다중 끌개를 갖는 셀룰라 오토마타를 이용한 패턴 분류기 생성

황윤희[*] · 조성진[**] · 최언숙[***]

Multiple Attractor CA Based Pattern Classifier

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

요 약

다중 클래스로 이루어진 패턴을 분류하는 것은 데이터 베이스 시스템에서 기록을 그룹화하거나 VLSI 회로에서 어디에 결함이 있는지를 찾는 것 등에서 중요한 역할을 한다. 본 논문에서는 주어진 다중 클래스 패턴을 MACA(Multiple Attractor Cellular Automata)와 부분공간의 개념을 이용하여 가능한 최소 메모리량을 필요로 하는 다중 클래스 패턴 분류기를 구성하는 알고리즘을 제안한다.

ABSTRACT

Classifying multi-class pattern plays an important role in grouping of records in database systems, detection of faults in the VLSI circuits and so on. In this paper, we propose an algorithm for the construction of multi-class pattern classifier with minimum memory capacity using MACA(Multiple Attractor Cellular Automata) and the subspace concept for given multi-class patterns.

키워드

MACA(Multiple Attractor Cellular Automata), Multi class pattern classifier, Subspace, Attractor

Ⅰ. 서 론

셀룰라 오토마타(Cellular Automata,이하 CA)는 1986년 말에 Wolfram에 의하여 처음으로 의사 난수열 생성기로 사용되었다[1]. 이러한 CA는 국소적 상호작용에 의해 상태가 전이되고 간단하고 규칙적이며 작은 단위로 확장 연결할 수 있는 구조를 가지고 있어 VLSI 구현에 적합하며 LFSR에 비하여 난수성이 뛰어나다[2]. CA는 상태를 전이를 시켰을 때 모든 상태가 사이클 상에 놓이는 그룹 CA와 그렇지 않은 비그룹 CA로 나뉜다[3]. 그룹 CA는 테스트 패턴 생성, 의사 난수열 생성 등에 응용되는 CA이며 특히 Cho 등은 주어진 기약다항식에 대한 90/150가 2개가 존재함을 밝혔다 [4]. 비그룹 CA는 이미지 압축, 해쉬 함수 등에 응용되고 있다[5-8]. Depth가 1인 MACA (D-1 Multiple Attractor CA)는 비그룹 CA의 한 분류로 상태 전이 그래프가 길이가 1인 사이클을 루트(root)로 갖는 분할된 집합으로 구성된다. 이것은 자연스러운 분류기(classifier)의 형태를 가진다. 다중 클래스로 이루어진 패턴을 분류하는 것은 데이터베이스 시스템에서 기록의 그룹화, VLSI(Very Large Scale Integration) 회로에서의 결함을 찾는 것 등과 같은 컴

* 부경대학교 응용수학과(yhhwang@pknu.ac.kr) ** 교신저자 : 부경대학교 응용수학과(sjcho@pknu.ac.kr)
*** 동명대학교 미디어공학과
접수일자 : 2010. 03. 09 심사(수정)일자 : 2010. 04. 30 게재확정일자 : 2010. 06. 14

퓨터 과학에 중요한 역할을 한다. [9]에서는 두 개의 클래스로 이루어진 패턴을 두 개의 클래스로 분류하는 분류기를 구성하는 방법으로 D-1 MACA를 사용하고 있다. [9]에서 제안한 패턴 분류기는 D-1 MACA에 기반으로 하여 기존의 방법보다 분류 시 저장해야 하는 메모리 량이 비교적 적은 장점을 가지고 있으나 이러한 D-1 MACA를 구성하는 방법이 BDD(Binary Decision Diagram)을 이용하기 때문에 복잡하고, 주어진 패턴 클래스에 대하여 메모리 량을 최소로 할 수 없다는 단점과 어떤 패턴의 경우에 분류기로써의 역할을 못하고 있다. 또한 세 클래스 이상으로 분류된 패턴을 분류할 수 있는 분류기도 두 클래스 패턴 분류기를 이용함으로써 앞에서 언급한 단점을 벗어나지 못한다.

이 논문에서는 주어진 패턴 집합이 세 개 이상의 클래스로 분할된 패턴들을 분류할 수 있는 다중 클래스 패턴 분류기를 D-1 MACA와 부분공간의 개념을 이용하여 주어진 패턴 클래스에 대하여 메모리 요구량이 최소가 되도록 구성하는 알고리즘을 제안한다.

II. 배경지식

이 절에서는 논문 전개에 필요한 배경지식으로 CA와 부분공간에 대하여 간단히 소개한다.

2.1. CA

비그룹 CA의 상태 전이 그래프에서 순환상태들 중 사이클의 길이가 1인 상태를 attractor라 하고, 임의의 도달 불가능한 상태에서 가장 가까운 순환상태로 가는데 걸리는 최소의 단계 수를 depth라 한다[3].

상태 전이 행렬 T를 갖는 n-셀 MACA에 대하여 $T\oplus I$의 차원 $\dim(T\oplus I)$이 r일 때 attractor의 수는 2^{n-r}개이고, 2^m개의 attractor를 갖는 CA에서 각 attractor에서 2^m개의 의사 전수 패턴(pseudo-exhaustive pattern)을 갖는 비트 위치가 존재한다[3]. 예를 들어 4-셀 CA의 상태 전이 행렬이 다음과 같을 때 $\dim(T\oplus I)=2$이다. 따라서 $2^{4-2}=4$개의 attractor를 가짐을 알 수 있다. 예를 들어 어떤 CA의 규칙이 <60, 204, 102, 60>라 하면 상태전이 행렬 T와 $T\oplus I$는 다음과 같다.

$$T=\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&1\\0&0&1&1\end{pmatrix},\quad T\oplus I=\begin{pmatrix}0&0&0&0\\0&0&0&0\\0&0&0&1\\0&0&1&0\end{pmatrix}$$

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

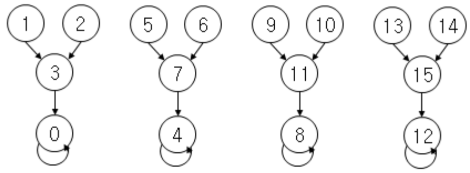

그림 1. MACA의 상태 전이 행렬
Fig.1 State-transition matrix of a MACA

여기서 attractor는 $\{0,4,8,12\}$로 2-비트의 의사 전수 패턴을 첫 번째 비트와 두 번째 비트를 택함으로써 다음과 같이 생성할 수 있다.

$$\underline{00}00$$
$$\underline{01}00$$
$$\underline{10}00$$
$$\underline{11}00$$

depth가 1인 D-1 MACA는 모든 도달 불가능한 상태들이 CA를 전이시켰을 때 한 번 만에 attractor에 도달할 수 있기 때문에 분류 시 필요한 시간을 줄일 수 있으므로 패턴 분류기를 디자인 하는데 유용하다.

depth가 d인 MACA에 대하여 D-1 MACA가 존재한다[9]. 즉, depth가 d인 MACA의 상태 전이 행렬이 T일 때 D-1 MACA의 상태 전이 행렬은 T^d가 된다. 예를 들어 그림 1에서 depth가 2이므로 T^2을 상태 전이 행렬로 둔다면 그림 2의 상태 전이 그래프를 갖는 D-1 MACA가 된다.

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

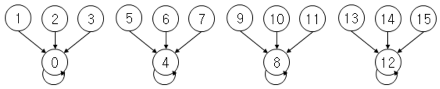

그림 2. D-1 MACA의 상태 전이 그래프
Fig. 2 State transition diagram of a D-1 MACA

2.2. 부분공간

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 하고, 집합 V의 부분집합 S가 V에서 정의된 $\oplus$에 대하여 벡터공간이 될 때, S를 V의 부분공간(subspace)이라 한다[10].

예를 들어 $S=\{0,1,3,5,7\} \subset \{0,1,2,\cdots,15\}$라 하면 $0001 \oplus 0101 = 0100 (=4) \not\in S$이므로 S는 전공간 $\{0,1,2,\cdots,15\}$의 부분공간이 아니다. $S'=\{0,1,4,5\} \subset \{0,1,2,\cdots,15\}$에 대하여 S'의 임의의 두 원소 x,y에 대하여 아래의 표에서 보는 바와 같이 $x \oplus y \in S'$이므로 S'는 $\{0,1,2,\cdots,15\}$의 부분공간이다.

$\oplus$	0	1	4	5
0	0	1	4	5
1	1	0	5	4
4	4	5	0	1
5	5	4	1	0

III. D-1 MACA 기반의 다중 클래스 패턴 분류기

m개의 attractor를 갖는 D-1 MACA에서 각 attractor에 대한 트리의 상태들은 m개로 분류된다. 그림 3에서와 같이 의사 전수 패턴을 생성하는 비트 위치의 값만으로 그 패턴의 클래스를 구별할 수 있다.

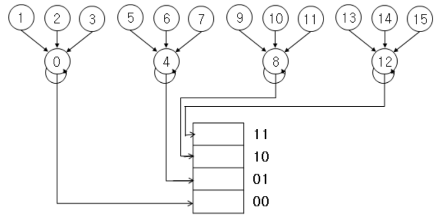

그림 3. MACA 기반의 분류기
Fig. 3 Classifier based on a MACA

주어진 n-비트 패턴 P가 세 개 이상의 클래스로 분할되는 경우, 즉

$$P = P_1 \cup P_2 \cup \cdots \cup P_k \ (k \geq 3)$$

이라 하자. 그러면 임의의 $i,j \in \{1,2,\cdots,k\}$에 대하여 다음 두 가지 경우를 고려하여야 한다.

i) $P_i \cap P_j = \varnothing$
ii) $P_i \cap P_j \neq \varnothing$

먼저 임의의 두 클래스 P_i와 P_j가 서로 같은 원소를 갖지 않을 때, 즉 $P_i \cap P_j = \varnothing$일 때 이 두 클래스를 분류하기 위한 패턴 분류기 T는 다음을 만족하여야 한다.

관계 1> P_1의 임의의 원소 $\boldsymbol{x}$와 P_2의 임의의 원소 $\boldsymbol{y}$에 대하여 $T \bullet \boldsymbol{x} \neq T \bullet \boldsymbol{y}$이 성립하여야 한다. 즉, $T \bullet (\boldsymbol{x} \oplus \boldsymbol{y}) \neq O$이다.
관계 2> depth가 1이므로 $T^2 = T$이여야 한다. 즉, $T(T \oplus I) = O$이 성립하여야 한다.

위에서 언급한 관계1과 관계2가 성립하는 D-1 MACA를 구성하기 위하여 [8]에서는 $P_1 \oplus P_2$의 원소들이 T의 영공간에 속하지 않게 BDD를 이용하여 T를 구성하고 있으므로 메모리량이 최소가 되는 D-1 MACA를 구성하는 것을 보장하지 못한다. 주어진 임의의 두 클래스 P_i와 P_j들을 분할될 때 $X := P_i \oplus P_j$를 생각해 보자. 여기

서 $P_i \oplus P_j$의 원소들은 관계 1에 의하여 구성되어질 패턴 분류기로써의 T의 영공간이여서는 안된다. 따라서 $(P_i \oplus P_j)^c := K$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 잡으면, O-트리에 들어가는 원소들의 개수를 최대가 되게 할 수 있다. 즉, 각 트리에 속하는 원소들의 개수 최대가 되게 할 수 있어서 각 패턴 클래스들의 원소들이 놓이는 트리의 수도 줄고 각 attractor에서 뽑아오는 의사 전수 패턴의 비트수가 줄게 되어 패턴 분류 시 요구되는 메모리 량이 최소가 된다. 또한 임의의 두 클래스 P_i와 P_j가 서로 같은 원소를 갖고 있을 때, 즉 $P_i \cap P_j \neq \varnothing$일 때도 P_i와 P_j을

$$P_i{}' = P_i - (P_i \cap P_j)$$
$$P_j{}' = P_j - (P_i \cap P_j)$$
$$P_{c_1} = P_i \cap P_j$$

로 분할하면 $P_i' \cap P_j' \cap P_{c_1} = \varnothing$이 된다. 따라서 앞에서 언급한 방법으로 패턴 분류기를 구성할 수 있다. 표 1은 위의 내용을 토대로 구성한 알고리즘이다.

예를 들어 다중 클래스가 다음과 같다고 하자.

$$P_1 = \{2,8,17\}$$
$$P_2 = \{2,4,9,20\}$$
$$P_3 = \{2,9,20\}$$

그러면 먼저 각각의 클래스에 겹치는 원소가 없도록 다음과 같이 다시 구성한다.

$$P_1{}' = \{8,17\},\ P_2{}' = \{4\},$$
$$P_{c_1} = \{2\},\ P_{c_2} = \{9,20\}$$

위에서 구성한 클래스를 이용하여 다음을 계산한다.

$$P_1{}' \oplus P_2{}' = \{12,21\},\quad P_1{}' \oplus P_{c_1} = \{10,19\},$$
$$P_1{}' \oplus P_{c_2} = \{1,5,24,28\},\quad P_2{}' \oplus P_{c_1} = \{6\},$$
$$P_2{}' \oplus P_{c_2} = \{13,16\},\quad P_{c_1} \oplus P_{c_2} = \{11,22\}$$

위의 6개의 집합들의 다 합한 집합을 X라 하면 X와 K는 다음과 같다.

$$X = \{1,5,6,10,11,12,13,16,19,21,22,24,28\}$$
$$K = \{0,2,3,4,7,8,9,14,15,17,18,20,23,25,,26,27,29,30,31\}$$

따라서 $K' = \{0,7,8,15\}$이고 이를 이용하여 다음과 같이 분류할 수 있다.

표 1. D-1 MACA 기반의 최소 메모리량을 요구하는 다중 클래스 패턴 분류기 구성 알고리즘
Table 1. Algorithm of construction of Multi-Class Pattern Classifier with minimum memory

알고리즘 :
D-1 MACA 기반의 다중 클래스 패턴 분류기

입력 : 주어진 다중 클래스 패턴 $P_1, P_2, \cdots, P_k$
출력 : 다중 클래스 패턴 분류기 T

Step 1. 각각이 클래스의 교집합이 없게 $P_i{}'$ 와 P_{C_i}를 구한다.
Step 2. $X_i := A \oplus B$,$(i = 1,2,\cdots)$를 구한다.
(여기서 A, B 는 Step 1에서 구한 클래스들이다.)
Step 3. $X := \bigcup_i X_i$ 를 구한다.
Step 4. $K := \{0,1,\cdots,2^n - 1\} - X$를 구한다.
Step 5. K 의 원소 들 중 원소가 최대가 되게 부분공간 K'를 구한다.
Step 6. K'를 이용하여 $\{0,1,\cdots,2^n - 1\}$를 분할한다.
Step 7. 분할된 분류를 이용하여 T 를 구성한다.

$[0]=\{0,7,8,15\}$, $[1]=\{1,6,9,14\}$
$[2]=\{2,5,10,13\}$, $[3]=\{3,4,11,12\}$
$[16]=\{16,23,24,31\}$, $[17]=\{17,22,25,30\}$
$[18]=\{18,21,26,29\}$, $[19]=\{19,20,27,28\}$

위와 같이 분류할 수 있는 다중 클래스 패턴 분류기 T는 다음과 같다.

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

그림 4는 T에 대한 상태 전이 그래프이다. 그림 4에서와 같이 각각의 클래스들이 다른 트리에 놓여 분류하고 있음을 알 수 있다.

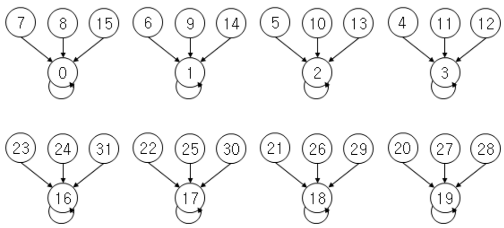

그림 4. T의 상태 전이 그래프
Fig. 4 State transition diagram of T

Ⅳ. 결 론

본 논문에서는 주어진 다중 클래스 패턴을 MACA와 부분공간의 개념을 이용하여 가능한 최소 메모리량을 필요로 하는 다중 클래스 패턴 분류기를 구성하는 알고리즘을 제안하였다.

감사의 글

이 논문은 2009학년도 부경대학교의 지원을 받아 수행된 연구임(PK-2009-26)

참고 문헌

[1] S. Wolfram, O. Martin and A. M. Odlyzko, “Algebraic properties of cellular automata,” Communications in Mathematical Physics, 3, pp. 219-258, 1984.
[2] P. Sarkar, "A brief history of cellular automata", ACM Comput. Surveys, vol. 32, no. 1, pp. 80-107, 2000.
[3] P. P. Chaudhuri, D. R. Chowdhury, S. Nandi, and S. Chattopadhyay, Additive Cellular Automata: Theory and Applications. Los Alamitos, CA: IEEE CS Press, vol. 1., 1997.
[4] S.J. Cho, U.S. Choi, H.D. Kim, Y.H. Hwang, J.G. Kim and S.H. Heo, New Synthesis of One-Dimensional 90/150 Linear Hybrid Group Cellular Automata, IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst., Vol. 26, No. 9, pp. 1720-1724, Sep. 2007.
[5] P.H. Bardell, “Analysis of cellular auto- mata used as pseudorandom pattern generators," *Proc. IEEE int. Test. Conf.*, pp. 762-767, 1990.
[6] 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.
[7] 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.
[8] 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.
[9] 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.
[10] R.A. Horn and C.R. Johnson, Matrix analysis, Cambridge university Press, 1985.

저자 소개

황윤희(Yoon-Hee Hwang)

2002년 2월 부경대학교 통계학과 졸업 (이학사)
2004년 2월 부경대학교 대학원 응용수학과 졸업(이학석사)
2008년 8월 : 부경대학교 대학원 정보보호학과 졸업 (공학박사)
※ 주관심분야 : 셀룰라 오토마타론, 정보보호, 유한체, 컴퓨터구조론

조성진(Sung-Jin Cho)

1979년 2월 강원대학교 수학교육과 졸업 (이학사)
1981년 2월 고려대학교 대학원 수학과 졸업(이학석사)
1988년 2월 고려대학교 대학원 수학과 졸업(이학박사)
부경대학교 응용수학과 교수
※ 주관심분야 : 셀룰라 오토마타론, 정보보호, 부호이론, 컴퓨터 구조론

최언숙(Un-Sook Choi)

1992년 2월 성균관대학교 산업공학과 졸업 (공학사)
2000년 2월 부경대학교 대학원 응용수학과 졸업(이학석사)
2004년 2월 부경대학교 대학원 응용수학과 졸업(이학박사)
2009년 8월 부경대학교 대학원 정보보호학과 졸업 (공학박사)
동명대학교 미디어공학과 전임강사
※ 주관심분야 : 셀룰라 오토마타론, 정보보호, 암호이론