



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2021년04월16일
(11) 등록번호 10-2241311
(24) 등록일자 2021년04월12일

(51) 국제특허분류(Int. Cl.)
G06N 3/08 (2006.01)

(52) CPC특허분류
G06N 3/08 (2013.01)

(21) 출원번호 10-2019-0073341

(22) 출원일자 2019년06월20일
심사청구일자 2019년06월20일

(65) 공개번호 10-2020-0145002

(43) 공개일자 2020년12월30일

(56) 선행기술조사문헌

JP2017182483 A*

KR101605078 B1*

KR1020100112742 A*

KR1020190060021 A*

*는 심사관에 의하여 인용된 문헌

(73) 특허권자

경상국립대학교산학협력단

경상남도 진주시 진주대로 501 (가좌동)

(72) 발명자

박균열

서울특별시 광진구 광나루로56길 32, 202동 202호(구의동, 구의현대2단지아파트)

(74) 대리인

김태현

전체 청구항 수 : 총 10 항

심사관 : 송근배

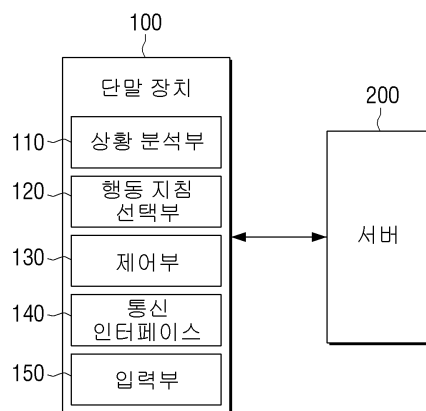
(54) 발명의 명칭 인공지능의 자율적 도덕 판단 및 수행을 위한 시스템

(57) 요약

인공지능의 자율적 도덕 판단 및 수행을 위한 시스템이 개시된다. 본 개시는 각각의 개인정보 및 데이터를 서버로 전송하는 단말 장치 및 단말 장치로부터 개인정보 및 데이터를 수신하는 서버를 포함하고, 서버는 데이터로부터 키워드 및 성향 관련 정보를 추출하고, 추출된 모든 키워드와 성향 관련 정보에 기초하여 범용 정체성 학습 데이터를 생성하며, 생성된 범용 정체성 학습 데이터에 기초하여 범용 정체성 딥러닝 학습 모델을 생성하고, 개인정보에 기초하여 복수의 그룹을 생성하며, 생성된 복수의 그룹 각각과 관련된 그룹 정체성 학습 데이터를 생성하고, 생성된 그룹 정체성 학습 데이터에 기초하여 복수의 그룹 각각에 대한 그룹 정체성 딥러닝 학습 모델을 생성하며, 생성된 딥러닝 학습 모델 중 사용자 요청에 대응되는 하나의 딥러닝 학습 모델을 단말 장치로 전송할 수 있다.

대표도 - 도1

1000



명세서

청구범위

청구항 1

각각의 개인식별 데이터, 개인성향 데이터 및 규칙 데이터를 서버로 전송하는 단말 장치; 및

상기 단말 장치로부터 상기 개인식별 데이터, 상기 개인성향 데이터 및 상기 규칙 데이터를 수신하는 서버;를 포함하고,

상기 서버는,

상기 수신된 개인 성향 데이터에 기초하여 키워드 및 성향 관련 정보를 추출하고, 상기 추출된 키워드 및 성향 관련 정보에 기초하여 범용 정체성 학습 데이터를 생성하고, 상기 규칙 데이터에 기초하여 특정 상황에 대응되는 행동 지침을 포함하는 행동 지침 학습 데이터를 생성하고, 상기 생성된 범용 정체성 학습 데이터 및 상기 생성된 행동 지침 학습 데이터에 기초하여 자율적 도덕 판단에 따른 행동 지침을 출력하는 범용 정체성 딥러닝 학습 모델을 생성하고,

상기 개인식별 데이터에 기초하여 복수의 그룹을 생성하며, 상기 생성된 복수의 그룹 각각과 관련된 그룹 정체성 학습 데이터를 생성하고, 상기 생성된 그룹 정체성 학습 데이터 및 상기 생성된 행동 지침 학습 데이터에 기초하여 자율적 도덕 판단에 따른 행동 지침을 출력하는 상기 복수의 그룹 각각에 대한 그룹 정체성 딥러닝 학습 모델을 생성하며,

상기 생성된 범용 정체성 딥러닝 학습 모델 및 상기 복수의 그룹 각각에 대한 그룹 정체성 딥러닝 학습 모델 중 사용자 요청에 대응되는 하나의 딥러닝 학습 모델을 상기 단말 장치로 전송하고,

상기 단말 장치는,

상기 서버로부터 상기 딥러닝 학습 모델 또는 그룹 정체성 딥러닝 학습 모델을 수신하고,

주변의 상황 및 환경에 대한 정보를 센싱하고,

상기 수신된 상기 딥러닝 학습 모델 또는 그룹 정체성 딥러닝 학습 모델에 기초하여 상기 센싱된 주변의 상황 및 환경에 대한 정보에 대응되는 행동 지침을 획득하고,

상기 획득된 행동 지침에 대응되는 동작을 수행하는, 인공지능의 자율적 도덕 판단 및 수행을 위한 시스템.

청구항 2

제1항에 있어서,

상기 복수의 그룹은,

인종, 국가, 연령, 종교, 이념, 성별을 포함하는 상기 개인식별 데이터 중 적어도 하나에 기초하여 생성되는, 인공지능의 자율적 도덕 판단 및 수행을 위한 시스템.

청구항 3

제1항에 있어서,

상기 행동 지침 학습 데이터는,

법률, 관습, 조리 및 복수의 윤리 규칙을 포함하는, 인공지능의 자율적 도덕 판단 및 수행을 위한 시스템.

청구항 4

제1항에 있어서,

상기 단말 장치는,

상기 획득된 행동 지침이 복수 개이면, 기 결정된 가치 위계의 점수에 따라 하나의 행동 지침을 식별하고,

상기 식별된 행동 지침에 대응되는 동작을 수행하는, 인공지능의 자율적 도덕 판단 및 수행을 위한 시스템.

청구항 5

제4항에 있어서,

상기 단말 장치는,

사용자 명령이 입력되는 경우, 사용자의 행동, 표정, 음성 및 상기 음성에 포함된 단어 중 적어도 하나를 분석하여 사용자의 감정, 심박수, 활동성, 신체상태 중 적어도 하나를 포함하는 사용자 상태를 파악하고,

상기 파악된 사용자 상태, 상기 주변의 상황 및 환경에 대하여 분석된 정보를 상기 사용자 명령과 매칭하여 경험적 행동 지침을 형성하고,

상기 형성된 경험적 행동 지침 및 상기 수신된 딥러닝 학습 모델에 기초하여 상기 행동 지침을 획득하는, 인공지능의 자율적 도덕 판단 및 수행을 위한 시스템.

청구항 6

단말 장치가 각각의 개인식별 데이터, 개인성향 데이터 및 규칙 데이터를 서버로 전송하는 단계;

상기 서버가 상기 단말 장치로부터 상기 개인식별 데이터, 상기 개인성향 데이터 및 상기 규칙 데이터를 수신하는 단계;

상기 서버가 상기 수신된 개인 성향 데이터로부터 키워드 및 성향 관련 정보를 추출하고, 상기 추출된 키워드 및 성향 관련 정보에 기초하여 범용 정체성 학습 데이터를 생성하는 단계;

상기 서버가 상기 규칙 데이터에 기초하여 특정 상황에 대응되는 행동 지침을 포함하는 행동 지침 학습 데이터를 생성하는 단계;

상기 서버가 상기 생성된 범용 정체성 학습 데이터 및 상기 생성된 행동 지침 학습 데이터에 기초하여 자율적 도덕 판단에 따른 행동 지침을 출력하는 범용 정체성 딥러닝 학습 모델을 생성하는 단계;

상기 서버가 상기 수신된 개인식별 데이터에 기초하여 복수의 그룹을 생성하는 단계;

상기 서버가 상기 생성된 복수의 그룹 각각과 관련된 그룹 정체성 학습 데이터를 생성하는 단계;

상기 서버가 상기 생성된 그룹 정체성 학습 데이터 및 상기 생성된 행동 지침 학습 데이터에 기초하여 자율적 도덕 판단에 따른 행동 지침을 출력하는 상기 복수의 그룹 각각에 대한 그룹 정체성 딥러닝 학습 모델을 생성하는 단계;

상기 서버가 상기 생성된 범용 정체성 딥러닝 학습 모델 및 상기 복수의 그룹 각각에 대한 그룹 정체성 딥러닝 학습 모델 중 사용자 요청에 대응되는 하나의 딥러닝 학습 모델을 상기 단말 장치로 전송하는 단계;

상기 단말 장치가 상기 서버로부터 상기 딥러닝 학습 모델 또는 그룹 정체성 딥러닝 학습 모델을 수신하는 단계;

상기 단말 장치가 주변의 상황 및 환경에 대한 정보를 센싱하는 단계;

상기 단말 장치가 상기 수신된 상기 딥러닝 학습 모델 또는 그룹 정체성 딥러닝 학습 모델에 기초하여 상기 센싱된 주변의 상황 및 환경에 대한 정보에 대응되는 행동 지침을 획득하는 단계; 및

상기 단말 장치가 상기 획득된 행동 지침에 대응되는 동작을 수행하는 단계;를 포함하는, 인공지능의 자율적 도덕 판단 및 수행 방법.

청구항 7

제6항에 있어서,

상기 복수의 그룹을 생성하는 단계는,

인종, 국가, 연령, 종교, 이념, 성별을 포함하는 상기 개인식별 데이터 중 적어도 하나에 기초하여 생성되는, 인공지능의 자율적 도덕 판단 및 수행 방법.

청구항 8

제6항에 있어서,
 상기 행동 지침 학습 데이터는,
 법률, 관습, 조리 및 복수의 윤리 규칙을 포함하는, 인공지능의 자율적 도덕 판단 및 수행 방법.

청구항 9

제6항에 있어서,
 상기 단말 장치가, 상기 획득된 행동 지침이 복수 개이면, 기 결정된 가치 위계의 점수에 따라 하나의 행동 지침을 식별하는 단계; 및
 상기 단말 장치가 상기 식별된 행동 지침에 대응되는 동작을 수행하는 단계;를 더 포함하는, 인공지능의 자율적 도덕 판단 및 수행 방법.

청구항 10

제9항에 있어서,
 상기 단말 장치가, 사용자 명령이 입력되는 경우, 사용자의 행동, 표정, 음성 및 상기 음성에 포함된 단어 중 적어도 하나를 분석하여 사용자의 감정, 심박수, 활동성, 신체상태 중 적어도 하나를 포함하는 사용자 상태를 파악하는 단계,
 상기 단말 장치가 상기 파악된 사용자 상태, 상기 주변의 상황 및 환경에 대하여 분석된 정보를 상기 사용자 명령과 매칭하여 경험적 행동 지침을 형성하는 단계; 및,
 상기 단말 장치가 상기 형성된 경험적 행동 지침 및 상기 수신된 딥러닝 학습 모델에 기초하여 상기 행동 지침을 획득하는 단계;를 더 포함하는, 인공지능의 자율적 도덕 판단 및 수행 방법.

발명의 설명

기술 분야

[0001] 본 개시는 인공지능의 자율적 도덕 판단 및 수행을 위한 시스템에 관한 것으로, 더욱 상세하게는 심화학습(딥러닝, deep learning) 등의 기계 학습 알고리즘을 활용하여 인간 두뇌의 인지, 정서, 행동 등의 기능을 학습하고, 학습된 기능을 기초로 윤리적 제어 메커니즘을 수행하는 인공지능(Artificial Intelligence, AI) 시스템 및 그 응용 기술에 관련된 것이다.

배경 기술

[0002] 근래에는 인간 수준의 지능을 구현하는 인공 지능 시스템이 다양한 분야에서 이용되고 있다. 인공 지능 시스템은 사용할수록 인식률이 향상되고 사용자 취향을 보다 정확하게 이해할 수 있게 발전되고 있다.

[0003] 인공 지능 기술은 기계학습(예로, 딥러닝) 및 기계학습을 활용한 요소 기술들로 구성된다.

[0004] 기계학습은 입력 데이터들의 특징을 스스로 분류/학습하는 알고리즘 기술이며, 요소 기술은 딥러닝 등의 기계학습 알고리즘을 활용하여 인간 두뇌의 인지, 판단 등의 기능을 모사하는 기술로서, 언어적 이해, 시각적 이해, 추론/예측, 지식 표현, 동작 제어 등의 기술 분야로 구성된다.

[0005] 인공 지능 기술이 응용되는 다양한 분야는 다음과 같다. 언어적 이해는 인간의 언어/문자를 인식하고 응용/처리하는 기술로서, 자연어 처리, 기계 번역, 대화 시스템, 질의 응답, 음성 인식/합성 등을 포함한다. 시각적 이해는 사물을 인간의 시각처럼 인식하여 처리하는 기술로서, 객체 인식, 객체 추적, 영상 검색, 사람 인식, 장면 이해, 공간 이해, 영상 개선 등을 포함한다. 추론 예측은 정보를 판단하여 논리적으로 추론하고 예측하는 기술로서, 지식/확률 기반 추론, 최적화 예측, 선호 기반 계획, 추천 등을 포함한다. 지식 표현은 인간의 경험정보를 지식데이터로 자동화 처리하는 기술로서, 지식 구축(데이터 생성/분류), 지식 관리(데이터 활용) 등을 포함한다. 동작 제어는 차량의 자율 주행, 로봇의 움직임을 제어하는 기술로서, 움직임 제어(항법, 충돌, 주행), 조

작 제어(행동 제어) 등을 포함한다.

발명의 내용

해결하려는 과제

[0006] 본 개시의 목적은 AI가 자율적 도덕 판단을 할 수 있도록 제어 역량을 갖출 수 있는 기초적인 판단 근거와 자체적인 심화학습을 할 수 있는 범주와 핵심 역량을 제공함에 있다.

과제의 해결 수단

[0007] 본 개시의 일 실시 예에 따른 인공지능의 자율적 도덕 판단 및 수행을 위한 시스템은, 각각의 개인정보 및 데이터를 서버로 전송하는 단말 장치 및 상기 단말 장치로부터 개인정보 및 데이터를 수신하는 서버를 포함하고, 상기 서버는 상기 데이터로부터 키워드 및 성향 관련 정보를 추출하고, 상기 추출된 모든 키워드와 성향 관련 정보에 기초하여 범용 정체성 학습 데이터를 생성하며, 상기 생성된 범용 정체성 학습 데이터에 기초하여 범용 정체성 딥러닝 학습 모델을 생성하고, 상기 개인정보에 기초하여 복수의 그룹을 생성하며, 상기 생성된 복수의 그룹 각각과 관련된 그룹 정체성 학습 데이터를 생성하고, 상기 생성된 그룹 정체성 학습 데이터에 기초하여 상기 복수의 그룹 각각에 대한 그룹 정체성 딥러닝 학습 모델을 생성하며, 상기 생성된 범용 정체성 딥러닝 학습 모델 및 상기 복수의 그룹 각각에 대한 그룹 정체성 딥러닝 학습 모델 중 사용자 요청에 대응되는 하나의 딥러닝 학습 모델을 상기 단말 장치로 전송할 수 있다.

[0008] 그리고, 상기 복수의 그룹은 인종, 국가, 연령, 종교, 이념, 성별을 포함하는 상기 개인정보 중 적어도 하나에 기초하여 생성될 수 있다.

[0009] 그리고, 상기 범용 정체성 딥러닝 학습 모델은 법률, 관습, 조리 및 복수의 윤리 규칙을 포함하는 행동 지침 학습 데이터 및 상기 범용 정체성 학습 데이터를 기초로 생성되고, 상기 그룹 정체성 딥러닝 학습 모델은 법률, 관습, 조리 및 복수의 윤리 규칙을 포함하는 행동 지침 학습 데이터 및 상기 그룹 정체성 학습 데이터를 기초로 생성될 수 있다.

[0010] 또한, 상기 단말 장치는 주변의 상황 및 환경에 대한 정보를 센싱하고, 상기 센싱된 정보를 분석하는 상황 분석부, 상기 수신된 딥러닝 학습 모델에 기초하여 상기 분석된 정보에 대응되는 행동 지침을 선택하는 행동 지침 선택부, 및 상기 선택된 행동 지침을 수행하도록 상기 단말 장치를 제어하는 제어부를 더 포함할 수 있다.

[0011] 그리고, 상기 상황 분석부는 사용자 명령이 입력되는 경우, 사용자의 행동, 표정, 음성 및 상기 음성에 포함된 단어 중 적어도 하나를 분석하여 사용자의 감정, 심박수, 활동성, 신체상태 중 적어도 하나를 포함하는 사용자 상태를 파악하고, 상기 파악된 사용자 상태, 상기 주변의 상황 및 환경에 대하여 분석된 정보를 상기 사용자 명령과 매칭하여 경험적 행동 지침을 형성하고, 상기 행동 지침 선택부는 상기 형성된 경험적 행동 지침 및 상기 수신된 딥러닝 학습 모델에 기초하여 상기 행동 지침을 선택할 수 있다.

[0012] 본 개시의 일 실시 예에 따른 인공지능의 자율적 도덕 판단 및 수행 방법은 단말장치가 각각의 개인정보 및 데이터를 서버로 전송하는 단계, 서버가 상기 단말 장치로부터 개인정보 및 데이터를 수신하는 단계, 서버가 상기 수신된 데이터로부터 키워드 및 성향 관련 정보를 추출하고, 상기 추출된 모든 키워드와 성향 관련 정보에 기초하여 범용 정체성 학습 데이터를 생성하는 단계, 서버가 상기 생성된 범용 정체성 학습 데이터에 기초하여 범용 정체성 딥러닝 학습 모델을 생성하는 단계, 서버가 상기 수신된 개인정보에 기초하여 복수의 그룹을 생성하는 단계, 서버가 상기 생성된 복수의 그룹 각각과 관련된 그룹 정체성 학습 데이터를 생성하는 단계, 서버가 상기 생성된 그룹 정체성 학습 데이터에 기초하여 상기 복수의 그룹 각각에 대한 그룹 정체성 딥러닝 학습 모델을 생성하는 단계 및 서버가 상기 생성된 범용 정체성 딥러닝 학습 모델 및 상기 복수의 그룹 각각에 대한 그룹 정체성 딥러닝 학습 모델 중 사용자 요청에 대응되는 하나의 딥러닝 학습 모델을 상기 단말 장치로 전송하는 단계를 포함할 수 있다.

[0013] 그리고, 상기 복수의 그룹을 생성하는 단계는 인종, 국가, 연령, 종교, 이념, 성별을 포함하는 상기 개인정보 중 적어도 하나에 기초하여 생성될 수 있다.

[0014] 그리고, 상기 범용 정체성 딥러닝 학습 모델을 생성하는 단계는 법률, 관습, 조리 및 복수의 윤리 규칙을 포함하는 행동 지침 학습 데이터 및 상기 범용 정체성 학습 데이터를 기초로 생성되고, 상기 그룹 정체성 딥러닝 학습 모델을 생성하는 단계는, 법률, 관습, 조리 및 복수의 윤리 규칙을 포함하는 행동 지침 학습 데이터 및 상기

그룹 정체성 학습 데이터를 기초로 생성될 수 있다.

[0015] 그리고, 주변의 상황 및 환경에 대한 정보를 센싱하는 단계, 상황 분석부가 상기 센싱된 정보를 분석하는 상황 분석 단계, 행동 지침 선택부가 상기 수신된 딥러닝 학습 모델에 기초하여 상기 분석된 정보에 대응되는 행동 지침을 선택하는 단계 및 제어부가 상기 선택된 행동 지침을 수행하도록 상기 단말 장치를 제어하는 단계를 더 포함할 수 있다.

[0016] 그리고, 상기 상황 분석 단계는 사용자 명령이 입력되는 경우, 사용자의 행동, 표정, 음성 및 상기 음성에 포함된 단어 중 적어도 하나를 분석하여 사용자의 감정, 심박수, 활동성, 신체상태 중 적어도 하나를 포함하는 사용자 상태를 파악하고, 상기 파악된 사용자 상태, 상기 주변의 상황 및 환경에 대하여 분석된 정보를 상기 사용자 명령과 매칭하여 경험적 행동 지침을 형성하고, 상기 행동 지침을 선택하는 단계는 상기 형성된 경험적 행동 지침 및 상기 수신된 딥러닝 학습 모델에 기초하여 상기 행동 지침을 선택할 수 있다.

발명의 효과

[0017] 본 개시의 다양한 실시 예에 의하면, 인공지능의 자율적 도덕 판단 및 수행에 있어서 윤리적 제어 메커니즘을 구현할 수 있다. 구체적으로 법률, 관습, 조리 및 복수의 윤리 규칙을 포함하는 행동 지침을 기초로 7가지 가치 위계와 그에 대한 지향 정보에 점수를 부여하는 방식으로 윤리적 제어 메커니즘을 형성할 수 있다.

[0018] 이에 따라, 윤리적 제어 메커니즘이 적용된 인공지능의 자율적 도덕 판단 및 수행이 가능할 수 있다.

도면의 간단한 설명

[0019] 도 1은 본 개시의 일 실시 예에 따른 인공지능의 자율적 도덕 판단 및 수행을 위한 시스템을 도시한 블록도이다.

도 2는 본 개시의 일 실시 예에 따른 단말 장치의 상황 분석부를 도시한 블록도이다.

도 3은 본 개시의 일 실시 예에 따른 범용 정체성 딥러닝 학습 모델이 형성되는 과정이 도시된 흐름도이다.

도 4는 본 개시의 일 실시 예에 따른 그룹 정체성 딥러닝 학습 모델이 형성되는 과정이 도시된 흐름도이다.

도 5는 본 개시의 일 실시 예에 따른 행동 지침이 선택 및 수행되는 과정이 도시된 흐름도이다.

도 6은 본 개시의 다른 실시 예에 따른 행동 지침이 선택 및 수행되는 과정이 도시된 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0020] 이하에서는 본 개시의 구성 및 효과를 충분히 이해하기 위하여, 첨부한 도면을 참조하여 본 개시의 바람직한 실시 예들을 상세하게 설명한다. 그러나 본 개시는 이하에서 개시되는 실시 예들에 한정되는 것이 아니라, 여러 가지 형태로 구현될 수 있고 다양한 변경을 가할 수 있다. 도면의 설명과 관련하여, 유사한 구성요소에 대해서는 유사한 참조 부호가 사용될 수 있다.

[0021] 제1, 제2 등과 같이 서수를 포함하는 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 구성요소들은 용어들에 의해 한정되지는 않는다. 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다. 예를 들어, 본 개시의 권리 범위를 벗어나지 않으면서 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소도 제1 구성요소로 명명될 수 있다. 및/또는 이라는 용어는 복수의 관련된 항목들의 조합 또는 복수의 관련된 항목들 중의 어느 하나의 항목을 포함한다.

[0022] 본 명세서에서 사용한 용어는 실시 예를 설명하기 위해 사용된 것으로, 본 개시를 제한 및/또는 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, 포함하다 또는 가지다 등의 용어는 명세서상에 기재된 특징, 숫자, 동작, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 동작, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0023] 실시 예에 있어서 '모듈' 혹은 '부'는 적어도 하나의 기능이나 동작을 수행하며, 하드웨어 또는 소프트웨어로 구현되거나 하드웨어 또는 소프트웨어의 결합으로 구현될 수 있다. 또한, 복수의 '모듈' 혹은 복수의 '부'는 특정한 하드웨어로 구현될 필요가 있는 '모듈' 혹은 '부'를 제외하고는 적어도 하나의 모듈로 일체화되어 적어도 하나의 프로세서로 구현될 수 있다.

- [0024] 도 1은 본 개시의 일 실시 예에 따른, 인공지능의 자율적 도덕 판단 및 수행을 위한 시스템을 도시한 블록도이다. 도 1에 도시된 바와 같이, 인공지능의 자율적 도덕 판단 및 수행을 위한 시스템(1000)은 단말 장치(100) 및 서버(200)를 포함할 수 있다.
- [0025] 도 1에 도시된 구성들은 본 개시의 실시 예를 구현하기 위한 예시도이며, 당업자에게 자명한 수준의 적절한 하드웨어/소프트웨어 구성들이 인공지능의 자율적 도덕 판단 및 수행을 위한 시스템(1000)에 추가로 포함될 수 있다.
- [0026] 인공지능의 자율적 도덕 판단 및 수행을 위한 시스템(1000)의 단말 장치(100)는 개인정보 및 데이터를 획득하여 서버(200)로 전송할 수 있고, 단말 장치(100)는 서버(200)에서 수신된 개인정보 및 데이터에 기초하여 생성된 딥러닝 학습 모델 중 사용자 요청에 대응되는 딥러닝 학습 모델을 수신할 수 있다. 단말 장치(100)는 주변의 상황 및 환경에 대한 정보를 센싱하고, 센싱된 정보를 분석할 수 있다. 그리고 단말 장치(100)는 서버로부터 수신된 딥러닝 학습 모델을 기초로, 분석된 정보에 대응되는 행동 지침을 선택하고, 선택된 행동 지침을 수행할 수 있다. 또한, 단말 장치(100)는 사용자 명령을 입력받고 사용자 명령에 대응되는 행동 지침을 수행할 수 있다.
- [0027] 본 개시의 일 실시 예에 따른 단말 장치(100)는 상황 분석부(110), 행동 지침 선택부(120), 제어부(130), 통신 인터페이스(140) 및 입력부(150)를 포함할 수 있다.
- [0028] 도 1에 도시된 구성들은 본 개시의 실시 예를 구현하기 위한 예시도이며, 당업자에게 자명한 수준의 적절한 하드웨어/소프트웨어 구성들이 단말 장치(100)에 추가로 포함될 수 있다.
- [0029] 단말 장치(100)는 개인정보 및 데이터를 획득하고, 획득한 각각의 개인정보 및 데이터를 서버로 전송할 수 있다. 단말 장치(100)는 개인정보 및 데이터를 오프라인 상으로 발행되는 간행물에 게재되어 있는 것뿐만 아니라, 온라인 상에 게시된 뉴스, 신문, SNS, 댓글, 칼럼, 블로그 등으로부터 획득할 수 있다.
- [0030] 여기서, 개인정보는 인종, 국가, 연령, 종교, 이념, 성별 등을 포함할 수 있고, 이외에도 개인을 특정할 수 있는 주소, 가족관계, 학력, 혈액형 등을 포함할 수 있다. 또한, 데이터는 개인의 성향을 파악할 수 있는 단어, 의견 및 진술을 포함할 수 있고, 사회의 법률, 관습, 조리 및 복수의 윤리 규칙을 포함할 수 있다.
- [0031] 단말 장치(100)는 입력부(150)를 포함할 수 있다. 단말 장치(100)는 입력부(150)를 이용하여 개인정보 및 데이터를 사용자로부터 입력 받을 수 있다. 여기서 입력되는 개인정보 및 데이터는 오프라인 상으로 발행되는 간행물에 게재되어 있는 것뿐만 아니라, 사용자가 단말 장치에서 제시하는 설문조사에 응답하여 개인정보 및 데이터를 제공하거나, 사용자의 도덕적 가치 지향 정도를 판단할 수 있는 문제를 해결하는 방법으로 개인정보 및 데이터를 제공할 수 있다.
- [0032] 또한, 입력부(150)는 사용자로부터 단말 장치 및 주변 장치를 제어하기 위한 사용자 명령을 입력 받을 수 있다. 예를 들어, 입력부(150)는 사용자의 물리적 명령을 인식할 수 있는 터치 센서, 압력 센서를 포함하는 물리적 버튼을 포함할 수 있고, 사용자의 모션 및 주변의 상황과 환경을 촬영할 수 있는 카메라, 사용자의 발화 및 주변의 상황과 환경을 입력 받을 수 있는 마이크 등을 포함할 수 있다. 이러한 입력부(150)의 예는 일 실시 예에 불과할 뿐, 다양한 전자 장치로 구현될 수 있다.
- [0033] 단말 장치(100)는 통신 인터페이스(140)를 포함할 수 있다. 단말 장치(100)는 통신 인터페이스(140)를 이용하여 획득한 개인정보 및 데이터를 서버로 전송할 수 있다. 통신 인터페이스(140)는 다양한 유형의 통신방식에 따라 다양한 유형의 외부 기기와 통신을 수행할 수 있다. 또한, 단말 장치(100)는 통신 인터페이스(140)를 이용하여 외부 기기, 주변의 IoT 장치 및 전자 장치를 제어할 수 있다.
- [0034] 여기서 통신 인터페이스(140)는 와이파이칩, 블루투스 칩, 무선 통신 칩, NFC 칩을 포함할 수 있다. 단말 장치(100)는 통신 인터페이스(140)를 이용하여 외부 서버 또는 각종 외부 기기와 통신을 수행할 수 있다.
- [0035] 단말 장치(100)는 개인정보 및 데이터를 서버(200)로 전송하고, 서버(200)는 단말 장치(100)로부터 수신한 개인정보 및 데이터를 기초로 정체성 학습 데이터 및 딥러닝 학습 모델을 생성한다. 서버(200)가 생성하는 정체성 학습 데이터 및 딥러닝 학습 모델의 생성에 대한 구체적인 설명은 후술한다.
- [0036] 단말 장치(100)는 서버(200)에서 생성된 범용 정체성 딥러닝 학습 모델 또는 복수의 그룹 각각에 대한 그룹 정체성 딥러닝 학습 모델 중 사용자의 요청에 대응되는 하나의 딥러닝 학습 모델을 서버(200)로부터 수신할 수 있다.
- [0037] 단말 장치(100)는 저장부(미도시)를 포함할 수 있는데, 저장부에는 서버로부터 수신한 딥러닝 학습 모델이 저장

될 수 있고, 딥러닝 학습 모델에는 주어진 특정 상황에서의 행동 지침이 포함될 수 있다. 또한, 저장부는 도 2에서 후술하는 상황 분석부(110)로부터 생성된 경험적 행동 지침을 저장할 수 있다.

- [0038] 단말 장치(100)는 상황 분석부(110)를 포함할 수 있다. 상황 분석부(110)는 학습부(미도시) 및 인식부(미도시) 중 적어도 하나를 포함할 수 있고, 또는 서버(200)에 학습부가 포함되어 있는 경우, 인식부는 단말 장치(100)에 포함되어 구현될 수 있다. 학습부는 소정의 상황 판단을 위한 기준을 갖는 인식 모델을 생성 또는 학습시킬 수 있다. 학습부는 수집된 학습 데이터를 이용하여 판단 기준을 갖는 인식 모델을 생성할 수 있다.
- [0039] 일 예로, 학습부는 객체가 포함된 이미지를 학습 데이터로서 이용하여 이미지에 포함된 객체가 어떤 것인지 또는 어떠한 동작을 취하는지 여부를 판단하는 기준을 갖는 객체 인식 모델을 생성, 학습 또는 갱신시킬 수 있다.
- [0040] 또한, 상황 분석부(110)는 상황 분석부에 포함된 각종 모듈을 이용하여 주변의 상황 및 환경을 인식하고 분석할 수 있다. 구체적인 과정은 도 2에서 후술한다.
- [0041] 단말 장치(100)는 행동 지침 선택부(120) 및 제어부(13)를 포함할 수 있는 데, 단말 장치(100)는 행동 지침 선택부(120) 및 제어부(13)를 이용하여 주변의 상황 및 환경에 따라 행동 지침을 선택하고, 선택된 행동 지침을 수행할 수 있다.
- [0042] 행동 지침 선택부(120)는 서버(200)로부터 수신한 딥러닝 학습 모델과 상황 분석부(110)를 기초하여 분석된 주변의 상황 및 환경에 대응되는 행동 지침을 선택할 수 있다. 여기서 행동 지침이란, 특정 상황에서 법률, 관습, 조리 및 복수의 윤리 규칙에 따라 생성된 행동 지침 학습 데이터에 따라 결정된 것으로, 예컨대 A라는 상황에 적절한 B라는 행동을 수행하도록 설정된 것을 의미한다. 따라서 분석된 주변의 상황 및 환경에 대응되는 행동 지침은 복수개 존재할 수 있고, 행동 지침이 존재하지 않을 수 있다. 분석된 주변의 상황 및 환경에서 수행할 수 있는 복수개의 행동 지침이 존재하는 경우는 도 5에서, 행동 지침이 존재하지 않은 경우는 도 6에서 구체적으로 후술한다. 또한, 상황 분석부(110)가 주변의 상황 및 환경에 대한 정보를 분석하는 방법 및 딥러닝 학습 모델을 생성하는 방법에 대하여는 후술한다.
- [0043] 제어부(130)는 단말 장치(100)의 전반적인 동작을 제어할 수 있다. 구체적으로, 제어부(130)는 단말 장치(100)의 입력부(150)로 사용자의 발화 또는 주변의 음성이 입력되는 경우, 상기 입력을 상황 분석부로 전송하여 처리할 수 있도록 하고, 행동 지침 선택부(120)가 행동 지침을 선택할 수 없는 경우 또는 수행할 수 있는 행동 지침이 없는 경우 단말 장치(100)를 제어하여 사용자에게 알릴 수 있다.
- [0044] 또한, 제어부(130)는 행동 지침 선택부(120)를 기초하여 특정 행동 지침이 선택되는 경우, 이를 수행될 수 있도록 단말 장치(100)를 제어할 수 있고, 제어부(130)는 통신 인터페이스(140)를 이용하여 외부 기기, 주변의 IoT 장치 및 전자 장치를 제어할 수 있다.
- [0045] 또한, 제어부(130)는 서버(200)로부터 수신한 딥러닝 학습 모델을 저장부(미도시)에 저장할 수 있고, 상황 분석부(110)로부터 생성된 경험적 행동 지침을 저장부에 저장할 수 있다.
- [0046] 여기서 제어부(130)는 RAM(미도시), ROM, 그래픽 처리부, 메인 CPU, 그래픽 처리부를 포함할 수 있다.
- [0047] 인공지능의 자율적 도덕 판단 및 수행을 위한 시스템(1000)은 머신 러닝 기반의 시스템에 의해 수행될 수 있으며, 본 개시에서는 뉴럴 네트워크(Neural Networks)에 기반한 일련의 기계학습 알고리즘에 의한 분류 시스템으로서, 딥러닝 기반의 시스템을 예로서 설명한다.
- [0048] 인공지능의 자율적 도덕 판단 및 수행을 위한 시스템(1000)의 서버(200)는 적어도 하나의 분류기(미도시)를 포함할 수 있으며, 분류기는 하나 또는 복수 개의 프로세서(미도시)에 해당할 수 있다. 프로세서는 다수의 논리 게이트들의 어레이(Array)로 구현될 수 있고, 범용적인 마이크로 프로세서와 이 마이크로 프로세서에서 실행될 수 있는 프로그램이 저장된 메모리의 조합으로 구현될 수도 있다.
- [0049] 분류기는 뉴럴 네트워크(Neural Network) 기반 분류기, SVM(Support Vector Machine), 에이다부스트 분류기(Adaboost Classifier), 베이시안 분류기(Bayesian Classifier) 및 퍼셉트론 분류기(Perceptron Classifier) 등으로 구현될 수 있다. 이하, 본 개시의 분류기는 컨볼루션 뉴럴 네트워크(Convolutional Neural Network, CNN) 기반 분류기로 구현되는 실시 예에 대하여 설명한다. 뉴럴 네트워크 기반 분류기는, 연결선으로 연결된 많은 수의 인공 뉴런들을 이용하여 생물학적인 시스템의 계산 능력을 모방하도록 구현된 연산모델로서, 연결 강도(가중치)를 갖는 연결선을 통해 인간의 인지작용이나 학습과정을 수행하게 된다. 그러나, 본 개시의 분류기기에 한정되는 것은 아니며, 상술한 다양한 분류기로 구현될 수 있음은 물론이다.

- [0050] 서버(200)는 단말 장치(100)로부터 개인정보 및 데이터를 수신하고, 수신된 개인정보 및 데이터를 기초로 범용 정체성 학습 데이터 및 그룹 정체성 학습 데이터, 행동 지침 학습 데이터를 생성하며, 생성된 학습 데이터들을 기초로 학습하여 정체성 딥러닝 학습 모델을 생성할 수 있다.
- [0051] 구체적으로, 서버(200)는 적어도 하나의 분류기(미도시)를 포함할 수 있고, 분류기를 이용하여 수신된 데이터로부터 키워드 및 성향 관련 정보만을 추출할 수 있다.
- [0052] 여기서 수신된 데이터는 오프라인 상으로 발행되는 간행물에 게재되어 있는 것뿐만 아니라, 온라인 상에 게시된 뉴스, 신문, SNS, 댓글, 칼럼, 블로그 등으로부터 획득한 것으로, 개인의 성향을 파악할 수 있는 단어, 의견 및 진술을 포함할 수 있고, 사회의 법률, 관습, 조리 및 복수의 윤리 규칙을 포함할 수 있다.
- [0053] 여기서, 추출된 키워드 및 성향 관련 정보란, 단말 장치(100)로부터 수신된 데이터가 도덕 판단의 기반이 되는 제1 내지 제7 단계의 가치 위계에 관련한 단어 및 문장으로 분리된 것일 수 있다.
- [0054] 여기서 제1 내지 제7 단계의 가치 위계란 제1 단계인 복종 및 처벌 정향(orientation) 및 곤경회피(trouble-avoiding)의 태도, 제2 단계인 이기주의 성향 및 소박한 평등주의, 제3 단계인 대인 평판에 근거한 착한 아이 정향, 제4 단계인 권위존중과 작은 공동체의 질서 유지 중시 정향, 제5 단계인 공리주의적 이익의 중시와 사회적 합의 중시 정향, 제6 단계인 정의 및 인간의 존엄성 등 눈에 보이지 않는 자율적 양심 및 원리 정향, 그리고 제7 단계인 초월적이고 신비스러운 숭고성 정향으로 분류할 수 있다.
- [0055] 또한, 서버(200)는 추출된 모든 키워드와 성향 관련 정보를 이용하여 제1 내지 제7 단계의 가치 위계에 대응되도록 키워드와 성향 관련 정보를 분류할 수 있다.
- [0056] 서버(200)는 7개 단계의 가치 위계에 따라 분류된 키워드와 성향 관련 정보에 대하여 긍정형, 부정형, 중간형의 문장 또는 단어로 분류할 수 있고, 분류된 문장 또는 단어에 기초하여 선호 빈도에 따라 각 단계의 기초적인 선호 점수를 도출할 수 있다.
- [0057] 서버(200)가 도출하는 각 단계의 기초적인 선호 점수는 도덕 판단의 판별에 대한 것으로 여섯 가지의 절차로 나뉠 수 있다.
- [0058] 첫번째 절차는 제1 단계 내지 제7 단계에 이르는 각 단계를 어느 정도 각각 지향하는 지를 파악할 수 있다. 7개 단계의 가치 위계에 대해 상정한 키워드와 성향 관련 정보를 이용하여 추출된 키워드와 성향 관련 정보를 매칭하여 각 단계에 대하여 '어느 정도로 긍정 하는지' 또는 '어느 정도로 부정 하는지'와 관련된 지향 정도를 기설정된 계산자의 알고리즘에 따라 기초 점수를 도출할 수 있다.
- [0059] 두번째 절차는 첫번째 절차의 기초 점수를 바탕으로, 제4 단계보다 상대적으로 낮은 단계의 가치 위계에 대해서는 낮은 가중치를 부여하고, 상대적으로 높은 단계의 가치 위계에 대해서는 높은 가중치를 부여하여, 높은 단계가 낮은 단계보다 높은 점수가 부여될 수 있다. 높은 단계에 높은 점수가 부여된 경우 높은 도덕 판단 역량을 갖춘 것으로 파악할 수 있으며, 그 반대의 경우도 마찬가지로 논리에 따라 파악될 수 있다.
- [0060] 세번째 절차는 7개 단계의 가치 위계에 대해 상정한 키워드와 성향 관련 정보를 기초로하여, 특정 주제에 찬성할 때의 경향성과 반대할 때의 경향성이 일관되는 정도에 따라 다른 가중치를 부여할 수 있다. 즉, 일관성이 높으면 높은 가중치를, 일관성이 낮으면 낮은 가중치를 부여할 수 있다.
- [0061] 네번째 절차는 낮은 단계에서부터 높은 단계에 이르기까지 지향성이 점진적으로 증가하면 높은 가중치가 부여될 수 있다. 즉, 가까운 단계끼리 지향성이 유사하면 상대적으로 점수가 낮게 나타나고, 가까운 단계끼리 지향성이 차이가 나면, 그 중 높은 단계의 지향성이 상대적으로 더 높은 단계의 지향성보다 낮을 경우, 높은 점수를 나타낼 수 있다.
- [0062] 다섯번째 절차는 마지막 단계인 제7 단계보다 제1 단계의 점수가 더 높거나, 단계가 낮아질수록 지향 점수가 높아지는 단계 하향의 지향성을 보일 경우, 도덕 판단 능력이 전도된 것으로 간주하여, 기설정된 계산자에 의한 점수 및 각 단계의 가치 위계에 대한 척도 값을 별도 구분처리 할 수 있다.
- [0063] 여섯번째 절차는 각 단계의 가치 위계에 대한 경향성이 주제가 달라져도 일관되게 나타나는가를 기준으로 가중치를 부여할 수 있다. 예컨대 정치 주제에서 위의 다섯가지의 절차에서 보여준 경향성이 경제, 문화, 군사, 역사 등의 다양한 주제에 일관되게 나타나게 되면 높은 가중치를 부여하고, 그 반대일 경우 낮은 가중치를 부여할 수 있다.
- [0064] 즉, 각 단계 가치 위계의 점수는 해당 가치 위계와 관련된 행동의 경중완급(輕重緩急), 대관소찰(大觀小察), 순

서조절(順序調節)을 나타내는 수치일 수 있다. 복합적인 도덕 판단에 있어서, 제4 단계를 기준으로 그보다 낮은 단계에 대해서는 층위별로 빈도를 반영하여 더 낮은 점수를 부여하고, 반대로 높은 가치 단계에 대해서는 층위별로 빈도를 반영하여 더 높은 점수를 각기 부여하여, 계산방식에 따라 가치점수를 반영할 수 있다.

- [0065] 즉, 서버(200)는 7 단계로 분류된 각각의 가치 위계에 대한 점수를 상술한 방식으로 산출하고, 산출된 점수의 통계 값을 바탕으로 각 단계의 가치 위계에 대한 척도 값을 산출할 수 있다.
- [0066] 서버(200)는 각 단계에 대하여 산출된 가치 위계 점수를 기초하여 범용 정체성 학습 데이터 또는 그룹 정체성 학습 데이터를 생성할 수 있다. 서버(200)는 범용 정체성 학습 데이터를 수집된 모든 키워드와 모든 성향 관련 정보에 대하여 낮은 단계에 대해서는 층위별로 빈도를 반영하여 더 낮은 점수를 부여하고, 반대로 높은 가치 단계에 대해서는 층위별로 빈도를 반영하여 더 높은 점수를 각기 부여하는 계산자에 의하여 제1 내지 제7 단계 가치 위계의 점수를 산출하고, 각 개인에 대하여 산출된 각 단계 가치 위계 점수를 종합하는 방법으로 형성할 수 있다. 즉, 수집된 모든 개인에 대한 평균이 범용 정체성 학습 데이터일 수 있다.
- [0067] 그리고, 그룹 정체성 학습 데이터는 개인정보에 기초하여 분류된 그룹에 따라 해당 그룹에 속한 개인에 대한 키워드와 성향 관련 정보를 수집하여 형성될 수 있다. 구체적으로, 서버(200)는 단말 장치(100)로부터 수신한 개인정보에 기초하여 복수의 그룹으로 분류하는 데, 여기서 개인정보란 인종, 국가, 연령, 종교, 이념, 성별 등을 포함하는 것으로 개인을 특정할 수 있는 정보가 포함될 수 있다. 서버(200)는 그룹을 구성하는 개인에 대하여 산출된 각 단계에 대한 가치 위계 점수를 기초하여 그룹 정체성 학습 데이터를 생성할 수 있다.
- [0068] 그리고, 서버(200)는 단말 장치(100)로부터 수신한 데이터로부터 행동 지침 학습 데이터를 생성할 수 있다. 여기서 데이터는 사회의 법률, 관습, 조리 및 복수의 윤리 규칙을 포함할 수 있다. 서버(200)는 데이터에 포함된 사회의 법률, 관습, 조리 및 복수의 윤리 규칙을 기초하여 학습 가능한 형태인 행동 지침 학습 데이터를 생성할 수 있다. 행동 지침 학습 데이터는 특정 상황에서 법률, 관습, 조리 및 복수의 윤리 규칙에 따라 행동을 수행하도록 설정된 것으로 예컨대 A라는 상황에 적절한 B라는 행동을 수행하도록 설정된 것을 의미할 수 있다.
- [0069] 서버(200)는 학습부(미도시)를 포함할 수 있다. 학습부는 생성된 행동 지침 학습 데이터 및 범용 정체성 학습 데이터를 기초로 학습하여 범용 정체성 딥러닝 학습 모델을 생성하고, 행동 지침 학습 데이터 및 그룹 정체성 학습 데이터를 기초로 그룹 정체성 딥러닝 학습 모델을 생성할 수 있다. 여기서 그룹 정체성 딥러닝 학습 모델은 인종, 국가, 연령, 종교, 이념, 성별을 포함하는 개인정보 중 특정 개인정보를 기초로 형성된 그룹에 대한 것을 의미할 수 있다.
- [0070] 또한, 범용 및 그룹 정체성 학습 데이터에는 각 가치 위계에 대한 점수가 포함되어 있으므로, 생성된 범용 및 그룹 정체성 딥러닝 학습 모델은 각 단계 가치 위계의 점수에 따라 해당 가치 위계와 관련된 행동의 경중완급, 대관소찰, 순서조절이 다르게 나타날 수 있다.
- [0071] 서버(200)는 학습부를 이용하여 범용 정체성 딥러닝 학습 모델 및 그룹 정체성 딥러닝 학습 모델을 형성할 수 있고, 서버(200)는 사용자 요청에 대응되는 하나의 딥러닝 학습 모델을 단말 장치로 전송할 수 있다. 서버(200)는 단말 장치(100)로 딥러닝 학습 모델을 전송한 후에도 학습부를 이용하여 지도 학습 및 비지도 학습을 계속하여 딥러닝 학습 모델을 갱신할 수 있다.
- [0072] 도 2는 본 개시의 일 실시 예에 따른, 단말 장치(100)의 상황 분석부(110)를 도시한 블록도이다.
- [0073] 도 2를 참조하면, 단말 장치(100)의 상황 분석부(110)는 행위 인식 모듈, 얼굴 인식 분석 모듈, 음성인식 분석 모듈, 자연언어 이해 지능 모듈, 이미지 분석 모듈을 포함할 수 있다.
- [0074] 상황 분석부(110)는 입력부(150)에 포함된 카메라(미도시)가 촬영한 사용자의 모션 및 주변의 상황과 환경을 분석할 수 있고, 입력부(150)에 포함된 마이크(미도시)가 입력 받은 사용자의 발화 및 주변의 상황과 환경을 분석할 수 있다.
- [0075] 상황 분석부(110)는 사용자의 행동, 표정, 음성, 음성에 포함된 단어를 분석하여 사용자의 감정, 심박수, 활동성, 신체상태를 포함하는 사용자 상태를 파악할 수 있다.
- [0076] 구체적으로, 상황 분석부(110)는 행위 인식 모듈을 이용하여 사물감지(Object detection) 기술을 기초로 사물과 사람을 분류할 수 있다. 행위 인식 모듈은 사물과 사람을 분류한 후, 행위 인식 모듈은 사람의 행위를 세그먼트로 나누어 카테고리화 하는 방식을 통하여 사용자의 행위를 인식 및 학습 할 수 있다.
- [0077] 상황 분석부(110)는 얼굴인식 분석 모듈을 이용하여 얼굴인식을 통하여 이미지에서 사람의 얼굴을 감지하여 찾

아내는 기능을 수행할 수 있다. 그리고, 얼굴인식 분석 모듈은 감지된 사람의 얼굴을 분석하여 분노(Anger), 경멸(Contempt), 불쾌(Disgust), 공포(Fear), 행복(Happiness), 중립/무관심(Neutral), 슬픔(Sadness), 놀라움(Surprise) 등 8가지 감정들을 수치화할 수 있다. 즉, 상황 분석부(110)는 수치화한 8가지 감정들을 기초로 사용자의 감정 및 신체상태를 포함하는 사용자 상태를 파악할 수 있다.

[0078] 또는 서버(100)는 수치화한 개별 8가지 감정들을 이용하여 다양한 형태로 작용하였는지 반영할 수 있다. 즉, 사용자의 수치화된 감정을 이요하여 경중완급, 대관소찰, 순서조절 등의 기능적인 작동과정에 어느 정도로 작용했는지를 반영할 수 있다.

[0079] 따라서, 서버(100)는 사용자의 8가지 표정과 그들 간의 복합적인 상호관계에 대한 데이터를 통해 도출된 값을 기초로, 도덕 판단 알고리즘에 데이터로 활용될 수 있다.

[0080] 상황 분석부(110)는 음성인식 분석 모듈 및 자연언어 이해 지능 모듈을 이용하여 사용자의 발화를 인식 및 분석할 수 있다. 음성인식 분석 모듈 및 자연언어 이해지능 모듈은 어텐션(Attention) 기반의 언어모델을 기초로 하여 입력 정보 중 중요한 단어에 집중해 사용자의 발언 의도를 분석할 수 있다. 중요한 단어들을 계속하여 업데이트하며 학습하고 사용자의 발화 의도와 문맥 분석에 집중하여 학습할 수 있다.

[0081] 상황 분석부(110)는 이미지 분석 모듈을 이용하여 이미지의 내용을 이해할 수 있다. 이미지 분석 모듈은 행위 인식모듈, 얼굴인식 분석 모듈, 음성인식 분석 모듈 및 자연언어 이해 지능 모듈로부터 분석 및 인식된 정보를 취합할 수 있다. 구체적으로, 행위 인식모듈을 이용하여 인식된 사용자의 행위 및 사용자의 활동성을 파악할 수 있고, 얼굴인식 분석 모듈을 통하여 사용자의 감정을 파악하고, 음성인식 분석 모듈 및 자연언어 이해 지능 모듈을 이용하여 음성 및 음성에 포함된 단어를 분석하여 사용자의 발화 의도를 파악할 수 있다. 이미지 분석 모듈은 각 모듈에 의하여 인식 및 분석된 정보를 취합하여 사용자 상태 및 사물을 포함하는 주변의 상황을 파악할 수 있다.

[0082] 상황 분석부(110)는 학습부(미도시) 를 포함할 수 있다. 학습부는 소정의 상황 판단을 위한 기준을 갖는 행동 지침을 딥러닝 학습 모델에 학습시킬 수 있다.

[0083] 일 예로, 학습부는 사용자 명령이 입력되는 경우, 사용자의 행동, 표정, 음성 및 음성에 포함된 단어를 분석하여 사용자의 감정, 심박수, 활동성, 신체상태 등을 파악할 수 있고, 파악된 사용자 상태, 주변의 상황 및 환경에 대하여 분석된 정보를 사용자 명령과 매칭하여 경험적 행동 지침을 형성할 수 있다. 형성된 경험적 행동 지침을 서버(200)부터 수신한 딥러닝 학습 모델에 학습 또는 갱신시킬 수 있다.

[0084] 학습부의 일부와 행위 인식 모듈, 얼굴인식 분석 모듈, 음성인식 분석 모듈, 자연언어 이해 지능 모듈, 이미지 분석 모듈(이하, "상황 분석부(110)에 포함된 모듈") 중 적어도 일부는, 소프트웨어 모듈로 구현되거나 적어도 하나의 하드웨어 칩 형태로 제작되어 전자 장치에 탑재될 수 있다. 예를 들어, 학습부 및 상황 분석부(110)에 포함된 모듈 중 적어도 하나는 인공 지능(AI; artificial intelligence)을 위한 전용 하드웨어 칩 형태로 제작될 수도 있고, 또는 기존의 범용 프로세서(예: CPU 또는 application processor) 또는 그래픽 전용 프로세서(예: GPU)의 일부로 제작되어 전술한 각종 전자 장치 또는 객체 인식 장치에 탑재될 수도 있다. 이 때, 인공 지능을 위한 전용 하드웨어 칩은 확률 연산에 특화된 전용 프로세서로서, 기존의 범용 프로세서보다 병렬 처리 성능이 높아 기계 학습과 같은 인공 지능 분야의 연산 작업을 빠르게 처리할 수 있다. 학습부 및 상황 분석부(110)에 포함된 모듈이 소프트웨어 모듈(또는, 인스트럭션(instruction) 포함하는 프로그램 모듈)로 구현되는 경우, 소프트웨어 모듈은 컴퓨터로 읽을 수 있는 판독 가능한 비일시적 판독 가능 기록매체(non-transitory computer readable media)에 저장될 수 있다. 이 경우, 소프트웨어 모듈은 OS(Operating System)에 의해 제공되거나, 소정의 애플리케이션에 의해 제공될 수 있다. 또는, 소프트웨어 모듈 중 일부는 OS(Operating System)에 의해 제공되고, 나머지 일부는 소정의 애플리케이션에 의해 제공될 수 있다.

[0085] 이 경우, 학습부 및 상황 분석부(110)에 포함된 모듈은 하나의 전자 장치에 탑재될 수도 있으며, 또는 별개의 전자 장치들에 각각 탑재될 수도 있다. 예를 들어, 학습부 및 상황 분석부에 포함된 모듈 중 하나는 단말 장치(100)에 포함되고, 나머지 하나는 서버(200)에 포함될 수 있다. 또한, 학습부 및 상황 분석부에 포함된 모듈은 유선 또는 무선으로 통하여, 학습부가 구축한 모델 정보를 상황 분석부에 포함된 모듈로 제공할 수도 있고, 상황 분석부에 포함된 모듈로 입력된 데이터가 추가 학습 데이터로서 학습부로 제공될 수도 있다.

[0086] 도 3 내지 도 6은 본 개시의 일 실시 예에 따른, 인공지능의 자율적 도덕 판단 및 수행을 위한 시스템의 흐름도이다.

[0087] 도 3을 참조하면, 본 개시의 일 실시 예에 따라 범용 정체성 딥러닝 학습 모델이 형성되는 과정이 도시되어 있

다.

- [0088] 단말 장치(100)가 수집한 각각의 개인정보 및 데이터를 서버로 송신할 수 있다(S310). 여기서, 개인정보 및 데이터는 오프라인 상으로 발행되는 간행물에 게재되어 있는 것뿐만 아니라, 온라인 상에 게시된 뉴스, 신문, SNS, 댓글, 칼럼, 블로그 등으로부터 획득할 수 있다.
- [0089] 또한, 개인정보는 인종, 국가, 연령, 종교, 이념, 성별 등을 포함할 수 있고, 이외에도 개인을 특정할 수 있는 주소, 가족관계, 학력, 혈액형 등을 포함할 수 있다. 데이터는 개인의 성향을 파악할 수 있는 단어, 의견 및 진술을 포함할 수 있고, 사회의 법률, 관습, 조리 및 복수의 윤리 규칙을 포함할 수 있다.
- [0090] 서버(200)가 단말 장치(100)로부터 개인정보 및 데이터를 수신하고(S320), 서버(200)는 적어도 하나의 분류기를 포함할 수 있고, 분류기를 이용하여 수신된 데이터로부터 키워드 및 성향 관련 정보만을 추출할 수 있다.
- [0091] 여기서, 추출된 키워드 및 성향 관련 정보란, 단말 장치(100)로부터 수신된 데이터가 도덕 판단의 기반이 되는 제1 내지 제7 단계의 가치 위계에 관련한 단어 및 문장으로 분리 된 것일 수 있다.
- [0092] 서버(200)는 범용 정체성 학습 데이터를 수집된 모든 키워드와 모든 성향 관련 정보에 대하여 제1 내지 제7 단계 가치 위계의 점수를 산출하여 평균을 내는 방법으로 형성할 수 있다. 즉, 수집된 모든 개인에 대한 평균이 범용 정체성 학습 데이터일 수 있다. 또한, 각 단계 가치 위계의 점수는 해당 가치 위계와 관련된 행동의 경중완급, 대관소찰, 순서조절을 나타내는 수치일 수 있다. 서버(200)는 추출된 키워드와 성향 관련 정보에 기초하여 범용 정체성 학습 데이터를 생성할 수 있다(S330).
- [0093] 서버(200)는 생성된 범용 정체성 학습 데이터에 기초하여 범용 정체성 딥러닝 학습 모델을 생성할 수 있다(S340). 범용 정체성 학습 데이터에는 각 가치 위계에 대한 점수가 포함되어 있으므로, 생성된 범용 정체성 딥러닝 학습 모델은 각 단계 가치 위계의 점수에 따라 해당 가치 위계와 관련된 행동의 경중완급, 대관소찰, 순서조절이 다르게 나타날 수 있다.
- [0094] 서버(200)는 생성된 범용 정체성 딥러닝 학습 모델을 단말 장치(100)로 전송할 수 있다(S350).
- [0095] 학습부(미도시)는 단말 장치(100) 또는 서버(200)에 포함될 수 있고, 단말 장치(100) 또는 서버(200)는 학습부를 이용하여 지도 학습 및 비지도 학습을 계속하여 딥러닝 학습 모델을 갱신 할 수 있다.
- [0096] 도 4를 참조하면, 본 개시의 일 실시 예에 따라 그룹 정체성 딥러닝 학습 모델이 형성되는 과정이 도시되어 있다.
- [0097] 단말 장치(100)가 수집한 각각의 개인정보 및 데이터를 서버로 송신할 수 있다(S410).
- [0098] 여기서, 개인정보 및 데이터는 오프라인 상으로 발행되는 간행물에 게재되어 있는 것뿐만 아니라, 온라인 상에 게시된 뉴스, 신문, SNS, 댓글, 칼럼, 블로그 등으로부터 획득할 수 있다.
- [0099] 또한, 개인정보는 인종, 국가, 연령, 종교, 이념, 성별 등을 포함할 수 있고, 이외에도 개인을 특정할 수 있는 주소, 가족관계, 학력, 혈액형 등을 포함할 수 있다. 데이터는 개인의 성향을 파악할 수 있는 단어, 의견 및 진술을 포함할 수 있고, 사회의 법률, 관습, 조리 및 복수의 윤리 규칙을 포함할 수 있다.
- [0100] 서버(200)가 단말 장치(100)로부터 개인정보 및 데이터를 수신할 수 있다(S420).
- [0101] 서버(200)는 적어도 하나의 분류기를 포함할 수 있고, 분류기를 이용하여 수신된 개인정보에 기초하여 복수의 그룹을 생성할 수 있다(S430).
- [0102] 복수의 그룹은 인종, 국가, 연령, 종교, 이념, 성별을 포함하는 개인정보에 기초하여 생성될 수 있다. 또한, 서버(200)는 데이터로부터 그룹 각각에 대한 키워드 및 성향 관련 정보만을 추출할 수 있다.
- [0103] 여기서, 추출된 키워드 및 성향 관련 정보란, 단말 장치(100)로부터 수신된 데이터가 도덕 판단의 기반이 되는 제1 내지 제7 단계의 가치 위계에 관련한 단어 및 문장으로 분리 된 것일 수 있다.
- [0104] 서버(200)는 그룹 정체성 학습 데이터를 수집된 모든 키워드와 모든 성향 관련 정보 중에서 그룹을 구성하는 개인에 대하여 제1 내지 제7 단계 가치 위계의 점수를 산출하여 평균을 내는 방법으로 형성할 수 있다. 즉, 그룹을 이루는 개인에 대한 평균이 그룹 정체성 학습 데이터일 수 있다. 또한, 각 단계 가치 위계의 점수는 해당 가치 위계와 관련된 행동의 경중완급, 대관소찰, 순서조절을 나타내는 수치일 수 있다. 예를 들어, 제2 단계의 지지/반대의 지향성이 2/10인데, 제4 단계의 지지/반대의 지향성이 4/10이면 제4 단계에 대하여 높은 도덕 판단을 보이는 것이다. 서버(200)는 추출된 키워드와 성향 관련 정보에 기초하여 그룹 정체성 학습 데이터를 생성할 수

있다(S440).

- [0105] 서버(200)는 생성된 그룹 정체성 학습 데이터를 기초로 학습하여 그룹 정체성 딥러닝 학습 모델을 생성할 수 있다(S450). 그룹 정체성 학습 데이터에는 각 가치 위계에 대한 점수가 포함되어 있으므로, 생성된 그룹 정체성 딥러닝 학습 모델은 각 단계 가치 위계의 점수에 따라 해당 가치 위계와 관련된 행동의 경중완급, 대관소찰, 순서조절이 다르게 나타날 수 있다.
- [0106] 서버(200)는 생성된 그룹 정체성 딥러닝 학습 모델 중 사용자 요청에 대응되는 하나의 딥러닝 학습 모델을 단말 장치(100)로 전송할 수 있다(S460). 그룹 정체성 딥러닝 학습 모델은 각 그룹별로 다르게 나타나므로, 사용자는 선호도에 따라 그룹을 다르게 선택할 수 있고, 단말 장치(100)는 선택된 그룹에 따라 딥러닝 학습 모델을 수신할 수 있다.
- [0107] 학습부(미도시)는 단말 장치(100) 또는 서버(200)에 포함될 수 있고, 단말 장치(100) 또는 서버(200)는 학습부를 이용하여 지도 학습 및 비지도 학습을 계속하여 딥러닝 학습 모델을 갱신 할 수 있다.
- [0108] 도 5는 참조하면, 본 개시의 일 실시 예에 따라 행동 지침이 선택 및 수행되는 과정이 도시되어 있다.
- [0109] 단말 장치(100)가 서버(200)로부터 사용자 요청에 대응되는 딥러닝 학습 모델을 수신할 수 있다(S510). 여기서 딥러닝 학습 모델은 범용 정체성 딥러닝 학습 모델 또는 그룹 정체성 딥러닝 학습 모델 중 하나 일 수 있다.
- [0110] 단말 장치(100)가 입력부(150)를 이용하여 주변의 상황 및 환경에 대한 정보를 센싱할 수 있다(S520).
- [0111] 여기서 입력부(150)는 사용자의 물리적 명령을 인식할 수 있는 터치 센서, 압력 센서를 포함하는 물리적 버튼을 포함할 수 있고, 사용자의 모션 및 주변의 상황과 환경을 촬영할 수 있는 카메라, 사용자의 발화 및 주변의 상황과 환경을 입력 받을 수 있는 마이크 등을 포함할 수 있다.
- [0112] 단말 장치(100)의 상황 분석부(110)가 입력부(150)로부터 센싱된 주변의 상황 및 환경에 대한 정보를 분석할 수 있다(S530). 상황 분석부(150)는 상황 분석부에 포함된 각종 모듈을 이용하여 주변의 상황 및 환경을 인식하고 분석할 수 있다.
- [0113] 단말 장치(100)의 행동 지침 선택부(120)는 서버(200)로부터 수신한 딥러닝 학습 모델과 상황 분석부(110)를 기초하여 분석된 주변의 상황 및 환경에 대응되는 행동 지침을 선택할 수 있다(S540). 단말 장치(100)의 제어부(130)가 선택된 행동 지침을 수행하도록 단말 장치(100)를 제어할 수 있다(S550).
- [0114] 여기서 행동 지침이란, 특정 상황에서 법률, 관습, 조리 및 복수의 윤리 규칙에 따라 생성된 행동 지침 학습 데이터에 따라 결정된 것으로, 예컨대 A라는 상황에 적절한 B라는 행동을 수행하도록 설정된 것을 의미한다. 따라서 분석된 주변의 상황 및 환경에 대응되는 행동 지침은 복수 개 존재할 수 있다.
- [0115] 분석된 주변의 상황 및 환경에서 수행할 수 있는 복수개의 행동 지침이 존재하는 경우란, 예를 들어, "도서관에서는 다른 사람에게 피해를 주지 않기 위해 조용하게 행동하거나 주의해야 한다."라는 행동 지침과 "C라는 책의 위치를 찾아달라는 사용자 명령에 따라 대응되는 정보를 제공하기"라는 행동 지침이 존재하는 경우가 있다. 이 경우 단말 장치(100)는 다른 사람에게 피해를 주지 않기 위해 기 설정된 데시벨 이하로 음성을 출력하여 C라는 책의 위치에 대한 정보를 제공하거나 디스플레이(미도시)로 정보를 출력 할 수 있다.
- [0116] 다른 예로, "도서관에서는 다른 사람에게 피해를 주지 않기 위해 조용하게 행동하거나 주의해야 한다."라는 행동 지침과 "'음악 틀어줘'라는 사용자 명령을 수행하기"라는 행동 지침이 존재하는 경우가 있다. 이 경우 단말 장치(100)는 사용자에게만 음악을 제공할 수 없는 경우 "도서관에서는 다른 사람에게 피해를 주지 않기 위해 조용하게 행동하거나 주의해야 한다."라는 행동 지침만을 수행할 수 있다. 즉, 복수개의 행동 지침이 존재하는 경우에 해당하나, 두 개의 행동 지침이 충돌하는 경우, 단말 장치(100)는 7단계의 가치 위계의 점수에 따라 수행할 행동 지침을 선택할 수 있다.
- [0117] 도 6을 참조하면, 사용자 명령에 따른 경험적 행동 지침을 생성하는 과정이 도시되어 있다.
- [0118] 단말 장치(100)가 서버(200)로부터 사용자 요청에 대응되는 딥러닝 학습 모델을 수신할 수 있다(S610). 여기서 딥러닝 학습 모델은 범용 정체성 딥러닝 학습 모델 또는 그룹 정체성 딥러닝 학습 모델 중 하나 일 수 있다.
- [0119] 단말 장치(100)가 입력부(150)를 이용하여 주변의 상황 및 환경에 대한 정보를 센싱할 수 있다(S620).
- [0120] 여기서 입력부(150)는 사용자의 물리적 명령을 인식할 수 있는 터치 센서, 압력 센서를 포함하는 물리적 버튼을 포함할 수 있고, 사용자의 모션 및 주변의 상황과 환경을 촬영할 수 있는 카메라, 사용자의 발화 및 주변의 상

황과 환경을 입력 받을 수 있는 마이크 등을 포함할 수 있다.

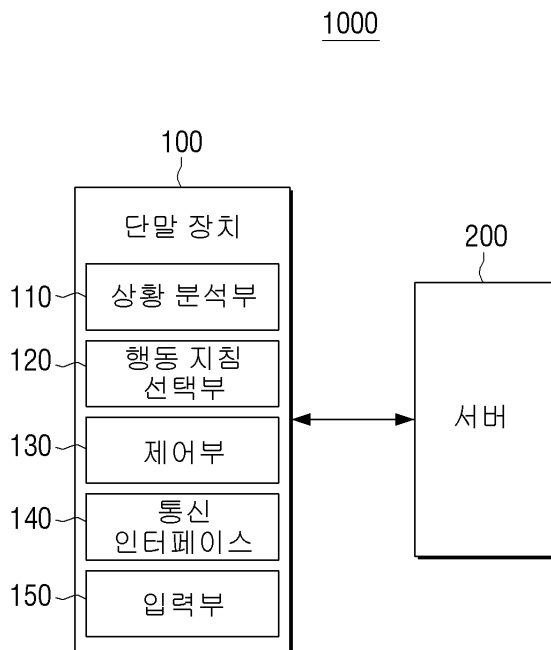
- [0121] 단말 장치(100)의 상황 분석부(110)가 센싱된 정보 및 사용자의 행동, 표정, 음성 및 상기 음성에 포함된 단어를 분석하여 사용자의 감정, 심박수, 활동성, 신체상태를 포함하는 사용자 상태를 파악할 수 있다(S630). 상황 분석부(150)는 상황 분석부에 포함된 행위 인식 모듈, 얼굴인식 분석 모듈, 음성인식 분석 모듈, 자연언어 이해 기능 모듈 및 이미지 분석 모듈을 이용하여 주변의 상황 및 환경을 인식하고 분석할 수 있다.
- [0122] 단말 장치(100)의 상황 분석부(110)는 파악된 사용자의 상태, 주변의 상황 및 환경에 대하여 분석된 정보를 사용자 명령과 매칭하여 경험적 행동 지침을 생성할 수 있다(S640). 단말 장치(100)는 학습부(미도시)를 포함할 수 있고, 학습부는 상황 분석부에 의하여 생성된 경험적 행동 지침을 학습할 수 있다.
- [0123] 경험적 행동 지침이란, 사용자 명령 또는 반복적인 상황에 대하여 단말 장치(100)가 수행해야 하는 행동 지침을 의미할 수 있다.
- [0124] 단말 장치(100)의 행동 지침 선택부(120)는 형성된 경험적 행동 지침 및 수신된 딥러닝 학습 모델에 기초하여 행동 지침을 선택할 수 있다(S650).
- [0125] 다만, 본 개시의 다른 실시예에 따른 행동 지침 선택부(120)는 주어진 상황에서 수행할 수 있는 행동 지침이 존재하지 않을 경우가 있을 수 있다. 이 경우 제어부(130)는 행동 지침 선택부(120)가 '선택할 수 있는 행동 지침이 없음' 또는 '사용자 명령을 수행할 수 없음'을 사용자에게 알리고 적절한 행동 지침을 입력 받을 수 있다. 또는, 다른 행동 지침을 수행하도록 다른 대안을 사용자에게 제시할 수 있다.
- [0126] 단말 장치(100)의 제어부(130)는 선택된 행동 지침을 수행하도록 단말 장치(100)를 제어할 수 있다(S660).
- [0127] 이상에서는 본 개시의 다양한 실시예를 각각 개별적으로 설명하였으나, 각 실시예들은 반드시 단독으로 구현되어야만 하는 것은 아니며, 각 실시예들의 구성 및 동작은 적어도 하나의 다른 실시예들과 조합되어 구현될 수도 있다.
- [0128] 일 실시 예에 따르면, 본 문서에 개시된 다양한 실시예들에 따른 방법은 컴퓨터 프로그램 제품(computer program product)에 포함되어 제공될 수 있다. 컴퓨터 프로그램 제품은 상품으로서 판매자 및 구매자 간에 거래될 수 있다. 컴퓨터 프로그램 제품은 기기로 읽을 수 있는 저장 매체(예: compact disc read only memory (CD-ROM))의 형태로, 또는 어플리케이션 스토어(예: 플레이 스토어™)를 통해 온라인으로 배포될 수 있다. 온라인 배포의 경우에, 컴퓨터 프로그램 제품의 적어도 일부는 제조사의 서버, 어플리케이션 스토어의 서버, 또는 중계 서버의 메모리와 같은 저장 매체에 적어도 일시 저장되거나, 임시적으로 생성될 수 있다.
- [0129] 또한, 이상에서는 본 개시의 바람직한 실시 예에 대하여 도시하고 설명하였지만, 본 개시는 상술한 특징의 실시 예에 한정되지 아니하며, 청구범위상에서 청구하는 본 개시의 요지를 벗어남이 없이 당해 발명이 속하는 기술분야에서 통상의 지식을 가진 자에 의해 다양한 변형실시가 가능한 것은 물론이고, 이러한 변형실시들은 본 개시의 기술적 사상이나 전망으로부터 개별적으로 이해되어서는 안될 것이다.

부호의 설명

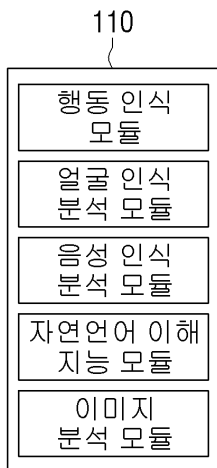
- [0130] 1000: 인공지능의 자율적 도덕 판단 및 수행을 위한 시스템
- 100: 단말 장치
- 200: 서버

도면

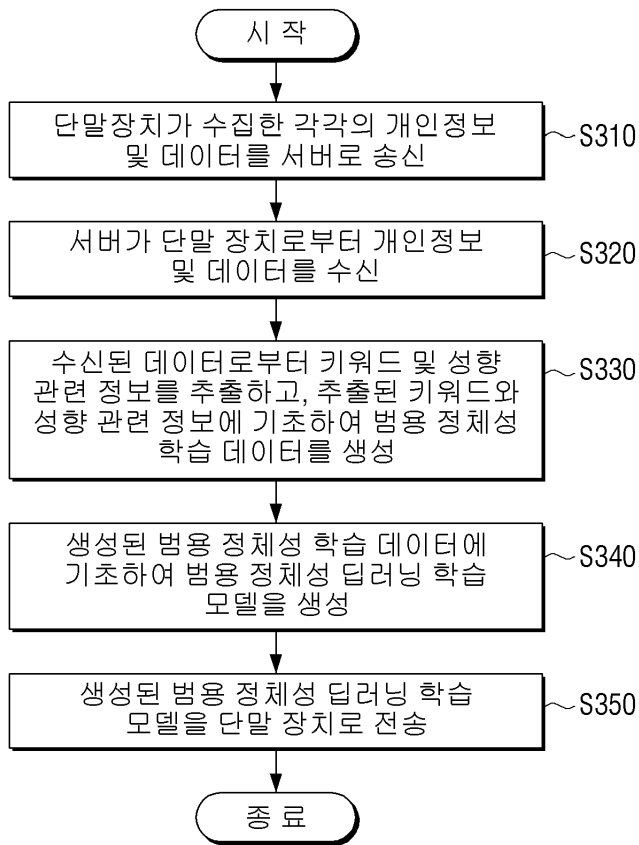
도면1



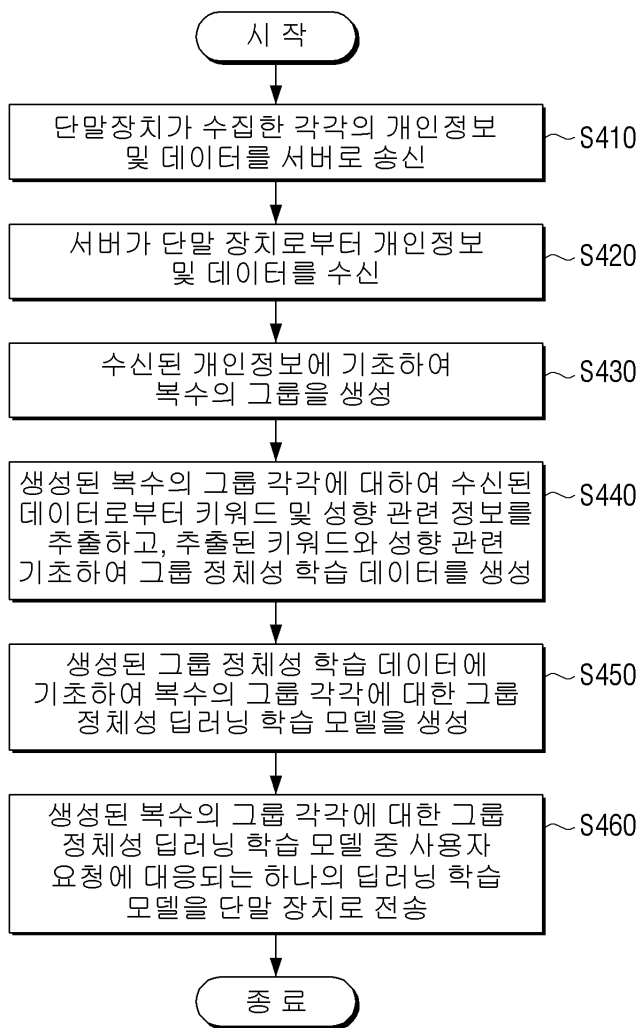
도면2



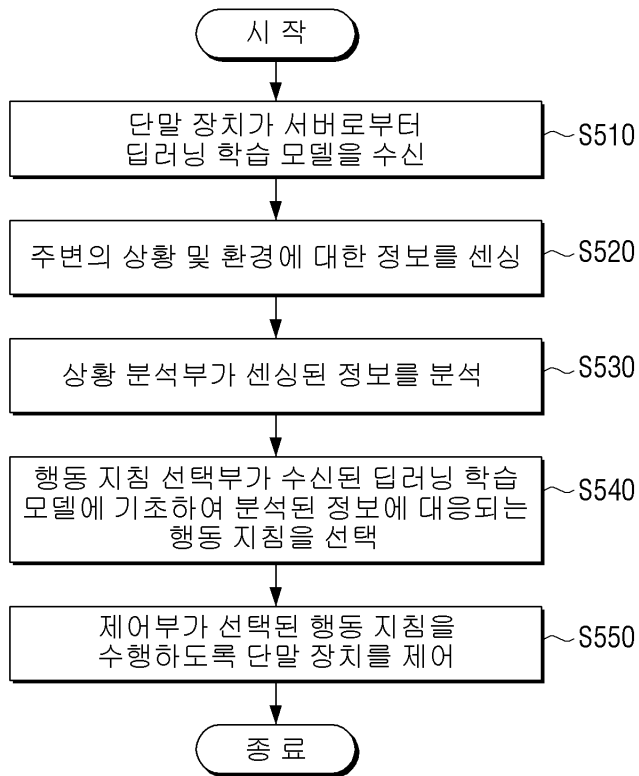
도면3



도면4



도면5



도면6

