(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0008664
(43) 공개일자 2024년01월19일

(51) 국제특허분류(Int. Cl.)
G06N 20/20 (2019.01) *G06N 20/10* (2019.01)
G06N 3/04 (2023.01) *G06N 3/08* (2023.01)
(52) CPC특허분류
G06N 20/20 (2021.08)
G06N 20/10 (2021.08)
(21) 출원번호 10-2022-0085785
(22) 출원일자 2022년07월12일
심사청구일자 2022년07월12일

(71) 출원인
조선대학교산학협력단
광주광역시 동구 필문대로 309 (서석동)
(72) 발명자
장일식
광주광역시 동구 밤실로 100 (산수동, 무등산광신프로그레스)
이득선
광주광역시 서구 월드컵강로 22
(74) 대리인
특허법인지원

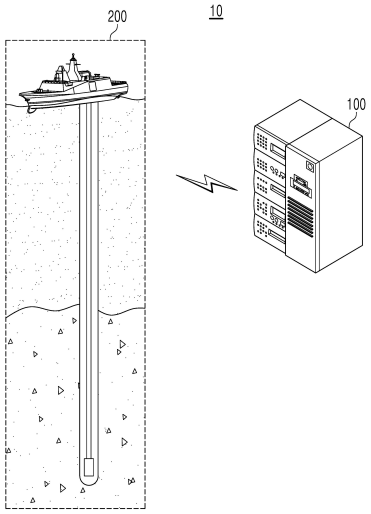
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 **예측 모델을 통한 예측 결과의 신뢰도를 평가하는 방법 및 장치**

(57) 요 약

본 발명의 일 실시 예는 예측 모델을 통한 예측 결과의 신뢰도를 평가하는 방법으로서, 예측 모델을 통한 예측에 사용되기 위한 입력 데이터를 수신하는 단계, 예측 모델을 훈련하는데 사용된 훈련 데이터를 이용하여 구축된 데이터 분류 알고리즘을 이용하여 입력 데이터를 분류하는 단계 및 분류된 입력 데이터 그룹마다 서로 다른 평가 신뢰도를 할당하는 단계를 포함할 수 있다.

대 표 도 - 도1



(52) CPC특허분류
G06N 3/045 (2023.01)
G06N 3/08 (2023.01)

이 발명을 지원한 국가연구개발사업
과제고유번호 1415174699
과제번호 20216110100070
부처명 산업통상자원부
과제관리(전문)기관명 한국에너지기술평가원
연구사업명 중소규모가스전및희소광물탐사·활용기술개발사업(R&D)
연구과제명 북미 주요 가스전의 지능형 현장 빅데이터 평가시스템 개발 및 사업화
기 여 율 1/2
과제수행기관명 에너지홀딩스그룹(주)
연구기간 2021.05.01 ~ 2022.01.31
이 발명을 지원한 국가연구개발사업
과제고유번호 1415176439
과제번호 20212010200010
부처명 산업통상자원부
과제관리(전문)기관명 한국에너지기술평가원
연구사업명 에너지수요관리핵심기술개발(에특)(R&D)
연구과제명 CO2 저장효율 향상 기술 개발
기 여 율 1/2
과제수행기관명 공주대학교산학협력단
연구기간 2021.11.01 ~ 2022.04.30

명 세 서

청구범위

청구항 1

각 단계의 적어도 일부가 프로세서에 의해 수행되는, 예측 모델을 통한 예측 결과의 신뢰도를 평가하는 방법으로서,

상기 예측 모델을 훈련하는데 사용된 훈련 데이터를 이용하여 커널 변수의 제1 값을 가지는 제1 데이터 분류 알고리즘을 구축하는 단계;

상기 예측 모델을 통한 예측에 사용되기 위한 입력 데이터를 상기 제1 데이터 분류 알고리즘을 이용하여 제1 외부 데이터 및 제1 내부 데이터로 분류하는 단계;

상기 예측 모델을 훈련하는데 사용된 훈련 데이터를 이용하여 커널 변수의 제2 값을 가지는 제2 데이터 분류 알고리즘을 구축하는 단계;

상기 제1 내부 데이터를 상기 제2 데이터 분류 알고리즘을 이용하여 제2 외부 데이터 및 제2 내부 데이터로 분류하는 단계; 및

상기 예측 모델을 이용하여 상기 제1 외부 데이터에 대해 예측된 예측 결과의 신뢰도를 제1 신뢰도로 결정하고, 상기 예측 모델을 이용하여 상기 제2 외부 데이터에 예측된 예측 결과의 신뢰도를 제2 신뢰도로 결정하는 단계를 포함하는,

예측 모델의 예측 결과 신뢰도 평가 방법.

청구항 2

제1항에 있어서,

상기 제1 및 제2 데이터 분류 알고리즘은 서포트 벡터 데이터 디스크립션(Support Vector Data Description; SVDD)이고 상기 커널 변수는 SVDD의 커널 너비를 조절하는 변수이며, 상기 커널 변수의 상기 제1 값은 상기 제2 값과 상이한 값으로 설정되는,

예측 모델의 예측 결과 신뢰도 평가 방법.

청구항 3

제2항에 있어서,

상기 제1 값은 상기 제2 값보다 작도록 설정되며,

상기 제1 신뢰도는 상기 제2 신뢰도보다 작은,

예측 모델의 예측 결과 신뢰도 평가 방법.

청구항 4

제2항에 있어서,

상기 SVDD의 구(sphere)는 하기 수학식 1의 가우시안 커널을 이용하여 생성되고,

[수학식 1]

$$K(x_i, x_j) = \exp\left(-g\|x_i - x_j\|^2\right)$$

여기서, x_i 및 x_j는 데이터 포인트이며, g는 커널 너비 조절 변수인,

예측 모델의 예측 결과 신뢰도 평가 방법.

청구항 5

제4항에 있어서,

상기 SVDD의 목적 함수는 하기 수학식 2와 같으며,

[수학식 2]

$$\min_{R,c,\xi} R^2 + C\sum_{i=1}^{n}\xi_i$$

$$\text{subject to } \| \Phi_k(x_i) - a \|_{F_k}^2 \le R^2 + \xi_i,\ \xi_i \ge 0 \ \text{ for all } i = 1,2,\cdots,n$$

여기서, R은 상기 구의 반경이고, a는 상기 구의 중심이며, ξ_i 는 소프트 마진을 위한 페널티 값이고, C는 트레이드 오프를 제어하기 위한 하이퍼파라미터이고, Φ_k 는 오리지널 데이터를 피처 공간에 맵핑하기 위한 함수이고, F_k는 피처 공간인,

예측 모델의 예측 결과 신뢰도 평가 방법.

청구항 6

각 단계의 적어도 일부가 프로세서에 의해 수행되는, 예측 모델을 통한 예측 결과의 신뢰도를 평가하는 방법으로서,

상기 예측 모델을 통한 예측에 사용되기 위한 입력 데이터를 수신하는 단계;

상기 예측 모델을 훈련하는데 사용된 훈련 데이터를 이용하여 구축된 데이터 분류 알고리즘을 이용하여 상기 입력 데이터를 분류하는 단계; 및

분류된 입력 데이터 그룹마다 서로 다른 평가 신뢰도를 할당하는 단계를 포함하는,

예측 모델의 예측 결과 신뢰도 평가 방법.

청구항 7

제6항에 있어서,

상기 입력 데이터를 분류하는 단계는,

상기 데이터 분류 알고리즘에 대한 커널 변수의 값을 변경시키면서 상기 데이터 분류 알고리즘을 상기 입력 데이터에 적용하여 상기 입력 데이터를 복수의 그룹으로 분류하는 단계를 포함하는,

예측 모델의 예측 결과 신뢰도 평가 방법.

청구항 8

제7항에 있어서,

상기 데이터 분류 알고리즘은 서포트 벡터 데이터 디스크립션(Support Vector Data Description; SVDD)이고 상기 커널 변수는 SVDD의 커널 너비 조절 변수이며, 상기 커널 변수의 값은 점차 증가하도록 설정되거나, 점차 감소하도록 설정되는,

예측 모델의 예측 결과 신뢰도 평가 방법.

청구항 9

제8항에 있어서,

상기 SVDD의 구(sphere)는 하기 수학식 1의 가우시안 커널을 이용하여 생성되고,

[수학식 1]

$$K(x_i, x_j) = \exp\left(-g\|x_i - x_j\|^2\right)$$

여기서, x_i 및 x_j는 데이터 포인트이며,g는 커널 너비 조절 변수인,

예측 모델의 예측 결과 신뢰도 평가 방법.

청구항 10

제9항에 있어서,

상기 SVDD의 목적 함수는 하기 수학식 2와 같으며,

[수학식 2]

$$\min_{R,a,\xi} R^2 + C\sum_{i=1}^{n}\xi_i$$

$$\text{subject to } \| \Phi_k(x_i) - a \|_{F_k}^2 \le R^2 + \xi_i,\ \xi_i \ge 0 \ \text{ for all } i = 1,2,\cdots,n$$

여기서, R은 상기 구의 반경이고, a는 상기 구의 중심이며, ξ_i 는 소프트 마진을 위한 페널티 값이고, C는 트레이드 오프를 제어하기 위한 하이퍼파라미터이고, Φ_k 는 오리지널 데이터를 피쳐 공간에 맵핑하기 위한 함수이고, F_k는 피쳐 공간인,

예측 모델의 예측 결과 신뢰도 평가 방법.

청구항 11

예측 모델을 통한 예측 결과의 신뢰도를 평가하는 장치로서,

메모리; 및

상기 메모리와 연결되고, 상기 메모리에 포함된 컴퓨터 판독가능한 명령들을 실행하도록 구성된 하나의 프로세서를 포함하고,

상기 적어도 하나의 프로세서는,

상기 예측 모델을 훈련하는데 사용된 훈련 데이터를 이용하여 커널 변수의 제1 값을 가지는 제1 데이터 분류 알고리즘을 구축하는 동작;

상기 예측 모델을 통한 예측에 사용되기 위한 입력 데이터를 상기 제1 데이터 분류 알고리즘을 이용하여 제1 외부 데이터 및 제1 내부 데이터로 분류하는 동작;

상기 예측 모델을 훈련하는데 사용된 훈련 데이터를 이용하여 커널 변수의 제2 값을 가지는 제2 데이터 분류 알고리즘을 구축하는 동작;

상기 제1 내부 데이터를 상기 제2 데이터 분류 알고리즘을 이용하여 제2 외부 데이터 및 제2 내부 데이터로 분류하는 동작; 및

상기 예측 모델을 이용하여 상기 제1 외부 데이터에 대해 예측된 예측 결과의 신뢰도를 제1 신뢰도로 결정하고, 상기 예측 모델을 이용하여 상기 제2 외부 데이터에 예측된 예측 결과의 신뢰도를 제2 신뢰도로 결정하는 동작을 수행하도록 설정되는,

예측 모델의 예측 결과 신뢰도 평가 장치.

청구항 12

제11항에 있어서,

상기 제1 및 제2 데이터 분류 알고리즘은 서포트 벡터 데이터 디스크립션(Support Vector Data Description; SVDD)이고 상기 커널 변수는 SVDD의 커널 너비 조절 변수이며, 상기 커널 변수의 상기 제1 값은 상기 제2 값과 상이하도록 설정되는,

예측 모델의 예측 결과 신뢰도 평가 장치.

청구항 13

제12항에 있어서,

상기 커널 변수의 제1 값은 상기 제2 값보다 크도록 설정되며,

상기 제1 신뢰도는 상기 제2 신뢰도보다 큰,

예측 모델의 예측 결과 신뢰도 평가 장치.

청구항 14

제12항에 있어서,

상기 SVDD의 구(sphere)는 하기 수학식 1의 가우시안 커널을 이용하여 생성되고,

[수학식 1]

$$K(x_i, x_j) = \exp\left(-g\|x_i - x_j\|^2\right)$$

여기서, x_i 및 x_j는 데이터 포인트이며, g는 커널 너비 조절 변수인,

예측 모델의 예측 결과 신뢰도 평가 장치.

청구항 15

제14항에 있어서,

상기 SVDD의 목적 함수는 하기 수학식 2와 같으며,

[수학식 2]

$$\min_{R,a,\xi} R^2 + C\sum_{i=1}^{n}\xi_i$$

$$\text{subject to } \| \Phi_k(x_i) - a \|_{F_k}^2 \le R^2 + \xi_i,\ \xi_i \ge 0 \ \text{ for all } i = 1,2,\cdots,n$$

여기서, R은 상기 구의 반경이고, a는 상기 구의 중심이며, ξ_i 는 소프트 마진을 위한 페널티 값이고, C는 트레이드 오프를 제어하기 위한 하이퍼파라미터이고, Φ_k 는 오리지널 데이터를 피쳐 공간에 맵핑하기 위한 함수이고, F_k는 피쳐 공간인,

예측 모델의 예측 결과 신뢰도 평가 장치.

청구항 16

예측 모델을 통한 예측 결과의 신뢰도를 평가하는 장치로서,

메모리; 및

상기 메모리와 연결되고, 상기 메모리에 포함된 컴퓨터 판독가능한 명령들을 실행하도록 구성된 하나의 프로세서를 포함하고,

상기 적어도 하나의 프로세서는,

상기 예측 모델을 통한 예측에 사용되기 위한 입력 데이터를 수신하는 동작;

상기 예측 모델을 훈련하는데 사용된 훈련 데이터를 이용하여 구축된 데이터 분류 알고리즘을 이용하여 상기 입력 데이터를 분류하는 동작; 및

분류된 입력 데이터 그룹마다 서로 다른 평가 신뢰도를 할당하는 동작을 수행하도록 설정되는,

예측 모델의 예측 결과 신뢰도 평가 장치.

청구항 17

제16항에 있어서,

상기 입력 데이터를 분류하는 동작은,

상기 데이터 분류 알고리즘에 대한 커널 변수의 값을 변경시키면서 상기 데이터 분류 알고리즘을 상기 입력 데이터에 적용하여 상기 입력 데이터를 복수의 그룹으로 분류하는 동작을 포함하는,

예측 모델의 예측 결과 신뢰도 평가 장치.

청구항 18

제17항에 있어서,

상기 데이터 분류 알고리즘은 서포트 벡터 데이터 디스크립션(Support Vector Data Description; SVDD)이고 상기 커널 변수는 SVDD의 커널 너비 조절 변수이며, 상기 커널 변수의 값은 점차 증가하도록 설정되거나 점차 감소하도록 설정되는,

예측 모델의 예측 결과 신뢰도 평가 장치.

청구항 19

제18항에 있어서,

상기 SVDD의 구(sphere)는 하기 수학식 1의 가우시안 커널을 이용하여 생성되고,

[수학식 1]

$$K(x_i, x_j) = \exp\left(-g\|x_i - x_j\|^2\right)$$

여기서, x_i 및 x_j는 데이터 포인트이며, g는 커널 너비 조절 변수인,

예측 모델의 예측 결과 신뢰도 평가 장치.

청구항 20

제19항에 있어서,

상기 SVDD의 목적 함수는 하기 수학식 2와 같으며,

$$\min_{R,c,\xi} R^2 + C\sum_{i=1}^{n}\xi_i$$

$$\text{subject to } \|\Phi_k(x_i) - a\|_{F_k}^2 \le R^2 + \xi_i,\ \xi_i \ge 0 \text{ for all } i = 1,2,\cdots,n$$

여기서, R은 상기 구의 반경이고, a는 상기 구의 중심이며, ξ_i 는 소프트 마진을 위한 페널티 값이고, C는 트레이드 오프를 제어하기 위한 하이퍼파라미터이고, Φ_k 는 오리지널 데이터를 피쳐 공간에 맵핑하기 위한 함수이고, F_k는 피쳐 공간인,

예측 모델의 예측 결과 신뢰도 평가 장치.

발명의 설명

기 술 분 야

[0001] 본 발명은, 예측 모델을 통한 예측 결과의 신뢰도를 평가하는 방법 및 장치에 관한 것으로서, 더욱 상세하게는 데이터 분류 알고리즘을 통해 예측 모델에 입력되는 입력 데이터를 분류하여 신뢰도를 평가하는 방법에 관한 것이다.

배 경 기 술

[0003] 이하에서 기술되는 내용은 본 발명의 실시 예와 관련되는 배경 정보를 제공할 목적으로 기재된 것일 뿐이고, 기술되는 내용들이 당연하게 종래기술을 구성하는 것은 아니다.

[0004] 머신러닝은 물리적 법칙을 사용하여 물리적 시스템을 분석하는 대신, 데이터 기반 입력변수인 특성자료와 출력변수인 목표값의 비선형성의 복잡한 상호관계를 데이터 기반으로 파악한 후 새로운 입력변수에 대한 출력값을 예측하는 기법이다.

[0005] 머신러닝의 분야에 속하는 대표적인 기술로서 딥러닝, CNN, RNN-LSTM, GAN 등으로 알려진 인공신경망, 무수히 많은 결정트리로 구성되어 앙상블을 형성하여 해를 구하는 랜덤포레스트, 마진벡터를 최대화하는 과정에서 분류 및 회귀분석이 이루어지는 서포트벡터머신 등이 있다.

[0006] 이러한 머신러닝 기술들은 다양한 분야에 적용되어 데이터기반 분석에 사용되고 있으며, 머신러닝 모델을 학습하는 방법으로는 먼저 수집한 데이터를 학습데이터, 검증데이터, 시험데이터로 나눈 후 모델학습과 검증에 학습데이터와 검증데이터를 사용하여 모델을 구축하고 마지막으로 시험데이터를 사용하여 구축된 모델의 예측성능을 측정한다.

[0007] 이렇게 구축한 모델은 새로운 입력데이터에 대해 결과를 예측하는 작업에 활용되나, 훈련에 사용된 데이터가 아닌 새로운 데이터에 대해 머신러닝 모델이 예측한 예측 결과가 얼마나 정확한지, 얼마나 신뢰할 수 있는지 파악하기 어려운 한계가 있다.

[0008] 전술한 배경기술은 발명자가 본 발명의 도출을 위해 보유하고 있었거나, 본 발명의 도출 과정에서 습득한 기술정보로서, 반드시 본 발명의 출원 전에 일반 공중에게 공개된 공지기술이라 할 수는 없다.

발명의 내용

해결하려는 과제

[0010] 본 발명의 실시 예는 예측 모델을 통한 예측 결과의 신뢰도를 평가하는 방법 및 장치를 제공하는 것이다.

[0011] 본 발명의 실시 예는 머신러닝을 활용하여 새로운 데이터에 대한 예측을 수행할 때 입력 데이터 중 머신러닝 모델 예측의 신뢰도가 높은 데이터와 신뢰도가 낮은 데이터를 머신러닝의 일종인 SVDD(Support Vector Data Description)기법으로 분류하는 방법으로 예측 신뢰도에 따라 입력자료를 분류하는 기능을 제공하고자 한다.

[0012] 또한, 본 발명의 실시 예는 물리검층 중 음파 검층이 미 측정된 지역의 음파 검층을 추정할 수 있는 음파 검층 추정 시스템 및 방법을 제공하는 것이다.

[0013] 또한, 본 발명의 실시 예는 측정 대상지 별로 각각 획득한 물리 검층 데이터에 따라 측정 대상지 별로 음파 검층을 추정할 수 있는 음파 검층 추정 시스템 및 방법을 제공하는 것이다.

[0014] 본 발명의 목적은 이상에서 언급한 과제에 한정되지 않으며, 언급되지 않은 본 발명의 다른 목적 및 장점들은 하기의 설명에 의해서 이해될 수 있고, 본 발명의 실시 예에 의해 보다 분명하게 이해될 것이다. 또한, 본 발명의 목적 및 장점들은 특허 청구 범위에 나타낸 수단 및 그 조합에 의해 실현될 수 있음을 알 수 있을 것이다.

과제의 해결 수단

[0016] 본 발명의 실시 예에 따른 예측 모델을 통한 예측 결과의 신뢰도를 평가하는 방법은, 예측 모델을 훈련하는데 사용된 훈련 데이터를 이용하여 커널 변수의 제1 값을 가지는 제1 데이터 분류 알고리즘을 구축하는 단계, 예측 모델을 통한 예측에 사용되기 위한 입력 데이터를 제1 데이터 분류 알고리즘을 이용하여 제1 외부 데이터 및 제1 내부 데이터로 분류하는 단계, 예측 모델을 훈련하는데 사용된 훈련 데이터를 이용하여 커널 변수의 제2 값을 가지는 제2 데이터 분류 알고리즘을 구축하는 단계, 제1 내부 데이터를 상기 제2 데이터 분류 알고리즘을 이용하여 제2 외부 데이터 및 제2 내부 데이터로 분류하는 단계 및 예측 모델을 이용하여 제1 외부 데이터에 대해 예측된 예측 결과의 신뢰도를 제1 신뢰도로 결정하고, 예측 모델을 이용하여 상기 제2 외부 데이터에 예측된 예측 결과의 신뢰도를 제2 신뢰도로 결정하는 단계를 포함할 수 있다.

[0017] 실시 예에 있어서, 상기 제1 및 제2 데이터 분류 알고리즘은 서포트 벡터 데이터 디스크립션(Support Vector Data Description; SVDD)이고 상기 커널 변수는 SVDD의 커널 너비를 조절하는 변수이며, 상기 커널 변수의 상기 제1 값은 상기 제2 값보다 작을 수 있다.

[0018] 실시 예에 있어서, 상기 제1 신뢰도는 상기 제2 신뢰도보다 작을 수 있다.

[0019] 실시 예에 있어서, 상기 SVDD의 구(sphere)는 하기 수학식 1과 같이 가우시안 커널을 이용하여 생성될 수 있고, 구체적으로 $K(x_i, x_j) = \exp\left(-g\|x_i - x_j\|^2\right)$ 에 의해 생성될 수 있다.

$$\min_{R\in\xi} R^2 + C\sum_{i=1}^{n}\xi_i$$

[0020] 실시 예에서, 상기 SVDD의 목적 함수는 $\text{subject to } \| \Phi_k(x_i) - a \|_{F_k}^2 \le R^2 + \xi_i,\ \xi_i \ge 0 \text{ for all } i = 1,2,\cdots,n$ 에 의해 정의될 수 있다.

[0021] 본 발명의 실시 예에 따른 예측 모델을 통한 예측 결과의 신뢰도를 평가하는 방법은, 상기 예측 모델을 통한 예측에 사용되기 위한 입력 데이터를 수신하고, 상기 예측 모델을 훈련하는데 사용된 훈련 데이터를 이용하여 구축된 데이터 분류 알고리즘을 이용하여 상기 입력 데이터를 분류한 후, 분류된 입력 데이터 그룹마다 서로 다른 평가 신뢰도를 할당하는 과정으로 수행될 수 있다.

[0022] 구체적으로 실시 예에 있어서, 상기 데이터 분류 알고리즘에 대한 커널 변수의 값을 변경시키면서 상기 데이터 분류 알고리즘을 상기 입력 데이터에 적용하여 상기 입력 데이터를 복수의 그룹으로 분류할 수 있다.

[0023] 실시 예에 있어서, 상기 데이터 분류 알고리즘은 서포트 벡터 데이터 디스크립션(Support Vector Data Description; SVDD)이고 상기 커널 변수는 SVDD의 커널 너비를 조절하는 변수이며, 상기 커널 변수의 값은 점차 증가하도록 변경될 수 있다. 예를 들어, 커널 변수는 SVDD의 커널 너비의 역수일 수도 있다.

[0024] 실시 예에 있어서, 상기 SVDD의 구(sphere)는 $K(x_i, x_j) = \exp\left(-g\|x_i - x_j\|^2\right)$ 에 의해 정의될 수 있다.

$$\min_{R\in\xi} R^2 + C\sum_{i=1}^{n}\xi_i$$

[0025] 실시 예에 있어서, 상기 SVDD의 목적 함수는 $\text{subject to } \| \Phi_k(x_i) - a \|_{F_k}^2 \le R^2 + \xi_i,\ \xi_i \ge 0 \text{ for all } i = 1,2,\cdots,n$ 에 의해 정의될 수 있다.

[0026] 본 발명의 실시 예에 따른 메모리 및 상기 메모리와 연결되고, 상기 메모리에 포함된 컴퓨터 판독가능한 명령들을 실행하도록 구성된 하나의 프로세서를 포함하고, 상기 적어도 하나의 프로세서는, 상기 예측 모델을 훈련하는데 사용된 훈련 데이터를 이용하여 커널 변수의 제1 값을 가지는 제1 데이터 분류 알고리즘을 구축하는 동작, 상기 예측 모델을 통한 예측에 사용되기 위한 입력 데이터를 상기 제1 데이터 분류 알고리즘을 이용하여 제1 외부 데이터 및 제1 내부 데이터로 분류하는 동작, 상기 예측 모델을 훈련하는데 사용된 훈련 데이터를 이용하여 커널 변수의 제2 값을 가지는 제2 데이터 분류 알고리즘을 구축하는 동작, 상기 제1 내부 데이터를 상기 제2 데이터 분류 알고리즘을 이용하여 제2 외부 데이터 및 제2 내부 데이터로 분류하는 동작 및 상기 예측 모델을 이용하여 상기 제1 외부 데이터에 대해 예측된 예측 결과의 신뢰도를 제1 신뢰도로 결정하고, 상기 예측 모델을 이용하여 상기 제2 외부 데이터에 예측된 예측 결과의 신뢰도를 제2 신뢰도로 결정하는 동작을 수행하도록 설정될 수 있다.

[0027] 전술한 것 외의 다른 측면, 특징, 및 이점이 이하의 도면, 청구범위 및 발명의 상세한 설명으로부터 명확해질 것이다.

발명의 효과

[0029] 본 발명의 실시 예는 예측 모델에 입력되는 데이터가 완전히 신규한 것으로 예측 모델의 결과에 대한 신뢰도를 알기 어려운 상황에서도 예측 모델을 통한 예측 결과의 신뢰도를 평가하는 방법을 제공할 수 있다.

[0030] 또한, 본 발명의 실시 예는 측정된 시추공의 물리 검층 데이터를 입력 받아 생성된 학습 데이터로 훈련된 음파 검층 추정 모델에 측정한 물리 검층 데이터를 입력하여 물리 검층 데이터를 획득한 대상지와 인접하며, 물리검층 중 음파 검층이 미 측정된 예측 대상지의 음파 검층을 추정할 수 있도록 한다.

[0031] 다수의 측정 대상지 별로 획득한 물리 검층 데이터에 따라 측정 대상지 별로 각각 학습 모델이 생성될 수 있다. 따라서, 측정 대상지와 일정 거리 이내에 위치한 예측 대상지의 음파 검층은 측정 대상지 별로 생성된 학습 모델에 기반하여 추정될 수 있기 때문에 대상지 별로 보다 정확한 음파 검층이 추정될 수 있다.

[0032] 본 발명의 효과는 이상에서 언급된 것들에 한정되지 않으며, 언급되지 아니한 다른 효과들은 아래의 기재로부터 당업자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

[0034] 도 1은 본 발명의 실시 예에 따른 음파 검층 추정 시스템 및 음파 검층 추정 시스템의 통신을 개략적으로 나타낸 도면이다.

도 2는 본 발명의 실시 예에 따른 음파 검층 추정 시스템의 구성을 나타내는 블록도이다.

도 3은 본 발명의 실시 예에 따른 음파 검층 추정 시스템과 통신 연결된 서버의 구성을 나타낸 블록도이다.

도 4는 본 발명의 실시 예에 따라 측정 대상지에서 획득한 물리 검층 데이터를 기반으로 음파 검층이 미 측정된 예측 대상지의 음파 검층을 예측한 실시 예를 도시한 도면이다.

도 5는 본 발명의 실시 예에 따른 음파 검층 추정 방법을 도시한 흐름도이다.

도 6은 본 발명의 실시 예에 따른 예측 모델을 통한 예측 결과의 신뢰도를 평가하는 방법을 도시한 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0035] 이하, 첨부된 도면을 참조하여 본 명세서에 발명된 실시 예를 상세히 설명하되, 도면 부호에 관계없이 동일하거나 유사한 구성요소는 동일한 참조 번호를 부여하고 이에 대한 중복되는 설명은 생략하기로 한다. 이하의 설명에서 사용되는 구성요소에 대한 접미사 "모듈" 및 "부"는 명세서 작성의 용이함만이 고려되어 부여되거나 혼용되는 것으로서, 그 자체로 서로 구별되는 의미 또는 역할을 갖는 것은 아니다. 또한, 본 명세서에 발명된 실시 예를 설명함에 있어서 관련된 공지 기술에 대한 구체적인 설명이 본 명세서에 발명된 실시 예의 요지를 흐릴 수 있다고 판단되는 경우 그 상세한 설명을 생략한다. 또한, 첨부된 도면은 본 명세서에 발명된 실시 예를 쉽게 이해할 수 있도록 하기 위한 것일 뿐, 첨부된 도면에 의해 본 명세서에 발명된 기술적 사상이 제한되지 않으며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다.

[0036] 제1, 제2 등과 같이 서수를 포함하는 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 상기 구성요소들은 상기 용어들에 의해 한정되지는 않는다. 상기 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다.

[0037] 어떤 구성요소가 다른 구성요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성요소가 다른 구성요소에 "직접 연결되어" 있다거나 "직접 접속되어" 있다고 언급된 때에는, 중간에 다른 구성요소가 존재하지 않는 것으로 이해되어야 할 것이다.

[0038] 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수개의 표현을 포함한다.

[0039] 본 출원에서, "포함한다" 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.

[0040] 도 1은 본 발명의 실시 예에 따른 음파 검층 추정 시스템 및 음파 검층 추정 시스템의 통신을 개략적으로 나타낸 도면이다.

[0041] 도 1을 참고하면, 음파 검층 추정 시스템(10)은 음파 검층을 추정하기 위해 물리 검층(예를 들어, 음파 검층, 감마 검층, 비저항 검층 등 다양한 검층을 포함)이 이루어지는 측정 대상지에서 획득한 물리 검층 데이터에 기반하여 음파 검층을 추정할 수 있도록 훈련된 머신 러닝 기반의 음파 검층 추정 모델이 저장된 메모리를 포함하거나 이러한 음파 검층 추정 모델이 저장된 서버와 통신할 수 있다. 예를 들어, 음파 검층 추정 시스템(10)은 훈련된 음파 검층 추정 모델을 사용하여 측정 대상지 주변의 음파 검층을 제외한 다른 종류의 물리 검층이 이루어진 예측 대상지의 음파 검층을 추정할 수 있다.

[0042] 본 발명의 실시예에서 실시하는 물리 검층은 음파 검층, 감마 검층, 비저항 검층 등 중 적어도 어느 하나를 포함할 수 있으며, 음파 검층을 추정하기 위해 측정 대상지에서 실행하는 물리 검층의 종류에 의해 본 발명이 제한되는 것은 아니다.

[0043] 본 발명의 실시 예들에 있어서 음파 검층 추정 시스템(10)은 시추공에 투입하는 물리 검층 장치(200) 등에 구현

될 수 있다. 또는 물리 검층 장치(200)와 유무선 통신으로 연결된 서버, 전자장치(예: 개인 단말, 서버 장치 등)으로도 구현될 수 있다.

[0044] 실시 예에서 음파 검층 추정 시스템(10)은 물리 검층 장치(200)를 통해 측정 대상지에서 물리 검층 데이터를 획득하고, 획득한 물리 검층 데이터를 입력으로 하여 물리 검층이 행해진 대상지 주변의 인근 지역 시추공의 물리 검층 중 음파검층이 미측정된 대상지(이하: 예측 대상지)의 음파 검층을 출력할 수 있도록 한다.

[0045] 한편, 본 발명의 실시 예에 있어서, 음파 검층 추정 시스템(10)은 설명한 바와 같이 물리 검층 장치(200), 물리 검층 장치(200)와 유무선으로 통신 연결된 단말, 서버 등 중 어느 하나에 구현될 수 있다.

[0046] 구체적으로는 물리 검층 장치(200)에서 측정한 측정 대상지의 물리 검층의 데이터가 서버(100)로 전송되고, 전송된 물리 검층 데이터(예를 들어, 음파 검층 데이터, 감마 검층 데이터, 비저항 검층 데이터 등을 포함)를 사용하여 음파 검층 추정 머신 러닝 기반의 모델을 훈련하여 측정 대상지 주변의 시추공에서 물리 검층이 이루어졌으나 음파 검층이 미 측정된 경우(예측 대상지)에 대하여 음파 검층 데이터를 추정할 수 있다.

[0047] 다른 실시 예에서 음파 검층 추정 시스템(10)은 물리 검층 장치(200)에서 구현될 수도 있다. 예를 들어 물리 검층 장치(200)에서는 측정 대상지에서 측정한 물리 검층을 머신 러닝 기반 모델의 훈련에 사용하여 측정 대상지와 인접한 예측 대상지의 음파 검층을 추정할 수 있는 구동 서버(미도시)를 통해 예측 대상지의 음파 검층을 추정할 수도 있다.

[0048] 이때, 구동 서버에 저장된 음파 검층 추정 모델은 물리 검층 장치 (200)에서 다운로드된 모델을 훈련한 학습모델이거나 서버(100)로부터 전송된 훈련 데이터를 이용하여 서버(100)에서 훈련된 학습 모델을 다운로드 한 것일 수도 있다.

[0050] 도 2는 본 발명의 실시 예에 따른 음파 검층 추정 시스템의 구성을 나타내는 블록도이고, 도 3은 본 발명의 실시 예에 따른 음파 검층 추정 시스템과 통신 연결된 서버의 구성을 나타낸 블록도이다.

[0051] 도면의 설명에 앞서 설명의 편의상 음파 검층 추정 시스템(10)은 서버(100)에서 구현된 예를 들어 설명하기로 하며, 물리 검층 장치(200)은 서버(100)와 통신 연결되어 물리 검층 장치(200)에서 측정한 측정 대상지의 물리 검층 데이터를 서버(100)로 전송하는 예를 들어 설명하기로 한다.

[0052] 도 2를 참고하면, 음파 검층 추정 시스템(10)은 물리 검층 장치(200)에서 측정한 측정 대상지의 물리 검층 데이터를 기반으로 측정 대상지와 인접하고, 물리 검층이 시행되었으나 음파 검층이 미측정된 예측 대상지의 음파 검층을 추정할 수 있는 구성이다. 이러한 음파 검층 추정 시스템(10)은 통신부(120), 메모리(160), 러닝 프로세서(150) 및 프로세서(180) 등을 포함하여 구성될 수 있다.

[0053] 음파 검층 추정 시스템(10)의 통신부(120)는 물리 검층 장치(200)와 통신 연결될 수 있는 구성이고, 물리 검층 장치(200)로 습득한 측정 대상지의 물리 검층 데이터를 송신할 수 있다.

[0054] 메모리(160)는 모델 저장부(161)를 포함할 수 있다. 모델 저장부(161)는 러닝 프로세서(150)를 통해 학습 전 또는 학습된 모델(161a)을 저장할 수 있다. 학습된 모델 또는 학습 전 모델을 포함하는 학습 모델은 인공 신경망 기반, 포레스트 모델 기반 등 특별히 그 종류를 한정하지 않는다.

[0055] 러닝 프로세서(150)는 학습 데이터를 이용하여 학습 모델(161a)을 학습시킬 수 있다. 학습 모델은 서버(100)나 물리 검층 장치(200)에 탑재된 상태에서 이용될 수 있다.

[0056] 학습 모델은 하드웨어, 소프트웨어 또는 하드웨어와 소프트웨어의 조합으로 구현될 수 있다. 학습 모델의 일부 또는 전부가 소프트웨어로 구현되는 경우 학습 모델을 구성하는 하나 이상의 명령어(instruction)는 메모리(160)에 저장될 수 있다.

[0057] 본 발명의 학습 모델은 머신 러닝을 기반으로 하는 음파 검층 추정 모델이며, 음파 검층 추정 모델은 물리 검층 장치(200)에서 측정한 측정 대상지의 물리 검층 데이터를 입력 받은 물리 검층 결과값으로 레이블링(labeling)한 훈련 데이터로 훈련될 수 있다. 즉, 음파 검층 추정 모델은 측정 대상지마다 각각 다른 물리 검층 데이터를 레이블링한 훈련 데이터로 훈련된 음파 검층 추정 모델을 이용함으로써, 지층의 구조 및 지층마다 석유 존재 여부 등을 보다 정확하게 추정할 수 있다.

[0058] 구체적으로 음파 검층 추정 모델에 적용하는 측정 대상지의 물리 검층 데이터는 측정 대상지의 지층에 삽입되는

시추 내 검증기를 통해 측정 대상지의 지층을 정량적으로 분석하기 위한 정보와 같은 데이터일 수 있다. 구체적으로 물리 검층은 지층 내에 석유가 있는지 여부, 석유가 있다면 얼마나 있는지를 보다 정확하게 알아내기 위한 작업을 의미할 수 있다.

[0059] 예를 들어, 시추에 의해 굴착한 시추공 내에 측정 장비를 투입하여 지층 내에서 자연적으로 발생하는 물리적인 양이나 측정 장비로부터 인공적으로 물질(예: 음파, 중성자, 전자, 감마선 등)을 지층에 발생시킬 수 있다. 특히, 지층으로 발생된 인공 물질을 지층을 통과한 후 다시 측정 장비로 돌아오게 되는데, 측정 장비로 돌아온 인공 물질의 양을 지층 깊이에 따라 연속적으로 측정할 수 있다. 이렇게 측정된 인공 물질의 양에 따라 측정 대상지 인근의 암석 성질, 지층의 석유 함유량, 시추공 벽의 상태 등을 조사할 수 있다.

[0060] 측정 대상지에서 물리 검층 데이터를 획득하면 획득한 측정 대상지의 물리 검층은 기 설정된 음파 검층 추정 모델에 입력될 수 있다. 이때, 음파 검층 추정 모델은 측정 대상지에서 측정한 물리 검층 데이터에 따라 레이블링된 훈련 데이터로 훈련된다.

[0061] 구체적으로, 훈련 데이터는 앞서 설명한 물리 검층 데이터와 같이 측정 대상지에서 물리 검층 과정에서 획득한 물리 검층 데이터를 기초로 생성할 수 있다.

[0062] 즉, 본 발명의 훈련 데이터는 기 설정된 지층 깊이 별로 시추된 시추공에 삽입된 측정 장비로 되돌아오는 인공(예: 음파, 중성자, 전자, 감마선 등) 또는/및 자연 물질의 양을 측정한 데이터를 기초로 생성할 수 있다.

[0063] 한편, 본 발명의 실시 예에 따른 음파 검층 추정 모델은 랜덤 포레스트 모델에 기반하고 음파 검층 추정 모델에 포함된 복수의 트리(tree)에 입력되어 각 트리로부터 획득된 값들 중 최대 빈도값(Majority-Voting)을 출력하도록 훈련된 모델일 수 있다.

[0064] 예를 들어, 측정 대상지에서 획득한 물리 검층 데이터(예: Volve 유전 내의 시추정 15/9-F-1A, 15/9-F-1B, 15/9-F-11A)가 레이블링된 훈련 데이터를 랜덤 샘플링하여 복수의 미리 설정되거나 훈련 데이터에 따른 깊이(depth)를 가지는 트리를 생성하도록 훈련될 수 있다. 훈련된 랜덤 포레스트 모델에 기반한 학습 모델에 예측 대상지(음파 검층이 미측정된 시추공)에서 획득한 물리 검층 데이터를 입력하고 복수의 트리에서 결정(출력)되는 응답값 중에서 최대 빈도값(Majority-Voting)을 최종 응답값(출력값)으로 결정하여 예측 대상지의 음파 검층을 추정할 수 있다.

[0065] 실시 예에서, 포레스트 모델에 기반하여 훈련된 학습 모델을 기반으로 음파 검층을 추정하기 앞서 이때, 음파 검층 추정 모델의 정확성을 위해 예측 신뢰도를 평가할 수 있다. 랜덤 포레스트 학습 모델의 예측 신뢰도 평가는 예측오차인 RMSE(root mean squared error)를 이용할 수 있으며, RMSE는 다음의 [수학식 1]로 정의될 수 있다.

[0066] [수학식 1]

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)}$$

[0067]

[0068] 여기서, y 는 참값(true value)이고, $\hat{y}$ 는 모델에서 추정한 값(predicted value) 이며, n은 데이터의 갯수일 수 있다. [수학식 1]에 의해, RMSE가 작을수록 예측 신뢰도는 커진다는 것을 의미할 수 있다.

[0069] 이와 같이 측정한 학습 모델의 예측 신뢰도는 훈련된 모델이 입력한 데이터에 대해 예측을 잘 하는지를 추정할 수 있다.

[0070] 이때, 학습 모델의 정확성을 위해 다양한 측정 대상지에서 물리 검층 데이터를 확보하여 훈련 데이터를 생성할 수 있도록 할 수 있다. 예를 들어 물리 검층 데이터에 포함된 음파 검층을 위한 데이터(DT_ sonic log 및 DTS_ shear sonic log)를 훈련 데이터에 입력하여 음파 검층이 이루어지지 않는 측정 대상지와 인접한 예측 대상지의 물리 검층 추정을 수행하다가 음파 검층 데이터를 수집하지 않는 예측 대상지에서 음파 검층 데이터를 수집하는 경우, 음파 검층 예측 대상지를 변경하고 훈련 데이터에 수집한 음파 검층 데이터를 추가하여 다시 학습 모델을 훈련할 수도 있다.

[0071] 프로세서(180)는 학습 모델을 이용하여 새로운 입력 데이터에 대하여 결과 값을 추론하고, 추론한 결과 값에 기초한 응답이나 제어 명령을 생성할 수 있다.

[0072] 또한, 프로세서(180)는 통신부(120)를 통해 물리 검층 장치(200)로부터 수신한 지층의 물리 검층 데이터에 학습 모델을 적용하여 물리 검층 데이터를 획득한 지층과 인접하며 음파 검층이 이루어지지 않은 인접 예측 대상지의 음파 검층을 추정할 수 있다.

[0073] 도 3을 참고하면, 물리 검층 장치(200)는 통신부(220), 검출부(240), 메모리(260) 및 프로세서(280)를 포함하여 구성될 수 있다.

[0074] 통신부(220)는 음파 검층 추정을 위한 서버(100)와 통신 연결될 수 있다. 구체적으로 통신부(220)는 물리 검층 장치(200)에서 획득한 지층의 물리 검층 데이터를 서버(100)로 송신하도록 한다.

[0075] 검출부(240)는 지층의 특징, 시추공 내부의 정보 등을 검출하는 구성이다. 구체적으로, 검출부(240)는 시추에 의해 굴착한 시추공 내에 측정 장비를 투입하여 지층 내에서 자연적으로 발생하는 물리적인 양이나 측정 장비로부터 인공적으로 물질(예: 음파, 중성자, 전자, 감마선 등)이 지층에서부터 측정 장비로 돌아올 때, 지층의 깊이에 따라 측정되는 인공 물질 양 변화를 검출할 수 있다.

[0076] 이때, 검출부(240)는 일정한 깊이 간격마다 인공 물질 및/또는 지층 내 자연 물질을 검출할 수 있으며, 수집된 물질 정보를 서버(100)로 송신할 수 있다. 송신된 물질 정보 양 변화를 기초로 지층 및 지층 인근의 암석 성질, 지층의 석유 함유량, 시추공 벽의 상태 등에 대한 정보를 검출할 수 있다.

[0077] 메모리(260)는 검출부(240)에서 검출한 측정 대상지의 물리 검층 데이터를 통해 측정 대상지와 인접하고, 음파 검층 데이터를 획득하지 않은 예측 대상지의 음파 검층을 추정하고자 하는 조건을 저장할 수 있다. 즉, 메모리(260)는 음파 검층이 이루어지지 않은 예측 대상지의 음파 검층을 추정할 수 있도록 측정 대상지에서 물리 검층이 수행될 수 있는 검출부(240) 동작을 위한 데이터들, 명령어들이 저장될 수 있다.

[0078] 프로세서(280)는 메모리(260)에 저장된 응용 프로그램과 관련된 동작 이외에 물리 검층 장치(200)의 전반적인 동작을 제어할 수 있다. 프로세서(280)는 앞서 살펴본 구성요소들을 통해 입력 또는 출력되는 신호, 데이터, 정보 등을 처리하거나 메모리(260)에 저장된 응용 프로그램을 구동하도록 하여 예측 대상지의 음파 검층 추정이 이루어질 수 있도록 한다.

[0080] 도 4는 본 발명의 실시 예에 따라 측정 대상지에서 획득한 물리 검층 데이터를 기반으로 음파 검층이 미 측정된 예측 대상지의 음파 검층을 예측한 실시 예를 도시한 도면이다.

[0081] 도면의 설명에 앞서, 본 발명의 실시 예의 학습 모델은 랜덤 포레스트 예측 모델을 기반으로 하고 있으며, 음파 검층이 미측정된 예측 대상지에서 획득한 물리 검층 데이터(예를 들어, 감마 검층 데이터, 비저항 검층 데이터, 밀도 검층 데이터 등을 포함)에 따른 랜덤 포레스트의 예측 신뢰도를 평가하기 위해 SVDD(support vector data description)를 이용할 수 있다.

[0082] 구체적으로, SVDD는 탐색 공간(feature space)에서 대부분의 정상 데이터를 둘러싸는 가장 작은 구(hypersphere)의 경계를 찾는 기법으로 새로운 데이터로부터 정상 데이터와 비정상 데이터를 분류하는데 사용될 수 있다. 본 발명의 실시 예에서는 음파 검층이 미측정된 예측 대상지의 물리 검층 데이터에 SVDD의 하이퍼파라미터 중 하나를 변화시켜 학습 모델의 음파 검층 예측 신뢰도를 평가하는 방식을 예를 들어 설명하기로 한다.

[0083] 여기서 SVDD의 하이퍼파라미터에 의해 학습 데이터의 중심에서부터 구(hypersphere)의 경계 사이 거리가 달라지고, 하이퍼파라미터 값이 클수록 학습 데이터에 가까운 구 경계가 설정될 수 있다. 이러한 특징을 기준으로 하이퍼파라미터 값을 증가시키면서 또는 감소시키면서 학습 데이터 경계를 설정할 수 있다.

[0084] 구체적으로, 학습 데이터 경계를 설정하는 과정에서, 음파검층을 제외한 감마, 중성자, 밀도, 비저항 검층 등 입력데이터로 사용되는 물리검층 자료를 좌표계에 나타낼 수 있다. 이후, SVDD 목적 함수의 Φ_k (수학식 3 참고, 오리지널 데이터를 피쳐 공간에 맵핑하기 위한 함수) 변환을 통해 입력 데이터 공간을 피쳐 공간(F_k)으로 변환할 수 있다. 입력 데이터 공간이 피쳐 공간으로 변환되면서 구 형태의 공간으로 변환될 수 있으며, 구의 반경(R)은 커널 너비 조절 변수인 g에 따라 결정될 수 있다. 이후, 피쳐 공간을 다시 물리 검층 입력 데이터 공간으로 역 변환하면서 내부 데이터 및 외부 데이터를 분리하는 경계가 생성될 수 있다. 여기서, 경계 내의 내부

데이터란 g 값에 따라서 설정되는 경계의 내부에 있는 데이터로 신뢰도 분석을 통해 예측모델의 입력자료로 사용되는 물리검층 데이터 중 상대적으로 신뢰도가 높은 데이터를 나타내고, 경계 바깥의 외부 데이터란 g 값에 따라 설정되는 경계의 외부에 있는 데이터로 상대적으로 신뢰도가 낮을 것으로 예상되는 데이터를 나타낼 수 있다. 이러한 과정을 기초로 본 발명의 실시 예는 하이퍼파라미터미터가 1, 3, 5로 증가하고, 학습 데이터가 15/9-F-1B 이고, 15/9-F-1A의 음파 검층을 예측하는 예를 들어 설명하기로 한다. 다만, 하이퍼파라미터는 5, 3, 1의 순서로 감소하는 실시예도 가능함은 물론이다.

표 1

[0086]

	g=1	g=3	g=5
학습데이터(15/9-F-1B)에 의해 형성되는 구 경계 내부에 위치하는 예측데이터(15/9-F-1A)의 비율 (= 경계 내부 데이터의 비율)	47.17 %	27.23 %	16.35 %
학습데이터(15/9-F-1B)에 의해 형성되는 구 경계 외부에 위치하는 예측데이터(15/9-F-1A)의 비율 (= 경계 외부 데이터의 비율)	52.83 %	72.77 %	83.65 %
(1)예측 데이터(15/9-F-1A)에 대한 학습 모델의 총 예측 오차 (total RMSE)	0.3973		
(2) 경계 내부 데이터의 RMSE	0.2491	0.2033	0.2004
(3) 경계 외부 데이터의 RMSE	0.4933	0.4488	0.4252
(1)과 (2)의 상대 차이 비율(relative difference) (= ((2)-(1))/(1)*100)	-37.30%	-48.83%	-49.56%
(1)과 (3)의 상대 차이 비율(relative difference) (= ((3)-(1))/(1)*100)	24.16%	12.96%	7.02%

[0087] 예시적으로 [표 1]를 참고하면, 하이퍼파라미터(g)를 1, 3, 5로 증가할 때 설정되는 구 경계 내부로 포함되는 데이터의 비율은 점점 감소하고, 경계 외부에 해당하는 데이터 비율은 증가하는 것으로 나타났다.

[0088] 즉, 전체 예측오차(total RMSE) 대비 경계 내부 데이터의 예측오차는 더 작으며, 경계 외부 데이터의 예측오차는 더 크게 나타났다. 또한, 하이퍼파라미터의 증가에 따라 경계 내부 데이터와 경계 외부 데이터의 예측오차는 모두 점차 감소하는 것으로 나타났다.

[0089] 즉, 하이퍼파라미터 값이 증가함에 따라 구 경계가 축소되기 때문에 경계 내부 데이터에 포함되어 있던 데이터들이 경계 외부 데이터로 이동하면서 경계 외부 데이터의 오차가 낮아지는 것으로 판단할 수 있다.

[0090] 이와 다르게, 하이퍼파라미터(g)를 5, 3, 1로 설정하여 신뢰도를 평가할 수도 있다. 이 경우, 구 경계 내부로 포함되는 데이터의 비율은 점점 증가할 수 있고, 경계 외부에 해당하는 데이터 비율은 감소할 수 있다.

[0091] 이와 유사하게, 15/9-F-1A와 15/9-F-1B를 학습자료로 사용하여 15/9-F-11A의 음파 검층을 추정할 수도 있다. 하기 [표 2]를 참고하면, 이 하이퍼파라미터(g)의 변화에 따라 경계 내부 데이터 비율이 감소할수록 경계 외부 데이터 비율은 증가하는 것으로 나타났다.

표 2

[0093]

	g=1	g=3	g=5
학습데이터(15/9-F-1A와 15/9-F-1B)에 의해 형성되는 구 경계 내부에 위치하는 예측데이터(15/9-F-11A)의 비율 (= 경계 내부 데이터의 비율)	77.51%	63.80%	50.93%
학습데이터(15/9-F-1A와 15/9-F-1B)에 의해 형성되는 구 경계 외부에 위치하는 예측데이터(15/9-F-11A)의 비율 (= 경계 외부 데이터의 비율)	22.49%	36.20%	49.07%
(1)예측 데이터(15/9-F-11A)에 대한 학습 모델의 총 예측 오차 (total RMSE)	0.220801182		
(2) 경계 내부 데이터의 RMSE	0.2027	0.1916	0.1772
(3) 경계 외부 데이터의 RMSE	0.2649	0.2604	0.2539
(1)과 (2)의 상대 차이 비율(relative difference) (= ((2)-(1))/(1)*100)	-7.10%	-12.19%	-18.79%

(1)과 (3)의 상대 차이 비율(relative difference) (= ((3)-(1))/(1)*100)	21.40%	19.34%	16.36%

[0095] [표 1] 및 [표 2]를 비교하면 15/9-F-1B에서 획득한 물리 검증 데이터를 학습하고, 15/9-F-1A의 음파 검증을 추정하는 추정 결과보다 두 개의15/9-F-1A와 15/9-F-1B 물리 검증 데이터를 학습하여 15/9-F-11A의 음파 검증을 추정할 때, 예측한 결과의 신뢰도가 더 높다는 것을 확인할 수 있다.

[0097] 이러한 특징을 기반으로 하이퍼파라미터의 변화 및 학습 데이터에 따라 음파 검증 추정 결과값은 하기의 [표 3]과 같이 도시될 수 있다.

표 3

[0099]

	g=1		g=3		g=5	
	경계 내부 데이터	경계 외부 데이터	경계 내부 데이터	경계 외부 데이터	경계 내부 데이터	경계 외부 데이터
	상대차이비율	상대차이비율	상대차이비율	상대차이비율	상대차이비율	상대차이비율
1A_1B[1)]	-29.97%	54.42%	-37.12%	34.88%	-40.39%	25.58%
1A_11A	-16.60%	38.09%	-26.38%	27.20%	-29.53%	20.74%
1B_1A	-37.30%	24.16%	-48.83%	12.96%	-49.56%	7.02%
1B_11A	-56.95%	17.77%	-64.26%	6.98%	-66.51%	3.66%
11A_1A	-11.96%	34.33%	-17.57%	11.14%	-25.09%	7.93%
11A_1B	3.66%	-13.75%	-18.64%	7.41%	-26.64%	5.79%
1A+1B_11A[2)]	-7.10%	21.40%	-12.19%	19.34%	-18.79%	16.36%
1A+11A_1B	-2.10%	22.98%	-4.21%	17.80%	-6.23%	12.96%
1B+11A_1A	-17.17%	65.20%	-17.94%	22.65%	-20.31%	13.59%
평균 상대오차	-19.50%	29.40%	-27.46%	17.82%	-31.45%	12.63%

[0100] 1) 1A_1B: 15/9-F-1A를 학습한 모델로 15/9-F-1B의 음파 검증 예측

[0101] 2) 1A+1B_11A: 15/9-F-1A와 15/9-1B를 학습한 모델로 15/9-F-11A의 음파 검증 예측

[0103] 구체적으로 도 4의 (a) 및 (b)를 참고하면 15/9-F-1B의 학습 자료에 대한 RMSE(root mean squared error)는 0.0792이고, 이에 대한 검증 결과의 RMSE는 0.1228로 검출되었다.

[0104] 이때, 학습 모델을 구축하기 위한 입력되는 물리 검증 자료로는 중성자, 밀도, 감마, 비저항, 광전자 검증이며, 출력 자료는 음파 검증이다. 설명한 바와 같이, 학습 모델을 구축하기 위해서는 모델의 입력자료로 중성자, 밀도, 감마, 비저항, 광전자 검증 등의 자료가 사용되며, 출력자료로 음파 검증이 레이블로 사용될 수 있다. 즉, 모델을 학습시킨다는 것은 중성자, 밀도, 감마, 비저항, 광전자 검증 값을 입력자료로 넣을 때 출력으로 음파 검증을 예측하도록 훈련한다는 것을 의미할 수 있다.

[0105] 구축된 학습 모델에 15/9-F-1A의 음파 검증을 제외한 검증 자료를 입력하여 15/9-F-1A의 음파 검증을 예측할 수 있으며, 도 4의 (c)에 도시된 바와 같이 15/9-F-1A의 음파검증에 대한 예측 오차는 RMSE는 0.3973으로 추정할 수 있다.

[0106] 이와 같이, volve 유전의 시추에 의해 측정한 물리 검증 데이터를 이용하여 음파 검증이 미측정된 예측 대상지의 음파 검증을 추정할 수 있다.

[0107] 이때, 학습 모델의 예측 신뢰도를 평가하기 위해 모델의 입력자료로 사용되는 물리 검증 데이터에 대해 SVDD를 이용할 수 있으며, SVDD의 하이퍼파라미터(g) 값을 변화시켜 내부 경계 데이터와 외부 경계 데이터 오차를 분석

할 수 있다. 전체 데이터에 대해 내부 경계 데이터의 오차는 하이퍼파라미터가 증가할수록 감소하는 것으로 나타난다. 따라서, 학습 데이터를 기준으로 구 경계를 가깝게 설정하여 경계 내부에 포함된 데이터를 예측할 때, 보다 높은 신뢰도를 도출할 수 있다. 이러한 특징에 기반하여 음파 검층 데이터가 미 측정된 예측 대상지의 음파 검층 추정 오차가 낮은 데이터와 높은 데이터를 분류할 수 있다.

[0109] 도 5는 본 발명의 실시 예에 따른 음파 검층 추정 방법을 도시한 흐름도이다.

[0110] 도 5를 참고하면, 본 발명의 실시 예에 따른 음파 검층 추정 방법은 물리 검층 장치(200)를 통해 깊이 변화에 따라 지층의 물리적 특징이 포함된 물리 검층 데이터를 획득할 수 있다(S110).

[0111] 획득한 물리 검층 데이터는 레이블링된 훈련 데이터를 랜덤 샘플링하여 복수의 미리 설정되거나 훈련 데이터에 따른 깊이(depth)를 가지는 트리를 생성하고, 생성된 학습 데이터로 학습 모델을 훈련할 수 있다.

[0112] 구체적으로 예측 대상지의 음파 검층 추정을 위해, 측정 대상지의 시추공에서 음파 검층을 포함한 물리 검층 데이터를 수집할 수 있다. 수집된 물리 검층 데이터를 훈련된 랜덤 포레스트 모델에 기반한 학습 모델에 입력할 수 있다. 물리 검층 데이터가 입력된 학습 모델은 복수의 트리에서 결정(출력)되는 응답값 중에서 최대 빈도값(Majority-Voting)을 최종 응답값(출력값)으로 결정하여 음파 검층 이외의 물리 검층이 이루어진 측정 대상지와 인접한 예측 대상지의 음파 검층을 추정할 수 있다.

[0113] 음파 검층을 예측하려는 예측 대상지의 예측 신뢰도를 분석할 수 있다. 신뢰도 분석은 SVDD(support vector data description)를 이용할 수 있으며, 예를 들어 SVDD의 하이퍼파라미터 중 하나를 변화시키며 물리 검층 데이터의 신뢰도를 평가할 수 있다. 평가된 신뢰도를 기반으로 물리 검층 자료 중 음파 검층이 포함되어 있지 않은 예측 대상지의 예측 대상지의 음파 검층을 추정할 수 있다.

[0114] 이러한 신뢰도 분석을 통해 예측모델의 입력자료로 사용되는 물리검층을 신뢰도가 높은 데이터(경계 내부 데이터)와 신뢰도가 낮을 것으로 예상되는 데이터(경계 외부 데이터)로 구분할 수 있다.

[0115] 이와 같이, 시추의 물리 검층 데이터를 입력 받아 생성된 학습 데이터로 훈련된 음파 검층 추정 모델에 측정한 물리 검층 데이터를 입력하여 물리 검층 데이터를 획득한 대상지와 인접하며, 음파 측정이 미 측정된 예측 대상지의 음파 검층을 추정할 수 있도록 한다.

[0116] 도 6은 본 발명의 실시 예에 따른 예측 모델을 통한 예측 결과의 신뢰도를 평가하는 방법을 도시한 흐름도이다.

[0117] 하기에서 설명되는 예측 모델을 통한 예측 결과의 신뢰도를 평가하는 방법은 각 단계의 적어도 일부가 프로세서에 의해 수행될 수 있다.

[0118] 본 발명의 실시 예는 도 6의 알고리즘과 같이 SVDD 기법을 적용하여 예측에 사용하고자 하는 데이터를 내부 데이터와 외부 데이터로 분류한 후, 내부 데이터는 모델 예측시 신뢰도가 높은 데이터로 분류하고 외부 데이터는 모델 예측시 신뢰도가 낮은 데이터 즉, 예측오차가 클 것으로 예상하는 데이터로 분류할 수 있다.

[0119] 여기서, SVDD의 커널변수 g에 여러값을 사용하면 예측 신뢰도의 정도에 따라 데이터를 여러 그룹으로 분류할 수 있다.

[0120] 본 발명의 방법이 시작되면(S200), 예측 모델을 훈련시키기 위한 학습 데이터, 검증 데이터, 시험 데이터가 준비될 수 있고, 머신러닝 모델이 학습 데이터를 통해 훈련되고, 검증 데이터로 검증된 후 시험 데이터에 의해 예측 성능이 평가될 수 있다(S210).

[0121] 여기서, 학습 데이터, 검증 데이터, 시험 데이터는 예측을 위해 입력되는 데이터와 입력된 데이터에 따라 출력되는 예측 데이터(타겟 데이터)를 포함할 수 있다.

[0122] 예를 들어, 하기의 표와 같이 입력되는 데이터가 NPHI, RHOB, GR, RT 및 PEF일 수 있고 이러한 데이터들에 따라 예측 대상이 되는 데이터는 DT일 수 있다.

표 4

[0124]

Input data	NPHI	중성자검층
	RHOB	밀도검층
	GR	감마검층
	RT	비저항검층
	PEF	광전자검층
Target	DT	음파검층

[0126] 학습 데이터에 의해 훈련되어 도출된 예측 모델은 측정된 값인 NPHI, RHOB, GR, RT 및 PEF를 입력받아 DT를 예측하는 모델일 수 있다.

[0127] 이러한 예측 모델은 일반적으로 입력 자료(입력 데이터로도 지칭)를 사용하여 목표 자료(타겟 데이터로도 지칭)를 예측하는 머신러닝 모델일 수 있다.

[0128] 입력 데이터와 타겟 데이터가 모두 존재하는 데이터는 모델학습을 위해 학습 데이터, 검증 데이터, 시험 데이터로 랜덤하게 분리될 수 있다. 이후, 머신러닝 모델을 위해 인공신경망, 랜덤 포레스트, 서포트 벡터머신 등 다양한 방법 중 하나를 선택해서 입력 데이터와 타겟 데이터를 위한 모델이 구성될 수 있다.

[0129] 초기 조건으로 구성된 모델에 학습 데이터가 입력된 후 출력부에서 나오는 결과와 타겟 데이터 사이의 오차가 감소되는 방향으로 모델이 학습될 수 있다.

[0130] 검증 데이터는 학습된 모델이 과적합되었는지 판단하는데 사용되며 과적합되지 않은 것이 확인되면 위의 과정을 반복하여 오차를 최대한 줄이도록 한다.

[0131] 모델에 대한 학습이 종료되면 시험 데이터를 사용하여 모델의 성능을 확인할 수 있다.

[0132] 모델의 학습이 완료되면 미측정된 영역의 타겟 데이터를 예측하는데 사용될 수 있다. 미측정된 영역이란 입력 데이터에 해당하는 검층자료들은 존재하지만 타겟 데이터에 해당하는 특정 검층자료는 여러 이유로 측정되지 못한 상태이다.

[0133] 측정된 검층자료를 머신러닝 모델에 입력하여 출력값이 산출되면 이 출력값이 바로 미측정 검층의 예측값이 되나, 참값을 알 수는 없으므로 예측된 타겟 값이 실제 값과 얼마나 오차가 큰지는 알 수 없다.

[0134] 훈련된 예측 모델에 의해 예측된 예측 결과의 신뢰도를 알 수 없다는 문제를 해결하기 위해 서포트 벡터 데이터 디스크립션(Support Vector Data Description; SVDD)이라는 기법이 활용될 수 있다.

[0135] 예측 결과의 신뢰도를 평가하기 위한 장치는 이후의 신규 입력 데이터를 분류하기 위한 데이터 분류 알고리즘, 예를 들어, 서포트 벡터 데이터 디스크립션(Support Vector Data Description; SVDD)을 예측 모델을 훈련할 때 사용한 학습 데이터를 이용하여 생성할 수 있다.

[0136] 예측 결과의 신뢰도를 평가하기 위한 장치는 예측 모델을 준비한 이후, 위의 데이터 분류 알고리즘을 구축하기 전에, 데이터 분류 알고리즘의 커널 변수인 g를 위한 g_1, g_2, 쪋 g_n을 준비할 수 있다(S220).

[0137] 최초 루프에서는 g_1(커널 변수의 제1 값으로 지칭될 수 있음)을 가지는 제1 데이터 분류 알고리즘을 위의 예측 모델을 훈련하는데 사용된 훈련 데이터를 이용하여 구축할 수 있다.

[0138] 훈련 데이터를 이용하여 구축된 SVDD는 훈련 데이터가 존재하는 영역 내부와 외부를 나눌 수 있는 경계를 찾아낼 수 있다. 이후에 훈련 데이터가 아닌 새로운 입력 데이터가 주어지면 SVDD는 이 경계를 기준으로 훈련 데이터의 입력 데이터와 새로운 입력 데이터가 유사한지 판단할 수 있고, 유사하면 영역 내부에 있는 것으로 판단하고, 유사하지 않고 차이가 크면 영역 외부에 있는 것으로 판정할 수 있다.

[0139] 새로운 입력 데이터가 SVDD에 의해 영역 내부에 있다는 것은 기존 모델의 학습에 사용한 입력 데이터와 유사도

가 높다는 것(영역 내부 데이터로 분류된 것)으로 학습된 머신러닝 모델에 입력하여 타겟을 예측하면 예측 신뢰도가 높을 가능성이 크다는 것을 의미할 수 있다.

[0140] 반면에, 새로운 입력 데이터가 SVDD에 의해 영역 외부에 있다는 것은 기존 모델의 학습에 사용한 입력 데이터와 유사도가 낮다는 것(영역 외부 데이터로 분류된 것)이며, 머신러닝 모델이 충분히 학습하지 못한 자료가 입력 데이터로 들어온다는 것이며, 이에 따라 머신러닝 모델의 예측 신뢰도가 낮을 것으로 예상될 수 있다.

[0141] SVDD는 Boundary-Based Novelty Detection의 일종으로 Binary Classification을 수행하기 위한 경계(Boundary)를 제공한다는 점에 있어서 SVM과 용도가 비슷하나 SVM이 원점에서 가장 멀리 떨어진 하이퍼플레인 바운더리를 찾는 것에 목적을 두었다면 SVDD는 정상 데이터를 감싸안는 최소 체적의 구(Sphere)를 찾는 것을 목적으로 하며, 이상치 탐지를 위한 경계를 제공하는 알고리즘이다.

[0142] SVDD의 구(Sphere)는 하기 수학식 2의 가우시안 커널을 이용하여 생성되고,

[0144] [수학식 2]

[0145] $$K(x_i, x_j) = \exp\left(-g\|x_i - x_j\|^2\right)$$

[0146] 여기서, x_i 및 x_j는 데이터 포인트이며, g는 커널 너비를 조절하는 변수이다.

[0147] 한편, SVDD의 목적 함수는 하기 수학식 3과 같다.

수학식 3

[0149] $$\min_{R,a,\xi} R^2 + C\sum_{i=1}^{n}\xi_i$$
$$\text{subject to } \|\Phi_k(x_i) - a\|_{F_k}^2 \le R^2 + \xi_i,\ \xi_i \ge 0 \text{ for all } i = 1,2,\cdots,n$$

[0150] 여기서, R은 상기 구의 반경이고, a는 상기 구의 중심이며, ξ_i 는 소프트 마진을 위한 페널티 값이고, C는 트레이드 오프를 제어하기 위한 하이퍼파라미터이고, Φ_k 는 오리지널 데이터를 피쳐 공간에 맵핑하기 위한 함수이고, F_k는 피쳐 공간일 수 있다.

[0151] SVDD를 통한 입력 데이터의 내부 영역과 외부 영역을 나누는 경계는 SVDD의 커널변수인 g에 의해 결정될 수 있다.

[0152] g는 경계를 정할 때 분산정도를 나타내는 변수로 g가 커질수록 기존 데이터에 매우 가까운 영역만 내부 영역으로 결정하는 특성이 있다. 즉, 작은 g값을 사용하면 훈련시 사용된 입력 데이터가 있는 영역에서 많이 떨어진 영역도 내부 영역으로 결정하고 g가 크면 훈련시 사용된 입력 데이터에 매우 근접한 영역에 대해 내부 영역으로 결정할 수 있다.

[0153] 따라서, 변동하는 g에 대해 항상 내부 영역으로 결정되는 데이터는 훈련시 사용된 입력 데이터와 매우 밀접한 위치에 있다는 것이고, 작은 g에 대해서는 내부 영역으로 나타나지만 g가 커지면서 외부 영역으로 결정되는 데이터는 훈련시 사용된 입력 데이터의 분포 가장자리에 위치한 경우로 판단될 수 있다.

[0154] 또한, g값에 상관 없이 외부영역으로 결정되는 데이터는 훈련시 사용된 입력 데이터와 무관하게 멀리 떨어진 위치에 있는 것으로 판단될 수 있다.

[0155] 따라서, 도 6의 흐름도에서 제시된 바와 같이 SVDD 모델을 사용하여 훈련에 사용된 입력 데이터의 특성을 파악하기 위해서 다양한 g값을 준비하고 각 g에 대해 SVDD를 적용시켜 훈련시 사용된 입력 데이터를 기준으로 하여 내부 영역과 외부 영역을 결정할 수 있다.

[0156] 이후, 각 g에 대해 훈련 데이터를 학습한 SVDD 모델들 각각에 예측에 사용되기 위한 새로운 입력 데이터를 입력하여 내부 데이터와 외부 데이터로 분류할 수 있다.

[0157] 이렇게 n개의 g값에 대해 새로운 입력 데이터의 각각의 데이터가 어떤 영역에 해당하는지 파악한 후 g에 따른 데이터 영역의 변동을 검토하면 새로운 입력 데이터 내의 각각의 데이터에 따른 예측의 신뢰도를 평가할 수 있다.

[0158] g 값이 증가함에도 불구하고 항상 내부 데이터로 표현되는 데이터는 머신러닝 모델의 훈련에 사용된 입력 데이터와 유사한 영역이므로 예측의 성능이 높을 것으로 평가될 수 있다.

[0159] 반면에, g 값에 관계 없이 항상 외부 데이터로 표현되는 데이터는 훈련시 사용된 입력 데이터와 멀리 떨어진 외부에 있는 것으로 이 영역에 대해서는 머신러닝 모델이 학습하지 못한 영역이므로 예측 성능이 낮을 것으로 평가될 수 있다.

[0160] 이러한 방식으로 머신러닝 모델을 새로운 입력 데이터에 적용하기 전 머신러닝 학습에 사용된 입력 데이터와의 관계를 파악하여 새로운 입력 데이터에 따른 예측 신뢰도를 평가할 수 있다.

[0161] 위와 같은 본 발명의 예측 모델의 예측 결과 신뢰도 평가 방식에 따른 반복 루프가 시작되는 최초의 루프에서 i는 1로 설정될 수 있다(S230).

[0162] 먼저, 첫번째 루프에서 예측 모델을 훈련하는데 사용된 훈련 데이터를 이용하여 커널 변수 g의 제1 값(g_1)을 가지는 제1 데이터 분류 알고리즘(SVDD)이 구축될 수 있다(S240).

[0163] 여기서, g_1은 가장 작은 값으로 제1 데이터 분류 알고리즘은 가장 느슨한 기준으로 새로운 입력 데이터와 훈련시 사용된 입력 데이터 사이의 유사성을 판단할 수 있다.

[0164] 제1 데이터 분류 알고리즘이 구축되면, 예측 모델을 통한 예측에 사용되기 위한 새로운 입력 데이터를 제1 데이터 분류 알고리즘을 이용하여 제1 외부 데이터 및 제1 내부 데이터로 분류할 수 있다(S250).

[0165] 여기서 제1 외부 데이터는 훈련에 사용된 입력 데이터와 유사도가 매우 낮은 데이터로 제1 외부 데이터에 예측 모델이 적용되어 도출되는 예측 결과는 낮은 신뢰도를 가질 것으로 평가될 수 있다.

[0166] 제1 내부 데이터와 제1 외부 데이터가 각각 예측 모델에 입력되어 예측결과가 출력될 수 있으며(S260), 제1 외부 데이터에 따른 예측 결과는 가장 낮은 신뢰도를 갖는 것으로 평가될 수 있다.

[0167] 이후에는, i값이 미리 결정된 회차인 n회차에 도달하였는지 판단하고(S270), i+1을 새로운 i 값으로 설정하고, 예를 들어, g_2를 가지는 제2 분류 데이터 알고리즘을 예측 모델을 훈련하는데 사용된 훈련 데이터를 이용하여 구축할 수 있으며, 상술된 S240 내지 S260이 다시 반복 수행될 수 있다. 여기서, g_1=1, g_2=3, g_3=5 로 설정될 수도 있다.

[0168] 이를 통해 위에서 도출된 제1 내부 데이터를 제2 데이터 분류 알고리즘을 이용하여 제2 외부 데이터 및 제2 내부 데이터로 분류할 수 있다.

[0169] 제1 내부 데이터 중 제2 외부 데이터에 대해 예측 모델이 적용된 예측 결과의 신뢰성은 제2 내부 데이터에 대해 예측 모델이 적용된 예측 결과의 신뢰성보다 낮을 것임이 이해될 수 있다.

[0170] 반면, 제2 외부 데이터에 대해 예측 모델이 적용된 예측 결과의 신뢰성은 제1 외부 데이터에 대해 예측 모델이 적용된 예측 결과의 신뢰성보다는 높을 것임이 이해될 수 있다.

[0171] 위와 같은 루프가 i=n이 될 때까지 반복될 수 있고, 새로운 입력 데이터는 n개의 외부 데이터와 1개의 제n 내부 데이터로 분류될 수 있다.

[0172] 이후 g_i 값에 따라 분류된 결과에 따라 n개의 외부 데이터와 1개의 제n 내부 데이터에 대한 머신러닝 모델의 예측 신뢰도 평가가 이루어질 수 있다.

[0173] 예를 들어, 예측 모델을 이용하여 제1 외부 데이터에 대해 예측된 예측 결과의 신뢰도를 제1 신뢰도로 결정하고, 예측 모델을 이용하여 제2 외부 데이터에 예측된 예측 결과의 신뢰도를 제2 신뢰도로 결정할 수 있으며, 여기서 제1 신뢰도는 제2 신뢰도보다 작게 결정될 수 있다.

[0174] 예를 들어, n이 3이라면, 제3 내부 데이터에 따른 예측 결과의 신뢰도는 95%로, 제3 외부 데이터에 따른 예측

결과의 신뢰도는 90%로, 제2 외부 데이터에 따른 예측 결과의 신뢰도는 85%로, 제1 외부 데이터에 따른 예측 결과의 신뢰도는 80%로 결정될 수 있다.

[0175] 즉, 가장 높은 분류 영역의 데이터에 대해 가장 높은 신뢰도 값을 설정하고, 외부 영역으로 분류될수록 일정 간격으로 신뢰도가 낮아지도록 설정될 수 있다.

[0176] 한편, 예측 모델을 통한 예측 결과의 신뢰도를 평가하는 방법은, 예측 모델을 통한 예측에 사용되기 위한 입력 데이터를 수신하고, 예측 모델을 훈련하는데 사용된 훈련 데이터를 이용하여 구축된 데이터 분류 알고리즘을 이용하여 새로운 입력 데이터를 분류하고, 분류된 입력 데이터 그룹마다 서로 다른 평가 신뢰도를 할당하는 단계를 포함하는 것으로 표현될 수 있다.

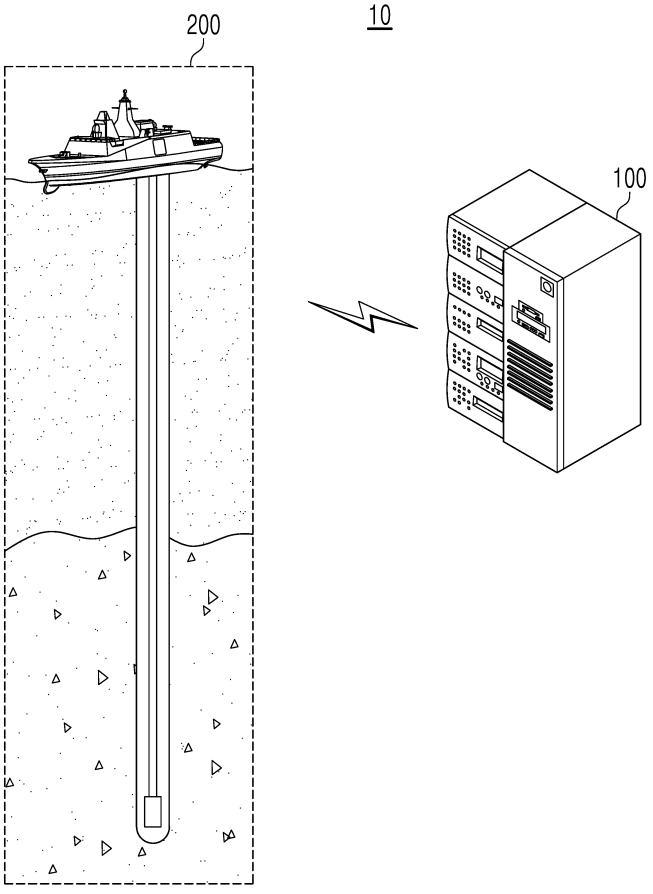
[0177] 여기서, 새로운 입력 데이터를 분류하는 단계는, 데이터 분류 알고리즘에 대한 커널 변수 g의 값을 변경시키면서 데이터 분류 알고리즘을 입력 데이터에 적용하여 입력 데이터를 복수의 그룹으로 분류하는 단계를 포함할 수 있다.

[0178] 이상 설명된 본 발명의 실시 예에 대한 설명은 예시를 위한 것이며, 본 발명이 속하는 기술분야의 통상의 지식을 가진 자는 본 발명의 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 쉽게 변형이 가능하다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시 예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다. 예를 들어, 단일형으로 설명되어 있는 각 구성 요소는 분산되어 실시될 수도 있으며, 마찬가지로 분산된 것으로 설명되어 있는 구성 요소들도 결합된 형태로 실시될 수 있다.

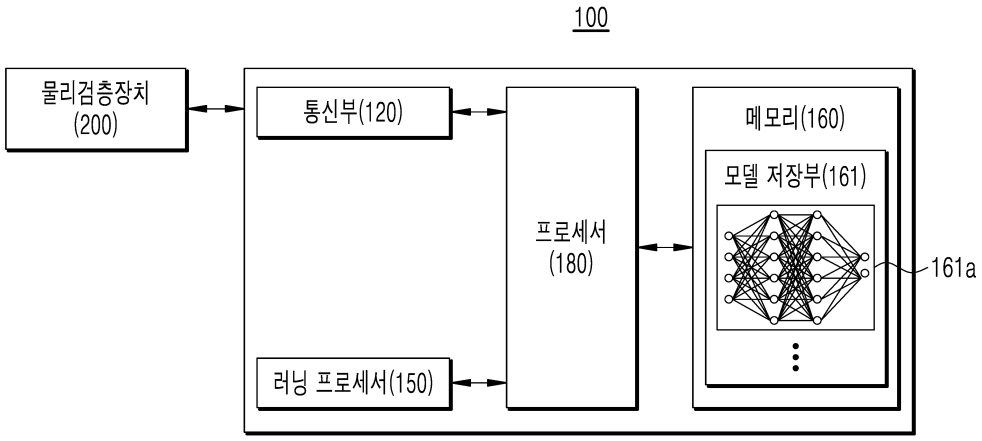
[0179] 본 발명의 범위는 상기 상세한 설명보다는 후술하는 청구범위에 의하여 나타내어지며, 청구범위의 의미 및 범위 그리고 그 균등 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야 한다.

도면

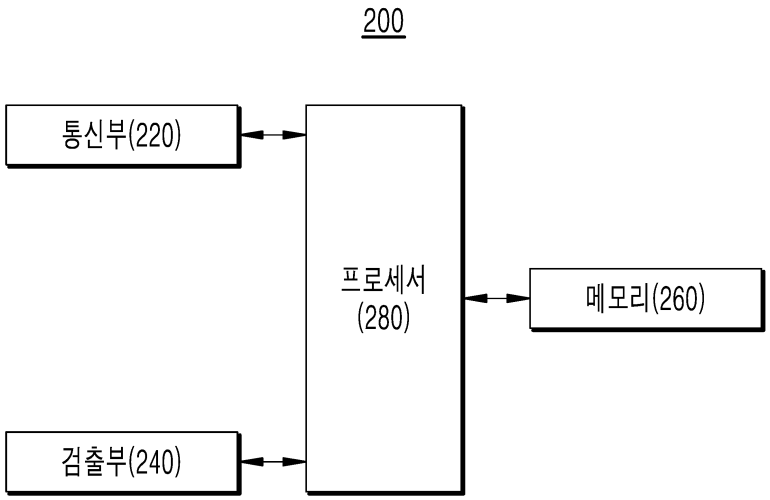
도면1



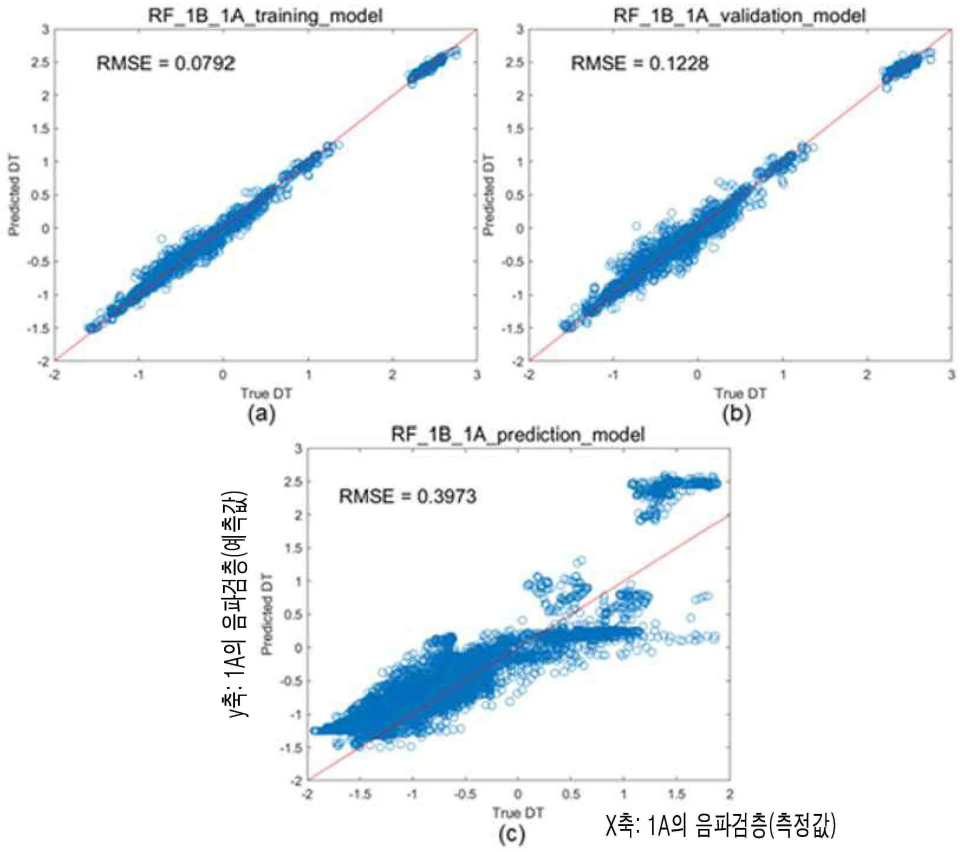
도면2


100
물리검증장치
(200)
통신부(120)
프로세서
(180)
메모리(160)
모델 저장부(161)
161a
러닝 프로세서(150)

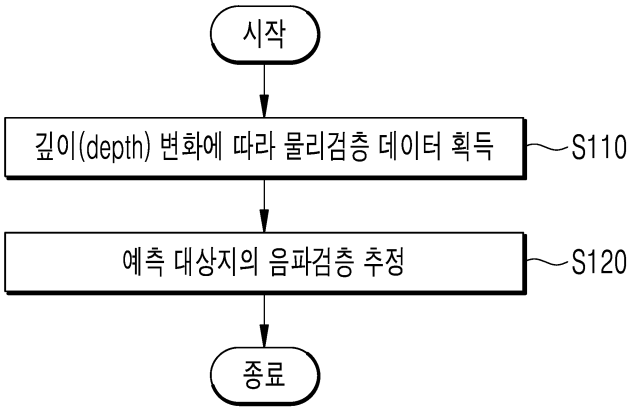
도면3


200
통신부(220)
프로세서
(280)
메모리(260)
검출부(240)

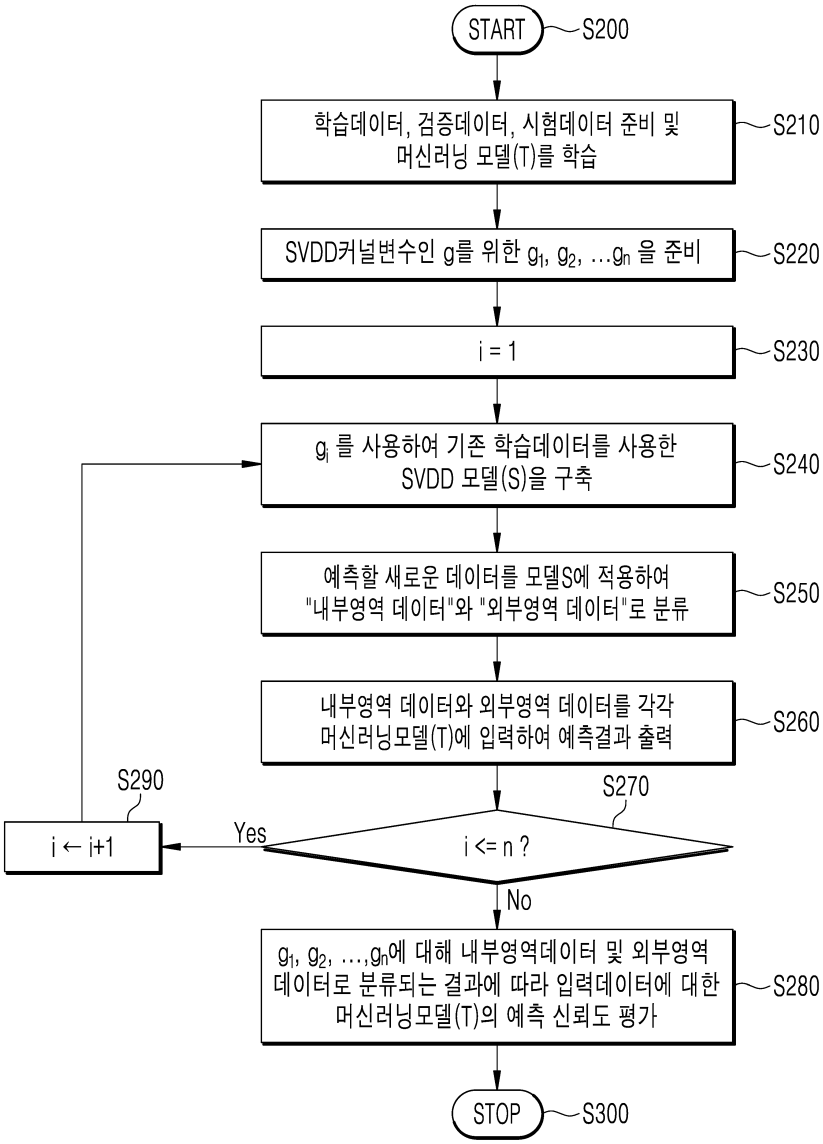
도면4


RF_1B_1A_training_model
RMSE = 0.0792
Predicted DT
True DT
(a)
RF_1B_1A_validation_model
RMSE = 0.1228
Predicted DT
True DT
(b)
RF_1B_1A_prediction_model
RMSE = 0.3973
y축: 1A의 음파검층(예측값)
Predicted DT
True DT
(c)
X축: 1A의 음파검층(측정값)

도면5


시작
깊이(depth) 변화에 따라 물리검층 데이터 획득
S110
예측 대상지의 음파검층 추정
S120
종료

도면6


START
S200
학습데이터, 검증데이터, 시험데이터 준비 및 머신러닝 모델(T)를 학습
S210
SVDD커널변수인 g를 위한 g_1, g_2, …g_n 을 준비
S220
i = 1
S230
g_i 를 사용하여 기존 학습데이터를 사용한 SVDD 모델(S)을 구축
S240
예측할 새로운 데이터를 모델S에 적용하여 "내부영역 데이터"와 "외부영역 데이터"로 분류
S250
내부영역 데이터와 외부영역 데이터를 각각 머신러닝모델(T)에 입력하여 예측결과 출력
S260
S270
i <= n ?
Yes
S290
i ← i+1
No
g_1, g_2, …,g_n에 대해 내부영역데이터 및 외부영역 데이터로 분류되는 결과에 따라 입력데이터에 대한 머신러닝모델(T)의 예측 신뢰도 평가
S280
STOP
S300