

인공지능 기반 소재설계 연구동향

한상수, 김동훈

한국과학기술연구원 계산과학연구센터

1. 서론

최근 소재산업에서 계산과학(예, 제일원리, 분자동역학, 몬테카를로 시뮬레이션 등)을 이용한 신소재 개발에 많은 관심을 보이고 있다. 2005년도 IDC Market Research White Paper에 의하면 모델링과 시뮬레이션 기술은 연구효율을 증진시켜 연구비용과 시간을 감소시키고 소재 개발 성공확률을 높여주는 등 \$1 투자당 \$3~\$9의 이익을 얻는 효과를 갖는 것으로 평가되고 있다. 특히 2011년 미국 Material Genome Initiative (MGI) 프로젝트가 시작되어 세계적으로 계산재료과학에 대한 관심이 증폭되었다. 더불어 이 MGI 프로젝트는 소재설계 분야에 새로운 패러다임을 가져다 주는 계기가 되기도 하였다. MGI 프로젝트 이후 소재분야에도 데이터베이스(DB)의 중요성이 인식되었고, 몇몇 연구그룹을 중심으로 소재 DB를 체계적으로 구축하려는 노력이 시작되었다. 그 후 DB를 구축한 연구그룹을 중심으로 소재설계 분야에 인공지능을 접목시키려는 연구가 진행되기 시작하였다.

최근 소재설계 분야에 인공지능기술에 대한 관심은 가히 폭발적이라 할 수 있다. 그럼, 왜 소재 연구자들이 이토록 인공지능에 열광하는가? 가장 큰 이유는 빠른 예측 속도에 있다. 예를 들어, 기존 계산과학기술로 소재 물성 예측에 몇 개월이 걸리는 시간을(예, 제일원리 계산으로 화학반응 경로 예측 등), 인공지능을 이용하면 몇 분(minute)으로 단축시킬 수도 있다. 그럼, 기존 계산과학기술에 비해 인공지능이 빠른 예측 속도를 제공할 수 있는가? 기본적으로 인공지능의 예측 방식은 기존 계산과학 방식과 다르기 때문이다. 인공지능의 예측 방식은 소재 물성 DB에서 소재(조성 및 구조 등) 정보와 물성간의 “상관관계(correlation)”를 찾고, 이를 기반으로 새로운 소재의 물성을 예측하는 방식이다[1]. 이때 양자역학, 고전역학, 통계역학과 같은 principle은 전혀 사용되지 않는다. 이런 이유로 인공지능 기반 소재설계기술은 소재개발 분야의 새로운 패러다임을 제공할 수 있다. 그럼 1은 소재인공지능 기술을 소재개발 4차 패러다임으로 규명하고 있다. 즉 경험 및 시행착오적 시험에 의존한 1차 패러다임, 열역학 기반의 2

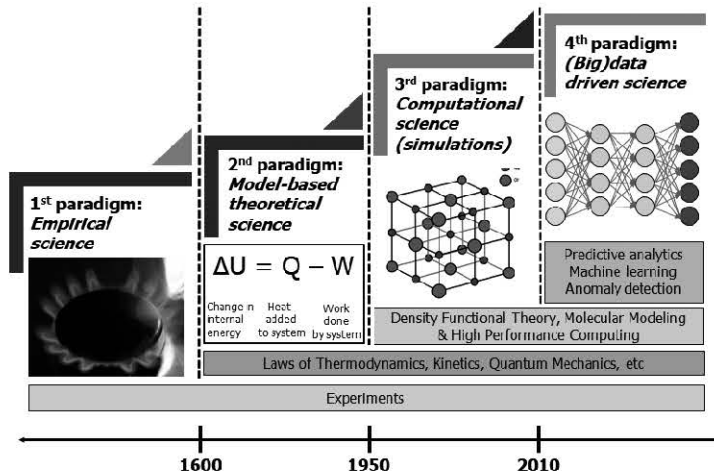


그림 1. 소재 개발 4차 패러다임 기술: 인공지능 기반 소재 개발 [2]

차 패러다임, 계산과학 기반의 3차 패러다임, 그리고 인공지능 기반의 4차 패러다임으로 분류되고 있다.

현재 소재설계 분야에 인공지능은 다양하게 활용되고 있다. 특히, 기존 계산과학기술의 단점을 극복하기 위한 새로운 이론적 방법론 개발 분야, 빠르게 소재 정보로부터 물성을 예측하는 분야(정방향 예측 기술), 그리고 원하는 물성을 입력하여 소재 정보를 출력하는 역방향 예측 기술 분야의 연구가 활발하게 진행되고 있다. 이에 본 기고에서는 이러한 소재설계 분야에 인공지능의 연구 동향에 관해 다루고자 한다.

2. 기존 계산과학 한계 극복형 인공지능기술 개발 분야

양자역학을 기반으로 하는 제일원리계산(ab initio calculation)은 현재 계산과학기술 중에서 가장 높은 정확성을 보이며 소재의 제약이 없다는 이유로 소재설계 분야에 가장 폭넓게 활용되고 있다. 하지만 제일원리계산은 계산시간이 많이 소요된다는 단점을 지니고 있다. 최근에 인공지능기술을 이용하여 이런 단점을 극

복하려는 시도가 많이 진행되고 있다.

양자역학적으로 소재의 물성은 그 소재의 전자구조에 의해 결정된다. 이러한 이유로 전자구조 예측은 나노소재, 전자재료뿐만 아니라 촉매와 같은 화학소재 분야에도 중요하게 여겨지고 있다. 인공지능기술을 이용하여 이런 전자구조를 예측하고자 하는 시도가 몇몇 연구그룹에서 진행되었다. 특히, 본 연구팀은 주성분분석법(Principal Component Analysis: PCA)을 이용하여 합금계 소재의 전자상태밀도(Electronic Density of States: DOS) 패턴(모양과 값)을 예측하는 기술을 최근에 발표하였다[3]. 소재 DOS 데이터는 복잡한 스펙트럼 형태를 지니며, 본 연구팀은 이 DOS 데이터를 이미지로 변환하였다. 즉, DOS 데이터를 임의의 $M \times N$ (예, 100×100) 그리드(grid)를 고려하면 복잡한 DOS 스펙트럼 데이터를 $M \times N$ 차원의 한 좌표(coordinate)로 표현이 가능하다(그림 2a). 여기서 그 좌표는 물리적으로 소재정보(원자구조 및 조성 등)와 전자구조 정보를 모두 포함하고 있는 것으로 생각할 수 있다. 만약 DB에 10개 소재의 DOS 전자구조 정보가 있다면 이들을 이미지 처리하면 $M \times N$ 공간상에 10개의 좌표가 분

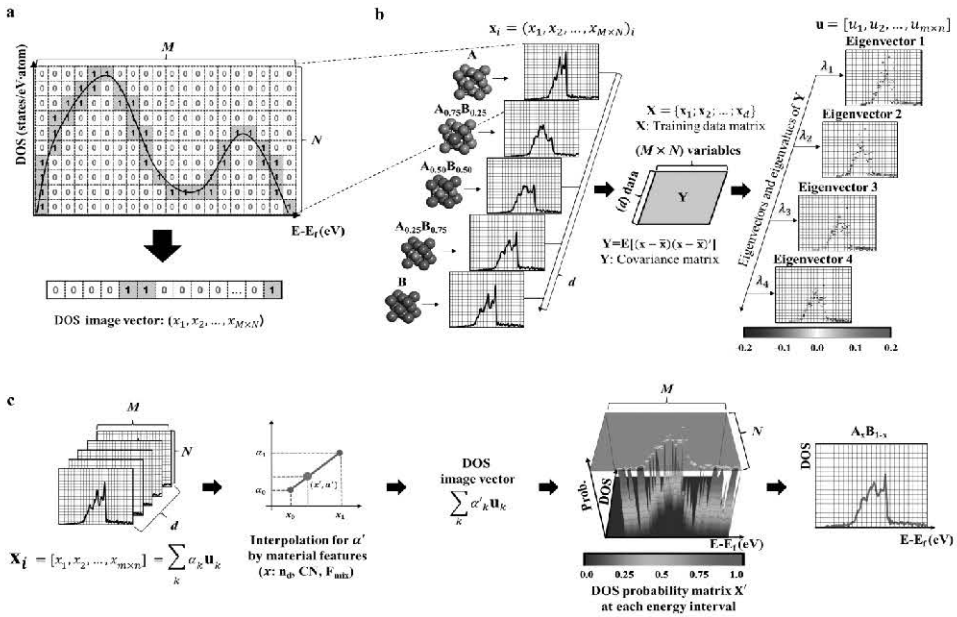


그림 2. KIST 연구팀에서 개발한 PCA 기반 소재의 DOS 전자구조 예측 기술 [3]

포되어 있는 것으로 여겨질 수 있다. 여기에 PCA를 이용하면 그 데이터 분포에서 가장 중요한 벡터(물리적으로 가장 linear correlation을 잘 표현하는 벡터), 2번째, 3번째, 총 $M \times N$ 개의 벡터를 찾을 수 있다(그림 2b). 이 벡터는 DB에 있는 10개의 DOS 전자구조 패턴을 잘 표현해주는 대표 패턴을 찾는 것으로 이해할 수 있다. PCA 후에 얻어진 그 벡터 정보로부터 우리는 소재의 DOS 패턴은 그 벡터들간의 선형조합(linear combination) 형태로 간단히 표현할 수 있게 된다. 그런 후, 임의소재의 DOS 패턴은 PCA 벡터들간의 선형조합에서의 계수(coefficient)를 결정하면 얻을 수 있고, 그 계수 결정하기 위해 소재의 원자구조와 조성과 관련된 피쳐(feature) 혹은 디스크립터(descriptor)를 기반으로 얻게 된다(그림 2c). 이 머신러닝 기술은 다양한 합금(3성분계까지 검증 완료) 및 표면의 DOS 패턴을 예측하는데 적용가능 하고 기존 제일원리기술 대비

정확성은 평균 >95%, 계산속도는 1,000배 빠른 성능을 보이고 있다. 최근에 본 연구팀은 이 기술을 확장하여 나노입자 구조의 DOS를 예측하는 기술을 개발하기도 하였다.

현재 소재의 전자구조를 예측하는 기술은 양자역학을 기반으로 한 다양한 방법(예, 제일원리, Tight binding 등)이 존재하며 그 방법들은 계산속도와 정확성사이의 trade-off 관계를 본질적으로 지니고 있다. 즉, 높은 정확성을 얻기 위해서는 많은 계산시간이 필요하고, 빠른 계산결과를 얻고자 하면 정확성을 희생할 수밖에 없다. 하지만 위에서 기술한 PCA 기반의 DOS 전자구조 예측기술은 높은 정확성과 빠른 계산속도를 동시에 얻을 수 있어 기존 principle 기반의 계산법에서 갖는 trade-off를 쫓 수 있는 옵션(option)으로 여겨질 수 있을 것으로 기대된다.

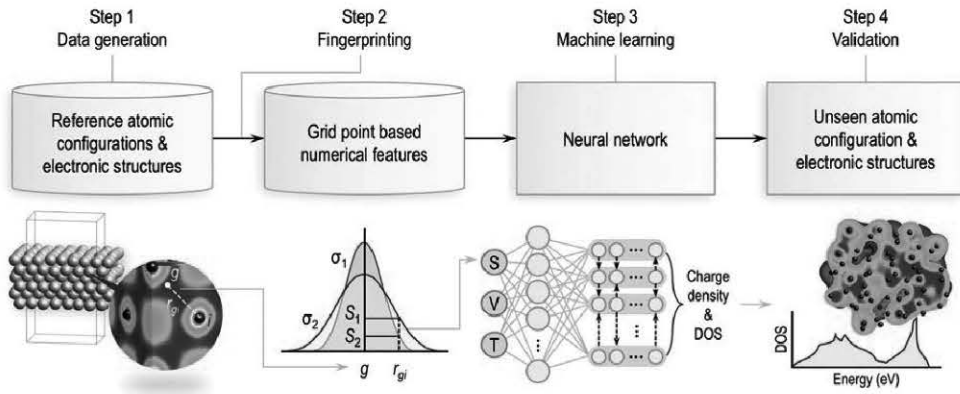


그림 3. 조지아텍 연구팀에서 개발한 뉴럴네트워크 기반 소재의 DOS 전자구조 예측 기술 [4]

위에서 기술한 PCA 기반 머신러닝 기술과 유사한 목적(빠른 전자구조 예측을 위한 인공지능기술 개발)으로 2019년 미국 조지아텍(Georgia Tech.)의 Ramprasad 교수팀은 뉴럴네트워크(Neural Network)를 이용하여 DOS를 예측할 수 있는 기술을 개발하기도 하였다 (그림 3)[4]. 일반 제일원리계산은 계산 시스템 크기(N)의 제곱에 계산속도가 증가하는 $[O(N^2)]$ 반면, 조지아텍에서 개발된 머신러닝 기술의 계산속도는 시스템 크기에 선형으로 증가한다 $[O(N)]$ 고 발표했다. 최근에 기존 제일원리계산을 느린 계산속도의 약점을 극복하기 위한 인공지능기술 개발 노력은 이 밖에도 여러 발표되기도 하였다[5-7].

제일원리계산뿐만 아니라 분자동역학 (Molecular Dynamics) 시뮬레이션도 계산재료과학분야에 폭넓게 활용되고 있다. 특히, 분자동역학 시뮬레이션은 시간의 따라 원자들의 움직임을 예측할 수 있고, 이러한 이유로 원자확산 및 상변태 등의 재료과학 문제를 이해하는데 주로 사용되고 있다. 분자동역학 시뮬레이션은 실제 나노 크기 시스템을 다룰 수 있다는 장점을 지니고 있지만, 정확성 측면에서 제일원리 계산을 따라가기에는 한계가 있다. 물론, 제일원리 분자동역학 기

술은 높은 정확성을 제공해주지만 일반 실험실 수준에서 원자 100개를 시뮬레이션을 구동하는데도 버거운 실정이다. 일반 분자동역학 시뮬레이션에서는 원자간 척력(repulsion)과 인력(attraction)의 상호작용을 나타내는 포텐셜 (interatomic potential) 혹은 포스필드 (force field)가 필요하며, 이것이 결국 시뮬레이션의 정확성을 크게 좌우하는 데 보통의 경우 제일원리 보다 다소 낮은 정확성을 보인다. 현재 원자간 포텐셜의 또 다른 약점 중에 하나라 낮은 transferability에 있다. 즉, 금속, 세라믹, 유기소재 등 다루고자 하는 시스템의 종류에 따라 사용되는 원자간 포텐셜은 달라지게 된다. 예로서, 금속의 경우 EAM (Embedded Atom Method), 세라믹의 경우 셸 모델 (Shell Model), 그리고 유기소재의 경우 유니버셜 포스필드 (Universal Force Field) 등이 많이 사용되고 있다. 하지만, 만약 금속/유기물 계면을 연구하고 싶다면 포텐셜 설정에 많은 어려움이 따르게 된다.

최근에는 인공지능 기술을 이용하여 이러한 어려움을 극복하려는 연구가 활발히 진행되고 있다. 특히 2007년 Behler와 Parrinello가 뉴럴네트워크를 기반으로 한 포텐셜을 발표하였다[8]. 이 방법은 원자구조

에 대한 정보를 대칭 함수로 맵핑한 후 제일원리로 계산된 에너지를 뉴럴네트워크로 학습시켜 포텐셜 라이브러리를 구축하는 것으로 양자역학 수준의 정확도를 유지하는 동시에 일반적인 제일원리계산 대비 약 100배 이상 빠른 계산속도를 확보할 수 있다[9]. 그 이후 유사한 방법론을 바탕으로 다양한 머신러닝 포텐셜이 발표되었는데, 대표적으로 브라운 대학교 Peterson 교수 그룹의 AMP [10], MIT Kolpak 교수 그룹의 PROPhet [11], 버클리 대학 Artrith 박사가 주도적으로

개발 중인 aenet [12], 서울대학교 한승우 교수 그룹의 SIMPLE-NN [13] 등이 있다.

개발된 머신러닝 포텐셜은 분자동역학 시뮬레이션 기술과 접목하여 다양한 분야로 응용되고 있다. 최근에는 실리콘(Si) 구조 내부에 vacancy 결함이 있을 때 구조 변화 [15], 리튬이온전지의 실리콘 음극재와 리튬과 반응했을 때 형성되는 다양한 구조 [16], 이종 합금 이산화탄소 환원 촉매 [17], 수소 환경에서의 준안정 상태의 백금 나노입자구조 예측 [18] 등 다양한 분야에 적용된 사례가 있다. 머신러닝 포텐셜 방법론의 자세한 설명과 추가적인 연구동향은 참고문헌 9에서 찾을 수 있다.

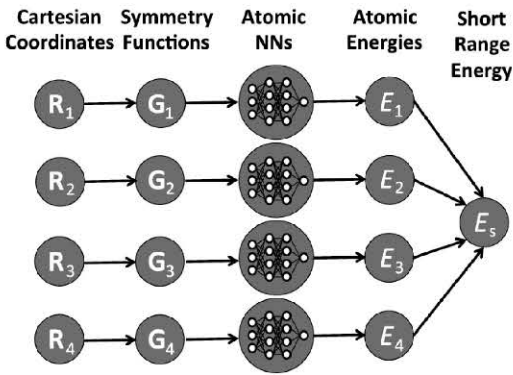


그림 4. 뉴럴네트워크 포텐셜 방법의 개념도 [14]

3. 소재 물성을 빠르게 예측하기 위한 인공지능 기술

2018년 미국 MIT 재료과 Grossman 교수 연구팀은 소재의 벌크 (bulk) 상태 물성 예측에 특화된 Crystal Graph Convolutional Neural Network (CGCNN) 인공지능 기술을 개발하였다[19]. 소재의 결정구조와 물성간의 상관관계를 훌륭하게 학습하기 위해서는 소재

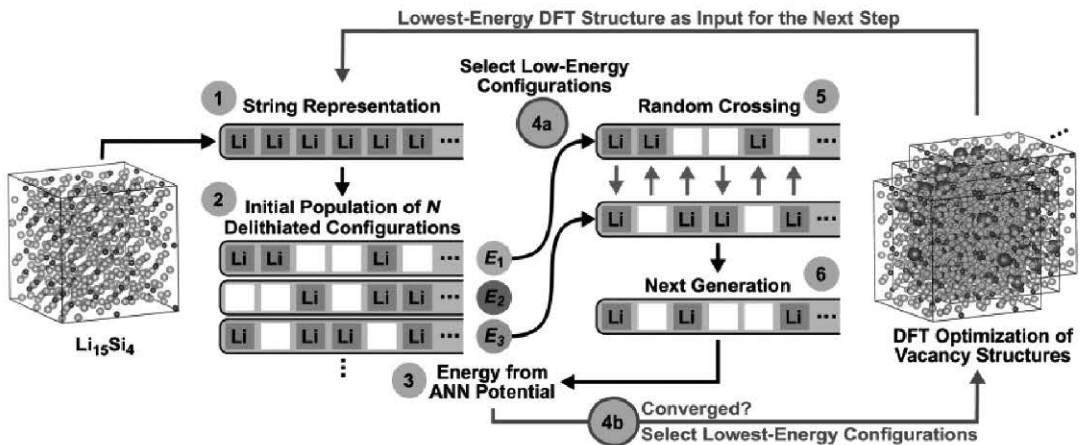


그림 5. 머신러닝 포텐셜을 기반으로 리튬이차전지 적용 사례 [16]

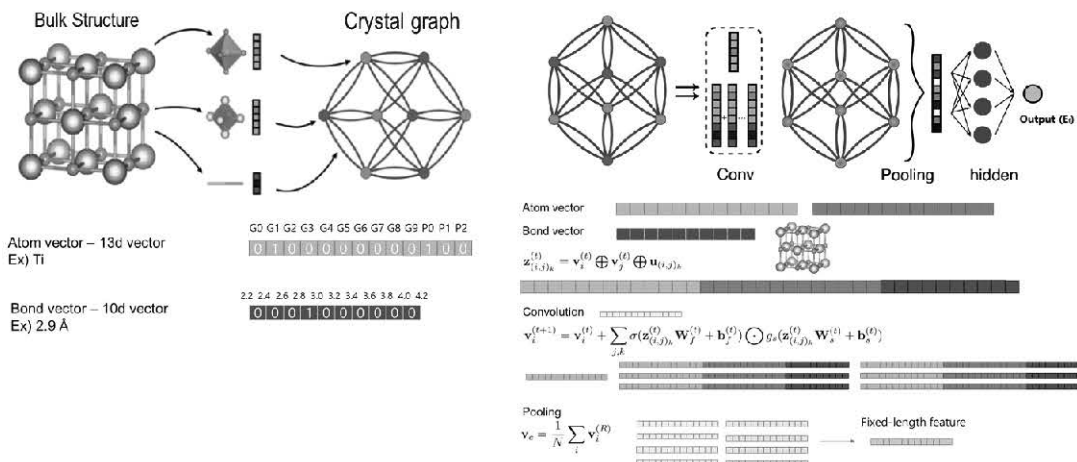


그림 6. MIT Grossman 교수 연구팀이 개발한 CGCNN 기술 [19]

의 구조정보를 암호화시키는 것이 첫 번째 단계이자 성공의 관건이다. Grossman 교수 연구팀은 Crystal Graph의 개념을 도입하여 결정성 소재의 구조정보를 매우 유연하게 암호화시키는데 성공하였다(그림 6). 그리고, Materials Project DB (미국 로렌스버클리국립연구소 주도하 개발된 소재 물성 DB) [20] 내에 있는 소재정보 전체(약 70,000여개)를 학습시켜서 소재의 가장 기초적인 고유물성인 열역학적, 전기적, 기계적 성질들을 주로 학습대상으로 정하였다. 예를 들면, 벌크 생성에너지(bulk formation energy), 절대 에너지(absolute energy), 밴드갭(band gap), 페르미 에너지(Fermi energy), 체적 탄성율(bulk moduli), 전단 탄성율(shear moduli), 포아송 비율(Poisson ratio) 등이 포함된다. 소재의 물성 예측 정확도는 전반적으로 매우 우수하게 나타났으며, 연구진은 제일원리계산이 가지는 실험대비 어려움보다 인공지능 어려가 더 낮음을 보여주어, 인공지능이 소재물성 예측에 주요 기술로 자리매김할 수 있음을 보였다. 여기서, CGCNN을 이용하면 주기율표상에서 손쉽게 얻을 수 있는 정보(예, group #, period #, electronegativity, covalent

radius 등)만을 입력하여 제일원리계산 수준의 소재 물성을 예측할 수 있다는 점에 주목할만하다.

CGCNN 기술은 소재 물성을 예측하는데 매우 효율적으로 사용될 수는 있지만, 벌크 구조에서만 적용할 수 있다는 한계를 지니고 있다. 하지만 촉매소재 설계를 위해서는 표면구조를 다룰 수 있는 기술이 필요로 한다. 본 연구팀은 이를 위해 CGCNN을 표면구조에도 적용할 수 있도록 이를 개량하여 Slab Graph Convolutional Neural Network (SGCNN)이란 인공지능 기술을 개발하였다(그림 7). SGCNN의 핵심 아이디어는 표면 구조는 슬라브(Slab) 형태로 간주하고, 다시 슬라브 구조는 벌크 영역과 표면 영역으로 나누어서 생각할 수 있다. 여기서, 벌크 영역의 원자구조 및 원소 정보 등은 Grossman 교수팀이 개발한 Crystal graph를 가지고 표현하고, 표면 영역의 원자구조 및 원소 등은 CGCNN과 유사한 방법으로 별개의 Surface graph를 가지고 표현하는 방식이다. surface graph 표현시 촉매 표면에 흡착된 화학반응 관여 분자/원자도 함께 고려한다. 그런 후 CGCNN으로 암호화된 벡터와 SGCNN으로 암호화된 벡터를 결합하면 분자/원자가

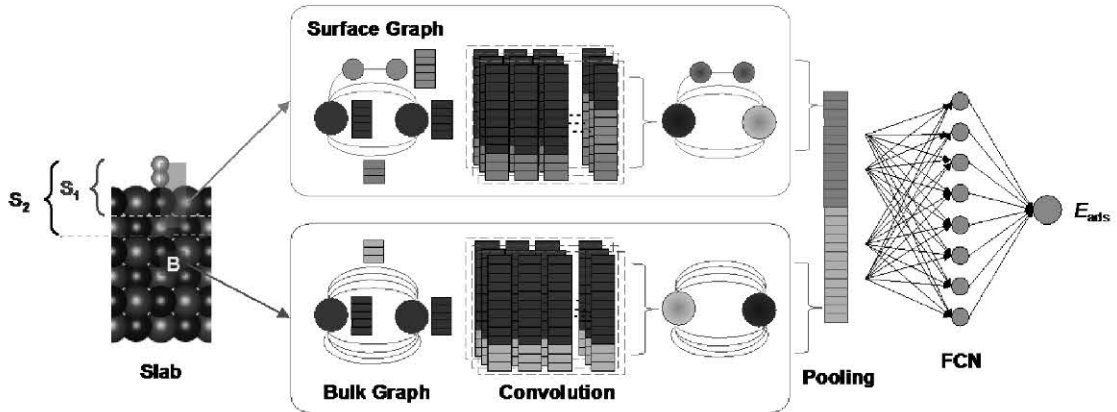


그림 7. 촉매 설계용 SGCNN 인공지능 기술 [21]

흡착된 표면 구조를 표현할 수 있게 된다. 그리고 이렇게 표면을 암호화한 벡터와 흡착에너지 간의 뉴럴네트워크로 엮어주면 촉매 표면에 흡착된 분자/원자의 흡착에너지를 예측할 수 있게 된다. 실제로 본 연구진은 이 SGCNN을 암모니아(NH₃) 합성용 합금촉매 설계분야에 적용하여 6종의 신규 촉매를 제안하였고[21], 현재는 연료전지용 합금촉매 설계 분야에 적용 중에 있다.

한편 인공지능 기술은 구조재료 개발 분야에도 활발히 이용되고 있다. 미국 스탠포드 대학교 Mehta 교수팀과 노스웨스턴 대학교 Wolverton 교수팀은 새로운 비정질 금속(metallic glass)을 발굴해내는 작업에 인공지능 기술을 도입하였다[22]. 비정질 금속은 결정성이 강한 금속 대비 높은 기계적 강도와 취성을 보이기 때문에 활용가치가 크다. 실제 3원계나 4원계의 금속들에서 비정질화되는 금속이 발견이 되는 경우가 많은

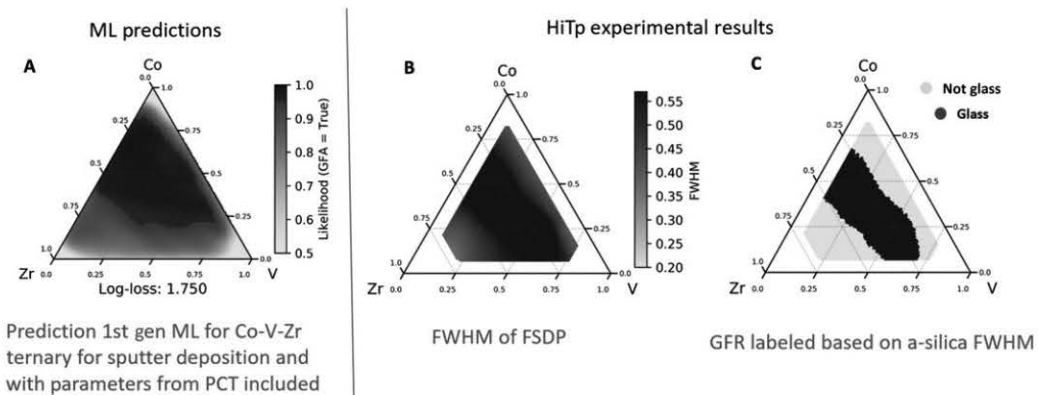


그림 8. 인공지능 기술을 이용하여 고성능의 3원계 비정질 금속 개발 사례 [22]

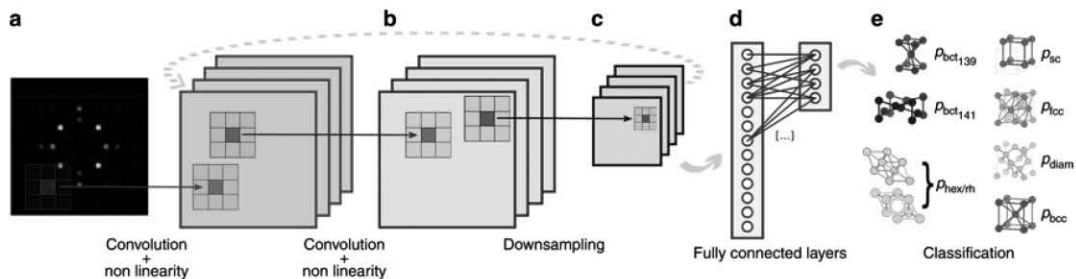


그림 9. 결정구조 구분화 연구에 사용된 인공지능 기술 도식도 [24]. (a) 미지 물질의 회절패턴. (b-d) 인공지능 기술에서 convolution, pooling, fully-connected layer 과정, (e) 결정구조 구분화 작업.

데, 탐색해야하는 재료영역(원소조합 및 조성)이 너무 방대해서 연구의 진행이 매우 더딘 실정이다. 연구진은 3원계 금속들이 매우 다양한 조성에서 물질이 비정질 형성능력(glass forming ability)를 가질 수 있는지를 예측하는 머신러닝 모델을 만들었다. 그림 8은 Co-Zr-Ti, Co-Zr-Fe, Co-Zr-V 계열 3원계 금속재료들이 어떤 조성에서 비정질 금속으로 존재할 수 있는 가능성이 큰지에 대한 인공지능을 이용하여 예측한 결과를 보여준다. 예측된 재료들은 실험검증을 통해 실제 비정질 형성체(glass former)가 실제로 될 수가 있음을 확인하였으며, 연구진은 해당 연구 분야 소재 발굴의 속도를 약 200배 개선시키는 효과가 있다고 밝혔다.

또한, 미국 텍사스 휴스턴 대학 Brgoch 교수팀은 초고강도의 기계적 성질을 가지는 신소재 탐색을 위해 인공지능을 활용하기도 하였다[23]. 연구팀은 2,572개의 재료를 샘플링(sampling)하여, 체적 탄성률과 전단 탄성률을 정확하게 예측할 수 있는 인공지능 기술을 개발하였고, 이를 이용하여 총 118,287개의 화학 결합물에 대해 인공지능 기술을 적용하여 체적 탄성률 300 GPa 이상, 전단 탄성률 150 GPa 이상인 재료를 선별하여 실제 실험으로 테스트하였고, 그 결과 $\text{Mo}_0.9\text{W}_1.1\text{BC}$, $(\text{Re}_0.5\text{W}_0.5)\text{C}$ 두 종류의 카바이드(carbide) 재료를 발굴해내었다. 실제 실험에서 약 12만개에 달하는 실험을 시행착오를 겪으면서 진행하는 것은 사실상 불가능

한데, 인공지능 기술을 통해서 실험에서 테스트 해야하는 재료를 높은 확률로 정해줄 수 있다. 이는 인공지능 기술이 소재개발 가속화를 실현한 좋은 예시 가운데 하나이다.

인공지능 기술은 이미지(Image) 분석처리 (얼굴 분석 등) 분야에 탁월한 성능을 보인다. 이런 이유로 이미지 형태의 소재데이터(예, 회절패턴 및 X-ray 이미지 등)에서 소재정보를 추출해내는 분야에도 인공지능이 활용되고 있다. 독일 프리츠하버(Fritz-Haber) 연구소의 Ziletti와 Scheffler 교수팀은 인공지능 기술을 이용하여 소재의 회절패턴(diffraction pattern)과 소재의 결정구조의 상관관계를 학습하였다[24]. 이러한 상관관계 학습이 성공적으로 이루어진다면, 미지의 재료에 대하여 회절패턴을 얻을 후 인공지능 기술을 통해 소재의 결정구조 정보를 알아낼 수 있어 학계 및 산업계에 미치는 영향이 매우 클 것으로 기대된다. 연구진의 핵심아이디어는 회절패턴을 이미지 그 자체로 바라본 것이다. 이미 설명한 바와 같이, 인공지능 알고리즘은 사실 소재연구 이전에 이미지 분석처리에 매우 특화되어 발전되어 왔으므로 현재 개발된 딥러닝 알고리즘(Convolutional Neural Network 기반)들은 이미지 분석 및 구분에 특화되어 있다. 연구진은 이러한 사실을 인지하고 회절패턴을 이미지화 시켜서 다양한 소재에 대한 결정구조를 구분해내는데 성공하였다. 개발된

알고리즘은 결함(defect)이 많이 존재하는 소재의 경우에도 적용이 가능함을 보여, 실제 학계나 산업계에서의 활용 가치는 매우 클 것으로 판단된다.

4. 역방향 소재설계 인공지능 기술

향후 소재설계 분야에서 인공지능 활용 연구는 어떤 흐름으로 진행될까? 불과 2~3년전까지만 하더라도, 인공지능 기술은 소재의 구조나 조성 정보로부터 소재의 물성을 예측하는 목적으로 주로 사용되어 왔다. 이는 기존 계산과학기술을 사용할 때보다 예측속도가 빨라 소재설계의 효율성 측면에서 상당한 장점을 지니고 있지만, 여전히 시행착오(trial-and-error)법에 기반한 에디슨이안(Edisonian) 접근 방식에 머물러 있다(그림 10). 즉, 현재의 인공지능 기술로는 본인이 원하는 소재 물성이 나올 때까지 소재 정보(조성 및 구조 등)를 바꿔보면서 예측해보는 방식이다. 하지만 내가 원하는 물성을 입력 정보로 하고 그 물성을 갖는 소재 정보를 출력해주는 소재설계기술이 개발이 되면 기존

에디슨이안 접근방식에서 탈피할 수 있고, 더 나아가 소재개발의 새로운 패러다임을 제공해줄 수도 있다. 기존 소재정보로부터 물성을 예측하는 방식을 “정방향” 설계방식(forward design)이라 한다면, 그 반대인 물성에서 소재정보를 예측하는 방식을 “역방향(inverse design)”이라 부를 수 있다. 향후에는 이런 역방향 소재설계에 인공지능이 할 일이 많아질 것으로 판단된다.

안타깝게도 역방향 소재설계는 정방향 설계보다 근본적으로 훨씬 어려운 문제이다. 기본적으로 입력 정보인 소재물성 변수보다 출력 정보인 소재정보(조성, 구조, 모양 등)의 변수가 많으며, 뿐만 아니라 같은 물성을 가지는 서로 다른 소재들이 여럿 존재할 수 있다. 즉, 인공지능 모델을 활용해서 함축된 정보로부터 많은 정보를 뽑아야 하는 어려움이 있을 수 밖에 없다.

역방향 소재설계의 성패는 다음 두 가지 요소를 어떻게 개발 및 사용하는가에 달려있다. 첫째는 머신러닝 모델에 인풋으로 활용되는 “재료 표현자 (representer)”이고, 둘째는 새로운 재료를 계속해서 만들어 낼 수 있는 “생성 모델 (generative model)”이다. 먼저

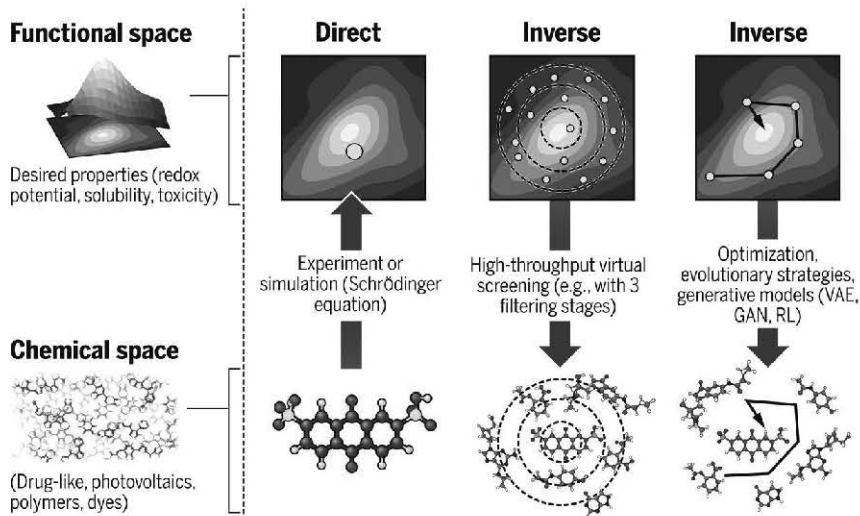


그림 10. 정방향 및 역방향 소재개발 방식의 개념도 [25]

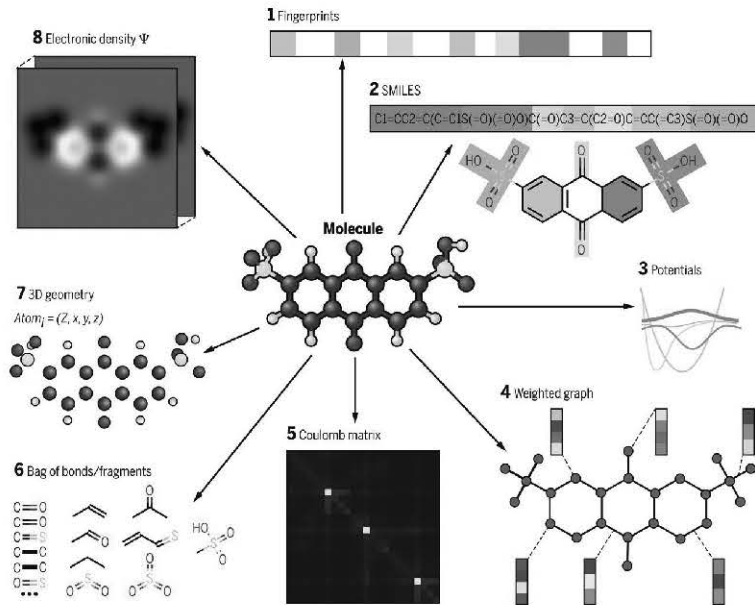


그림 11. 머신러닝 모델에 이용되는 다양한 재료 표현자. (시계방향) 1. 분자환경의 유/무를 이용한 fingerprint vector, 2. SMILES string, 3. Potential energy function, 4. 그래프 표현법, 5. Coulomb 매트릭스, 6. Bag of bonds/fragments, 7. 3차원 좌표 및 전하상태, 8. 전자구름 [25]

재료 표현자란, 재료를 머신러닝 모델 내에서 어떤 포맷으로 표현할 것인가를 의미한다. 그림 11에 나와있듯이 이미 다양한 표현자들이 시도되고 있다. 현재 가장 많이 사용되고 있는 표현자는 SMILES string (1차원 텍스트 인코딩) [26] 과 그래프 표현법 (원자와 본드 정보의 수학적 그래프화) [27] 이다. 그러나 어떠한 표현자가 가장 훌륭한가는 타겟 물성의 종류에 따라 충분히 달라질 수 있다. 역방향 설계에서 가장 바람직한 표현자는 invertibility (역으로 도치가 되는 성질)를 가지고 있어야 한다. 즉, 재료의 구조나 조성을 표현자로 전환하는 것은 물론이고, 표현자를 역으로 재료의 구조 및 조성으로 역전환하는 것이 가능함을 말한다. 재료를 표현자로 전환하는 과정에서 압축을 통한 데이터 손실이 있기 때문에, 역전환을 1:1로 해내는 과정은 매우 도

전적인 문제이다. 때문에 더 나은 표현자를 개발하고자 하는 노력이 다방면으로 이루어지고 있다.

다음으로 중요한 요소는, 새로운 재료를 계속해서 출력해줄 수 있는 “생성 모델 (generative model)”이다. 생성 모델이란 기존의 방대한 양의 재료들을 학습함으로써, 학습 데이터베이스에는 없는 새로운 (그렇지만 매우 닮았으며 그럴듯한) 재료들을 만들어주는 모델을 말한다. 그림 12는 실제 사용되고 있는 몇 가지 생성 모델들이 소개되어 있다. 현재 가장 효과적이고 널리 사용되고 있는 두가지 모델은 Variational Autoencoder (VAE) [28]와 Generative Adversarial Networks (GAN) [29]이다. VAE 모델은 Autoencoder (AE)의 변형된 형태이다. AE 모델은 인코더(encoder)와 디코더(decoder)로 구성되어 있으며, 인코더는 재료 표현

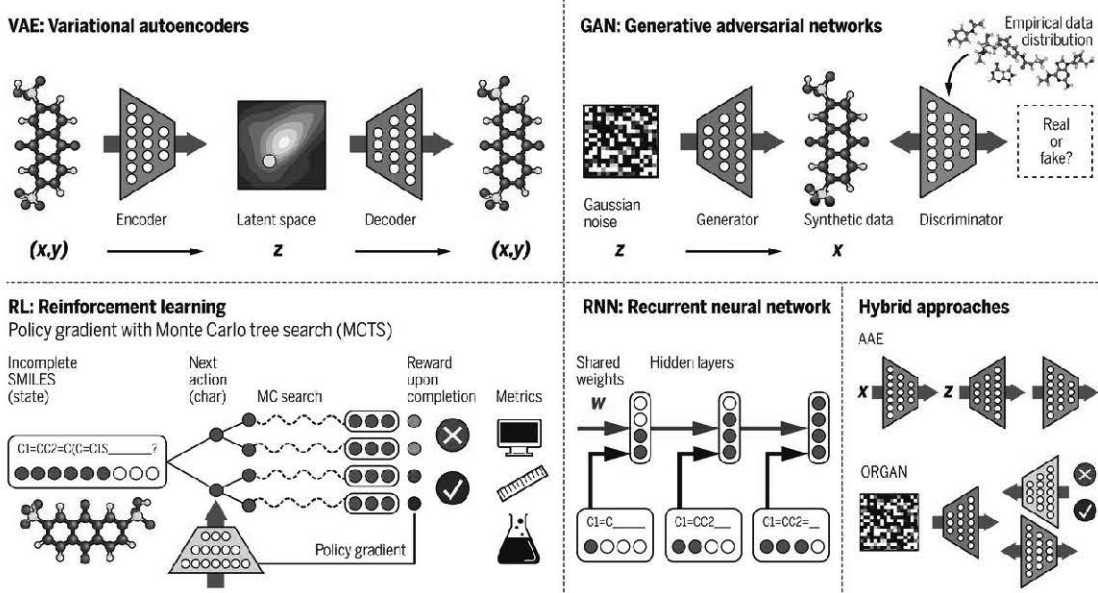


그림 12. 역방향 소재 설계에 사용되는 생성 모델 종류 [25]

자를 저차원 공간(latent space)로, 그리고 디코더는 이를 다시 재료 표현자로 복귀시키는 형태이다. VAE 모델은 가운데 latent space에 확률분포 함수를 가미 시킴으로써, 데이터베이스에 없는 새로운 재료를 샘플링 및 출력시킬 수 있는 원리이다. 다음으로, GAN 모델은 크게 생성기(generator)와 구별기(discriminator)로 구성되어 있다. 생성기는 새로운 재료를 만들어 내고, 구별기는 만들어진 재료에 대해 트레이닝 데이터베이스에 있는 재료와의 유사성을 평가한다. 이 두가지 기작의 경쟁을 통해, 트레이닝 데이터들과 매우 유사한 새로운 재료들을 출력할 수 있는 원리이다. 이외에도 Recurrent Neural Network (RNN), 강화학습(Reinforcement Learning) 등이 보조적으로 사용되어 생성모델의 성능을 증가시키는데 사용되고 있다.

2018년 이후, 여러 연구팀에서 역방향 소재 설계에 성공한 연구들을 문헌에 보고하고 있다. 다양한 사례들을 아래 표1에 정리하였다. 이 중에 세 가지의 연구를

대표로 소개하고자 한다. 역방향 설계 연구들이 다양한 응용분야(바이오, 신약개발, OLED, 다공소재, 촉매 등)에 적용되고 있으므로, 소재 종류별로 선별하여 신약 유기물 1건, 다공소재 1건, 메타물질 1건의 연구들을 아래에 구체적으로 소개한다.

첫 번째 대표 사례는 캐나다 토론토 대학의 Aspuru-Guzik 교수팀의 연구이다 (그림 13) [30]. 그 연구는 우리가 원하는 물성을 가지는 새로운 분자(신약개발용)를 찾아주는 것을 목표로 한다. 이 연구는 VAE 알고리즘을 재료/바이오 분야 역방향 설계에 처음으로 도입하였기 때문에 학계와 산업계에서 매우 파급력이 크다. VAE 알고리즘을 분자의 SMILES string (표현자)에 적용하면 인코더와 디코더 사이에 latent space라 불리는 저차원 공간이 생성된다. 이 연구에서 가장 흥미로운 부분은 단연 latent space이다. SMILES string은 텍스트 인코딩법이기 때문에 이산적(discrete)인 성격을 가지지만, VAE 네트워크를 통

표 1. 소재 및 바이오 분야, 인공지능 기반 역방향 소재개발 연구 사례

연구팀	타겟 물성	타겟 물질군	Generative model	사용 DB	참고 문헌
U. Toronto A. Aspuru-Guzik 교수팀	LogP (partition coefficients), SAS (synthetic accessibility score) 포함 5개	Drug molecules	VAE	ZINC, QM9	30
KAIST 김우연 교수팀	MP (molecular weight), LogP 포함 5개	Drug molecules	Conditional VAE	ZINC	31
삼성중합기술원	HOMO, LUMO (광흡수 파장)	Organic molecule (OLED 용도)	RNN	PubChem	32
KAIST 김지한 교수팀	흡착에너지 (메탄분자)	Zeolite (실리카)	GAN	IZA, PCOD	33
KAIST 정유성 교수팀	생성 에너지	Inorganic solid (바나듐 산화물계)	VAE	Materials Project	34
U. South Carolina J. Hu 교수팀	생성 에너지, 밴드갭	Inorganic solid	GAN	Materials Project, OQMD, ICSD	35
파나소닉 Y. Sawada 박사팀	생성 에너지	Inorganic solid	Conditional VAE/GAN	Materials Project	36
Stony Brook 대학교 Frenkel 교수팀	X-ray 흡수 스펙트럼	금속 나노입자	ANN	Ab initio simulation (자체수집)	37
Wisconsin 대학교 Z. Yu 교수팀	광학 스펙트럼 (투과율 등)	메타물질	ANN	EM simulation (자체수집)	38
Georgia 공과대학교 W. Cai 교수팀	광학 스펙트럼 (투과율 등)	메타물질	GAN	FEM simulation (자체수집)	39

해 latent space라는 연속적이고 (continuous) 미분 가능한 (differentiable) 공간으로 이동시킬 수 있다. 따라서 이 연속적인 공간 내에서 거리를 점차 늘려가면 새로운 분자들을 샘플링하는 것이 가능하다. 본 연구는 바이오 분야에서 잘 알려진 QM9과 ZINC 데이터베이스에서 각각 108,000개, 250,000개의 데이터를 활용하여 학습을 진행하였으며, 신약개발에서 중요한 물성인 partition function (logP), synthetic accessibility score (SAS), quantitative estimation of drug-likeness (QED) 등을 관심 물성으로 선택하였다.

두 번째 사례는 KAIST 김지한 교수팀의 연구이다

(그림14) [33]. 그 연구에서는 VAE 알고리즘이 아닌 GAN 알고리즘을 활용하여, 원하는 물성을 가지는 새로운 다공소재를 개발하는데 목표가 있다. 연구팀이 개발한 ZeoGAN (zeolite GAN)은 그림 14(a)에 도식화되어있다. ZeoGAN은 순수한 실리카 제올라이트 (Si, O 두가지 원소종류만 활용)에 한정된 모델이며, 연구팀은 이 모델을 메탄분자의 흡착문제를 예시로 적용하였다. 즉, ZeoGAN 모델의 인풋 (또는 표현자)은 Si, O 각각의 원자 좌표 그리고 메탄 포텐셜 에너지를 함께 사용하였다. 총 31,713개의 알려진 실리카 제올라이트를 학습시킨 결과, 총 14개의 새로운 (알려지지 않은)

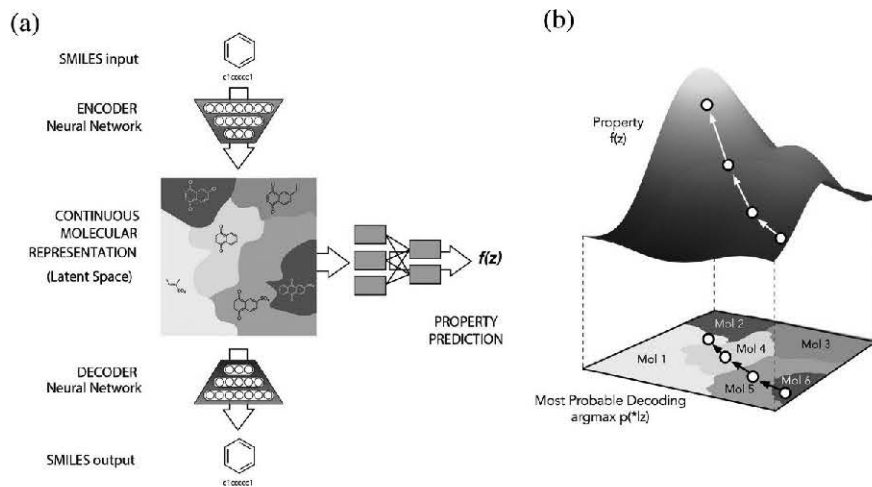


그림 13. Aspuru-Guzik 교수팀 연구의 도식도 [30]. (a) 바이오 분자에 VAE 모델을 적용한 도식도. Latent space에서 소재 물성 예측도 함께 학습과정에 포함됨. (b) Chemical latent space와 property space간의 1:1 맵핑. 연속적 공간에서 원하는 물성을 가지는 분자를 찾아가는 과정을 도식화 함.

제올라이트를 생성해내는데 성공했다. 특히 사용자가 원하는 물성 (예: 18–22 kJ/mol 범주의 메탄 흡착 에너지)을 가지는 재료를 만들어내는 데모를 제시함으로써 모델의 효과성을 입증했다 (그림 14(b)). 본 연구는 바이오 분자가 아닌 반복성 (periodicity)과 결정성 (crystallinity)이 있는 재료를 생성해내는데 성공했다는 점에서 큰 가치가 있다. 다만, Si, O 두 가지 원소 종류뿐만 아니라, 다원계의 복잡한 시스템으로의 확장은 도전적인 과제로 남아있다.

마지막 사례는 미국 조지아공과대학교 Wenshan Cai 교수팀의 연구이다 (그림 15) [39]. 메타물질 연구의 기존 접근방식은 물질이 가질 수 있는 다양한 패턴들에 대해 반복적으로 맥스웰 방정식 (Maxwell's equation)을 풀어냄으로써 원하는 광학 스펙트럼을 찾아내는 것이다. 그러나, 그 연구에서는 우리가 원하는 물성 (광학 스펙트럼)을 가지는 메타물질의 기하학적 패턴을 역생성해내는데 성공하였다. 그 연구에서도

GAN을 활용하였는데, GAN이 사람 얼굴과 같은 이미지 데이터를 생성해내는데 탁월한 만큼 메타표면의 기하학적 패턴을 만들어내는데도 훌륭한 성능을 보인다. 그 연구는 앞서 설명한 두 가지의 사례와는 다르게 관심 물성이 하나의 스칼라 값이 아닌 훨씬 더 복잡한 형태의 스펙트럼 데이터라는 점에서 가치가 크다.

아직은 인공지능 기반 역방향 소재설계는 매우 초기 단계라 할 수 있다. 그러나 그 발전속도는 매우 빠르다. 위의 사례들에서도 보여지듯, 매우 빠른 속도로 바이오, 신약, 다공소재, 광학 등 분야에 침투하고 있음을 확인할 수 있다. 특히, 빅데이터가 오랜 시간 탄탄하게 구축되어 있는 바이오나 신약 연구분야에 머물지 않고, 더 어려운 문제들로 점차 확장됨에 주목할 필요가 있다. 특히, 데이터의 양이 비교적 많이 부족했던 분야나, 반복성/결정성이 있는 무기물 소재 (다공소재 등)나, 관심 있는 물성이 스칼라 값이 아닌 스펙트럼 (광학 등)인 경우 등 난제로 여겨져 왔던 문제들도 점차 해결할

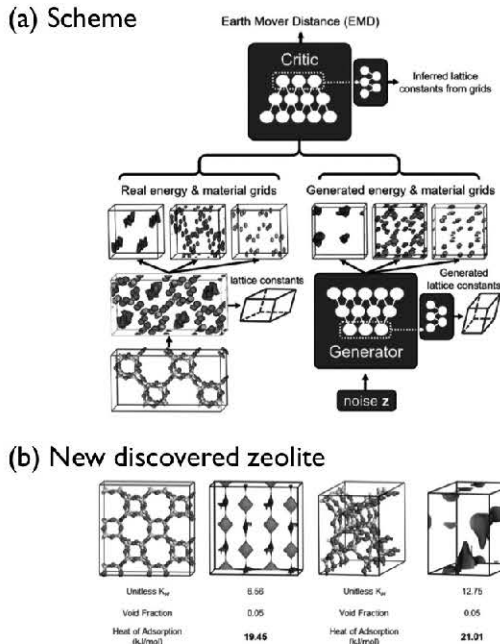


그림 14. KAIST 김지한 교수팀의 다공소재 분야의 역방향 설계 인공지능 기술 [33]. (a) ZeoGAN 모델의 도식도. Si, O 원자의 좌표 및 메탄 포텐셜 에너지를 인풋으로 활용함. Generator와 Critic의 경쟁을 통해, 학습된 제올라이트와 유사한 신재료를 생성함. (b) 원하는 물질(메탄 흡착 에너지)을 가지는 새로운 다공소재의 예시.

수 있는 가능성을 보고하는 논문들이 이어지고 있다. 그러나 아직은 특정 재료군 (예: 특정 원소조합에 한정)에 한정되어 연구되는 경우가 많으며 새롭게 만들어지는 소재들의 물리학적/화학적으로 무의미한 경우들이 많아 성공률이 낮다. 개선의 여지가 크게 남아있다.

5. 맺음말

본 기고에서는 인공지능 기술이 소재설계 분야에 어떻게 활용되고 있는지에 대해 알아보았다. 비록 인공지능 기술이 소재분야에 본격적으로 활용된 지는 불과 몇 년이 되지 않았지만 급속도로 빠르게 발전하고 있다.

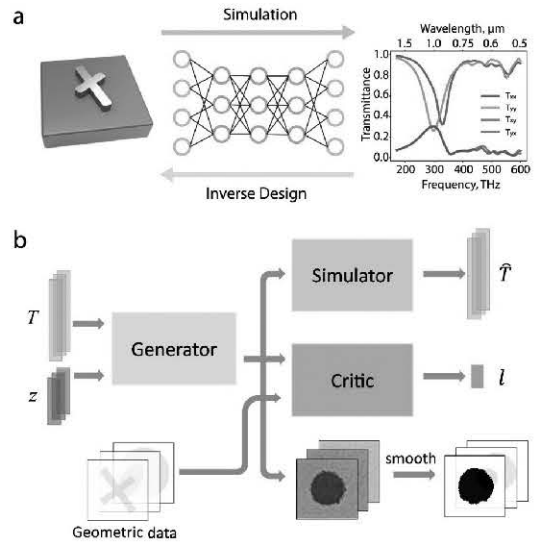


그림 15. 미국 조지아공과대학 Wenshan 교수팀의 메타물질 역방향 설계 인공지능 기술 [39]

현재 인공지능을 이용한 소재설계연구 대부분은 소재 정보로부터 물성을 예측하는 “정방향” 설계 방식이다. 향후 몇 년 내에 많은 소재분야에서 이러한 정방향 소재설계 인공지능 기술이 개발 될 것이고, 많은 연구자들이 손쉽게 이용할 수 있을 것이다. 하지만 그 반대인 사용자가 원하는 물성을 입력하여 소재 정보를 출력하는 “역방향” 설계 방식은 아직 세계적으로도 초기이다. 역방향 설계 기술이 개발이 되면 소재 개발의 패러다임을 바꿀 수 있는 큰 파급력을 지니고 있다. 따라서 이 분야에 대한 많은 투자가 필요하다.

향후 소재설계 인공지능 기술은 컴퓨터 세상에만 머무르지 않을 것이다. 소재설계 인공지능 기술은 결국에는 로봇틱스(Robotics)와 만나게 될 것으로 생각된다. 즉, 인공지능이 최적의 소재를 설계하고, 로봇이 결국 그 소재를 합성하는 시대가 곧 우리에게 다가올 것으로 생각된다. 최근 미국 MIT에서는 사용자가 원하는 유기소재 정보를 입력하면, 인공지능 로봇이 그 유기소

재를 합성하기 위한 합성반응 경로를 찾고 실제 합성에 성공한 결과를 발표한 바 있다[40]. 이는 시작에 불과하다. 향후 다양한 소재분야에 이러한 인공지능 로봇이 생겨날 것으로 기대하고 있어, 소재분야의 새로운 흐름에 우리는 준비 할 필요가 있다.

인공지능의 성패는 DB가 결정한다 해도 과언이 아니다. 미래는 소재 DB를 보유하는 소수의 연구하는 소수 연구팀이 소재분야를 리드할 가능성이 높다. 이런 점에서 우리도 이에 대한 체계적인 준비를 해야 한다. 본인의 최근 학회에 참석해보면 소재분야 인공지능 기술의 발전 속도는 학교/연구소에 비해 기업이 상대적으로 빠르다는 느낌을 받는다. 이는 기업에서 보유한 소재 DB의 양이 학교나 연구소의 단일 연구팀의 것보다 근본적으로 차이 나기 때문이다. 따라서 단일 연구 그룹에서 해결하기 보다는 대학, 연구소, 수요기업 등이 연대해 데이터와 기술을 상호 보완하는 방식으로 접근할 필요가 있다. 소재개발에 필요한 데이터의 경우 물론 많은 양을 수집하는 것도 중요하지만, 질적인 측면도 중요하다. 데이터 처리능력과 이에 소요되는 시간과 노력을 고려하면 연구개발의 방향에 따라 초기부터 데이터의 질적인 측면을 고려해 수집하는 선택과 집중 전략에 맞는 개발이 필요하다.

참고 문헌

1. 한상수, 김동훈, 융합연구리뷰 **5**, 4(2019).
2. A. Agrawal and A. Choudhary, *APL Materials* **4**, 053208 (2016).
3. B. C. Yeo, D. Kim, C. Kim, and S. S. Han, *Scientific Reports* **9**, 5879 (2019).
4. A. Chandrasekaran, D. kamal, R. Batra, C. Kim, L. Chen, and R. Ramprasad, *npj Computational Materials* **5**, 22 (2019).
5. F. Brockherde, L. Vogt, L. Li, M. E. Tuckerman, K. Burke, and K.-R. Müller, *Nature Communication* **8**, 872 (2017).
6. G. Carleo and M. Troyer, *Science* **355**, 602 (2017).
7. A. Grisafi, A. Fabrizio, B. Meyer, D. M. Wilkins, C. Cominboeuf, and M. Ceriotti, *ACS Central Science* **5**, 57 (2019).
8. J. Behler and M. Parrinello, *Physical Review Letters* **98**, 146401 (2007).
9. 강준희, 한명찬, *공업화학 전망* **22**, 13 (2019).
10. A. Khorshidi and A. A. Peterson, *Computer Physics Communications* **207**, 310 (2016).
11. B. Kolb, L. C. Lentz, and A. M. Kolpak, *Scientific Reports* **7**, 1 (2017).
12. N. Artrith and A. Urban, *Computational Materials Science* **114**, 135 (2016).
13. K. Lee, D. Yoo, W. Jeong, and S. Han, *Computer Physics Communications* **242**, 95 (2019).
14. J. Behler, *Journal of Physics: Condensed Matter* **26**, 183001 (2014).
15. W. Jeong, K. Lee, D. Yoo, D. Lee, and S. Han, *Journal of Physical Chemistry C* **122**, 22790 (2018).
16. N. Artrith, A. Urban, and G. Ceder, *The Journal of Chemical Physics* **148**, 241711 (2018).
17. Z. W. Ulissi, M. T. Tang, J. Xiao, X. Liu, D. A. Torelli, M. Karamad, K. Cummins, C. Hahn, N. S. Lewis, T. F. Jaramillo, K. Chan, J. K. Nørskov, *ACS Catalysis* **7**, 6600 (2017).
18. G. Sun and P. Sautet, *Journal of the American Chemical Society* **140**, 2812 (2018).
19. T. Xie and J. C. Grossman, *Physical Review Letters* **120**, 145301 (2018).
20. A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholla, D. Gunter, D. Skinner, G. Ceder, K. A. Persson, *APL Materials* **1**, 011002 (2013).

21. M. Kim, B. C. Yeo, Y. Park, H. M. Lee, S. S. Han, and D. Kim, arXiv: 1812.02949 (2018).
22. F. Ren, L. Ward, T. Williams, K. J. Laws, C. Wolverton, J. Hattrick-Simpers, A. Mehta, *Science Advances* **4**, eaq1566 (2018).
23. A. M. Tehrani, A. O. Oliynyk, M. Parry, Z. Rizvi, S. Couper, F. Lin, L. Miyagi, T. D. Sparks, J. Brgoch, *Journal of the American Chemical Society* **140**, 9844 (2018).
24. A. Ziletti, D. Kumar, M. Scheffler, and L. M. Ghiringhelli, *Nature Communications* **9**, 2775 (2018).
25. B. Sanchez-Lengeling and A. Aspuru-Guzik, *Science* **361**, 360 (2018).
26. D. Weininger, *J. Chem. Inf. Comput. Sci.* **28**, 31 (1998).
27. D. duvenaud, D. Maclaurin, J. Aguilera-Iparraguirre, R. Gomez-Bombarelli, T. Hirzel, A. Aspuru-Guzik, R. P. Adams, arXiv: **1509**, 09292.
28. P. Baldi, *JMLR: Workshop and Conference Proceedings* **27**, 37 (2012).
29. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Couville, Y. Bengio, *Advances in Neural Information Processing Systems*, 2672 (2014).
30. R. Gomez-Bombarelli, J. N. Wei, D. duvenaud, J. M. Hernandez-Lobato, B. Sanchez-Lengeling, D. Sheberla, J. Aguilera-Iparraguirre, T. D. Hirzel, R. P. Adams, and A. Aspuru-Guzik, *ACS Central Science* **4**, 268 (2018).
31. J. Lim, S. Ryu, J. W. Kim, and W. Y. Kim, *Journal of Cheminformatics* **10**, 31 (2018).
32. K. Kim, S. Kang, J. Yoo, Y. Kwon, Y. Nam, D. Lee, I. Kim, Y.-S. Choi, Y. Jung, S. Kim, W.-J. Son, J. Son, H. S. Lee, S. Kim, J. Shin, and S. Hwang. *npj Computational Materials* **4**, 67 (2018).
33. B. Kim, S. Lee, and J. Kim, chemrxiv. 7987475 (to appear in *Science Advances*).
34. J. Noh, J. Kim, H. S. Stein, B. Sanchez-Lengeling, J. M. Gregoire, A. Aspuru-Guzik, Y. Jung. *Matter* **1**, 1370 (2019).
35. Y. Dan, Y. Zhao, X. Li, S. Li, M. Hu, J. Hu. arXiv: 1911.05020.
36. Y. Sawada, K. Morikawa, M. Fujii. arXiv: 1910.11499.
37. J. Timoshenko, D. Lu, Y. Lin, A. I. Frenkel. *Journal of Physical Chemistry Letters* **8**, 5091 (2017).
38. D. Liu, Y. Tan, E. Khoram, Z. Yu. *ACS Photonics* **5**, 1365 (2018).
39. Z. Liu, D. Zhu, S. P. Rodrigues, K.-T. Lee, W. Cai. *Nano Letters* **18**, 6570 (2018).
40. C. W. Coley, D. A. Thomas III, J. A. M. Lummiss, J. N. Jaworski, C. P. Breen, V. Schultz, T. Hart, J. S. Fishman, L. Rogers, H. Gao, R. W. Hicklin, P. P. Plehiers, J. Byington, J. S. Piotti, W. H. Green, A. J. Hart, T. F. Jamison, and K. F. Jensen, *Science* **365**, eaax1566 (2019).