

# 하이퍼네트워크 분류기의 분자컴퓨팅 구현 방법

류제환<sup>01</sup> 장병탁<sup>1,2,3</sup>

<sup>1</sup>서울대 뇌과학 협동과정

<sup>2</sup>서울대 생물정보학 협동과정

<sup>3</sup>서울대 컴퓨터공학부

{jhryu, btzhang}@bi.snu.ac.kr

## Molecular Computing Realization Method of Hypernetworks Classifier

Je-Hwan Ryu<sup>01</sup> DaSom Baek<sup>1</sup> Ji-Hoon Lee<sup>2</sup> Byoung-Tak Zhang<sup>3</sup>

<sup>1</sup>Interdisciplinary Program in Neuroscience, Seoul National University

<sup>2</sup>Interdisciplinary Program in Bioinformatics, Seoul National University

<sup>3</sup>Department of Computer Science & Engineering, Seoul National University

### 요 약

분자 컴퓨팅은 기존의 실리콘 기반 컴퓨팅과 비교하여 대규모의 메모리를 분자의 종류 및 농도로 표현할 수 있으며, 동시다발적인 분자간의 화학 반응을 통해 대규모 병렬 처리를 자연스럽게 수행할 수 있다. 따라서 분자 컴퓨팅은 빅 데이터를 사용하는 기계학습 및 패턴인식 문제를 푸는 데에 좋은 단서를 제공할 수 있다. 분자 컴퓨팅은 보통 화학 반응 네트워크를 통해 그 형태가 표현되고 미래가 예측된다. 하지만 일반적인 화학 반응 네트워크는 반응이 단방향으로만 진행되어 복잡한 기계학습 알고리즘을 표현하기 어려우므로, 본 연구에서는 2상 화학 반응 네트워크라는 시스템을 제안하고 이를 통해 하이퍼네트워크 분류 알고리즘을 분자 컴퓨팅적인 관점에서 해석 및 구현하였다.

### 1. 서론

분자 컴퓨팅은 DNA, RNA, 단백질 등의 각종 화학 물질을 매체로 컴퓨팅을 수행하려는 학문이다. 실제 분자의 화학 반응을 컴퓨팅에 사용한다는 점 때문에 프로세스의 조절과 정밀한 조작이 힘들지만, 화학 반응의 동시다발성에 의하여 본질적인 병렬 컴퓨팅이 가능하며 분자 메모리의 스케일이 거대하여(1mol = 10<sup>23</sup>개의 분자) 컴퓨팅의 새로운 가능성을 제시하고 있다. 이러한 특성을 이용하여 초기의 분자 컴퓨팅에서는 TSP 등의 문제를 상수 시간 알고리즘으로 푼 바가 있다[1]. 특히 많은 데이터를 사용하는 기계학습 및 패턴인식 등의 계산집약적인 문제는 분자 컴퓨팅의 대규모 병렬 컴퓨팅을 이용하기에 적합하다. 보통 이러한 분자 컴퓨팅은 화학 반응 네트워크(CRN)라는 비선형 시스템 형식에 의해 정형적으로 표현되며, 분자간 화학반응을 모사하여 시간에 따른 이 시스템의 행동을 예측할 수 있다. 하지만 화학 반응 네트워크는 자발적인 반응에 의해 진행되기 때문에[2] 시스템이 단순한 행동을 보이지

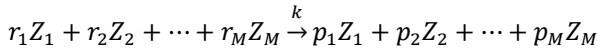
쉬운 반면, 기계학습은 학습과정에서 변수를 갱신하는 과정이 필요하여 복잡한 행동이 요구된다. 따라서 본 논문에서는 2상 화학 반응 네트워크라는 시스템을 제안하고 이 형식을 기계학습 분류 알고리즘에 적용하여 결과를 확인하고자 한다.

하이퍼네트워크 모델은 본디 분자 컴퓨팅의 아이디어에서 출발하였기 때문에[3] 분자 컴퓨팅과 알고리즘적으로 호환이 자연스러우며, 데이터의 분포 표현과 추론 과정 중에 동시다발적인 대용량의 처리가 필요하다. 따라서 하이퍼네트워크 알고리즘은 화학 반응 네트워크를 통해 표현하기가 수월하므로, 본 연구에서는 2상 화학 반응 네트워크를 적용하는 기계학습 알고리즘으로 하이퍼네트워크 알고리즘을 선택하였다.

### 2. 2상 화학 반응 네트워크

다음 모델은 M개의 화학종  $z = \{Z_1, Z_2, \dots, Z_M\}$  이 어떻게 반응하며 그 농도가 시간에 따라 어떻게 변화하는지 나타낸다. 각 반응은 다음과 같은 일반적

형태를 가진다.



이 일반적 표현식에서 대부분의  $r_m \in \mathbb{N}, p_m \in \mathbb{N}$  은 0의 값을 가진다. 왼쪽의 0이 아닌 원소들은 반응물로 불리며 반응 도중 소비된다. 오른쪽의 0이 아닌 원소들은 생성물로 불리며 반응 도중 생성된다. 반응에 참여하지만 생성이나 소비가 일어나지 않는 종은( $r_m = p_m$ ) 촉매라고 불린다.

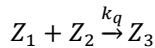
화학종의 집합이 주어진 상황에서 화학 반응 시스템은 다음과 같이 정의된다[4].

$$CRN = \{S, C, R\}.$$

이 때  $R$  은  $Q$ 개의 반응의 집합  $R = \{R_1, R_2, \dots, R_Q\}$  으로 구성된다. 각 반응은 세 개의 반응 파라미터로 이루어져 있다.

$$R_q = (r^q, k_q, p^q)$$

예를 들어, 화학종  $Z_1$ 과  $Z_2$ 가 반응하여  $Z_3$ 가 생성되는 반응  $R_q \in R$ 이 있다고 했을 때, 반응물 벡터  $r^q$ 는  $r_1^q = r_2^q = 1$  을 제외한 나머지는 모두 0이며, 생성물 벡터  $p^q$ 는  $p_3^q = 1$  을 제외한 나머지가 모두 0이다. 반응에 참여하지 않는 화학종을 모두 제거하고 반응식을 다시 써 보면 다음과 같다.



본 연구에서 제안하는 2상 화학 반응 네트워크는 위의 화학 반응 시스템에서 두 번째 반응 집합과 전환 시간 집합이 추가된 형태이다.

$$2\text{phase CRN} = \{S, C^1, C^2, R^1, R^2, T\}.$$

이 때  $t_n \in T, t_0 = 0$  에 대해서  $n$ 이 자연수일 때 실제 반응 집합  $R$ 은

$$R(t) = \begin{cases} R_1 & \text{if } t_{2n} \leq t < t_{2n+1} \\ R_2 & \text{if } t_{2n+1} \leq t < t_{2(n+1)} \end{cases}$$

로 시간에 따라 변하게 된다.

### 3. 하이퍼네트워크 분류 알고리즘의 표현

하이퍼네트워크 모델에서 분류 알고리즘은 다음과 같다.  $L$ 을 학습된 하이퍼네트워크의 하이퍼에지 라이브러리며,  $X_n$ 가  $n$ 번째 테스트 데이터에서 추출한 하이퍼에지들의 집합이라고 생각하면

1.  $L$ 에서  $X_n$ 와 일치하는 하이퍼에지를  $M$ 에 저장한다
2.  $M$ 에서  $y = 0$ 인 하이퍼에지를  $M_0$ 에 저장하고,  $y = 1$ 인 하이퍼에지를  $M_1$ 에 저장한다.
3.  $y^* = \text{argm } ax_{y \in \{0,1\}} |M^y|/|M|$

이는 2상 화학 반응 네트워크에서 다음과 같이

표현될 수 있다[5,6].

$$S = L \cup X \cup \{d\}$$

$$C_1 = \{(s_1, s_2) | s_1, s_2 \in S, m \text{ at } (s_1, s_2) = \text{True}\}$$

$$C_2 = \{s | s \in S\}$$

$$R_1$$

$$= \{(r, k, p) \mid \text{for } (s_1, s_2) \in C_1, \text{if } s = s_1 \text{ or } s_2, r_s = 1, \text{ else } r_s = 0 \text{ if } s = d, p_s = 1, \text{ else } p_s = 0\}$$

$$R_2$$

$$= \{(r, k, p) \mid \text{for } s_1 \in C_2, \text{if } s = s_1 \neq d, r_s = 1, \text{ else } r_s = 0 \text{ } p = 0\}$$

$$T = \{t_{\text{start}}, t_{\text{shift}}, t_{\text{end}}\}$$

이 때  $R_1$ 은 라이브러리와 테스트 데이터간의 일치와 저장을 나타내는 반응이며,  $R_2$ 는  $d$ 를 제외한 화학종이 붕괴하는 반응이다. 최종적으로는 일치하지 않은 모든 하이퍼에지는 소멸하게 되며, 이 때 분류 결정은

$$y^* = \text{argm } ax_{y \in \{0,1\}} |d^y|/|d|$$

으로 판단할 수 있다.

### 4. 실험 결과

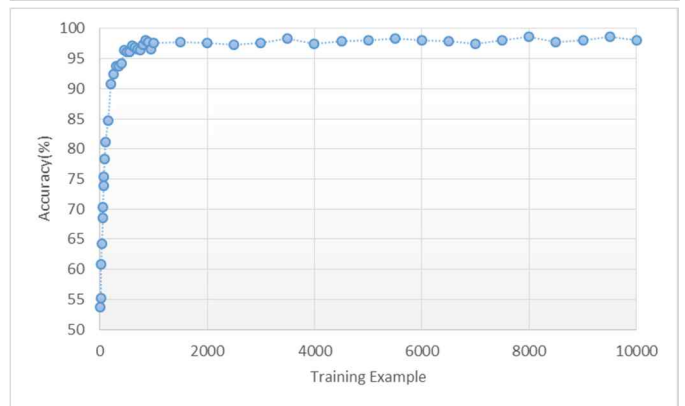
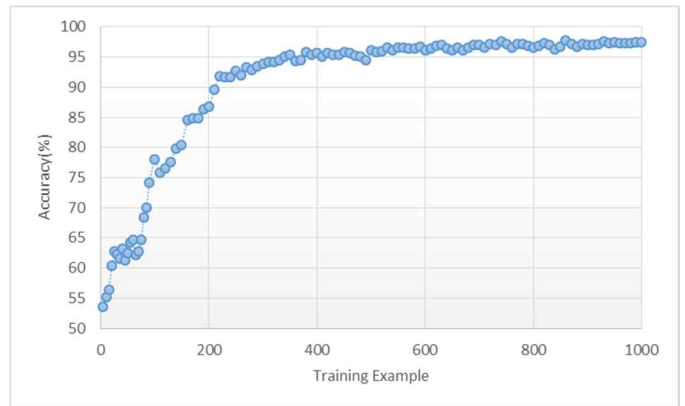


Figure 1. 하이퍼네트워크의 필기체 인식 학습 곡선 (위) 0-1000의 학습 데이터 (아래) 전체 학습 데이터

실험에 쓰인 데이터는 MNIST 필기체 데이터 중 6과 7에 해당하는 학습 및 평가 데이터이다. 각 레이블별로 학습 데이터는 약 5000개씩이었으며, 평가 데이터는 약 1000개씩이었다. 다음은 사용된 학습 데이터가 늘어날 때의 분류 정확도를 나타낸 그래프이다. 각 이미지에서 1000개의 하이퍼에지를 추출하였으며, 모든 하이퍼에지는 3-order로 고정되었다. ~1000 학습 데이터에서 95%의 정확도를 보이며, 최종 정확도는 약 98%에 달한다. 위와 같이 학습된 하이퍼네트워크를 2상 화학 반응 네트워크로 전환시켜 반응을 지켜본 결과는 다음과 같다.

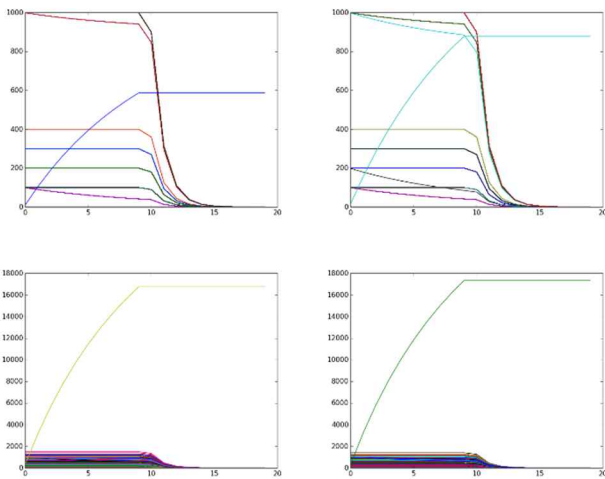


Figure 2. (상) training=500일 때 class별 반응 농도  
(하) training=10000일 때 class별 반응 농도

반응 농도 그래프는 전반부와 후반부로 확연히 구분되며, 최종적으로는 목표된 반응 결과물만 실험관 내에 남고 다른 화학종은 사멸한다. 분류는 남게 된 반응 결과물의 농도로 쉽게 이루어진다.

### 5. 결론

본 논문에서는 복잡한 화학 반응에 적합한 2상 화학 반응 네트워크 시스템을 제안하고 이를 통하여 하이퍼네트워크 기계학습 모델을 분자 컴퓨팅적인 관점에서 해석할 수 있음을 보였다. 나아가 기존의 알고리즘이 분자컴퓨팅 알고리즘으로 변환이 가능하며 실험관에 존재하는 화학종의 농도를 통해 분류 과정이 정상적으로 진행될 수 있음을 확인했다. 이후에는 발전된 연구로서 학습 및 분류 과정이 한 분자적 알고리즘 안에서 이루어지는 것이 필요하다.

### 감사의 글

이 논문은 삼성전자 미래기술육성센터의 지원(SRFC-IT1401-12)을 받아 수행된 연구이다.

### 참고문헌

[1] Adleman, L. M., Molecular computation of solutions to combinatorial problems, *Science*, 266(5187), p. 1021-1024, 1994.  
 [2] Zhang, B. T., and Park, C. H., Self-assembling hypernetworks for cognitive learning of linguistic memory, *International Journal of Intelligent Technology*, 3(1), 2008.  
 [3] Zhang, B.-T., Hypernetworks: A molecular evolutionary architecture for cognitive learning and memory, *Computational Intelligence Magazine*, IEEE3.3, p. 49-63, 2008.  
 [4] Napp, N. E., and Adams, R. P., Message Passing Inference with Chemical Reaction Networks, In *Advances in Neural Information Processing Systems*, p. 2247-2255, 2013.  
 [5] Lee, J.-H., Lee, S.H., Chung, W.-H, Lee, E.S., Park, T.H., R. Deaton, and Zhang, B.-T., A DNA assembly model of sentence generation, *BioSystems*, 106, p. 51-56, 2011.  
 [6] Lee, J.-H., Kim, J.-W., R. Deaton, Lee, S. H., Park, T. H., and Zhang, B.-T., Molecular machine learning in vitro, *International Conference on DNA Computing and Molecular Programming (DNA 18)*, p.55, 2012.  
 [7] Kim, J.-K. and Zhang, B.-T., Evolving hypernetworks for pattern classification, *IEEE Congress on Evolutionary Computation (CEC 2007)*, pp.1856~1862, 2007.  
 [8] Zhang, B.-T. and Jang, H.-Y., A Bayesian algorithm for in vitro molecular evolution of pattern classifiers, *International Conference on DNA Computing and Molecular Programming (DNA10)*, pp.294-303, 2004.