

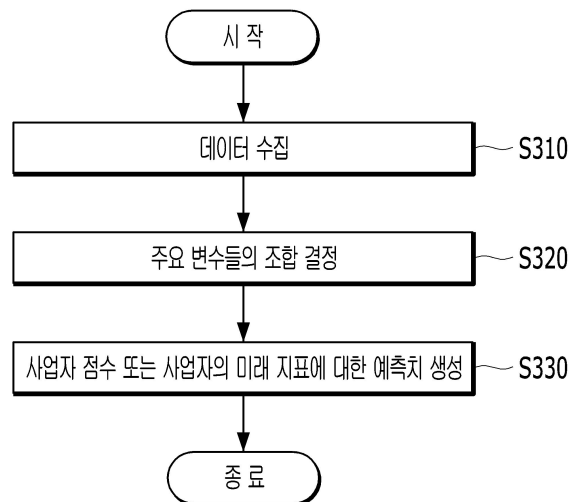
**(19) 대한민국특허청(KR)**
(12) 공개특허공보(A)**(11) 공개번호** 10-2025-0050854
(43) 공개일자 2025년04월15일

- (51) 국제특허분류(Int. Cl.)
G06Q 40/03 (2023.01) **G06N 20/00** (2019.01)
G06N 3/08 (2023.01) **G06Q 10/04** (2023.01)
G06Q 10/063 (2023.01) **G06Q 30/0201** (2023.01)
- (52) CPC특허분류
G06Q 40/03 (2023.01)
G06N 20/00 (2021.08)
- (21) 출원번호 **10-2025-0046647(분할)**
(22) 출원일자 **2025년04월10일**
심사청구일자 **2025년04월10일**
(62) 원출원 **특허 10-2022-0016814**
원출원일자 **2022년02월09일**
심사청구일자 **2022년02월09일**
- (30) 우선권주장
1020210180082 2021년12월15일 대한민국(KR)
- (71) 출원인
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
- (72) 발명자
이경호
서울특별시 강남구 선릉로 222, 105동 501호(대치동, 대치아이파크아파트)
오준형
서울특별시 성북구 안암로9가길 71(안암동5가)
(뒷면에 계속)
- (74) 대리인
이대호, 박건홍

전체 청구항 수 : 총 12 항

(54) 발명의 명칭 사업자 활동 데이터 분석 방법**(57) 요약**

본 개시의 일 실시예에 따라 사업자 활동을 분석하기 위한 방법 및 이를 이용한 컴퓨터 판독가능 저장 매체에 저장된 컴퓨터 프로그램이 개시된다. 구체적으로, 본 개시에 따르면, 컴퓨팅 장치가, 사업자 활동 데이터를 수집하고, 수집된 상기 사업자 활동 데이터를 기반으로 분석 대상인 조합과 예측 알고리즘의 쌍을 결정하고, 상기 조합을 기반으로, 사업자 점수 데이터를 산출하거나 예측 모델을 이용하여 사업자의 미래 지표에 대한 예측치 정보를 생성한다.

대표도 - 도3

(52) CPC특허분류

G06N 3/08 (2023.01)

G06Q 10/04 (2023.01)

G06Q 10/063 (2023.01)

G06Q 30/0201 (2023.01)

(72) 발명자

윤현식

서울특별시 송파구 올림픽로4길 15, 5동 904호(잠실동, 아시아선수촌아파트)

강해영

경기도 안산시 단원구 당곡로 33, 102동 307호(고잔동)

명세서

청구범위

청구항 1

사업자 활동을 분석하기 위해 컴퓨팅 장치에 의해 수행되는 방법으로서,

사업자 활동 데이터를 수집하는 단계;

수집된 상기 사업자 활동 데이터를 기반으로, 하나 이상의 상품 카테고리를 식별하는 단계;

변수 선택 모델 및 예측 모델을 기초로, 사업자의 판매 활동과 관련된 정보를 포함하는 사업자 활동 지표에 영향을 미치는 변수 중 상기 하나 이상의 상품 카테고리에 각각 대응되는 하나 이상의 주요 변수의 조합을 결정하는 단계; 및

상기 결정된 하나 이상의 주요 변수의 조합을 기반으로, 사업자 점수 데이터를 산출하는 동작 또는 예측 모델을 이용하여 사업자의 미래 지표에 대한 예측치 정보를 생성하는 동작 중 적어도 하나를 수행하는 단계;

를 포함하고,

상기 하나 이상의 상품 카테고리에 각각 대응되는 하나 이상의 주요 변수의 조합을 결정하는 단계는:

하나 이상의 변수 선택 알고리즘들로 구성된 상기 변수 선택 모델을 기초로, 사업자 활동 지표에 영향을 미치는 변수 중 N개의 주요 변수들을 추출하는 단계;

상기 N개의 주요 변수들을 기초로, 각각 하나 이상의 주요 변수를 포함하는, M개의 주요 변수 조합들을 생성하는 단계;

상기 주요 변수 조합들 각각을 하나 이상의 예측 알고리즘들로 구성된 상기 예측 모델을 통해 분석하는 단계; 및

상기 예측 모델을 통해 분석한 결과에 기초하여 상기 하나 이상의 상품 카테고리에 대하여 가장 정확한 예측을 한 것으로 분석된 주요 변수 조합과 예측 알고리즘의 쌍을 도출하고, 도출된 주요 변수 조합을 최종 분석 대상으로 결정하고, 도출된 예측 알고리즘을 최종 분석에 사용할 예측 알고리즘으로 결정하는 단계;

를 포함하는,

방법.

청구항 2

제 1 항에 있어서,

상기 사업자 활동 데이터를 수집하는 단계는:

하나 이상의 사업자 데이터베이스로부터 상기 사업자 활동 데이터를 일정 주기 간격으로 크롤링하여 수집하는 단계; 및

수집된 상기 사업자 활동 데이터를 상기 크롤링 주기에 따라 업데이트하는 단계;

를 포함하는,

방법.

청구항 3

제 2 항에 있어서,

상기 사업자 활동 데이터는,

상기 사업자 데이터베이스에 기록된, 상품 정보, 사업자 정보 또는 매출 정보 중 적어도 하나와 관련된 데이터;

를 포함하는,
방법.

청구항 4

제 2 항에 있어서,
상기 사업자 활동 데이터는,
상기 사업자 데이터베이스로에 포함된 정보에 관한 증감 추이 데이터;
를 더 포함하는,
방법.

청구항 5

제 1 항에 있어서,
상기 하나 이상의 변수 선택 알고리즘들로 구성된 변수 선택 모델을 기초로, 사업자 활동 지표에 영향을 미치는 변수 중 N개의 주요 변수들을 추출하는 단계는:
상기 변수들 각각이 사업자 활동 지표에 미치는 영향을 상기 하나 이상의 변수 선택 알고리즘들을 통해 분석하고 표준점수의 형태로 도출하는 단계;
상기 변수들 각각에 대하여 상기 변수 선택 알고리즘별 표준점수를 합산하는 단계; 및
합산된 상기 변수 선택 알고리즘별 표준점수를 비교하여 상기 변수 중 상위 N개의 변수들을 주요 변수로서 추출하는 단계;
를 포함하는,
방법.

청구항 6

제 1 항에 있어서,
상기 N개의 주요 변수들을 기초로, 각각 하나 이상의 주요 변수를 포함하는, M개의 주요 변수 조합들을 생성하는 단계는,
상기 N개의 변수들 중 순서에 상관없이 K개의 주요 변수들을 추출하는 동작을 M회 수행하여, M개의 주요 변수 조합들을 생성하는 단계;
방법.

청구항 7

제 1 항에 있어서,
상기 예측 모델을 통해 분석한 결과에 기초하여 상기 하나 이상의 상품 카테고리에 대하여 가장 정확한 예측을 한 것으로 분석된 주요 변수 조합과 예측 알고리즘의 쌍을 도출하는 것은:
상기 M개의 주요 변수 조합들을 상기 예측 모델에 속하는 하나 이상의 예측 알고리즘들을 통해 분석한 결과를 표준화하는 것; 및
상기 표준화된 점수가 가장 높게 판정된 하나의 조합과 그 조합의 분석에 사용한 상기 예측 알고리즘 쌍을 도출하는 것;
을 포함하는,
방법.

청구항 8

제 7 항에 있어서,

상기 M개의 주요 변수 조합들을 상기 예측 모델에 속하는 하나 이상의 예측 알고리즘들을 통해 분석한 결과를 표준화하는 것은,

정확도(Accuracy);

정밀도(Precision);

재현율(Recall); 또는

F1점수;

중 적어도 하나의 정보를 활용하는 것을 포함하는,

방법.

청구항 9

제 1 항에 있어서,

상기 사업자 점수 데이터는:

상기 결정된 주요 변수 조합, 상기 조합에 포함된 주요 변수의 중요도 및 상기 조합을 결정하기 위해 활용된 변수 선택 모델을 기반으로 계산된 정규화된 데이터;

를 포함하는,

방법.

청구항 10

제 1 항에 있어서,

상기 예측 모델을 이용하여 사업자의 미래 지표에 대한 예측치 정보를 생성하는 동작은:

상기 결정된 하나 이상의 주요 변수 조합과 상기 예측 모델을 기반으로 사업자의 미래 지표에 대한 예측치 정보를 계산하는 단계; 및

계산된 상기 사업자의 미래 지표에 대한 예측치 정보를 복수의 카테고리 중 하나의 카테고리로 분류하는 단계;

를 포함하는,

방법.

청구항 11

컴퓨팅 장치로 하여금 동작들을 수행하도록 하는 명령들을 포함하는 컴퓨터 판독가능 저장 매체에 저장된 컴퓨터 프로그램으로서, 상기 동작들은:

사업자 활동 데이터를 수집하는 동작;

수집된 상기 사업자 활동 데이터를 기반으로, 하나 이상의 상품 카테고리를 식별하는 동작;

변수 선택 모델 및 예측 모델을 기초로, 사업자의 판매 활동과 관련된 정보를 포함하는 사업자 활동 지표에 영향을 미치는 변수 중 상기 하나 이상의 상품 카테고리에 각각 대응되는 하나 이상의 주요 변수 조합을 결정하는 동작; 및

상기 결정된 하나 이상의 주요 변수 조합을 기반으로, 사업자 점수 데이터를 산출하는 동작 또는 예측 모델을 이용하여 사업자의 미래 지표에 대한 예측치 정보를 생성하는 동작 중 적어도 하나를 수행하는 동작;

을 포함하고,

상기 하나 이상의 상품 카테고리에 각각 대응되는 하나 이상의 주요 변수 조합을 결정하는 동작은:

하나 이상의 변수 선택 알고리즘들로 구성된 상기 변수 선택 모델을 기초로, 사업자 활동 지표에 영향을 미치는 변수 중 N개의 주요 변수들을 추출하는 동작;

상기 N개의 주요 변수들을 기초로, 각각 하나 이상의 주요 변수를 포함하는, M개의 주요 변수 조합들을 생성하는 동작;

상기 주요 변수 조합들 각각을 하나 이상의 예측 알고리즘들로 구성된 상기 예측 모델을 통해 분석하는 동작; 및

상기 예측 모델을 통해 분석한 결과에 기초하여 상기 하나 이상의 상품 카테고리에 대하여 가장 정확한 예측을 한 것으로 분석된 주요 변수 조합과 예측 알고리즘의 쌍을 도출하고, 도출된 주요 변수 조합을 최종 분석 대상으로 결정하고, 도출된 예측 알고리즘을 최종 분석에 사용할 예측 알고리즘으로 결정하는 동작;

을 포함하는,

컴퓨터 판독가능 저장 매체에 저장된 컴퓨터 프로그램.

청구항 12

컴퓨팅 장치로서,

적어도 하나의 프로세서; 및

메모리를 포함하고,

상기 적어도 하나의 프로세서는,

사업자 활동 데이터를 수집하고;

수집된 상기 사업자 활동 데이터를 기반으로, 하나 이상의 상품 카테고리를 식별하고;

변수 선택 모델 및 예측 모델을 기초로, 사업자의 판매 활동과 관련된 정보를 포함하는 사업자 활동 지표에 영향을 미치는 변수 중 상기 하나 이상의 상품 카테고리에 각각 대응되는 하나 이상의 주요 변수 조합을 결정하고; 그리고

상기 결정된 하나 이상의 주요 변수 조합을 기반으로, 사업자 점수 데이터를 산출하는 동작 또는 예측 모델을 이용하여 사업자의 미래 지표에 대한 예측치 정보를 생성하는 동작 중 적어도 하나를 수행하도록 구성되고,

상기 하나 이상의 상품 카테고리에 각각 대응되는 하나 이상의 주요 변수 조합을 결정하는 것은:

하나 이상의 변수 선택 알고리즘들로 구성된 상기 변수 선택 모델을 기초로, 사업자 활동 지표에 영향을 미치는 변수 중 N개의 주요 변수들을 추출하는 것;

상기 N개의 주요 변수들을 기초로, 각각 하나 이상의 주요 변수를 포함하는, M개의 주요 변수 조합들을 생성하는 것;

상기 주요 변수 조합들 각각을 하나 이상의 예측 알고리즘들로 구성된 상기 예측 모델을 통해 분석하는 것; 및

상기 예측 모델을 통해 분석한 결과에 기초하여 상기 하나 이상의 상품 카테고리에 대하여 가장 정확한 예측을 한 것으로 분석된 주요 변수 조합과 예측 알고리즘의 쌍을 도출하고, 도출된 주요 변수 조합을 최종 분석 대상으로 결정하고, 도출된 예측 알고리즘을 최종 분석에 사용할 예측 알고리즘으로 결정하는 것;

을 포함하는,

컴퓨팅 장치.

발명의 설명

기술 분야

[0001]

본 발명은 사업자 활동 데이터를 분석하는 방법에 관한 것으로, 보다 구체적으로 인공지능 모델을 활용하여 사업자의 영업에 관련된 활동을 분석하는 방법에 관한 것이다.

배경 기술

- [0002] 근래 전자상거래 시장의 성장으로 소규모의 온라인 사업자의 유입이 폭발적으로 증가하고 있다. 예를 들어, 쇼핑 플랫폼 중 하나인 네이버 스마트스토어의 전체 사업자 중 중소 사업자와 온라인 소상공인은 전체의 14%를 차지하고 있고, 매출이 발생하지 않은 1년 미만 사업자는 전체의 48%를 차지할 정도이다. 이러한 신규 소규모 온라인 사업자들의 운영자금 마련을 위한 금융지원의 필요성이 대두되고 있으나, 기술 환경의 변화로 인해 소규모의 사업자에 대하여 기존의 신용평가방식을 이용하는 것은 한계가 있음이 드러나고 있다.
- [0003] 소규모 온라인 사업자들은 매출이 적고 일정하지 않으며 재무제표를 갖추고 있지 않기 때문에, 기존에 신용도를 평가하는 방법에 의할 경우 사업 활동 대비 신용도가 낮은 것으로 분석되어 은행의 중금리 대출을 이용하기 어렵다. 따라서 사업자들은 자금을 융통하지 못하여 사업 확장에 난항을 겪거나, 어쩔 수 없이 금리가 더 높은 대출을 활용하게 된다. 따라서 소규모 온라인 사업자들의 운영자금 확보와 매출 확장을 도모하기 위해, 소규모 온라인 사업자들의 신용도를 평가하기에 적합한 신용평가 기준의 필요성이 대두된다.

선행기술문헌

특허문헌

- [0004] (특허문헌 0001) 국내 등록특허 제2109162(2019. 11. 19.)는 인공지능을 이용한 소상공인 대출가능여부 및 경영분석 시스템에 관하여 개시한다.

발명의 내용

해결하려는 과제

- [0005] 본 개시는 전술한 배경기술에 대응하여 안출된 것으로, 재무제표가 아닌 사업자의 영업활동 데이터를 기반으로 사업자의 현재 및 미래의 영업활동에 대해 정확히 예측할 수 있는 방법을 만드는 것을 목적으로 한다.

과제의 해결 수단

- [0006] 전술한 바와 같은 과제를 실현하기 위한 본 개시의 일 실시예에 따라, 사업자 활동 데이터를 분석하기 위한 방법이 개시된다. 상기 방법은, 사업자 활동을 분석하기 위해 컴퓨팅 장치에 의해 수행되는 방법으로서, 사업자 활동 데이터를 수집하는 단계; 수집된 상기 사업자 활동 데이터를 기반으로, 분석 대상인 주요 변수의 조합을 결정하는 단계; 상기 결정된 주요 변수의 조합을 기반으로, 사업자 점수 데이터를 산출하는 동작 또는 예측 모델을 이용하여 사업자의 미래 지표에 대한 예측치 정보를 생성하는 동작 중 적어도 하나를 수행하는 단계를 포함할 수 있다.
- [0007] 대안적 실시예에서, 상기 사업자 활동 데이터를 수집하는 단계는, 하나 이상의 사업자 데이터베이스로부터 상기 사업자 활동 데이터를 일정 주기 간격으로 크롤링하여 수집하는 단계; 및 수집된 상기 사업자 활동 데이터를 상기 크롤링 주기에 따라 업데이트하는 단계;를 포함할 수 있다.
- [0008] 대안적 실시예에서, 상기 사업자 활동 데이터는, 상기 사업자 데이터베이스에 기록된, 상품 분류 정보 및 매출과 관련된 데이터; 및 상기 사업자 데이터베이스에 기록된, 상품 정보, 사업자 정보 또는 매출 정보 중 적어도 하나와 관련된 데이터를 포함할 수 있다.
- [0009] 대안적 실시예에서, 상기 사업자 활동 데이터는, 상기 사업자 데이터베이스에 포함된 정보에 관한 증감 추이 데이터를 더 포함할 수 있다.
- [0010] 대안적 실시예에서, 수집된 상기 사업자 활동 데이터를 기반으로, 분석 대상인 주요 변수의 조합을 결정하는 단계는, 하나 이상의 변수 선택 알고리즘으로 구성된 변수 선택 모델을 기초로, 사업자 활동 지표에 영향을 미치는 변수 중 N개의 주요 변수들을 추출하는 단계; 상기 N개의 주요 변수들로부터 M개의 주요 변수의 조합(subset)들을 생성하는 단계; 상기 조합들 각각을 하나 이상의 예측 알고리즘으로 구성된 예측 모델을 통해 분석하는 단계; 및 상기 예측 모델을 통해 분석한 결과에 기초하여 가장 정확한 예측을 한 것으로 분석된 조합을 도출하고, 도출된 조합을 최종 분석 대상으로 결정하는 단계를 포함할 수 있다.
- [0011] 대안적 실시예에서, 상기 하나 이상의 변수 선택 알고리즘으로 구성된 변수 선택 모델을 기초로, 사업자 활동

지표에 영향을 미치는 변수 중 N개의 주요 변수를 추출하는 단계는: 상기 변수들 각각이 사업자 활동 지표에 미치는 영향을 상기 하나 이상의 변수 선택 알고리즘을 통해 분석하고 표준점수의 형태로 도출하는 단계; 상기 변수들 각각에 대하여 상기 변수 선택 알고리즘별 표준점수를 합산하는 단계; 및 합산된 상기 변수 선택 알고리즘별 표준점수를 비교하여 상기 복수의 변수 중 상위 N개의 변수를 추출하는 단계를 포함할 수 있다.

[0012] 대안적 실시예에서, 상기 N개의 주요 변수로부터 M개의 주요 변수의 조합을 생성하는 단계는, 상기 N개의 변수들 중 순서에 상관없이 K개의 주요 변수들을 추출하는 동작을 M회 수행하여, M개의 조합들을 생성하는 단계를 포함할 수 있다.

[0013] 대안적 실시예에서, 상기 예측 모델을 통해 분석한 결과에 기초하여 가장 정확한 예측을 한 것으로 분석된 조합을 도출하고, 도출된 조합을 최종 분석 대상으로 결정하는 단계는, 상기 예측 모델을 통해 분석한 결과에 기초하여 가장 정확한 예측을 한 것으로 분석된 조합과 예측 알고리즘의 쌍을 도출하고, 도출된 조합을 최종 분석 대상으로 결정하고, 도출된 예측 알고리즘을 최종 분석에 사용할 예측 알고리즘으로 결정하는 단계를 포함할 수 있다.

[0014] 대안적 실시예에서, 상기 예측 모델을 통해 분석한 결과에 기초하여 가장 정확한 예측을 한 것으로 분석된 조합과 예측 알고리즘의 쌍을 도출하는 것은: 상기 M개의 조합을 상기 예측 모델에 속하는 하나 이상의 예측 알고리즘을 통해 분석한 결과를 표준화하는 것; 및 상기 표준화된 점수가 가장 높게 판정된 하나의 조합과 그 조합의 분석에 사용한 상기 예측 알고리즘 쌍을 도출하는 것을 포함할 수 있다.

[0015] 대안적 실시예에서, 상기 M개의 조합을 상기 예측 모델에 속하는 하나 이상의 예측 알고리즘을 통해 분석한 결과를 표준화하는 것은, 정확도(Accuracy); 정밀도(Precision); 재현율(Recall); 또는 F1점수; 중 적어도 하나의 정보를 활용하는 것을 포함할 수 있다.

[0016] 대안적 실시예에서, 상기 사업자 점수 데이터는: 상기 결정된 주요 변수의 조합, 상기 조합에 포함된 주요 변수의 중요도, 및 상기 조합을 결정하기 위해 활용된 변수 선택 모델을 기반으로 계산된 정규화된 데이터를 포함할 수 있다.

[0017] 대안적 실시예에서, 상기 예측 모델을 이용하여 사업자의 미래 지표에 대한 예측치 정보를 생성하는 동작은: 상기 결정된 주요 변수의 조합과 상기 예측 모델을 기반으로 사업자의 미래 지표에 대한 예측치 정보를 계산하는 단계; 및 계산된 상기 사업자의 미래 지표에 대한 예측치 정보를 복수의 카테고리 중 하나의 카테고리로 분류하는 단계를 포함할 수 있다.

[0018] 전술한 바와 같은 과제를 실현하기 위한 본 개시의 일 실시예에 따라, 사업자 활동 데이터를 분석하기 위한 컴퓨터 프로그램이 개시된다. 상기 프로그램은, 사업자 활동 데이터를 수집하는 동작; 수집된 상기 사업자 활동 데이터를 기반으로, 분석 대상인 주요 변수의 조합을 결정하는 동작; 및 상기 결정된 주요 변수의 조합 기반으로, 사업자 점수 데이터를 산출하는 동작 또는 예측 모델을 이용하여 사업자의 미래 지표에 대한 예측치 정보를 생성하는 동작 중 적어도 하나를 수행하는 동작; 을 포함할 수 있다.

[0019] 전술한 바와 같은 과제를 실현하기 위한 본 개시의 일 실시예에 따라, 사업자 활동 데이터를 분석하기 위한 컴퓨팅 장치가 개시된다. 상기 컴퓨팅 장치는 적어도 하나의 프로세서 및 메모리를 포함하되, 상기 적어도 하나의 프로세서는, 사업자 활동 데이터를 수집하고; 수집된 상기 사업자 활동 데이터를 기반으로, 분석 대상인 주요 변수들의 조합을 결정하고; 그리고 상기 결정된 주요 변수의 조합을 기반으로, 사업자 점수 데이터를 산출하는 동작 또는 예측 모델을 이용하여 사업자의 미래 지표에 대한 예측치 정보를 생성하는 동작 중 적어도 하나를 수행하도록 구성될 수 있다.

발명의 효과

[0021] 본 개시는 인공지능 모델에 기초하여 사업자 활동 데이터를 분석하고, 그 결과물을 제공할 수 있다.

도면의 간단한 설명

[0022] 본 개시의 실시 예의 설명에 이용되기 위하여 첨부된 아래 도면들은 본 개시의 실시 예들 중 단지 일부일 뿐이며, 본 개시가 속한 기술분야에서 통상의 지식을 가진 사람(이하 "통상의 기술자"라 함)에게 있어서는 신규한 발명에 이르는 노력 없이도 이 도면들에 기초하여 다른 도면들이 얻어질 수 있다.

도 1은 본 개시의 일 실시예에 따른 사업자 활동 데이터를 분석하기 위한 동작을 수행하는 컴퓨팅 장치의 블록 구성도이다.

도 2는 본 개시의 일 실시예에 따른 네트워크 함수를 나타낸 개략도이다.

도 3은 본 개시의 일 실시예에 따른 사업자 활동 데이터를 분석하고 사업자 점수 또는 사업자의 미래 지표에 대한 예측치를 생성하는 방법에 대한 순서도이다.

도 4는 본 개시의 일 실시예에 따른 조합을 도출하기 위한 과정을 나타낸 개념도이다.

도 5는 본 개시의 일 실시예에 따른 컴퓨팅 장치의 블록 구성도이다.

발명을 실시하기 위한 구체적인 내용

- [0023] 본 개시는 사업자 활동 데이터를 수집하여 사업자 활동 데이터에 큰 영향을 미치는 주요 변수들을 선정하고, 그 주요 변수들을 기반으로 하여 사업자 점수 데이터 또는 사업자의 미래 지표에 대한 예측치 정보를 제공함으로써 사업자의 활동을 평가할 수 있는 지표를 만드는 방법에 대해 개시한다.
- [0024] 다양한 실시예들이 이제 도면을 참조하여 설명된다. 본 명세서에서, 다양한 설명들이 본 개시의 이해를 제공하기 위해서 제시된다. 그러나, 이러한 실시예들은 이러한 구체적인 설명 없이도 실행될 수 있음이 명백하다.
- [0025] 본 명세서에서 사용되는 용어 "컴포넌트", "모듈", "시스템" 등은 컴퓨터-관련 엔티티, 하드웨어, 펌웨어, 소프트웨어, 소프트웨어 및 하드웨어의 조합, 또는 소프트웨어의 실행을 지칭한다. 예를 들어, 컴포넌트는 프로세서 상에서 실행되는 처리과정(procedure), 프로세서, 객체, 실행 스레드, 프로그램, 및/또는 컴퓨터일 수 있지만, 이들로 제한되는 것은 아니다. 예를 들어, 컴퓨팅 장치에서 실행되는 애플리케이션 및 컴퓨팅 장치 모두 컴포넌트일 수 있다. 하나 이상의 컴포넌트는 프로세서 및/또는 실행 스레드 내에 상주할 수 있다. 일 컴포넌트는 하나의 컴퓨터 내에 로컬화 될 수 있다. 일 컴포넌트는 2개 이상의 컴퓨터들 사이에 분배될 수 있다. 또한, 이러한 컴포넌트들은 그 내부에 저장된 다양한 데이터 구조들을 갖는 다양한 컴퓨터 판독가능한 매체로부터 실행할 수 있다. 컴포넌트들은 예를 들어 하나 이상의 데이터 패킷들을 갖는 신호(예를 들면, 로컬 시스템, 분산 시스템에서 다른 컴포넌트와 상호작용하는 하나의 컴포넌트로부터의 데이터 및/또는 신호를 통해 다른 시스템과 인터넷과 같은 네트워크를 통해 전송되는 데이터)에 따라 로컬 및/또는 원격 처리들을 통해 통신할 수 있다.
- [0026] 더불어, 용어 "또는"은 배타적 "또는"이 아니라 내포적 "또는"을 의미하는 것으로 의도된다. 즉, 달리 특정되지 않거나 문맥상 명확하지 않은 경우에, "X는 A 또는 B를 이용한다"는 자연적인 내포적 치환 중 하나를 의미하는 것으로 의도된다. 즉, X가 A를 이용하거나; X가 B를 이용하거나; 또는 X가 A 및 B 모두를 이용하는 경우, "X는 A 또는 B를 이용한다"가 이들 경우들 어느 것으로도 적용될 수 있다. 또한, 본 명세서에 사용된 "및/또는"이라는 용어는 열거된 관련 아이템들 중 하나 이상의 아이템의 가능한 모든 조합을 지칭하고 포함하는 것으로 이해되어야 한다.
- [0027] 또한, "포함한다" 및/또는 "포함하는"이라는 용어는, 해당 특징 및/또는 구성요소가 존재함을 의미하는 것으로 이해되어야 한다. 다만, "포함한다" 및/또는 "포함하는"이라는 용어는, 하나 이상의 다른 특징, 구성요소 및/또는 이들의 그룹의 존재 또는 추가를 배제하지 않는 것으로 이해되어야 한다. 또한, 달리 특정되지 않거나 단수 형태를 지시하는 것으로 문맥상 명확하지 않은 경우에, 본 명세서와 청구범위에서 단수는 일반적으로 "하나 또는 그 이상"을 의미하는 것으로 해석되어야 한다.
- [0028] 그리고, "A또는 B 중 적어도 하나"이라는 용어는, "A만을 포함하는 경우", "B만을 포함하는 경우", "A와 B의 구성으로 조합된 경우"를 의미하는 것으로 해석되어야 한다.
- [0029] 당업자들은 추가적으로 여기서 개시된 실시예들과 관련되어 설명된 다양한 예시적 논리적 블록들, 구성들, 모듈들, 회로들, 수단들, 로직들, 및 알고리즘 단계들이 전자 하드웨어, 컴퓨터 소프트웨어, 또는 양쪽 모두의 조합들로 구현될 수 있음을 인식해야 한다. 하드웨어 및 소프트웨어의 상호교환성을 명백하게 예시하기 위해, 다양한 예시적 컴포넌트들, 블록들, 구성들, 수단들, 로직들, 모듈들, 회로들, 및 단계들은 그들의 기능성 측면에서 일반적으로 위에서 설명되었다. 그러한 기능성이 하드웨어로 또는 소프트웨어로서 구현되는지 여부는 전반적인 시스템에 부과된 특정 어플리케이션(application) 및 설계 제한들에 달려 있다. 숙련된 기술자들은 각각의 특정 어플리케이션들을 위해 다양한 방법들로 설명된 기능성을 구현할 수 있다. 다만, 그러한 구현의 결정들이 본 개시내용의 영역을 벗어나게 하는 것으로 해석되어서는 안된다.
- [0030] 제시된 실시예들에 대한 설명은 본 개시의 기술 분야에서 통상의 지식을 가진 자가 본 개시를 이용하거나 또는

실시할 수 있도록 제공된다. 이러한 실시예들에 대한 다양한 변형들은 본 개시의 기술 분야에서 통상의 지식을 가진 자에게 명백할 것이다. 여기에 정의된 일반적인 원리들은 본 개시의 범위를 벗어남이 없이 다른 실시예들에 적용될 수 있다. 그리하여, 본 개시는 여기에 제시된 실시예들로 한정되는 것이 아니다. 본 개시는 여기에 제시된 원리들 및 신규한 특징들과 일관되는 최광의의 범위에서 해석되어야 할 것이다.

- [0031] 본 개시에서 네트워크 함수와 인공 신경망 및 뉴럴 네트워크(neural network)는 상호 교환 가능하게 사용될 수 있다.
- [0032] 한편, 본 개시의 상세한 설명 및 청구항들에 걸쳐 이용된 "사업자 활동 데이터" 또는 "사업자 데이터"라는 용어는 온라인 쇼핑몰에서 직접 크롤링하거나 온라인 쇼핑몰이 제공하는 API(Application Programming Interface)를 이용하여 크롤링된 사업자의 온라인 상점 정보, 판매하는 상품의 카테고리, 판매하는 상품의 명칭, 상품별 매출, 상품별 배송 정보, 상품별 대기 물량, 상품별 리뷰, 상품 페이지 방문자 수에 대한 데이터를 의미할 수 있다. 다만 본 개시는 이에 한정하지 아니한다.
- [0033] 도 1은 본 개시의 일 실시예에 따른 사업자 활동 데이터를 분석하는 컴퓨팅 장치의 블록 구성도를 도시한 도면이다.
- [0034] 도 1에 도시된 컴퓨팅 장치(100)의 구성은 간략화 하여 나타난 예시일 뿐이다. 본 개시의 일 실시예에서 컴퓨팅 장치(100)는 컴퓨팅 장치(100)의 컴퓨팅 환경을 수행하기 위한 다른 구성들이 포함될 수 있고, 개시된 구성들 중 일부만이 컴퓨팅 장치(100)를 구성할 수도 있다.
- [0035] 컴퓨팅 장치(100)는 프로세서(110), 메모리(130), 네트워크부(150)를 포함할 수 있다.
- [0036] 프로세서(110)는 하나 이상의 코어로 구성될 수 있으며, 컴퓨팅 장치의 중앙 처리 장치(CPU: central processing unit), 범용 그래픽 처리 장치 (GPGPU: general purpose graphics processing unit), 텐서 처리 장치(TPU: tensor processing unit) 등의 데이터 분석, 딥러닝을 위한 프로세서를 포함할 수 있다. 프로세서(110)는 메모리(130)에 저장된 컴퓨터 프로그램을 판독하여 본 개시의 일 실시예에 따른 기계 학습을 위한 데이터 처리를 수행할 수 있다. 본 개시의 일 실시예에 따라 프로세서(110)는 신경망의 학습을 위한 연산을 수행할 수 있다. 프로세서(110)는 딥러닝(DL: deep learning)에서 학습을 위한 입력 데이터의 처리, 입력 데이터에서의 피쳐 추출, 오차 계산, 역전파(backpropagation)를 이용한 신경망의 가중치 업데이트 등의 신경망의 학습을 위한 계산을 수행할 수 있다. 프로세서(110)의 CPU, GPGPU, 및 TPU 중 적어도 하나가 네트워크 함수의 학습을 처리할 수 있다. 예를 들어, CPU 와 GPGPU가 함께 네트워크 함수의 학습, 네트워크 함수를 이용한 데이터 분류를 처리할 수 있다. 또한, 본 개시의 일 실시예에서 복수의 컴퓨팅 장치의 프로세서를 함께 사용하여 네트워크 함수의 학습, 네트워크 함수를 이용한 데이터 분류를 처리할 수 있다. 또한, 본 개시의 일 실시예에 따른 컴퓨팅 장치에서 수행되는 컴퓨터 프로그램은 CPU, GPGPU 또는 TPU 실행가능 프로그램일 수 있다.
- [0037] 본 개시의 일 실시예에 따르면, 프로세서(110)는 복수의 온라인 쇼핑 플랫폼에서 사업자 활동 데이터를 수집할 수 있다. 예를 들어, 온라인 쇼핑 플랫폼은 G마켓, SSG, 네이버 스마트스토어 등 쇼핑몰이 될 수 있다. 대안적인 실시예에서, 프로세서(110)는 데이터를 수집하는 데 있어서 네이버 스마트스토어에서 직접 사업자 데이터를 크롤링하거나 네이버 데이터랩 API를 활용하여 네이버 스마트스토어에 등록된 온라인 사업자들의 상품에 대한 정보를 수집하여 메모리(120)에 저장하는 방법을 사용할 수 있으나, 본 개시는 이에 한하지 아니한다.
- [0038] 프로세서(110)는 수집된 데이터 및 주요 변수를 업데이트하기 위하여 일정한 주기로 데이터를 다시 수집할 수 있다. 예를 들어, 프로세서는 하루에 한 번 네이버 스마트스토어에서 데이터를 수집하고, 아래에 서술할 변수의 증감 데이터를 생성하는 기초 데이터로서 사용하는 한편 사업자 활동 데이터에 영향을 미치는 주요 데이터의 변화 양상을 추적할 수 있다.
- [0039] 본 개시의 일 실시예에 따르면, 프로세서(110)는 상기 사업자 활동 데이터로부터 사업자의 활동 지표, 예를 들어 매출에 영향을 미치는 변수들을 추출할 수 있다. 예를 들어 네이버 스마트스토어의 경우, 변수들은 '상품 등록 기간', '상품 가격 할인률', '누적 상품 찜 수', '리뷰 점수', '스토어 총 방문자 수', '상품 평균 배송기간', '스토어 찜 수', '방문자 수 증감', '평균 리뷰 수', '평균 방문자 수', '상품가격 변화율', '상품 찜 수 증감', '리뷰 점수 증감', '리뷰 수 증감', '스토어 찜 수 증감', '업력', '고용인원 수', '당월 국민연금 고지금액', '상품 카테고리별 광고 노출 순위'등을 포함할 수 있다.
- [0040] 어떤 변수는 데이터 수집 주기마다 변화하는 변수의 값을 반영한 변수의 추이일 수 있다. 예를 들어, 사업자의 활동 지표에 영향을 미치는 변수들 중 '방문자 수 증감', '리뷰 점수 증감', '스토어 찜 수 증감', '상품 찜 수

증감'등은 데이터 수집 주기마다 변화하는 데이터의 증감 또는 추이를 표시하는 변수들일 수 있다.

- [0041] 본 개시의 일 실시예에 따르면, 프로세서(110)는 상기 변수들 중 사업자 활동 데이터와 연관성이 높은 변수들을 주요 변수로서 추출할 수 있다. 이 과정에서, 프로세서(110)는 주요 변수를 추출하기 위해 변수 선택 모델을 사용하되, 추출 과정에서의 신뢰도를 높이기 위해 변수 선택 모델에 포함된 하나 이상의 알고리즘들을 사용하여 각 변수가 사업자 활동 데이터에 미치는 영향을 분석할 수 있다. 예를 들어, 변수 선택 모델에 포함된 알고리즘은 'Correlation Analysis', 'Backward Elimination', 'Recursive Feature Elimination', 'Chi-Square'를 포함할 수 있다.
- [0042] 위와 같이 사업자 활동 지표에 영향을 미치는 변수 전체가 아닌 주요 변수를 추출하여 연산을 수행하므로, 본 발명에서는 사업자 활동 지표에 영향을 미치는 요인을 파악하기 위한 데이터를 처리하는 데 있어서 time complexity와 space complexity를 최소화하는 동시에 사업자 활동 지표에 대한 높은 수준의 추론이 가능하게 된다.
- [0043] 프로세서(110)는 변수 선택 모델을 사용하여 각 알고리즘별로 각 변수들이 사업자 활동에 미치는 영향을 표준화된 점수의 형태로 계산하고, 하나의 변수에 대한 알고리즘별 점수를 합산 후 표준화하여 각 변수들이 사업자 활동에 미치는 영향을 점수화할 수 있다.
- [0044] 프로세서(110)는 각 변수들이 사업자 활동에 미치는 영향이 큰 순서대로 N개의 변수를 주요 변수로서 선택할 수 있다. 예를 들어, 위와 같은 방법으로 점수화된 사업자 활동에 영향을 미치는 19개의 변수 중 점수가 가장 높은 8개의 변수가 주요 변수로서 선택될 수 있다. 이렇게 선택된 주요 변수는 상품 또는 상품 카테고리마다 상이할 수 있으며, 주요 변수가 해당 상품 또는 카테고리에 영향을 미치는 정도 또한 상품 또는 상품 카테고리마다 다를 수 있다.
- [0045] 이렇게 선택된 주요 변수들은 최초 한 번의 선택으로 고정되는 것이 아니라, 데이터의 크롤링 주기에 따라 업데이트된 사업자 활동 데이터를 반영하여 변동될 수 있다.
- [0046] 프로세서(110)는 추출된 주요 변수들 중 어떤 조합이 사업자 활동에 미치는 영향이 가장 큰지를 검증하기 위하여, N개의 주요 변수 중 정해진 개수의 일부 변수를 선택하여 M개의 조합(Subset)을 생성할 수 있다. 예를 들어, 프로세서(110)는 8개의 주요 변수 중 6개를 순서와 관계없이 선택하는 동작을 수 차례 수행하여 조합을 생성할 수 있다. 이 경우, 생성되는 조합의 개수는 8C_6 을 계산한 결과인 28개이다.
- [0047] 본 개시의 일 실시예에 따르면 프로세서(110)는 기계학습 알고리즘을 사용하여 생성된 조합이 상품별 사업자 활동에 얼마나 큰 영향을 미치는지 검증할 수 있다. 이 때, 업종별, 상품 카테고리별 상품의 특징과 구매자의 특성이 다르기 때문에 한 가지 기계학습 알고리즘을 통해 모든 카테고리의 상품을 분석하는 것은 신뢰성을 담보할 수 없으므로, 수 개의 기계학습 알고리즘이 분석에 사용될 수 있다. 예를 들어, 프로세서(110)는 선택된 조합이 사업자 활동에 영향을 미치는 정도를 파악하기 위해 'Random Forest', 'Catboost', 'Logistic Regression'과 같은 예측 알고리즘을 사용하여 분석의 신뢰성을 높일 수 있다.
- [0048] 각 조합을 각 기계학습 알고리즘이 분석한 결과가 도출되면, 프로세서(110)는 그 중 사업자 활동에 가장 큰 영향을 미치는 조합과 예측 알고리즘 쌍을 선택할 수 있다. 예를 들어, 주요 변수 8개 중 6개를 선택하여 조합 28개를 생성하고, 분석을 위해 3종류의 예측 알고리즘을 사용한 경우, 한 종류의 상품 또는 카테고리에 대하여 총 84개의 결과가 생성될 수 있다. 그 후 프로세서(110)는 84개의 결과 중 예측 모델을 통해 분석한 결과를 기초로 가장 정확한 예측을 한 조합과 예측 알고리즘 쌍을 채택하여, 도출된 조합을 최종 분석 대상으로 결정하고, 도출된 예측 알고리즘을 최종 분석에 사용할 예측 알고리즘으로 결정할 수 있다.
- [0049] 사업자 활동 데이터를 가장 정확하게 예측한 조합과 예측 알고리즘을 파악하기 위해, 각 분석 결과는 예측 결과와 실제 수치가 부합하는 정도에 따라 점수의 형태로 표준화될 수 있다. 여기서, 표준화된 점수를 도출하는 평가 요소로서 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), F1점수 중 적어도 하나의 정보가 활용될 수 있으나, 본 개시는 이와 같은 평가 요소로만 한정하지 아니한다. 여기서 F1점수란 다음 수식을 의미한다.

$$F1\ score = 2 \times \frac{precision \times recall}{precision + recall}$$

[0050]

- [0051] 이렇게 채택된 조합과 예측 알고리즘 쌍은, 본 발명이 사업자 활동 데이터를 일정 주기로 크롤링하여 데이터를 업데이트하기 때문에, 주기마다 다른 값이 될 수 있다. 이에 따라 사업자 활동에 가장 큰 영향을 미치는 지표

와 분석 알고리즘이 동적인 형태를 띠므로 본 발명은 사업자의 변화하는 상황에 대해 실시간으로 대응할 수 있고, 오래된 데이터를 계속 저장할 필요가 없다. 이로써 예측에 있어 정확성을 유지하는 한편 데이터 저장의 효율성 또한 향상시킬 수 있다.

[0052] 본 개시의 일 실시예에 따르면, 프로세서(110)는 사업자 활동 데이터의 증감을 평가할 수 있는 사업자 점수를 생성할 수 있다. 사업자 점수는 사업자 활동 데이터를 기반으로 도출된 점수로써 사업자의 영업활동 상황을 지표의 형태로 보여주는 점수일 수 있다.

[0053] 사업자 점수는 평가 방법을 통해 도출된 주요 변수의 조합과 조합에 포함된 주요 변수의 중요도, 그리고 상기 조합을 결정하기 위해 활용된 변수 선택 모델을 기반으로 계산된 정규화된 데이터일 수 있다. 구체적으로, 사업자 점수는 이전 과정에서 결정된 주요 변수의 조합(즉, 이전 과정에서 선택된 사업자 활동 데이터를 가장 잘 예측하는 것으로 결정된 주요 변수의 조합)을 구성하는 각각의 주요 변수들에 대하여, "변수 선택 과정에서 도출된 변수 선택 알고리즘별 점수를 합산하고 표준화한 점수(가중치)"와, "각 사업자의 해당 변수에 대한 표준 점수"의 곱을 합산한 결과물일 수 있다.

[0054] 예를 들어, '건강식품' 카테고리의 상품의 데이터를 가장 잘 예측하는 주요 변수의 조합으로 채택된 6개의 변수인 '리뷰 수 증감', '상품 점수', '평균 리뷰 수', '상품 점수 증감', '카테고리 순위' 및 '평균 배송일'이 결정된 경우를 가정할 수 있다. 여기서 각 사업자의 리뷰 수 증감에 대한 표준 점수(Fs_1)는 사업자의 리뷰 수 증감에 대한 데이터를 0에서부터 5사이의 실수값으로 표준화한 형태일 수 있다. 여기서 리뷰 수 증감에 대한 가중치(Fw_1)는 사업자의 영업활동에 있어서 리뷰 수 증감이라는 변수가 차지하는 중요도(예를 들어, 해당 변수에 대하여, 변수 선택 과정에서 도출된 변수 선택 알고리즘별 점수를 합산하는 것에 의해 정량화된 수치일 수 있다)를 0에서부터 5사이의 실수값으로 표준화한 형태일 수 있다. 같은 방법으로, 각 사업자의 상품 점수에 대한 표준 점수(Fs_2)는 사업자의 상품 점수에 대한 데이터를 0에서부터 5사이의 실수값으로 표준화한 형태일 수 있다. 여기서 상품 점수에 대한 가중치(Fw_2)는 사업자의 영업활동에 있어서 상품 점수라는 변수가 차지하는 중요도를 0에서부터 5사이의 실수값으로 표준화한 형태일 수 있다. 이와 같이 주요 변수 조합을 구성하는 6개의 변수 각각에 대하여 표준 점수(Fs_n)에 변수 선택 과정에서 도출된 변수 선택 알고리즘별 점수를 합산하고 표준화한 점수를 가중치(Fw_n)로서 곱한 다음, 곱한 6개의 값을 전부 합산하여 표준화한 값(Y)이 사업자 점수일 수 있다. 이 경우 최종적으로 계산된 사업자 점수는 아래와 같은 식으로 표현될 수 있으나, 본 개시는 이와 같은 계산방법에 한정하지 아니한다.

[0055] 사업자 점수 $Y = Fs_1 \times Fw_1 + Fs_2 \times Fw_2 + Fs_3 \times Fw_3 + Fs_4 \times Fw_4 + Fs_5 \times Fw_5 + Fs_6 \times Fw_6$

[0056] (Fs_n =n번째 변수의 표준 점수, Fw_n =n번째 변수의 가중치)

[0057] 사업자 점수는 각 사업자들의 변수별 활동 정도와 주요 변수 각각의 중요도 모두를 반영하여 계산된 점수에 해당한다. 따라서 이렇게 계산된 사업자 점수는 해당 사업자의 사업자 활동을 평가할 수 있는 척도, 예를 들어 대안적인 신용평가의 기초 자료로서 사용될 수 있다.

[0058] 본 개시의 일 실시예에 따르면, 사업자 활동 데이터를 가장 잘 예측하는 것으로 판명된 주요 변수의 조합과 예측 모델을 기반으로 사업자의 미래 지표에 대한 예측치 정보를 생성할 수 있다. 이렇게 생성된 사업자의 미래 지표에 대한 예측치 정보는 미리 생성된 복수의 카테고리 중 하나의 카테고리로 분류되어, 향후 사업자 활동을 예측하기 위한 가이드라인으로서 사용될 수 있다.

[0059] 예를 들어, 사업자의 미래 지표에 대한 예측치 정보는 '대폭 감소', '소폭 증감', '대폭 증가' 세 가지의 카테고리로 분류되어 사업자 활동을 예측할 수 있는 정보를 제공할 수 있으나, 본 개시는 이와 같은 카테고리에 한정하지 아니한다.

[0060] 본 개시의 일 실시예에 따르면, 메모리(130)는 프로세서(110)가 생성하거나 결정한 임의의 형태의 정보 및 네트워크부(150)가 수신한 임의의 형태의 정보를 저장할 수 있다.

[0061] 본 개시의 일 실시예에 따르면, 메모리(130)는 플래시 메모리 타입(flash memory type), 하드디스크 타입(hard disk type), 멀티미디어 카드 마이크로 타입(multimedia card micro type), 카드 타입의 메모리(예를 들어 SD

또는 XD 메모리 등), 램(Random Access Memory, RAM), SRAM(Static Random Access Memory), 롬(Read-Only Memory, ROM), EEPROM(Electrically Erasable Programmable Read-Only Memory), PROM(Programmable Read-Only Memory), 자기 메모리, 자기 디스크, 광디스크 중 적어도 하나의 타입의 저장매체를 포함할 수 있다. 컴퓨팅 장치(100)는 인터넷(internet) 상에서 상기 메모리(130)의 저장 기능을 수행하는 웹 스토리지(web storage)와 관련되어 동작할 수도 있다. 전술한 메모리에 대한 기재는 예시일 뿐, 본 개시는 이에 제한되지 않는다.

[0062] 본 개시의 일 실시예에 따른 네트워크부(150)는 공중전화 교환망(PSTN: Public Switched Telephone Network), xDSL(x Digital Subscriber Line), RADSL(Rate Adaptive DSL), MDSL(Multi Rate DSL), VDSL(Very High Speed DSL), UADSL(Universal Asymmetric DSL), HDSL(High Bit Rate DSL) 및 근거리 통신망(LAN) 등과 같은 다양한 유선 통신 시스템들을 사용할 수 있다.

[0063] 또한, 본 명세서에서 제시되는 네트워크부(150)는 CDMA(Code Division Multi Access), TDMA(Time Division Multi Access), FDMA(Frequency Division Multi Access), OFDMA(Orthogonal Frequency Division Multi Access), SC-FDMA(Single Carrier-FDMA) 및 다른 시스템들과 같은 다양한 무선 통신 시스템들을 사용할 수 있다.

[0064] 본 개시에서 네트워크부(150)는 임의의 형태의 유무선 통신 시스템을 사용할 수 있다.

[0065] 본 명세서에서 설명된 기술들은 위에서 언급된 네트워크들뿐만 아니라, 다른 네트워크들에서도 사용될 수 있다.

[0066] 도 2는 본 개시의 일 실시예에 따라 네트워크 함수를 나타낸 개략도이다.

[0067] 본 명세서에 걸쳐, 연산 모델, 신경망, 네트워크 함수, 뉴럴 네트워크(neural network)는 상호 교환 가능한 의미로 사용될 수 있다. 신경망은 일반적으로 노드라 지칭될 수 있는 상호 연결된 계산 단위들의 집합으로 구성될 수 있다. 이러한 노드들은 뉴런(neuron)들로 지칭될 수도 있다. 신경망은 적어도 하나 이상의 노드들을 포함하여 구성된다. 신경망들을 구성하는 노드(또는 뉴런)들은 하나 이상의 링크에 의해 상호 연결될 수 있다.

[0068] 신경망 내에서, 링크를 통해 연결된 하나 이상의 노드들은 상대적으로 입력 노드 및 출력 노드의 관계를 형성할 수 있다. 입력 노드 및 출력 노드의 개념은 상대적인 것으로서, 하나의 노드에 대하여 출력 노드 관계에 있는 임의의 노드는 다른 노드와의 관계에서 입력 노드 관계에 있을 수 있으며, 그 역도 성립할 수 있다. 상술한 바와 같이, 입력 노드 대 출력 노드 관계는 링크를 중심으로 생성될 수 있다. 하나의 입력 노드에 하나 이상의 출력 노드가 링크를 통해 연결될 수 있으며, 그 역도 성립할 수 있다.

[0069] 하나의 링크를 통해 연결된 입력 노드 및 출력 노드 관계에서, 출력 노드의 데이터는 입력 노드에 입력된 데이터에 기초하여 그 값이 결정될 수 있다. 여기서 입력 노드와 출력 노드를 상호 연결하는 링크는 가중치(weight)를 가질 수 있다. 가중치는 가변적일 수 있으며, 신경망이 원하는 기능을 수행하기 위해, 사용자 또는 알고리즘에 의해 가변될 수 있다. 예를 들어, 하나의 출력 노드에 하나 이상의 입력 노드가 각각의 링크에 의해 상호 연결된 경우, 출력 노드는 상기 출력 노드와 연결된 입력 노드들에 입력된 값들 및 각각의 입력 노드들에 대응하는 링크에 설정된 가중치에 기초하여 출력 노드 값을 결정할 수 있다.

[0070] 상술한 바와 같이, 신경망은 하나 이상의 노드들이 하나 이상의 링크를 통해 상호 연결되어 신경망 내에서 입력 노드 및 출력 노드 관계를 형성한다. 신경망 내에서 노드들과 링크들의 개수 및 노드들과 링크들 사이의 연관관계, 링크들 각각에 부여된 가중치의 값에 따라, 신경망의 특성이 결정될 수 있다. 예를 들어, 동일한 개수의 노드 및 링크들이 존재하고, 링크들의 가중치 값이 상이한 두 신경망이 존재하는 경우, 두 개의 신경망들은 서로 상이한 것으로 인식될 수 있다.

[0071] 신경망은 하나 이상의 노드들의 집합으로 구성될 수 있다. 신경망을 구성하는 노드들의 부분 집합은 레이어(layer)를 구성할 수 있다. 신경망을 구성하는 노드들 중 일부는, 최초 입력 노드로부터의 거리들에 기초하여, 하나의 레이어(layer)를 구성할 수 있다. 예를 들어, 최초 입력 노드로부터 거리가 n인 노드들의 집합은, n 레이어를 구성할 수 있다. 최초 입력 노드로부터 거리는, 최초 입력 노드로부터 해당 노드까지 도달하기 위해 거쳐야 하는 링크들의 최소 개수에 의해 정의될 수 있다. 그러나, 이러한 레이어의 정의는 설명을 위한 임의적인 것으로서, 신경망 내에서 레이어의 차수는 상술한 것과 상이한 방법으로 정의될 수 있다. 예를 들어, 노드들의 레이어는 최종 출력 노드로부터 거리에 의해 정의될 수도 있다.

[0072] 최초 입력 노드는 신경망 내의 노드들 중 다른 노드들과의 관계에서 링크를 거치지 않고 데이터가 직접 입력되는 하나 이상의 노드들을 의미할 수 있다. 또는, 신경망 네트워크 내에서, 링크를 기준으로 한 노드 간의 관계에 있어서, 링크로 연결된 다른 입력 노드들을 가지지 않는 노드들을 의미할 수 있다. 이와 유사하게, 최종 출

력 노드는 신경망 내의 노드들 중 다른 노드들과의 관계에서, 출력 노드를 가지지 않는 하나 이상의 노드들을 의미할 수 있다. 또한, 히든 노드는 최초 입력 노드 및 최종 출력 노드가 아닌 신경망을 구성하는 노드들을 의미할 수 있다.

[0073] 본 개시의 일 실시예에 따른 신경망은 입력 레이어의 노드의 개수가 출력 레이어의 노드의 개수와 동일할 수 있으며, 입력 레이어에서 히든 레이어로 진행됨에 따라 노드의 수가 감소하다가 다시 증가하는 형태의 신경망일 수 있다. 또한, 본 개시의 다른 일 실시예에 따른 신경망은 입력 레이어의 노드의 개수가 출력 레이어의 노드의 개수보다 적을 수 있으며, 입력 레이어에서 히든 레이어로 진행됨에 따라 노드의 수가 감소하는 형태의 신경망일 수 있다. 또한, 본 개시의 또 다른 일 실시예에 따른 신경망은 입력 레이어의 노드의 개수가 출력 레이어의 노드의 개수보다 많을 수 있으며, 입력 레이어에서 히든 레이어로 진행됨에 따라 노드의 수가 증가하는 형태의 신경망일 수 있다. 본 개시의 또 다른 일 실시예에 따른 신경망은 상술한 신경망들의 조합된 형태의 신경망일 수 있다.

[0074] 딥 뉴럴 네트워크(DNN: deep neural network, 심층신경망)는 입력 레이어와 출력 레이어 외에 복수의 히든 레이어를 포함하는 신경망을 의미할 수 있다. 딥 뉴럴 네트워크를 이용하면 데이터의 잠재적인 구조(latent structures)를 파악할 수 있다. 즉, 사진, 글, 비디오, 음성, 음악의 잠재적인 구조(예를 들어, 어떤 물체가 사진에 있는지, 글의 내용과 감정이 무엇인지, 음성의 내용과 감정이 무엇인지 등)를 파악할 수 있다. 딥 뉴럴 네트워크는 컨볼루션 뉴럴 네트워크(CNN: convolutional neural network), 리커런트 뉴럴 네트워크(RNN: recurrent neural network), 오토 인코더(auto encoder), GAN(Generative Adversarial Networks), 제한 볼츠만 머신(RBM: restricted boltzmann machine), 심층 신뢰 네트워크(DBN: deep belief network), Q 네트워크, U 네트워크, 삼 네트워크, 적대적 생성 네트워크(GAN: Generative Adversarial Network) 등을 포함할 수 있다. 전술한 딥 뉴럴 네트워크의 기재는 예시일 뿐이며 본 개시는 이에 제한되지 않는다.

[0075] 본 개시의 일 실시예에서 네트워크 함수는 오토 인코더(autoencoder)를 포함할 수도 있다. 오토 인코더는 입력 데이터와 유사한 출력 데이터를 출력하기 위한 인공 신경망의 일종일 수 있다. 오토 인코더는 적어도 하나의 히든 레이어를 포함할 수 있으며, 홀수 개의 히든 레이어가 입출력 레이어 사이에 배치될 수 있다. 각각의 레이어의 노드의 수는 입력 레이어의 노드의 수에서 병목 레이어(인코딩)라는 중간 레이어로 축소되었다가, 병목 레이어에서 출력 레이어(입력 레이어와 대칭)로 축소와 대칭되어 확장될 수도 있다. 오토 인코더는 비선형 차원 감소를 수행할 수 있다. 입력 레이어 및 출력 레이어의 수는 입력 데이터의 전처리 이후에 차원과 대응될 수 있다. 오토 인코더 구조에서 인코더에 포함된 히든 레이어의 노드의 수는 입력 레이어에서 멀어질수록 감소하는 구조를 가질 수 있다. 병목 레이어(인코더와 디코더 사이에 위치하는 가장 적은 노드를 가진 레이어)의 노드의 수는 너무 작은 경우 충분한 양의 정보가 전달되지 않을 수 있으므로, 특정 수 이상(예를 들어, 입력 레이어의 절반 이상 등)으로 유지될 수도 있다.

[0076] 뉴럴 네트워크는 교사 학습(supervised learning), 비교사 학습(unsupervised learning), 반교사학습(semi supervised learning), 또는 강화학습(reinforcement learning) 중 적어도 하나의 방식으로 학습될 수 있다. 뉴럴 네트워크의 학습은 뉴럴 네트워크가 특정한 동작을 수행하기 위한 지식을 뉴럴 네트워크에 적용하는 과정일 수 있다.

[0077] 뉴럴 네트워크는 출력의 오류를 최소화하는 방향으로 학습될 수 있다. 뉴럴 네트워크의 학습에서 반복적으로 학습 데이터를 뉴럴 네트워크에 입력시키고 학습 데이터에 대한 뉴럴 네트워크의 출력과 타겟의 에러를 계산하고, 에러를 줄이기 위한 방향으로 뉴럴 네트워크의 에러를 뉴럴 네트워크의 출력 레이어에서부터 입력 레이어 방향으로 역전파(backpropagation)하여 뉴럴 네트워크의 각 노드의 가중치를 업데이트 하는 과정이다. 교사 학습의 경우 각각의 학습 데이터에 정답이 라벨링되어있는 학습 데이터를 사용하며(즉, 라벨링된 학습 데이터), 비교사 학습의 경우는 각각의 학습 데이터에 정답이 라벨링되어 있지 않을 수 있다. 즉, 예를 들어 데이터 분류에 관한 교사 학습의 경우의 학습 데이터는 학습 데이터 각각에 카테고리가 라벨링 된 데이터 일 수 있다. 라벨링된 학습 데이터가 뉴럴 네트워크에 입력되고, 뉴럴 네트워크의 출력(카테고리)과 학습 데이터의 라벨을 비교함으로써 오류(error)가 계산될 수 있다. 다른 예로, 데이터 분류에 관한 비교사 학습의 경우 입력인 학습 데이터가 뉴럴 네트워크 출력과 비교됨으로써 오류가 계산될 수 있다. 계산된 오류는 뉴럴 네트워크에서 역방향(즉, 출력 레이어에서 입력 레이어 방향)으로 역전파 되며, 역전파에 따라 뉴럴 네트워크의 각 레이어의 각 노드들의 연결 가중치가 업데이트 될 수 있다. 업데이트 되는 각 노드의 연결 가중치는 학습률(learning rate)에 따라 변화량이 결정될 수 있다. 입력 데이터에 대한 뉴럴 네트워크의 계산과 에러의 역전파는 학습 사이클(epoch)을 구성할 수 있다. 학습률은 뉴럴 네트워크의 학습 사이클의 반복 횟수에 따라 상이하게 적용될 수 있다. 예를 들어, 뉴럴 네트워크의 학습 초기에는 높은 학습률을 사용하여 뉴럴 네트워크가 빠르게 일정 수준의 성능을 확보하도록 하

여 효율성을 높이고, 학습 후기에는 낮은 학습률을 사용하여 정확도를 높일 수 있다.

- [0078] 뉴럴 네트워크의 학습에서 일반적으로 학습 데이터는 실제 데이터(즉, 학습된 뉴럴 네트워크를 이용하여 처리하고자 하는 데이터)의 부분집합일 수 있으며, 따라서, 학습 데이터에 대한 오류는 감소하나 실제 데이터에 대해서는 오류가 증가하는 학습 사이클이 존재할 수 있다. 과적합(overfitting)은 이와 같이 학습 데이터에 과하게 학습하여 실제 데이터에 대한 오류가 증가하는 현상이다. 예를 들어, 노란색 고양이를 보여 고양이를 학습한 뉴럴 네트워크가 노란색 이외의 고양이를 보고는 고양이임을 인식하지 못하는 현상이 과적합의 일종일 수 있다. 과적합은 머신러닝 알고리즘의 오류를 증가시키는 원인으로 작용할 수 있다. 이러한 과적합을 막기 위하여 다양한 최적화 방법이 사용될 수 있다. 과적합을 막기 위해서는 학습 데이터를 증가시키거나, 레귤라이제이션(regularization), 학습의 과정에서 네트워크의 노드 일부를 비활성화하는 드롭아웃(dropout), 배치 정규화 레이어(batch normalization layer)의 활용 등의 방법이 적용될 수 있다.
- [0079] 도 3은 본 개시의 일 실시예에 따른 사업자 활동 데이터 분석 방법에 대한 순서도이다.
- [0080] S310단계에서, 컴퓨팅 장치(100)의 프로세서(110)는 적어도 하나 이상의 온라인 쇼핑 플랫폼으로부터 사업자 활동 데이터를 수집할 수 있다. 이 때 온라인 쇼핑 플랫폼은 G마켓, SSG, 네이버 스마트스토어 등 현재 국내외에서 서비스를 개시 중인 쇼핑 플랫폼이 될 수 있다. 예를 들어, 프로세서(110)는 네이버 스마트스토어에서 직접 사업자 데이터를 크롤링하거나 네이버 데이터랩 API를 활용하여 네이버 스마트스토어에 등록된 온라인 사업자들의 상품에 대한 정보를 수집하여 메모리(120)에 저장할 수 있다.
- [0081] S320단계에서, 프로세서(110)는 S310단계에서 수집된 사업자 활동 데이터를 기반으로 하여 사업자 활동 데이터에 영향을 미치는 하나의 주요 변수 조합과 예측 알고리즘의 쌍을 결정할 수 있다. 최종 분석 대상인 주요 변수 조합과 예측 알고리즘의 쌍을 결정하는 구체적인 과정은 도 4를 참조하여 후술한다.
- [0082] S330단계에서, 프로세서(110)는 사업자 활동 데이터에 영향을 미치는 주요 변수의 조합에 기초하여 사업자 점수 또는 사업자의 미래 지표에 대한 예측치를 생성할 수 있다. 예를 들어, 사업자 점수는 S320단계에서 도출된 조합을 구성하는 변수 6개 각각에 대하여, "사업자의 표준점수" 및 "가중치(예를 들어, 변수 선택 과정에서 도출된 변수 선택 알고리즘별 점수를 합산한 형태로 산출된 가중치)"를 곱한 다음, 각 변수에 대하여 곱해진 점수들을 합하여 표준화한 점수일 수 있다.
- [0083] 사업자의 미래 지표에 대한 예측치는 표준화된 점수로 표현될 수 있고 점수의 구간에 따라 세분화된 카테고리의 이름, 예를 들어 '대폭 증가', '소폭 증감', '대폭 감소'등으로 표현될 수 있으나, 본 개시는 이에 한정하지 아니한다.
- [0084] 도 4는 사업자 활동 데이터로부터 주요 변수의 조합 및 예측 알고리즘의 쌍을 도출하는 과정을 나타낸 개념도이다.
- [0085] 변수 선택 모델(420)은 사업자 활동 데이터로부터 사업자 활동 데이터에 영향을 미치는 변수를 추출할 수 있다. 이 때 변수 선택 모델(420)은 하나 이상의 변수 선택 알고리즘(421)들로 이루어져 있을 수 있다. 예를 들어, 변수 선택 모델(420)은 'Correlation Analysis', 'Backward Elimination', 'Recursive Feature Elimination', 'Chi-Square'등의 알고리즘들을 포함할 수 있다.
- [0086] 변수 선택 모델(420)이 사업자 활동 데이터(410)에 영향을 미치는 변수를 추출한 후, 프로세서(110)는 그 변수들로부터 사업자 활동 데이터에 큰 영향을 미치는 N개의 복수의 주요 변수(411)를 추출한다. 이 때 주요 변수들을 추출하기 위해서, 각 변수별로 변수 선택 알고리즘(421)각각에 대하여 변수와 사업자 활동 데이터의 연관성을 표준화된 점수로 표시할 수 있다. 그 후 프로세서(110)는 표준화된 변수 선택 알고리즘(421)별 점수를 모두 합하여 각 변수별로 점수를 계산한다. 그 후 프로세서(110)는 점수가 가장 높은 적어도 일부의 상위 변수들을 주요 변수(412)로서 채택할 수 있다.
- [0087] 그 후, 프로세서는 선택된 N개의 주요 변수들 중 정해진 수(예컨대, K개)의 일부 변수를 순서와 관계없이 반복적으로 M회 선택하여 M개의 조합(413)들을 생성할 수 있다.
- [0088] 예측 모델(430)은 프로세서(110)가 생성한 각 주요 변수의 조합들이 사업자 활동 데이터에 미치는 영향을 정량적으로 예측할 수 있다.
- [0089] 예측 모델(430)의 신뢰성을 확보하기 위해, 예측 모델은 하나 이상의 기계학습 알고리즘인 예측 알고리즘(431)들로 구성될 수 있다. 예를 들어, 예측 알고리즘은(431) 'Random Forest', 'Catboost', 'Logistic Regression'

과 같은 기계학습 알고리즘들을 포함할 수 있다.

[0090] 프로세서(110)는 조합(413)이 사업자 활동 데이터에 미치는 영향을 각 예측 알고리즘(431)별로 분석하고, 예측치를 표준화한 점수를 생성할 수 있다. 이에 따라 각 조합별, 예측 알고리즘별로 사업자 활동 데이터에 미치는 영향에 관한 예측치(414)를 얻을 수 있다. 예를 들어, 8개의 주요 변수들 중 순서와 상관없이 변수 6개를 고르는 동작을 반복하여 조합을 생성하고 3개의 예측 알고리즘들로 이루어진 예측 모델을 사용한 경우 얻어지는 결과(414)는 ${}^8C_6 \times 3$ 으로 총 84개가 된다. 프로세서는 이 중 가장 정확한 예측을 한 것으로 분석된 조합과 예측 알고리즘의 쌍(415)을 도출할 수 있으며, 도출된 조합을 최종 분석 대상으로 결정하고, 도출된 예측 알고리즘을 최종 분석에 사용할 예측 알고리즘으로 결정할 수 있다.

[0091] 여기서 예측치는 점수의 형태로 표준화될 수 있으며, 분석의 요소로서 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), 또는 F1점수와 같은 패러미터를 고려할 수 있다.

[0092] 이러한 과정을 통해 얻어진 조합과 예측 알고리즘 쌍은, 데이터를 이루는 카테고리에 따라 상이할 수 있으며, 각각의 사업자 활동 데이터의 카테고리를 가장 잘 설명할 수 있는 구성에 해당한다. 예를 들어, 사업자가 판매하는 '건강식품' 카테고리의 상품에 가장 큰 영향을 주는 주요 변수 조합은 '리뷰 수 증감', '상품 점수', '평균 리뷰 수', '상품 점수 증감', '카테고리 순위', '평균 배송일'의 조합일 수 있으며, 분석하기 가장 적합한 예측 알고리즘은 Catboost에 해당할 수 있다.

[0093] 주요 변수의 조합과 예측 알고리즘이 최종 분석 대상으로서 선택되어 분석을 거친 결과는, "8개의 주요 변수 중 6개를 선택한 조합들 중 '리뷰 수 증감', '상품 점수', '평균 리뷰 수', '상품 점수 증감', '카테고리 순위', '평균 배송일'의 조합은 Catboost 알고리즘을 사용했을 때 '건강식품' 카테고리의 현재의 사업자 데이터를 85%의 정확도로 재현할 수 있다"와 같은 형태를 가질 수 있다.

[0094] 따라서, 이와 같은 방법으로 주요 변수 조합과 예측 알고리즘 쌍을 활용하여 현재의 사업자 활동 데이터로부터 미래의 사업자 활동을 예측하고, 사업자의 영업 활동을 보조하는 한편 금융기관으로 하여금 사업자의 장래 가능성을 가늠할 수 있는 지표로서 기능할 수 있다.

[0095] 한편, 본 개시의 일 실시예에 따라 데이터 구조를 저장한 컴퓨터 판독가능 매체가 개시된다.

[0096] 데이터 구조는 데이터에 효율적인 접근 및 수정을 가능하게 하는 데이터의 조직, 관리, 저장을 의미할 수 있다. 데이터 구조는 특정 문제(예를 들어, 최단 시간으로 데이터 검색, 데이터 저장, 데이터 수정) 해결을 위한 데이터의 조직을 의미할 수 있다. 데이터 구조는 특정한 데이터 처리 기능을 지원하도록 설계된, 데이터 요소들 간의 물리적이거나 논리적인 관계로 정의될 수도 있다. 데이터 요소들 간의 논리적인 관계는 사용자 정의 데이터 요소들 간의 연결관계를 포함할 수 있다. 데이터 요소들 간의 물리적인 관계는 컴퓨터 판독가능 저장매체(예를 들어, 영구 저장 장치)에 물리적으로 저장되어 있는 데이터 요소들 간의 실제 관계를 포함할 수 있다. 데이터 구조는 구체적으로 데이터의 집합, 데이터 간의 관계, 데이터에 적용할 수 있는 함수 또는 명령어를 포함할 수 있다. 효과적으로 설계된 데이터 구조를 통해 컴퓨팅 장치는 컴퓨팅 장치의 자원을 최소한으로 사용하면서 연산을 수행할 수 있다. 구체적으로 컴퓨팅 장치는 효과적으로 설계된 데이터 구조를 통해 연산, 읽기, 삽입, 삭제, 비교, 교환, 검색의 효율성을 높일 수 있다.

[0097] 데이터 구조는 데이터 구조의 형태에 따라 선형 데이터 구조와 비선형 데이터 구조로 구분될 수 있다. 선형 데이터 구조는 하나의 데이터 뒤에 하나의 데이터만이 연결되는 구조일 수 있다. 선형 데이터 구조는 리스트(List), 스택(Stack), 큐(Queue), 데크(Deque)를 포함할 수 있다. 리스트는 내부적으로 순서가 존재하는 일련의 데이터 집합을 의미할 수 있다. 리스트는 연결 리스트(Linked List)를 포함할 수 있다. 연결 리스트는 각각의 데이터가 포인터를 가지고 한 줄로 연결되어 있는 방식으로 데이터가 연결된 데이터 구조일 수 있다. 연결 리스트에서 포인터는 다음이나 이전 데이터와의 연결 정보를 포함할 수 있다. 연결 리스트는 형태에 따라 단일 연결 리스트, 이중 연결 리스트, 원형 연결 리스트로 표현될 수 있다. 스택은 제한적으로 데이터에 접근할 수 있는 데이터 나열 구조일 수 있다. 스택은 데이터 구조의 한 쪽 끝에서만 데이터를 처리(예를 들어, 삽입 또는 삭제)할 수 있는 선형 데이터 구조일 수 있다. 스택에 저장된 데이터는 늦게 들어갈수록 빨리 나오는 데이터 구조(LIFO-Last in First Out)일 수 있다. 큐는 제한적으로 데이터에 접근할 수 있는 데이터 나열 구조로서, 스택과 달리 늦게 저장된 데이터일수록 늦게 나오는 데이터 구조(FIFO-First in First Out)일 수 있다. 데크는 데이터 구조의 양 쪽 끝에서 데이터를 처리할 수 있는 데이터 구조일 수 있다.

[0098] 비선형 데이터 구조는 하나의 데이터 뒤에 복수개의 데이터가 연결되는 구조일 수 있다. 비선형 데이터 구조는

그래프(Graph) 데이터 구조를 포함할 수 있다. 그래프 데이터 구조는 정점(Vertex)과 간선(Edge)으로 정의될 수 있으며 간선은 서로 다른 두개의 정점을 연결하는 선을 포함할 수 있다. 그래프 데이터 구조 트리(Tree) 데이터 구조를 포함할 수 있다. 트리 데이터 구조는 트리에 포함된 복수개의 정점 중에서 서로 다른 두개의 정점을 연결시키는 경로가 하나인 데이터 구조일 수 있다. 즉 그래프 데이터 구조에서 루프(loop)를 형성하지 않는 데이터 구조일 수 있다.

[0099] 본 명세서에 걸쳐, 연산 모델, 신경망, 네트워크 함수, 뉴럴 네트워크(neural network)는 동일한 의미로 사용될 수 있다. 이하에서는 신경망으로 통일하여 기술한다. 데이터 구조는 신경망을 포함할 수 있다. 그리고 신경망을 포함한 데이터 구조는 컴퓨터 판독가능 매체에 저장될 수 있다. 신경망을 포함한 데이터 구조는 또한 신경망에 의한 처리를 위하여 전처리된 데이터, 신경망에 입력되는 데이터, 신경망의 가중치, 신경망의 하이퍼 파라미터, 신경망으로부터 획득한 데이터, 신경망의 각 노드 또는 레이어와 연관된 활성화 함수, 신경망의 학습을 위한 손실 함수 등을 포함할 수 있다. 신경망을 포함한 데이터 구조는 상기 개시된 구성들 중 임의의 구성 요소들을 포함할 수 있다. 즉 신경망을 포함한 데이터 구조는 신경망에 의한 처리를 위하여 전처리된 데이터, 신경망에 입력되는 데이터, 신경망의 가중치, 신경망의 하이퍼 파라미터, 신경망으로부터 획득한 데이터, 신경망의 각 노드 또는 레이어와 연관된 활성화 함수, 신경망의 학습을 위한 손실 함수 등 전부 또는 이들의 임의의 조합을 포함하여 구성될 수 있다. 전술한 구성들 이외에도, 신경망을 포함한 데이터 구조는 신경망의 특성을 결정하는 임의의 다른 정보를 포함할 수 있다. 또한, 데이터 구조는 신경망의 연산 과정에 사용되거나 발생하는 모든 형태의 데이터를 포함할 수 있으며 전술한 사항에 제한되는 것은 아니다. 컴퓨터 판독가능 매체는 컴퓨터 판독가능 기록 매체 및/또는 컴퓨터 판독가능 전송 매체를 포함할 수 있다. 신경망은 일반적으로 노드라 지칭될 수 있는 상호 연결된 계산 단위들의 집합으로 구성될 수 있다. 이러한 노드들은 뉴런(neuron)들로 지칭될 수도 있다. 신경망은 적어도 하나 이상의 노드들을 포함하여 구성된다.

[0100] 데이터 구조는 신경망에 입력되는 데이터를 포함할 수 있다. 신경망에 입력되는 데이터를 포함하는 데이터 구조는 컴퓨터 판독가능 매체에 저장될 수 있다. 신경망에 입력되는 데이터는 신경망 학습 과정에서 입력되는 학습 데이터 및/또는 학습이 완료된 신경망에 입력되는 입력 데이터를 포함할 수 있다. 신경망에 입력되는 데이터는 전처리(pre-processing)를 거친 데이터 및/또는 전처리 대상이 되는 데이터를 포함할 수 있다. 전처리는 데이터를 신경망에 입력시키기 위한 데이터 처리 과정을 포함할 수 있다. 따라서 데이터 구조는 전처리 대상이 되는 데이터 및 전처리로 발생하는 데이터를 포함할 수 있다. 전술한 데이터 구조는 예시일 뿐 본 개시는 이에 제한되지 않는다.

[0101] 데이터 구조는 신경망의 가중치를 포함할 수 있다. (본 명세서에서 가중치, 파라미터는 동일한 의미로 사용될 수 있다.) 그리고 신경망의 가중치를 포함한 데이터 구조는 컴퓨터 판독가능 매체에 저장될 수 있다. 신경망은 복수개의 가중치를 포함할 수 있다. 가중치는 가변적일 수 있으며, 신경망이 원하는 기능을 수행하기 위해, 사용자 또는 알고리즘에 의해 가변될 수 있다. 예를 들어, 하나의 출력 노드에 하나 이상의 입력 노드가 각각의 링크에 의해 상호 연결된 경우, 출력 노드는 상기 출력 노드와 연결된 입력 노드들에 입력된 값들 및 각각의 입력 노드들에 대응하는 링크에 설정된 가중치에 기초하여 출력 노드에서 출력되는 데이터 값을 결정할 수 있다. 전술한 데이터 구조는 예시일 뿐 본 개시는 이에 제한되지 않는다.

[0102] 제한이 아닌 예로서, 가중치는 신경망 학습 과정에서 가변되는 가중치 및/또는 신경망 학습이 완료된 가중치를 포함할 수 있다. 신경망 학습 과정에서 가변되는 가중치는 학습 사이클이 시작되는 시점의 가중치 및/또는 학습 사이클 동안 가변되는 가중치를 포함할 수 있다. 신경망 학습이 완료된 가중치는 학습 사이클이 완료된 가중치를 포함할 수 있다. 따라서 신경망의 가중치를 포함한 데이터 구조는 신경망 학습 과정에서 가변되는 가중치 및/또는 신경망 학습이 완료된 가중치를 포함한 데이터 구조를 포함할 수 있다. 그러므로 상술한 가중치 및/또는 각 가중치의 조합은 신경망의 가중치를 포함한 데이터 구조에 포함되는 것으로 한다. 전술한 데이터 구조는 예시일 뿐 본 개시는 이에 제한되지 않는다.

[0103] 신경망의 가중치를 포함한 데이터 구조는 직렬화(serialization) 과정을 거친 후 컴퓨터 판독가능 저장 매체(예를 들어, 메모리, 하드 디스크)에 저장될 수 있다. 직렬화는 데이터 구조를 동일하거나 다른 컴퓨팅 장치에 저장하고 나중에 다시 재구성하여 사용할 수 있는 형태로 변환하는 과정일 수 있다. 컴퓨팅 장치는 데이터 구조를 직렬화하여 네트워크를 통해 데이터를 송수신할 수 있다. 직렬화된 신경망의 가중치를 포함한 데이터 구조는 역직렬화(deserialization)를 통해 동일한 컴퓨팅 장치 또는 다른 컴퓨팅 장치에서 재구성될 수 있다. 신경망의 가중치를 포함한 데이터 구조는 직렬화에 한정되는 것은 아니다. 나아가 신경망의 가중치를 포함한 데이터 구조는 컴퓨팅 장치의 자원을 최소한으로 사용하면서 연산의 효율을 높이기 위한 데이터 구조(예를 들어, 비선형 데이터 구조에서 B-Tree, Trie, m-way search tree, AVL tree, Red-Black Tree)를 포함할 수 있다. 전술한 사항

은 예시일 뿐 본 개시는 이에 제한되지 않는다.

- [0104] 데이터 구조는 신경망의 하이퍼 파라미터(Hyper-parameter)를 포함할 수 있다. 그리고 신경망의 하이퍼 파라미터를 포함한 데이터 구조는 컴퓨터 판독가능 매체에 저장될 수 있다. 하이퍼 파라미터는 사용자에 의해 가변되는 변수일 수 있다. 하이퍼 파라미터는 예를 들어, 학습률(learning rate), 비용 함수(cost function), 학습 사이클 반복 횟수, 가중치 초기화(Weight initialization)(예를 들어, 가중치 초기화 대상이 되는 가중치 값의 범위 설정), Hidden Unit 개수(예를 들어, 히든 레이어의 개수, 히든 레이어의 노드 수)를 포함할 수 있다. 전술한 데이터 구조는 예시일 뿐 본 개시는 이에 제한되지 않는다.
- [0105] 도 5는 본 개시의 실시예들이 구현될 수 있는 예시적인 컴퓨팅 환경에 대한 간략하고 일반적인 개략도이다.
- [0106] 본 개시가 일반적으로 컴퓨팅 장치에 의해 구현될 수 있는 것으로 전술되었지만, 당업자라면 본 개시가 하나 이상의 컴퓨터 상에서 실행될 수 있는 컴퓨터 실행가능 명령어 및/또는 기타 프로그램 모듈들과 결합되어 및/또는 하드웨어와 소프트웨어의 조합으로써 구현될 수 있다는 것을 잘 알 것이다.
- [0107] 일반적으로, 프로그램 모듈은 특정의 태스크를 수행하거나 특정의 추상 데이터 유형을 구현하는 루틴, 프로그램, 컴포넌트, 데이터 구조, 기타 등등을 포함한다. 또한, 당업자라면 본 개시의 방법이 단일-프로세서 또는 멀티프로세서 컴퓨터 시스템, 미니컴퓨터, 메인프레임 컴퓨터는 물론 퍼스널 컴퓨터, 핸드헬드(handheld) 컴퓨팅 장치, 마이크로프로세서-기반 또는 프로그램가능 가전 제품, 기타 등등(이들 각각은 하나 이상의 연관된 장치와 연결되어 동작할 수 있음)을 비롯한 다른 컴퓨터 시스템 구성으로 실시될 수 있다는 것을 잘 알 것이다.
- [0108] 본 개시의 설명된 실시예들은 또한 어떤 태스크들이 통신 네트워크를 통해 연결되어 있는 원격 처리 장치들에 의해 수행되는 분산 컴퓨팅 환경에서 실시될 수 있다. 분산 컴퓨팅 환경에서, 프로그램 모듈은 로컬 및 원격 메모리 저장 장치 둘 다에 위치할 수 있다.
- [0109] 컴퓨터는 통상적으로 다양한 컴퓨터 판독가능 매체를 포함한다. 컴퓨터에 의해 액세스 가능한 매체는 그 어떤 것이든지 컴퓨터 판독가능 매체가 될 수 있고, 이러한 컴퓨터 판독가능 매체는 휘발성 및 비휘발성 매체, 일시적(transitory) 및 비일시적(non-transitory) 매체, 이동식 및 비-이동식 매체를 포함한다. 제한이 아닌 예로서, 컴퓨터 판독가능 매체는 컴퓨터 판독가능 저장 매체 및 컴퓨터 판독가능 전송 매체를 포함할 수 있다. 컴퓨터 판독가능 저장 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보를 저장하는 임의의 방법 또는 기술로 구현되는 휘발성 및 비휘발성 매체, 일시적 및 비-일시적 매체, 이동식 및 비이동식 매체를 포함한다. 컴퓨터 판독가능 저장 매체는 RAM, ROM, EEPROM, 플래시 메모리 또는 기타 메모리 기술, CD-ROM, DVD(digital video disk) 또는 기타 광 디스크 저장 장치, 자기 카세트, 자기 테이프, 자기 디스크 저장 장치 또는 기타 자기 저장 장치, 또는 컴퓨터에 의해 액세스될 수 있고 원하는 정보를 저장하는 데 사용될 수 있는 임의의 기타 매체를 포함하지만, 이에 한정되지 않는다.
- [0110] 컴퓨터 판독가능 전송 매체는 통상적으로 반송파(carrier wave) 또는 기타 전송 메커니즘(transport mechanism)과 같은 피변조 데이터 신호(modulated data signal)에 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터 등을 구현하고 모든 정보 전달 매체를 포함한다. 피변조 데이터 신호라는 용어는 신호 내에 정보를 인코딩하도록 그 신호의 특성들 중 하나 이상을 설정 또는 변경시킨 신호를 의미한다. 제한이 아닌 예로서, 컴퓨터 판독가능 전송 매체는 유선 네트워크 또는 직접 배선 접속(direct-wired connection)과 같은 유선 매체, 그리고 음향, RF, 적외선, 기타 무선 매체와 같은 무선 매체를 포함한다. 상술된 매체들 중 임의의 것의 조합도 역시 컴퓨터 판독가능 전송 매체의 범위 안에 포함되는 것으로 한다.
- [0111] 컴퓨터(1102)를 포함하는 본 개시의 여러가지 측면들을 구현하는 예시적인 환경(1100)이 나타내어져 있으며, 컴퓨터(1102)는 처리 장치(1104), 시스템 메모리(1106) 및 시스템 버스(1108)를 포함한다. 시스템 버스(1108)는 시스템 메모리(1106)(이에 한정되지 않음)를 비롯한 시스템 컴포넌트들을 처리 장치(1104)에 연결시킨다. 처리 장치(1104)는 다양한 상용 프로세서들 중 임의의 프로세서일 수 있다. 듀얼 프로세서 및 기타 멀티프로세서 아키텍처도 역시 처리 장치(1104)로서 이용될 수 있다.
- [0112] 시스템 버스(1108)는 메모리 버스, 주변장치 버스, 및 다양한 상용 버스 아키텍처 중 임의의 것을 사용하는 로컬 버스에 추가적으로 상호 연결될 수 있는 몇 가지 유형의 버스 구조 중 임의의 것일 수 있다. 시스템 메모리(1106)는 판독 전용 메모리(ROM)(1110) 및 랜덤 액세스 메모리(RAM)(1112)를 포함한다. 기본 입/출력 시스템(BIOS)은 ROM, EPROM, EEPROM 등의 비휘발성 메모리(1110)에 저장되며, 이 BIOS는 시동 중과 같은 때에 컴퓨터(1102) 내의 구성요소들 간에 정보를 전송하는 일을 돕는 기본적인 루틴을 포함한다. RAM(1112)은 또한 데이터를 캐싱하기 위한 정적 RAM 등의 고속 RAM을 포함할 수 있다.

- [0113] 컴퓨터(1102)는 또한 내장형 하드 디스크 드라이브(HDD)(1114)(예를 들어, EIDE, SATA)-이 내장형 하드 디스크 드라이브(1114)는 또한 적당한 새시(도시 생략) 내에서 외장형 용도로 구성될 수 있음-, 자기 플로피 디스크 드라이브(FDD)(1116)(예를 들어, 이동식 디스켓(1118)으로부터 판독을 하거나 그에 기록을 하기 위한 것임), 및 광 디스크 드라이브(1120)(예를 들어, CD-ROM 디스크(1122)를 판독하거나 DVD 등의 기타 고용량 광 매체로부터 판독을 하거나 그에 기록을 하기 위한 것임)를 포함한다. 하드 디스크 드라이브(1114), 자기 디스크 드라이브(1116) 및 광 디스크 드라이브(1120)는 각각 하드 디스크 드라이브 인터페이스(1124), 자기 디스크 드라이브 인터페이스(1126) 및 광 드라이브 인터페이스(1128)에 의해 시스템 버스(1108)에 연결될 수 있다. 외장형 드라이브 구현을 위한 인터페이스(1124)는 USB(Universal Serial Bus) 및 IEEE 1394 인터페이스 기술 중 적어도 하나 또는 그 둘 다를 포함한다.
- [0114] 이들 드라이브 및 그와 연관된 컴퓨터 판독가능 매체는 데이터, 데이터 구조, 컴퓨터 실행가능 명령어, 기타 등등의 비휘발성 저장을 제공한다. 컴퓨터(1102)의 경우, 드라이브 및 매체는 임의의 데이터를 적당한 디지털 형식으로 저장하는 것에 대응한다. 상기에서의 컴퓨터 판독가능 매체에 대한 설명이 HDD, 이동식 자기 디스크, 및 CD 또는 DVD 등의 이동식 광 매체를 언급하고 있지만, 당업자라면 zip 드라이브(zip drive), 자기 카세트, 플래쉬 메모리 카드, 카트리지, 기타 등등의 컴퓨터에 의해 판독가능한 다른 유형의 매체도 역시 예시적인 운영 환경에서 사용될 수 있으며 또 임의의 이러한 매체가 본 개시의 방법들을 수행하기 위한 컴퓨터 실행가능 명령어를 포함할 수 있다는 것을 잘 알 것이다.
- [0115] 운영 체제(1130), 하나 이상의 애플리케이션 프로그램(1132), 기타 프로그램 모듈(1134) 및 프로그램 데이터(1136)를 비롯한 다수의 프로그램 모듈이 드라이브 및 RAM(1112)에 저장될 수 있다. 운영 체제, 애플리케이션, 모듈 및/또는 데이터의 전부 또는 그 일부분이 또한 RAM(1112)에 캐싱될 수 있다. 본 개시가 여러가지 상업적으로 이용가능한 운영 체제 또는 운영 체제들의 조합에서 구현될 수 있다는 것을 잘 알 것이다.
- [0116] 사용자는 하나 이상의 유선/무선 입력 장치, 예를 들어, 키보드(1138) 및 마우스(1140) 등의 포인팅 장치를 통해 컴퓨터(1102)에 명령 및 정보를 입력할 수 있다. 기타 입력 장치(도시 생략)로는 마이크, IR 리모콘, 조이스틱, 게임 패드, 스타일러스 펜, 터치 스크린, 기타 등등이 있을 수 있다. 이들 및 기타 입력 장치가 종종 시스템 버스(1108)에 연결되어 있는 입력 장치 인터페이스(1142)를 통해 처리 장치(1104)에 연결되지만, 병렬 포트, IEEE 1394 직렬 포트, 게임 포트, USB 포트, IR 인터페이스, 기타 등등의 기타 인터페이스에 의해 연결될 수 있다.
- [0117] 모니터(1144) 또는 다른 유형의 디스플레이 장치도 역시 비디오 어댑터(1146) 등의 인터페이스를 통해 시스템 버스(1108)에 연결된다. 모니터(1144)에 부가하여, 컴퓨터는 일반적으로 스피커, 프린터, 기타 등등의 기타 주변 출력 장치(도시 생략)를 포함한다.
- [0118] 컴퓨터(1102)는 유선 및/또는 무선 통신을 통한 원격 컴퓨터(들)(1148) 등의 하나 이상의 원격 컴퓨터로의 논리적 연결을 사용하여 네트워크화된 환경에서 동작할 수 있다. 원격 컴퓨터(들)(1148)는 워크스테이션, 컴퓨팅 디바이스 컴퓨터, 라우터, 퍼스널 컴퓨터, 휴대용 컴퓨터, 마이크로프로세서-기반 오락 기기, 피어 장치 또는 기타 통상의 네트워크 노드일 수 있으며, 일반적으로 컴퓨터(1102)에 대해 기술된 구성요소들 중 다수 또는 그 전부를 포함하지만, 간략함을 위해, 메모리 저장 장치(1150)만이 도시되어 있다. 도시되어 있는 논리적 연결은 근거리 통신망(LAN)(1152) 및/또는 더 큰 네트워크, 예를 들어, 원거리 통신망(WAN)(1154)에의 유선/무선 연결을 포함한다. 이러한 LAN 및 WAN 네트워킹 환경은 사무실 및 회사에서 일반적인 것이며, 인트라넷 등의 전사적 컴퓨터 네트워크(enterprise-wide computer network)를 용이하게 해주며, 이들 모두는 전세계 컴퓨터 네트워크, 예를 들어, 인터넷에 연결될 수 있다.
- [0119] LAN 네트워킹 환경에서 사용될 때, 컴퓨터(1102)는 유선 및/또는 무선 통신 네트워크 인터페이스 또는 어댑터(1156)를 통해 로컬 네트워크(1152)에 연결된다. 어댑터(1156)는 LAN(1152)에의 유선 또는 무선 통신을 용이하게 해줄 수 있으며, 이 LAN(1152)은 또한 무선 어댑터(1156)와 통신하기 위해 그에 설치되어 있는 무선 액세스 포인트를 포함하고 있다. WAN 네트워킹 환경에서 사용될 때, 컴퓨터(1102)는 모뎀(1158)을 포함할 수 있거나, WAN(1154) 상의 통신 컴퓨팅 디바이스에 연결되거나, 또는 인터넷을 통하는 등, WAN(1154)을 통해 통신을 설정하는 기타 수단을 갖는다. 내장형 또는 외장형 및 유선 또는 무선 장치일 수 있는 모뎀(1158)은 직렬 포트 인터페이스(1142)를 통해 시스템 버스(1108)에 연결된다. 네트워킹된 환경에서, 컴퓨터(1102)에 대해 설명된 프로그램 모듈들 또는 그의 일부분이 원격 메모리/저장 장치(1150)에 저장될 수 있다. 도시된 네트워크 연결이 예시적인 것이며 컴퓨터들 사이에 통신 링크를 설정하는 기타 수단이 사용될 수 있다는 것을 잘 알 것이다.
- [0120] 컴퓨터(1102)는 무선 통신으로 배치되어 동작하는 임의의 무선 장치 또는 개체, 예를 들어, 프린터, 스캐너, 데

스크린 및/또는 휴대용 컴퓨터, PDA(portable data assistant), 통신 위성, 무선 검출가능 태그와 연관된 임의의 장비 또는 장소, 및 전화와 통신을 하는 동작을 한다. 이것은 적어도 Wi-Fi 및 블루투스 무선 기술을 포함한다. 따라서, 통신은 종래의 네트워크에서와 같이 미리 정의된 구조이거나 단순히 적어도 2개의 장치 사이의 애드혹 통신(ad hoc communication)일 수 있다.

[0121] Wi-Fi(Wireless Fidelity)는 유선 없이도 인터넷 등으로의 연결을 가능하게 해준다. Wi-Fi는 이러한 장치, 예를 들어, 컴퓨터가 실내에서 및 실외에서, 즉 기지국의 통화권 내의 아무 곳에서나 데이터를 전송 및 수신할 수 있게 해주는 셀 전화와 같은 무선 기술이다. Wi-Fi 네트워크는 안전하고 신뢰성 있으며 고속인 무선 연결을 제공하기 위해 IEEE 802.11(a, b, g, 기타)이라고 하는 무선 기술을 사용한다. 컴퓨터를 서로에, 인터넷에 및 유선 네트워크(IEEE 802.3 또는 이더넷을 사용함)에 연결시키기 위해 Wi-Fi가 사용될 수 있다. Wi-Fi 네트워크는 비인가 2.4 및 5GHz 무선 대역에서, 예를 들어, 11Mbps(802.11a) 또는 54 Mbps(802.11b) 데이터 레이트로 동작하거나, 양 대역(듀얼 대역)을 포함하는 제품에서 동작할 수 있다.

[0122] 본 개시의 기술 분야에서 통상의 지식을 가진 자는 정보 및 신호들이 임의의 다양한 상이한 기술들 및 기법들을 이용하여 표현될 수 있다는 것을 이해할 것이다. 예를 들어, 위의 설명에서 참조될 수 있는 데이터, 지시들, 명령들, 정보, 신호들, 비트들, 심볼들 및 칩들은 전압들, 전류들, 전자기파들, 자기장들 또는 입자들, 광학장들 또는 입자들, 또는 이들의 임의의 결합에 의해 표현될 수 있다.

[0123] 본 개시의 기술 분야에서 통상의 지식을 가진 자는 여기에 개시된 실시예들과 관련하여 설명된 다양한 예시적인 논리 블록들, 모듈들, 프로세서들, 수단들, 회로들 및 알고리즘 단계들이 전자 하드웨어, (편의를 위해, 여기에서 소프트웨어로 지칭되는) 다양한 형태들의 프로그램 또는 설계 코드 또는 이들 모두의 결합에 의해 구현될 수 있다는 것을 이해할 것이다. 하드웨어 및 소프트웨어의 이러한 상호 호환성을 명확하게 설명하기 위해, 다양한 예시적인 컴포넌트들, 블록들, 모듈들, 회로들 및 단계들이 이들의 기능과 관련하여 위에서 일반적으로 설명되었다. 이러한 기능이 하드웨어 또는 소프트웨어로서 구현되는지 여부는 특정한 애플리케이션 및 전체 시스템에 대하여 부과되는 설계 제약들에 따라 좌우된다. 본 개시의 기술 분야에서 통상의 지식을 가진 자는 각각의 특정한 애플리케이션에 대하여 다양한 방식으로 설명된 기능을 구현할 수 있으나, 이러한 구현 결정들은 본 개시의 범위를 벗어나는 것으로 해석되어서는 안 될 것이다.

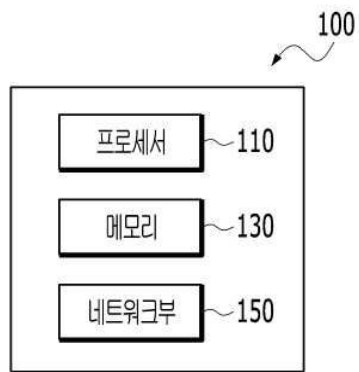
[0124] 여기서 제시된 다양한 실시예들은 방법, 장치, 또는 표준 프로그래밍 및/또는 엔지니어링 기술을 사용한 제조물품(article)으로 구현될 수 있다. 용어 제조물품은 임의의 컴퓨터-판독가능 저장장치로부터 액세스 가능한 컴퓨터 프로그램, 캐리어, 또는 매체(media)를 포함한다. 예를 들어, 컴퓨터-판독가능 저장매체는 자기 저장 장치(예를 들면, 하드 디스크, 플로피 디스크, 자기 스트림, 등), 광학 디스크(예를 들면, CD, DVD, 등), 스마트 카드, 및 플래쉬 메모리 장치(예를 들면, EEPROM, 카드, 스틱, 키 드라이브, 등)를 포함하지만, 이들로 제한되는 것은 아니다. 또한, 여기서 제시되는 다양한 저장 매체는 정보를 저장하기 위한 하나 이상의 장치 및/또는 다른 기계-판독가능한 매체를 포함한다.

[0125] 제시된 프로세스들에 있는 단계들의 특정한 순서 또는 계층 구조는 예시적인 접근들의 일례임을 이해하도록 한다. 설계 우선순위들에 기반하여, 본 개시의 범위 내에서 프로세스들에 있는 단계들의 특정한 순서 또는 계층 구조가 재배열될 수 있다는 것을 이해하도록 한다. 첨부된 방법 청구항들은 샘플 순서로 다양한 단계들의 엘리먼트들을 제공하지만 제시된 특정한 순서 또는 계층 구조에 한정되는 것을 의미하지는 않는다.

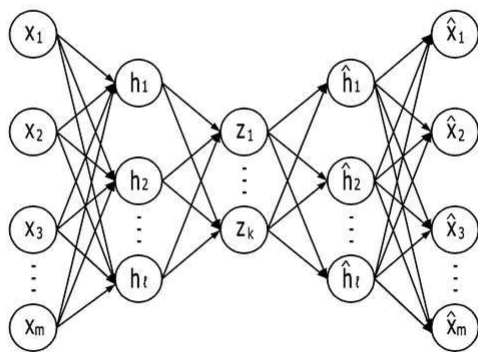
[0126] 제시된 실시예들에 대한 설명은 임의의 본 개시의 기술 분야에서 통상의 지식을 가진 자가 본 개시를 이용하거나 또는 실시할 수 있도록 제공된다. 이러한 실시예들에 대한 다양한 변형들은 본 개시의 기술 분야에서 통상의 지식을 가진 자에게 명백할 것이며, 여기에 정의된 일반적인 원리들은 본 개시의 범위를 벗어남이 없이 다른 실시예들에 적용될 수 있다. 그리하여, 본 개시는 여기에 제시된 실시예들로 한정되는 것이 아니라, 여기에 제시된 원리들 및 신규한 특징들과 일관되는 최광의의 범위에서 해석되어야 할 것이다.

도면

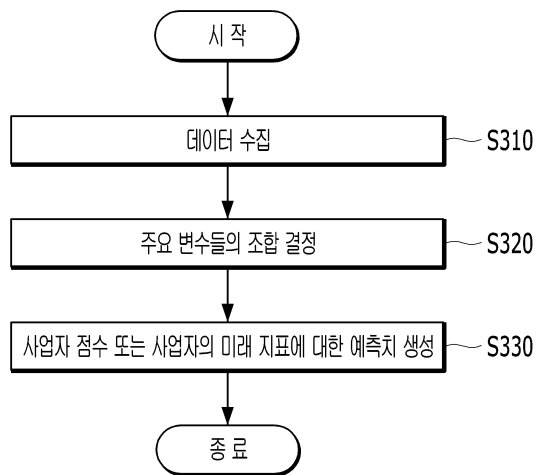
도면1



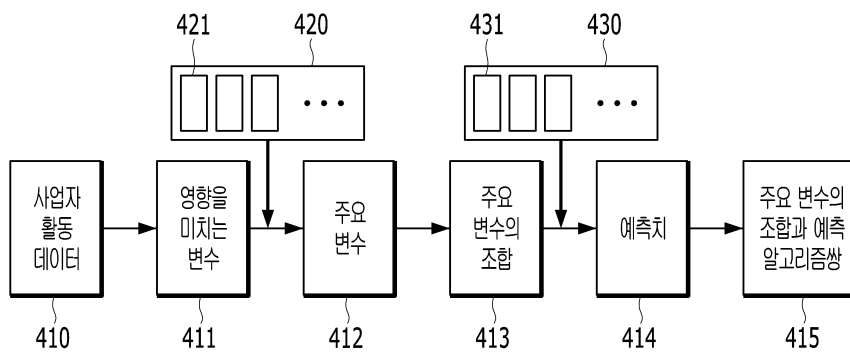
도면2



도면3



도면4



도면5

