



2019. 9. 23.(월)

문의: 정연준 교수(02-880-4085, yjjung@snu.ac.kr)
연구책임자: 정연준 교수(02-880-4085) / 교신저자
연구진: 임현태 연구원(02-880-4369) / 제1저자

화학부 정연준 교수팀

인공지능 활용 용매화 자유에너지 예측 모델 개발

-세계적 화학 학술지 Chemical Science 표지 논문 선정-

□ 내용

- 용해는 용질과 용매가 서로 섞여서 용액을 만드는 가장 근본적인 화학과정으로, 화학 반응, 신약 개발, 소재 개발 등 많은 분야에서 핵심적인 역할을 한다. 이러한 중요성 때문에 그동안 다양한 계산화학 방법을 활용하여 용해 과정에 수반되는 용매화 자유에너지를 예측하는 연구가 진행되어 왔다.
- 그러나 전통적인 계산화학 방법은 막대한 수치 연산을 필요로 하기 때문에 전산 자원과 시간의 소모가 크다는 단점을 가지고 있다. 이러한 단점을 극복하기 위한 대안으로 최근 인공지능 방법을 적용하여 용매화 자유에너지를 계산하는 방법이 연구되고 있다.
- 최근 서울대학교 화학부 정연준 교수 연구팀은 자연 언어 처리(natural language processing) 기법과 인공 신경망을 결합하여, 용매화 자유에너지를 계산할 수 있는 심층 학습(deep learning) 모델을 개발하였다.
(모델명: **델포스, Delfos: Deep learning model for prediction of solvation free energies in generic organic solvents**)
- 이를 활용하면 기존의 복잡한 계산과정을 거치지 않고 단순히 용매와 용질의 분자 구조만 이용하여 용매화 자유 에너지를 알아낼 수 있으며, 약 2,500 가지의 다양한 용매-용질 조합에 대해서 테스트해 본 결과 기존 계

산방법보다 훨씬 빠른 속도를 보이면서도 정확도면에서는 유사하거나 우수한 성질을 보이는 것으로 나타났다.

- 또한 인공신경망의 작동과정 분석 결과 기계학습 예측 모델이 자유에너지를 판단하는 근거가 인간의 화학 지식과 유사한 원리를 가진다는 사실을 밝혀내었다.
- 본 연구는 기계학습 방법을 활용하여 일반적인 용매계의 자유 에너지를 계산할 수 있는 최첨단 방법론으로, 향후 소재 개발, 신약 개발 등 용액 상에서 일어나는 수많은 화학 반응에 대한 연구에 큰 기여를 할 수 있을 것으로 기대된다.
- 본 연구는 왕립 화학회(Royal Society of Chemistry)에서 발간하는 세계적인 화학 학술지 Chemical Science (Impact Factor 9.556)에 2019년 8월 20일(화)자 온라인으로 출판되었으며, 논문의 우수성과 독창성을 인정받아 후면 표지논문(outside back-cover article)으로 선정되었으며 추후 오프라인으로 출판될 예정이다.
- 본 연구는 한국연구재단에서 시행하는 미래소재디스커버리사업의 지원으로 수행되었다.
- 출판 논문 : Delfos: deep learning model for prediction of solvation free energies in generic organic solvents, Hyuntae Lim and YounJoon Jung, *Chemical Science* 2019 DOI: 10.1039/C9SC02452B (2019년 8월 20일 온라인 출판, 추후 오프라인 출판 예정)

[붙임] 1. 연구결과 2. 용어설명 3. 그림설명
4. 연구진 이력사항

연구결과

Delfos: deep learning model for prediction of solvation free energies in generic organic solvents

Hyuntae Lim (임현태) and YounJoon Jung (정연준)

(Chemical Science, DOI: 10.1039/C9SC02452B)

물은 생체 내에서 존재하는 유일한 용매이기 때문에 물질의 수용해도나 수화 에너지를 예측하는 연구는 많이 진행되어 왔다. 하지만, 유기 용매를 포함하는 비수계 용매계는 많은 화학반응에서 중요한 역할을 함에도 불구하고 용매화 자유에너지에 대해서 많은 연구가 이루어지지 않고 있다. 본 연구에서는 다양한 유기용매-유기용질 조합에서 범용적으로 용매화 자유 에너지를 계산할 수 있는 구조-성질 정량 관계(QSPR) 모델인 Delfos(Deep learning model for solvation free energies in generic organic solvents)를 개발하였다. 이 모델은 두 개의 독립된 용매 인코더와 용질 인코더 신경망이 단어 임베딩과 순환 신경망으로 분자의 구조적 특징을 추출하며, 중요한 세부구조를 판별할 수 있도록 어텐션 알고리즘으로 보강된 추측 신경망이 추출한 특징으로부터 목표로 하는 자유 에너지를 계산한다. 완성된 인공신경망 모델로 2,495가지 용매-용질 조합에 대한 용매화 자유 에너지를 계산한 결과 본 모델의 정확도는 계산화학의 최신 전산모사 방법과 비견할 수 있는 것으로 나타남은 물론 어느 세부구조들이 자유에너지 결정에 중요한 역할을 하는지에 대한 정보 또한 얻을 수 있었다.

1. 연구의 필요성

- 용질이 용매에 용해되면서 생성되는 용매화 자유 에너지는 용매 내에서 일어나는 수많은 화학반응에서 핵심적인 영향을 미치는 성질로서, 생체 내의 생화학 반응과 이차전지 내에서의 전기화학 반응 등을 연구하는데 중요한 요소이다. 이러한 중요성 때문에 분자동역학과 밀도범함수 이론 등 다양한 이론, 전산모사 접근을 통하여 자유 에너지를 계산하는 연구가

이루어지고 있으며 최근 인공지능 기술의 발달로 이를 응용하여 물질의 각종 화학적 성질들을 예측하는 방안이 소개되고 있다.

- 인공지능 기술의 발달에 따라 기계학습을 이용하여 용매화 자유 에너지를 예측하는 방법 또한 최근 연구가 이루어지고 있지만, 이미 소개된 대부분의 모델은 주로 수계에 대한 수화 에너지와 용해도만을 얻을 수 있는 등 특정 용매계에서만 작동할 수 있기 때문에 보다 범용적으로 용매의 종류에 구애받지 않고 이용할 수 있는 기계학습 모델의 개발이 필요하다.

2. 연구의 내용

- 연구팀은 단어 임베딩 방법 등 인공신경망을 이용한 자연어 처리 과정에 주로 이용되는 방법과 인공신경망 기계번역기의 구조를 이용하여 용매와 용질 분자들의 구조를 입력받아 그들이 가지는 용매화 자유 에너지를 예측할 수 있는 기계학습 구조-성질 정량 관계(QSPR) 모델을 개발하였다.
- 기계학습 모델의 구조는 각각 용매와 용질의 구조식을 입력받아 순환 신경망을 통해 분자구조를 학습하고 주요한 구조적 특징을 추출하는 부분과 추출한 구조적 특징을 다시 입력받아 이로부터 목표로 하는 자유에너지를 계산하는 부분으로 나눌 수 있다.
- 연구진이 개발한 기계학습 모델을 약 2,500가지 용매-용질 조합의 용매화 자유 에너지 실험값으로 학습을 실시하였을 때, 기대할 수 있는 정확도는 학습이 실시된 구조에 대해서 약 0.3 kcal/mol, 실시되지 못한 구조에 대해서 0.9 kcal/mol 수준으로, 비교적 적은 학습량에도 불구하고 일반적인 실험의 오차로 여겨지는 1.0 kcal/mol 이내의 값임은 물론 양자역학을 기반으로 한 밀도범함수 계산 결과와 동등 혹은 더 우수한 예측을 할 수 있는 것으로 분석되었다.
- 또한, 니트로메탄 용질이 네 가지 극성이 다른 용매에서 가지는 자유 에너지 계산을 통해 인공신경망의 작동 과정을 분석한 결과, 극성이 큰 니트로기의 중요도가 용매의 극성이 커질수록 점점 증가하는 양상을 보이는 등 개발된 모델이 용매화 자유 에너지를 판단하는 근거가 인간의 직관적인 화학 지식과 유사함을 발견하였다.

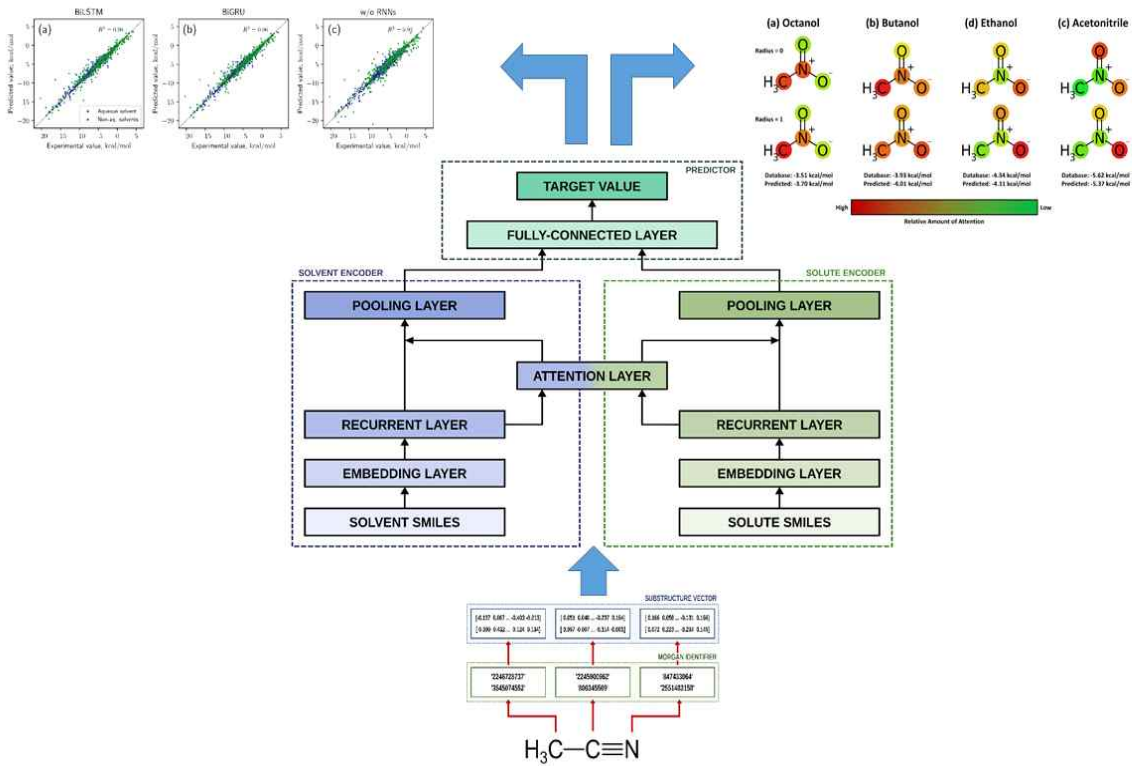


그림: 인공지능 기반 용매화 자유에너지 예측 모델인 Delfos의 작동 원리를 나타내는 모식도

3. 연구의 성과 및 의의

- 기존에 실시되었던 많은 기계학습 연구에서는 특정 용매 내에서의 용해 관련 성질들만을 예측할 수 있었으나, 본 연구로 이러한 한계를 극복하여 다양한 용매계에서 범용적으로 사용할 수 있는 기계학습 모델의 실현이 가능함이 증명되었다.
- 완성된 기계학습 모델의 정확도는 실험을 통한 측정이나 양자역학을 기반으로 한 전산모사 방법과 비교할 만한 수준에 이르렀으며, 기계학습의 특성 상 학습이 이루어지고 난 뒤부터는 매우 빠른 속도로 결과를 내놓는 것이 가능하기 때문에 다양한 분야에 이르는 관련 연구에 큰 기여를 할 수 있을 것으로 기대된다.
- 또한 인공지능망을 이용한 기계학습은 그 범용성이 매우 높아 용해 관련 성질에 국한되지 않고 이온성 액체의 물리/화학적 성질, 화학 반응 예측 등 두 가지 이상의 화학종 간의 상호작용으로 발생하는 성질에 관련된 다양한 응용 분야에 적용될 수 있을 것으로 예상된다.

용 어 설 명

1. 용매화 자유 에너지

- 기체상태에 있던 용질 분자가 액체상태의 용질 내로 이동하면서 수반하는 자유 에너지의 변화를 의미하며, 이 과정에서 용매 분자들이 용질 분자와의 상호작용으로 인한 재배열이 이루어지면서 변화하는 에너지 및 엔트로피가 주된 원인이다. 물질의 용해도와 용매 내에서의 산성도 등 기본적인 용해 관련 성질과 직접적인 연관이 있으며 이를 예측하는 방법론의 개발과 응용은 계산화학의 주요한 분야 중 하나이다.

2. 구조-성질 정량 관계

- 화합물의 분자 구조와 화합물이 가지는 각종 물리/화학/생물학적 특성을 선형 회귀법 혹은 로지스틱 회귀법을 통해 연관성을 분석하는 방법으로, 엄밀하게 정의된 이론적 배경 없이도 손쉽게 각종 성질을 빠르게 예측할 수 있다. 최근 기계학습을 통한 회귀분석이 적용되어 빠른 발전이 이루어지고 있다.

3. 자연어 처리

- 기계번역이나 문장의 내용 분석 등 컴퓨터를 통해 인간의 언어를 처리하는 컴퓨터 과학의 분야 중 하나이다. 최근 인공지능 기술의 발달로 인해 급격한 발전 추세를 보이고 있으며 인공지능의 최신 기술들이 가장 먼저 연구 및 적용되고 있는 분야이다.

4. 단어 임베딩

- 자연어 처리에서 인간 언어의 문장을 인공신경망에 입력하기 위한 기법으로, 문장 내 단어들의 분포를 분석하여 이를 토대로 단어별로 미리 정해진 길이를 가지는 벡터를 생성한다. 임베딩이 적절히 이루어지면 공간 내 벡터의 위치를 토대로 단어의 언어적 의미를 예측할 수 있으며, 이를 응용하여 분자 구조를 인공신경망에 입력하는 연구가 매우 최근에 시도되고 있다.

5. 순환 신경망

- 시계열 등 특정 순서에 따라 배열된 입력값을 처리하는데 적합하며 학습 과정에서 앞뒤의 입력값을 참고하기 때문에 자연어 처리 분야에서 문장의 문맥을 고려할 수 있는 등의 장점을 가지고 있다.

Chemical Science

www.rsc.org/chemicalscience



ISSN



EDGE ARTICLE

Chemical Science 저널의 후면 표지로 선정된 논문으로 인공지능망을 통해 용매-용질 혼합물(용액)의 성질을 예측하는 과정을 표현한다.

연구자 이력사항

<교신저자>

1. 인적사항

- 성명 : 정연준
- 소속 : 서울대학교 화학부 교수
- 전화 : 02-880-4085
- E-mail : yjjung@snu.ac.kr



2. 학력

- 1990 - 1994 서울대학교 학사
- 1994 - 1997 서울대학교 석사
- 1997 - 2002 MIT 박사

3. 경력사항

- 2002 - 2005 University of California, Berkeley 밀러 펠로우
- 2005 - 2006 Northwestern University 박사후 연구원
- 2006 - 서울대학교 화학부 교수
- 2012-2013 University of California, Berkeley 방문 교수

4. 기타 정보

- 2010. 9. 서울대학교 자연과학대학 교육부문 우수교수 선정
- 2016. 9. 삼성미래기술육성사업 미래기술연구과제 선정
- 2019. 4. 대한화학회 신국조 학술상
- 2019. 9. 서울대학교 자연과학대학 우수 강의상

<제 1저자>

1. 인적사항

- 성명 : 임현태
- 소속 : 서울대학교 화학부
계산나노바이오화학 연구실
- 전화 : 02-880-4369
- E-mail : ht0620@snu.ac.kr



2. 학력

- 2006 - 2010 서울대학교 화학부 학사
- 2010 - 서울대학교 화학부 박사과정