

시험관 내 경쟁 반응을 통한 가중치 합의 분자 계산

임희웅⁰¹, 이승환², 양경애³, 유석인¹, 박태현², 장병탁³

서울대학교, 전기컴퓨터공학부, 바이오지능연구실², 인공지능 & 컴퓨터비전연구실¹

서울대학교, 화학생물공학부, 세포 및 미생물공학연구실³

hwlim@bi.snu.ac.kr; skulsh78@snu.ac.kr; kayang@bi.snu.ac.kr; siyoo@bi.snu.ac.kr;

thpark@plaza.snu.ac.kr; btzhang@bi.snu.ac.kr

Molecular Computation of Weighted Sum via Competitive Reaction *In Vitro*

Hee-Woong Lim⁰¹, Seung Hwan Lee², Kyung Ae Yang³, Suk-In Yoo¹,

Tai Hyun Park², and Byoung-Tak Zhang³

Seoul National University, School of Computer Science and Engineering,

Biointelligence Lab.², Artificial Intelligence and Computer Vision Lab.¹,

Seoul National University, School of Chemical and Biological Engineering,

Cell and Microbial Engineering Lab.³

요 약

최근 생체 분자를 이용한 정보 처리 기술의 발전은 분자 수준에서의 생화학적 정보 처리의 가능성을 제시하였으며 더 나아가 유전자 분석이나 질병진단과 같은 의료 목적의 응용 개발을 목표로 하고 있다. 이러한 분자 정보 처리 기술은 기존의 실리콘 컴퓨터 기반의 방법론과는 달리 생화학적 정보를 전기적 신호로의 변환 과정 없이 시험관 내에서 직접 처리하며 결과적으로 정보 손실을 최소화하며 최종 결과를 얻기까지의 절차를 단축시킨다. 그러나 현실적으로 응용 가능한 시스템의 개발은 해당 분자 정보의 특성에 잘 맞는 새로운 모델과 정보 표현 방법, 그리고 새로운 기본 연산자의 도입을 필요로 한다. 본 논문에서는 다양한 패턴 분석 모델의 기본연산이라 할 수 있는 가중치 합 연산의 분자 컴퓨팅 버전을 제시하는데, 분자간의 특이적인 혼성화와 그에 따른 경쟁 반응을 이용하여 가중치를 표현하는 일반화된 모델을 제안한다. 그리고 간단한 바이오-랩 실험 결과를 통해서 본 모델의 가능성을 제시한다.

1. 서 론

최근 생체분자와 이에 대한 생화학적인 실험 과정을 이용하여 정보를 표현하고 처리하는 분자 컴퓨팅(molecular computing)이 새로운 계산 패러다임으로 대두되었다. 특히 DNA나 RNA와 같은 인간의 유전정보를 구성하는 물질들이 이러한 분자 컴퓨팅에 많이 사용되었는데, 이는 해당 분자들간의 상보성에 따른 특이적인 결합 및 혼성화라는 특징에 기인한다. A, T, G, C로 구성되는 이러한 분자의 염기서열을 이용한 정보 표현은 그 무수히 많은 조합론적 다양성으로 인해 정보 표현 매체로서의 뛰어난 이점을 가지고 있으며, 뿐만 아니라 해당 염기 서열을 적절히 디자인 함으로써 분자들간의 결합 반응을 조절하여 원하는 연산을 수행 가능 하도록 하는 것이다[1].

초창기 분자 컴퓨팅은 대용량 병렬성에 영감을 받아 NP 문제와 같은 복잡한 계산 문제의 해결에 대한 연구가

많이 이루어 졌다. 그러나 화학 반응의 특성상 시험관 내에서 생화학 분자로 표현된 정보의 처리를 위한 실험 방법이 사람의 수작업으로 이루어지고 불확정성을 안고 있는 화학 반응의 특성으로 인해 아직까지 현실적인 크기의 문제에는 적용되지 못하고 있다. 한편 최근 연구의 흐름은 이러한 분자 컴퓨팅이 연산 매체로서 생화학 분자를 사용한다는 사실에 주안점을 두어 생물학적인 문제 해결을 위한 새로운 방법론으로서의 분자 컴퓨팅에 대한 연구가 활발하게 이루어지고 있다. 그 예로서, 유전자 발현 패턴의 분석을 위한 방법[2], 유전자 발현의 논리적 조절을 위한 분자 오토마타[3]를 들 수 있다. 게놈 프로젝트의 완료 이후 관심을 받고 있는 유전자 정보의 분석은 정보처리의 관점에서 볼 때 생화학 분자 형태로 존재하는 수치 정보로 생각할 수 있다. 해당 유전자의 발현 수준에 따라 수치의 높고 낮음이 결정되는 것이다. 지금까지의 생물정보학적 방법론들은 마이크로 어레이[4][5] 방법과 같은 측정 방법을 이용하여 생화학

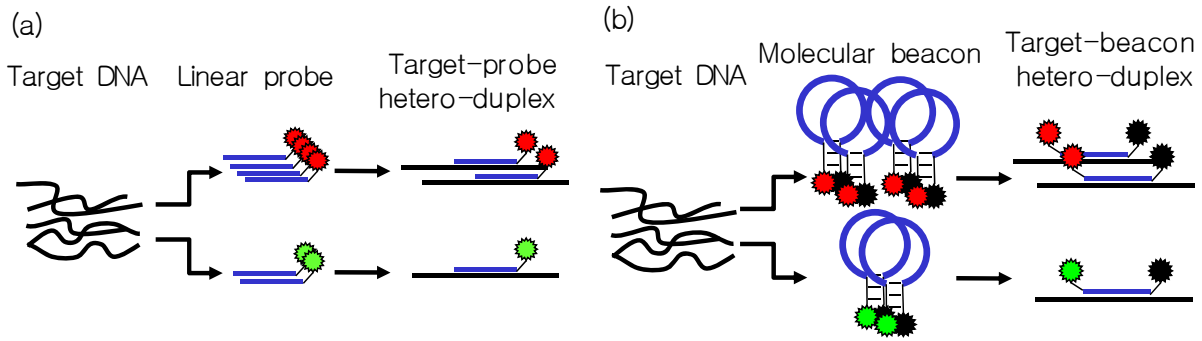


그림 1. 경쟁 반응의 예. (a) 목표 DNA와 서로 다른 형광으로 라벨링된 선형 프로브 (linear probe) 분자들 간의 경쟁적 혼성화 반응. (b) 목표 DNA와 서로 다른 형광을 가진 분자 비컨 (molecular beacon) 간의 경쟁적인 반응.

적 분자 정보를 전기적인 형태로 변환한 후, 전산 및 통계 분야의 다양한 패턴 분석 방법을 이용하여 결과를 얻어내는 것을 기본으로 하고 있다. 그러나 그러한 생화학 분자를 연산 매체로 사용하는 분자 컴퓨팅 방법을 이용할 경우 정보의 변환 과정을 생략하고 분자 수준에서 직접 유전 정보를 처리하여 더 빠르고 효과적인 문제 해결을 가능하게 하는 것이다.

효과적인 분자 패턴 분석을 위해서는 해당 패턴 분석 모델의 시험관 내 구현을 위한 수치 정보의 표현 방법과 그러한 정보의 처리가 가능한 대수적 연산이 필요하다. 과거 벡터나 행렬 연산을 위한 DNA 기반의 방법들이 제안된 바 있으나[6][7] 그 구현을 위한 실험 절차가 복잡하다는 단점이 있어 아직 실험적으로 검증되지 못했다. 이러한 한계를 극복하기 위해 범용의 복잡한 행렬 연산 대신 다양한 패턴 분석 모델에 응용 가능한 간단한 기본연산의 구현에 집중하여[8], DNA 기반의 가중치 합 연산이 제안되어 실험적으로 구현된 바 있고 효과적인 수치정보 표현을 위한 열역학 기반의 방법이 제안되었다[9][10]. 본 논문에서는 이러한 분자 기반의 가중치 합 연산을 확장하여 경쟁적인 혼성화 반응을 바탕으로 하는 일반화된 가중치 합 연산 모델을 제시하는데, 기존의 가중치 합 연산이 선형 DNA 간의 단순 혼성화 반응에 바탕을 두고 있는 반면 일반화된 모델은 분자 비컨이나 [11] 압타머와 [12] 같은 특이적인 결합이 가능한 다양한 생화학 분자로 확장하고 있다.

본 논문은 다음과 같이 구성되어 있다. 다음 절에서는 가중치 합 연산의 기본 단위를 이루는 경쟁 반응과 개념적인 모델을 설명하고, 3절에서는 이러한 경쟁 반응을 어떻게 일반화된 가중치 합 연산으로 표현 가능한 지에 대해서 기술한다. 그리고 4절에서는 바이오-랩 실험 결과를 제시하고 5절에서 결론을 맺는다.

2. 경쟁 반응

현재의 실리콘 기반의 전자식 컴퓨터는 AND, OR, NOT

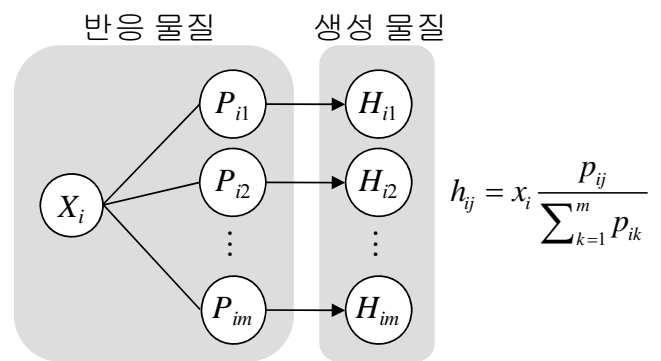


그림 2. 경쟁적 혼성화 반응. X_i 는 입력 분자, P_{ij} 는 X_i 에 특이적으로 결합하는 프로브 분자, H_{ij} 는 X_i 와 P_{ij} 가 결합하여 생성되는 헤테로 분자. 각 프로브 분자들은 목표 분자와 결합하는데 있어서 열역학적으로 동등함을 가정한다. 수식의 소문자는 해당 분자들의 양을 나타낸다.

등의 Boolean 논리 게이트만으로 정보처리에 필요한 모든 연산을 구현할 수 있다는 이론을 바탕으로 하여 모든 정보를 0과 1의 이진 정보로 표현하여 처리한다. 여기에서 실제 물리적인 구현은 전기적인 신호에 의한 이진 정보의 표현과 실리콘 기반의 반도체 칩 상의 Boolean 논리 회로에 의해서 이루어진다. 한편, 이와 같이 정보 처리의 이론적인 측면과 구현의 측면을 분리해서 생각해 볼 때 정보 처리의 매체는 실리콘 기반의 전기적 신호에 제한되지 않고 자연계에 존재하는 어떠한 현상도 이용 가능하다는 것이 바로 자연 컴퓨팅 (natural computing)의 기본 아이디어이다.

분자 컴퓨팅은 자연 컴퓨팅의 한 분야로서 생화학 분자를 이용하여 정보를 표현하고 그들 간의 화학 반응을 통해서 정보를 처리한다. DNA나 RNA와 같은 핵산 (nucleic acid) 분자는 해당 분자를 구성하는 염기 서열에 따라서 상호간에 특이적으로 결합하는 성질이 있으며 그 염기 서열을 적절하게 디자인 함으로서 반응을 조절할 수 있다. 이러한 생체 분자간의 결합 반응은 시험관 내에서 무수히 많은 분자들에 의해서 병렬적으로 수행

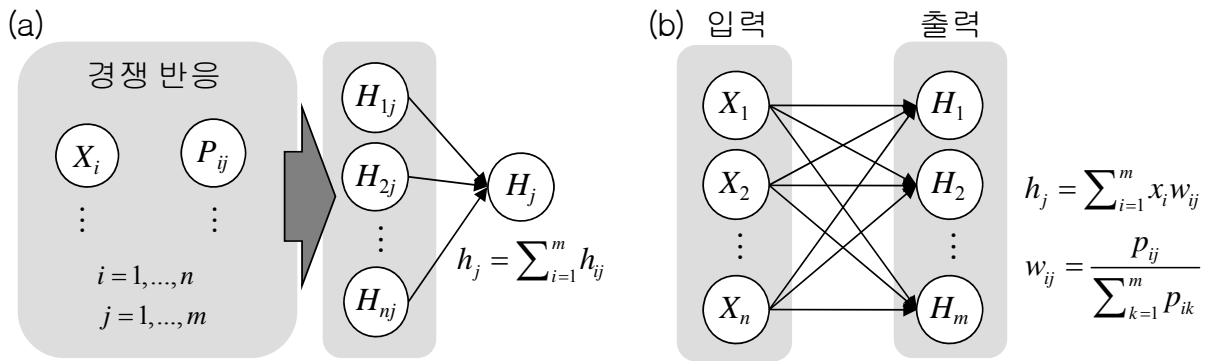


그림 3. 가중치 합 연산으로서의 경쟁 반응. (a) 가중치 합 계산 결과를 표현하는 목표 분자는 H_j 에 의해서 표현되며 이는 모든 i 에 대해 H_{ij} 의 양을 합한 것이다. (b) H_j 는 X_i 분자들에 의해 표현되는 입력에 대한 가중치 합 연산 결과를 표현하는데, 이때 해당 가중치는 X_i 에 대한 기본 반응에 참가하는 프로브들의 전체 양에 대한 개별 프로브의 비율로서 표현된다.

되는데, 반응에 참여하는 분자들을 어떻게 디자인 하느냐에 따라서 특정 목표 분자에 대해 서로 다른 몇몇 분자들이 경쟁적으로 결합하는 시스템을 구성할 수 있다. 경쟁 반응의 예는 다양하게 찾아 볼 수 있는데, 유전자 발현 패턴을 측정하기 위한 마이크로어레이 방법에서와 같이 서로 다른 형광 물질로 표지된 선형 프로브 분자가 목표 RNA 혹은 cDNA에 경쟁적으로 혼성화되거나, 혹은 목표 분자에 결합함으로써 형광을 발하여 목표 분자의 양적 정보를 얻을 수 있도록 하는 분자 비컨 이 서로 다른 형광 물질을 가지고서 공통의 목표 분자에 결합하는 것을 들 수 있다(그림 1). 뿐만 아니라 경쟁적 혼성화 반응은 단지 핵산 분자간의 반응뿐만 아니라 단백질과 해당 목표 단백질에 특이적으로 결합하는 압타머 간의 반응에서도 찾아볼 수 있다.

그림 2는 이러한 경쟁 반응에서의 반응 물질과 생성 물질간의 관계와 각 해당 물질들의 양적 관계를 보여주고 있다. 한 종의 목표 m 종의 프로브 분자 P_{ij} 가 결합하여 헤테로 분자 H_j 를 생성한다고 할 때 반응 물질과 생성물질 간의 양적 관계는 다음 수식으로 나타낼 수 있다.

$$h_j = x_i \frac{P_{ij}}{\sum_{k=1}^m P_{ik}} \quad (1)$$

여기에서 x_i 는 입력 분자의 양을 나타내므로 항상 0 이상의 값을 갖는다. 이 관계는 반응에 참가하는 프로브 분자 P_{ij} 들이 목표 X_i 에 결합함에 있어서 그 열역학적인 특성이 모두 동등하다는 것을 전제로 하고 있다.

3. 가중치 합 연산

앞서 기술한 경쟁 반응은 반응 물질과 생성 물질에서

수치 정보로의 적절한 mapping과정을 통해서 가중치 합 연산으로 규정될 수 있다. 먼저 가중치 정보의 표현 방법에 대해서 살펴보자면, 그림 1에서 보여주는 개별 경쟁 반응에서 생성되는 각 헤테로 분자 H_j 의 양을 결정하는 요소는 다음의 두 가지이다.

- a) 초기 입력 분자의 양, x_i
- b) X_i 와 반응하는 프로브 P_{ik} ($k = 1, \dots, m$)의 전체 양에 대한 P_{ij} 의 비율

이 때, 전체 P_{ik} ($k = 1, \dots, m$)에 대한 P_{ij} 의 비율을 다음과 같이 w_{ij} 라 정의하자.

$$w_{ij} = \frac{P_{ij}}{\sum_{k=1}^m P_{ik}} \quad (2)$$

수식 (1)과 (2)에서 볼 수 있는 바와 같이, w_{ij} 는 h_{ij} 를 생성함에 있어 입력 X_i 에 대한 가중치로 생각할 수 있고, 따라서 가중치 합 연산에서의 가중치를 이와 같이 표현할 수 있다.

이렇게 계산된 개별 입력 x_i 에 대한 가중치 곱 결과는 모두 합해져서 m 개의 가중치 합 h_j ($j=1, \dots, m$)이 되며, 구체적인 과정은 다음과 같다. 각 프로브 분자 P_{ij} 는 두 가지 특성을 가지는데, 하나는 X_i 에 특이적으로 결합하는 성질이고 다른 하나는 계산 결과 측정에 사용되는 형광 특성이다. 이 때, 경쟁 반응의 결과 생성되는 헤테로 분자들을 각 형광 특성에 따라 분류하여 공통의 형광 특성을 갖는 헤테로 분자들을 H_j 로 묶어서 생각할 수 있다. 결과적으로 가중치 합 h_j 는 다음과 같이 계산된다.

$$h_j = \sum_{i=1}^n x_i w_{ij}, \quad (j=1, \dots, m) \quad (3)$$

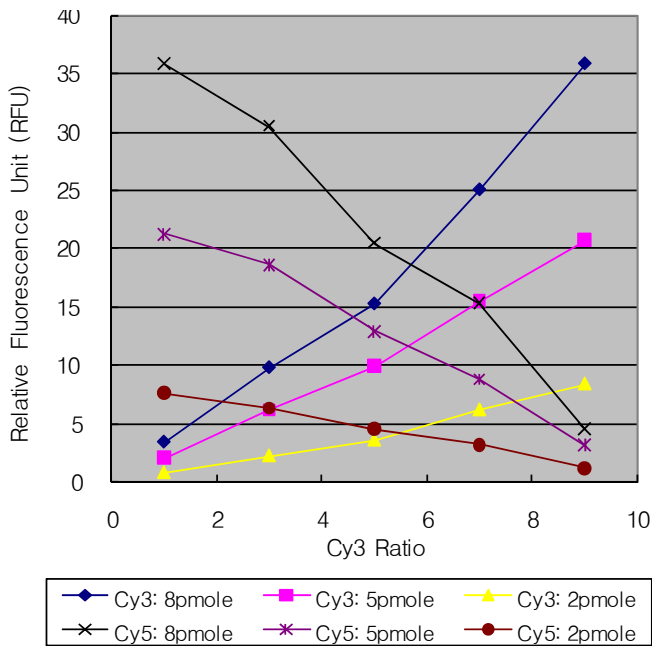


그림 4. 분자 비컨을 이용한 가중치 표현 검증 실험 결과. 각각 Cy3와 Cy5로 라벨링 된 분자 비컨들을 다양한 비율로 섞어서 3가지 레벨 (2, 5, 8 pmole) 의 입력 분자와 결합 한 결과 측정된 형광 신호는 입력 분자의 양에 비례하고, 해당 형광의 비율에 비례하는 결과를 보였다.

단, $x_i \geq 0$ 그리고 $w_{ij} \in [0,1]$. 여기에서 계산되는 가중치 합 연산은 입력이 항상 0 이상의 값이고 가중치가 0과 1의 사이의 값을 갖는다는 한계가 있으나 생물학 목적의 문제 해결에 있어서 목적 패턴은 항상 양적 정보로서 양의 값을 가지고 있으며, 음수의 가중치가 필요한 경우 양의 가중치와 음의 가중치를 분리하여 계산 함으로서 극복할 수 있다.

4. 실험 결과 및 고찰

여기에서는 앞서 기술한 경쟁 반응을 이용한 가중치합의 계산의 기본 요소라 할 수 있는 단일 입력 분자에 대한 프로브들의 경쟁 반응을 실제 바이오-랩 기술을 이용하여 시험 함으로써, 프로브의 혼합비율에 따라 생성되는 헤테로 분자들의 양이 비례해서 결정됨을 보인다. 구체적으로는 선형 DNA가 입력 분자 사용되며 프로브로는 Cy5와 Cy3로 각각 라벨링 되어있는 두 종류의 분자 비컨이 사용되었다. 이것은 단일 입력 x_i 에 대한 가중치 곱 연산, 즉 $x_i \times w_{ij}$ 연산으로서, 수식 (3)에서 $n = 1$

그리고 $m = 2$ 인 경우로 볼 수 있다. x_i 는 총 3가지 레벨(2, 5, 8 pmole)이 사용되었고 각각의 입력에 레벨에 대해서

Cy3와 Cy5의 5가지 혼합 비율 (Cy3: Cy5) = {(1:9), (3:7), (5:5), (7:3), (9:1)}을 이용하여 경쟁 반응을 수행하였다.

그림 4는 실험에서 측정된 계산 결과를 보여주는데, 그림에서 볼 수 있는 바와 같이 입력 레벨이 증가할수록, 그리고 해당 형광의 비율이 증가할수록 반응 결과 측정되는 형광 값이 비례해서 증가하였다. 이는 서로 다른 형광으로 라벨링 된 비컨의 혼합 비율이 목표로 하는 가중치를 잘 표현함을 뜻한다. 본 실험에서 수행된 단일 입력에 대한 경쟁 반응은, 여러 입력 분자를 사용하되 각 입력 분자마다 특이적으로 결합하는 비컨이 다른 입력에 결합하는 cross-talk이 발생하지 않도록 비컨의 서열을 디자인 함으로서 일반적인 가중치 합 연산으로 확장 가능하다. 또한 본 실험에서 사용된 DNA-비컨의 결합 반응은 가중치 합 연산을 구현하기 위해 사용 가능한 경쟁 반응의 한 예이고 앞서 언급한 바와 같이 RNA-비컨, 단백질-암타머 등의 결합 반응으로 확장 가능하며, 사용 되는 프로브들 간의 열역학적 특성에 따라 단순한 선형 합이 아닌 더 복잡한 수식의 구현에 사용될 수도 있다. 뿐만 아니라 비컨을 이용함으로써 기존의 선형 프로브를 이용한 방법에서 잉여 프로브의 제거 작업을 위해 필요로 했던 친화분리과정을 건너 뛸 수 있어 연산을 위한 실험적 복잡도를 낮추는 효과가 있다.

5. 결 론

본 논문에서 우리는 생체 분자를 이용하여 시험관 내에서 가중치 합 연산을 수행 가능한 구체적인 방법으로서 분자들 간의 경쟁 반응을 제시하였고, 입력 분자와 프로브 분자들간의 양적 관계를 통해서 입력 및 가중치 정보를 표현하는 일반화된 모델을 제시하였다. 또한 바이오-랩 실험을 통해서 DNA 입력과 분자 비컨을 사용하는 가중치 표현 방법을 검증하였고, 성공적인 가중치 표현을 확인했다. 본 논문에서 제안된 분자 컴퓨팅 방법은 유전자 진단과 같은 생물학 목적의 문제 해결에 응용 가능하며[13][14], 더 나아가 생체 분자를 이용한 기계 학습 방법[15]과의 결합을 통해 질병 진단을 위한 유전자 발현 패턴의 학습과 진단에 응용 가능하다.

감사의 글

본 연구는 교육인적자원부 BK21-IT 및 산업자원부 차세대 신기술 개발 사업의 분자 진화 컴퓨팅(MEC) 과제에 의하여 일부 지원되었다. 또한 이 연구를 위해 장비를 지원하고 공간을 제공한 서울대학교 컴퓨터연구소에도 감사 드린다.

참고문헌

- [1] G. Păun, G. Rozenberg and A. Salomaa, DNA Computing: New Computing Paradigms. Berlin Heidelberg: Springer-Verlag, 1998.
- [2] A. P. Mills Jr., "Gene expression profiling diagnosis through DNA molecular computation," *Trends Biotechnol.*, vol. 20, pp. 137-140, 2002.
- [3] Y. Benenson, B. Gil, U. Ben-Dor, R. Adar and E. Shapiro, "An autonomous molecular computer for logical control of gene expression," *Nature*, vol. 429, pp. 423-429, 2004.
- [4] D. J. Duggan, M. Bitter, Y. Chen, P. Meltzer and J. M. Trent, "Expression profiling using cDNA microarrays", *Nat. Genet.*, vol. 21, pp. 10-14, 1999.
- [5] R. J. Lipshutz, S. P. A. Fodor, T. R. Gingeras and D. J. Lockhart, "High density synthetic oligonucleotide arrays," *Nat. Genet.*, vol. 21, pp. 20-24, 1999.
- [6] J. S. Oliver, "Matrix multiplication with DNA," *J. Mol. Evol.*, vol. 45, pp. 161-167, 1997.
- [7] A. P. Mills Jr., B. Yurke and P. M. Platzman, "Article for analog vector algebra computation," *BioSystems*, vol. 52, pp. 175-180, 1999.
- [8] C. M. Bishop, *Pattern Recognition and Machine Learning*, New York: Springer Science+Business Media, 2006.
- [9] 임희웅, 유석인, 장병탁, "Thermodynamics-based weight encoding methods for improving reliability of biomolecular perceptrons", *정보과학회논문지: 소프트웨어 및 응용*, vol. 34(12), pp 1056-1064, 2007.
- [10] H.-W. Lim, J. Y. Lee, S. N. Kim, S.-I. Yoo, T. H. Park, and B.-T. Zhang, "DNA-Based Perceptron and Its Application to Binary Classification of Gene Expression," in *Proc. Tenth Int. Meeting on DNA Based Comput.*, pp. 439, 2004.
- [11] S. Tyagi and F. R. Kramer, "Molecular beacons: probes that fluoresce upon hybridization," *Nat. Biotech.*, vol. 14, pp. 303-308, 1996.
- [12] S. Klussmann, *The Aptamer Handbook: Functional Oligonucleotides and Their Applications*, WILEY-VCH Verlag GmbH & Co. KGaA, 2006.
- [13] S. Ramaswamy et al., "Multiclass cancer diagnosis using tumor gene expression signatures," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 98, pp. 15149-15154, 2001.
- [14] J. Khan et al., "Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks," *Nat. Med.*, vol. 7, pp.673-679, 2001.
- [15] B.-T. Zhang, "Hypernetworks: a molecular evolutionary architecture for cognitive learning and memory," *IEEE Comput. Intell. Mag.*, vol. 3, pp. 49-63, 2008.