

컴퓨터 면역 시스템 개발을 위한 생체 면역 시스템 모델링

Biological Immune System Modeling for Development of Computer Immune System

선상준, 이채원, 심재윤, 심귀보

중앙대학교 전자전기공학부

Sang-Joon Sun, Chae-Won Lee, Jae-Yoon Shim, Kwee-Bo Sim,

School of Electrical and Electronic Engineering, Chung-Ang University

E-mail : kbsim@cau.ac.kr

요 약

생체 면역 시스템은 외부의 물질로부터 생명체를 지켜주는 방어 시스템으로 자기와 구별되는 외부 물질을 인식하여 생명체를 지켜준다. 본 논문에서는 이러한 생체 면역 시스템의 자기와 비자기를 구분하는 특징을 컴퓨터에 적용해서 자기(Self) 파일과 구분되는 비자기(Nonself) 파일을 인식 가능하도록 생체 면역 시스템의 T세포를 모델링하였다. 모델링한 T세포는 자기를 인식하는 MHC인식부와 비자기를 인식하는 항원인식부로 구성되어 있으며, 자기-비자기 검사를 요하는 파일과의 매칭을 통해 자기-비자기를 구분한다. MHC인식부와 항원인식부로 구성된 T세포의 자기-비자기 인식률이 항원인식부만을 사용한 자기-비자기 인식률에 비해 높아짐을 입증하였다.

1. 서 론

생체 면역 시스템은 외부의 물질로부터 생명체의 몸을 지키는 방어 시스템으로 외부의 물질을 인식하고 자기와 비자기를 구분하며 학습 및 기억 능력과 분산계로서 전체의 조화를 유지하는 능력 등을 가지고 있는 복잡하고 정교한 시스템이다[1, 2]. 그 중 자기-비자기(Self-NonSelf) 인식은 외부에서 침입한 존재나 내부에서 변형된 존재를 자기 외의 존재로 인식함으로써 생체 면역 시스템의 감시작용을 한다. 이러한 생체 면역 시스템의 자기-비자기 인식은 새로운 컴퓨터 바이러스 등에 능동적으로 대처할 수 없는 현재의 바이러스 검출 및 치료 방법에 응용이 가능해 많은 연구가 활발히 진행 중에 있다[3-7].

생체 면역 시스템을 자기와 비자기의 구분에 적용한 사례는 Stephanie Forrest와 Lawrence Allen의 항원인식에 의한 자기-비자기 구분 모델이 있다[3]. 본 저자는 위의 항원인식에 의한 자기-비자기 구분 모델에서 고려하지 않은 MHC인식부를 이용하여 생체 면역 시스템의 자기-비자기 구분 모델을 제안한다. 제안한 모델을 이용하여 시뮬레이션한 결과 항원인식부만을 사용한 경우에 비해 MHC인식부를 같이 사용한 경우의 자기-비자기 인식률이 높아짐을 입증하였다.

2. 자기-비자기 구분 모델

2.1 생체 면역 시스템

외부에서 침입한 존재를 감지하고 이를 제거하

는 생체 면역 시스템을 구성하는 기본 요소는 B세포(B-cell)와 T세포(T-cell)이다. B세포는 생체에 침입하거나 들어온 외부 물질인 항원에 대해 항체를 분비함으로써 항원을 제거하는 역할을 한다. T세포는 기능에 따라 몇 가지의 구분이 가능하다. B세포를 활성화시켜 항체의 분비를 촉진하는 역할의 보조 T세포(Helper T-cell)와 항원에 의해 감염된 자기세포를 식별하여 죽이는 역할을 하는 세포독성 T세포(Cytotoxic T-cell), 그리고 항체에 의한 면역 시스템이 활성화 된 이후 시스템을 억제하는 역할인 억제 T세포(Suppressor T-cell)로 구분된다. 특히 세포독성 T세포는 각각의 세포가 자기세포임을 알리는 MHC(Major Histocompatibility Complex) 단백질을 인식하는 부분과 외부 물질인 항원을 인식하는 항원인식부분으로 구성되어서 세포가 자기세포인지 아닌지, 그리고 항원에 의해 감염된 세포인지 아닌지를 구분할 수 있다.

2.2 생체 면역 시스템 모델링

본 논문에서는 자기와 비자기를 구분하여 인식하는 생체 면역 시스템의 세포독성 T세포의 특성을 컴퓨터의 면역 시스템에 적용 가능하게 모델링하였다. 컴퓨터에서 자기를 인식하고 있으면 외부의 침입으로 인한 자기의 변화를 감지할 수 있다.

모델링 된 T세포는 두 부분의 인식부로 형성된다. 자기를 구분하는 MHC인식부와 비자기를 구분하는 항원인식부로 구성된다. 이러한 T세포의 MHC인식부는 자기로 인식되는 파일의 일부를 이용해 구성하며 항원인식부는 랜덤스트링을 매칭해 구성한다[그림 1].

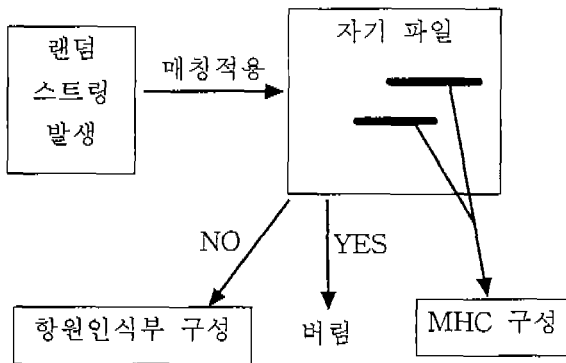


그림 1. T세포의 인식부 구성

이렇게 구성된 각각의 인식부를 이용해 자기-비자기 검색을 요하는 파일과의 매칭을 통해 자기-비자기를 결정하게 된다[그림 2].

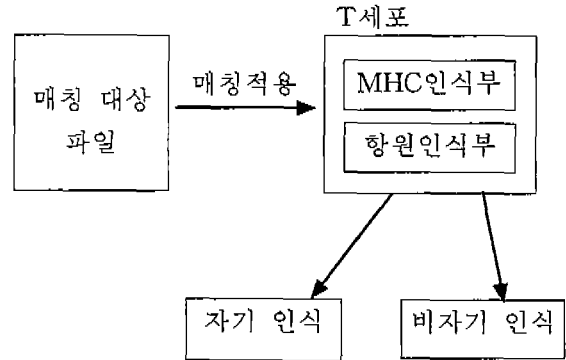


그림2. T세포를 이용한 자기-비자기 검색

2.2.1 MHC인식부

MHC는 생체 면역 시스템에서 세포들 자신이 자기임을 알리는 단백질로 이를 이용하여 T세포는 그 세포가 자기인지를 구분한다. 모델링 된 MHC인식부는 자기로 인식되는 파일의 일부를 스트링 단위로 M 개를 선택하여 구성한다. 구성된 MHC인식부로 자기-비자기 검색을 했을 때 자기로 인식할 확률 P_M 은

S : 자기 파일의 전체 스트링의 갯수

M : 구성된 MHC인식부의 갯수

S_p : 검색 파일 내의 자기 파일과 같은 스트링의 갯수(유사도)

P_M : MHC인식부로 매칭하였을 때 자기파일로 인식할 확률

$$P_M \approx \frac{s_p C_M}{s C_M} = \frac{S_p!(S-M)!}{S!(S_p-M)!} \quad (1)$$

이다.

표1. S , S_p , M 을 구성했을 때 MHC인식부가 자기로 인식할 확률 P_M

S	S_p	M	P_M
1600	1440	5	0.5901
1600	1440	10	0.3476
1600	1440	20	0.1200
1600	1520	10	0.5978
1600	1520	20	0.3562
1600	1520	30	0.2116

2.2.2 항원인식부

생체 면역 시스템에서 항원인식부는 MHC를 인식하지 못하게 구성되어 자기를 인식하지 못한다. 따라서 항원인식부는 비자기를 인식하기 위한 부분인 것이다. 모델링 된 항원인식부는 자기 파일과 매칭이 안 되는 A개의 스트링들로 구성되어 자기와의 매칭이 안되게 하여 비자기를 인식하게 한다.

자신에게 1:1로 완전히 매칭이 안 되는 항원인식부를 찾을 경우에는 자기 파일의 크기를 줄여서 설정하거나 항원인식부를 만드는 자기 파일이 작아야만 한다. 항원인식부를 구성할 때의 이러한 점을 보완하기 위해 매칭시 r 만큼의 연속적인 수가 매칭됨을 검사하여 항원인식부를 구성한다. 다음은 r개의 연속적인 심벌을 이용하여 매칭시 자기 파일로 설정이 가능한 스트링의 수를 계산하였다. 언제나 항원인식부를 찾을 수 있는 자기 파일 스트링의 최대값 S_R 은

A : 항원인식부의 갯수

m : 사용한 심벌의 갯수

r : 매칭시 검색요인인 연속적인 심벌 개수

S_R : r개의 연속적인 심벌을 이용한 매칭시

자기 파일 스트링의 수

$$\text{if } P(A)=1, \text{ then } S_R < m^r - A \quad (2)$$

이다.

표2. A, m, r을 결정했을 때 자기 파일의 사용가능한 최대값 S_R

m	r	A	S_R
256	1	20	236
256	2	20	65516
256	3	20	16777196

항원인식부는 랜덤하게 스트링을 하나 만들고 이를 자기 파일과 연속적인 수를 이용해 매칭한다. 매칭되는 스트링은 자기 파일을 구분할 수 없는 스트링이므로 버리고 또 다른 랜덤한 스트링을 만들어 다시 매칭한다. 이러한 단계를 거쳐 구성된 항원인식부는 자기 파일과는 매칭이 안된다. 이렇게 구성된 항원인식부로 검사를 원하는 파일을 검사하여 매칭이 안되면

자기로 인식하며 매칭이 되는 경우는 비자기로 인식해 자기와 비자기를 구분한다.

다음에서 r개의 연속적인 심벌을 이용한 매칭시 두 개의 랜덤 스트링이 매칭될 확률은

S_{ALL} : 구성 가능한 모든 스트링 집합

l : 스트링 길이

P_r : r개의 연속적인 심벌을 이용한 매칭시

두 개의 랜덤 스트링이 매칭될 확률

P_A : 구성된 항원인식부를 이용하여 매칭시

자기로 인식할 확률

$$P_r \approx \frac{(l-r)(m-1) + 1}{m^r} \quad (3)$$

이며 $m^r \ll 1$ 대만 근사화할 수 있다[3].

항원인식부를 스트링 1개로 구성했을 경우의 자기로 인식될 확률 $P_{A(1)}$ 는

$$P_{A(1)} = (1 - P_r)^{(S - S_b)} \quad (4)$$

이다.

항원인식부로 다수의 스트링을 사용할 경우에 자기로 인식될 확률 P_A 는

$$P_A \leq (1 - P_r)^{A(S - S_b)} \quad (5)$$

이다.

이때 등호는 항원인식부가 매칭할 수 있는 집합들이 독립일 때만 가능하다.

표3. m, r, l을 결정했을 때 두 개의 랜덤스트링이 매칭될 확률 P_r

m	l	r	P_r
256	32	2	0.000471
256	16	2	0.000228
256	8	2	0.000106
256	32	8	~0.0
256	16	8	~0.0

표4. P_r, A, S, S_b 를 결정했을 때 항원인식부가 자기로 인식할 확률 P_A 의 값

P_r	A	S	S_b	P_A
0.000471	10	1600	1440	0.47058
0.000471	20	1600	1440	0.22145
0.000471	20	1600	1520	0.47058
0.000471	30	1600	1520	0.32281
0.000228	10	1600	1440	0.69430
0.000228	20	1600	1440	0.48206

2.2.3 자기-비자기 구분

자기-비자기 구분은 T세포의 두 인식부로 각각 매칭 검사를 하여 결과가 자신인지 아닌지를 결정함으로써 이루어진다. 이 경우 어느 하나라도 자신이 아니라는 결과를 가지고 있으면 전체적인 결과는 자신이 아닌 것으로 인식하게 된다. 각각의 매칭결과를 검사하기 때문에 결국 P_M 과 P_A 는 독립적인 사건이 된다. P_S 는 MHC인식부와 항원인식부를 가지고 매칭 했을 경우에 자신이라고 인식될 확률로

$$P_S = P_M \cap P_A = P_M \times P_A \quad (7)$$

이다.

P_M 은 자기를 검색할 경우에 1이 되며 그 외의 경우에는 1보다 작게 된다. 따라서 MHC인식부와 항원인식부를 이용해 자기-비자기를 구분할 경우의 P_S 의 값은 항원인식부만을 가지고 자기-비자기를 구분하는 경우인 P_A 보다 작아지게 되어 자기-비자기를 구분하는 인식 확률이 높아진다.

3. 시뮬레이션 결과

컴퓨터에서 적용이 가능한 자기와 비자기를 구분할 수 있게 모델링 된 T세포를 이용하여 시뮬레이션에 적용하였다. 시뮬레이션에서 사용 가능한 심벌의 개수로는 이진 8bit인 256개로 결정하였다. 이는 컴퓨터의 기본 입출력 단위가 1byte단위인 Character변수이기 때문이다. 스트링의 길이는 32, 자기 파일의 스트링의 개수는 1600으로 구성하였다. 자기 파일과의 유사도는 90%로 설정하여 자기-비자기 구분의 가능성을 검사하였다. 또한 T세포의 MHC인식부와 항원인식부는 각각 20개, 10개, 5개로 구성하여 시뮬레이션을 실행하였다.

표5. 시뮬레이션 결과

M, A	P_M (이론)	P_M (결과)	P_A (이론)	P_A (결과)	P_S (이론)	P_S (결과)
5	0.5901	0.8969	0.68600	0.6799	0.4048	0.6105
10	0.3476	0.00	0.47058	0.4553	0.1636	0.00
20	0.1200	0.00	0.22145	0.2085	0.0266	0.00

시뮬레이션을 실행한 결과 MHC인식부의 크기에 따라 기여도가 달라지지만 일정 크기 이상의 MHC인식부가 구성되면 항원인식부만을 사용하여 자기-비자기를 인식하는 비율보다 높은 비율을 가진다.

4. 결 론

본 논문은 생체 면역 시스템의 자기와 비자기를 구분하여 인식하는 세포독성 T세포의 특징을 이용하여 컴퓨터에서 자기 파일과 비자기 파일을 구분할 수 있는 T세포를 모델링하였다. 모델링 된 T세포는 MHC인식부와 항원인식부로 구성되었으며 각각의 인식부를 이용해 비자기 파일에 적용시켜 자기-비자기 구분이 가능함을 입증하였다.

감사의 글

본 연구는 한국과학재단 산학협력연구과제 연구비(2000-30300-003-1)지원으로 수행되었으며 지원에 감사드립니다.

5. 참 고 문 헌

- [1] I. Roitt, J. Brostoff, D. Male, Immunology, 4th edition, Mosby, 1996
- [2] 이동욱, 실귀보, "인공 면역 시스템 기반 자율 분산로봇 시스템의 협조 전략과 군행동", 한국 퍼지 및 지능시스템학회논문지 Vol. 9, No. 6 p. 627-633, 1999.
- [3] Stephanie Forrest, Lawrence Allen, Alan S. Perelson, Rajesh Cherukuri "Self-Nonself Discrimination in a Computer" IEEE Symposium on Research in Security and Privacy, 1994.
- [4] Ja Beom Gu, Dong Wook Lee, Kwee-Bo Sim, Se Hyun Park, "An Immunity-based Security Layer Model", Genetic and Evolutionary Computation Conference Workshop Program p.47 - 48, 2000.
- [5] Dipankar Dasgupta, "An Immune Agent Architecture for Intrusion Detection", Genetic and Evolutionary Computation Conference Workshop Program p.42 - 44, 2000.
- [6] Paul K. Harmer, Gary B. Lamont, "An Agent Based Architecture for a Computer Virus Immune System", Genetic and Evolutionary Computation Conference Workshop Program p.45 - 46, 2000.
- [7] 구자범, 이동욱, 박세현, 실귀보, "생체 면역 시스템 기반의 새로운 보안 항체 계층 모델", 한국 퍼지 및 지능시스템학회 논문지 Vol.10, No.2, p.122-128, 2000.