

애너그램 문제 해결과정의 분자적 시뮬레이션

천효선^{1,2}, 이지훈^{1,3}, 류제한^{1,4}, 백다솜¹, 장병탁^{1,2,3,4,5}

서울대학교 ¹바이오지능 연구실, ²컴퓨터공학부, ³생물정보학 협동과정, ⁴뇌과학 협동과정, ⁵인지과학 협동과정
 {hschun, jhlee, jhyu, dsbaek, btzhang}@bi.snu.ac.kr

Molecular Experimental Simulation for Cognitive Anagram Solving

Hyo-Sun Chun^{1,2}, Ji-Hoon Lee^{1,3}, Je-Hwan Ryu^{1,4}, Christina Baek¹, Byoung-Tak Zhang^{1,2,3,4,5}

¹Biointelligence Laboratory, ²Department of Computer Science and Engineering, ³Graduate Program in Bioinformatics, ⁴Brain Science Program, ⁵Cognitive Science Program, Seoul National University

요약

본 연구에서는 인간이 애너그램(anagram) 문제를 해결하는 과정을 분자컴퓨터를 이용하여 시뮬레이션 하였다. 애너그램은 주어진 문자들을 재배열하여 숨겨진 단어를 찾아내는 놀이로, 문제를 빨리 풀어내는 사람들은 제약 만족 네트워크(constraint satisfaction network)의 병렬적 탐색에 의해 문제를 해결한다. 본 연구에서는 이러한 인지적 현상을 모델링한 분자 애너그램 모델을 실험을 통해 검증하였다. 문자와 bi-gram, 단어들을 DNA 서열로 인코딩하고, hybridization을 통해 제약네트워크를 병렬적으로 탐색해나간다. 탐색한 결과를 전기영동으로 확인한 결과 정답과 오답을 구분해낼 수 있었다.

1. 서론

Adleman에 의해 시작된 DNA 컴퓨팅은 DNA 분자가 가진 자가조립적(self-assembly) 성질과 그로 인한 병렬성(parallelism), 그리고 뛰어난 정보저장능력으로 인해 기존의 실리콘 기반 컴퓨터가 풀기 힘든 NP-hard 문제들을 풀어내는 것으로 주목받아왔다[1][2]. 최근에는 이러한 능력과 계산매체의 본질적인 생체 친화성에 주목하여 인간의 인지적 현상을 DNA 컴퓨팅으로 구현하려는 시도가 계속되고 있다[3][4].

본 연구에서는 애너그램 문제를 해결하는 인지적 과정을 분자컴퓨터를 이용하여 시뮬레이션 하였다. 애너그램(anagram)이란 주어진 문자들의 순서를 바꾸어서 숨겨진 새로운 단어를 찾아내는 단어놀이를 말한다. Novic의 연구에 따르면, 애너그램 문제를 푸는 전략은 search와 pop-out으로 나눌 수 있다[5]. Search는 주어진 문자들의 모든 배열을 순차적으로 확인해보는 방법으로, 문제를

늦게 풀어내는 사람들에게서 많이 나타난다. 반면, pop-out은 제약 만족 네트워크(constraint satisfaction network)를 병렬적으로 탐색하는 과정으로, 문제를 빨리 풀어내는 사람들에게서 많이 나타난다. pop-out은 기호적(symbolic) 처리과정과 하위기호적(subsymbolic) 처리과정의 상호작용에 의해 일어나는데, 이러한 특성으로 인해 인간의 인지 현상을 분석에 중요한 역할을 한다[6].

그동안 분자컴퓨터 이용해 애너그램 문제의 인지적 해결과정을 모사하는 연구가 꾸준히 진행되어왔다[7][8]. 본 연구에서는 기존의 분자 애너그램 모델에서 나아가 구체적인 분자실험 방법을 제시하고, 실험을 통해 결과를 검증하였다.

2. 분자 애너그램 모델

그림 1은 분자 애너그램 모델의 문제 해결 과정을 간략히 요약한 것이다. 우선 문제로 주어진 문자들을 학습을 통해 형성한 bi-gram 사전과 비교하여 문제를 푸는데 필요한 bi-gram를 찾는다. 이 과정은 제약조건을 만족하는 요소들을 찾아내는 과정으로, 문제를 빨리 풀어내는 사람은 늦게 푸는 사람들에 비해 양질의 제약조건들을 더 많이 학습했기 때문에 답을 빨리 찾아내게 된다. 두 번째 단계에서는 찾아낸 bi-gram을 학습을 통해 형성한 단어 사전과 비교하여 답이 될 수 있는 단어를 찾는다.

분자 애너그램 알고리즘의 기본 틀은 [8]을 기반으로 한다. 각 문자들을 일정한 길이의 DNA 서열로 인코딩하고, bi-gram과 단어는 그들을 이루는 문자들에 대해 상보적인 염기서열 조합으로 구성하였다. 각 단계에서 DNA hybridization을 통해 bi-gram 사전과 단어 사전을 병렬적으로 탐색한다. 탐색한 결과는 전기영동을 통해 확인할 수 있다.

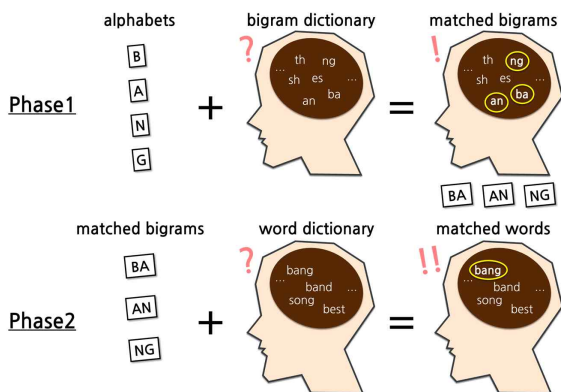


그림 1. Pop-out process. (Phase1) find matched constraints(or bigrams) from alphabets and bigram dictionary. (Phase2) find correct answer from matched constraint and word dictionary.

3. 실험 및 결과

문제로 주어지는 알파벳이 {B, A, N, G}이고, 분자컴퓨터가 가진 bi-gram 사전은 {BA, AN, NG, ND}, 단어 사전은 {BANG, BAND}일 경우에 대해 실험을 수행하였다. 실험이 제대로 되었다면 문자 B, A, N, G로부터 정답인 단어 BANG과 오답인 단어 BAND를 구분할 수 있어야한다.

첫 번째 단계의 실험에서 문자와 bi-gram 사전을 결합하여 조건에 만족하는 bi-gram과 만족하지 않는 bi-gram을 구분할 수 있음을 확인했다. (그림2-(a)). 따라서 조건에 만족하는 bi-gram만을 추출해낼 수 있다(그림2-(b)). 두 번째 단계의 실험에서는 찾아낸 bi-gram들을 단어 사전과 결합시켜서 정답(BANG)과 오답(BAND)을 구별해낼 수 있었다(그림3-(a)).

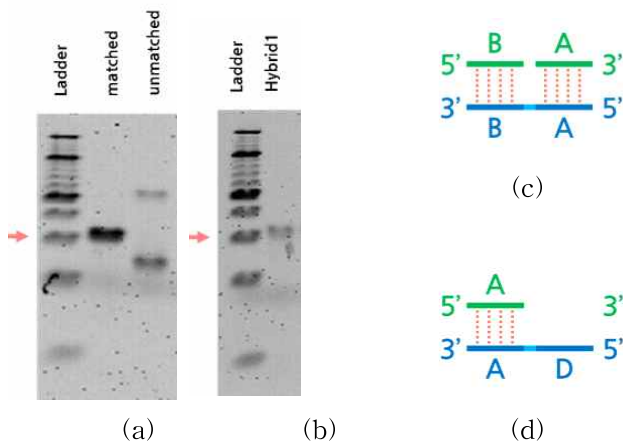


그림 2. (a) Comparison between matched constraint and unmatched constraint (b) Result of a process for searching satisfied constraints (c) Structure of matched constraint (d) Structure of unmatched constraint

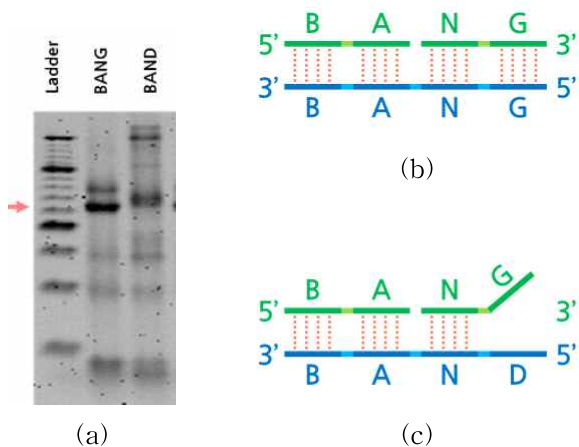


그림 3. Result of matched word(BANG) and unmatched word(BAND) (b) Structure of correct answer (c) Structure of incorrect answer

4. 결론

본 논문에서는 분자컴퓨터를 이용하여 인간의 애너그램 문제 해결과정을 모사할 수 있음을 보였다. 현재의 분자

애너그램 모델이 기존 컴퓨터를 이용한 풀이에 비해 성능이 좋지 못한 것은 분명한 사실이다. 그러나 본 연구의 가장 큰 의의는 분자컴퓨터로 인지 현상을 시뮬레이션 한 것에 있다. 문제의 크기가 늘어날수록 기존 컴퓨터가 필요로 하는 연산은 매우 많아지지만, 분자컴퓨터의 경우에는 병렬성으로 인해 크게 늘어나지 않는다. 또한 현재의 모델에 제약조건에 대한 학습 가중치를 활용하는 과정을 추가한다면 분자컴퓨터를 이용한 풀이의 장점이 보다 명확히 드러날 것으로 본다.

감사의 글

이 논문은 미공군연구소의 지원(FA2386-12-1-4087)과 정부(미래창조과학부)의 재원으로 한국연구재단의 지원(NRF-2010-0017734-Videome)을 받아 수행된 연구임

참고문헌

[1] Adleman, L. M., "Molecular computation of solutions to combinatorial problems", Science, Vol. 266, No. 5187, pp. 1021-1024, 1994.

[2] Braich, R. S., Chelyapov, N., Johnson, C., Rothmund, P. W. & Adleman, L., "Solution of a 20-variable 3-SAT problem on a DNA computer", Science, Vol. 296, No. 5567, pp. 499-502, 2002.

[3] Qian, L., Winfree, E. & Bruck, J., "Neural network computation with DNA strand displacement cascades", Nature, Vol. 475, No. 7356, pp. 368-372, 2011.

[4] Lakin, M. R., Minnich, A., Lane, T. & Stefanovic, D., "Towards a biomolecular learning machine", In Proceedings of Unconventional Computation and Natural Computation, pp. 152-163, 2012.

[5] Novick, L. R. & Sherman, S. J., "The effects of superficial and structural information on online problem solving for good versus poor anagram solvers", The Quarterly Journal of Experimental Psychology, Vol. 61, No. 7, pp. 1098-1120, 2008.

[6] Grimes, D. B. & Mozer, M. C., "The interplay of symbolic and subsymbolic processes in anagram problem solving", Advances in neural information processing systems, pp. 17-23, 2001.

[7] 이은석, 윤지은, 장병탁, "DNAGram: Anagram 문제 해결에 관한 분자컴퓨팅 연구", 한국인지과학회, 2003.

[8] 이지훈, 천효선, 이은석, 류제환, 장병탁, "분자컴퓨터를 통한 애너그램 시뮬레이션", 한국정보과학회 가을 학술발표 논문집, 제40권, pp. 723-726, 2013.