確な一致が問題にされる。②は概念間の「近さ」に基づいておりボタン処理に適している。

本論文では、「記号」と「ボタン」とを別のものとしてとらえるのではなく、より「記号」的あるいは「ボタン」的であるというように相対的にとらえる。すなわち、ある表現がより「記号」的であると、概念がより抽象化され、より簡潔に表現されていることである。一方、より「ボタン」的であると、より空間的あるいは時間的広がりがあり、より広い範囲がひとまとまりとして意味をもつことである。ボタンのような「面」的に広がり、記号的であるほど「点」に近づく。

「点」としての記号をボタンへ変換する方法には、1つの記号を多数の“microfeature”の組み合わせで表現する方法がある。しかし、この方法は、結局は“microfeature”という新たな記号を定義していることになる。

論理的关系と連想的関係の両方を処理するには、それぞれ別個に処理する方法がある。たとえば、筆者らの自然言語処理システムでは、コネクショニスト・モデルを用いて概念間の興奮と抑制による相互作用により多義文の文脈処理を行ない、一方、構文解析はプロダクション・システムを用いている[田村87a]。

これに対して、本論文では、それぞれ別個に対応するのではなく、記号表現としての性質もボタン表現としての性質も兼ね備えた表現形式を考える。すでに、意味の空間的表現についての研究[堀86]がなされているが、その表現の生成方式は、最も大きな矛盾を引き起こしている「点」を対象にした逐次的なアルゴリズムである。本研究では記号のボタンの表現の生成方式を並列分散処理としてとらえて検討する。また、制約プログラミングでは、現在、厳密な論理的关系を対象とした制約「論理」プログラミングに主眼が置かれているが、本研究は制約条件を大まかに満たす解を導くための手法であるということもできる。

3. 動的相互作用表現

本論文では、記号間に非類似度(距離の公理を満たすとは限らないので、本論文では非類似度とよぶ)が与えられたときの情報の表現方法として、記号の「点」を空間的な「面」上に配置したボタンを考える。記号は各記号間の非類似度という制約をできるだけ満たすように空間的に配置される。その結果、記号間の空間的な「近さ」はなんらかの意味的な近さを表現することになる。

記号群を空間的に配置する際には、まず、「点」としての各記号をユークリッド空間中のある初期位置にそれぞれ配置する。ここで、それぞれの「点」に対して、空間内の2点間の距離と、制約としての非類似度との差に比例した「力」がはたらくと仮定する。m個の点が存在するとき、自分以外の(m-1)個の点のうち自分との間に制約の課せられた点からこの「力」がはたらき、その「力」のベクトル和により時間Δt後の点の位置が決まる。たとえば、図1では、記号Aと記号B、および記号Bと記号Cが非類似度との関係から現在よりも近くに配置されるべきであり、記号Aと記号Cは現在よりも遠くに配置されるべきであるという状態であり、次に動くべき方向が2つのベクトル和として表現されている。このΔt時間刻みの処理を繰り返して各点の位置の変更が十分小さくなれば、収束したものとみなして処理を終了する。このとき、準最適解が得られている。

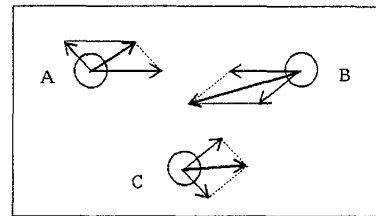


図1 「点」としての記号にはたらく「力」

具体的には、第i番目の記号 $P_i$ と第j番目の記号 $P_j$ との間に非類似度 $d(i, j)$ が与えられているとき、この非類似度による制約をできるだけ満たして記号をn次元空間中に配置することを考える。ここで、 $P_i$ の第k軸の座標値を $p(i, k)$ で表わし、その時点での $P_i$ と $P_j$ との距離を $\delta(i, j)$ で表わす。このとき記号 $P_i$ の第k軸方向の補正值 $\Delta(i, k)$ は、

$$\Delta(i, k) = c * \sum (p(j, k) - p(i, k)) * (1 - d(i, j) / \delta(i, j))$$

で与えられる。cは定数である。なお、 $\Sigma$ は $P_i$ に対して制約条件が与えられているjについての総和を意味する。

このようにして、記号群のボタンは時間とともに変形してある状態で安定するが、この記号群ボタンの時間的・空間的表現を動的相互作用表現とよぶことにする。

本方式では、記号間の非類似度のセットとして制約条件を与えると、その制約条件をできるだけ満たすような、記号の空間的配置が生成される。したがって、記号群の部分パタンの切り出しや、非類似度の値の変更により、対象領域や制約条件を変更しても、その時点で新たな表現を生成することができる。つまり、動的相互作用表現の「相互作用」には、利用者あるいは制御部と表現系との相互作用と、その結果与えられた条件の下での記号間の相互作用とがある。

本方式は、一つにはヒューマンインタフェースとして用いることができる。この場合、似ている概念を表わす記号が近くに配置されるので、利用者が概念をうまく言葉で表現できない場合にも、「それらしき」記号群パターンを指示することで、利用者の概念をシステムに伝えることができる。

また、本方式は、計算機内部の知識表現形式として用いることもできる。この場合、計算機がある記号群を処理しているとき、強く「想起」している記号ほど近くに配置されることになる。したがって、推論の対象を絞るときに、制御部は「空間的近傍」を探索すればよい。

以下では、情報検索の利用者インタフェースへの応用例について述べる。

#### 4. 情報検索への応用

情報検索では、キーワードや分類による論理的な関係に基づいた検索手がかりが用いられている。そのため、利用者がもつ概念（情報要求）と検索文書との「妥当性」よりも、利用者が与えた表現（検索手がかり）と検索文書との「関係性」の方に重点が置かれていた（図2）。

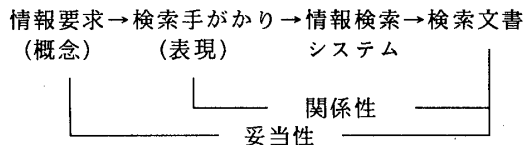


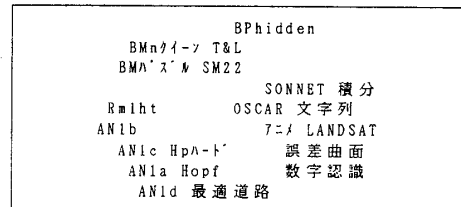
図2 情報検索における妥当性と関係性

妥当性を高めるために、筆者らは連想的関係を重視した方式[田村87b]をすでに提案しているが、概念（ユニット）間の関係の度合い（リンクの重み）の付け方に明確な指針がなかった。ここでは、文献間の非類似度を表わす尺度として、共引用度の逆数を用いる。共引用度とは計量書誌学における引用分析で提案されているもので、文献*i*と文献*j*の共引用度は  $N(C_i \cap C_j) / N(C_i \cup C_j)$  で与えられる[Garf80]。ここで、 $C_i$ 、 $C_j$ はそれぞれ文献*i*、*j*が引用している文献の集合である。共引用度を用いれば、共引用度は対象分野の知識がなくても引用情報だけから動的相互作用表現を生成でき、その結果、キーワードや分類以外の検索手がかりを作り出すことができる。この表現を利用者に提示すれば、従来、キーワードや分類で表現しにくかった概念をうまく表現でき、検索文書の妥当性を高めることができる。

#### 5. 実行例

共引用度に基づいて文献を空間的に表示するため

の簡易実験システムをPC-9801上にBASICで構築した。このシステムにより、前回の全国大会におけるニューラルネット関係の論文[情処89]を本方式によって表現した例を図3に示す。表示対象は関連論文36件のうち他の論文と共引用文献がないもの14件を除外した22件である。この22件の引用文献は延べ81件あり、そのうち8件が複数の論文から引用されている。



(重なった部分は一部修正・削除してある。)

図3 動的相互作用表現の一例

図3中、ボルツマンマシン関係の論文("BMnクイ-ツ", "BMn'ス'ル")が左上に、バックプロパゲーション関係の論文("BPhidden", "T&L")が中央上部に、その下にシミュレータの論文("SM22", "SONNET")が配置されている。ホップフィールドネットの考察("Hopf")やそのハードウェア化についての論文("AN1a"~"AN1d", "HpA-ト")が左下に、それぞれ固まっていることがわかる。また、その他の応用などが右側に集まっている。

なお、今回の実験では、対象文献が全国大会の論文であり、引用論文数が一論文あたり平均3.7件と少なく、共引用文献が少ないので、初期値や各種パラメータ（各時刻の変更量係数、極小値判定定数、共引用度が0のときの非類似度など）により若干異なった配置結果が得られる。

#### 6. おわりに

記号の空間的表現方式として動的相互作用表現を提案し、これを情報検索に応用した。文献だけでなく、人、組織、商品などさまざまな対象を表現する応用が考えられる。また、新しい知識表現形式としても発展させてゆきたい。今後は、実験を重ねて本モデルを洗練させてゆく予定である。

#### [参考文献]

- [Garf80]Garfield, E.: "ABCs of Cluster Mapping; part 1. Most Active Fields in the Life Sciences in 1978". Current Contents, No.40, pp.5-12 (1980).
- [堀86]堀浩一: 単語の意味の学習について, コトブキ7, Vol.3, No.4, pp.65-72 (1986).
- [情処89]情報処理学会第38回全国大会講演論文集 (I), (II), (III) (1989).
- [田村87a]田村淳, 安西祐一郎: Connectionist Modelを用いた自然言語処理システム, 情報処理学会論文誌, Vol.28, No.2, pp.202-210 (1987).
- [田村87b]田村淳, 原良憲, 金子朝男: Connectionist Modelによる「対話」としての情報検索システム, 人工知能学会全国大会(第1回), pp.159-162 (1987).