

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

並列分散処理による、記号の空間的表現の生成方式を提案する。記号を空間中の「点」としてとらえ、記号間の類似度という制約をできるだけ満たすように「点」としての記号を空間中に配置する。バネからなる、摩擦力を含む力学系では、系のもつ力学的エネルギーは時間とともに減少するが、各バネの自然長を各制約条件に、系全体の力学的エネルギーを制約条件全体の充足の程度として考えると、極小値に収束した状態は準最適解を意味し、似た概念を表す記号が互いに近くに配置されている。本方式の1次元空間の場合の解析、および2次元空間においてすべての2点間を等距離にするという制約のもとでのシミュレーションを行なった。さらに、情報検索の利用者インタフェースに応用し、その有効性を示した。

A Method for Generating a Spatial Representation of Symbols

Atsushi Tamura

C&C Information Technology Research Labs., NEC Corp.

This paper proposes a method for generating a spatial representation of symbols by parallel distributed processing. Symbols are arranged in space with constraint of pairwise similarities between symbols. The process corresponds to a spring system with friction. Whenever symbols interact and move, the mechanical energy decreases by the effect of friction. As the degree of constraint satisfaction can be defined by the energy, the state of local minimum becomes a quasi-optimal solution.

One of its application is human interface for information retrieval systems. In this case, co-citation ratios are used as similarity measures between documents. The concept which is hard to represent only by words can be represented spatially by this method.

1 はじめに

記号の表現手法について考えてみると、従来のフレームや意味ネットワークなどの手法では、あらかじめ設定された属性の枠内で表現しなければならないため、新しい概念や、ことばで表現しにくい概念に対して、必ずしも柔軟な表現能力があるとはいえなかった。

本論文では、記号とボタンとを比較した上で、「ボタン的な記号」を生成する方式を提案する。すなわち、記号を空間中の「点」とみなし、その「点」としての記号の間で、並列分散処理の概念に基づいた相互作用により、その空間的配置を決定するメカニズムを提案する。

2 記号とボタン

本論文では、表現の構成単位が一つだけで意味をもつ情報を「記号」とよび、構成単位が一つだけでは意味をもたず、構成単位間の関係が意味をもつ情報を「ボタン」とよぶことにする。このようにとらえると、表現の構成単位が一つだけでも意味をもち、かつ構成単位間の関係も意味をもつ情報も考えることができる。言い換えれば、「記号」としての性質も「ボタン」としての性質も兼ね備えた表現を考えることができる。

このような「ボタン」的な「記号」による表現については、すでに、意味の空間的表現の研究[1]がなされている。本論文で提案する方式で得られる表現も、結果としては文献[1]と同様に、似た概念を表す「記号」が近くに配置され、その「記号」が単独で意味をもつと同時に、その周辺領域が全体として何らかの意味をもった「ボタン」にもなる。ただし、文献[1]では、この表現を生成するために、最も大きな矛盾を引き起こしている「記号」を対象にした逐次的なアルゴリズムを用いているが、本研究では並列分散処理によって記号の空間的表現を生成する方式を提案する。

3 動的相互作用表現の基本構想

本論文では、記号という「点」どうしの間に「望ましい長さ」が与えられたときの情報の表現方法として、それぞれの「点」を空間的な「面」に広げた配置パターンを考える。記号は、各記号間の「望ましい長さ」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。初期位置は、ランダムに決めてもよいし、人間が大まかに指定してもよい。ここで、それぞれの「点」に対して、空間内の2点間のその時点での距離と、制約としての「望ましい長さ」との差に比例した「力」がはたらくと仮定する。 n 個の点が存在するとき、自分以外の $(n-1)$ 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微小時間 Δt 後におけるそ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「望ましい長さ」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。詳しくは次章以降で述べる。

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、この記号群の空間的表現を「動的相互作用表現」とよぶことにする。ところで、たとえば、2次元平面上では、3点までならば、それぞれの点を等距離に配置することができるが、4点以上になるとこの制約条件を完全に満たすことはできない。本方式では、それぞれの制約条件を完全ではなくても、できるだけ満たすように、記号を空間的

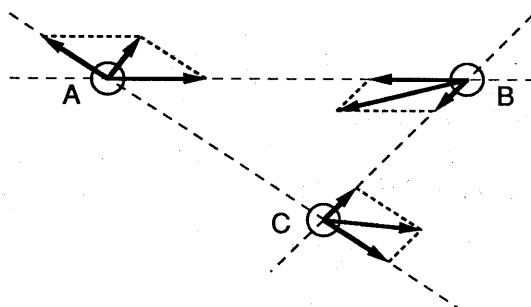


図 1: 「点」としての記号にはたらく「力」

に配置する。

4 力学系との対応

n 個の点 $P_1, P_2, \dots, P_n$ を次の制約条件をできるだけ満たして空間に配置することを考える。制約条件は、点の対 (P_i, P_j) に対して与えられた「望ましい長さ」 d_{ij} を満足することである。この制約条件はすべての点の対に対して与えられる必要はない。なお、「望ましい長さ」 d_{ij} には、対称性、すなわち、

$$d_{ij} = d_{ji}$$

が成立すると仮定する。また、

$$d_{ii} = 0$$

とする。なお、「望ましい長さ」の間には、三角不等式

$$d_{ij} \leq d_{ik} + d_{kj}$$

が成立するとは限らないので、「望ましい長さ」は数学的な意味での距離とはいえない。

この解を求めるために、バネを組み合わせた力学系を考える。ここで、バネ定数 k のバネが x だけ変位したときに加わる力 F は、

$$F = kx$$

であり、そのときのエネルギー E は、

$$E = \frac{1}{2} kx^2$$

である。

このバネの系からのアナロジーに基づき、 n 個の点の空間配置を求める問題において、それぞれの「点」を質量 m の質点と考え、点 P_i と点 P_j とが自然長 d_{ij} のバネでつながれていると考える。この自然長 d_{ij} は i 番目の記号と j 番目の記号との間の「望ましい長さ」である。

点 P_i と点 P_j の時刻 t における位置ベクトルをそれぞれ $\mathbf{p}_j(t), \mathbf{p}_i(t)$ とすると、バネは $(\mathbf{p}_j(t) - \mathbf{p}_i(t))$ の方向にあり、長さは $|\mathbf{p}_j(t) - \mathbf{p}_i(t)|$ であるから、点 P_i に対して点 P_j から、

$$\mathbf{F}_{ij}(t) = k(|\mathbf{p}_j(t) - \mathbf{p}_i(t)| - d_{ij})\mathbf{e}_{ij}(t)$$

で与えられる力 F_{ij} が加えられる。ここで、

$$e_{ij}(t) = \frac{\mathbf{p}_j(t) - \mathbf{p}_i(t)}{|\mathbf{p}_j(t) - \mathbf{p}_i(t)|}$$

は点 P_i から点 P_j へ向かう単位ベクトルである。

点 P_i は自分自身以外の点 $P_j (j \neq i)$ のうち、制約が課せられたそれぞれの点から力を受けるので、点 P_i が時刻 t で受ける力 $\mathbf{F}_i(t)$ は、 $\mathbf{F}_{ij}(t)$ のベクトル和として、

$$\begin{aligned} \mathbf{F}_i(t) &= \sum_{j \neq i} \mathbf{F}_{ij}(t) \\ &= k \sum_{j \neq i} (|\mathbf{p}_j(t) - \mathbf{p}_i(t)| - d_{ij}) \mathbf{e}_{ij}(t) \end{aligned}$$

で与えられる。その結果、点 P_i は速度 $\frac{d\mathbf{p}_i(t)}{dt}$ で動き出すが、そのときこの速度に比例した摩擦力 $-\mu \frac{d\mathbf{p}_i(t)}{dt}$ がはたらくと考える (μ : 動摩擦係数)。各点に加わる力を、図 2 に示す。

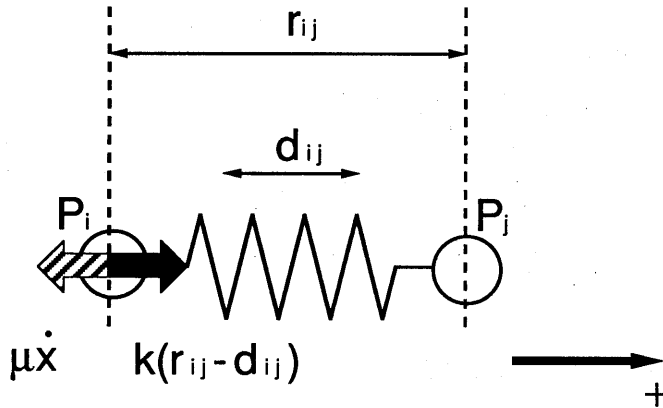


図 2: バネにはたらく力

このときの運動方程式は、

$$m \frac{d^2 \mathbf{p}_i(t)}{dt^2} = k \sum_{j \neq i} (|\mathbf{p}_j(t) - \mathbf{p}_i(t)| - d_{ij}) \mathbf{e}_{ij}(t) - \mu \frac{d\mathbf{p}_i(t)}{dt}$$

より、

$$m \frac{d^2 \mathbf{p}_i(t)}{dt^2} + \mu \frac{d\mathbf{p}_i(t)}{dt} + k \sum_{j \neq i} (d_{ij} - r_{ij}(t)) \mathbf{e}_{ij} = 0$$

と書くことができる。ただし、 $r_{ij}(t)$ は点 P_i と点 P_j の間の距離で、

$$r_{ij}(t) = |\mathbf{p}_j(t) - \mathbf{p}_i(t)|$$

で表される。

また、このようなバネの組み合わせからなる系全体の時刻 t における力学的エネルギー $E(t)$ は、弾性エネルギーと運動エネルギーの和から、

$$E(t) = \frac{1}{2} \sum_{i < j} (|\mathbf{p}_j(t) - \mathbf{p}_i(t)| - d_{ij})^2 + \frac{1}{2} m \sum_i \left(\frac{d\mathbf{p}_i(t)}{dt} \right)^2$$

である。この系の力学的エネルギーは、時刻 t が進むにつれて、摩擦力による熱エネルギーの分だけ減少してゆく。

以下では、質量 m を 0 に限りなく近づけた極限について考える。これは、物理的には減衰振動を起こさないということであり、1階の微分方程式で表現することができる。

このバネの問題は多体問題であって、解析的に解くことができないが、次節では Euler 法による近似解の求め方について述べる。その後で、この計算によって弾性エネルギーの極小値が求められることを示す。

5 Euler 法による解法

本節では、 $m \rightarrow 0$ とした結果得られる、1階微分方程式

$$\mu \frac{d\mathbf{p}_i(t)}{dt} + k \sum_{j \neq i} (d_{ij} - r_{ij}(t)) \mathbf{e}_{ij}(t) = 0$$

の解を、前進型 Euler 公式を用いて求める。

まず、微分をきざみ幅 Δt で 1 次近似すると、

$$\mu \frac{\mathbf{p}_i(t + \Delta t) - \mathbf{p}_i(t)}{\Delta t} + k \sum_{j \neq i} (d_{ij} - r_{ij}(t)) \mathbf{e}_{ij}(t) = 0$$

となる。これを変形すると、

$$\mathbf{p}_i(t + \Delta t) = \frac{k}{\mu} \Delta t \sum_{j \neq i} (r_{ij}(t) - d_{ij}) \mathbf{e}_{ij}(t) + \mathbf{p}_i(t)$$

となるが、ここで、 $c = \frac{k}{\mu} \Delta t$ とおけば、

$$\mathbf{p}_i(t + \Delta t) = c \sum_{j \neq i} (r_{ij}(t) - d_{ij}) \mathbf{e}_{ij}(t) + \mathbf{p}_i(t)$$

となる。すなわち、時間 Δt 後の点 P_i の位置は、時刻 t における点 P_i と点 P_j の間の距離 $r_{ij}(t)$ とその望ましい距離 d_{ij} との差に応じて得られる、点 P_j の方向への微小なベクトルを、 j について合成することによって求められる。

6 弾性エネルギーの単調減少性

この系のもつ力学的エネルギーは、摩擦力によって熱エネルギーとして散逸する分、減少してゆく。点 P_i と関係するバネの弾性エネルギー $E_i(t)$ は、

$$E_i(t) = \frac{1}{2} k \sum_{j \neq i} (r_{ij}(t) - d_{ij})^2$$

と書くことができる。 k はバネ定数である。点 P_i 以外の点が固定されていると考えると、点 P_i と関係するバネの弾性エネルギー $E_i(t)$ の変化は、

$$E_i(t+1) - E_i(t) = \frac{1}{2} k \sum_{j \neq i} \{(r_{ij}(t+1) - d_{ij})^2 - (r_{ij}(t) - d_{ij})^2\}$$

となる。ここで、時刻 t における $r_{ij}(t)$ の変化を $\Delta r_{ij}(t)$ とすると、

$$r_{ij}(t+1) = r_{ij}(t) + \Delta r_{ij}(t)$$

である。 $|\Delta p_i|$ が十分小さいとすると、 $\Delta r_{ij}(t)$ は、

$$\Delta r_{ij}(t) \simeq -\mathbf{e}_{ij}(t) \cdot \Delta \mathbf{p}_i(t)$$

と近似できる (図 3)。

また、 $\Delta r_{ij}(t)$ は $c(r_{ij}(t) - d_{ij})$ のオーダーであるから、 $(r_{ij}(t) - d_{ij})$ に比べて無視することができる。し

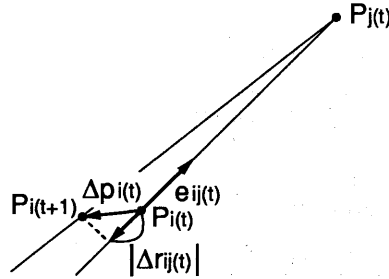


図 3: $\Delta r_{ij}(t)$ の近似

たがって、エネルギー $E_i(t)$ の変化分は、

$$\begin{aligned} E_i(t+1) - E_i(t) &= \frac{1}{2}k \sum_{j \neq i} (2r_{ij}(t) + \Delta r_{ij} - 2d_{ij}) \Delta r_{ij}(t) \\ &\simeq k \sum_{j \neq i} (r_{ij}(t) - d_{ij}) \Delta r_{ij}(t) \\ &\simeq -k \sum_{j \neq i} (r_{ij}(t) - d_{ij}) \mathbf{e}_{ij}(t) \cdot \Delta \mathbf{p}_i(t) \\ &= -ck \sum_{j \neq i} \sum_{k \neq i} (r_{ij}(t) - d_{ij})(r_{ik}(t) - d_{ik}) \mathbf{e}_{ij}(t) \cdot \mathbf{e}_{ik}(t) \end{aligned} \quad (1)$$

と書くことができる。

$$\begin{aligned} \tilde{\mathbf{r}}_{ij}(t) &= (r_{ij}(t) - d_{ij}) \mathbf{e}_{ij}(t) \\ \tilde{\mathbf{R}}_i(t) &= \sum_{j \neq i} \tilde{\mathbf{r}}_{ij}(t) \end{aligned}$$

とおけば、式(1)は、

$$\begin{aligned} E_i(t+1) - E_i(t) &= -ck \sum_{j \neq i} \sum_{k \neq i} \tilde{\mathbf{r}}_{ij}(t) \cdot \tilde{\mathbf{r}}_{ik}(t) \\ &= -ck |\tilde{\mathbf{R}}_i|^2 \leq 0 \end{aligned}$$

となるから、弾性エネルギー $E_i(t)$ は広義の単調減少性をもち、この系はエネルギー E の極小値を求めるように挙動することがわかる。ところで、弾性エネルギー $E(t)$ は、理想値 d_{ij} と実際値 r_{ij} の誤差の自乗和に比例した値を表している。すなわち、制約としての「望ましい長さ」とそれに対応する各点間の距離との差の自乗和の極小値を求めており、この意味で制約条件をできるだけ満たすようにこの系は動作するといえる。

7 1次元の場合の解析

この系の挙動を簡単に示すために、以下では、 n 個の点を1次元に配置する場合について考察する。ここで、任意の時刻 t において、

$$p_1(t) < p_2(t) < \cdots < p_n(t)$$

が成立すると仮定すると、

$$\frac{p_j(t) - p_i(t)}{|p_j(t) - p_i(t)|} = \begin{cases} 1 & \text{if } i < j \\ -1 & \text{if } i > j \end{cases}$$

であるから、点 P_i の時刻 t における変位 $\Delta p_i(t)$ は、

$$\Delta p_i(t) = c \left\{ - \sum_{j < i} (p_i(t) - p_j(t) - d_{ij}) + \sum_{j > i} (p_j(t) - p_i(t) - d_{ij}) \right\}$$

となる。したがって、次の時刻 $(t+1)$ における点 P_i の位置 $p_i(t+1)$ は、

$$\begin{aligned} p_i(t+1) &= p_i(t) + \Delta p_i(t) \\ &= \{1 - (n-1)c\} p_i(t) + c \left\{ \sum_{j \neq i} p_j(t) + \sum_{j < i} d_{ij} - \sum_{j > i} d_{ij} \right\} \end{aligned}$$

となるから、時刻 $(t+1)$ における点 P_i と点 P_j の間の距離 $r_{ij}(t+1)$ は、

$$\begin{aligned} r_{ij}(t+1) &= p_j(t+1) - p_i(t+1) \\ &= (1 - nc)(p_j(t) - p_i(t)) + c \left(\sum_{k < j} d_{jk} - \sum_{k > j} d_{jk} - \sum_{k < i} d_{ik} + \sum_{k > i} d_{ik} \right) \end{aligned} \quad (2)$$

となる。ここで、

$$D = \sum_{k < j} d_{jk} - \sum_{k > j} d_{jk} - \sum_{k < i} d_{ik} + \sum_{k > i} d_{ik}$$

とおくと、

$$D = 2d_{ij} + \sum_{k < i} (d_{kj} - d_{ki}) + \sum_{i < k < j} (d_{ik} + d_{kj}) + \sum_{k > j} (d_{ik} - d_{jk}) \quad (3)$$

と変形することができる。 D を用いて式(2)を書き直すと、

$$r_{ij}(t+1) = (1 - nc)r_{ij}(t) + cD$$

となり、さらに、

$$r_{ij}(t+1) - \frac{D}{n} = (1 - nc) \left(r_{ij}(t) - \frac{D}{n} \right)$$

と書ける。したがって、数列 $\{r_{ij}(t) - D/n\}$ は初項 $(r_{ij}(0) - D/n)$ 、等比 $(1 - nc)$ の等比数列であるから、時刻 t では、

$$r_{ij}(t) = (1 - nc)^t \left(r_{ij}(0) - \frac{D}{n} \right) + \frac{D}{n}$$

である。ここで、 $0 < c < 1/n$ とすれば、 $0 < 1 - nc < 1$ であるから、

$$t \rightarrow \infty \text{ のとき } r_{ij}(t) \rightarrow \frac{D}{n}$$

となるが、式(3)から、 D/n は、点 P_i と点 P_j の間の距離に関する項の平均である。すなわち、 D の第1項は、点 P_i, P_j の直接の距離、第2項は点 P_k ($p_k < p_i$) を中継した間接的な距離、第3項は点 P_k ($p_i < p_k < p_j$) を中継した間接的な距離、第4項は点 P_k ($p_j < p_k$) を中継した間接的な距離をそれぞれ示している (図4)。ただし、 d_{ij} のみ2倍の加重がかかっている。

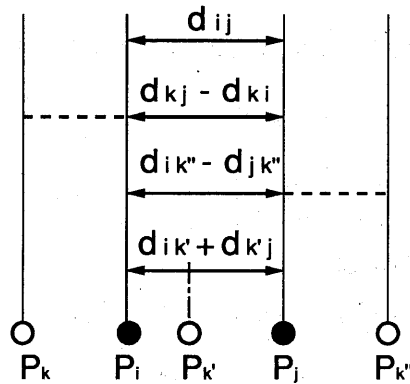


図 4: 1次元の場合の解の意味

8 制約が一様なとき

ここで、すべての2点間に制約が課され、かつその制約条件（2点間の望ましい距離）が等しい場合を考える。すなわち、

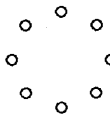
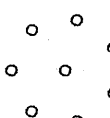
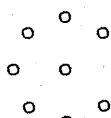
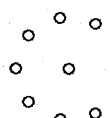
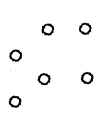
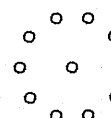
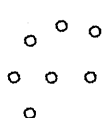
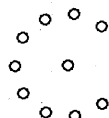
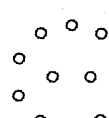
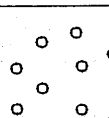
$$d_{ij} = \begin{cases} d & (i \neq j) \\ 0 & (i = j) \end{cases}$$

とする ($1 \leq i \leq n, 1 \leq j \leq n$)。Sun 4上でCを用いたシステムを構築して、このように制約条件が一様なときのシミュレーションを行なった。その結果の一部を図5に示す。図中“ $i; j$ ”は i 角形の中に j 角形があることを表す。したがって、たとえば、“ $(n-2); 2$ ”は中心に2点が入った $(n-2)$ 角形を示す。図中、右下に示した数値は、ランダムな初期値から始めて100回のシミュレーションを行なったとき、そのパターンが出現した頻度を表している。点の数 n が小さいうちは、パターン“ n ”（正 n 角形）が出現するが、次第にその出現比率が減り、代わってパターン“ $(n-1); 1$ ”が増え、次に、パターン“ $(n-2); 2$ ”へと代わってゆくことがわかる。ただし、一般に、同じ“ $i; j$ ”であっても、 i 角形と j 角形との間の位相関係によってさらに分類できるが、その詳細については省略する。なお、このように密に制約が課せられている場合には、計算量は点の数を n とすると n^2 のオーダーになる。

9 動的相互作用表現の情報検索への応用

動的相互作用表現は、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互いに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、利用者がもつ概念をシステムに伝えることができる。

情報検索へ応用するために、ここでは、文献間の「望ましい長さ」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 [2]。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。例として、情報処理学会第38回全国大会におけるニューラルネット関係の論文 [3] を本方式によって表現した結果を図6に示す [4]。図6から関連する文献が互いに近くに集まっていることがわかる。

	"n"	"(n-1) : 1"	"(n-2) : 2"	"(n-3) : 3"
8	 7	 93		
9		 100		
10		 65	 35	
11		 17	 83	
12		 6	 75	 19

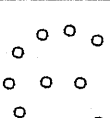
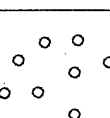
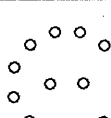
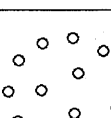
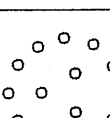
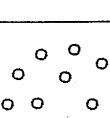
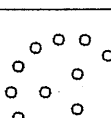
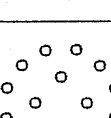
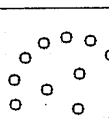
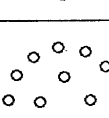
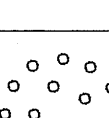
	"(n-2) : 2"	"(n-3) : 3"	"(n-4) : 4"	"(n-5) : 5"
13	 37	 63		
14	 6	 94		
15		 57	 43	
16		 22	 78	
17		 1	 76	 23

図 5: 制約が一樣なときのシミュレーション結果

	BPhidden
BMnクイーン	T&L
BMハースル	SM22
	SONNET 積分
Rmlht	OSCAR 文字列
AN1b	アニメ LANDSAT
AN1c Hpハート	誤差曲面
AN1a Hopf	数字認識
AN1d	最適道路

図 6: 動的相互作用表現の情報検索への応用例 [4]

10 おわりに

並列分散処理を用いて、記号を空間的に表現するための動的相互作用表現を提案し、これを情報検索に応用した。本方式では、記号間の「望ましい長さ」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に記号が空間的に配置されるので、似た概念を表わす記号を空間的近傍に配置することができ、「点」だけでなく、「面」として概念を表現することができる。

本方式は、文献に限らず、広く一般に情報を柔軟に表現するために利用することができるので、ヒューマンインタフェースだけでなく、新しい知識表現形式としても発展させてゆきたい。また、本論文では質量 m 、バネ定数 k 、動摩擦係数 μ は、それぞれ、質点やバネなどによらず同一と仮定している。今後は、これらの定数が異なった一般的な場合を対象にして、理論的解析および実験を解析を進め、より柔軟な方式にしてゆく必要がある。

参考文献

- [1] 堀浩一: "単語の意味の学習について", コンピュータソフトウェア, Vol.3, No.4, pp.65-72 (1986).
- [2] 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).
- [3] 情報処理学会: "第 38 回全国大会講演論文集", (I), (II), (III) (1989).
- [4] 田村淳: "記号の空間的表現について", 情報処理学会第 39 回全国大会講演論文集, pp.251-252 (1989).