



(19) 대한민국특허청(KR)  
(12) 등록특허공보(B1)

(45) 공고일자 2023년09월20일  
(11) 등록번호 10-2581404  
(24) 등록일자 2023년09월18일

- (51) 국제특허분류(Int. Cl.)  
G16C 20/30 (2019.01) G16C 20/10 (2019.01)  
G16C 20/60 (2019.01) G16C 20/70 (2019.01)
- (52) CPC특허분류  
G16C 20/30 (2019.02)  
G16C 20/10 (2019.02)
- (21) 출원번호 10-2021-0123337
- (22) 출원일자 2021년09월15일  
심사청구일자 2021년09월15일
- (65) 공개번호 10-2023-0040120
- (43) 공개일자 2023년03월22일
- (56) 선행기술조사문헌

- (73) 특허권자  
포항공과대학교 산학협력단  
경상북도 포항시 남구 청암로 77 (지곡동)
- (72) 발명자  
이동화  
경상북도 포항시 남구 청암로 77(지곡동) 포항공과대학교 신소재공학과  
도경연  
경상남도 창원시 의창구 사림로8번길 57 (사림동)  
(뒷면에 계속)
- (74) 대리인  
특허법인이상

D. Tarasov et al., Application of Bayesian Artificial Neural Networks for Modeling the Dependence of Nickel-based Superalloys Ultimate Strength on their Chemical Composition, 2019 ICCAIRO(2020)\*  
J. Jung et al., Microstructure design using machine learning generated low dimensional and continuous design space, Materialia Vol.11, 100690(2020.05.)\*  
T. Long et al., CCDCGAN: Inverse design of crystal structures, arXiv:2007.11228v1(2020.07.22.)\*

\*는 심사관에 의하여 인용된 문헌

전체 청구항 수 : 총 7 항

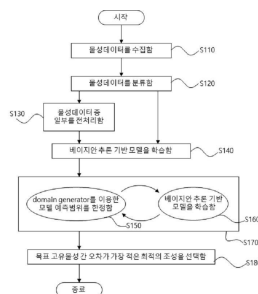
심사관 : 권계민

(54) 발명의 명칭 계산과학과 베이지안 추론 기반의 소재 역설계를 통한 화합물의 조성 예측 방법

(57) 요약

본 발명은 화합물의 조성 예측 방법에 관한 것으로, 계산 데이터를 활용해 실험으로 수집한 물성 관련 데이터의 한계를 극복하여 넓은 조성 영역에서 정확하게 새로운 조성의 화합물을 예측하는 것을 목적으로 하며, 부족한 실험 데이터로부터 발생할 수 있는 예측의 한계를 제일원리 계산을 통해 보완하고, 확장된 실험-계산 데이터의 조합을 바탕으로 베이지안 추론을 통해 최적의 물성을 가질 수 있는 조성 영역을 예측함으로써 실험과 계산 데이터의 단점을 모두 극복하여 오차가 가장 작은 조성을 예측하는 방법을 제공할 수 있다.

대표도 - 도2



- (52) CPC특허분류  
G16C 20/60 (2019.02)  
G16C 20/70 (2019.02)

**김선규**

경상북도 포항시 남구 연일읍 유강길9번길 13 드림  
하이빌 301호

- (72) 발명자  
**김신현**  
부산광역시 북구 의성로78번길 15(덕천동) 336-10

이 발명을 지원한 국가연구개발사업

과제고유번호	1415169732
과제번호	20012196
부처명	산업통상자원부
과제관리(전문)기관명	한국산업기술평가관리원
연구사업명	산업기술알키미스트프로젝트(R&D)
연구과제명	AI 기반 초임계 소재
기 여 율	1/1
과제수행기관명	연세대학교 산학협력단
연구기간	2020.09.01 ~ 2021.04.30

---

## 명세서

### 청구범위

#### 청구항 1

실험을 통해 물성 관련 데이터를 수집하는 단계;

상기 물성 관련 데이터를 분류하는 단계;

상기 분류된 물성 관련 데이터 중 조성 데이터를 전처리하여 계산값으로 고유물성 데이터를 얻는 단계;

상기 계산으로 획득된 고유물성 데이터와 상기 실험을 통한 물성 관련 데이터 중 공정 조건, 측정 조건 및 미세구조를 입력하여 베이지안 추론 기반 모델을 학습하는 단계;

상기 베이지안 추론 기반 모델 학습 후 얻은 물성 관련 데이터를 영역 생성 프로그램에 입력하여 상기 고유물성 데이터의 예측 범위를 한정하는 단계;

상기 고유물성 예측 범위 중 목표물성을 나타낼 가능성이 가장 높은 고유물성을 목표로 베이지안 추론 기반 모델을 학습하는 단계;

상기 베이지안 추론 기반 모델에서 선택된 고유물성을 목표로 영역 생성 프로그램 및 베이지안 추론 기반 모델의 반복학습을 통한 베이지안 최적화를 수행하는 단계; 및

상기 베이지안 최적화를 수행한 후, 목표 고유물성 간 오차가 가장 작은 최적의 조성을 예측하는 단계;를 포함하는 화합물의 조성 예측 방법.

#### 청구항 2

제1항에 있어서,

상기 계산으로 획득된 고유물성 데이터와 상기 공정 조건, 측정 조건 및 미세구조를 입력하여 베이지안 추론 기반 모델을 학습하는 단계와 상기 고유물성 데이터의 예측 범위를 한정하는 단계 사이에,

상기 계산으로 획득된 고유물성 중 목표 물성을 가장 잘 예측하는 조합을 선택하는 단계;를 더 포함하는 것을 특징으로 하는 화합물의 조성 예측 방법.

#### 청구항 3

제1항에 있어서,

상기 물성 관련 데이터는 조성, 공정 조건, 측정 조건, 고유물성, 미세구조 및 외적물성을 포함하는 화합물의 조성 예측 방법.

#### 청구항 4

제1항에 있어서,

상기 전처리는 상기 물성 관련 데이터 중 조성을 제일원리계산에 의해 처리하여 계산값으로 고유물성을 얻는 것인 화합물의 조성 예측 방법.

#### 청구항 5

제1항에 있어서,

상기 베이지안 추론 기반 모델은 데이터를 확률변수로 취급하여 확률변수간 확률분포를 학습할 수 있는 수학적 모델인 것을 특징으로 하는 화합물의 조성 예측 방법.

#### 청구항 6

제1항에 있어서,

상기 베이지안 추론 기반 모델의 반복학습은 베이지안 최적화에서 예측한 조성과 상기 조성에 대한 제일원리 계산으로 예측한 고유물성 데이터를 추가 보완한 데이터베이스를 활용하는 것인 화합물의 조성 예측 방법.

**청구항 7**

제1항에 있어서,

상기 베이지안 최적화를 통한 학습은 5 내지 10회 반복하여 목표로 하는 고유물성과 오차가 가장 작은 조성을 예측하는 것인 화합물의 조성 예측 방법.

**발명의 설명**

**기술 분야**

[0001] 본 발명은 소재 역설계를 통한 화합물의 조성 예측 방법에 관한 것으로, 더욱 상세하게는 제한된 실험 물성 데이터를 계산과학 데이터를 이용해 보완하고 이를 베이지안 추론과 연계하여 목표하는 물성을 만족하는 화합물의 조성을 찾는 목적 지향형 소재 설계에 관한 것이다.

**배경 기술**

[0002] 기계학습 기반으로 소재가 가질 수 있는 물성을 학습하고 이를 바탕으로 최적의 물성을 갖는 새로운 조성의 소재를 예측하는 소재 역설계 기법은 실험으로 진행된 연구 결과들을 데이터베이스화 하여 이를 기반으로 실험 조건에 따른 물성 변화를 학습하고 그 결과를 바탕으로 최적의 조건을 추론하는 방법이다. 하지만 실험에 요구되는 시간과 비용의 한계로 실험결과의 데이터베이스는 제한적일 수밖에 없고 이로 인하여 학습의 정확도 및 학습이 가능한 범위가 제한되는 문제가 발생한다. 이러한 문제를 극복하기 위해서 우리는 제한적인 실험결과 데이터베이스가 가진 가능성을 극대화 하는 두 가지 전략을 활용한다. 데이터베이스와 기계학습모델을 통한 예측 과정에서 발생하는 불확실성을 정량화 하여 다루는 베이지안 추론의 활용과 제일원리 계산을 통해 얻어진 물성 데이터베이스를 추가 보완함으로써 실험 결과만으로 학습을 진행할 때 발생할 수 있는 한계를 극복하고 넓은 범위에 걸쳐 학습을 진행하고 정확하게 최적의 조성 조합을 예측할 수 있는 방법론에 대해 제안하고자 한다.

[0003] 예측 과정에서 발생하는 불확실성의 원인은 크게 데이터베이스와 기계학습모델로 나눌 수 있다. 상기, 데이터베이스에서 불확실성은 관찰 대상 자체, 관찰 기기의 한계, 측정값을 데이터베이스화 하는 과정에서 입력 오류에 기인한다. 한편, 상기 기계학습모델에서 불확실성은 모델의 일반화 가능성 대비 학습된 데이터가 부족할 경우 모델의 자유도에 비례해서 나타난다.

[0004] 화합물의 양자적 현상에 따른 물성을 예측하는 시뮬레이션 방법으로 제일원리 계산법(Ab-initio calculations)을 사용한다. 상기 제일원리 계산법은 양자역학에 기초한 계산과학(computational science)의 방법으로, 크게 하트리폭(Hartree-Fock) 방법과 밀도 범함수 이론(density functional theory)에 기초한 방법으로 나눌 수 있다.

[0005] 상기 하트리폭 방법은 다체 시스템에서 바닥 상태의 파동함수와 에너지를 구하는 근사 방법으로 양자역학에서 나타나는 교환에너지를 정확히 계산하는 반면, 시스템의 전자 개수에 따라 연산 시간이 기하급수적으로 길어져 일반 화합물에 적용하기는 어렵다. 상기 밀도 범함수 이론에 기초한 방법은 파동함수 대신 전자밀도 함수로 대체하는 방법으로, 하트리폭 방법보다 빠르게 계산할 수 있다.

[0006] 그러나 상기 밀도 범함수 이론에 기초한 방법은 적용하는 밀도 범함수의 종류에 따라 계산 결과가 상이할 수 있고, 화합물의 물성을 예측시 각 화합물에 따라 정확도를 높일 수 있는 적절한 방법을 결정할 선행적 근거가 부족하다.

[0007] 한편, 소재 역설계란 목표로 하는 물성을 만족하는 화합물의 조성을 찾는 것을 의미한다. 실험이나 계산과학 모두 주어진 조성에 대하여 해당 물질의 물성을 예측하는 것은 쉽지만 역으로 물성으로부터 조성을 예측하는 것은 어렵다. 정보적인 관점에서 조성에서 물성으로 가는 관계는 고차원 공간에서 저차원 공간으로 사상하는 것이기 때문에 결정적이지만 물성에서 조성으로 가는 관계는 저차원 공간에서 고차원 공간으로 사상이기 때문에 결정적이지 못하다. 즉, 예측하는 값뿐만 아니라 몇 개를 예측해야 하는지 알아야 한다.

[0008] 최근 머신 러닝을 이용한 구조-특성 관계의 학습과 무작위 샘플을 활용한 역설계 방법이 제안되었다. 그러나, 산업적으로 유의미한 외적 요인을 포함한 물성의 경우 제안된 방법으로 고려할 수 없기 때문에 보완이 필요한

실정이다.

**선행기술문헌**

**특허문헌**

[0009] (특허문헌 0001) 한국특허등록번호 제10-2214634호

**발명의 내용**

**해결하려는 과제**

[0010] 상기와 같은 문제점을 해결하기 위한 본 발명의 목적은 제한된 실험 데이터의 한계를 극복하기 위해 제일원리 계산을 통해 원하는 목적에 맞는 데이터를 생산하고 이를 통해 실험 데이터의 한계를 극복하고 베이지안 추론에 기반한 외적 요인에 대한 일반화 모델을 학습함으로써 목표 물성을 나타낼 가능성이 높은 화합물의 조성을 예측하는 방법을 제공하는 데 있다.

[0011] 본 발명의 기술적 과제들은 이상에서 언급한 기술적 과제로 제한되지 않으며, 언급되지 않은 또 다른 기술적 과제들은 아래의 기재로부터 당업자에게 명확하게 이해될 수 있을 것이다.

**과제의 해결 수단**

[0012] 상기 목적을 달성하기 위한 본 발명의 일 실시예에 따른 화합물의 조성 예측 방법은 실험을 통해 물성 관련 데이터를 수집하는 단계, 상기 물성 관련 데이터를 분류하는 단계, 상기 분류한 데이터 중 일부를 전처리하는 단계, 상기 물성 관련 데이터 및 상기 전처리한 데이터를 입력하여 베이지안 추론에 기반한 모델을 학습하는 단계, 상기 모델을 학습 후 얻은 물성 관련데이터를 영역 생성 프로그램에 입력하여 고유물성 예측 범위를 설정하는 단계, 상기 고유물성 예측 범위 중 목표물성을 나타낼 가능성이 가장 높은 고유물성을 목표로 베이지안 추론에 기반한 모델을 학습하는 단계, 상기 베이지안 추론 기반 모델에서 선택된 고유물성을 목표로 영역 생성 프로그램 및 베이지안 추론 기반 모델의 반복학습을 통한 베이지안 최적화를 수행하는 단계 및 상기 베이지안 최적화를 수행한 후, 목표 고유물성 간 오차가 가장 작은 최적의 조성을 예측하는 단계를 포함할 수 있다.

[0013] 상기 물성 관련 데이터 및 전처리한 데이터를 입력하여 베이지안 추론에 기반한 모델을 학습하는 단계에서, 계산된 고유물성 중 목표 물성을 가장 잘 예측하는 조합을 선택하는 단계를 더 포함할 수 있다.

[0014] 상기 물성 관련 데이터는 조성, 공정 조건, 측정 조건, 고유물성, 미세구조 및 외적물성을 포함할 수 있다.

[0015] 상기 전처리는 상기 물성 관련 데이터 중 조성을 제일원리계산에 의해 처리하여 계산값으로 고유물성을 얻는 것일 수 있다.

[0016] 상기 베이지안 추론 기반 모델은 데이터를 확률변수로 취급하여 확률변수간 확률분포를 학습할 수 있는 수학적 모델인 것을 특징으로 할 수 있다.

[0017] 상기 베이지안 추론 기반 모델의 반복학습은 베이지안 최적화에서 예측한 조성과 상기 조성에 대한 제일원리 계산으로 예측한 고유물성 데이터를 추가 보완한 데이터베이스를 활용하는 것일 수 있다.

[0018] 상기 베이지안 최적화를 통한 학습은 5 내지 10회 반복하여 목표로 하는 고유물성과 오차가 가장 작은 조성을 예측하는 것일 수 있다.

**발명의 효과**

[0019] 상술한 바와 같이 본 발명의 일 실시예에 따르면 계산과학과 베이지안(Bayesian) 추론을 통해 화합물의 목표 물성과 조성 사이의 관계를 파악할 수 있는 일반화 모델을 학습함으로써 목표로 하는 물성을 갖는 화합물의 조성을 예측할 수 있다.

[0020] 그러나, 본 발명의 효과들은 이상에서 언급한 효과로 제한되지 않으며, 언급되지 않은 또 다른 효과들은 아래의 기재로부터 당업자에게 명확하게 이해될 수 있을 것이다.

**도면의 간단한 설명**

- [0021] 도 1은 물질 데이터의 계층구조를 나타낸 모식도이다.  
 도 2는 본 발명의 일 실시예에 따른 화합물의 조성을 예측하는 방법을 차례로 보여주는 플로우 차트이다.  
 도 3은 물성 데이터를 통해 화합물의 조성을 예측하는 최적화된 알고리즘을 나타낸 모식도이다.  
 도 4는 본 발명의 일 실시예에 따라 극저온 합금에 대하여 수집된 물성 데이터로 베이지안 추론 기반 모델 중 하나인 베이지안 신경망(Bayesian neural networks)을 학습한 후 영역 생성 프로그램(domain generator)를 활용하여 목표 고유물성 예측 범위를 한정된 것을 나타낸 그래프이다.  
 도 5는 본 발명의 일 실시예에 따른 극저온 합금에 대하여 목표하는 물성을 만족할 가능성이 높은 상위 3개의 고유물성 조합을 나타낸 표이다.  
 도 6은 본 발명의 일 실시예에 따른 극저온 합금에 대하여 항복강도가 가장 높을 것으로 예상되는 고유물성 조합과 해당 조합에 가장 가까운 새로운 화합물의 조성 조합을 찾기 위해 베이지안 최적화 방법(bayesian optimization)으로 예측한 조성의 고유물성 오차를 나타낸 그래프이다.

**발명을 실시하기 위한 구체적인 내용**

- [0022] 본 발명은 다양한 변경을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 상세한 설명에 상세하게 설명하고자 한다. 그러나, 이는 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다. 각 도면을 설명하면서 유사한 참조부호를 유사한 구성요소에 대해 사용하였다.
- [0023] 본 출원에서 사용한 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 출원에서, "포함하다" 또는 "가지다" 등의 용어는 명세서상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.
- [0024] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가지고 있다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥상 가지는 의미와 일치하는 의미를 가지는 것으로 해석되어야 하며, 본 출원에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.
- [0026] 이하, 본 발명에 따른 바람직한 실시예를 첨부된 도면을 참조하여 상세하게 설명한다.
- [0027] 도 1은 물성 관련 데이터의 계층구조를 나타낸 모식도이다. 도 1을 참조하면, 상기 물성 관련 데이터는 조성, 공정 조건(합성) 및 측정 조건과 같은 독립변수와 고유물성, 미세구조 및 외적물성과 같은 종속변수로 구분될 수 있다.
- [0028] 상기 고유물성은 물질의 양, 크기, 형태 및 외부 상호작용과 무관하게 나타나는 물성을 의미하며 조성에 의존할 수 있고, 상기 미세구조는 고유물성 및 공정 조건에 의존하는 것일 수 있다. 상기 외적물성은 물성 중 고유물성을 제외한 나머지 물성을 의미하며, 고유물성, 미세구조 및 측정 조건에 의존하는 것일 수 있다. 또한, 상기 외적물성은 특정 미세구조와 측정 조건 하에서 고유물성을 매개변인으로 하여 조성에 의존할 수 있다.
- [0029] 상기 고유물성은 계산과학의 발달로 제일원리 기반의 계산을 통해 높은 수준으로 예측 가능할 수 있다. 상기 제일원리계산은 물질의 상태를 기술하는 슈뢰딩거 방정식의 해를 구하는 것으로 실험 없이도 플랑크 상수나 전자의 질량과 같은 기본 상수로부터 물질의 특성을 계산할 수 있다. 또한 상기 제일원리계산은 전자를 직접적으로 기술하므로 전자 수준의 시공간에서 실험과 독립적으로 응집물질의 고유물성을 예측할 수 있어, 외적 요인이 배제된 고유물성과 조성 사이의 관계를 파악하는 것이 가능할 수 있다.
- [0030] 구체적으로, 상기 제일원리계산은 슈뢰딩거 방정식의 해를 구하는 다양한 Ab-initio 계산 방법이나 First-Principles 계산 방법을 총칭하는 것으로 QMC(quantum monte carlo) method, GW method, DFT(density functional theory) method, HF (Hartree Fock) method 및 EMTO-CPA(exact muffin-tin orbital coherent potential approximation) 방법 등을 포함하며 이를 통해 복잡한 조성의 다양한 고유물성을 정확하게 예측할 수 있다.

- [0031] 하지만 외적물성은 고유물성 외에도 점결합, 전위, 입계 등 결함으로 나타나는 미세구조 까지 영향을 받는 물성으로 상기 제일원리계산 만을 가지고 예측하는 데에는 한계가 있고, 실험을 통해 측정된 외부요인의 효과를 데이터에 추가하는 것이 정확한 외적물성 예측에 유리할 수 있다. 따라서 상기 데이터를 계층구조의 형태로 실험과 계산 데이터를 상호 보완하여 조성과 외적물성을 연계할 수 있는 베이지안 추론 기법이 제안될 수 있다.
- [0032] 도 2는 본 발명의 일 실시예에 따른 화합물의 조성을 예측하는 방법을 차례로 보여주는 플로우 차트이다. 본 발명의 일 실시예에 따른 원하는 물성을 갖는 최적의 화합물 조성을 예측하는 방법은, 실험을 통해 물성 관련 데이터를 수집하는 단계(S110); 상기 수집한 물성 관련 데이터를 분류하는 단계(S120); 상기 분류한 데이터 중 일부를 전처리하는 단계(S130); 상기 물성 관련 데이터 및 상기 전처리한 데이터를 입력하여 베이지안 추론(Bayesian inference) 기반 모델을 학습하는 단계(S140); 상기 모델 학습 후 얻은 데이터를 영역 생성 프로그램(Domain generator)에 입력하여 고유물성 예측 범위를 설정하는 단계(S150); 상기 고유물성 예측 범위 중 목표 물성을 나타낼 가능성이 가장 높은 고유물성을 목표로 베이지안 추론 기반 모델을 학습하는 단계(S160); 상기 베이지안 추론 기반 모델에서 선택된 고유물성을 목표로 조성과 고유물성 데이터로부터 영역 생성 프로그램 및 베이지안 추론 기반 모델의 반복학습하여 베이지안 최적화를 수행하는 단계(S170); 및 상기 베이지안 최적화를 수행한 후, 목표 고유물성 간 오차가 가장 작은 최적의 조성을 예측하는 단계(S180)을 포함할 수 있다.
- [0033] 또한, 상기 물성 관련 데이터 및 전처리한 데이터를 입력하여 베이지안 추론에 기반한 모델을 학습하는 단계(S140)와 상기 모델 학습 후 얻은 데이터를 영역 생성 프로그램(Domain generator)에 입력하여 고유물성 예측 범위를 설정하는 단계(S150) 사이에, 계산된 고유물성 중 목표 물성을 가장 잘 예측하는 조합을 선택하는 단계;를 더 포함할 수 있다.
- [0034] 도 2를 참조하면, 본 발명의 일 실시예에 따른 최적의 조성을 예측하는 방법은 먼저 실험을 통해 물성 관련 실험 데이터를 수집(S110)한 후, 상기 실험 데이터를 조성, 공정 조건(합성), 측정 조건, 미세구조 및 외적물성으로 분류(S120)할 수 있다. 분류한 데이터 중 상기 조성 데이터는 제일원리계산 중 EMPO-CPA method를 통한 전처리 과정을 거쳐 계산값으로 고유물성을 얻는 것(S130)일 수 있다.
- [0035] 상기 고유물성(S130)과 상기 분류(S120)한 데이터 중 공정 조건, 측정 조건 및 미세구조를 입력하여 베이지안 추론 기반 모델 중 베이지안 신경망을 통한 학습을 수행(S140)할 수 있다. 상기 베이지안 신경망을 통한 학습은 베이지안 추론을 기반으로 행하여지며, 학습을 수행한 후 외적물성을 출력값으로 얻는 것일 수 있다.
- [0036] 상기 베이지안 추론은 베이스 정리에 따라 가설에 대한 확률을 새로운 정보를 통해 갱신하는 확률적 추론 방법으로, 추론 대상의 사전 확률과 추가적인 정보를 통해 해당 대상의 사후 확률을 추론하는 것일 수 있다. 구체적으로, 주어진 조성과 대응하는 물성의 관계가 1:1인 함수 모델을 가지고 상기 물질 데이터로부터 조성 및 물성 간의 관계를 예측할 수 있다. 이 때, 모든 데이터는 오차를 가질 가능성이 존재하며, 상기 함수 모델로 학습할 경우, 학습과정에서 데이터 오차가 최소화되도록 학습이 진행되지만 데이터 자체적인 오차는 함수 모델에 포함되지 않을 수 있다. 또한, 데이터가 균일하게 분포하지 못한 경우에는 데이터가 밀집된 영역에서는 과적합되고, 데이터가 소한 영역은 적합하지 않게 될 수 있다.
- [0037] 상기 베이지안 신경망에 적용된 함수 모델이 자체적으로 오차를 학습할 경우, 상기 데이터의 분포가 균일하지 못한 문제를 고려하여 분포 모델을 생각할 수 있다. 상기 분포 모델에서는 조성과 대응하는 물성의 관계가 1:1로 정해지지 않고, 모든 가능성을 고려하여 어떤 물성이 나타날 가능성이 높고 낮은지를 구분할 수 있다.
- [0038] 상기 조성과 물성이 대응하는 관계가 분명한 경우, 특정 물성 값에서 높은 확률이 나타나고 나머지 물성 값에서는 0에 가까운 확률을 보일 수 있다. 이와 반대로 상기 조성과 물성이 대응하는 관계가 분명하지 않은 경우, 여러 물성 값에서 비슷한 확률이 고르게 분포할 수 있다. 따라서, 상기 베이지안 추론을 통해 가설에 해당하는 상기 분포 모델을 새로운 정보인 물성 데이터를 통해 한정할 수 있다. 또한, 상기 분포 모델에서 분포의 평균과 표준편차를 각각 물성값과 오차로 해석하여 소재 역설계에 활용할 수 있다.
- [0039] 한편, 상기 전처리 과정(S130)을 거쳐 계산값으로 얻은 고유물성 데이터를 영역 생성 프로그램을 통해 고유물성 예측 범위를 설정(S150)할 수 있다. 상기 영역 생성 프로그램은 데이터의 공간을 유효 범위만큼 한정함으로써 예측 성능을 높일 때 적용하는 것일 수 있다. 상기 영역 생성 프로그램을 이용함으로써 실험을 통한 데이터 획득시 비용이 높은 점과 화합물의 개발시 데이터 포인트가 군집화 되어있어, 물성 데이터는 추가 획득 비용이 크고 많은 데이터 오차를 포함하여 군집 영역을 벗어난 기계학습시 학습률이 떨어지는 단점을 보완할 수 있다.
- [0040] 상기 영역 생성 프로그램은 구체적으로 Convex hull 방식 및 ball 방식을 이용할 수 있다. 상기 Convex hull 방식은 임의의 차원의 공간에 여러 개의 점이 주어졌을 때, 모든 점을 포함하는 볼록 껍질 내부의 영역과 볼록 껍

질로 이루어진 다면체의 최대 지름의 0배수에서 1배수에 사이에 해당하는 거리만큼 볼록 꺾질 바깥으로 뺀 나간 거리 사이의 영역 내부를 이루는 점들을 구하는 것일 수 있다. 상기 ball 방식은 임의의 차원의 공간에 여러 개의 점이 주어졌을 때, 모든 점을 하나의 그래프로 연결할 수 있는 최소 거리의 절반을 반지름으로 하여, 모든 점에 대하여 상기 최소 거리의 절반을 반지름으로 하는 ball 내부를 이루는 점들을 구하는 것일 수 있다.

- [0041] 상기 영역 생성 프로그램을 통해 설정(S150)한 고유물성 예측 범위와 공정 조건, 측정 조건 및 미세구조를 입력하여 베이지안 신경망을 통한 학습을 수행(S140)할 때 입력한 데이터인 합성(공정 조건), 측정 조건 및 미세구조의 조합에 대하여 참값과 예측값의 상관관계가 가장 높은 조합을 선택할 수 있다. 상기 선택된 조합을 입력 데이터로 하여 베이지안 신경망을 통한 학습을 수행하고, 데이터의 오차가 평균에서 표준편차 이상인 샘플에 대하여 입력 오류를 보정하고 입력 오류가 아닌 경우 배제하는 것일 수 있다.
- [0042] 상기 미세구조 및 측정조건을 고정하고, 상기 예측된 고유물성을 영역 생성 프로그램으로 생성된 조성의 범위 내에서 베이지안 추론 기반 모델 중 GP 모델을 통한 학습(S160)을 수행하고, 상기 GP 모델에서 선택된 고유물성을 목표로 조성과 고유물성 데이터로부터 영역 생성 프로그램 및 베이지안 추론 기반 모델을 반복학습하여 최적의 화합물의 조성을 예측하는 베이지안 최적화(S170)를 수행할 수 있다.
- [0043] 상기 베이지안 추론 기반 모델의 반복학습은 베이지안 최적화에서 예측한 조성과 상기 조성에 대한 제일원리 계산으로 예측한 고유물성 데이터를 추가 보완한 데이터베이스를 활용하는 것일 수 있다.
- [0044] 상기 베이지안 최적화를 통한 학습은 2 내지 20회, 구체적으로 5 내지 10회 반복하여 목표로 하는 고유물성과 오차가 가장 작은 조성을 선택(S180)함으로써, 원하는 물성을 갖는 화합물의 조성을 예측할 수 있다.
- [0045] 상기 베이지안 최적화 방법은 단순히 무작위 추출을 반복하는 것보다, 기존에 추출되어 평가된 결과를 바탕으로 앞으로 탐색할 범위를 더욱 좁혀 효율적으로 시행하는 아이디어로부터 시작된 것으로, 상기 아이디어를 베이지안 이론(bayesian theory) 및 GP 모델(Gaussian process model)을 통해 구현한 것일 수 있다. 상기 베이지안 최적화 방법은 알려지지 않은 목적 함수를 최대, 최소로 하는 최적해를 찾는 기법으로 surrogate model 및 acquisition function으로 구성될 수 있다. 상기 surrogate model은 조사된 입력값과 함수값 점들을 바탕으로 함수를 추정하는 확률모델로 보통 GP 모델이 주로 사용될 수 있다.
- [0046] 상기 GP 모델은 데이터를 기반으로 모델을 학습할 수 있는 방법일 수 있다. 데이터를 얻기 위해 실험을 수행하면 너무 많은 비용이 들어가 함부로 실험을 진행하기 어려울 수 있다. 따라서, 초기에 실험 방향을 설정하고 각각의 실험에 맞는 설정값들을 미리 조정해서 계획성 있게 실행해야 실험 비용을 줄이면서도, 모델을 완성하기 위한 유효한 데이터를 얻을 수 있다.
- [0047] 초기에 몇 번의 실험을 통해 데이터를 얻고, 임의(grid search)로 설정값을 정하여 수행할 수 있다. 상기 설정값을 수행하여 얻은 실험값들을 모아서 최초의 GP 모델을 만들 수 있다. 상기 GP 모델은 평균 및 표준편차 각각의 예측값을 제시해줄 수 있는 모델로, 상기 GP 모델을 기반으로 하여 어떤 위치의 설정값들을 더 실험해볼 것인지 정하여 실험한 결과를 다시 GP 모델을 수정하는데 사용할 수 있다. 이 때, 다음 위치의 설정값을 찾기 위한 반복된 계산과정을 베이지안 최적화 방법이라 할 수 있다.
- [0048] 상기 베이지안 최적화 방법은 몇 번의 임의 설정 실험으로 데이터를 얻는 제1단계; 주어진 최소 데이터에 기반하여 GP 모델을 만들거나 또는 임의의 선형 모델로 초기화하는 제2단계; 베이지안 최적화 방법에 의해 실험 설정값을 추천받는 제3단계; 상기 추천받은 설정값으로 실험하는 제4단계; 실험 데이터를 얻는 제5단계; 및 제2단계부터 반복하는 단계를 포함할 수 있다.
- [0049] 상기 주어진 최소 데이터를 기반으로 하여 만들어진 GP 모델을 사전분포(prior distribution)이라 할 수 있고, 추가적인 실험에 의해 얻어진 정보를 우도(likelihood)라고 할 수 있다. 상기 사전분포 및 우도를 조합하여 다시 모델을 업데이트하는데 이를 사후분포(posterior distribution)이라 할 수 있으므로 이러한 전체 최적화 과정은 베이지안 이론이 적용된 것으로 볼 수 있다.
- [0050] 상기 실험 설정값을 추천받는 제3단계는 현재 모델에서 가장 실험에 필요한 부분을 exploitation해야하는 문제와 예상치 못한 미지의 설정을 해봄으로써 예측하지 못한 모델의 성능이 급격하게 좋아지게 만드는 설정값의 exploration이 필요할 수 있다. 이 때, 상기 exploitation은 현재 있는 정보로 최선의 선택을 하는 것이며, 상기 exploration은 정보를 모으기 위해 실험적인 실행을 하는 것일 수 있다. 상기 exploitation 및 exploration의 적절한 조화를 통해 설정값을 제안받아야 하며, 이를 위해 exploitation 및 exploration 간의 수식적인 표현이 필요할 수 있다. 이를 acquisition function이라고 하며, 상기 acquisition function의 대표적인 방법은 maximum probability of improvement(MPI) 및 이를 개선한 expected improvement(EI), 또 다른 방식의 upper



confidence bound(UCB)를 포함할 수 있다.

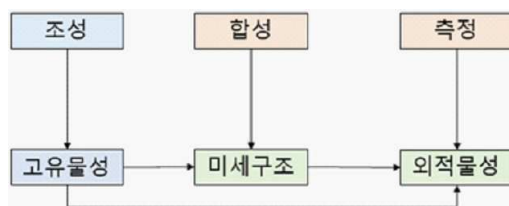
- [0051] 도 3은 물성 데이터를 통해 화합물의 조성을 예측하는 최적화된 알고리즘을 나타낸 모식도이다. 상기 알고리즘은 계산과학, 베이지안 추론 및 영역 생성 프로그램(domain generator)을 이용하여 학습하여 입력한 데이터를 가공하여 추출하는 것일 수 있다.
- [0052] 도 3을 참조하여 상기 최적화된 알고리즘을 살펴보면, 먼저 물성 관련된 실험 데이터를 수집한 후, 상기 실험 데이터를 조성, 공정 조건, 측정 조건, 미세구조 및 외적물성으로 분류할 수 있다. 상기 실험 데이터 중 조성(C)는 계산과학에 의해 고유물성인 계산( $P_1$ )값을 얻고 데이터베이스에 추가할 수 있다. 상기 계산( $P_1$ )값과 환경(E)값을 입력하여 베이지안 신경망을 학습하여 외적물성( $P_E$ )을 출력값으로 얻을 수 있다. 이 때, 상기 환경(E)값은 공정 조건, 측정 조건 및 미세구조일 수 있다.
- [0053] 영역 생성 프로그램을 이용하여 상기 계산( $P_1$ )값인 고유물성의 예측범위를 한정할 수 있으며, 데이터 입력 조합에 대하여 참값과 예측값의 상관관계가 가장 높은 조합을 선택할 수 있다. 이 때, 상기 선택한 데이터 입력 조합은 미세구조, 공정조건 및 측정조건을 포함하며, 고정값( $E_{fix}$ )을 입력하는 것일 수 있다.
- [0054] 상기 고정값( $E_{fix}$ )에 의해 한정된 고유물성의 범위( $P_E^*$ )를 입력하여 베이지안 최적화 방법으로 예측된 최적화된 고유물성( $P_1^*$ ) 값을 얻을 수 있다. 상기 베이지안 최적화 방법은 상기 영역 생성 프로그램으로 한정된 고유물성 범위에 대하여 acquisition function을 계산하여 목표물성을 나타낼 가능성이 가장 높은 최적화된 고유물성( $P_1^*$ ) 값을 선택하는 것일 수 있다.
- [0055] 상기 최적화된 고유물성( $P_1^*$ )값 및 조성(C)값으로부터 영역 생성 프로그램으로 생성된 범위를 입력하고 계산과학과 베이지안 최적화 방법으로 최적의 조성( $C^*$ ) 값을 예측할 수 있다. 이 때, 상기 영역 생성 프로그램으로 조성(C) 값에 대하여 범위를 생성하는 경우, 구체적으로 ball 방식의 영역 생성 프로그램이 사용될 수 있으며, 최종 산출물인 최적의 조성( $C^*$ ) 값의 예측시 구체적으로 GP 모델을 통해 베이지안 최적화 방법을 수행하는 것일 수 있다.
- [0057] 이하, 본 발명의 이해를 돕기 위하여 바람직한 실시예를 제시한다. 다만, 하기의 실시예는 본 발명의 이해를 돕기 위한 것일 뿐, 본 발명이 하기의 실시예에 의해 한정되는 것은 아니다.
- [0059] 실시예
- [0060] 극저온 합금에 대한 조성, 합성 조건, 측정 조건, 미세구조 및 외적 물성(실험물성)을 수집하여 데이터베이스를 구축하였다. 상기 데이터베이스 중 조성에 대하여 제일원리계산 중 EMTO-CPA 기법으로 계산 가능한 모든 고유물성을 예측하여 데이터베이스에 포함하였다. 계산하여 수득한 고유물성, 합성 조건, 측정 조건, 미세구조 및 외적 물성으로 베이지안 신경망을 통하여 학습하였다. 학습 후 얻어진 데이터의 오차 중 표준편차 이상의 데이터는 입력 과정에서 오류가 있었는지 확인하고, 입력 오류인 경우 정정하고 입력 오류가 아닌 경우는 배제하였다. 모든 고유물성 조합에 대하여 상기 베이지안 신경망을 통한 학습을 진행하고, 측정 물성과 예측 물성 사이의 상관관계가 가장 높은 조합을 선별하였다. 미세구조와 측정 조건을 고정하고 고유물성에 대하여 convex hull 방식의 영역 생성 프로그램을 활용하여 예측 범위를 한정하였다. 한정된 고유물성 범위에 대하여 acquisition function을 계산하고, 목표물성을 나타낼 가능성이 가장 높은 고유물성을 선택하였다. 선택한 고유물성을 목표로 조성과 고유물성 데이터로부터 ball 방식의 영역 생성 프로그램과 GP 모델을 통해 베이지안 최적화 방법을 수행하였다. 상기 베이지안 최적화 방법을 10회 반복 후 목표 고유물성 간 오차가 가장 작은 예측 조성을 제안하였다.
- [0062] 도 4는 본 발명의 일 실시예 따라 극저온 합금에 대하여 수집된 물성 데이터로 베이지안 추론 기반 모델 중 하나인 베이지안 신경망(Bayesian neural networks)을 학습한 후 영역 생성 프로그램(domain generator)를 활용하여 목표 고유물성 예측 범위를 한정된 것을 나타낸 그래프로, 도 3의 베이지안 최적화 방법 수행 중 (a)베이지안 신경망 학습결과, (b)데이터오차와 모델오차의 비교, (c)(a)에서 모델 오차만 표시, (d)(a)에서 데이터 오차만 표시, (e)합수모델인 신경망 결과, (f)다른 분포 모델인 GP 모델을 적용한 결과 (g)베이지안 신경망에서 영역 생성프로그램으로 제한된 범위에 대한 산점도(scatter plot)에 의해 생성된 실험 결과 데이터 집합들에 기

초하여 예측된 조성의 예측 및 실제 조성을 비교하여 나타낸 것이다.

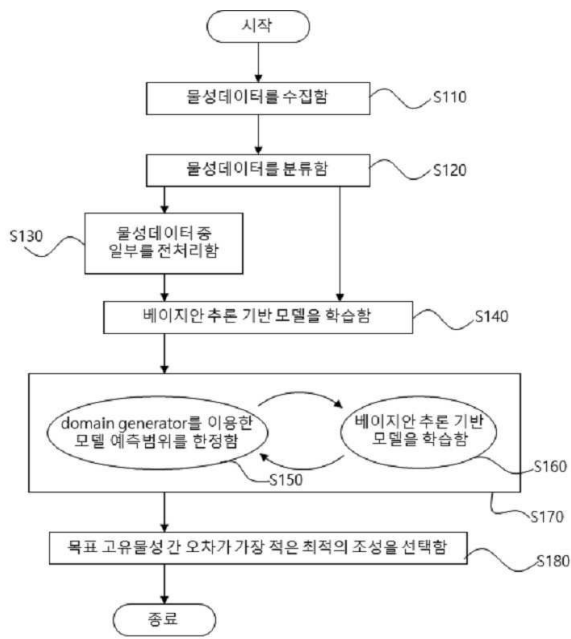
- [0063] 도 4를 참조하면, 본 발명의 일 실시예에 따라 극저온 합금에 대하여 수집된 물성 데이터로 베이지안 추론 기반 모델 중 하나인 베이지안 신경망을 학습한 후 영역 생성 프로그램을 활용하여 화합물의 조성 예측 범위를 한정하여 얻은 조성값과 상기 극저온 합금 소재의 예측된 조성값 사이의 편차는 약 6 %인 것으로 나타났다. 한편, 실제 제작되어 측정된 극저온 합금 소재의 조성값과 상기 극저온 합금 소재의 예측된 조성값 사이의 편차는 약 12 %인 것을 확인하였다.
- [0064] 도 5는 본 발명의 일 실시예에 따른 극저온 합금에 대하여 측정온도가 77 K이고 입계크기가 25.893 um일 때, 항복강도가 가장 높을 것으로 예상되는 상위 3개의 고유물성 조합을 나타낸 표이다.
- [0065] 도 6은 도 5의 극저온 합금에 대하여 항복강도가 가장 높을 것으로 예상되는 상위 3개의 고유물성 조합에 대하여, 해당 조합에 가장 가까운 새로운 조성 조합을 찾기 위해 베이지안 최적화 방법으로 예측한 조성의 고유물성 오차를 나타낸 그래프이다.
- [0066] 도 6을 참조하면, 실제 제작되어 측정된 극저온 합금 소재의 조성 값에서 상대적 오차의 최소값이 약 3.9%로 나타났다. 한편, 본 발명의 일 실시예에 따라 극저온 합금에 대하여 수집된 물성 데이터를 베이지안 최적화 방법으로 예측하여 도출된 극저온 합금 소재의 조성 값에서 상대적 오차의 최소값은 약 1.8%로 나타났다. 따라서, 상기 극저온 합금에 대하여 실험 데이터로부터 얻은 조성값에 비해 수집된 물성 데이터를 베이지안 최적화 방법으로 예측한 조성값의 정밀도가 더 높음을 알 수 있다.
- [0067] 상술한 본 발명의 실시예들은 소재의 실험적 결과 데이터에 포함된 매개변수의 결측값을 대체하고, 이를 기계학습의 입력 특징으로서 사용하여 상이한 고유 특성 및 외부 특성에 따라 그들의 성능을 예측하는 시스템으로 제공될 수도 있다. 또한, 상술한 본 발명의 실시예들은, 실험적 결과 데이터를 기반으로 기계학습의 입력 기능을 적용하는 것에 의해, 이론적 데이터 세트를 기반으로 한 입력 기능과 비교하여 계산 결과와 실험 결과 사이의 간격을 좁힌다.
- [0068] 상기에서 설명한 본 발명의 기술적 사상은 바람직한 실시예에서 구체적으로 기술되었으나, 상기 실시예는 그 설명을 위한 것이며 그 제한을 위한 것이 아님을 주의하여야 한다. 또한, 본 발명의 기술적 분야의 통상의 지식을 가진 자라면 본 발명의 기술적 사상의 범위 내에서 다양한 실시예가 가능함을 이해할 수 있을 것이다. 따라서 본 발명의 진정한 기술적 보호 범위는 첨부된 청구범위의 기술적 사상에 의해 정해져야 할 것이다.

**도면**

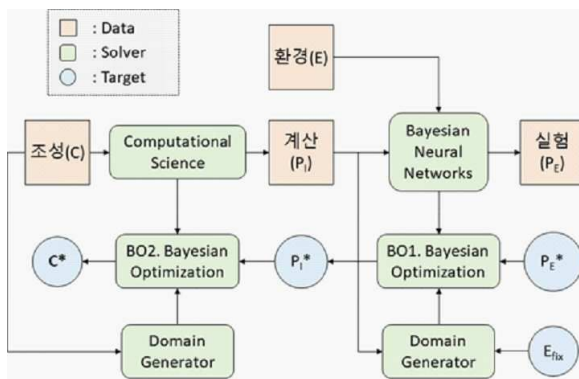
**도면1**



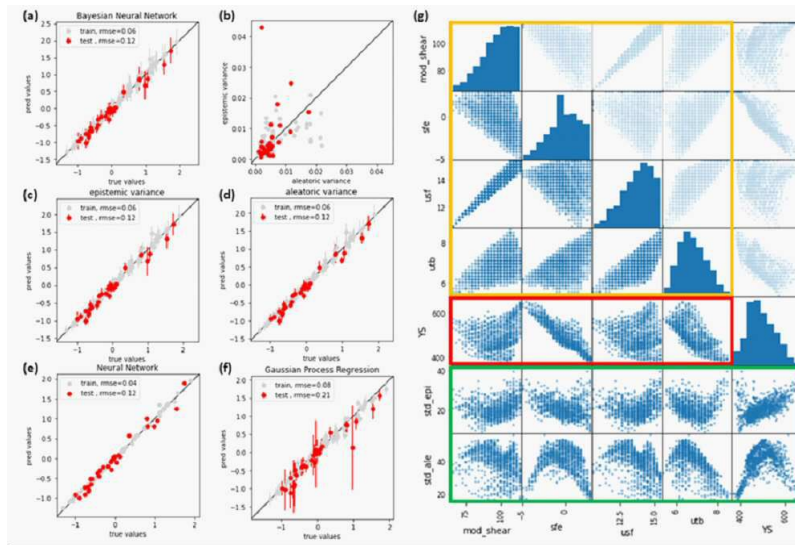
도면2



도면3



도면4



도면5

실험조건		계산물성				실험물성	
Test Temp	grain size	shear modulus	sfe	usf	utb	YS(GPa)	Expectation value
77	25.893	109.062	-3.831	13.558	6.014	546.146	0.092
77	25.893	86.384	-2.43	11.698	6.014	518.147	0.083
77	25.893	111.329	-1.729	13.558	6.474	552.235	0.08

도면6

AIM data			
mod_shear	ISF	USF	UTB
109.062 GPa	-3.381 mJ/m <sup>2</sup>	13.558 mJ/m <sup>2</sup>	6.014 mJ/m <sup>2</sup>

