2Ｈ－３

確率的探索手法による写像と記号表現の同時探索

甲元　洋[+]　　岡部　洋一[++]
＋日立製作所　日立工場　＋＋東京大学　先端科学技術研究センター

１．はじめに

帰納法を有する問題、たとえば無限級数の恒等式証明では、その両辺の間に存在する帰納的手続きを予め仮説として知る必要がある。その為には、第１ステップとして、両辺の対応写像を探索し、第２ステップとして対応写像を記号化しなければならない。

我々は、２つの集合間に、帰納的手続きが存在する場合に、確率的探索法によって第１ステップの対応写像を導出することを試み、低次項で一致する結果を得た。[3)]

本研究では、第２ステップに相当する写像の記号表現を得るため、問題を木構造データの確率的探索問題として定式化した。形式的べき級数の恒等式両辺に成立する合理的対応写像の記号表現探索を例にして、アルゴリズムの考え方を提示する。

２．対応写像の確率的探索

最初に対応写像の探索を定式化する。

値域を$A=\{x_{A,i}\}$、定義域を$B=\{x_{B,j}\}$とし、$X_{A,\infty}={}^t(x_{A,0},\cdots,x_{A,i},\cdots)$

$X_{B,\infty}={}^t(x_{B,0},\cdots,x_{B,j},\cdots)$とする。

（以下、添字 i, j, k, q は整数とする。）

ここで、$x_{A,i}$、$x_{B,j}$は一般には、無限次元のベクトルである。求めるべき写像は無限次元行列$V_\infty(i,j)$であり次式を満たす。

$$V_\infty(i,j)\ X_{B,\infty}=X_{A,\infty}\quad\cdots\quad(1)$$

行列要素$V_\infty(i,j)$は以下を満たすものとする。すなわち、

$V_\infty(i,j)=1$　　$x_{B,j}$が$x_{A,i}$の構成要素

$V_\infty(i,j)=0$　　$x_{B,j}$が$x_{A,i}$の非構成要素

$\cdots$　(2)

Probabilistic search of reasonable mapping and its symbolic representation.

＋ Hiroshi Kohmoto, Hitachi,Ltd. Hitachi Works, 3-1-1 Saiwai,Hitachi, Ibaraki 317,Japan
＋＋ Youichi Okabe, Research Center for Advanced Science and Technology,The University of Tokyo 4-6-1 Komaba,Meguro,Tokyo 153,Japan

実際の探索は、有限次元のベクトルデータX_AならびにX_Bによって、有限次元行列$V(i,j)$を求めるように定式化される。例題としている形式的べき級数の恒等式を次式に示す。

$$A(x)=1+\frac{x^2}{1-x^2}+\frac{x^6}{(1-x^2)(1-x^4)}+\frac{x^{12}}{(1-x^2)(1-x^4)(1-x^6)}+\cdots$$

$x_{A,0}$　x_{A1}　x_{A2}　x_{A3}

$$B(x)=(1+x^2)(1+x^4)(1+x^6)\cdot\ =$$

$$1+x^2+x^4+x^2\cdot x^4+x^6+x^2\ x^6+\cdots$$

$x_{B,0}\ x_{B,1}\ x_{B,2}$　$x_{B,3}\ x_{B,4}$　$x_{B,5}$

$\cdots$(3)

$V(i,j)$の探索のため、確率的緩和法の一種であるシミュレーテッドアニーリング（以下ＳＡ）を適用する。遷移確率を求めるためのエネルギーは、次のように設定する。

$u=u1+u2+u3+u4$

$u1=b*$(べき級数間のｴﾈﾙｷﾞｰ)

$u2=g*$(写像の束縛ｴﾈﾙｷﾞｰ)

$u3=c*$(相関の束縛ｴﾈﾙｷﾞｰ)

$u4=t*$(記述長の束縛ｴﾈﾙｷﾞｰ)

$\cdots$(4)

ここで、b, g, c, t は調整パラメータである。

第１項のべき級数間のエネルギーは、(限定された)形式的べき級数の集合に導入される対数的な距離として定義される。第２項の写像の束縛エネルギーは、$x_{B,j}$から$x_{A,i}$への写像が一価関数となることの制約条件である。

第１項と第２項を合わせて、全単射すなわちいかなる$x_{B,j}$も$x_{A,i}$の中のいずれかの要素すべてに対応することの制約条件となる。第３項と第４項については、記号表現の確率的探索の中で説明する。

３．記号表現の確率的探索

$V(i,j)$を表す関数の記号表現を得ることが目的である。まず、$V(i,j)$を表す関数の記号

表現方法について定義する。表１は例題に関する4-16行列の場合のV(i,j)の解である。

表１　V(i,j)とタグ要素（○印）

①	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	①	0	①	0	0	0	①	0	0	0	0	0	0	0
0	0	0	①	0	1	①	0	0	1	1	①	0	0	0	0
0	0	0	0	0	0	0	①	0	0	0	0	1	1	①	0

○印を付けた要素は、タグ要素と称して、記号表現の冗長性を低減する目的で導入する。つまり、ある列に１が連続している場合に、先頭の要素のみに着目する。集合Ａの添字をｉ、集合Ｂの添字をｊ、補助添字をｋとする。V(i,j)の記号表現｛R(i,k)｝の元は整数関数であり、あるi,kに対し次式にて定義される。

$$R(i,k)=j(i,k) \qquad \cdots \quad (5)$$

たとえば、式(3)の場合には、求める解は、次のように書かれる。

$$R(i,k)=0(k=0)$$
$$R(i,k)=2^{i-k-1}+\sum_{n=1}^{k-1}{}_{i+k-1}C_n(k\geq 1) \qquad \cdots(6)$$

基本的な考え方は、(6)式を導出するために、(i,j)上の部分集合上で制約条件を満たす関数を生成しつつ、より大きな部分集合上で制約条件を満たす関数が探索されるようなエネルギー関数を設定して、確率的緩和法が適用できるよう定式化することにある。

図１に、探索データの模式図を示す。

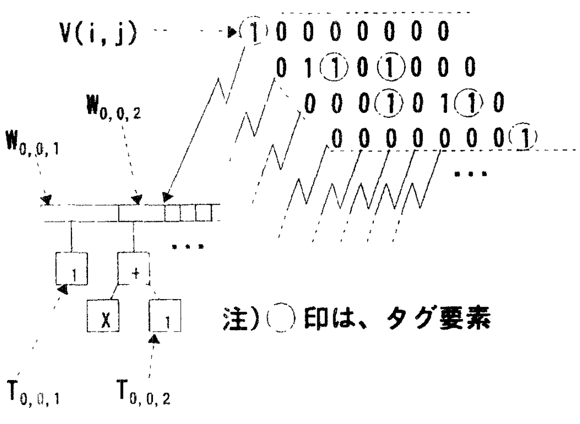

図１　探索データの模式図

ここで、$T_{i,j,q}(i,k)$は V(i,j)＝１を満たすR(i,k)の候補である。また、Ｔの添字のｑは、その優先順位を示すものである。その中で、タグ要素に付随する記号表現の集合｛$T_{i,j,q}(i,k)$｝が、求める記号表現の候補として定義される。

$T_{i,j,q}(i,k)$は、別途設定される重み付け $W_{i,j,q}$ に従って選択され、その後 $T_{i,j,q}(i,k)$自体が更新もしくは保持される。

(4)式のｕ３は、ある$T_{i,j,q}(i,k)$が異なるタグ要素 $V(i_1,j_1)$をも説明できる時、相関項として系のエネルギーを下げる効果を持たせる。また、(4)式のｕ４は、系の記述長を最小にするための項である。ｕ４は、実際の関数記号表現から計算される数値である。

アルゴリズムは、概略以下のようになる。（但し、温度更新ルーチンについては省略）

１）(i_0,j_0)を任意に選ぶ。

２）$V(i_0,j_0)$＝１とする確率を計算する。

a)$W_{i0,j0,q}$に従い、ランダムに$T^{old}_{i0,j0,q}$を選択する。

b)適当な操作で、$T^{new}_{i0,j0,q}$を生成。

c)$V_{new}(i,j)$と$T^{new}_{i0,j0,q}$からエネルギーを計算する。

d)遷移確率を計算する。

３）遷移確率に従ってＶとＴを遷移させる。

４）重み$W_{i,j,q}$を計算し、$T_{i,j,q}$をソートする。

５）終了条件に基づき１）または停止へ。

４．結　論

２つの集合間に帰納的手続きが存在する問題に対して、合理的な対応写像とその記号表現を探索するためのＳＡをベースとした確率的探索手法を構成した。

本手法は、相関と記述長を考慮していることから、合理的な対応写像とその記号表現を探索する能力が期待できる。

５．今後の課題

予備的計算を継続中であり、計算量評価を含めた性能の確認は今後の課題である。

６．参考文献

1)Robert S.Boyer, J Strother Moore : A Computational Logic Handbook (1988)

2)Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images:IEEE Trans. Pattern Anal. Machine Intell. Vol.PAMI-6,pp.721-741 (1984)

3)甲元洋、岡部洋一：確率的探索手法を用いた集合間の合理的な対応写像探索、情報処理学会第54回全国大会 1-389 (1997)