DNA 컴퓨팅을 이용한 혼 절 논리 정리 증명

박의준[0] 남진우 이인희 장병탁
서울대학교 대학원 인지과학 협동과정[0]
서울대학교 컴퓨터공학부 바이오지능 연구실
{ejpark[0], jwnam, ihlee, btzhang}@bi.snu.ac.kr

Theorem Proving in Horn-Clause Logic Using DNA Computing

Eui-Jun Park[0] Jin-Wu Nam In-Hee Lee Byoung-Tak Zhang
Interdisciplinary Program in Cognitive Science[0],
Biointelligence Laboratory, School of Computer Science and Engineering
Seoul National University, Seoul 151-742, Korea

요 약

원숭이와 바나나 문제는 논리적 추론에 의한 문제 해결 과정을 설명하기 위해 사용되는 대표적 예제이다. 본 논문에서는 전통적인 접근 방식과는 달리, 이 문제를 그래프 탐색의 그것으로 이해한 후 DNA 컴퓨팅에 근거한 너비 우선 탐색(breadth-first search, BFS)을 통해 해들을 발견하고자 한다. 그 결과, 최적해(optimal solution)를 포함한 최소 4개 이상의 다양한 해들이 실제 DNA 생화학 실험을 통해 확인되었다.

1. 서 론

원숭이와 바나나 문제(Monkey & Banana Problem, MBP)는 논리적 추론에 의한 문제 해결 과정을 설명하기 위해 인공지능 관련 문헌에서 자주 언급되는 예제이다. 예컨대 선언적 언어(declarative language)인 Prolog를 사용하면 이것은 상태 공간 내 상태 벡터들의 순차적인 탐색을 통해 쉽게 해결된다[1]. Prolog는 일종의 혼 절(Horn-Clause) 논리이므로 결국 MBP는 혼 절 논리의 한 정리(theorem)로 간주될 수 있다.

한편 DNA 컴퓨팅을 이용하여 혼 절 논리의 정리들을 증명하고자 한 몇몇 연구들이 있어 왔다[2,3,4]. 그러나 이들 모두는 그 제시하는 방법의 복잡함으로 인해 DNA 분자들을 이용한 실험으로는 구현되지 못한 상태이다. 반면, 본 논문의 실험 절차는 현재 널리 쓰이는 기법들, 즉 hybridization, ligation, PCR, Gel 전기영동만을 포함하는 단순한 것이고, 따라서 실제 실험이 가능하다.

MBP의 해를 찾는 것은 상태 공간의 탐색을 통한 풀이 경로의 발견으로 이해할 수 있다. 전통적인 규칙 기반 인공지능에서는 이 경우 순차적인 탐색으로 최단 경로를 거쳐 목표 상태에 도달하게 된다. 이에 반해, 대규모 병렬 처리라는 DNA 컴퓨팅의 장점을 이용하면, 상태 공간 내 거의 모든 상태들을 원리적으로 상수 시간 내에 탐색할 수 있다. 따라서 최적해 외에 다양한 해들의 발견이 기대된다. 본 논문의 접근 방식이 갖는 또 하나의 부가적인 장점은 혼 절 논리에서 절들의 순서를 더 이상 고려하지 않아도 되므로, 백트래킹(backtracking) 기법이나 추론 중 무한 루프에 빠지는 위험을 사전에 자연스럽게 배제시킬 수 있다는 것이다.

2. 원숭이와 바나나 문제

MBP는 다음과 같이 주어진다:

어떤 방에 원숭이, 상자, 그리고 바나나가 있다. 바나나를 얻기 위해 원숭이는 어떤 동작들을 수행하여야 하는가? 즉 원숭이는 상자 있는 곳으로 가서(walk), 상자를 바나나 있는 곳까지 밀고(push), 상자 위로 올라가(climb), 바나나를 잡아야(grasp) 하는 것이다.

MBP의 가능한 상태는 다음과 같이 벡터로 표현된다:

(w, x, y, z)
(이 때, w: 원숭이 위치, x: 원숭이가 상자 위에 있는가, y: 상자 위치, z: 원숭이가 바나나를 가졌는가)

이제 이것에 대해 다음과 같은 4개의 연산을 정의함으로써 원숭이의 행동을 나타낼 수 있게 된다:

① *walk*(u): $(w, 0, y, z) \rightarrow (u, 0, y, z)$
② *push*(v): $(w, 0, w, z) \rightarrow (v, 0, v, z)$
③ *climb*: $(w, 0, w, z) \rightarrow (w, 1, w, z)$
④ *grasp*: $(c, 1, c, 0) \rightarrow (c, 1, c, 1)$

따라서 초기 상태와 목표 상태는 그림 1과 같다. 여기서 원숭이, 상자, 바나나의 초기 위치 조건은 각각 a와 b와 c이다.

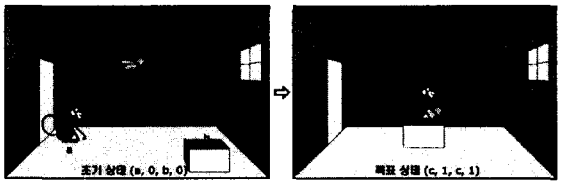

그림 1. MBP의 초기 상태와 목표 상태

예컨대 MBP를 Prolog로 구현할 경우 해답이 도출되기까지의 경로는 그림 2에 나타나 있다. 이 때, 찾아진 경로는 최단 경로이며, 또 오직 최단 경로만이 찾아진다. 이는 연산자 적용 순서, 곧 절(clause)의 순서 때문이다.

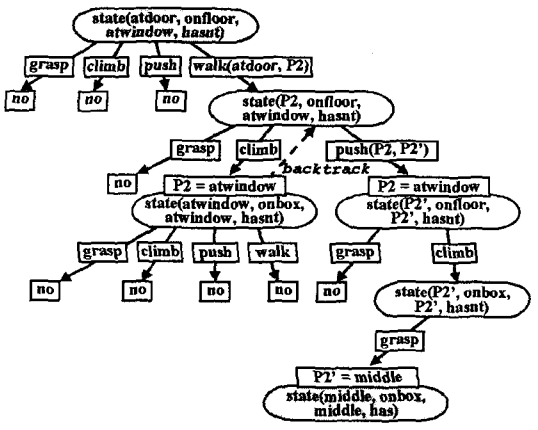

그림 2. Prolog 프로그램의 추론 과정

3. DNA 컴퓨팅 기법을 이용한 MBP 표현 및 해결

3.1. MBP 상태 공간의 표현

본 논문에서는 원숭이의 위치와 상자의 위치가 취할 수 있는 값을 '창 옆(at window)'과 '방 가운데(middle)'의 2 가지로만 제한한다. 따라서 가능한 상태와 연산의 개수는 각각 6 개와 8 개이다. 만약 필요하다면 상태 공간의 크기는 얼마든지 늘일 수 있으므로, 이러한 제한이 문제 풀이의 일반성을 잃게 만드는 것은 아니다.

상태들에 대한 DNA 시퀀스들은 표 1에 나타나 있다. 실제 시퀀스 디자인 작업은 서울대 컴퓨터공학부 바이오지능 연구실에서 개발한 시뮬레이터인 NACST를 써서 수행되었다[5].

표 1. 6 가지 상태들을 나타내는 DNA 시퀀스들 (a: at door, b: middle, 상태 0: 초기 상태, 상태 5: 목표 상태)

상태 (번호)	시 퀀 스 (5' - 3')
a 0 a (0)	GACCCTGCTG GAGAGACTAC CCCTTTAGAC
a 0 b (1)	CCCGCCCTAG CCATTCAGCA CTGTTACGCA
a 1 a (2)	GTAGTTCATG CCACAAGCAC ACTATTTGGA
b 0 a (3)	CAACTTGAGA TTGTCACGGC TGGGCGAGGC
b 0 b (4)	GTTCGTTGCG GTTCCTCGGC CCTAAAAGTC
b 1 b (5)	ATGCACTATG ATTGCGCTCT AGTCACACAC

연산에 대한 시퀀스 또한 표 1에서 알 수 있다. 왜냐하면 하나의 연산-시퀀스는 그것이 연결시켜 주는 두 상태-시퀀스 각각의 절반에 대해 상보적이어야 하기 때문이다. 예컨대 표 1의 상태 0에서 상태 3으로 걸어가는 연산에 대한 시퀀스는 그림 3과 같다.

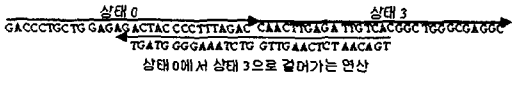

그림 3. 연산-시퀀스의 한 예

3.2. 실험 설계

실험의 대략적인 과정은 그림 4와 같다.

(1) 테스트 튜브에서 DNA 시퀀스들 혼합
(2) hybridization & ligation: 원숭이가 이동할 수 있는 다양한 경로 생성
(3) PCR: 초기 상태와 목표 상태를 모두 포함하는 경로들만 분리
(4) gel 전기영동: PCR 결과 생성된 해들 확인

그림 4. 실험 절차 개요

단계 (2)를 통해 MBP의 해가 생성될 수 있다. 예컨대 그림 5는 6 개의 상태로 구성된 하나의 해를 보여 준다.

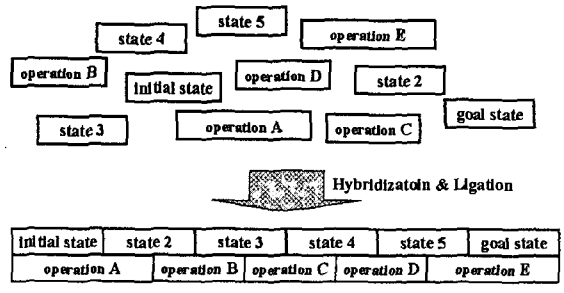

그림 5. 해 생성 과정 도식

이와 같이 생성된 해들은 단계 (3)을 거쳐 증폭된 후, 단계 (4)에서 그 존재가 확인된다.

3.3. 실험 프로토콜

그림 4의 절차를 보다 상세히 기술하면 다음과 같다.

(1) 올리고의 Hybridization

각 올리고를 100 pmol 농도로 섞은 후, 초기 denature 위해 95 도에서 5 분 간 유지한다. 다음으로 hybridization을 위해 iCycler thermal cycler (Bio-rad, USA)를 이용해 온도를 분당 1 도씩 내려 16 도까지 낮춘다.

(2) Hybridization 거친 올리고의 ligation

Ligation 과정은 T4 DNA ligase를 처리한 후 16 도에서 12 시간을 iCycler thermal cycler를 이용하여 유지한다. 반응 완충용액은 50 mM Tris-HCl (pH 7.8), 10 mM MgCL2, 5 mM DTT, 1 mM ATP, and 2.5 g/ml BSA. 3.2.6을 포함하고 있다.

(3) 올리고 정제를 위한 전기영동

Ligation된 올리고는 2 % Agarose gel에서 전기영동을 진행한다. 전기영동 완충용액은 100 mM Tris-HCl, pH 8.3, 89 mM boric acid, and 2 mM EDTA을 포함하고 있다. 전기영동 동안에는 Bio-rad Model Power PAC 3000 electrophoresis unit을 이용하여 100 W (6V/cm)를 유지한다. 그 결과, gel 상에서 정확한 정답에 대응하는 사이즈의 밴드(band)들과 답이 아닌 사이즈의 밴드들을 모두 확인할 수 있었다. 이것들을 PCR을

거치게 하여, 정답의 밴드들만을 다시 확인할 수 있다.

(4) 올리고 정제

올리고 정제는 QIAquick Gel Extraction Kit (QIAGEN, USA)를 이용하여 진행된다. 사용된 정제 kit에는 QG buffer, PE buffer, EB buffer, PB buffer가 포함되어 있다. 정제를 위해 전기영동을 진행한 gel에서 정답을 포함할 수 있는 모든 범위의 gel을 잘라 tube에 넣은 후 QG 완충용액을 3 배의 부피로 넣어준다. 다음 vortexing 후 50 도에서 10 분간 유지하고, 다시 isopropanol을 처리한다. 이 tube를 13000 rpm에서 원심 분리한 후 PE 완충용액 0.75 ml을 첨가하여 씻어준다. 다음으로 다시 13000 rpm에서 원심 분리한다. 원심 분리된 샘플을 elution을 위해 EB완충용액(10 mM Tric-Cl, pH 8.5)이나 증류수를 넣어 보관한다.

(5) 정답의 PCR 증폭

PCR을 위해 primer와 template를 10 pmol 농도로 PCR premix(10 mM Tris-HCL, pH 7.8, containing 50 mM KCl, 1.5 mM MgCL2, 0.1% Triton X-100, 0.2 mM dNTP, and 1 U DNA Taq polymerase (Genenmed, Korea)와 섞어 총 20 ul 부피로 만든 후 PCR 반응시킨다.

(6) 결과 확인을 위한 전기영동

증폭된 PCR 결과물을 다시 2% gel 상에서 전기영동하여 확인한다.

3.4. 실험 결과

실험 결과는 그림 6과 같다. 여기서 M은 사이즈 마커(size marker)를, L1부터 L6까지는 3 개의 테스트 튜브와 2 종류의 온도 조건에 따른 서로 다른 PCR 결과물들을 나타낸다. L1/L2, L3/L4, L5/L6 각각은 동일한 튜브 내 올리고들에 대해 전기영동을 건 결과인데, 이 때 L1, L3, L5는 PCR 시 Anealing의 온도 조건을 58℃로 한 것이고, L2, L4, L6는 62 ℃로 한 것이다. 그 결과, 전기영동 상 4 개의 밴드가 확인되었고, 이것은 특히 L5에서 가장 뚜렷하게 보인다.

관찰된 4 개의 밴드는 미리 예측한 해답과 모두 일치한다. 예컨대, 원숭이가 바나나를 얻게 되는 가장 간단한 경로, 즉 최적해는 방문 옆에 있던 원숭이가 상자로 가서, 그 상자 위로 바로 올라가는 것인데 이것은 3 개의 상태 (at door, on floor, middle), (middle, on floor, middle), (middle, on box, middle)로 구성되어 있다. 본 실험에서 하나의 상태는 30 bp로 코딩된 시퀀스로 표현되고, 초기 상태와 목표 상태를 나타내는 PCR primer로는 각각 25 bp의 시퀀스가 사용되므로 가장 짧은 해답-시퀀스는 80 bp인데, 이것은 그림 6에서 보이는 가장 아래 밴드에 정확히 상응한다. 실험 전 예상되는 밴드의 길이는 80 bp, 140 bp, 200 bp, 260 bp, 320 bp 등인데, 이들 중 140 bp 길이의 시퀀스를 제외한 나머지 4 가지 종류가 검출되었다. 그러나 아주 희미하게나마 140 bp 밴드도 존재하는 것을 알 수 있다(특히 L1, L2).

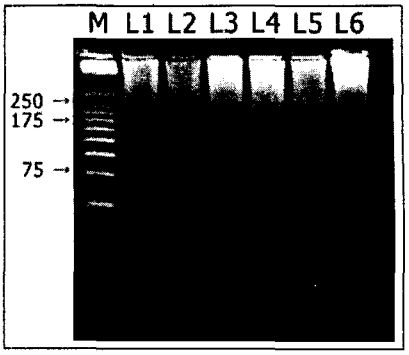

그림 6. 결과에 대한 electrophoretogram

4. 결 론

지금까지 DNA 분자들을 이용한 생물학 실험을 통해 MBP의 다양한 해들이 생성됨을 확인하였다. DNA 컴퓨팅 시스템 자체는 고정된 구조를 갖고 있지 않은 극단적인 연결주의 시스템으로 간주될 수 있으며, 여기서 발생하는 확률적인 생화학 반응은 복수 해의 탐색과 추론 과정 중 절차적 관점의 배제를 규칙 기반의 전통적 접근에 비해 용이하게 하여 준다. 요컨대 본 논문에서는 기호적 표상의 하위 차원에 아날로그 방식을 적용시킨 새로운 혼합 시스템의 예와 그것의 특성들을 제시하였다.

감사의 글

본 연구는 산자부 차세대 신기술 사업 수퍼 지능칩 과제의 Molecular Evolutionary Computing(MEC)에 의해 지원되었음. 이 연구를 위해 연구장비를 지원하고 공간을 제공한 서울대학교 컴퓨터연구소에 감사 드립니다.

참고 문헌

[1] Bratko, I., *Prolog Programming for Artificial Intelligence*, 2nd Ed., Addison-Wesley, 1990.

[2] Mihalache, V., Prolog Approach to DNA Computing, *Proceedings of the International Conference on Evolutionary Computation*, pp. 249-254, 1997.

[3] Kobayashi, S., Horn Clause Computation with DNA Molecules, *Journal of Combinatorial Optimization*, 3, pp. 277-299, 1999.

[4] Uejima, H., Hagiya, M. and Kobayashi, S., Horn Clause Computation by Self-Assembly of DNA Molecules, *Preliminary Proceedings of the Seventh International Meeting on DNA Based Computers*, pp. 63-71, 2001.

[5] Zhang, B.-T. and Shin, S.-Y., Molecular Algorithms for Efficient and Reliable DNA Computing, *Proceedings of the Third Annual Genetic Programming Conference*, pp. 735-742, Morgan Kaufmann, 1998.