



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2018-0080097  
(43) 공개일자 2018년07월11일

(51) 국제특허분류(Int. Cl.)  
G06N 99/00 (2010.01) G06F 15/18 (2018.01)  
G06F 17/30 (2006.01)  
(52) CPC특허분류  
G06N 99/005 (2013.01)  
G06F 15/18 (2018.05)  
(21) 출원번호 10-2017-0133079  
(22) 출원일자 2017년10월13일  
심사청구일자 2017년11월01일  
(30) 우선권주장  
1020170000689 2017년01월03일 대한민국(KR)

(71) 출원인  
한국전자통신연구원  
대전광역시 유성구 가정로 218 (가정동)  
(72) 발명자  
강현중  
경상남도 진주시 수곡면 원창길 116  
김현재  
인천광역시 부평구 굴포로 158, 502동 2001호 (삼산동, 행복한마을서혜그랑블)  
(뒷면에 계속)  
(74) 대리인  
특허법인지명

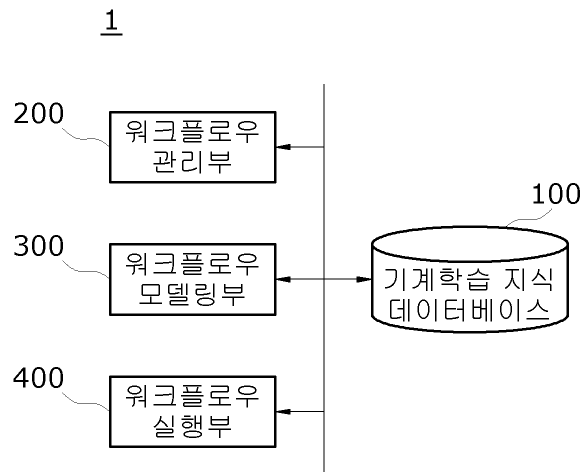
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 기계학습 지식 및 자동화된 기계 학습 절차 기반의 자가 학습 시스템

**(57) 요약**

본 발명에 따른 기계학습을 자동으로 수행하기 위한 자가 학습 시스템은 기계학습 지식이 저장된 기계학습 지식 데이터베이스 및 사용자의 요청 정보에 따른 기계학습을 자동으로 수행하기 위한 프로그램이 저장된 메모리 및 상기 메모리에 저장된 프로그램을 실행시키는 프로세서를 포함하되, 상기 프로세서는 상기 프로그램을 실행시킴에 따라, 상기 기계학습 지식 데이터베이스에 저장된 기계학습 지식에 기초하여 사용자의 요청 정보에 대응하는 하나 이상의 기계학습절차를 생성 또는 추천하고, 상기 생성 또는 추천된 기계학습절차를 실행하기 위한 실행코드를 생성한다.

**대표도** - 도1



(52) CPC특허분류

**G06F 17/30286** (2013.01)

(72) 발명자

**이호성**

대전광역시 유성구 배울1로 119, 1209동 1001호 (용산동, 대덕테크노밸리12단지아파트)

**권순현**

인천광역시 서구 검암로 53, 105동 204호 (검암동, 풍림아이원2차)

**김귀훈**

대전광역시 서구 만년로 25, 105동 1002호 (만년동, 강변아파트)

**김영민**

대전광역시 유성구 전민로 30, 101호 (전민동, 아이캐슬)

**김은주**

대전광역시 유성구 대덕대로590번길 11-10, 701호 (도룡동, 더포엠)

**박홍규**

대전광역시 서구 월평새뜸로4번길 53, 302호 (월평동, 채움빌라)

**배지훈**

세종특별자치시 새롬남로 18, 104동 1002호 (새롬동, 새뜸마을1단지)

**오세원**

대전광역시 서구 둔지로 75, 103동 705호 (둔산동, 파랑새아파트)

**유재학**

대전광역시 유성구 배울1로 35, 405동 1101호 (관평동, 대덕테크노밸리쌍용스윗닷홈4단지아파트)

**이연희**

대전광역시 유성구 노은로 353, 304동 1701호(하기동, 송림마을아파트3단지)

**김내수**

대전광역시 대덕구 동춘당로 178, 104동 1303호 (법동, 보람아파트)

**김선진**

대전광역시 유성구 왕가봉로 23, 1101동 602호 (노은동, 열매마을아파트 11단지)

**조성익**

대전광역시 유성구 죽동로 321, 108동 501호 (죽동, 금성백조예미지)

**표철식**

세종특별자치시 새롬남로 102, 1204동 1102호 (새롬동, 새뜸마을12단지)

이 발명을 지원한 국가연구개발사업

과제고유번호 CRC-15-05-ETRI

부처명 국가과학기술연구회

연구관리전문기관 국가과학기술연구회

연구사업명 융합연구사업

연구과제명 자가학습형 지식융합 슈퍼브레인 핵심기술개발

기 여 율 1/1

주관기관 한국전자통신연구원

연구기간 2015.12.01 ~ 2016.11.30

## 명세서

### 청구범위

#### 청구항 1

기계학습을 자동으로 수행하기 위한 자가 학습 시스템에 있어서,

기계학습 지식이 저장된 기계학습 지식 데이터베이스 및 사용자의 요청 정보에 따른 기계학습을 자동으로 수행하기 위한 프로그램이 저장된 메모리 및

상기 메모리에 저장된 프로그램을 실행시키는 프로세서를 포함하되,

상기 프로세서는 상기 프로그램을 실행시킴에 따라, 상기 기계학습 지식 데이터베이스에 저장된 기계학습 지식에 기초하여 사용자의 요청 정보에 대응하는 하나 이상의 기계학습절차를 생성 또는 추천하고, 상기 생성 또는 추천된 기계학습절차를 실행하기 위한 실행코드를 생성하는 것인 자가 학습 시스템.

#### 청구항 2

제 1 항에 있어서,

상기 기계학습 지식 데이터베이스는 사용자 타입에 따른 기계학습절차의 범위를 지식화한 사용자 지식, 분석 대상 도메인의 특성에 따른 기계학습절차의 범위를 지식화한 도메인 지식, 기계학습절차의 단계 생성을 위한 정보 구조를 정의한 가이드 지식 및 사용자 타입 및 도메인 타입 별로 적용 가능한 기계학습절차를 지식화한 워크플로우 지식 중 하나 이상을 포함하는 것인 자가 학습 시스템.

#### 청구항 3

제 2 항에 있어서,

상기 프로세서는 상기 사용자 지식, 도메인 지식, 가이드 지식 및 워크플로우 지식 중 하나 이상을 기반으로 상기 사용자의 요청 정보에 대응하는 하나 이상의 기계학습절차를 생성하는 것인 자가 학습 시스템.

#### 청구항 4

제 2 항에 있어서,

상기 사용자 지식은 사용자 타입 정보와, 사용자 구동 환경 정보와, 사용자 및 사용자 타입 별 기계학습절차의 사용자 설정 범위 또는 자동 설정 범위를 정의하는 설정 깊이 정보를 구조화되는 것인 자가 학습 시스템.

#### 청구항 5

제 2 항에 있어서,

상기 도메인 지식은 도메인 타입 정보, 상기 도메인 타입에서 해결하고자 하는 문제의 유형을 나타내는 문제 타입 정보를 포함하여 구조화되는 것인 자가 학습 시스템.

#### 청구항 6

제 2 항에 있어서,

상기 가이드 지식은 위치 정보 지식, 데이터 조건 지식, 모델 제한 지식, 실행 제한 지식, 사용 경험 지식 중 하나 이상을 포함하여 구조화되는 것인 자가 학습 시스템.

#### 청구항 7

제 6 항에 있어서,

상기 위치 정보 지식은 상기 기계학습절차를 수행하는 데 필요한 데이터 저장 위치 및 소프트웨어 패키지의 접근 경로 중 하나 이상을 포함하고,

상기 데이터 조건 지식은 상기 기계학습절차를 정의하는 특정 워크플로우, 특정 모델 요소, 특정 클래스에 대한 입력 및 출력 데이터 조건의 정보 중 하나 이상을 포함하며,

상기 모델 제한 지식은 수행 가능한 기계학습절차 또는 수행 가능한 기계학습모델을 제한하기 위한 지식을 포함하고,

상기 실행 제한 지식은 특정 기계학습 모델에 대한 도메인 제한 지식, 데이터 제한 지식, 메모리 제한 지식 및 하드웨어 제한 지식 중 하나 이상을 포함하며,

상기 사용 경험 지식은 예측 타입, 기계학습모델의 사용 빈도, 레이블 및 레이블의 필요 여부 정보 중 하나 이상을 포함하는 것인 자가 학습 시스템.

**청구항 8**

제 6 항에 있어서,

상기 가이드 지식은 상기 모델 제한 지식 및 상기 실행 제한 지식에 대하여 If, Then, Else 형태로 구조화되고,

상기 프로세서는 상기 기계학습절차의 실행 결과 정보를 통해 자동으로 상기 제한 지식을 지식화하는 것인 자가 학습 시스템.

**청구항 9**

제 2 항에 있어서,

상기 워크플로우 지식은 상기 기계학습절차를 구성하는 각 단위 기능을 정의하는 복수의 노드 및 상기 노드의 속성 정보와, 노드 간 연결 정보가 구조화되는 것인 자가 학습 시스템.

**청구항 10**

제 9 항에 있어서,

상기 복수의 노드는 작업 시작 노드, 데이터 처리 노드, 조건 분기 노드 및 작업 종료 노드 중 둘 이상의 노드를 포함하는 것인 자가 학습 시스템.

**청구항 11**

제 2 항에 있어서,

상기 기계학습 지식 데이터베이스는 상기 기계학습절차에서 이용 가능한 기능을 지식화한 논리적 지식을 더 포함하되,

상기 프로세서는 상기 논리적 지식을 기반으로 상기 생성한 기계학습절차를 논리적 지식 수준으로 구체화하는 것인 자가 학습 시스템.

**청구항 12**

제 11 항에 있어서,

상기 논리적 지식은 0개 이상의 물리적 지식과 매핑되는 것인 자가 학습 시스템.

**청구항 13**

제 2 항에 있어서,

상기 기계학습 지식 데이터베이스는 상기 기계학습절차에서 이용 가능한 소프트웨어 라이브러리 수준에서 모델 요소를 정의한 물리적 지식을 더 포함하되,

상기 프로세서는 상기 물리적 지식을 기반으로 상기 기계학습절차의 실행코드를 생성하는 것인 자가 학습 시스템.

**청구항 14**

제 13 항에 있어서,

상기 프로세서는 분석을 요청한 사용자 타입 및 분석 대상 도메인 타입을 포함하는 상기 사용자의 요청 정보를 수집하고, 상기 기계학습 지식 데이터베이스를 기반으로 상기 사용자의 요청 정보에 대응하는 하나 이상의 기계학습절차를 생성 또는 추천하며,

상기 기계학습 지식 데이터베이스에 포함된 물리적 지식에 기초하여 상기 실행코드를 생성하는 것인 자가 학습 시스템.

**청구항 15**

제 14 항에 있어서,

상기 프로세서는 상기 실행코드를 생성하기 이전에, 상기 기계학습 지식 데이터베이스에 포함된 논리적 지식에 기초하여, 상기 추천한 하나 이상의 기계학습절차를 논리적 지식 수준으로 구체화하고, 상기 구체화된 기계학습 절차를 상기 실행코드 수준으로 변환하는 것인 자가 학습 시스템.

**청구항 16**

제 14 항에 있어서,

상기 프로세서는 상기 생성한 실행코드에 기초하여 상기 하나 이상의 기계학습절차를 실행하고, 상기 하나 이상의 기계학습절차에 대한 결과를 피드백하여 상기 기계학습 지식 데이터베이스를 갱신하는 것인 자가 학습 시스템.

**청구항 17**

제 16 항에 있어서,

상기 프로세서는 상기 기계학습 지식 데이터베이스에 상기 사용자의 요청 정보에 대응하는 기계학습절차가 존재하지 않는 경우, 상기 사용자의 요청 정보에 포함된 분석 대상 도메인 타입에 적용 가능한 복수의 기계학습절차를 생성하고,

상기 생성한 기계학습절차의 실행 결과를 비교하여 성능을 분석하여 상기 복수의 기계학습절차 중 상기 추천할 하나 이상의 기계학습절차를 선택하여 제공하는 것인 자가 학습 시스템.

**청구항 18**

기계학습을 자동으로 수행하기 위한 자가 학습 방법에 있어서,

분석을 요청한 사용자 타입 및 분석 대상 도메인 타입을 포함하는 사용자의 요청 정보를 수신하는 단계;

기계학습 지식 데이터베이스에 저장된 기계학습 지식에 기초하여 상기 사용자의 요청 정보에 대응하는 하나 이상의 기계학습절차를 생성 또는 추천하는 단계; 및

상기 생성 또는 추천된 기계학습을 실행하기 위한 실행코드를 생성하는 단계를 포함하는 자가 학습 방법.

**청구항 19**

제 18 항에 있어서,

상기 기계학습 지식 데이터베이스는 사용자 타입에 따른 기계학습절차의 범위를 지식화한 사용자 지식, 분석 대상 도메인의 특성에 따른 기계학습절차의 범위를 지식화한 도메인 지식, 기계학습절차의 단계 생성을 위한 정보 구조를 정의한 가이드 지식, 사용자 타입 및 도메인 타입 별로 적용 가능한 기계학습절차를 지식화한 워크플로우 지식, 상기 기계학습절차에서 이용 가능한 기능을 지식화한 논리적 지식 및 상기 기계학습절차에서 이용 가능한 소프트웨어 라이브러리 수준에서 모델 요소를 정의한 물리적 지식 중 하나 이상을 포함하는 것인 자가 학습 방법.

**청구항 20**

제 19 항에 있어서,

상기 기계학습절차를 생성 또는 추천하는 단계는,

상기 사용자 지식, 도메인 지식, 가이드 지식 및 워크플로우 지식 중 하나 이상을 기반으로 상기 사용자의 요청 정보에 대응하는 하나 이상의 기계학습절차를 생성하는 단계; 및

상기 논리적 지식을 기반으로 상기 생성한 기계학습절차를 논리적 지식 수준으로 구체화하는 단계를 포함하고,

상기 실행코드를 생성하는 단계는, 상기 물리적 지식을 기반으로 상기 논리적 지식 수준으로 구체화된 기계학습 절차의 실행코드를 생성하는 것인 자가 학습 방법.

## 발명의 설명

### 기술 분야

[0001] 본 발명은 기계학습에 관련된 것으로서, 더욱 상세하게는 기계학습절차를 구조화된 지식에 저장하며, 지식을 기반으로 새로운 기계학습절차를 생성하고, 생성한 기계학습절차의 실행 결과를 다시 지식으로 반영하는 선순환을 통해서 사용자의 개입이나 요구되는 사전지식을 최소화할 수 있는 기계학습 지식 및 자동화된 기계 학습 절차 기반의 자가 학습 시스템에 관한 것이다.

### 배경 기술

[0002] 일반적으로 기계학습 절차를 구성하기 위해서는 전문적인 지식이 요구되며, 일반 사용자 등이 사용하기에는 많은 어려움이 존재한다.

[0003] 뿐만 아니라, 종래 기술의 경우 기존의 학습에 대한 설정 정보 및 그에 따른 결과 정보, 적합한 도메인이나 특징 등에 대한 정보의 공유가 원활하지 못하였다.

[0004] 이에 따라, 종래 기술의 경우 기계학습 전문가라 하더라도 목적하는 결과를 얻기 위해서는 하이퍼 파라미터 등과 같은 다양한 속성을 변경해 가면서 스스로 정보를 수집해야만 하는 문제가 있었다.

## 발명의 내용

### 해결하려는 과제

[0005] 본 발명의 일 실시예는 최적 기계학습절차의 생성에 있어서의 어려움을 해소하기 위하여 제안된 것으로서, 기계학습과 관련된 지식을 저장하고 또 다양한 수준의 사용자가 이용하기 위한 표준화된 구조를 제시하며, 해당 구조상에서 저장된 지식을 기반으로 최적의 기계학습절차를 추천하고, 추천한 기계학습절차의 실행 결과를 다시 지식으로 반영하는 선순환을 통해서 기계학습절차의 생성에 요구되는 사용자의 개입이나 요구되는 사전지식을 최소화할 수 있는 기계학습 지식 및 자동화된 기계 학습 절차 기반의 자가 학습 시스템을 제공하고자 한다.

[0006] 특히, 본 발명은 기계학습절차의 생성 및 결정을 위한 기술 및 단계적 복잡성으로 인해 비전문가 또는 해당 기계학습절차를 생성하지 아니한 자가 동일한 효과의 기계학습절차를 생성하는데 있어서의 어려움을 해소할 수 있는 기계학습 지식 및 자동화된 기계 학습 절차 기반의 자가 학습 시스템을 제공하고자 한다.

### 과제의 해결 수단

[0007] 상술한 기술적 과제를 달성하기 위한 기술적 수단으로서, 본 발명의 제 1 측면에 따른 기계학습을 자동으로 수행하기 위한 자가 학습 시스템은 기계학습 지식이 저장된 기계학습 지식 데이터베이스 및 사용자의 요청 정보에 따른 기계학습을 자동으로 수행하기 위한 프로그램이 저장된 메모리 및 상기 메모리에 저장된 프로그램을 실행시키는 프로세서를 포함한다. 이때, 상기 프로세서는 상기 프로그램을 실행시킴에 따라, 상기 기계학습 지식 데이터베이스에 저장된 기계학습 지식에 기초하여 사용자의 요청 정보에 대응하는 하나 이상의 기계학습절차를 생성 또는 추천하고, 상기 생성 또는 추천된 기계학습절차를 실행하기 위한 실행코드를 생성한다.

[0008] 상기 기계학습 지식 데이터베이스는 사용자 타입에 따른 기계학습절차의 범위를 지식화한 사용자 지식, 분석 대상 도메인의 특성에 따른 기계학습절차의 범위를 지식화한 도메인 지식, 기계학습절차의 단계 생성을 위한 정보 구조를 정의한 가이드 지식 및 사용자 타입 및 도메인 타입 별로 적용 가능한 기계학습절차를 지식화한 워크플로우 지식 중 하나 이상을 포함할 수 있다.

- [0009] 상기 프로세서는 상기 사용자 지식, 도메인 지식, 가이드 지식 및 워크플로우 지식 중 하나 이상을 기반으로 상기 사용자의 요청 정보에 대응하는 하나 이상의 기계학습절차를 생성할 수 있다.
- [0010] 상기 사용자 지식은 사용자 타입 정보와, 사용자 구동 환경 정보와, 사용자 및 사용자 타입 별 기계학습절차의 사용자 설정 범위 또는 자동 설정 범위를 정의하는 설정 깊이 정보를 구조화될 수 있다.
- [0011] 상기 도메인 지식은 도메인 타입 정보, 상기 도메인 타입에서 해결하고자 하는 문제의 유형을 나타내는 문제 타입 정보를 포함하여 구조화될 수 있다.
- [0012] 상기 가이드 지식은 위치 정보 지식, 데이터 조건 지식, 모델 제한 지식, 실행 제한 지식, 사용 경험 지식 중 하나 이상을 포함하여 구조화될 수 있다.
- [0013] 상기 위치 정보 지식은 상기 기계학습절차를 수행하는 데 필요한 데이터 저장 위치 및 소프트웨어 패키지의 접근 경로 중 하나 이상을 포함하고, 상기 데이터 조건 지식은 상기 기계학습절차를 정의하는 특정 워크플로우, 특정 모델 요소, 특정 클래스에 대한 입력 및 출력 데이터 조건의 정보 중 하나 이상을 포함하며, 상기 모델 제한 지식은 수행 가능한 기계학습절차 또는 수행 가능한 기계학습모델을 제한하기 위한 지식을 포함하고, 상기 실행 제한 지식은 특정 기계학습 모델에 대한 도메인 제한 지식, 데이터 제한 지식, 메모리 제한 지식 및 하드웨어 제한 지식 중 하나 이상을 포함하며, 상기 사용 경험 지식은 예측 타입, 기계학습모델의 사용 빈도, 레이블 및 레이블의 필요 여부 정보 중 하나 이상을 포함할 수 있다.
- [0014] 상기 가이드 지식은 상기 모델 제한 지식 및 상기 실행 제한 지식에 대하여 If, Then, Else 형태로 구조화되고, 상기 프로세서는 상기 기계학습절차의 실행 결과 정보를 통해 자동으로 상기 제한 지식을 지식화할 수 있다.
- [0015] 상기 워크플로우 지식은 상기 기계학습절차를 구성하는 각 단위 기능을 정의하는 복수의 노드 및 상기 노드의 속성 정보와, 노드 간 연결 정보가 구조화될 수 있다.
- [0016] 상기 복수의 노드는 작업 시작 노드, 데이터 처리 노드, 조건 분기 노드 및 작업 종료 노드 중 둘 이상의 노드를 포함할 수 있다.
- [0017] 상기 기계학습 지식 데이터베이스는 상기 기계학습절차에서 이용 가능한 기능을 지식화한 논리적 지식을 더 포함할 수 있다. 이때, 상기 프로세서는 상기 논리적 지식을 기반으로 상기 생성한 기계학습절차를 논리적 지식 수준으로 구체화할 수 있다.
- [0018] 상기 논리적 지식은 0개 이상의 물리적 지식과 매핑될 수 있다.
- [0019] 상기 기계학습 지식 데이터베이스는 상기 기계학습절차에서 이용 가능한 소프트웨어 라이브러리 수준에서 모델 요소를 정의한 물리적 지식을 더 포함할 수 있다. 이때, 상기 프로세서는 상기 물리적 지식을 기반으로 상기 기계학습절차의 실행코드를 생성할 수 있다.
- [0020] 상기 프로세서는 분석을 요청한 사용자 타입 및 분석 대상 도메인 타입을 포함하는 상기 사용자의 요청 정보를 수집하고, 상기 기계학습 지식 데이터베이스를 기반으로 상기 사용자의 요청 정보에 대응하는 하나 이상의 기계학습절차를 생성 또는 추천하며, 상기 기계학습 지식 데이터베이스에 포함된 물리적 지식에 기초하여 상기 실행코드를 생성할 수 있다.
- [0021] 상기 프로세서는 상기 실행코드를 생성하기 이전에, 상기 기계학습 지식 데이터베이스에 포함된 논리적 지식에 기초하여, 상기 추천한 하나 이상의 기계학습절차를 논리적 지식 수준으로 구체화하고, 상기 구체화된 기계학습절차를 상기 실행코드 수준으로 변환할 수 있다.
- [0022] 상기 프로세서는 상기 생성한 실행코드에 기초하여 상기 하나 이상의 기계학습절차를 실행하고, 상기 하나 이상의 기계학습절차에 대한 결과를 피드백하여 상기 기계학습 지식 데이터베이스를 갱신할 수 있다.
- [0023] 상기 프로세서는 상기 기계학습 지식 데이터베이스에 상기 사용자의 요청 정보에 대응하는 기계학습절차가 존재하지 않는 경우, 상기 사용자의 요청 정보에 포함된 분석 대상 도메인 타입에 적용 가능한 복수의 기계학습절차를 생성하고, 상기 생성한 기계학습절차의 실행 결과를 비교하여 성능을 분석하여 상기 복수의 기계학습절차 중 상기 추천할 하나 이상의 기계학습절차를 선택하여 제공할 수 있다.
- [0024] 또한, 본 발명의 제 2 측면에 따른 기계학습을 자동으로 수행하기 위한 자가 학습 방법은 분석을 요청한 사용자 타입 및 분석 대상 도메인 타입을 포함하는 사용자의 요청 정보를 수신하는 단계; 기계학습 지식 데이터베이스에 저장된 기계학습 지식에 기초하여 상기 사용자의 요청 정보에 대응하는 하나 이상의 기계학습절차를 생성 또

는 추천하는 단계; 및 상기 생성 또는 추천된 기계학습을 실행하기 위한 실행코드를 생성하는 단계를 포함한다.

[0025] 상기 기계학습 지식 데이터베이스는 사용자 타입에 따른 기계학습절차의 범위를 지식화한 사용자 지식, 분석 대상 도메인의 특성에 따른 기계학습절차의 범위를 지식화한 도메인 지식, 기계학습절차의 단계 생성을 위한 정보 구조를 정의한 가이드 지식, 사용자 타입 및 도메인 타입 별로 적용 가능한 기계학습절차를 지식화한 워크플로우 지식, 상기 기계학습절차에서 이용 가능한 기능을 지식화한 논리적 지식 및 상기 기계학습절차에서 이용 가능한 소프트웨어 라이브러리 수준에서 모델 요소를 정의한 물리적 지식 중 하나 이상을 포함할 수 있다.

[0026] 상기 기계학습절차를 생성 또는 추천하는 단계는, 상기 사용자 지식, 도메인 지식, 가이드 지식 및 워크플로우 지식 중 하나 이상을 기반으로 상기 사용자의 요청 정보에 대응하는 하나 이상의 기계학습절차를 생성하는 단계; 및 상기 논리적 지식을 기반으로 상기 생성한 기계학습절차를 논리적 지식 수준으로 구체화하는 단계를 포함할 수 있다. 그리고 상기 실행코드를 생성하는 단계는, 상기 물리적 지식을 기반으로 상기 논리적 지식 수준으로 구체화된 기계학습절차의 실행코드를 생성할 수 있다.

**발명의 효과**

[0027] 본 발명은 기계학습과 관련된 지식을 기계학습 지식 데이터베이스에 누적하고, 누적된 지식을 통해 다양한 사용자와 도메인 별로 적절한 기계학습절차를 자동 생성하거나 추천할 수 있으며, 기 생성된 기계학습절차를 다양한 알고리즘을 적용하여 운용, 평가, 저장 및 공유 할 수 있다.

[0028] 더 구체적으로, 본 발명은 데이터 전처리, 학습, 예측, 지식 증강과 같은 기계학습절차의 각 기능을 추상화하여 지식화하고, 이를 기반으로 생성한 기계학습절차의 운용에 따른 결과를 피드백 받아 지식을 강화할 수 있으며, 강화된 지식을 이용하여 다른 사용자가 기계학습절차 생성 시 추천 정보를 제공할 수 있다.

[0029] 또한, 본 발명은 기계학습을 위한 시소러스를 구성 및 갱신하며, 전문가, 비전문가, 엔지니어 등의 사용자 타입 및 헬스케어, 공장, 에너지, 홈, 빌딩, 오피스 등의 도메인 타입에 따라서 기계학습절차와 기능을 추상화 및 계층화하고, 기계학습 엔진의 라이브러리와 매핑함으로써, 다양한 계층의 사용자가 최소한의 지식으로 목적에 적합한 기계학습 절차를 생성 및 운용하도록 지원할 수 있다.

[0030] 또한, 본 발명은 사용자로부터 입력되는 데이터, 도메인, 학습 요구 사항 등이 사전에 기계학습절차로 지식화되어 있지 않은 경우라도, 적용 가능한 절차와 설정들을 모두 고려하여 복수의 기계학습절차를 생성하고, 각각의 기계학습절차를 독립적으로 또는 상호 연동하여 실행 후, 그 결과를 분석 평가함으로써, 최적의 기계학습절차를 추천할 수 있다.

[0031] 그 결과, 본 발명은 기존에 사용자 별로 독립적으로 생성 및 운용되던 기계학습 절차를 표준화 및 지식화하여 다양한 사용자들이 공유하도록 하며, 사용자 및 도메인 별 지식이 선순환을 통해 증강되도록 함으로써, 보다 많은 사용자가 기계학습절차를 용이하게 생성하여 운용할 수 있도록 한다.

**도면의 간단한 설명**

[0032] 도 1은 본 발명에 따른 기계학습 지식 및 자동화된 기계 학습 절차 기반의 자가 학습 시스템의 기본 구성을 나타낸 블록도이다.

도 2는 본 발명에 따른 기계학습 지식 및 자동화된 기계 학습 절차 기반의 자가 학습 시스템에서, 기계학습 지식 데이터베이스의 구조를 설명하는 도면이다.

도 3a 및 도 3b는 본 발명에 따른 기계학습 지식 및 자동화된 기계 학습 절차 기반의 자가 학습 시스템에서, 사용자 지식 및 도메인 지식을 설명하는 도면이다.

도 4a 내지 도 4c는 본 발명에 따른 기계학습 지식 및 자동화된 기계 학습 절차 기반의 자가 학습 시스템에 의해 추상화된 워크플로우 지식을 설명하기 위한 도면이다.

도 5는 본 발명에 따른 기계학습 지식 및 자동화된 기계 학습 절차 기반의 자가 학습 시스템에서, 논리적 지식과 물리적 지식 간의 관계를 설명하는 도면이다.

도 6은 본 발명에 따른 기계학습 지식 및 자동화된 기계 학습 절차 기반의 자가 학습 시스템에서, 기계학습절차의 생성 과정을 설명하는 도면이다.

도 7은 본 발명에 따른 기계학습 지식 및 자동화된 기계 학습 절차 기반의 자가 학습 시스템에서, 사용자 타입



및 도메인 타입 별 기계학습절차의 차별화된 추천 예를 설명하는 도면이다.

도 8은 본 발명에 따른 기계학습 지식 및 자동화된 기계 학습 절차 기반의 자가 학습 시스템의 동작을 설명하기 위한 흐름도이다.

도 9는 본 발명의 일 실시예에 따른 자가 학습 시스템의 구성도이다.

**발명을 실시하기 위한 구체적인 내용**

- [0033] 이하 본 발명의 바람직한 실시 예를 첨부한 도면을 참조하여 상세히 설명한다. 다만, 하기의 설명 및 첨부된 도면에서 본 발명의 요지를 흐릴 수 있는 공지 기능 또는 구성에 대한 상세한 설명은 생략한다. 또한, 도면 전체에 걸쳐 동일한 구성 요소들은 가능한 한 동일한 도면 부호로 나타내고 있음에 유의하여야 한다.
- [0034] 이하에서 설명되는 본 명세서 및 청구범위에 사용된 용어나 단어는 통상적이거나 사전적인 의미로 한정해서 해석되어서는 아니 되며, 발명자는 그 자신의 발명을 가장 최선의 방법으로 설명하기 위한 용어의 개념으로 적절하게 정의할 수 있다는 원칙에 입각하여 본 발명의 기술적 사상에 부합하는 의미와 개념으로 해석되어야만 한다. 따라서 본 명세서에 기재된 실시 예와 도면에 도시된 구성은 본 발명의 가장 바람직한 일 실시 예에 불과할 뿐이고, 본 발명의 기술적 사상을 모두 대변하는 것은 아니므로, 본 출원 시점에 있어서 이들을 대체할 수 있는 다양한 균등물과 변형 예들이 있을 수 있음을 이해하여야 한다.
- [0035] 또한, 제1, 제2 등과 같이 서수를 포함하는 용어는 다양한 구성요소들을 설명하기 위해 사용하는 것으로, 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용될 뿐, 상기 구성요소들을 한정하기 위해 사용되지 않는다. 예를 들어, 본 발명의 권리 범위를 벗어나지 않으면서 제2 구성요소는 제1 구성요소로 명명될 수 있고, 유사하게 제1 구성요소도 제2 구성요소로 명명될 수 있다.
- [0036] 더하여, 어떤 구성요소가 다른 구성요소에 "연결되어" 있다거나 "접속되어" 있다고 언급할 경우, 이는 논리적 또는 물리적으로 연결되거나, 접속될 수 있음을 의미한다. 다시 말해, 구성요소가 다른 구성요소에 직접적으로 연결되거나 접속되어 있을 수 있지만, 중간에 다른 구성요소가 존재할 수도 있으며, 간접적으로 연결되거나 접속될 수도 있다고 이해되어야 할 것이다.
- [0037] 또한, 본 명세서에서 사용한 용어는 단지 특정한 실시 예를 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 또한, 본 명세서에서 기술되는 "포함 한다" 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.
- [0038] 본 발명은 기계학습 지식을 기반으로 다양한 타입의 사용자가 다양한 분야의 도메인에 적합한 기계학습절차를 자동으로 생성하고, 생성한 기계학습절차를 다양한 알고리즘을 적용하여 구현할 수 있도록 하기 위한 것이다.
- [0039] 기계학습(Machine Learning, ML)은 인공지능의 한 분야로서, 주어진 데이터를 설명하는 공식을 찾아내는 일련의 과정이라고 할 수 있다.
- [0040] 기계학습의 핵심은 표현(representation)과 일반화(generalization)에 있으며, 여기서, 표현은 데이터의 평가를 의미하고, 일반화는 학습 이후에 새롭게 주어진 아직 알 수 없는 데이터에 대한 처리를 나타낸다. 여기에서 주어진 데이터는 답을 포함하고 있을 수도 있고, 없을 수도 있다. 만약에 답이 주어지게 되면, 수학적 연산을 통해 제공되는 데이터가 의미하는 바를 추측하고, 추측 값을 답과 비교한 후, 그 차이를 반영하여 추측함수를 갱신하는 과정을 반복적으로 수행하게 된다. 여기서, 기계학습의 수행을 통해 데이터를 설명하는 공식을 찾아가는 일련의 과정을 기능요소 별로 모듈화 하고 서로 연결한 것을 기계학습절차(Machine Learning Workflow, 이하 Workflow)라 한다.
- [0041] 이러한 기계학습절차를 생성하기 위해서는 다양한 기계학습 용어에 대한 이해뿐만 아니라, 기계학습 관련 기술에 대한 깊이 있는 이해가 필요하다.
- [0042] 특히, 기계학습절차는, 데이터 전처리, 학습, 평가, 예측, 지식화, 검색 등의 과정을 포함하는 것으로서, 동일한 목적을 위해 다양한 기계학습절차가 존재할 수 있으며, 다양한 기계학습절차를 비교하여 최적의 기계학습절차를 결정하기 위해서는 전문지식과 함께 많은 노력이 소요된다.
- [0043] 또한, 기계학습에서 가장 중요하고 가장 많은 시간을 소모하는 과정은, 수 백 만개 이상의 특징(feature) 또는

속성(attribute)들을 기반으로 하는 특징 추출(feature engineering), 모델 최적화(model optimization) 또는 최적 모델 선택(optimal model selecting) 과정이다.

- [0044] 한편, 본 발명의 일 실시예에서의 기계학습절차는 기계학습을 수행하는 전체 과정을 의미하는 것으로서, 이는 데이터 수집, 데이터 전처리, 기계학습, 기계학습의 평가, 기계학습의 결과 검증, 기계학습 결과 기반의 예측, 기계학습 절차에서 습득한 정보의 지식화를 포함할 수 있다.
- [0045] 예를 들어, 기계학습절차는, 저장공간 또는 장치에 있는 데이터를 실시간 또는 배치로 가져오고, 전 처리하여 빈 값을 대체하고, literal 값을 숫자 값으로 변환하여 컴퓨터가 연산할 수 있는 구조로 변경하며, 그 결과를 기반으로 5가지 타입으로 classification 할 수 있도록 학습하고, 학습을 통해 생성된 결과물인 모델을 이용하여, 동일한 구조의 새로운 데이터를 이용해 예측을 수행하며, 각 과정에서 발생하는 추가적 데이터를 기반으로 관계 추론 및 새로운 관계 연결 등을 통해 다시 지식에 반영하는 전반적인 과정을 정의한 것이다.
- [0046] 먼저, 도 1을 참조하여 본 발명에 따른 자가 학습 시스템(1)의 전반적인 구성을 개략적으로 설명한다.
- [0047] 도 1은 본 발명에 따른 기계학습 지식 및 자동화된 기계 학습 절차 기반의 자가 학습 시스템(1)의 기본 구성을 나타낸 블록도이다.
- [0048] 도 1을 참조하면, 본 발명에 따른 자가 학습 시스템(1)은, 기계학습 지식 데이터베이스(100), 워크플로우 관리부(200), 워크플로우 모델링부(300) 및 워크플로우 실행부(400)를 포함한다.
- [0049] 본 발명의 일 실시예에 따른 자가 학습 시스템(1)은 이들을 통해서 기계학습절차를 지식화하고, 이를 이용하여 사용자 타입 및 도메인 타입에 따른 적절한 기계학습절차를 생성한다. 또한, 생성한 기계학습절차를 다양한 알고리즘을 이용하여 구체화하게 되며, 이렇게 구체화된 기계학습절차를 실행코드 수준으로 변환하여 실행하고, 그 실행 결과를 다시 지식화한다.
- [0050] 기계학습 지식 데이터베이스(100)는 기계학습을 수행하기 위한 선택 가능한 모든 정보를 포함하는 기계학습지식(Machine Learning Knowledge)의 집합체이다. 구체적으로 기계학습절차가 추상화되고, 추상화된 기계학습절차를 사용자 및 도메인 타입에 따라서 구분하여 지식화하면, 기계학습 지식 데이터베이스(100)에는 추상화된 기계학습절차의 각 기능을 구현하기 위한 관련 지식이 저장된다.
- [0051] 예를 들어, 상기 기계학습 지식 데이터베이스(100)에는 기계학습절차를 생성하기 위한 소프트웨어와 관련된 지식, 기계학습절차의 생성을 위한 알고리즘과 관련된 지식, 기계학습 절차 생성을 위한 도메인 별 지식, 기계학습절차 생성을 위한 목적 별 지식, 각 도메인 별 지식 간의 관계, 기계학습에 이용되는 데이터, 기계학습절차의 수행 결과를 반영한 지식, 기계학습절차의 평가 정보 등이 포함된다. 그리고 이러한 지식들은 구조화되어 저장 가능한 모든 방법, 예를 들어, 온톨로지(ontology), 관계형 데이터베이스(Relational database, RDB), RDF(Resource Description Format) repository, 파일시스템(file system)을 통해 구축될 수 있으며, 상기 구축된 지식들에는 추상화된 정보의 구조(스키마 또는 온톨로지 등)와 이에 대한 인스턴스(Instance)가 포함된다.
- [0052] 도 2는 본 발명에 따른 기계학습 지식 및 자동화된 기계 학습 절차 기반의 자가 학습 시스템(1)에서, 기계학습 지식 데이터베이스(100)의 구조를 설명하는 도면이다.
- [0053] 특히, 도 2를 참조하면, 본 발명의 일 실시예에 따른 기계학습 지식 데이터베이스(100)에 저장된 기계학습절차의 자동화를 위한 지식들은, 사용자 지식(User knowledge)(110), 도메인 지식(Domain Knowledge)(120), 가이드 지식(Guide Knowledge)(130), 워크플로우 지식(Workflow knowledge)(140), 논리적 지식(Logical knowledge)(150) 및 물리적 지식(Physical knowledge)(160) 중 하나 이상을 포함할 수 있다. 이때, 사용자 지식(110), 도메인 지식(120), 가이드 지식(130), 워크플로우 지식(140), 논리적 지식(150) 및 물리적 지식(160)은 하나 이상의 다른 지식과 연관관계를 가질 수 있다.
- [0054] 이하에서는 도 3 내지 도 5를 참조하여, 기계학습 지식 데이터베이스(100)에 구조화된 지식들 각각 및 그 연관관계에 대해 구체적으로 설명하도록 한다.
- [0055] 도 3a 및 도 3b는 본 발명에 따른 기계학습 지식 및 자동화된 기계 학습 절차 기반의 자가 학습 시스템(1)에서, 사용자 지식 및 도메인 지식을 설명하는 도면이다. 도 4a 내지 도 4c는 본 발명에 따른 기계학습 지식 및 자동화된 기계 학습 절차 기반의 자가 학습 시스템(1)에 의해 추상화된 워크플로우 지식을 설명하기 위한 도면이다. 도 5는 본 발명에 따른 기계학습 지식 및 자동화된 기계 학습 절차 기반의 자가 학습 시스템(1)에서, 논리적 지식과 물리적 지식 간의 관계를 설명하는 도면이다.

- [0056] 먼저, 사용자 지식(110)은 사용자 타입에 따른 기계학습절차의 범위를 지식화한 것으로, 사용자 타입(User type) 및 설정 깊이(Setting depth), 사용자 구동 환경에 대한 정보를 포함한다.
- [0057] 여기서, 사용자 타입에 대한 정보는, 기계학습 및 도메인에 관련된 지식의 수준에 따라서 구분되는 것으로서, 예를 들어, 일반 사용자, 시스템 엔지니어, ML전문가 등을 포함할 수 있다. 이를 위해서, 상기 사용자 지식(110)은 일반적인 사용자 지식과 함께 사용자 수준에 따라 다른 구조로 재정의된 기계학습 지식의 기능을 포함할 수 있다.
- [0058] 그리고, 설정 깊이에 대한 정보는 해당 사용자 타입의 기계학습 지식 수준에 따라서 전체 기계학습절차에 대하여 어느 단계까지 사용자 설정을 허용하고, 어느 단계까지를 지식 기반으로 자동 설정할 것인지를 정의한다.
- [0059] 그리고, 사용자 구동 환경에 대한 정보는 사용자가 보유한 하드웨어 정보, 운영체제 정보를 포함한다.
- [0060] 예를 들어, 도 3a를 참조하면, ‘기계학습 전문가(P1)’의 경우, 기계학습절차의 각 기능의 모든 속성을 설정하는 것으로 제공할 수 있다. 반면, ‘시스템 엔지니어(P2)’의 경우, 일부 속성을 설정할 수 있도록 제공하고, 일반사용자(P3)에 대해서는 가급적 사용자 개입이 필요 없도록 최소 속성 설정 범위를 제공할 수 있다.
- [0061] 이와 같이, 특정 기계학습절차의 생성과 관련하여, 사용자 타입(예를 들어, 전문가, 시스템 엔지니어, 일반사용자 등)에 따라서 사용자가 설정할 수 있는 속성 정보의 범위는 사용자 지식(110)에 지식화되어 정의될 수 있다.
- [0062] 다음으로, 도메인 지식(120)은 기계학습이 적용되는 분야의 도메인 특성에 따른 기계학습절차의 상세 범위를 지식화한 것이다. 도메인 지식(120)은 도메인에 따라 다른 구조로 재정의된 기계학습 기능(Machine Learning Function)을 포함한다.
- [0063] 구체적으로, 도메인 지식(120)은 도메인 타입(Domain type)과, 문제 타입(Problem type)에 대한 정보를 포함하여 구조화된다.
- [0064] 상기 문제 타입은 기계학습을 통해서 해결하고자 하는 문제의 유형을 나타내는 것이다. 문제 타입은 일 예로, 지도 학습(supervised learning), 자율 학습(unsupervised learning), 강화 학습(reinforcement learning) 타입 등으로 구분되거나, 다른 기준으로는 군집화(clustering), 회귀분석(regression analysis), 분류(classification) 등으로 구분될 수도 있다.
- [0065] 이때, 도메인 지식(120)에는 기계학습이 적용되는 영역인 도메인 별로 구분되는 기계학습 관련 정보를 지식화한 것으로서, 각 도메인에서 사용하는 기계학습절차와 관련된 용어들이 포함될 수 있다. 또한, 도메인 지식(120)에는 동일한 목적의 기계학습 기능이나 속성이 도메인 별로 다르게 사용되는 경우 해당되는 기능이나 속성으로 매핑하는 정보들도 포함될 수 있다.
- [0066] 따라서, 본 발명의 일 실시예에 따른 자가 학습 시스템(1)은 도 3b와 같이, 도메인 지식(120)에 기반하여, 도메인 타입 별(P4-P6)로 각 기계학습에서 다르게 사용되는 언어나 용어를 적용하여 각 기계학습절차를 다르게 표현할 수 있다.
- [0067] 다음으로, 가이드 지식(130)은 기계 학습 절차의 단계적 생성을 위한 정보 구조와 제한 조건을 지식으로 정의한 것이다.
- [0068] 구체적으로 가이드 지식(130)은 해결하고자 하는 문제에 대해 대표성을 갖는 기능들을 기반으로 구조화된다. 이에 따라 가이드 지식(130)은 사용자가 제공한 입력에 따라 확정된 기계학습절차나 기능이 워크플로우 지식(140)에 의해 워크플로우로 추상화되고, 논리적 지식(150)에 의해 모델 요소로 변환될 수 있다. 또한, 가이드 지식(130)은 물리적 지식(160)에 의해 클래스로 변환되는 과정이 적절하게 수행되는 데 필요한 지식을 포함할 수 있다.
- [0069] 구체적으로 가이드 지식(130)은 위치 정보 지식, 데이터 조건 지식, 모델 제한(model restriction) 지식, 실행 제한(execution restriction) 지식, 사용 경험 지식을 포함할 수 있다.
- [0070] 상기 위치 정보 지식은 기계학습 절차를 수행하는 데 필요한 데이터 저장 위치, 소프트웨어 패키지의 접근 경로(access route) 등을 포함할 수 있다.
- [0071] 상기 데이터 조건 지식은 기계학습 절차를 정의하는 특정 워크플로우, 특정 모델 요소, 특정 클래스에 대한 입력 데이터 조건 및 출력 데이터 조건 등을 포함할 수 있다.
- [0072] 상기 모델 제한 지식은 수행 가능한 기계학습절차 또는 수행 가능한 기계학습 모델을 제한하기 위한 지식을 포

함할 수 있다. 예를 들어, 모델 제한 지식은 이미지 인식을 위한 데이터에 대해 정답에 해당하는 레이블이 없으면 "적절한 모델이 존재하지 않는다" 라고 제한하거나, "이미지 분류에 대한 회귀(Regression) 모델을 선택할 수 없다" 라고 제한할 수 있다. 즉, 데이터와 문제의 특성에 따라 활용 가능한 대상만이 추상화가 가능한 대상으로 제공되게 된다.

[0073] 상기 실행 제한 지식은 특정 기계학습 모델(특정 노드, 특정 모델 요소, 특정 클래스가 해당될 수 있음)에 대한 도메인 제한, 데이터 제한, 메모리 제한, 하드웨어 제한 지식 등을 포함할 수 있다. 도메인 제한 지식은 기계학습 모델이 적용한 가능한 도메인을 제한하기 위한 지식이고, 데이터 제한 지식은 기계학습 모델이 입출력 가능한 데이터를 제한하기 위한 지식이다. 그리고 메모리 제한 지식은 기계학습 모델의 실행에 필요한 최소 메모리를 제한하기 위한 지식이고, 하드웨어 제한 지식은 기계학습 모델이 실행될 수 있는 하드웨어를 제한하기 위한 지식이다.

[0074] 상기 사용 경험 지식은 예측 타입(prediction type), 기계학습 모델의 사용 빈도(frequently used workflow), 레이블(label) 및 레이블 필요여부 등의 정보를 포함할 수 있다. 상기 예측 타입은 기계학습을 통해 예측하고자 하는 정보의 타입을 의미하는 것으로서, 예를 들어, 정보의 타입은, 가부(true or false), 품질(quality), 수량(quantity), 가치(values), 구조(structure), 이상(anomaly), 범주(category) 중 하나 이상을 포함할 수 있다. 상기 기계학습 모델의 사용 빈도는 특정 기계학습 모델(특정 노드, 특정 모델 요소, 특정 클래스가 해당될 수 있음)이 사용된 빈도를 의미한다. 그리고 레이블은, 기계학습에서의 정답에 해당되는 것으로서, 문제를 해결한 후 기계학습에서 도출된 결과와 레이블로 정의된 정답과의 차이를 계산하여 문제 해결을 위한 절차를 보완할 수 있다. 상기 가이드 지식(130)의 예측 타입과 사용 빈도는 예측 타입 별로 빈번하게 사용되는 워크플로우 지식(140)과 연결될 수 있다.

[0075] 상기의 가이드 지식(130)은 구조화되는 제약사항에 대해 'If', 'Then', 'Else' 형태로 구조화 될 수 있다. 예를 들어 이미지(Image) 인식 문제에 대해, "입력 데이터(Input Data)의 샘플 사이즈(Sample size)가 100Mb 보다 작으면, 딥 러닝(Deep learning) 알고리즘을 사용할 수 없다." 와 같은 형식으로 구조화 될 수 있다. 이러한 제약사항은 워크플로우 실행을 통해 발생하는 결과 정보를 통해 자동적으로 지식화될 수도 있으며, 전문가의 설정에 의해 기본적인 구조를 생성하고, 운용 기록을 통해 자동적으로 갱신될 수도 있다.

[0076] 워크플로우 지식(140)은 사용자 타입 및 도메인 타입 별로 적용 가능한 기계학습절차를 추상화하여 노드-링크로 표현한 지식이다. 이때 각 기계학습절차는, 사용자 지식(110) 및 도메인 지식(120)에서 정의되는 특정 사용자 타입 및 도메인 타입과 연관될 수 있다.

[0077] 본 발명의 일 실시예에 따른 워크플로우 지식(140)에서, 기계학습절차는, 도 4a에 도시된 바와 같이, 기계학습 절차를 위한 하나 이상의 기계학습의 기능 요소를 노드(P7)라는 단위로 상호 연결한 워크플로우(W)로 정의된다. 노드(P7)는 기계학습절차를 구성하는 개별 기능을 정의하는 단위로서, 입력(input)과 출력(output)을 통해 복수의 노드를 연결하여 노드 간의 흐름을 구성할 수 있다. 하나의 워크플로우(P8)는 작업 시작 노드, 데이터 처리 노드, 조건 분기 노드, 작업 종료 노드 등을 포함하여 구성할 수 있다. 워크플로우(P8)는 기계학습 절차에 대한 설명 정보(description information)를 더 포함할 수 있다. 한편, 워크플로우(P8)에 포함되는 모든 노드의 수행에 필요한 제한 조건은 가이드 지식에 의해 정의된다.

[0078] 여기에서 작업 시작 노드는 최초로 작업이 시작되는 노드이고, 작업 종료 노드는 최후로 작업이 종료되는 노드이다. 데이터 처리 노드는 하나 이상의 노드가 출력하는 결과를 입력 받아 하나 이상의 노드로 출력하는 기능을 수행한다. 조건 분기 노드는 하나 또는 그 이상의 노드가 출력하는 결과를 입력받아 조건을 판단하여 둘 또는 그 이상의 노드로 선별적으로 출력하는 기능을 수행한다.

[0079] 상기 노드(P7)는 하나의 연산(computation)이 일어나는 동작(operation)을 수행하는 것으로서, 도 4b에 도시된 바와 같이, 노드(P7)의 양 엣지(edge)에서 데이터의 입출력이 발생한다.

[0080] 또한 도 4b에 도시한 바와 같이, 노드는 입력 및 출력과 함께, 연산의 동작을 설명하기 위한 정보들을 포함한다. 여기에 포함 가능한 정보는 노드의 이름, 타입정보, 카테고리 정보, 연산의 동작의 명세하는 각종 속성과 파라미터 정보들이 해당된다.

[0081] 또한 도 4c에 도시된 바와 같이, 노드의 종류는 작업 시작 노드(F0), 데이터 처리 노드(F1, F2, F3), 조건 분기 노드(C1), 작업 종료 노드(F4) 등으로 구분할 수 있다. 이때, 입력 엣지가 없는 노드는 작업 시작 노드(F0)를 의미하고, 출력 엣지가 없는 노드는 작업 종료 노드(F4)를 의미한다. 마름모 형태로 표시되어 있으며, 하나 또는 그 이상의 입력을 받아 둘 또는 그 이상의 출력으로 분기하도록 기능 타입이 정의된 노드는 조건 분기 노드



(C1)를 의미한다. 조건 분기 노드가 아니면서 입력 엣지와 출력 엣지가 모두 존재하는 노드는 하나 또는 그 이상의 입력 엣지를 통해 입력을 받아 기능 타입에 따른 기능을 수행하고 출력 엣지를 통해 하나의 출력을 하도록 정의된 데이터 처리 노드(F1, F2, F3)를 의미한다.

[0082] 상기 노드(P7)에서 입력 엣지는 노드의 입력 데이터 세트(input data set), 입력 조건(input condition), 입력 데이터의 특성에 대한 설명, 입력에 연결할 수 있는 노드의 식별자(input node identifier), 입력 명칭(input name), 입력 타입(input type), 입력 형식(input format) 등의 정보를 포함할 수 있다. 출력 엣지는 노드(P7)의 출력 데이터 세트(output data set), 출력 조건(output condition), 출력 데이터의 특성에 대한 설명, 출력에 연결할 수 있는 노드(P7)의 식별자(output node identifier), 출력 명칭(output name), 출력 타입(output type), 출력 형식(output format) 등의 정보를 포함할 수 있다. 상기 데이터 처리 노드(F1, F2, F3)의 기능에 대한 정의는 기능 식별자(function identifier), 기능 타입(function type), 기능 명칭(function name) 등을 포함할 수 있다. 속성(attribute)은 노드 식별자(node identifier), 노드 명칭(node name), 기능을 수행할 수 있는 하드웨어 디바이스의 종류, 실행 횟수(execution count) 정보 등을 포함할 수 있다. 이때 실행 횟수(execution count) 정보는 해당 기계학습 절차를 얼마나 많이 실행했는지에 대한 숫자 값으로, 이는 사용자, 도메인 등 이용 대상 구분자를 기준으로 독립적으로 저장될 수도 있다.

[0083] 상기 기계학습 절차에 대한 설명 정보는 워크플로우와 관련된, 키워드(keyword), 명칭(name), 식별자(identifier), 생성자(creator), 생성 시간(created time), 최종 수정 시간(last modified time), 지원되는 기계학습 엔진(supported machine learning engine), 관련 문제(related problem), 관련도(relevance rate) 등의 정보를 포함할 수 있다.

[0084] 이때, 상기 관련 문제 정보는 해당 기계학습 절차와 연관된 문제 타입(problem type)을 의미할 수 있으며, 관련도는 특정 기능이나 타입과의 연관도 혹은 유사도를 나타내는 것이 될 수 있다. 상기 관련 문제 정보와 관련된 가이드 지식(140)을 이용하여 적합한 기계 학습 절차를 탐색할 수 있다.

[0085] 다음으로, 논리적 지식(150)은 가이드 지식(140)에 의한 제한 조건을 준수하면서, 워크플로우의 노드를 하나 또는 그 이상의 모델 요소로 변환하기 위한 지식이다. 이러한 논리적 지식(150)은 기계학습을 위한 데이터 수집, 데이터 전처리, 기계학습, 학습결과 기반의 예측 과정 등의 기계학습절차에서 이용 가능한 기능을 지식화한 것이며, 용어 수준으로 기계학습절차에 사용되는 모델 요소를 정의한 것이다. 예를 들어, 논리적 지식(150)은 “Ensemble 알고리즘에는 Random Forest, Gradient Boosting Machines, AdaBoost, Gradient Boosted Regression Trees 등이 있다” 는 것 등을 구조화 할 수 있다.

[0086] 구체적으로, 논리적 지식(150)은 각 노드에 대한 기능 정보(function information), 설명 정보(description information) 등을 포함할 수 있다. 또한 논리적 지식(150)에서 정의되는 각 기능은 상위와 하위 관계 정보를 가지고 있으며, 범주(category), 패밀리(family), 그룹(group) 등을 통해 유사군을 구분할 수 있다.

[0087] 논리적 지식(150)의 설명 정보는 해당 모델 요소에 대한 명칭(name), 식별자(identifier), 버전(version), 소유자(owner), 사용자(users), 생성 시간(created time), 최종 수정 시간(last modified time) 등을 정보를 포함할 수 있다.

[0088] 논리적 지식(150)의 기능 정보는 기계학습의 기능을 정의하는데 필요한 상위 기능과 하위 기능의 관계로 정의되는 계층의 구조 정보(structure information)를 포함할 수 있다. 각 계층의 구조 정보에는 상위 계층 및 하위 계층에 대한 연결 정보, 연관 가능한 도메인 정보(relatable domain information), 타입(type), 식별자(identifier), 명칭(name), 설명(description) 등 정보를 포함할 수 있다. 연관 가능한 도메인 정보는 특정 기능 또는 그 기능을 포함하는 계층적 구조를 활용할 수 있는 도메인이 한정되는 경우에만 정의된다.

[0089] 또한, 논리적 지식(150)의 기능 정보는, 데이터의 위치와 접근 방법을 정의하는 데이터 소스(data source) 정보, 어떠한 방법으로 데이터를 수집할지를 정의하는 데이터 수집(data gathering) 정보, 수집 데이터 정보를 검토하기 위한 데이터 샘플링(data sampling) 정보, 수집된 데이터를 벡터화하기 위한 데이터 전처리(data preprocessing) 정보, 벡터화된 데이터를 기반으로 학습을 수행하는 학습(learning) 정보, 학습이 잘 되고 있는지 검증하는 학습 검증(learning test) 정보, 학습 결과에 대한 평가(learning evaluation) 정보, 학습결과로 도출된 학습모델을 이용해 새로운 데이터를 판단하는 예측(prediction) 정보, 전반적인 기능에서 개선되거나 새롭게 발견된 지식을 저장하는 지식화(save as Knowledge) 정보 중 하나 이상을 포함할 수 있다.

[0090] 더하여 논리적 지식(150)의 각 모델 요소는 도 5에 도시된 바와 같이, 0개 또는 그 이상의 물리적 지식(160)과 매칭될 수 있다. 이때, 논리적 지식(150)이 0개의 물리적 지식(160)과 매칭된다는 것은, 해당 기능과 매칭되는

물리적 지식(160)이 존재하지 않음을 의미한다. 이 경우, 사용자는 해당 기능에 대응되는 물리적 지식(160)을 추가할 수 있다. 상기 모델 요소가 지시하는 물리적 지식(160)은 작업 지식의 집합 구성, 물리적 지식(160)을 구성하는 기계학습 엔진의 특성이나 구조에 따라 달라질 수 있다.

- [0091] 마지막으로 물리적 지식(160)은, 가이드 지식에 의한 제한 조건을 준수하면서, 특정 소프트웨어 패키지에 대해 논리적 지식(150)의 모델 요소를 하나 또는 그 이상의 클래스로 변환하기 위한 지식이다. 상기 모델 요소는 기계학습을 위한 데이터 수집, 데이터 전처리, 기계학습, 학습결과 기반의 예측 과정 등의 기계학습절차에서 이용 가능한 기능을 논리적 수준에서 정의한 것이다.
- [0092] 한편, 본 발명의 일 실시예에 따른 자가 학습 시스템(1)은 워크플로우의 노드에 가이드 지식(130)의 제한 조건과 논리적 지식(150)을 적용하여 하나 또는 그 이상의 모델 요소를 생성하고, 상기 모델 요소에 가이드 지식(130)의 제한 조건과 물리적 지식(160)을 적용하여 하나 또는 그 이상의 클래스로 변환할 수 있다. 이때, 상기 클래스는 상기 워크플로우가 수행하는 데이터 처리 과정을 특정 소프트웨어 패키지의 제어 코드(control code) 또는 실행 코드(execution code)로 변환한 것이다.
- [0093] 물리적 지식(160)은 논리적 지식(150)의 모델 요소를 클래스로 변환하는 데 필요한 코드 변환 지식 사전을 포함할 수 있다. 코드 변환 지식 사전은 특정 모델 요소를 특정 소프트웨어 패키지에 적용하는 데 필요한 클래스 정보를 포함할 수 있다.
- [0094] 상기 클래스 정보는 특정 모델 요소에 대한 클래스가 지정되어 있지 않거나, 하나의 클래스만이 지정되어 있거나, 또는 두 개 이상의 클래스로 구성된 작업 프로세스가 지정될 수 있다. 이때 클래스가 지정되지 않는 경우는 특정 모델 요소에 대응하는 클래스가 해당 소프트웨어 패키지에 존재하지 않음을 의미한다. 이 경우, 사용자는 해당 기능에 맞는 클래스를 추가할 수 있으며, 추가된 클래스는 물리적 지식(160)에 자동으로 등록되어 지식 층장이 구현된다.
- [0095] 물리적 지식(160)은 코드 변환 대상이 되는 소프트웨어 패키지에 대한 설명 정보(description information), 속성 정보(attribute information), 해당 소프트웨어 패키지의 구성을 정의하는 클래스 정보(class information)를 포함할 수 있다. 물리적 지식(160)에 포함되는 각 소프트웨어 패키지는 하나 이상의 클래스로 구성될 수 있고, 각 클래스는 하나 이상의 기능(function)으로 구성될 수 있으며, 각 기능은 하나 이상의 인수(argument)를 포함하여 구성될 수 있다. 물리적 지식(160)은 이러한 설명, 클래스, 기능, 인수 등의 정보들을 구조화한 것일 수 있다.
- [0096] 상기 설명 정보는 해당 소프트웨어 패키지의 명칭(name), 식별자(identifier), 버전(version), 기능 명세, 설치 위치, 기관(organization), 지원 운영체제(supported operating system), 지원 장치 타입(supported device type), API 래퍼 언어(Application Program Interface wrapper language) 등의 정보를 포함할 수 있다.
- [0097] 상기 속성 정보는 추상화 타입, 오리지널과의 일치 여부, 수행 이력, 이력별 결과 정보 등을 포함할 수 있다.
- [0098] 상기 클래스 정보는 클래스 식별자(class identifier), 클래스 명칭(class name), 기능 타입(function type), 점수(score) 상위 클래스(parent class), 하위 클래스(child class), 범주(category), 인수 식별자(argument identifier), 인수 명칭(argument name), 인수 차순(argument order), 인수 상세(argument description), 반환 타입(return type), 반환 명칭(return name), 반환 상세(return description) 등의 정보를 포함할 수 있다.
- [0099] 이때, 상기 설명 정보에 포함된 기관(organization) 정보는 소프트웨어 패키지를 개발한 제작사, 소프트웨어 패키지에 대한 기술적 지원을 제공하는 지원 조직 등의 정보를 포함할 수 있다. 또한, 상기 지원 장치 타입(supported device type) 정보는 CPU(Central Processing Unit), GPU(Graphics Processing Unit), TPU(Tensor Processing Unit), ASIC(Application Specific Integrated Circuit), FPGA(Field Programmable Gate Array), 뉴로칩(Neuromorphic Chip, Neurosynaptic Chip) 등의 정보를 포함할 수 있다. 상기 API 래퍼 언어(API wrapper language) 정보는 Java, Python, Scala 등과 같이 해당 API를 서술한 언어 정보를 포함할 수 있다. 한편, 클래스를 구성하는 상기 기능은 입력 인수, 작업 수행 함수, 출력 인수 등의 정보를 포함할 수 있다.
- [0100] 본 발명의 일 실시예에 따른 자가 학습 시스템(1)은 논리적 지식(150)을 토대로 워크플로우의 노드를 모델 요소로 변환하고, 물리적 지식(160)을 토대로 모델 요소를 클래스로 변환할 수 있다.
- [0101] 또한, 본 발명의 일 실시예에 따른 자가 학습 시스템(1)은 물리적 지식(160)을 통해 최종 생성된 기계학습 제어 코드 또는 실행 코드에 대한 유효성 검증을 수행할 수 있다.
- [0102] 이때, 유효성 검증은 기계학습 제어 코드 또는 실행 코드와 같은 상기 코드가 실행될 소프트웨어 패키지에 공급

자 또는 사용자가 지정한 특정 데이터 세트와 가이드 지식에 정의된 기계학습 수행 환경을 적용하여 수행될 수 있다.

- [0103] 여기에서 특정 데이터 세트는 기계학습 모델의 개발 과정에서 널리 사용되는 사실 표준(de facto standards) 데이터세트 또는 공급자 또는 사용자가 이용하는 훈련 및 평가용 데이터세트 등을 포함할 수 있다.
- [0104] 본 발명의 일 실시예에 따른 자가 학습 시스템(1)은 유효성 검증이 완료된 코드를 소프트웨어 패키지에 적용하여 기계학습절차를 수행할 수 있다. 이때, 기계학습절차를 수행하도록 생성된 제어 코드 또는 실행 코드는 기계학습 엔진, 데이터 프로세싱 엔진, 데이터 스토리지 엔진, 데이터베이스 관리시스템에 연동하여 수행된다.
- [0105] 기계학습 엔진은 기계학습 기능을 수행할 수 있는 소프트웨어 패키지를 의미한다. 예를 들어, 널리 사용되는 기계학습 패키지에는 Cloudera Oryx, CUDA-Convnet, SciPy 등이 있고, 딥러닝 패키지에는 TensorFlow, Caffe, Theano, Keras 등이 있다.
- [0106] 그리고 대용량 데이터 처리를 수행할 수 있는 소프트웨어 패키지인 데이터 프로세싱 엔진의 일 예시로는 Hadoop MapReduce, Spark 등이 있다.
- [0107] 그리고 데이터 삽입, 추출, 갱신, 삭제 등의 기능을 수행할 수 있는 소프트웨어 패키지인 데이터 스토리지 엔진의 일 예시로는 Hbase, Cassandra, MongoDB, Apache Jena 등이 있다.
- [0108] 또한, 데이터베이스 관리시스템(database management system)은 데이터베이스에 저장된 데이터에 대한 접근 기능을 수행할 수 있는 소프트웨어 패키지를 의미한다.
- [0109] 이상의 설명에 있어서, 각 지식의 구조화 정보는 예시로서, 기계학습절차가 더 다양해지고 복잡해짐에 따라서 추가 및 갱신될 수 있다.
- [0110] 또한, 상술한 바와 같은 구조의 기계학습 지식 데이터베이스(100)는 사용자 혹은 지식화를 위해 기 설정된 기계학습장치를 통해 구축될 수 있다.
- [0111] 다시 도 1을 참조하면, 워크플로우 관리부(200)는 기계학습절차를 생성하고, 생성된 기계학습절차의 관리, 분석, 갱신을 수행하기 위한 구성이다. 구체적으로, 워크플로우 관리부(200)는 사용자와의 상호 작용을 통해서 사용자 요구사항을 수집하고, 상기 기계학습 지식 데이터베이스(100)의 지식들을 기반으로 상기 사용자 요구사항에 맞는 기계학습절차를 하나 이상을 생성하여 사용자에게 제공한다. 이때, 사용자와의 상호작용을 위해서, 상기 워크플로우 관리부(200)는 NLP(Neuro-Linguistic Programming) 등과 같이 사람의 언어나 행동을 해석하는 기능을 포함할 수 있다. 더하여, 워크플로우 관리부(200)는 워크플로우 모델링부(300) 및 워크플로우 실행부(400)로부터 워크플로우 정보 및 실행 결과를 피드백 받아, 상기 기계학습 지식 데이터베이스(100)의 지식을 갱신한다.
- [0112] 워크플로우 모델링부(300)는 상기 워크플로우 관리부(200)가 생성한 하나 이상의 기계학습절차를 상기 논리적 지식(150)을 기반으로 구체화하여 기계학습 모델을 생성한다. 이때, 상기 워크플로우 모델링부(300)에 의해 모델링 된 기계학습절차의 정보는 지식화를 위해 상기 워크플로우 관리부(200)로 피드백 된다.
- [0113] 워크플로우 실행부(400)는 상기 워크플로우 모델링부(300)가 구체화한 하나 이상의 기계학습절차를 실행하기 위한 구성이다. 구체적으로 워크플로우 실행부(400)는 기계학습 엔진의 라이브러리에 기반한 물리적 지식(160)을 기반으로 상기 모델링 된 기계학습 절차를 실행 코드 수준으로 변환하고, 기계학습 실행에 필요한 리소스를 배치하여 실행한다. 더하여, 상기 워크플로우 실행부(400)는 기계학습절차의 실행이 완료되면, 이를 지식화할 수 있도록 그 실행 결과와 이벤트 로그를 상기 워크플로우 관리부(200)로 피드백 한다.
- [0114] 이때, 워크플로우 실행부(400)는 평가 및 관리의 목적으로, 기계학습 수행 전반 및 각 컴포넌트에서 발생하는 각 반복학습(Iteration)별 결과 이력, 전체 지연 시간, 각 컴포넌트 시간 지연 이력, 오류 이력, 각 컴포넌트의 입출력 값, 상태 정보 등을 수집하여 상기 워크플로우 관리부(200)로 피드백 할 수 있다.
- [0115] 또한, 워크플로우 실행부(400)는 다수개의 기계학습절차를 동시에 생성하여 실행할 수 있으며, 각 기계학습절차의 실행 결과를 생성하고, 자동 혹은 사용자 정의에 따라서 속성을 매핑하여, 비교 정보를 제공할 수 있다.
- [0116] 이와 같은 구성으로 이루어지는 자가 학습 시스템에서, 기계학습절차를 구성하는 각 요소는 기계 학습 지식으로서 기계학습 지식 데이터베이스(100)에 저장되며, 기계학습절차를 생성하고자 하는 사용자의 타입, 적용되는 도메인 등에 따라서 다른 수준으로 제공될 수 있다.

- [0117] 상술한 바와 같이 구성되는 자가 학습 시스템(1)에서의 기계학습 지식 데이터베이스(100)를 기반으로 기계학습 절차의 생성 과정을 도 6을 참조하여 설명한다.
- [0118] 도 6은 본 발명에 따른 기계학습 지식 및 자동화된 기계 학습 절차 기반의 자가 학습 시스템(1)에서, 기계학습 절차의 생성 과정을 설명하는 도면이다.
- [0119] 본 발명의 일 실시예에 따른 자가 학습 시스템(1)을 이용하는 사용자는 기계학습을 수행하고자 하는 사용자의 사용자 타입 정보 및 적용 분야에 대한 도메인 정보를 포함하는 사용자 요구 사항을 입력하며, 기계학습절차의 생성을 지시할 수 있다. 상기 사용자 요구 사항은, 워크플로우 관리부(200)로 입력된다.
- [0120] 이에, 워크플로우 관리부(200)는 지식베이스(100)의 사용자 지식(110), 도메인 지식(120) 및 가이드 지식(130)에 기반하여, 사용자 타입 및 도메인 타입에 적합한 하나 이상의 기계학습절차를 생성한다.
- [0121] 도 7은 본 발명에 따른 기계학습 지식 및 자동화된 기계 학습 절차 기반의 자가 학습 시스템(1)에서, 사용자 타입 및 도메인 타입 별 기계학습절차의 차별화된 추천 예를 설명하는 도면이다.
- [0122] 노드들 N1~N8이 연결되어 이루어진 기계학습절차가 존재할 때, 'B 도메인 전문가(P10)'에 대해서는 각 노드의 속성 및 옵션을 해당 전문가 설정할 수 있도록 그대로 제공되는 반면, 다른 'A 도메인 전문가(P9)'에 대해서는 특정 노드(N1)가 A 도메인의 라이브러리에 매칭되고, 임의의 다른 노드(N3)는 최적 옵션으로 자동 설정되거나 가이드되어 제공될 수 있다.
- [0123] 또한, 사용자 타입이 '시스템 엔지니어(P11)'인 경우, 상기 기계학습절차 중 일부 노드 'N2~N4', 'N6~N7'이 자동으로 설정되어 제공되는 반면, 사용자 타입이 '일반 사용자(P12)'인 경우, 기계학습 절차의 대부분의 노드 'N2~N7'이 자동 생성되어 제공됨으로써, 사용자 수준에 맞추어 지원될 수 있다.
- [0124] 이렇게 생성된 기계학습절차의 각 노드는 워크플로우 모델링부(300)에 의해 도 6에 도시된 바와 같이, 논리적 기능 수준(150)으로 구체화되고, 각각의 기능은 워크플로우 실행부(400)에 의해 매칭되는 물리적 지식의 라이브러리(160)를 이용하여 실행 코드 수준으로 변환된다.
- [0125] 이렇게 실행 코드 수준으로 변환된 기계학습절차는 워크플로우 실행부(400)에 의해 실행되고, 기계학습절차의 실행이 완료되면, 상기 워크플로우 모델링부(300) 및 워크플로우 실행부(400)의 처리 결과는 워크플로우 관리부(200)를 통해 지식화되어 기계학습 지식 데이터베이스(100)에 저장한다.
- [0126] 이후에, 다른 사용자가 동일한 시나리오로 기계학습절차 계획(Workflow Plan)을 작성하고자 할 때, 개선된 버전의 기계학습절차 계획을 상기 다른 사용자에게 제공할 수 있다. 여기서, 기계학습절차 계획은, 기계학습에 대한 목표 및 예상 결과를 포함하는 것으로서, 상기 목표 및 예상 결과를 만족하기 위한 하나 이상의 기계학습 기능이 존재할 수 있으므로, 이에 의해 하나의 기계학습절차 계획에는 하나 이상의 기계학습절차가 존재할 수 있다.
- [0127] 이하에서는 도 8을 참조하여 본 발명의 일 실시예에 따른 자가 학습 시스템(1)에서의 기계학습절차의 생성 과정을 더 구체적으로 설명하도록 한다.
- [0128] 도 8은 본 발명에 따른 기계학습 지식 및 자동화된 기계 학습 절차 기반의 자가 학습 시스템(1)의 동작을 설명하기 위한 흐름도이다.
- [0129] 도 8을 참조하면, 워크플로우 관리부(200)는 기계학습절차의 생성을 요청한 사용자로부터 사용자 요청 정보를 입력 받는다(S105). 상기 사용자 요청 정보에는, 사용자 식별 정보, 사용자 타입 정보와 같은 사용자 정보, 분석 대상 도메인 타입 등을 포함하는 도메인 정보 및 분석할 데이터를 포함할 수 있다.
- [0130] 여기서, 사용자 타입은, 기계학습 지식을 기준으로 한 사용자 수준에 따라 구분되는 정보로서, 일반 사용자, 도메인 전문가, ML 전문가, 시스템 엔지니어 등으로 구분될 수 있다. 그리고 도메인 타입은, 응용 분야에 따라서 구분되는 정보로서, 예를 들어, 헬스, 시설, 에너지 등으로 구분될 수 있다.
- [0131] 다음으로, 워크플로우 관리부(200)는 상기 입력된 사용자 요청 정보에 기초하여 기계학습 지식 데이터베이스(100)에 관련 지식을 조회한다(S110).
- [0132] 여기서, 지식의 조회는 질의 및 응답 형태로 이루어질 수 있으며, 구체적으로, 워크플로우 관리부(200)는 사용자 타입에 대응하는 사용자 지식(110)과, 분석 대상 도메인 타입에 따른 도메인 지식(120) 및 가이드 지식(130)을 확인하고, 확인된 사용자 지식(110), 도메인 지식(120) 및 가이드 지식(130)에 기반하여 관련된 하나 이상의 워크플로우 지식(140)을 추출한다.



- [0133] 다음으로, 워크플로우 관리부(200)는 상기 조회된 지식을 기반으로 사용자 타입 및 도메인 타입에 적합한 하나 이상의 기계학습절차를 생성한다(S115). 여기서, 워크플로우 관리부(200)는 기계학습 지식 데이터베이스(100)로부터 사용자 타입 및 도메인 타입에 적합한 하나 이상의 기계학습절차를 조회하여 상기 기계학습절차의 옵션(속성 등)을 설정하는 형태로 하나 이상의 기계학습절차를 생성할 수 있다.
- [0134] 한편, 기계학습 지식 데이터베이스(100)에 사용자가 요구한 사용자 타입 및 도메인 타입에 대한 기계학습절차가 존재하지 않는 경우, 워크플로우 관리부(200)는 상기 도메인 타입에 적용 가능한 모든 지식을 요청하여, 복수의 기계학습절차를 생성할 수 있다.
- [0135] 다음으로, 워크플로우 관리부(200)는 생성한 하나 이상의 기계학습절차를 워크플로우 모델링부(300)에 전달한다(S120).
- [0136] 이어, 워크플로우 모델링부(300)는 상기 워크플로우 관리부(200)로부터 전달받은 각 기계학습절차를 논리적 지식 수준으로 구체화하기 위해, 질의 및 응답을 통해서 각 노드의 기능과 관련된 논리적 지식을 기계학습 지식 데이터베이스(100)에서 조회한다(S125). 이때, 상기 질의 및 응답은 워크플로우 관리부(200)를 통해 이루어질 수도 있다. 더하여, 워크플로우 모델링부(300)는 성능 최적화를 위해 하이퍼파라미터(Hyperparameter) 튜닝 기능을 기계학습절차에 추가할 수 있다.
- [0137] 다음으로, 워크플로우 모델링부(300)는 기계학습 지식 데이터베이스(100)로부터 전달받은 논리적 지식을 기반으로, 상기 하나 이상의 기계학습절차를 기능 수준으로 구체화한 후, 이를 실제 구동을 위해 워크플로우 실행부(400)로 전달한다(S130).
- [0138] 워크플로우 실행부(400)는, 기계학습 지식 데이터베이스(100)에 상기 구체화된 기계학습절차의 각 기능별 매칭되는 물리적 지식을 조회하여 수신한다(S135).
- [0139] 다음으로, 워크플로우 실행부(400)는 상기 물리적 지식 수준의 라이브러리를 이용하여 상기 하나 이상의 기계학습절차를 실행 코드 수준으로 변환하고(S140), 이를 실행한다(S145).
- [0140] 상기 워크플로우 실행부(400) 및 워크플로우 모델링부(300)의 실행 결과는 워크플로우 관리부(200)로 전달되고(S150), 워크플로우 관리부(200)는 기 생성한 하나 이상의 워크플로우의 실행 결과에 기반한 갱신 정보를 추출하여 기계학습 지식 데이터베이스(100)에 전달한다(S155). 여기서, 워크플로우 모델링부(300) 및 워크플로우 실행부(400)를 통해서 복수의 기계학습절차는 각각 모델링 및 실행된다. 그리고 워크플로우 모델링부(300) 및 워크플로우 실행부(400) 실행 결과를 분석하여 정확도, 학습시간, 예측 시간, 고려 사항 등의 성능 정보를 도출하고, 이를 비교하여 해당 사용자 타입 및 도메인 타입에 적합한 기계학습절차를 선별하여 제공한다.
- [0141] 이렇게 피드백된 기계학습절차 관련 지식은 기계학습 지식 데이터베이스(100)에 누적되어 지식확장이 이루어진다(S160).
- [0142] 도 9는 본 발명의 일 실시예에 따른 자가 학습 시스템(1)의 구성도이다.
- [0143] 한편, 본 발명의 일 실시예에 따른 자가 학습 시스템(1)은 통신모듈(10), 메모리(20) 및 프로세서(30)를 포함하여 구성될 수 있다.
- [0144] 통신모듈(10)은 외부 디바이스 및 장치들과 데이터를 송수신하기 위한 구성으로, 유선 통신 모듈 및 무선 통신 모듈을 모두 포함할 수 있다. 유선 통신 모듈은 전력선 통신 장치, 전화선 통신 장치, 케이블 홈(MoCA), 이더넷(Ethernet), IEEE1294, 통합 유선 홈 네트워크 및 RS-485 제어 장치로 구현될 수 있다. 또한, 무선 통신 모듈은 WLAN(wireless LAN), Bluetooth, HDR WPAN, UWB, ZigBee, Impulse Radio, 60GHz WPAN, Binary-CDMA, 무선 USB 기술 및 무선 HDMI 기술 등으로 구현될 수 있다.
- [0145] 메모리(20)에는 워크플로우 기반의 자가 학습 시스템(1)을 구동하기 위한 프로그램이 저장된다. 이때, 메모리(20)는 전원이 공급되지 않아도 저장된 정보를 계속 유지하는 비휘발성 저장장치 및 휘발성 저장장치를 통칭하는 것이다.
- [0146] 예를 들어, 메모리(20)는 콤팩트 플래시(compact flash; CF) 카드, SD(secure digital) 카드, 메모리 스틱(memory stick), 솔리드 스테이트 드라이브(solid-state drive; SSD) 및 마이크로(micro) SD 카드 등과 같은 낸드 플래시 메모리(NAND flash memory), 하드 디스크 드라이브(hard disk drive; HDD) 등과 같은 마그네틱 컴퓨터 기억 장치 및 CD-ROM, DVD-ROM 등과 같은 광학 디스크 드라이브(optical disc drive) 등을 포함할 수 있다.

- [0147] 프로세서(30)는 메모리(20)에 저장된 프로그램을 실행시킨다. 프로세서(30)는 상기 프로그램을 실행시킴에 따라, 도 1 내지 도 8에서 상술한 기계학습 지식 데이터베이스(100)를 관리하고, 워크플로우 관리부(200), 워크플로우 모델링부(300) 및 워크플로우 실행부(400)의 각 기능을 실행시킬 수 있다.
- [0148] 본 발명에 자가 학습 시스템(1)의 각 구성 및 동작은 다양한 컴퓨터 수단을 통하여 관독 가능한 소프트웨어 형태로 구현되어 컴퓨터로 관독 가능한 기록매체에 기록된 후, 하나 이상의 프로세서로 하여금 실행함에 의해 구현될 수 있다. 여기서, 기록매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 기록매체에 기록되는 프로그램 명령은 본 발명을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 예컨대 기록매체는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(Magnetic Media), CD-ROM(Compact Disk Read Only Memory), DVD(Digital Video Disk)와 같은 광 기록 매체(Optical Media), 플롭티컬 디스크(Floptical Disk)와 같은 자기-광 매체(Magneto-Optical Media), 및 롬(ROM), 램(RAM, Random Access Memory), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치를 포함한다. 프로그램 명령의 예는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함할 수 있다. 이러한 하드웨어 장치는 본 발명의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.
- [0149] 본 발명에 따른 자가 학습 시스템을 구현하는 프로세서는 본 발명에 따른 방법을 실행하기 위한 프로그램 명령을 처리할 수 있다. 일 구현 예에서, 이 프로세서는 싱글 스레드(Single-threaded) 프로세서일 수 있으며, 다른 구현 예에서 본 프로세서는 멀티 스레드(Multithreaded) 프로세서일 수 있다. 나아가 본 프로세서는 메모리 혹은 저장 장치 상에 저장된 명령을 처리하는 것이 가능하다.
- [0150] 나아가 본 발명에 따른 자가 학습 시스템은 서버 팜(Server Farm)과 같이 네트워크에 걸쳐서 분산형으로 구현될 수 있으며, 혹은 단일의 컴퓨터 장치에서 구현될 수도 있다.
- [0151] 비록 본 명세서와 도면에서는 예시적인 장치 구성을 기술하고 있지만, 본 명세서에서 설명하는 기능적인 동작과 주제의 구현물들은 다른 유형의 디지털 전자 회로로 구현되거나, 본 명세서에서 개시하는 구조 및 그 구조적인 등가물들을 포함하는 컴퓨터 소프트웨어, 펌웨어 혹은 하드웨어로 구현되거나, 이들 중 하나 이상의 결합으로 구현 가능하다. 본 명세서에서 설명하는 주제의 구현물들은 하나 이상의 컴퓨터 프로그램 제품, 다시 말해 본 발명에 따른 장치의 동작을 제어하기 위하여 혹은 이것에 의한 실행을 위하여 유형의 프로그램 저장매체 상에 인코딩된 컴퓨터 프로그램 명령에 관한 하나 이상의 모듈로서 구현될 수 있다. 컴퓨터로 관독 가능한 매체는 기계로 관독 가능한 저장 장치, 기계로 관독 가능한 저장 기판, 메모리 장치, 기계로 관독 가능한 전파형 신호에 영향을 미치는 물질의 조성물 혹은 이들 중 하나 이상의 조합일 수 있다.
- [0152] 특히, 본 발명에서 "시스템"이라는 용어는 예컨대 프로그래머블 프로세서, 컴퓨터 혹은 다중 프로세서나 컴퓨터를 포함하여 데이터를 처리하기 위한 모든 기구, 장치 및 기계를 포괄할 수 있다. 시스템은, 하드웨어에 부가하여, 예컨대 프로세서 펌웨어를 구성하는 코드, 프로토콜 스택, 데이터베이스 관리 시스템, 운영 체제 혹은 이들 중 하나 이상의 조합 등 요청 시 컴퓨터 프로그램에 대한 실행 환경을 형성하는 코드를 포함할 수 있다.
- [0153] 본 발명에 따른 기능을 실행하는 컴퓨터 프로그램(프로그램, 소프트웨어, 소프트웨어 어플리케이션, 스크립트 혹은 코드로도 알려져 있음)은 컴파일 되거나 해석된 언어나 선형적 혹은 절차적 언어를 포함하는 프로그래밍 언어의 어떠한 형태로도 작성될 수 있으며, 독립형 프로그램이나 모듈, 컴포넌트, 서브루틴 혹은 컴퓨터 환경에서 사용하기에 적합한 다른 유닛을 포함하여 어떠한 형태로도 전개될 수 있다. 컴퓨터 프로그램은 파일 시스템의 파일에 반드시 대응하는 것은 아니다. 프로그램은 요청된 프로그램에 제공되는 단일 파일 내에, 혹은 다중의 상호 작용하는 파일(예컨대, 하나 이상의 모듈, 하위 프로그램 혹은 코드의 일부를 저장하는 파일) 내에, 혹은 다른 프로그램이나 데이터를 보유하는 파일의 일부(예컨대, 마크업 언어 문서 내에 저장되는 하나 이상의 스크립트) 내에 저장될 수 있다. 컴퓨터 프로그램은 하나의 사이트에 위치하거나 복수의 사이트에 걸쳐서 분산되어 통신 네트워크에 의해 상호 접속된 다중 컴퓨터나 하나의 컴퓨터 상에서 실행되도록 전개될 수 있다.
- [0154] 본 명세서에서 설명한 주제의 구현물은 예컨대 데이터 서버와 같은 백엔드 컴포넌트를 포함하거나, 예컨대 어플리케이션 서버와 같은 미들웨어 컴포넌트를 포함하거나, 예컨대 사용자가 본 명세서에서 설명한 주제의 구현물과 상호 작용할 수 있는 웹 브라우저나 그래픽 유저 인터페이스를 갖는 클라이언트 컴퓨터와 같은 프론트엔드 컴포넌트 혹은 그러한 백엔드, 미들웨어 혹은 프론트엔드 컴포넌트의 하나 이상의 모든 조합을 포함하는 연산 시스템에서 구현될 수 있다. 시스템의 컴포넌트는 예컨대 통신 네트워크와 같은 디지털 데이터 통신의 어떠한

형태나 매체에 의해서도 상호 접속 가능하다.

[0155] 본 명세서는 다수의 특정한 구현물의 세부사항들을 포함하지만, 이들은 어떠한 발명이나 청구 가능한 것의 범위에 대해서도 제한적인 것으로서 이해되어서는 안되며, 오히려 특정한 발명의 특정한 실시형태에 특유할 수 있는 특징들에 대한 설명으로서 이해되어야 한다. 개별적인 실시형태의 문맥에서 본 명세서에 기술된 특정한 특징들은 단일 실시형태에서 조합하여 구현될 수도 있다. 반대로, 단일 실시형태의 문맥에서 기술한 다양한 특징들 역시 개별적으로 혹은 어떠한 적절한 하위 조합으로도 복수의 실시형태에서 구현 가능하다. 나아가, 특징들이 특정한 조합으로 동작하고 초기에 그와 같이 청구된 바와 같이 묘사될 수 있지만, 청구된 조합으로부터의 하나 이상의 특징들은 일부 경우에 그 조합으로부터 배제될 수 있으며, 그 청구된 조합은 하위 조합이나 하위 조합의 변형물로 변경될 수 있다.

[0156] 마찬가지로, 특정한 순서로 도면에서 동작들을 묘사하고 있지만, 이는 바람직한 결과를 얻기 위하여 도시된 그 특정한 순서나 순차적인 순서대로 그러한 동작들을 수행하여야 한다거나 모든 도시된 동작들이 수행되어야 하는 것으로 이해되어서는 안 된다. 특정한 경우, 멀티태스킹과 병렬 프로세싱이 유리할 수 있다. 또한, 상술한 실시형태의 다양한 시스템 컴포넌트의 분리는 그러한 분리를 모든 실시형태에서 요구하는 것으로 이해되어서는 안되며, 설명한 프로그램 컴포넌트와 시스템들은 일반적으로 단일의 소프트웨어 제품으로 함께 통합되거나 다중 소프트웨어 제품에 패키징될 수 있다는 점을 이해하여야 한다.

[0157] 본 명세서에서 설명한 주제의 특정한 실시형태를 설명하였다. 기타의 실시형태들은 이하의 청구항의 범위 내에 속한다. 예컨대, 청구항에서 인용된 동작들은 상이한 순서로 수행되면서도 여전히 바람직한 결과를 성취할 수 있다. 일 예로서, 첨부도면에 도시한 프로세스는 바람직한 결과를 얻기 위하여 반드시 그 특정한 도시된 순서나 순차적인 순서를 요구하지 않는다. 특정한 구현 예에서, 멀티태스킹과 병렬 프로세싱이 유리할 수 있다.

[0158] 본 기술한 설명은 본 발명의 최상의 모드를 제시하고 있으며, 본 발명을 설명하기 위하여, 그리고 당업자가 본 발명을 제작 및 이용할 수 있도록 하기 위한 예를 제공하고 있다. 이렇게 작성된 명세서는 그 제시된 구체적인 용어에 본 발명을 제한하는 것이 아니다. 따라서, 상술한 예를 참조하여 본 발명을 상세하게 설명하였지만, 당업자라면 본 발명의 범위를 벗어나지 않으면서도 본 예들에 대한 개조, 변경 및 변형을 가할 수 있다.

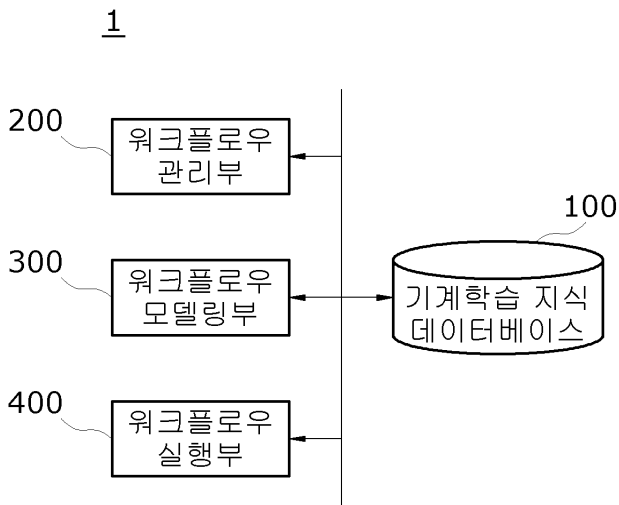
[0159] 따라서 본 발명의 범위는 설명된 실시 예에 의하여 정할 것이 아니고 특허청구범위에 의해 정하여져야 한다.

### 부호의 설명

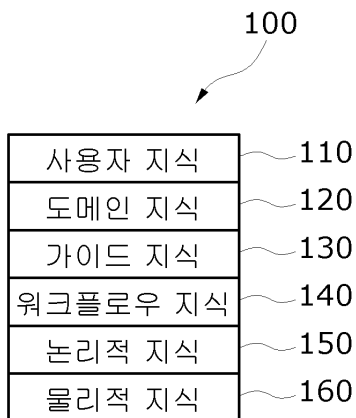
- [0160] 1: 자가 학습 시스템
- 10: 통신모듈
- 20: 메모리
- 30: 프로세서
- 100: 기계학습 지식 데이터베이스
- 200: 워크플로우 관리부
- 300: 워크플로우 모델링부
- 400: 워크플로우 실행부

도면

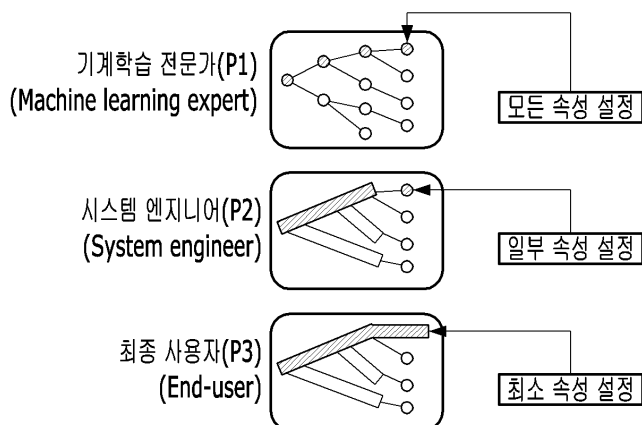
도면1



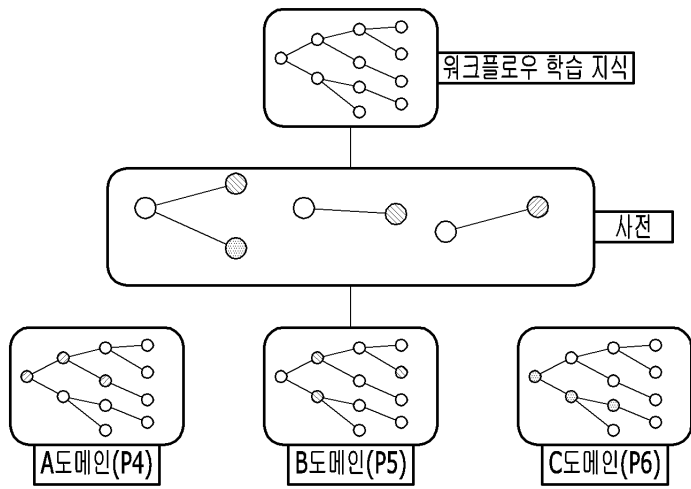
도면2



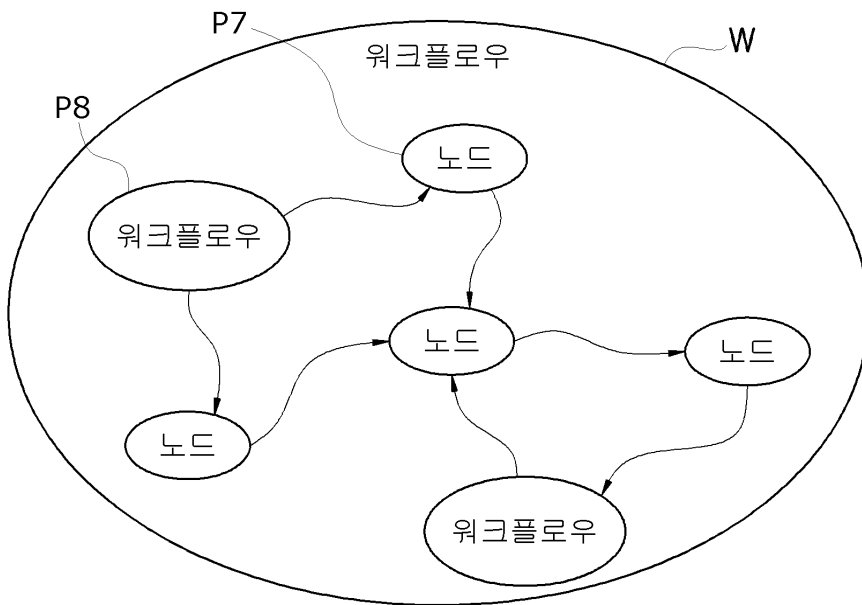
도면3a



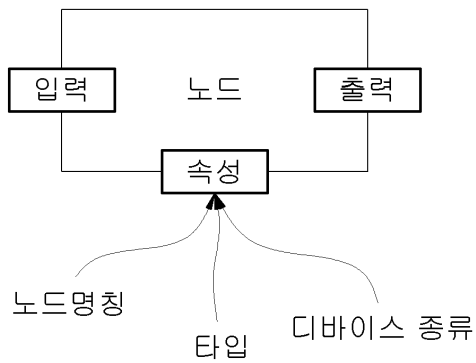
도면3b



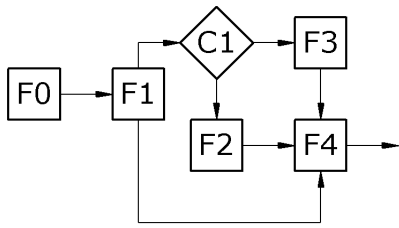
도면4a



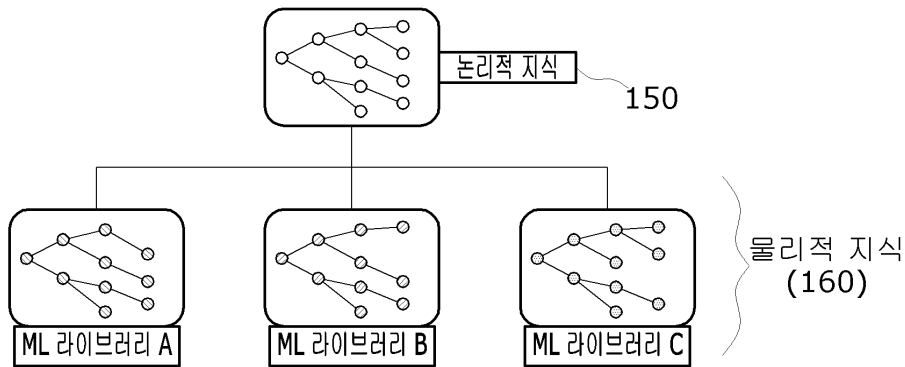
도면4b



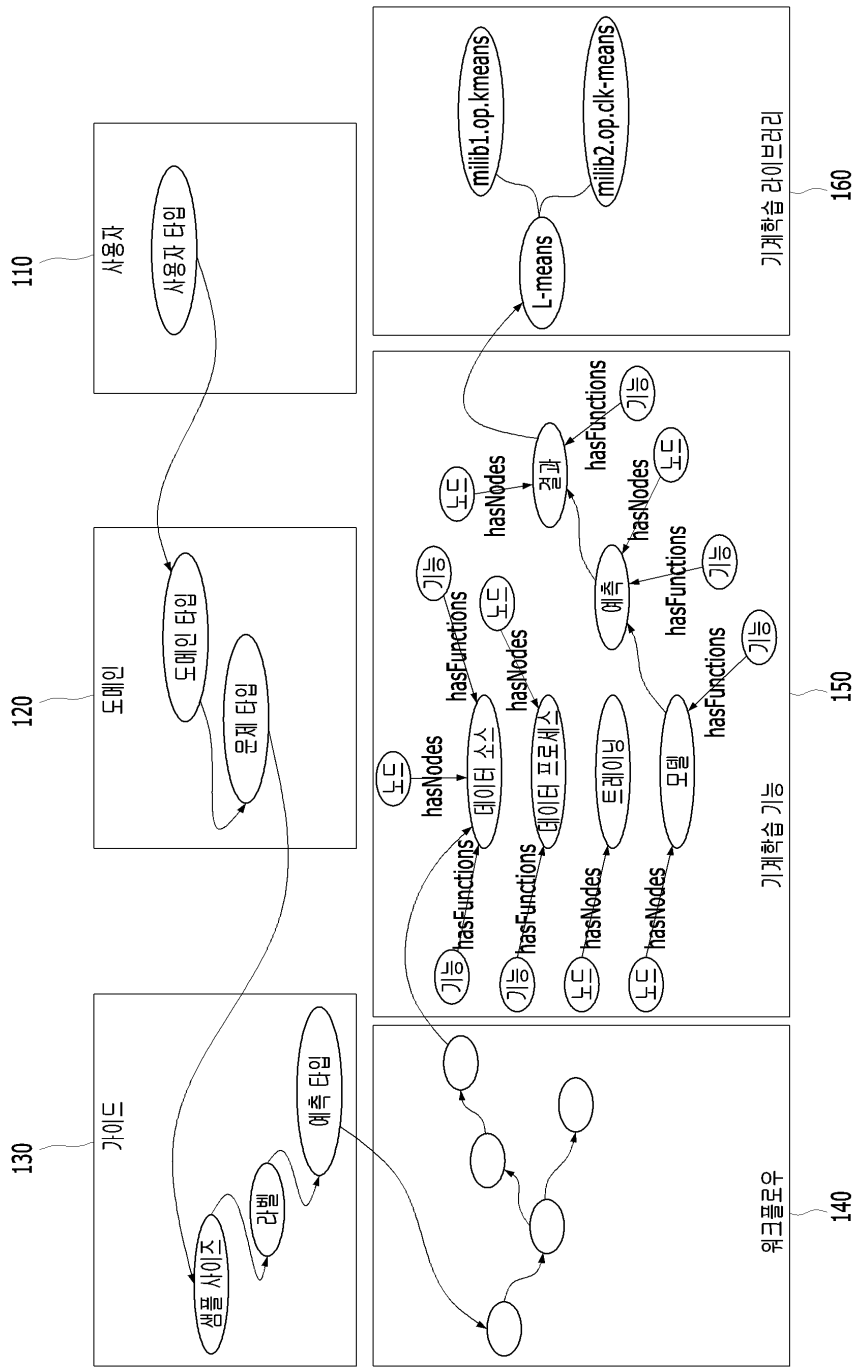
도면4c



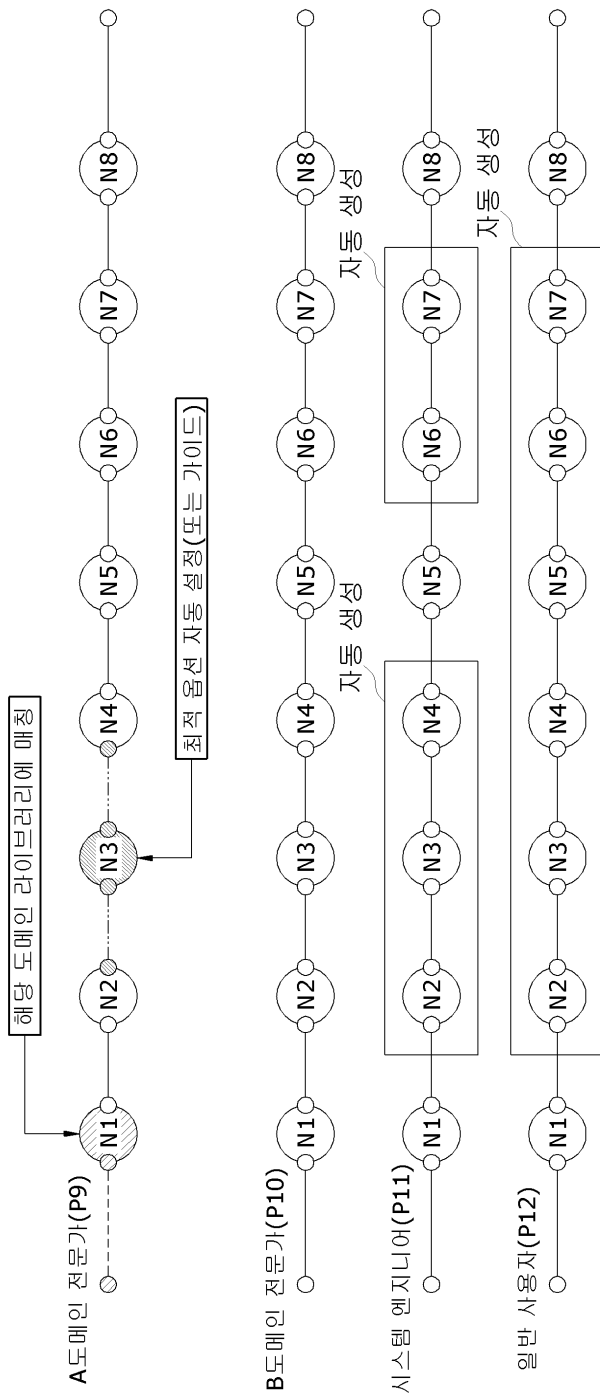
도면5



도면6

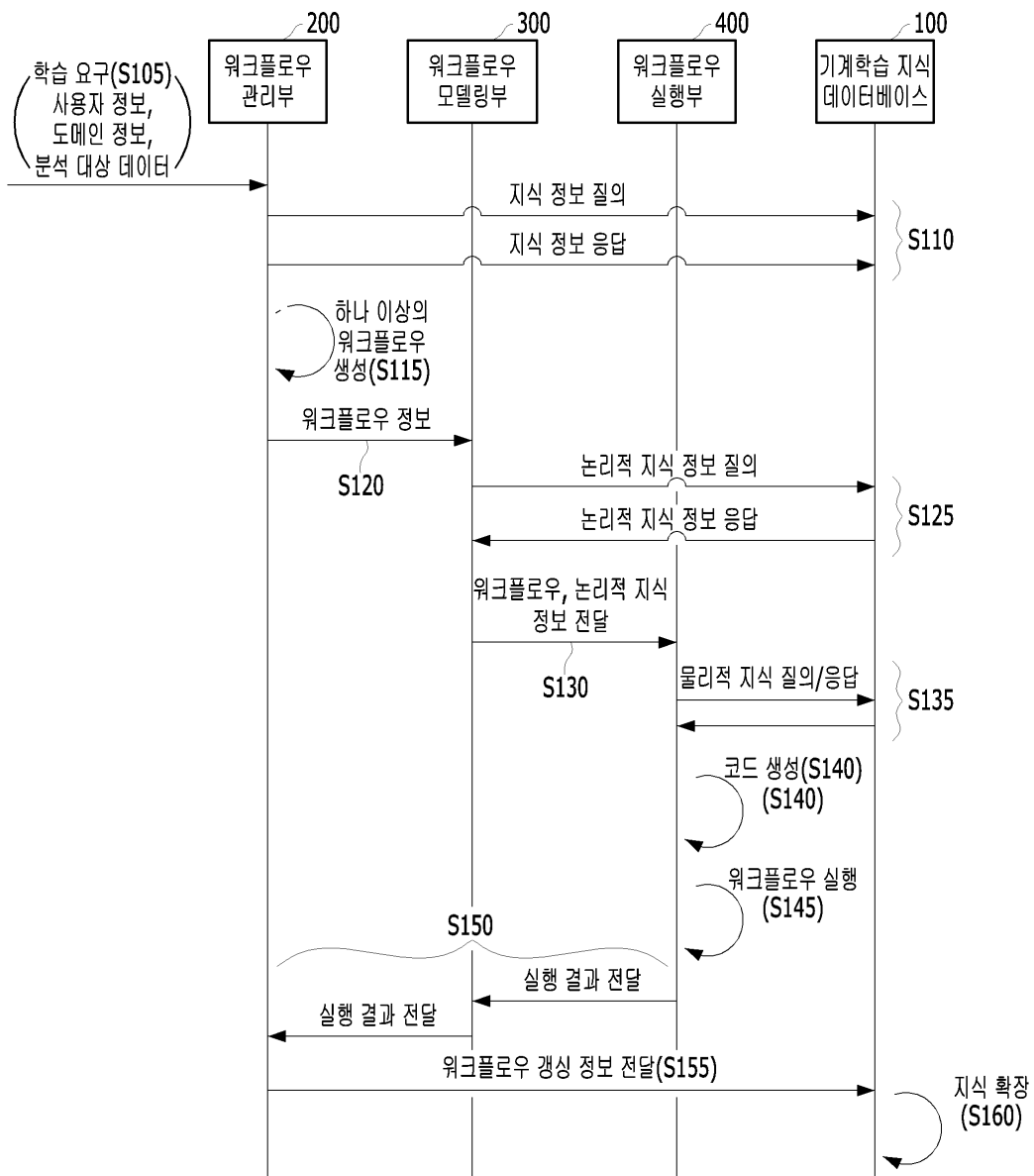


도면7





도면8



도면9

