

분자 면역 컴퓨팅에 의한 숫자 인식

김수동² 신기루¹ 장병탁^{1, 2}

¹서울대학교 컴퓨터공학부, ²서울대학교 대학원 인지과학 협동과정
{sdkim, krshin, btzhang}@bi.snu.ac.kr

Digit Recognition by Molecular Immunocomputing

Su Dong Kim², Ki-Roo Shin¹, Byoung-Tak Zhang^{1, 2}

¹School of Computer Science, Seoul National University

²Interdisciplinary Program in Cognitive Science, Seoul National University

요약

신경과학이 발전함에 따라 기억현상의 분자생물학적, 세포생물학적 메커니즘이 밝혀지고 있으며, 신경계의 특징을 계산학적 측면으로 응용한 신경망 분야는 상당한 연구성과가 축적되어 하나의 학문 분야로까지 자리매김하고 있다. 면역학이 발전함에 따라 면역현상의 분자생물학적, 세포생물학적 메커니즘이 밝혀지고 있으며, 특히 면역기억현상의 산술인 항체를 이용한 여러 가지 생물학적 실험 방법이 고안되어 사용되고 있는 한편, 면역 현상의 특징을 계산학적 측면으로 응용하려는 다양한 시도가 근래 이루어지고 있다. 본고에서는 항원과 항체 분자를 이용한 면역생물학적 실험 방법을 적용하여 문자 인식 문제를 해결하는 분자 면역 컴퓨팅의 개념을 도입하고, 이 개념을 도입하여 숫자를 인식하는 문제에 적용하는 사례를 제시하였다.

1. 서론

일반적으로, 기억(memory)이라는 용어는 정보를 기호화하여 저장하고, 인출하는 일련의 과정을 뜻한다. 일반적인 의미로 통용되는 기억에 대해서는 인지과학, 신경심리학, 신경생물학, 컴퓨터 과학 등 다양한 분야에서 많은 연구가 이루어져 왔다 [1].

한편, 생체방어기전인 면역현상을 연구하는 면역생물학적 연구가 진척됨에 따라, 클론선택에 기초한 적응면역의 발달에서 가장 중요한 생물학적 현상으로 면역기억(immunological memory)이 존재함이 밝혀졌으며, 면역기억현상의 세포생물학적, 분자생물학적 기초가 확립되었다 [2]. 이러한 성과를 낳기까지 다양한 실험방법이 고안되었으며, 면역현상을 이용한 실험방법이 다른 목적의 생물학적 연구에 폭넓게 응용되기도 하였다. 본고에서는 면역기억현상을 이용한 실험방법을 토대로 컴퓨터 과학적 측면에 대한 효과적인 응용가능성을 모색하고자 하며, 다음과 같은 차례에 따라 각 내용을 언급할 것이다.

첫째, 면역기억현상의 세포 및 분자 수준에서의 작용기전에 대해 면역생물학의 연구성과에 입각하여 약술하고, 둘째, 면역현상을 이용한 실험방법 중 항체를 이용하는 실험방법을 소개하고, 셋째, 이러한 실험방법의 컴퓨터 과학적 측면에 대한 응용가능성을 살펴보기로 한다.

2. 면역 컴퓨팅

2.1 컴퓨터 과학적 측면에서 바라본 면역계의 특징

면역계는 강력한 정보처리능력을 지니고 있기 때문에 컴퓨터 과학적 측면으로의 응용가능성에 대한 관심이 높아지고 있다. 특히 컴퓨터 과학적 측면에서 관심을 끄는 면역계의 특징을 들자면, 우선, 면역계는 다수의 복잡한 연산을 분산시켜 병렬적으로 수행한다. 또한, 면역계 곳곳에서 일어나는 지엽적 반응의 총합으로 신체를 감염으로부터 보호하는 면역계의 전체적 행동

양상이 창출(emergent)된다.

면역계를 구성하는 세포들은 과거에 경험한 항원에 대한 기억을 지닐 뿐더러, 처음 대하는 새로운 패턴의 항원에 대해서도 대항할 수 있으므로, “제 2의 뇌(second brain)”로서 기능한다고까지 말하는 학자도 있다. 더구나, 시간의 경과에 따른 면역반응의 발달 양상과 진화 과정은 그 자체가 하나의 동역학계로서, 주목할만한 연구주제이다. 컴퓨터 과학적 측면에서 면역계의 정보처리과정은 다음과 같은 중요한 특징을 지닌다 [3, 4].

- ① 패턴 인식(pattern recognition)
- ② 특징 추출(feature extraction)
- ③ 다양성(diversity)
- ④ 학습(learning)
- ⑤ 기억(memory)
- ⑥ 분산 탐지(distributed detection)
- ⑦ 자체 조절(self-regulation)
- ⑧ 역치 메커니즘(threshold mechanism)
- ⑨ 공동 자극(co-stimulation)
- ⑩ 역동적 방어기전(dynamic protection)
- ⑪ 확률적 탐지(probabilistic detection)

2.2 면역계의 컴퓨터 과학적 응용

면역 현상을 설명하고자 몇몇 컴퓨터 과학적 이론과 수학적 모델이 제시된 바 있으며, 생물학의 관점에서 면역계의 여러 구성 요소들을 컴퓨터로 흉내내려는 시도가 늘어나고 있다. 미분방정식 모델, 확률론적 모델, 세포자동자(celluar-automata) 모델, 모양-공간(shape-space) 모델 등이 그 예이다. 면역계는 컴퓨터 과학적 문제 해결을 위한 훌륭한 방법론의 원천이 될 잠재력이 풍부하지만, 이에 대한 연구는 미진한 편이다. 즉각 활용 가능한 정보처리 메커니즘을 면역계로부터 찾아내고자 시도된 연구로서는, 다음과 같은 사례를 들 수 있다 [4].

- ① 면역 네트워크 모델(immune network model)
- ② 음성선택 알고리즘(negative selection algorithm)
- ③ 컴퓨터 보안(computer security)
- ④ 시계열 데이터의 이상 탐지(anomaly detection in time series data)
- ⑤ 오류 진단(fault diagnosis)
- ⑥ 패턴 인식(pattern recognition)

2.3 분자 면역 컴퓨팅

생체 분자를 이용하여 정보처리를 하려는 분자정보처리기술 중에서 현재 가장 큰 실용화 가능성을 지니고 있는 분자 컴퓨팅 기술을 DNA 분자를 이용하는 DNA 컴퓨팅 기술이다 [5]. DNA 기반 분자 컴퓨팅은 기본적으로 DNA 염기서열에 정보를 저장하고, DNA 분자가 가진 화학적 특성을 이용하여 정보를 처리하는 연산방식이다 [6]. DNA 컴퓨팅과 비슷한 원리로, 단백질의 조각인 펩티드(peptide)와 항체(antibody)의 반응을 이용한 펩티드 컴퓨팅(peptide computing)도 가능성이 제시되었으며 [7], 이를 확장하여 적용하는 문제도 고려된 바 있다 [8].

최근 펩티드 컴퓨팅의 개념을 추상화하여 면역 컴퓨팅(immunocomputing)이라 명명하고 컴퓨터에 의한 패턴인식(pattern recognition) 문제에 적용하려는 시도에 이어 [9], 이 기법을 실제 바이오칩(biochip)화하는 데까지 성공한 사례가 보고되었다 [10].

지금까지 제시된 펩티드 컴퓨팅의 단점은, 실제 생체 분자를 이용하여 구현하기에 어려운 측면이 있으므로 개념 제시에 그치고 있는 데 있다. 면역 컴퓨팅의 개념은 비록 분자 컴퓨팅 기술로써 구현하는 실험에 성공한 사례가 보고되고 있지만, 펩티드 컴퓨팅이 지니고 있는 장점인 정보처리개념이 과도하게 추상화되어 있으므로 컴퓨터 과학의 문제를 면역 컴퓨팅으로 직접 풀어내는 데 한계를 지닌다.

본고에서는 DNA 컴퓨팅 및 펩티드 컴퓨팅이 지니는 정보저장방식의 장점과 면역학적 실험기법을 이용하여 실현 가능성에서 앞서 있는 면역 컴퓨팅의 장점을 모두 취한 분자 면역 컴퓨팅(molecular immunocomputing)의 개념을 제시하고자 한다.

우선, 항원으로 이용될 펩티드 서열을 디자인한다. 이 펩티드는 특정 위치에 보통 네 가지의 모든 가능한 뉴클레오티드(nucleotide)가 존재하도록 화학적으로 합성된 DNA 서열에 대해 상보적인 RNA 서열로부터 번역되는 짧은 펩티드이다. 따라서 생체에서 생성되는 20 가지 아미노산들이 모두 포함되어 있다. 다양한 위치의 숫자에 따라 가능한 펩티드의 종류도 급격히 증가하는데, 예컨대, 여덟 개의 아미노산 서열을 갖는 펩티드의 경우, 2×10^{10} 이상의 다양한 종류가 만들어질 수 있다 [11].

여기서, 풀고자 하는 문제에 대해 가능한 질의는 짧은 펩티드 항원서열의 정보를 담고 있는 DNA 염기서열로 표현 가능하며, 단순히 아미노산 서열로 표현할 수도 있다.

특정 질의에 대한 정보를 항원의 발현 형태로 담고 있는 펩티드를 KLH(keyhole limpet hemacyanin) 또는 BSA(bovine serum albumin)와 같이 항원성이 적다고 알려진 운반체 단백질과 결합(conjugation)시켜 면역능을 지닌 토끼, 또는 생쥐 등의 실험동물에 주입하면, 결합 단백질(conjugated protein) 항원이 면역 반응을 유도하게 된다 [12]. 특정 항원의 발현 형태로 질의를 받은 실험동물은 이에 대한 항체를 생성하게 되는데, 이때 생성되는 특정 항원에 대한 특정 항체 단백질이 질의에 대한 해가 된다. 실험동물의 종에 따라 확립된 체내 적응 기간을 두고 3차에 이르는 면역반응을 통해 생성된 항체 해는 혈청을 얻은 뒤, 친화성 크로마토그래피법을 통해 분리, 정제할 수 있다.

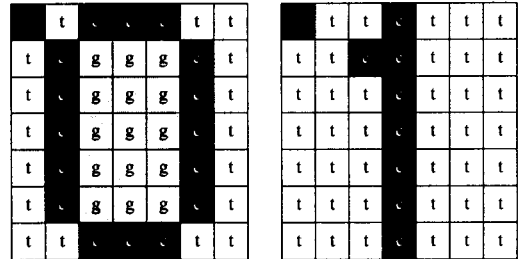
이 단계에서 얻은 해가 특정 질의 항원에 대한 특정 항체 해인지 여부는 대조 질의 항원에 대한 항원-항체 반응을 비교하여 확인하거나, ELISA(enzyme-linked immunosorbent assay)

등의 방법을 통해 정량화 할 수 있다 [11].

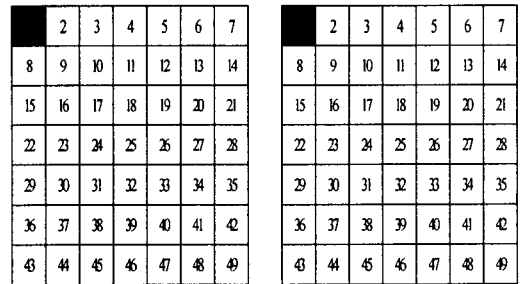
이러한 분자 면역 컴퓨팅 기법은 패턴 인식 문제에 적용될 수 있다.

3. 분자 면역 컴퓨팅 기법에 의한 숫자 인식

분자 면역 컴퓨팅 기법을 이용하여 숫자 "0" 과 "1" 을 구분하는 간단한 문자 인식문제를 해결하는 과정은 다음과 같다. 숫자 "0" 과 "1" 을 질의로 삼는 경우, 숫자 "0" 과 "1" 에 대한 정보를 각기 다른 항원성을 지니는 항원발현 형태의 펩티드 정보로 담고 있는 DNA 서열을 각각 디자인한다. 여기에서는 2차원 7×7 격자에 나타나는 문자를 인식하는 예를 들어 설명하고자 한다.



[그림 1.] 숫자 "0" 과 "1" 의 2차원 격자 DNA 인코딩



"0"-DNA template sequence:
 5'- a tc ctt ttc tgc ggc ttc ggg ctt cgg gct tgc ggc ttc ggg ctt tcc ctt - 3'
 "0"-RNA transcript sequence:
 3' - ug aag gaa agc agc ccc aag ccc gaa gcc cga agc ccc aag ccc gaa agg gaa - 5'
 "1"-DNA template sequence:
 5'- a ttc ttt ttc ctt ttt tct ttt ttc ttt ttt tct ttt ttc ttt - 3'
 "1"-RNA transcript sequence:
 3' - ug aag aaa aag gaa aaa aga aaa aag aaa aaa gaa aaa aga aaa aag aaa - 5'
 [그림 2.] DNA 에서 RNA 로의 전사(transcription)

"0"-RNA transcript sequence:
 5' - aag gga aag ccc gaa gcc cga agc ccg aag ccc gaa gcc cga aag gga ug - 3'
 "0"-peptide sequence:
 NH₂ - K G K P E A R S P K P K P E A R K G C - COOH
 "1"-RNA transcript sequence:
 5' - aaa gaa aaa aga aaa aag aaa aaa gaa aaa aga aaa aag gaa aaa gaa ug - 3'
 "1"-peptide sequence:
 NH₂ - K E K R K K K K E K R K K E K E C - COOH
 [그림 3.] RNA 에서 peptide 로의 번역(translation)

두 가지 서로 다른 문자 "0" 과 "1" 이 2차원 7×7 격자의 음영정보로 입력되면, 어두운 격자는 시토신(cytosine), 밝은 격자는 티민(thymine), 어두운 격자로 둘러싸인 격자는 구아닌(guanine)으로 변환시킨다 [그림 1]. 정보 변환시 아데닌(adenine) 사용을 기피하는 이유는 DNA 서열 정보가 펩티드 서

열 정보로 변환되는 도중에 정지 코돈(STOP codon)이 나타나 것을 꺼리기 때문이다. 7×7 격자 각각에 고유번호를 매겨 격자상의 2차원 DNA 배열을 1차원 DNA 서열정보로 변환한다. 그러면, 1차원으로 변환된 DNA 서열정보는 곧바로 그에 상응하는 1차원 RNA 서열정보로 옮길 수 있다 [그림 2]. 전사된 1차원 RNA 서열정보를 유전암호표(codon table)의 규칙(rule)에 따라 1차원 펩티드 서열정보로 변환한다 [그림 3]. 이때, 각 코돈(codon)의 세번째 염기위치, 즉 1차원 RNA 서열정보의 각 3의 배수째에 해당하는 DNA 서열정보는 noise에 tolerant한 격자 위치가 된다. 2차원 격자배열정보를 1차원화할 때 발생하는 noise-tolerant한 격자가 특정부분에 편재되지 않도록, 2차원 격자배열정보를 두 차례에 걸쳐 1 뉴클레오티드(nucleotide)씩 프레임쉬프트(frameshift)한 정보배열도 갖추어 둔다. 1차원 펩티드 서열정보로 전환할 때 C-말단부위에 나타나는 불완전한 코돈은 가급적 시스테인(cysteine)에 대한 코드(code)로 지정하여 마무리한다. 이는 펩티드 서열로 번역한 항원을 운반체 단백질과 합성할 때 이황화 결합(disulfide bond)을 통해 합성을 용이하게 하고자 함이다. 따라서 7×7 격자에서 첫 번째 격자의 정보만은 특별히 아데닌(adenine)으로 고정한다.

“1” 항원에 의해 생성된 항체는 항원과 항원-항체 결합에 의한 응집반응을 일으키는 반면, “0” 항원과는 응집반응을 일으키지 않을 것이다. 또한, 생성된 해는 주어진 질의와 유사한, 약간 변형된 형태의 질의에 대해서도 응집반응을 일으킬 것이다. 보다 높은 강도의 응집도 측정법으로서 ELISA의 활용도 고려할 수 있다.

4. 결론 및 고찰

지금까지 기술한 분자 면역컴퓨팅 기법에 의한 문자 인식 원리는 단순히 숫자 “0” 과 “1”의 구분에 그치지 않고, 7×7 격자에서 구현되는 다양한 문자와 기호의 인식에도 곧바로 적용된다. 하나의 항원으로 주어진 질의에 대해 생성된 항체 해는 noise-tolerant한 특성으로 인해 유사한 질의에도 답할 수 있으므로 다양한 필기체 인식 문제에 적용될 수 있는 잠재성을 지니고 있다.

앞의 예에서 7×7 격자를 정보입력의 틀로서 채택한 까닭은, 이 틀에서부터 생성되는 펩티드의 길이가 마지막 시스테인(cysteine) 잔기를 포함하여 17 개의 아미노산으로 구성되므로, 합텐(hapten)으로서 항원성을 가장 잘 유도하는 것으로 알려진 길이인 15 ~ 20 개의 아미노산으로 구성된 펩티드의 범주에 맞추기 위함이다 [12]. 따라서, 정보처리용량의 확장을 고려할 때에는 2차원 7×7 격자단위를 정보처리의 기본 단위로 삼는 것이 바람직하다. 안정적으로 정보처리용량의 확장이 가능할 경우, 분자 면역 컴퓨팅은 이론적으로 문자열 처리도 가능할 것으로 보이며, 이는 암호학 분야에 응용될 수 있다.

지금까지 제안된 분자 면역 컴퓨팅 기법의 가장 큰 약점은, 여러 차례의 면역 반응을 거쳐 문제에 대한 해를 얻기까지 약 10 주, 또는 그 이상의 오랜 시간이 걸린다는 점이다. 이러한 약점은 파지 발현 라이브러리(phage display libraries) 기법을 이용하면 2 주 남짓이라는 시간까지 단축시킴으로써 어느 정도까지는 보완할 수 있다 [13]. 하지만 이렇게 항체를 얻는 데 시간이 오래 걸리는 약점에도 불구하고 생체가 복잡한 연산을 수행하는 장점을 취하기 위해서는, ABO-방식의 혈액검사의 경우처럼, 표준화된 항원-항체 반응을 확립하고 이를 시행하기 위한 시약 등을 사전에 키트(kit)화 또는 칩(chip)화 하여, 문제가 주어졌을 때 즉각 응집반응 발생 여부를 확인하여 해답을 얻을 수 있는 방법을 강구하는 것이다 [10].

면역 컴퓨팅 분야에서 질병 진단 분야에 적용할 수 있는 면역 칩(immuno chip)을 변환하여 바이오컴퓨터(biocomputer)를 만들어내는 개념이 논의되고 있는 시점에서 [10], 분자 면역 컴퓨팅 기술은 면역 컴퓨팅 기법에 펩티드 컴퓨팅 개념을 효율적으로 도입하는 데 기여할 수 있을 것이다.

5. 참고 문헌

- [1] D. L. Schacter and E. Tulving (ed.), *Memory systems 1994*. Cambridge, The MIT Press, 1994.
- [2] R. A. Goldsby, et al., *Kuby immunology*, New York, W. H. Freeman and Company, 2000.
- [3] S. Forrest and S. A. Hofmeyr, Immunology as Information Processing. In L.A. Segel and I. R. Cohen (ed.), *Design Principles for Immune System & Other Distributed Autonomous Systems*, pp. 361-387, Oxford, Oxford Univ. Press, 2000.
- [4] D. Gupta (ed.), *Artificial Immune Systems and Their Application*, Berlin, Springer-Verlag., 1998.
- [5] L. M. Adleman, Molecular computation of solutions to combinatorial problems, *Science*, 266, pp. 1021-1024, 1994.
- [6] 장병택, 분자정보처리기술, *전자공학회지*, 29 (3), pp. 270 - 277, 2002.
- [7] H. Hug and R. Schuler, Strategies for the development of a peptide computer, *Bioinformatics* 17 (4), pp. 364-368, 2001.
- [8] M. S. Balan et al., Peptide computing - *Universality and Complexity*, 7th International Workshop on DNA-Based Computers, DNA7, pp. 290-299, 2002.
- [9] A. O. Tarakanov and V. A. Skormin, Pattern Recognition by Immunocomputing, *Proceedings of the 2002 Congress on Evolutionary Computation*, pp. 938-943, 2002.
- [10] A. O. Tarakanov et al., Immunocomputing for Bioarrays, *Proceedings of the 1st International Conferences on Artificial Immune Systems 2002*, pp. 32-40, 2002.
- [11] C. Janeway et al., *Immunology: the immune system in health and disease*, 5th ed., 2001.
- [12] E. Harlow and D. Lane, *Antibodies: A Laboratory Manual*, New York, Cold Spring Harbor Laboratory, 1988.
- [13] R. de Bruin et al., Selection of high-affinity phage antibodies from phage display libraries, *Nature Biotechnology* 17, pp. 397-399, 1999.