D1-MACA 기반의 최소 메모리량을 갖는 두 패턴 분류기의 구성

황윤희* ·조성진**· 최언숙***

Construction of Two-Class Classifier based on D1-MACA with minimum memory

Yoon-Hee Hwang* · Sung-Jin Cho** · Un-Sook Choi***

이 논문은 한국과학재단 특정기초연구지원사업(R01-2006-000-10260-0)에 의해 수행하였음.

요 약

분류의 문제는 데이터 베이스 시스템에서 기록을 그룹화, VLSI 회로에서 결함을 찾는 것이나 이미지 프로세싱 등에서 중요한 역할을 하고 있다. 이 논문에서는 주어진 패턴 집합을 부분공간의 개념을 이용하여 최소의 메모리량을 갖는 분류기로써의 D1-MACA를 구성하는 알고리즘을 제안한다. 또한 attractor의 수가 2개가 되게 D1-MACA를 구성할 수 있는 패턴 집합의 조건을 분석한다.

ABSTRACT

Classification problem plays a major role in grouping of the records in database systems, detection of faults in VLSI circuits, image processing, and so on. In this paper, we propose the algorithm constructing D1-MACA as a two-class classifier with minimum memory for given pattern sets using the concepts of subspace. Also we analyze the condition that is designed a two-class classifier D1-MACA with two attractors.

키워드

셀룰라 오토마타, D1-MACA, 패턴 분류기, 부분공간, attractor

* 부경대학교
** 부경대학교(교신저자)
*** 동명대학교

접수일자 2008. 11. 25
심사완료일자 2008. 12. 28

Ⅰ. 서 론

셀룰라 오토마타(Cellular Automata, 이하 CA)는 간단하고 규칙적이며 작은 단위로 확장 연결이 가능한 구조이기 때문에 하드웨어 구현에 적합하여 LFSR(Linear Feedback Shift Register)의 대안으로 제안되었다. 이러한 CA는 상태 전이 그래프에서 모든 상태가 사이클에 놓이는 그룹 CA와 그렇지 않은 비그룹 CA로 나뉜다. 그룹 CA는 테스트 패턴 생성, 의사 난수열 생성, 오류정정부호의 설계 등에 응용되면서 활발하게 연구되어 왔다[1,2]. 또한 비그룹 CA는 이미지 압축, 해쉬 함수, 부울 방정식의 해법 등에 응용되면서 최근 연구가 활발히 이루어지고 있다[3,4]. 특히 비그룹 CA의 상태 전이 그래프는 길이가 1인 사이클을 루트(root)로 갖는 분할된 집합으로 구성된다. 이것은 자연스러운 분류기(classifier)의 형태를 가진다. 분류의 문제는 데이터베이스 시스템에서 기록의 그룹화, VLSI(Very Large Scale Integration) 회로에서의 결함을 찾는 것, 이미지 프로세싱과 같은 컴퓨터 과학에 중요한 역할을 한다. [5]에서는 주어진 패턴 집합이 두 개의 클래스로 분류되는 패턴들을 분류할 수 있는 분류기를 구성하는 방법으로 D1-MACA (Depth 1 Multiple Attractor CA)를 사용하고 있다. D1-MACA에 기반한 분류기는 기존의 방법보다 저장해야 하는 메모리량이 적은 장점을 가지고 있으나 [5]에서는 이러한 CA를 구성하는 방법이 BDD(Binary Decision Diagram)을 이용하기 때문에 다소 복잡하고, 주어진 패턴 클래스에 대하여 메모리량을 최소로 할 수 없다는 단점과 어떤 패턴의 경우에 분류기로써의 역할을 못하고 있다.

이 논문에서는 주어진 패턴 집합이 두 개의 클래스로 분할된 패턴들을 분류할 수 있는 분류기를 여러 개의 attractor를 갖고 depth가 1인 D1-MACA와 부분공간의 개념을 이용하여 주어진 패턴 클래스에 대하여 메모리 요구량이 최소가 되게 효율적으로 구성한다. 또한 attractor의 수가 2개 가 되어 메모리량이 가장 작게 되는 D1-MACA를 구성할 수 있는 패턴 집합의 조건을 분석한다.

Ⅱ. 셀룰라 오토마타와 부분공간

이 절에서는 본 논문에서 사용되는 CA의 용어, 기본 성질과 부분공간의 개념에 대하여 간단히 언급한다.

<정의 1> attractor란 비그룹 CA의 상태 전이 그래프에서 순환상태들 중 사이클의 길이가 1인 상태를 말한다[6].

<정의 2> depth란 비그룹 CA의 상태 전이 그래프에서 임의의 도달 불가능한 상태에서 가장 가까운 순환상태로 가는데 걸리는 최소의 단계 수를 말한다[6].

<정리 1> 상태 전이 행렬 T를 갖는 n-셀 MACA에 대하여 $T\oplus I$의 차원이 r일 때 attractor의 수는 2^{n-r}개이다[6].

<정리 2> 2^m개의 attractor를 갖는 CA에서 각 attractor에서 2^m개의 의사 전수 패턴(pseudo-exhaustive pattern)을 갖는 비트 위치가 존재한다[6].

<예제 1> 4-셀 CA의 상태 전이 행렬이 다음과 같을 때 $T\oplus I$의 차원은 2이다. 따라서 2^2개의 attractor를 가짐을 알 수 있다.

$$T=\begin{pmatrix}1100\\1100\\0010\\0001\end{pmatrix},\quad T\oplus I=\begin{pmatrix}0100\\1000\\0000\\0000\end{pmatrix}$$

이 CA의 상태 전이 그래프를 그리면 그림 1과 같음을 확인할 수 있다.

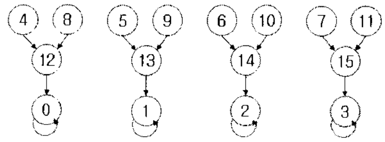

그림 1. MACA의 상태 전이 행렬
Fig.1. State-transition matrix of MACA

여기서 attractor 집합은 $\{0, 1, 2, 3\}$로 2-비트의 의사 전수 패턴을 세 번째 비트와 네 번째 비트를 택함으로써 다음과 같이 생성할 수 있다.

0000
0001
0010
0011

depth가 1인 MACA를 D1-MACA라 하는데 이러한 CA는 분류기를 디자인 하는데 유용하다. 왜냐하면 D1-MACA는 모든 도달 불가능한 상태들이 CA를 돌렸을 때 한 번 만에 attractor에 도달할 수 있기 때문에 분류 시 필요한 시간을 줄일 수 있다.

<정리 3> depth가 d인 MACA에 대하여 D1-MACA가 존재한다[5].

depth가 d인 MACA의 상태 전이 행렬이 T일 때 D1-MACA의 상태 전이 행렬은 T^d가 된다. 예를 들어 예제 1에서 depth가 2이므로 T^2을 상태 전이 행렬로 둔다면 그림 2의 상태 전이 그래프를 갖는 D1-MACA가 된다.

$$T^2 = \begin{pmatrix} 0&0&0&0 \\ 0&0&0&0 \\ 0&0&1&0 \\ 0&0&0&1 \end{pmatrix}$$

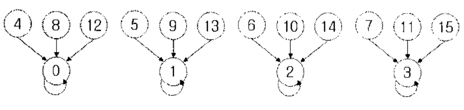

그림 2. D1-MACA의 상태 전이 그래프
Fig.2. State transition diagram of D1-MACA

<정의 3> 집합 $V \subseteq \{0,1,2,\cdots,2^n-1\}$에 대하여, 임의의 $\boldsymbol{x}, \boldsymbol{y} \in V$가 $\boldsymbol{x} \oplus \boldsymbol{y} \in V$일 때, $\oplus$(XOR)이 정의된 연산체계 V를 벡터공간(vector space)라 한다[7].

<정의 4> 집합 V의 부분집합 S가 V에서 정의된 $\oplus$에 대하여 벡터공간이 될 때, S를 V의 부분공간(subspace)라 한다[7].

<예제 2> $S=\{0,1,2,3,4\} \subset \{0,1,2,\cdots,15\}$는 S의 원소 1과 4를 $\oplus$하면 $0001 \oplus 0100 = 0101(=5) \not\subset S$이므로 부분공간이 아니다. $S' = \{0,1,2,3\} \subset \{0,1,2,\cdots,2^4-1\}$에 대하여 S'의 임의의 두 원소 x, y에 대하여 아래의 표에서 보는 바와 같이 $x \oplus y \in S'$이므로 S'는

$\oplus$	0	1	2	3
0	0	1	2	3
1	1	0	3	2
2	2	3	0	1
3	3	2	1	0

$\{0,1,2,\cdots,2^4-1\}$의 부분공간이다.

Ⅲ. D1-MACA 기반의 패턴 분류기

m개의 attractor를 갖는 n-비트 MACA는 분류기로 볼 수 있다. 각 attractor에 대한 트리의 상태들은 m개의 분할된 분류로 나눠진다. 그림 3에서와 같이 의사 전수 패턴을 생성하는 비트 위치의 값만으로 그 패턴의 클래스를 구별할 수 있다.

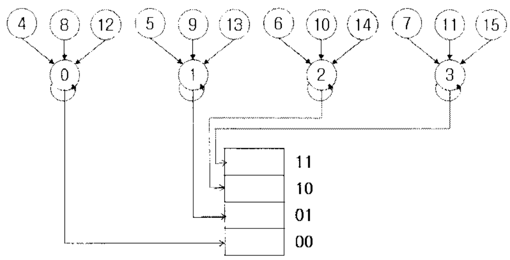

그림 3. MACA 기반의 분류기
Fig.3. Classifier based on a MACA

주어진 패턴 집합 P가 단지 두 개의 분할된 패턴 분류 P_1와 P_2로 구성된다고 하자. 즉, $P=P_1 \cup P_2$이고 $P_1 \cap P_2 = \varnothing$이 성립한다. 만약 분류기가 P를 분류한다면, 주어진 두 패턴 $x \in P_1$, $y \in P_2$은 다른 트리에 놓이게 된다. 직관적인 접근 방법으로 생각해 보자. 주어진 n-비트 패턴 집합 P에 대하여 P_1에 속한 패턴의 개수가 n_1개이고, P_2에 속한 패턴의 개수가 n_2개라면 이들을 분류하기 위한 메모리량은 다음과 같다.

$$M_{SFA} = \min(n_1, n_2) \times n + 1$$

그렇다면 D1-MACA의 경우를 생각해 보자. 만약 T 가 위의 조건을 만족하는 분류를 수행하는 D1-MACA의 상태 전이 행렬이면 다음의 두 관계를 만족하여야 한다.

관계 1> P_1의 임의의 원소 x와 P_2의 임의의 원소 y에 대하여 $T \bullet (x) \neq T \bullet (y)$이 성립하여야 한다. 즉, $T \bullet (x \oplus y) \neq O$이다.

관계 2> depth가 1이므로 $T^2 = T$이여야 한다. 즉, $T(T \oplus I) = O$이 성립하여야 한다.

위 두 관계를 만족하는 D1-MACA를 구성하는 것에는 각 패턴이 어떻게 구성되는 상황에 의하여 2개 이상의 attractor를 가지게 된다. 만약 P_1의 원소가 D1-MACA의 a_1개의 트리에 놓여있고, P_2의 원소가 D1-MACA의 a_2개의 트리에 놓여있다고 하자. 그 때 의사 전수 패턴을 생성하는 m비트가 있다고 가정한다면 요구되는 메모리량 다음과 같다[5].

$$M_{CA} = \min(a_1, a_2) \times m + 1$$

이는 M_{SFA}에 비하여 작은 량이다.

이 때 메모리를 가장 적게 차지하는 분류기를 가장 선호하게 되는데 이는 attractor가 두 개로만 구성된 D1-MACA일 경우가 된다.

주어진 패턴 집합에 대하여 attractor가 두 개로만 구성된 D1-MACA를 구성할 수 있는 패턴 클래스의 조건은 다음과 같다.

<정리 4> 주어진 패턴 집합 P에 대하여 두 클래스가 P_1과 P_2일 때

$$\{0,1,2,\cdots,2^n - 1\} - \{P_1 \oplus P_2\}$$

이 부분공간이면 P_1, P_2를 attractor가 2개인 D1-MACA로 분류할 수 있다.

위의 정리를 만족하지 않는다면 attractor의 수는 2를 넘게 된다.

위에서 언급한 관계1과 관계2를 성립하는 D1-MACA를 구성하기 위하여 [5]에서는 $P_1 \oplus P_2$의 원소들이 T의 영공간에 속하지 않게 BDD를 이용하여 T를 구성하고 있으므로 메모리량이 최소가 되는 D1-MACA를 구성하는 것을 보장하지 못한다. 다음 절에서는 이를 보장하는 D1-MACA를 부분공간의 개념을 사용하여 구성한다.

Ⅳ. 최소의 메모리량을 갖는 분류기

주어진 패턴 집합 P가 P_1, P_2들의 클래스로 분할될 때 $P_1 \oplus P_2$을 생각해 보자. 여기서 $P_1 \oplus P_2$의 원소들은 관계 1에 의하여 구성되어질 패턴 분류기로써의 T의 영공간이여서는 안된다. $(P_1 \oplus P_2)^c := K$는 패턴 O(영벡터)를 반드시 가지고 있으며 K의 원소들은 영공간에 속해도 되고 속하지 않아도 된다. 따라서 K의 원소들 중 원소의 개수가 최대가 되는 부분공간 K'을 잡는다. 그러면 0-트리에 들어가는 원소를 최대로 할 수 있다. 즉, 각 트리에 속하는 원소를 최대로 할 수 있어 각 패턴 클래스들의 원소들이 놓이는 트리의 수도 줄고 각 attractor에서 뽑아오는 의사 전수 패턴의 비트수가 줄게 되어 요구되어지는 메모리 량이 최소가 된다. 다음을 이를 바탕으로 구성한 알고리즘이다.

표 1. 최소 메모리 량을 갖는 두 패턴 분류기 알고리즘
Table 1. Algorithm of construction of two-Class Classifier with minimum memory

알고리즘 : 최소 메모리 요구량을 갖는 두 패턴 분류기 입력 : 주어진 두 패턴 클래스 P_1, P_2 출력 : 두 패턴 분류기 T Step 1. $P_1 \oplus P_2$를 구한다. Step 2. $(P_1 \oplus P_2)^c := K$를 구한다. Step 3. 원소가 최대가 되게 부분공간 K'를 구한다. Step 4. K'가 영공간이 되는 T를 구성한다.

<예제 3> 주어진 4-비트 패턴 집합 P 는 다음 과 같은 분할된 두 패턴 분류 P_1, P_2 로 구성되어 있다고 가정하자.

$$P=\{0,1,2,5,6,7,10,11,15\}$$
$$P_1=\{0,1,6,7,15\}$$
$$P_2=\{2,5,10,11\}$$

step1) $P_1 \oplus P_2=\{2,3,4,5,10,11,12,13\}$을 구한다.

step2) $K=\{0,1,2,\cdots,2^4-1\}-(P_1\oplus P_2)=(P_1\oplus P_2)^c$
$=\{0,1,6,7,8,9,14,15\}$를 구한다.

(여기서, K가 부분공간이므로 정리 4에 의하여 2개의 attractor를 갖는 D1-MACA를 분류기로 구성할 수 있다.)

step3) K 의 원소 중에서 원소의 개수가 최대가 되는 부분공간 K'를 택한다.

$$K'=\{0,1,6,7,8,9,14,15\}$$

step4) K'를 영공간으로 가지는 D1-MACA의 상태 전이 행렬 T_1를 다음과 같이 구성한다.

$$T_1=\begin{pmatrix}0000\\0110\\0000\\0000\end{pmatrix}$$

그림 4는 주어진 패턴을 분류하는 T_1를 상태 전이 행렬로 갖는 D1-MACA의 상태 전이 그래프이다.

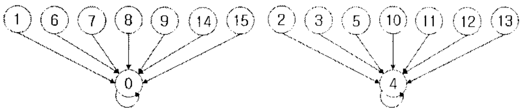

그림 4. P_1과 P_2를 분류한 D1-MACA
Fig.4. D1-MACA classified P_1 and P_2

여기서 T는 다음과 같이 구성하여도 된다.

$$T=\begin{pmatrix}0000\\0000\\0110\\0000\end{pmatrix},\quad T=\begin{pmatrix}0000\\0110\\0110\\0000\end{pmatrix}$$

여기서 요구되는 메모리 량은 직관적인 방법의 메모리 요구량인 $M_{SFA}=\min(5,4)\times 4+1=17$에 비하여 D1-MACA의 메모리 요구량은 $M_{CA}=\min(1,1)\times 1+1=2$로 줄어든다.

표 2는 본 논문에서 제안된 알고리즘에 의하여 구성된 패턴 분류기의 메모리 량의 실험 결과 및 [5]의 알고리즘에 의해 얻어진 패턴 분류기의 요구되는 메모리 량과의 비교이다.

표 2에서 M_{SFA}는 사전식 방법에 의하여 패턴 분류기를 구성했을 때 요구되는 메모리 량이며, M_{CA1}은 [5]에서 제안된 알고리즘에 의해 합성된 패턴 분류기의 요구되는 메모리 량이고, M_{CA2}는 본 논문에서 제안된 알고리즘에 의해 합성된 패턴 분류기의 요구되는 메모리 량이다.

표 2. 요구 메모리 량 비교
Table 2. Comparison of memories required

pattern size	# of patterns in class-1 (n_1)	# of patterns in class-2 (n_2)	MSFA (bits)	MCA1 (bits)	MCA2 (bits)
7	18	14	99	79	56
8	17	15	121	106	46
9	16	25	145	61	41
10	15	25	151	217	79
11	15	25	166	118	73
12	15	25	181	127	79
13	35	50	456	265	337
14	35	50	491	265	197
15	15	25	226	79	41

V. 결 론

이 논문에서는 주어진 패턴 집합이 두 개의 클래스로 분류되는 패턴들을 분류할 수 있는 분류기를 여러 개의 attractor를 갖고 depth가 1인 D1-MACA와 부분공간의 개념을 이용하여 메모리 요구량을 최소가 되게 효율적으로 구성하는 알고리즘을 제안하였다. 또한 메모리 요구량이 가장 작은 경우인 attractor의 수가 2개 되게 D1-MACA를 구성할 수 있는 패턴 집합의 조건을 분석하였다.

참고문헌

[1] P.H. Bardell, “Analysis of cellular auto- mata used as pseudorandom pattern generators,” *Proc. IEEE int. Test. Conf.*, pp. 762~767, 1990.

[2] S.U. Guan and S.K. Tan, "Pseudorandom number generation with self-programmable cellular automata," *IEEE Transaction on computer-aided design of integrated circuits and systems,* 23(7), pp. 1095-1101, 2004.

[3] S. Chakraborty, D.R. Chowdhury and P.P. Chaudhuri, "Theory and application of nongroup cellular automata for synthesis of easily testable finite state machines," *IEEE, Trans, Computers,* 45, pp. 769-781, 1996.

[4] S.J. Cho, U.S. Choi, H.D. Kim, Y.H. Hwang and J.K, Kim, "Analysis of 90/150 two predecessor nongroup cellular automata," *LNCS* 5191, pp. 128-135, 2008.

[5] S. Chattopadhyay, S. Adhikari, S. Sengupta and M. Pal, "Highly regular, modular, and cascadable design of cellular automata-based pattern classifier," *IEEE Transactions on VLSI systems,* 8(6), pp 724-734, 2000.

[6] P.P. Chaudhuri, D.R. Chowdhury, S. Nandy and Chattopadhyay, *Additive cellular automata theory and application, 1,* IEEE Computer Society Press, California, 1997.

[7] R.A. Horn and C.R. Johnson, *Matrix analysis,* Cambridge university Press, 1985.

[8] S.J. Cho, H.D. Kim and U.S. Choi, "Behavior of complemeted cellular automata derived from a linear cellular automata," *Mathematical and Computer Modelling,* 36, pp. 979-986, 2002.

저자소개

황 윤 희 (Yoon-Hee Hwang)

2002년 2월 : 부경대학교 통계학과 학사
2004년 2월 : 부경대학교 응용수학과 석사
2008년 8월 : 부경대학교 정보보호학과 박사
※관심분야 : 셀룰라 오토마타론, 정보보호, 유한체, 컴퓨터 구조론, VLSI

조 성 진 (Sung-Jin Cho)

1979년 2월: 강원대학교 수학교육과 학사
1981년 2월: 고려대학교 수학과 석사
1988년 2월: 고려대학교 수학과 박사
1988년 ~ 현재 : 부경대학교 수리과학부 정교수
※ 관심분야 : 셀룰라 오토마타론, 정보보호, 부호이론, 컴퓨터 구조론, VLSI

최 언 숙 (Un-Sook Choi)

1992년 성균관대학교 산업공학과 학사
2000년 부경대학교 응용수학과 석사
2004년 부경대학교 응용수학과 박사
2004년~2006년 2월 영산대학교 자유전공학부 단임교수
2006년 3월~현재 동명대학교 멀티미디어공학과 전임강사
※ 관심분야 : 셀룰라 오토마타론, 정보보호, 부호이론, 컴퓨터 구조론, VLSI