

言語を認識する 1 次元セルオートマトンの漸次学習による合成

中村 克彦[†] 菅家 悠喜[‡]

[†] 東京電機大学理工学部情報システム工学科

[‡] 東京電機大学理工学研究科

1 まえがき

セルオートマトン (CA : cellular automaton) とは同一のセル (順序機械) を規則的に配置, 接続したシステムである. 各セルの内部状態は近傍のセルの状態によって同期して推移する. CA による言語認識は情報科学において重要な研究テーマである. 1 次元 CA による各種の言語の並列 (実時間・線形時間) 認識能力が明らかにされている [2],

この報告では, 形式言語を実時間で受理する 1 次元 CA を学習によって自動生成する方法について述べる. CA の学習には遺伝的アルゴリズム (GA) が多く使われているが, われわれは Balzer [1] による一斉射撃 CA の探索法, および, これまで進めてきた文脈自由文法 (CFG) の漸次学習方式 [3] を応用した方式を用いている. 学習システムは与えられた目的の言語のサンプルである正例と負例から, 必要最小限の推移規則を追加することにより, 言語を認識する 1 次元 CA を生成する.

2 セルオートマトンと言語の実時間認識

2 方向 CA(S) は $S = (K, \#, f, A)$ によって定義される, ここで, K はセル状態の有限集合, $\# \in K$ は境界の状態, $f: K^3 \rightarrow K - \{\#\}$ は局所関数, $A \subseteq K$ は受理状態集合である.

局所関数 f は, $lcr \rightarrow f(l, c, r)$ なる形式の規則の集合によって表される. 1 方向 CA は, すべての $u, v, w, x \in K$ に対して $f(u, v, w) = f(u, v, x)$ なる局所関数 f をもつ CA である.

状相は記号列 $\#u\#, u \in (K - \{\#\})$, で表される. 局所関数 f は次の関係によって状相の推移関数に拡張される.

$$f(\#a_1a_2 \cdots a_n\#) = \#b_1b_2 \cdots b_n\#,$$

Incremental Learning of One-Dimensional Cellular Automata for Recognizing Languages

Katsuhiko NAKAMURA[†] and Yuuki KANKE[‡]

[†]Department of Computers and Systems Engineering, Tokyo Denki University

[‡]Graduate School of Science and Engineering, Tokyo Denki University 350-0394, Saitama, Japan

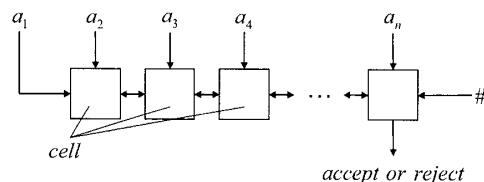


図 1: 2 方向 CA における言語の実時間受理

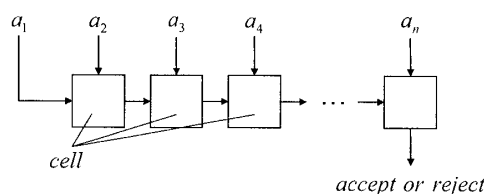


図 2: 1 方向 CA における言語の実時間受理

ここで, $b_i = f(a_{i-1}, a_i, a_{i+1})$, $(i \geq 1, i \leq n)$ かつ, $a_0 = a_{n+1} = \#$ である. さらに, f^n とはすべての状相 c に対して再帰的に $f^0(c) = c$, $f^n(c) = f(f^{n-1}(c))$: $n \geq 1$ と定義される.

CA による言語 $L \subseteq \Sigma^+$ を線形時間で認識するとは, $\Sigma \in K$, かつ, 任意の文字列 $w \in \Sigma^+$, $|w| \geq 2$ に対して,

$$w \in L \leftrightarrow f^{c(|w|-1)}(\#w\#) \in K^+A\{\# \}.$$

になることである (図 1). ただし, c は定数である. $c = 1$ のとき実時間 (real time) で言語を受理するという.

3 CA の学習システム

CA の学習システムは, 正例の記号列の順序集合と負例の記号列集合を入力として, すべての正例を受理し, 負例は受理しないような CA の規則の集合を探索する.

表 1: 1 方向 CA の生成結果

言語	規則数	合成時間 (sec)	GR
$\{a^n b^n n \geq 1\}$	10	7.14	289
括弧言語	31	12	771

【CA 学習の非決定性手続き】この手続きは、正例の順序集合 P_s , 負例の集合 N_s , 状態集合 K に対して、規則の集合 R を返す。

1. 各正例 $p_i \in P_s$ に対して順に、 R の規則の並列的な適用により状態を推移させ p_i が受理可能か否かを判別する。

このとき、ある連続した記号列 l, c, r に対して適用できる規則が R に含まれないならば、次のように規則を作成する。

- (a) $lcr \rightarrow k$ なる規則を R に加える。ただし、 k は状態集合 K から非決定的に選択する。
- (b) R がどの負例も受理しないことを確かめる、もし負例を受理するときは終了（失敗）。

P_i が受理されなければ終了（失敗）。

2. すべての正例を満足する規則集合が発見されれば終了（成功）。

4 CA の学習実験結果

表 1 に 1 方向 CA の学習結果を示す、GR は CA が得られるまでに生成された全規則数であり、処理系に依存しない計算量の尺度として用いている。実験には 2GHz の Intel Core 2 duo プロセッサおよび SWI-Prolog を用いた。

学習された CA における状態推移の例を図 3 に示す。図 4 は入力正例数に対し生成される規則数の関係（学習曲線）を示している。正例は $ab, aabb, aaabbb, aaaabbbb$ のように短い記号列から順に与えた。このグラフは十分な正例と負例を与えると一定の規則集合に収束することを示している。

5 むすび

この報告では、CA の言語学習システムの局所関数集合探索による CA の言語学習方式とその学習結果について述べた。

これからの研究課題として次のものがある。

- 生成時間を短くするための学習方法の改良。

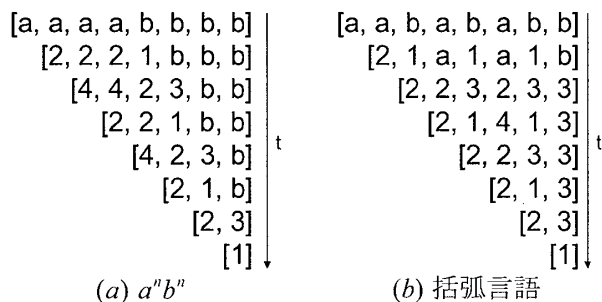


図 3: 学習された 1 方向 CA の推移

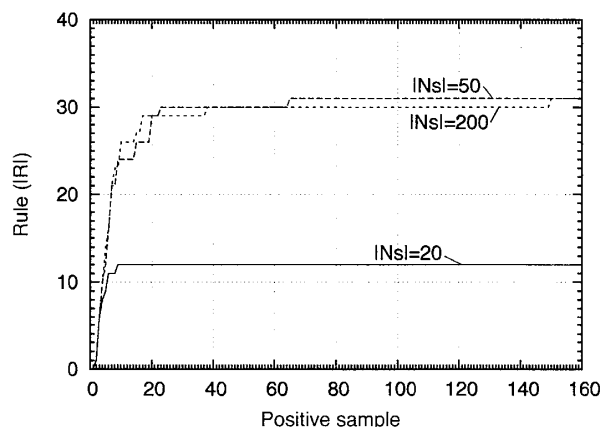


図 4: 括弧言語の CA 合成における正例数に対する規則数の関係

- 本学習方式の 2 次元 CA への拡張、この応用として囲碁の死活パターン認識・学習がある。

参考文献

- [1] R. Balzar, An 8-states minimal time solution to the firing squad synchronization problem., Information and Control, 10, 22-42, 1967.
- [2] K. Nakamura, Languages not recognizable in real time by one-dimensional cellular automata, JCSS 74, 1095-1102, 2008.
- [3] K. Nakamura, Incremental Learning of Context Free Grammars by Bridging Rule Generation and Search for Semi-optimum Rule Sets, LNAI 4201, 72-84, 2006