

바이오지능(Biointelligence) = 생물정보학 + 인공지능

서울대학교 장병탁*

1. 서론

Bioinformatics는 생물을 의미하는 bio와 정보학 즉 컴퓨터과학을 뜻하는 informatics¹⁾의 합성어로서 '생물정보학'을 뜻한다. Bioinformatics는 한편으로는 "생물학을 위한 정보학" (informatics for biology)의 의미로 이해할 수 있으며, 다른 한편으로는 "생물학에 기반한 정보학" (biology-based informatics)로 해석할 수 있다. 전자의 해석이 컴퓨터를 도구로 하는 생물학의 연구 분야를 의미하는 반면, 후자는 생물학에 기반한 컴퓨터를 연구하는 전산학의 한 분야를 의미한다. 현재 주로 통용되는 해석은 전자이며 실제로 본 특집호를 포함한 많은 논문들이 생물학 특히 분자생물학 분야에서 발생하는 대규모의 분자서열 데이터를 분석하기 위한 알고리즘이나 데이터베이스 기법 및 인공지능 기술에 관하여 다루고 있다. 이에 반해서, 여기서는 후자의 측면 즉 생물학 시스템의 정보처리 원리를 이해하여 이에 바탕을 둔 새로운 컴퓨터 기술을 개발하는 연구 방향에 관해 초점을 맞추고자 한다.

본 고에서는 특히 인공지능 관점에서, Bioinformatics에 대응되는 개념으로 생물학과 인공지능이 결합된 학제적 연구 분야로서 Biointelligence를 제안하며 앞으로 이를 '바이오지능'이라 부르기로 한다(그림 1 참조). 즉 Bioin-

formatics가 "생물학을 위한 정보학"과 "생물학 기반의 정보학"의 양면성을 지니는 것처럼 Biointelligence도 "생물학을 위한 인공지능"과 "생물학에 기반한 인공지능"을 연구하는 전산학의 한 연구 분야로 정의할 수 있다. 바이오지능에 대한 보다 구체적인 개념 정립과 그 세부 연구 대상은 이 원고의 후반부에서 다루기로 한다. 이 특집호의 논문[25]에서 "생물학을 위한 인공지능 기술"에 관한 예들을 살펴보았다. 여기서는 주로 "생물학에 기반한 인공지능 기술"에 관해 주로 다루고자 한다.

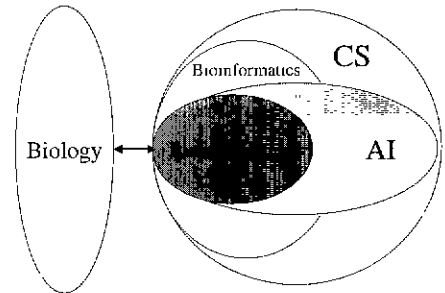


그림 1 생물학, 컴퓨터과학(CS), 인공지능(AI), 생물정보학, 바이오지능의 관계

본 논문은 다음과 같이 구성되어 있다. 제 2절에서는 생물학과 인공지능의 관계에 대해서 살펴본다. 특히 기초기반의 인공지능 연구와 생물학 기반의 인공지능 연구를 역사적으로 살펴본다. 제 3절에서는 "생물학에 기반한 인공지능"에 대한 구체적인 예로서 신경망, 진화알고리즘, 분자 컴퓨팅 방식을 기술한다. 제 4절에서는 현재 진행되고 있는 인공지능의 패러다임 변화를 살펴보고, 이로부터 "적용, 학습, 추론을 위한 생물 정

*종신회원

1) 영어로 computer science에 해당하는 독일어, 불어, 이태리어는 각각 Informatik, informatique, informatica이며, Informatik은 정보를 의미하는 Information과 수학이라는 의미의 Mathematik의 합성어이다

보처리 기반의 지능 기술”(Bio-inspired intelligence technologies for adaptation, learning, and inference), 즉 바이오지능(Biointelligence)의 개념을 도입한다. 바이오지능 기술을 다시 분자수준, 세포수준, 개체수준, 개체군수준에서의 지능 기술로 세분하고 각각의 특성 및 연구 주제에 대하여 서술한다. 마지막으로 제 5절에서는 현재 또는 단기적으로 바이오지능이 추구해야 할 발전 방향과, 나아가 인간 수준의 또는 인간을 능가하는 슈퍼지능(superintelligence)의 실현이라는 궁극적인 인공지능의 목표 달성을 위해서 장기적으로 추구해 나가야 할 연구 방향에 대해서 토론한다.

2. 인공지능과 생물학

인공지능(Artificial Intelligence, AI)은 일반적으로 “컴퓨터를 보다 지능적으로 만들기 위한 컴퓨터과학의 한 연구 분야” 또는 “현재로서 컴퓨터보다 사람이 하면 더 잘 할 수 있는 일들을 컴퓨터로 하여금 수행할 수 있도록 하는 연구 분야” 등으로 이해된다. 인공지능이란 용어는 1956년에 처음 만들어졌으나 사실은 이미 1940년대에 컴퓨터의 발명과 더불어 계산과 지능의 관계에 대한 많은 논의가 있어 왔다. 컴퓨터의 아버지로 일컬어지는 튜링은 1950년에 “Computing Machinery and Intelligence”라는 논문에서 이미 컴퓨터와 지능의 관계를 언급하며 기계의 지능을 테스트하기 위한 방법으로 소위 튜링 테스트를 제안하였다[21]. 튜링 테스트는 장막에 가려진 두 개의 방에 한 쪽에는 컴퓨터를, 다른 한 쪽에는 사람을 넣어 두고 밖에서 질문자가 자판을 통해 양쪽에 질문을 하여 어느쪽이 사람이고 어느쪽이 기계인지를 구별할 수 없으면 그 기계는 지능적이라고 판단하는 방법이다. 아직까지 튜링 테스트를 통과한 기계는 없다.

초기의 인공지능 연구는 다양한 지능형 행동을 나타낼 수 있는 몇 가지 단순한 원리를 찾아내고자 하였다. 합목적적인 행동을 하기 위해서 현재 상태와 목표 상태 사이의 차이점을 찾아내고 이 차이를 줄여주기 위한 수단을 찾음으로써 문제를 점차적으로 해결해나가는 Means-Ends Analysis는 이러한 일반적인 원리 중의 하나이다. 그러나 연구가 진행되고 이를 실세계 문제에 적용하

는 과정에서 이러한 일반적인 지능(general intelligence)의 실현은 애초에 생각했던 것보다 훨씬 어렵다는 것을 깨닫게 되었다. 따라서 70년대에 들어서는 목표를 줄이고 문제의 범위를 축소하는 대신 제한된 분야에서의 전문 지식을 시스템에 대량으로 주입함으로써 지능적인 행동을 보이도록 하려는 지식기반 시스템(knowledge-based system) 또는 전문가 시스템(expert systems)이 성행하게 되었다[28].

이러한 지식기반의 인공지능 기술은 문제 해결에 필요한 지식을 주로 기호논리 기반의 표현 방법을 사용하여 컴퓨터 상에 주입하고 그로부터 연역적인 추론에 의해 문제에 대한 답을 찾는데 관심을 집중하여 왔다[12]. 이 휴리스틱 기호 프로그래밍 방식은 제한된 분야에 대한 실용적인 시스템을 쉽게 구축할 수 있는 장점이 있다. 그러나 기호기반의 접근 방법은 문제 해결에 많은 지식을 필요로 하며 문제의 범위가 주어진 지식의 범위를 조금만 벗어나면 성능이 급속히 저하된다는 단점을 지닌다.

초기 인공지능 연구를 통해서 얻은 교훈 중의 하나는 컴퓨터로 하여금 어른 수준의 전문적인 지식을 요하는 문제를 해결하도록 하는 것은 비교적 쉬우나, 오히려 유아 수준의 지능을 흉내내는 것은 매우 어렵다는 것이다. 그 이유 중의 하나는 아마도 인간의 뇌가 오랜 동안 진화를 통해서 찾아낸 정보 처리 구조와 어린 아이가 성장하면서 경험으로부터 지식을 습득해온 학습 과정을 고려하지 않고 단지 그 결과, 즉 지식 그 자체만을 컴퓨터 프로그램에 주입하려고만 하였기 때문 일 것이다[23].

자연지능(natural intelligence)에 대해서 생각해 보면 인간은 부모로부터 유전 형질을 물려 받아, 세포의 성장과정을 통하여 뇌의 구조가 형성되며, 일생을 통해서 많은 경험과 교육을 통하여 지식과 상식을 쌓아간다. 그러나 이런 지식을 기계에 단순히 주입만 한다고만 할 때 근본적인 한계에 부딪힐 수밖에 없다. 더 나아가 인간의 뇌는 오랜 세월에 걸친 진화의 산물이다. 최근에 와서는 이러한 유기체의 발생 및 진화 과정[11]이나 뇌의 정보 처리 원리를 모사함으로써 지능형 시스템을 구축하려는 시도가 일어나고 있는 것도 이러한 이유에서이다[18].

Symbolic AI	Biological AI
<ul style="list-style-type: none"> • 1943 Production rules • 1956 "Artificial Intelligence" • 1958 LISP AI language • 1965 Resolution theorem proving • 1970 PROLOG language • 1971 STRIPS planner • 1973 MYCIN expert system • 1982-92. Fifth generation computer systems project • 1986: Society of mind • 1994 Intelligent agents 	<ul style="list-style-type: none"> • 1943 McCulloch-Pitts's neurons • 1959 Perceptron • 1965 Cybernetics • 1966 Simulated evolution • 1966 Self-reproducing automata • 1975 Genetic algorithm • 1982 Neural networks • 1986 Connectionism • 1987 Artificial life • 1992 Genetic programming • 1994 DNA computing

그림 2 고전적 인공지능과 생물학적 인공지능

인공지능의 역사를 돌아보면 사실 생물체를 모방함으로써 지능형 시스템을 구축하려는 시도는 예전부터 있어왔다. 그림 2의 왼쪽과 오른쪽 표에 기호기반의 인공지능(Symbolic AI)과 생물학 기반의 인공지능(Biological AI)의 역사를 대조적으로 나열하였다. 제시된 바와 같이 이미 인공지능 역사의 초기부터 생물체나 자연계의 정보 처리 기작에서 동기를 부여받은 지능형 정보 처리 모델이 꾸준히 연구되어 왔다. 기호적 인공지능 시스템의 이론적인 바탕이 된 Post 규칙 시스템이 발표된 1943년 같은 해에 신경세포가 논리 연산을 수행할 수 있다는 신경망(Neural Networks)의 이론적 기반이 수립되었다. 그 후 기호 기반의 시스템이 주류를 형성하던 1960-70년대에도 일부에서는 신경망과 Cybernetics에 대한 연구가 진행되고 있었으며 진화 연산(Evolutionary Computation) 방식[7]에 의한 인공지능 모델들도 이 시기에 제안된 바 있다. 1980년대 중반부터는 신경망 기반의 인공지능과 인지과학적 연구가 활발해졌다. 신경망은 특히 학습 특성이 뛰어나 기호기반 인공 지능이 어려움을 겪던 패턴인식 등과 같은 문제들은 학습을 통해 잘 할 수 있음을 보여줌으로써 연결론적 인공지능(Connectionist AI)이 점차 고전적 인공지능(Classical AI)에 뿌리를 내리기 시작하였다 [14, 15, 22].

아래의 3절에서는 생물학기반의 인공지능 기술의 대표적인 예로서 신경망, 진화연산, 분자컴퓨팅에 대하여 살펴본다. 이어서 4절에서는 이러한 기존의 연구에서 더 나아가 분자 생물학과 생물 정보학의 연구 결과에 기반한 바이오지능 기술

(Biointelligence Technologies) 즉 분자 수준, 세포 수준, 유기체 수준, 개체군 수준에서의 지능 기술에 대해서 살펴본다.

3. 생물학 기반 인공지능 기술의 예

3.1 신경망: 두뇌 구조를 모방한 계산 모델

신경연산(neural computation) 또는 신경망(neural networks)은 인간이나 동물의 뇌의 구조를 모방한 계산 모델이다. 이는 뇌라는 컴퓨터가 각각의 프로세서(뉴런, 신경세포)는 아주 단순한 일을 하지만 약 10^{11} 개의 아주 많은 수가 약 10^{14} 개의 연결선으로 상호 연결되어 병렬분산 방식으로 정보 처리를 한다는 점에 착안한 것이다. 신경망은 기존의 폰 노이만 방식 컴퓨터가 막대한 계산 속도와 메모리 기능을 가지는 소수의 프로세서를 사용하여 순차적인 계산을 하는 방식과 대조적이다.

신경세포가 계산을 수행할 수 있다는 것은 이미 1943년에 McCulloch와 Pitts에 의해 밝혀졌으며, 그 후 1950-60년대를 통해 여러 가지 신경망 학습 방식들이 연구된 바 있다. 그러나 그 당시로서는 대규모 실험에 필요한 컴퓨터 하드웨어 기술이 발달하지 못한데다 상대적으로 기호

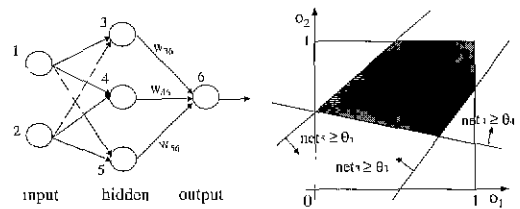


그림 3 신경망의 구조

기반의 인공지능 시스템들이 우세하게 됨에 따라 1970년대에는 연구가 부진하였다. 그러나 80년대 중반에 와서 VLSI 반도체를 비롯한 하드웨어 기술이 급진전하고 보다 막강한 능력을 지닌 다층 구조 신경망에 대한 학습 알고리즘이 개발되어 그 동안 지적되어온 기호 기반의 인공지능 시스템들의 여러 가지 단점을 신경망이 잘 보완해 줄 수 있음이 검증됨에 따라 다시 연구가 활성화되었다[19].

신경망은 여러 사례들로부터 입력력간에 존재

하는 일반적인 사상 관계를 잘 학습하는 능력을 지닌다. 이 과정은 보통 주어진 신경망 구조상에서 신경유닛들간의 연결강도를 재조정함으로써 수행되는데 이에 사용되는 학습 알고리즘으로는 헤비안 학습 규칙, 오류 역전파(error backpropagation) 방법, 볼츠만 학습법 등이 있다[14]. 기본적으로 이런 알고리즘들은 입력 패턴이 주어지면, 뉴런들이 활성화되어 출력값을 계산하고 원하는 출력값과 비교를 통해 에러를 계산한 후 이 에러를 최소화하는 방향으로 뉴런들의 연결가중치를 변경함으로써 학습이 이루어진다.

신경망은 디지털 방식이라기 보다는 원리적으로 아날로그 방식으로 연산을 수행하고, 순차 처리 보다는 병렬 처리에 가까우며, 지식이 어떤 특정 주소에 기억되어 있다기 보다는 여러 장소에 분산 저장되어 있다. 이러한 특성으로 인해 많은 데이터로부터 일반적인 특성을 추출해 내는 학습 능력이 강하고, 또한 데이터에 잡음이 있거나 몇 개의 뉴런유닛이 고장이 나도 원래의 기능을 수행하는 능력이 있으며, 일단 학습이 되면 아주 빠른 속도로 추론할 수 있어 실시간 응용에 적합하다. 지금까지의 응용 예를 몇 가지 들자면, DNA 서열 분석, 고객의 은행 신용도 판별, 주가 변동 예측, 패턴 인식, 음성 인식, 무인 자동차 제어, 공장의 생산 라인 감시, 항공사 좌석 예약 관리 등이 있다.

3.2 진화 연산: 자연 현상을 모방한 연산 방식

현재로서도 진화가 진행되는 메커니즘이 완전히 규명된 것은 아니지만, 적어도 분자 생물학적인 관점에서 자연 선택(natural selection)은 염색체의 교환(crossover)과 돌연변이(mutation)로부터 발생하는 변화에 의해서 이루어진다는 것은 잘 알려진 사실 중의 하나이다. 염색체 교환은 기존의 유전자를 새로운 조합으로 결합시킨다. 돌연변이는 이전에 염색체에 포함되지 않았던 새로운 유전자를 생성한다.

유전자 알고리즘(genetic algorithms) 또는 진화 알고리즘(evolutionary algorithms)은 이러한 자연 세계의 진화 과정을 컴퓨터상에서 시뮬레이션을 통하여 복잡한 실세계의 문제를 해결하고자 하는 계산 모델이다[26]. 진화 알고리즘은 특히

적응적 탐색과 학습 및 최적화를 통한 공학적인 문제의 해결에 많이 응용된 바 있으며, 최근 들어 특히 신경망과 퍼지 로직과의 결합으로 그 응용 범위는 점점 늘어나고 있다.

진화 알고리즘은 기계학습 패러다임 중에서 극도로 경험적인 입장을 나타낸다. 진화 알고리즘은 풀고자 하는 문제에 대한 가능한 해들을 염색체(chromosomes)로 표현한 다음 이들을 점차적으로 변형함으로써 점점 더 좋은 해들을 생성한다. 각각의 가능한 해를 하나의 개체(individual)로 보며 이들의 집합을 개체군(population)이라 한다. 하나의 개체는 보통 한 개 또는 여러 개의 염색체로 구성되며 염색체를 변형하는 연산자들을 유전 연산자(genetic operators)라 한다. 전형적인 유전 연산자로는 복제, 교차, 돌연변이 연산자가 있다(그림 4). 진화 알고리즘이 기존의 탐색 방법과 구별되는 점 중의 하나는 점(point)에 의한 탐색이 아니라 군(population)에 의한 탐색이라는 것이다. 진화적 탐색에 있어서는 임의의 값들로 초기화된 가능한 해들의 집합 즉 탐색군으로 시작하여 세대 교체를 반복함으로써 점점 더 우수한 해의 집합들을 생성하려 시도한다.

- Reproduction: make copies of chromosome (the fitter the chromosome, the more copies)

10000100	→	10000100
		10000100
- Crossover: exchange subparts of two chromosomes

100100100	→	10011111
111111111		11100100
- Mutation: randomly flip some bits

00000100	→	00000000
----------	---	----------

그림 4 유전 연산자

진화 알고리즘 계열은 오랜 동안 다른 기계학습 접근 방법과는 비교적 독립적으로 발전해 왔으며, 따라서 자신만의 분석적 도구, 응용 분야와 학술회의를 발전시켜 왔다. 그러나 많은 근본적인 문제점들과 기법들은 귀납법의 주류 및 연결주의적 패러다임과 일맥상통한다. 예를 들어, 모든 경험적인 학습에서와 같이 목적 함수로써 측정되는 수행 성능의 변화에 대해 보상치(reward, credit)나 벌점(penalty, blame)을 설정하는 것은

어렵고 간접적으로 수행되는 일이다. 귀납적 접근 방법에는 이러한 문제를 해결하기 위한 많은 방법들이 있으며, 그 기원은 Samuel의 체커게임 학습 시스템으로 거슬러 올라간다. 유전자 알고리즘에서도 일종의 강화학습으로 bucket brigade 알고리즘을 발전시켰다. 보상치와 벌점의 부여 문제는 역전파 기법과 같은 모든 연결주의적 학습 방법에서 단연 중심적인 요소이다.

진화 알고리즘은 염색체를 표현하는 방법과 사용되는 유전 연산자의 종류 및 특성에 따라서 다시 여러 가지 모델로 구분된다. 그 하나는 흔히 GA라고 하는 협의의 유전자 알고리즘(genetic algorithm)으로서 보통 고정된 길이의 이진 스트링을 염색체로 사용한다[10]. 이에 반해 진화 전략(evolution strategy, ES)은 실수의 값을 취하는 유전자들로 구성된 벡터를 염색체로 사용한다. 그 밖에도 그래프와 트리를 염색체 표현에 사용하는 진화 프로그래밍(evolutionary programming, EP)과 유전 프로그래밍(genetic programming, GP) 등이 있다. 진화적 탐색에 사용되는 연산자로 EP와 ES는 돌연변이(mutation), GA와 GP는 교차(crossover) 연산자를 주로 사용한다. 역사적으로 볼 때 EP, ES, GA는 1960년대와 70년대에 개발되었으며 GP는 90년대에 들어와 연구되기 시작한 분야이다[8].

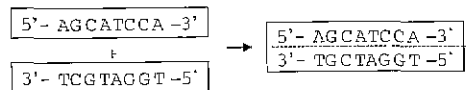
3.3 분자 컴퓨팅: 생체분자 기반의 정보처리 방식

DNA 컴퓨팅은 분자 생물학 분야의 발전에 힘입어 개발된 연구 분야로, 현재까지 밝혀진 여러 가지 유전자의 기능과 정보들을 기반으로 하여, 실제 생체 분자(bio-molecule)를 이용해 복잡한 문제들을 해결하고자 하는 새로운 계산 방법이다 [1]. 기존의 인공지능에서 연구해 오던 방법들은 진화연산이나 신경망처럼 대부분 생물학에서 개념만을 도입해서 계산 모델을 만들고, 이를 소프트웨어적으로 구현하여 시뮬레이션하는 방식이었다. 그러나 DNA 컴퓨팅은 이와는 달리 실제 생체 분자로 정보를 표현하고 저장하며 이를 이용해서 계산을 수행한다. 즉, 생물학적 방법들을 직접적으로 이용하고, 구현도 소프트웨어가 아닌 생물학적 기법을 사용하는 방식이다.

DNA 컴퓨팅 기법에서는 DNA 염기인

A(Adenine), C(Cytosine), G(Guanine), T(Thymine)로 정보를 표현하고, DNA 구조인 이중 가닥(double strand)이나 단일 가닥(single strand)으로 정보를 저장한다. 기존의 컴퓨터와 비교하자면 2 비트로 계산을 하던 것을 4 비트로 확장한 것으로 볼 수 있다. 또한 정보의 표현 방법이 전혀 다르기 때문에 연산자도 기존의 AND, OR, NOT 등과 같은 것은 사용할 수 없고, 생물학 실험실에서 실제로 사용하는 실험 방법들과 DNA의 특성들을 연산자로 사용한다. 예를 들어, Watson-Crick 상보 결합에 의해서 A는 T와 결합하고, C는 G와 결합하여 하이브리드를 형성(hybridization)하는 것이 가장 기본적인 연산자이며, 이 외에도 결합법(ligation), 중합효소 연쇄반응(PCR), 전기 영동법(gel electrophoresis) 등이 대표적인 연산자가 된다(그림 5와 6 참조). 이 외에도 여러 가지 효소(enzyme)들도 많이 이용되며, 기타 여러 가지 방법들이 바로 연산자로 사용될 수도 있다.

• Hybridization : 상보적인 염기쌍들이 결합을 형성



• Ligation : 점착말단(sticky end)들이 서로 결합하여 하나의 긴 염기배열을 형성

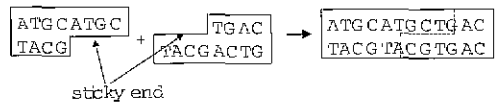


그림 5 DNA 컴퓨팅을 위한 분자 연산자

• PCR 특정 염기배열을 가지고 있는 DNA를 2ⁿ배만큼 증가시킴 (n 반복횟수)

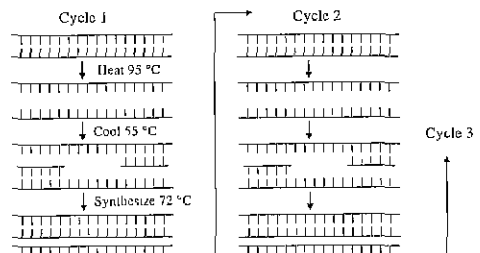


그림 6 DNA 컴퓨팅을 위한 PCR 방법

DNA 컴퓨팅 방식을 개략적으로 설명하자면, 우선 주어진 문제를 DNA 염기배열에 효과적으로 표현할 수 있도록 코딩 방법을 결정한다[24]. 그리고 절찰법 등을 이용해서 아주 많은 수의 DNA 염기 배열을 만들고, 중합효소 연쇄반응, 전기 영동법 등으로 주어진 문제의 조건을 만족하는 염기 배열만을 선택하여 주어진 문제의 해를 찾아내는 과정을 거친다. 이러한 과정을 효과적으로 수행하기 위한 여러 가지 새로운 bio-lab 기술들이 현재 연구되고 있다.

DNA 컴퓨팅이 부각되고 있는 이유는 DNA 분자가 가지고 있는 막대한 병렬성에 있다. 불과 1g 정도의 액체에 대략 3×10^{22} 개 정도의 DNA 분자가 존재하고, 이러한 막대한 수의 DNA 분자 하나 하나를 이용해서 정보를 저장하며 계산을 하기 때문에 그 병렬성은 아주 크다. 이런 병렬성을 이용하면 현재 기술로 해결하지 못하고 있는 여러 가지 난제들(NP-complete 문제 등)을 해결할 수 있을 것으로 기대되기 때문이다. 또한 계산 효율도 우수하다. 일반 PC는 대략 초당 10^6 번의 계산이 가능하고, 슈퍼컴퓨터라도 초당 10^{12} 번의 계산이 가능하다. 그러나 DNA 컴퓨터는 대략 초당 10^{14} 번의 계산이 가능하다.

실제로 DNA 컴퓨터 또는 분자 컴퓨터(molecular computer)를 만들기 위해서 튜링 기계(turing machine)나 유한 상태 기계(finite state machine), 불리안 회로(boolean circuit)를 생체분자를 이용하여 구현하는 연구가 진행되고 있다. 그 외 DNA의 정보 저장 능력을 이용해서 연상 기억 장치를 구현하고자 하는 연구와 전문가 시스템을 위한 초병렬 추론엔진을 DNA 컴퓨팅 기법으로 구현하는 연구도 있다.

4. 바이오지능: 인공지능의 미래

기존의 기호기반 인공지능의 단점을 해결하기 위한 방법으로서 최근에 와서 기계학습에 관한 많은 연구가 이루어지고 있다[14]. 또한 보다 새로운 접근 방식으로 생물학에 기반한 신경망 기반의 연결론적 인공지능, 유전자 알고리즘 기반의 진화론적 인공지능, 인공생명 기반의 인공지능[11, 20] 등이 시도되고 있다. 이러한 상향식(bottom-up) 또는 비기호적인(subsymbolic) 접

표 1 인공지능 연구 패러다임의 변화

Symbolic	==>	Subsymbolic
Knowledge-based	==>	Learning-based
Deduction	==>	Induction
Model-driven	==>	Data-driven
Top-down	==>	Bottom-up
High-level	==>	Low-level
Reflective	==>	Reflexive
Deep-thought	==>	Reactive behavior
Individual	==>	Collective
Syntactic	==>	Semantic
Discrete	==>	Continuous
Deterministic	==>	Stochastic
Logic	==>	Probabilistic

근 방법의 기반에는 학습과 진화의 개념이 내재되어 있다. 이것은 “지능적인 행동을 하도록” 프로그래밍을 하던 방식에서 탈피하여, “지능적인 행동이 발현되도록” 프로그래밍을 하는 방식으로의 사고의 전환을 의미한다. 기타 현재 변화하고 있는 컴퓨터기술 전반[4]과 관련되어 인공지능 연구 패러다임의 변화를 요약하면 표 1과 같다.

4.1 인공지능 연구 패러다임의 변화

진화와 적응학습 능력을 갖춘 인공 지능을 실현하기 위해서는 이에 대해 우수한 특성을 지닌 생물체 시스템의 정보 처리에 대하여 살펴볼 필요가 있다. 그림 7은 생물체와 기계의 계산 속도 및 기억 용량을 비교하고 있다[16]. 이 비교에서

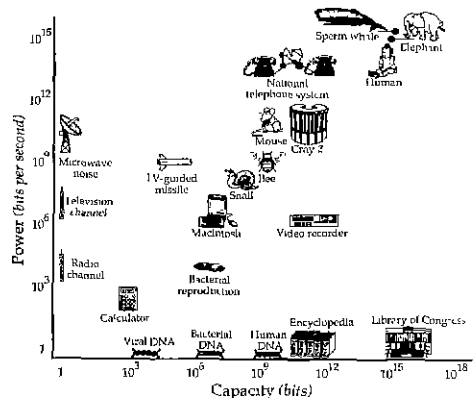


그림 7 컴퓨터와 생명체의 계산 속도와 기억 용량의 비교[16]

알 수 있는 바와 같이 슈퍼컴퓨터의 계산 속도나 기억 용량은 벌이나 쥐 보다 뛰어나지만 아직 이들 동물만큼 “지능적인 행동”을 보여주는 컴퓨터 프로그램 또는 로봇을 만들지는 못하고 있다. 이것은 인공지능의 실현이 막강한 계산 속도나 기억 용량뿐만 아니라 현존하는 자연지능 시스템의 구조와 정보처리 원리에 대한 보다 근본적인 이해를 바탕으로 해야 한다는 점을 시사한다.

지난 반세기 동안 생명 과학과 유전 공학의 급속한 발달로 인하여 점점 더 인간을 비롯한 생물체에 대한 비밀이 밝혀지고 있으며 이러한 분자(molecule)나 세포(cell) 수준에서의 과학적 발견에 기반한 새로운 정보 처리 모델들이 보다 고차원의 인공지능을 실현할 수 있는 새로운 지능 기술의 개발을 더욱 가속화시킬 것으로 기대된다. 또한 사회성을 지닌 동물이나 곤충들의 행동을 시뮬레이션하거나 생태계(ecology)의 변화를 컴퓨터상에서 합성함으로써 자연의 생명 현상과 생태계의 특성을 분석하고 이를 공학적으로 응용하려는 연구가 이론 생물학자뿐만 아니라 인공지능 관련 컴퓨터과학자와 로봇 공학자들의 많은 관심을 끌고 있다.

4.2 바이오지능 기술의 분류

모든 생물체는 DNA에 있는 정보로부터 단백질을 합성하여 세포를 형성하고, 세포분열을 통하여 발생과 성장 과정을 거친다. 동물들의 경우 경험으로부터 학습하며 주변의 환경에 능동적으로 적응하며 일생을 살아간다. 또한 짝짓기를 통해 자손을 낳으며 이 때 부모의 유전 형질이 후손에게 전달되고 변형됨으로써 진화해 나간다. 본 절에서는 이와 같은 생물학적 정보처리 원리에 기반한 지능형 정보처리 기술에 대하여 살펴보고자 한다.

생물체의 정보처리 소자의 단위는 분자, 세포, 개체, 개체군으로 구분할 수 있으며 이에 따라 바이오지능 기술을 구분하면 각각 molecular, cellular, organismic, ecological intelligence로 나눌 수 있다. 표 2는 이 구분을 요약하고 있으며, 이에 따른 바이오지능 기술의 특성은 표 3에 제시되어 있다. 지금까지 개체와 개체군 단계에서의 정보처리 과정을 수학적으로 모델링한 바이오지능 기술은 비교적 많이 연구되어 왔다. 그러

나 분자와 세포 단계에서의 바이오지능 기술의 개발과 그 응용은 아직 초보적인 수준에 머물러 있다. 아래에서는 분자 수준으로부터 개체군 수준에 이르기까지의 바이오지능 기술을 개괄적으로 살펴본다.

표 2 바이오지능 기술의 4가지 단계

Four Levels of Biointelligence
Molecular Intelligence (분자 지능)
Cellular Intelligence (세포 지능)
Organismic Intelligence (개체 지능)
Ecological Intelligence (군집 지능)

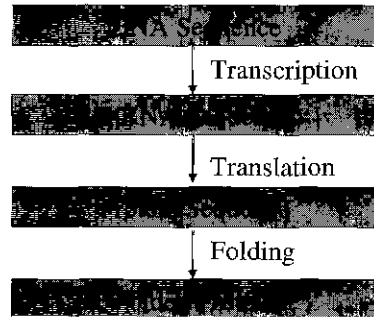


그림 8 분자수준의 생물 정보처리

분자수준의 바이오지능 기술인 Molecular Intelligence 기술은 막강한 정보 저장 능력과 생화학반응에 의한 초병렬 연산 속도를 가지는 분자들을 그 기본 단위로 한다. 분자들은 lock-key 메카니즘에 의해서 스스로 결합하는 self-assembly 능력과 자기조직 능력이 있으며[6] 용액 상태에서의 이와 같은 화학 반응을 연산 기작으로 활용함으로써 기존의 실리콘기반의 계산 방식 훨씬 우수한 연산 능력 및 분자수준에서의 적응학습 능력을 발휘할 수 있다. 대표적인 연산 모델로서 DNA 분자를 기반으로 하는 DNA 컴퓨팅이 있으며[24], 그 외에도 효소와 같은 기타 단백질의 합성과정을 기반으로한 분자컴퓨팅 방식이 연구되고 있다[2]. 분자수준에서의 지능형 정보처리 모델의 또다른 형태는 유전자 발현 과정을 시뮬레이션하는 Genomic Algorithm이다 [9]. 이는 그림 8에 보인 바와 같이 DNA에 저장된 유전 정보과 RNA를 통해서 전사되고 이를

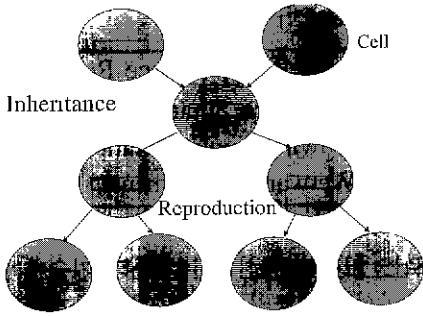


그림 9 세포 수준의 생물 정보처리 기작

번역하여 단백질이 합성되는 과정을 묘사하는 알고리즘으로서 현재 Genomics와 Proteomics의 핵심 연구 결과들을 바탕으로 한 바이오지능 기술이다. 이러한 계산 모델들은 반대로 Bioinformatics에서 유전자 발현 과정을 밝히는 계산 도구로 사용될 수 있어 분자생물학의 발전에 다시 기여할 수 있다[9].

세포수준에서의 바이오지능 기술인 Cellular Intelligence 기술은 세포간에 일어나는 정보처리 과정을 묘사하는 바이오지능 기술이다. 이는 Molecular Intelligence가 하나의 세포 안에서 일어나는 정보처리 과정을 묘사하는 것과는 대조적이다. 세포들이 복제되어 하나의 개체를 형성해 나가는 과정을 묘사한 Cellular Automata 모델은 이미 1960년대에 von Neumann에 의해 제안된 바 있으며[17], 최근 들어 이를 응용함으로써 self-repair 능력을 갖는 재구성 하드웨어 칩을 개발하려는 embryonics 분야가 새로이 개척

되고 있다[13]. 이러한 재구성 하드웨어는 자기진단이 가능하고 하드웨어 구성을 스스로 변경함으로써 칩의 고장이나 노화 및 환경의 변화에 대처할 수 있는 특징을 가진다. 세포수준에서의 바이오지능 기술의 또 다른 예는 Immune Network이다. 생체의 면역 시스템은 외부로부터 바이러스 등의 항원이 침입할 경우 이를 인지하고 항체를 만들어 이에 대처할 수 있는 세포 방어 체계를 가지고 있다. 사람의 경우 약 10만개 정도의 단백질로 구성된 자신의 세포와 분자를 가지고 약 10^{16} 개의 외부 침입 분자를 인식해 낼 수 있다고 알려져 있다. 이 면역 시스템은 과거의 경험으로부터 학습하고 기억하는 능력을 가지고 있어서 외부 세포의 공격을 방어할 수 있다. 이러한 연구는 이미 컴퓨터 보안과 바이러스 퇴치 등에 응용되고 있다[3].

개체수준에서의 Organismic Intelligence 기술은 하나의 유기체에서 일어나는 정보처리 과정에 초점을 두는 지능형 정보처리 기술이다. 특히 인공지능의 학습 모델로서 적합하며 대표적인 예로 신경망을 들 수 있다. 앞에서 언급한 바와 같이 기존의 전통적인 인공지능 시스템도 주로 개체 수준의 지능을 흉내내려 한 점에서는 유사하나, 여기서는 논리나 규칙 등의 기호기반의 표현 방식이 아니라 동물의 두뇌 구조와 기능을 모방하려고 하는 점이 다르다. 또한 신경망이 Neurophysiology에 기반한 모델인데 반해서 전통적인 인공지능에서의 semantic networks도 인간의 기억 구조를 모방하려 한 점에 있어서는

표 3 바이오지능 기술의 특성 비교

	Molecular Intelligence	Cellular Intelligence	Organismic Intelligence	Ecological Intelligence
Basic unit	molecules	cells	organism	population
Biology	molecular biology	cell biology	neurobiology	ecology
Phenomenon	self-assembly	development	learning	evolution
Time (typical)	seconds	days	months	years
Communication	lock-key mechanism	electrochemical signals	neuro-transmitters	audiovisual, symbolic
Basic operation	ligation hybridization	cell division differentiation	excitation inhibition	cooperation competition
Computational models	DNA/molecular computing	cell-automata immune nets	neural nets semantic nets	evolutionary algorithms
Chips	DNA chips	embryonic chips	neurochips	evolvable hardware

(다른 기호 기반의 지능 시스템 모델보다는) 바이오지능 기술에 더욱 유사하다고 할 수 있다. 최근 들어서 neural network과 semantic network의 구조를 융합한 형태로 볼 수 있는 확률신경망 모델들(probabilistic neural networks, graphical models)이 기억, 인지, 학습 모델로서 연구되고 있으며 이는 인지과학적으로 의미있는 신경심리학적인 바이오지능 기술로 분류될 수 있다. 확률신경망 모델은 복잡한 데이터에 대한 학습 능력이 뛰어나 인터넷 정보검색, 데이터마이닝, 사용자 인터페이스 등의 폭넓은 분야에 적용되고 있다.

개체군에 기반한 Ecological Intelligence 기술은 바이오지능 기술의 최상위 단계로서 다수의 개체들로 구성된 집단을 단위로 하는 지능형 정보처리 모델이다. 이러한 모델은 생태계의 변화나 사회성 동물이나 곤충들의 세계를 모사함으로써, 복잡한 공학적인 문제를 해결하거나 사회현상을 설명하는 방법으로 사용될 수 있다[11]. 핵심 기반 기술은 진화 연산으로 볼 수 있으며, 기존의 분석적인 방법이나 수식으로 표현이 불가능한 복잡 시스템을 연구하는 데 잘 적용되고 있다. 특히 이러한 방법론은 경제학에서는 경제현상을 모델링하기 위해 사용되고 있으며 (evolutionary economics), 최근 물리학에서 complex adaptive systems의 emergent behavior 연구, 공학적으로 환경의 오염이나 교통체증 현상의 원인 분석, 정보 기술 분야에서 복잡한 multi-agent systems의 협력 체계 구축, 인터넷 게임이나 virtual community 형성 과정 등을 연구하는데 응용되고 있다.

이와 같은 분자, 세포, 유기체 및 생태계를 모사하는 지능기술의 연구는 미래에 정보 처리 방식에서 뿐만 아니라 계산 매체에 있어서도 변화가 예상되며 지금처럼 소프트웨어에 의해 지능적 행동을 흉내내려는 시도뿐만 아니라 하드웨어(실리콘) 상에서 직접 온라인으로 학습하고 진화하는 진화 하드웨어에 관한 연구가 활발히 진행 중이다. 여기서 한 걸음 더 나아가 유기체의 특성을 자연스럽게 이용한 바이오웨어(탄소) 기반의 인공지능 실현도 시도될 것이다. 이러한 연구는 한편으로는 획기적인 컴퓨터 신기술의 창조와 새로운 응용 분야의 개척을 유도할 것이며, 다른

표 4 바이오지능 연구 방법

Methods for Biointelligence Research
Software Simulation
Hardware Implementation
Wetware Experiments

한편으로는 이러한 혁신적인 컴퓨터 기술에 힘입어 보다 향상된 인공지능 구현 기술의 개발을 가져올 것이다. 또한 발달된 바이오지능 기술은 다시 생물학의 연구 특히 Bioinformatics(생물정보학) 분야에 적용되어 더욱 많은 생물학적 발견을 가져올 것이다.

5. 요약 및 전망

본 고에서는 생물학과 생물정보학 및 인공지능의 학제적 연구 분야로서 바이오지능(Biointelligence)을 정의하고 네 가지의 구체적인 연구 영역을 소개하였다. 이것은 분자 수준의 Molecular Intelligence, 세포 수준의 Cellular Intelligence, 개체 수준의 Organismic Intelligence, 개체군 수준의 Ecological Intelligence이며, 각각 화학반응에 의한 초병렬적 추론(Inference), 다세포 유기체의 구조형성과정(Morphogenesis), 신경계의 학습과정(Learning) 및 개체군의 진화 과정(Evolution)을 모사하는데 특히 적합함을 제시하였다.

앞서 언급한 바와 같이 현재의 바이오지능 기술은 신경망 기반의 개체 지능 기술과 진화 연산 방식의 개체군 지능 기술에 집중되고 있으며, 상대적으로 분자나 세포 수준의 바이오지능 기술에 대한 연구는 아직 초보적인 단계에 머물러 있다. 그러나 앞으로 분자생물학과 생물정보학, 분자전자공학(molecular electronics), 생명공학, 유전공학, 바이오칩(biochip) 기술 등의 성과에 힘입어 장기적으로는 이 분야에서도 획기적인 발전을 이룩할 가능성이 높다. 특히 Genomics와 Proteomics의 연구 성과에 힘입어 분자수준의 생체 정보처리 기작이 밝혀지고 이를 조작하는 방법들이 개발되면 그 파급 효과가 아주 큰 획기적인 지능형 정보 기술을 기대할 수 있을 것이다.

마지막으로 인공지능의 관점에서 부언할 사항은 생물학기반의 지능 기술에 관한 연구가 중요

하고 필연적이라고 하여서 기존의 기호기반의 지능형 시스템에 대한 연구가 필요하지 않다는 뜻은 아니라는 점이다. 오히려 단기적인 응용 가능성, 산업적 기대 효과에 있어서 인간의 지적 정보처리 능력을 보조해 주는 여러 가지 지능형 에이전트의 역할은 앞으로도 계속 중요할 것이다. 왜냐하면, 인간의 많은 지적 활동이나 의사소통이 주로 언어와 그림 등의 기호에 기반해 있기 때문이다. 그러나 단순한 기계적인 기호 처리 방식의 단점과 궁극적인 한계를 뛰어넘기 위해서는 겉으로 드러나는 지능적 행동(behavior)뿐만 아니라 지능을 발현시키는 시스템의 기반 물질의 성분(substrate)과 구조(architecture) 및 그 표상의 형성 과정(process)에 관한 기반 연구가 함께 수행되어야 할 것이다[5]. 즉, 궁극적으로 인간의 지능을 잘 모사하기 위해서는 인간의 기억 구조와 기억을 형성하는 생물학적 하부 구조의 이해에 바탕을 둔 새로운 기억 장치와 정보처리 기작의 이해에 기반하여 유기체의 특성으로부터 나오는 적응, 학습, 발생, 진화 현상에 대해서 연구되어야 할 것이다. 왜냐하면 궁극적으로 생물체인 인간의 지능을 가장 자연스럽게 흉내낼 수 있는 것은 유기체인 인간 그 자체일 것이기 때문이다.

참고문헌

- [1] Adleman, L.M., Molecular computation of solutions to combinatorial problems, *Science*, 266:1021-1024, 1994.
- [2] Conrad, M. and Zauner, H.-P., Molecular computing with artificial neurons, *Communications of KISS*, Special issue on Bioinformatics, August 2000.
- [3] Dasgupta, D. (ed.), *Artificial Immune Systems and Their Applications*, Springer-Verlag, 1998.
- [4] Denning, P.J. and Metcalfe, R.M., *Beyond Calculation: The Next Fifty Years of Computing*, Copernicus, 1997.
- [5] Dreyfus, H. L., *What Computers Still Can't Do: A Critique of Artificial Reason*, MIT Press, 3rd Ed., 1992.
- [6] Eigen, M. and Schuster, P., *The Hypercycle: A Principle of Natural Self-Organization*, Springer-Verlag, 1979.
- [7] Fogel, L., Owens, A., and Walsh, M., *Artificial Intelligence Through Simulated Evolution*, Wiley, 1966.
- [8] Fogel, D., *Evolutionary Computation: The Fossile Record*, IEEE Press, 1998.
- [9] Gilbert, R. J., Rowland, J. J., and Kell, D. B., Genomic computing: Explanatory modelling for functional genomics, *Proc of Genetic and Evolutionary Computation Conference*, Morgan Kaufmann, pp. 551-557, 2000.
- [10] Holland, J.H., *Adaptation in Natural and Artificial Systems*, The University of Michigan Press, 1975.
- [11] Langton, C.G., *Artificial Life: An Overview*, MIT Press, 1995.
- [12] Luger, G.F. (Ed.), *Computation & Intelligence*, MIT Press, 1995.
- [13] Mange, D., Sipper, M., Stauffer, A., and Tempesti, D., Toward robust integrated circuits: The embryonic approach, *Proceedings of the IEEE*, Vol. 88, No. 4, April 2000.
- [14] Mitchell, T.M., *Machine Learning*, McGraw-Hill, 1997.
- [15] Minsky, M., *The Society of Mind*, MIT Press, 1985.
- [16] Moravec, H., *Mind Children*, Harvard University Press, 1988.
- [17] Neumann, J. von, *Theory of Self-Reproducing Automata*, (edited and published by A.W. Burks), Urbana, 1966.
- [18] Nilsson, N.J., *Artificial Intelligence: A New Synthesis*, Addison-Wesley, 1998.
- [19] Rumelhart, D. and McClelland, J.L., *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*, MIT Press, 1986.
- [20] Steels, L. and Brooks, R. (Eds.), *The Artificial Life Route to Artificial*

Intelligence, Lawrence Erlbaum, 1995.

- [21] Turing, A.M., Computing machinery and intelligence, *Mind*, 59:422-460, 1950.
- [22] Winston, P.H., *Artificial Intelligence*, Addison Wesley, 3rd Ed., 1992.
- [23] Zhang, B.T., *Learning by Genetic Neural Evolution*, ISBN 3-929037-16-5, 268 pages, Infix-Verlag, 1992.
- [24] Zhang, B.T. and Shin, S.-Y., Molecular Algorithms for Efficient and Reliable DNA Computing, *Proceedings of the Third Annual Genetic Programming Conference (GP-98)*, Morgan Kaufmann, pp. 735-742, 1998.
- [25] 김성동, 장병탁, "바이오데이터 마이닝을 위한 기계학습 기법", 정보과학회지, Bioinformatics 특집호, 2000년 8월.
- [26] 장병탁, "인공 진화에 의한 학습 및 최적화", *제어자동화시스템학회지*, 제1권 제3호, pp. 52-61, 1995.
- [27] 장병탁, "분자생물학 기반의 인공지능 기술에 관한 연구", 서울대학교 공과대학 연구보고서, 1997.

- [28] 장병탁, "인공지능의 과거 현재 미래", 교학사, 최양희 외 16인 공저, *21세기의 컴퓨터공학*, 제11장, pp. 176-200, 1998. <http://scai.snu.ac.kr/~btzhang/>

장 병 탁



- 1986 서울대학교 컴퓨터공학과 학사
- 1988 서울대학교 컴퓨터공학과 석사
- 1992 독일 Bonn 대학교 컴퓨터과학과 박사
- 1988~1992 Bonn 대학교 AI Lab 연구원
- 1992~1995 독일국립정보기술연구소(GMD) 연구원
- 1995~1997 건국대학교 컴퓨터공학과 조교수
- 1997~현재 서울대학교 컴퓨터공학과 부 조교수

관심분야: 인공지능, 기계학습, 신경망, 진화연산, 분자프로그래밍
 E-mail: btzhang@scai.snu.ac.kr
<http://scai.snu.ac.kr/~btzhang>

● 2000년 하계 컴퓨터통신 워크숍 ●

- 일 자 : 2000년 8월 24 ~ 25일
- 장 소 : 상록리조트(천안)
- 주 체 : 정보통신연구회
- 문 의 처 : 연세대학교 전자공학과 이재용 교수
Tel. 02-361-2873 E-mail: jyl@nasla.yonsei.ac.kr