



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2023년07월21일
(11) 등록번호 10-2557695
(24) 등록일자 2023년07월17일

(51) 국제특허분류(Int. Cl.)
G06Q 30/08 (2012.01) G06N 3/08 (2023.01)
G06Q 50/16 (2012.01)
(52) CPC특허분류
G06Q 30/08 (2023.01)
G06N 3/08 (2023.01)
(21) 출원번호 10-2021-0082994
(22) 출원일자 2021년06월25일
심사청구일자 2021년06월25일
(65) 공개번호 10-2023-0001047
(43) 공개일자 2023년01월04일
(56) 선행기술조사문헌
KR102008992 B1*
KR1020210073086 A*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
주식회사 지지옥션
서울특별시 용산구 청파로49길 3 (청파동2가)
연세대학교 산학협력단
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
(72) 발명자
강은
서울특별시 강남구 삼성로57길 45 대치삼성2차아파트 201동 502호
오경주
서울특별시 마포구 상암산로1길 92, 706-1204호
(74) 대리인
김인철

전체 청구항 수 : 총 14 항

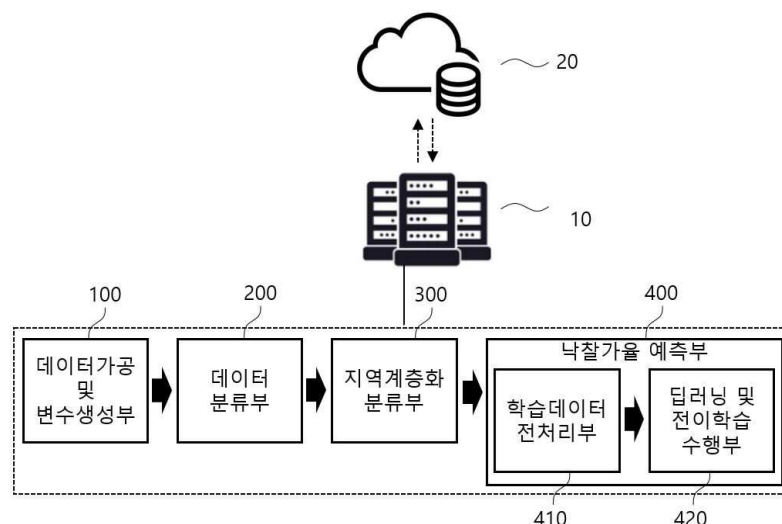
심사관 : 박성웅

(54) 발명의 명칭 계층적 분석 및 전이학습을 이용한 부동산 경매 낙찰가율 예측시스템 및 예측방법

(57) 요약

본 발명은 제어서버(10) 및 부동산 경매물건 데이터베이스(20)가 네트워크로 연결되고, 컴퓨터를 포함하는 연산 처리수단에 의해 경매물건 데이터를 딥러닝하여 낙찰가율을 예측하도록 수행되는 부동산 경매 낙찰가율 예측시스템으로서, 상기 제어서버(10)는 경매물건 데이터를 가공하고, 변수를 생성하는 데이터가공 및 변수생성부(100); 가공된 경매물건 데이터를 지역별 및 용도별로 분류하는 데이터 분류부(200); 분류된 경매물건 데이터를 지역의 광협을 기준으로 복수개의 지역계층으로 분류하는 지역계층화 분류부(300); 및 지역계층화된 경매물건 데이터를 딥러닝 및 전이학습하여 학습모델을 생성하고 경매물건의 낙찰가율을 예측하는 낙찰가율 예측부(400)를 포함하는 것을 특징으로 하는 계층적 분석 및 전이학습을 이용한 부동산 경매 낙찰가율 예측시스템이다.

대표도 - 도1



(52) CPC특허분류

G06Q 50/16 (2013.01)

(72) 발명자

강경모

경기도 화성시 동탄숲속로 19, 893동 1904호 (능동, 동탄숲속마을 능동역 리제 더 포레스트)

박해일

서울특별시 도봉구 도봉로 551-10, B01호 (쌍문동, 금강맨션)

정승환

경기도 용인시 기흥구 사은로126번길 46

류호선

서울특별시 영등포구 버드나루로 102-1, 1205호

김상희

서울특별시 서대문구 창천동 99-26 미르빌 505호

김정훈

인천광역시 중구 송월로 1-6 탐빌라 201호

박태현

서울특별시 강서구 곰달래로 113, 501호

이상규

서울시 서초구 사임당로 137, 신동아아파트 5동 908호

이상현

충청북도 청주시 흥덕구 대농로 45, 302-2706

이정택

경기도 고양시 일산동구 백마로 223, 1113호

주성준

경기도 고양시 일산동구 노루목로 80, 313동 1301호 (장항동, 호수마을3단지아파트)

최연지

서울특별시 서대문구 성산로 398, 311호

이 발명을 지원한 국가연구개발사업

과제고유번호 1711120415

과제번호 2020-0-02162-001

부처명 과학기술정보통신부

과제관리(전문)기관명 정보통신기획평가원

연구사업명 비대면비즈니스디지털혁신기술개발(R&D)

연구과제명 비대면 경매 정보 제공 서비스를 위한 딥러닝 기반 부동산 경매 적정가 예측시스템

개발

기 여 율 1/1

과제수행기관명 (주)지지옥션

연구기간 2020.08.01 ~ 2021.07.31

명세서

청구범위

청구항 1

제어 서버 및 부동산 경매물건 데이터베이스가 네트워크로 연결되고, 컴퓨터를 포함하는 연산처리수단에 의해 경매물건 데이터를 딥러닝하여 낙찰가율을 예측하도록 수행되는 부동산 경매 낙찰가율 예측시스템으로서,

상기 제어 서버는 경매물건 데이터를 가공하고, 변수를 생성하는 데이터가공 및 변수생성부; 가공된 경매물건 데이터를 지역별 및 용도별로 분류하는 데이터 분류부; 분류된 경매물건 데이터를 지역의 광협을 기준으로 복수개의 지역계층으로 분류하는 지역계층화 분류부; 및 지역계층화된 경매물건 데이터를 딥러닝 및 전이학습하여 학습모델을 생성하고 경매물건의 낙찰가율을 예측하는 낙찰가율 예측부를 포함하며,

상기 낙찰가율 예측부는 학습모델의 학습을 위해 데이터 전처리를 수행하는 학습데이터 전처리부 및 전처리된 데이터셋을 이용하여 딥러닝 및 전이학습을 수행하는 딥러닝 및 전이학습 수행부를 포함하며,

상기 딥러닝 및 전이학습 수행부는 상기 지역계층화 분류부의 지역계층 $L(t:t$ 는 기 설정된 양의 정수)에 대응되는 데이터셋 $S(t)$ 을 구비하며, 최초에는 데이터셋 $S(t)$ 에 초기 가중치 $W(t)$ 를 무작위로 부여하여 학습시켜 학습모델 $M(t)$ 를 생성하고, 데이터셋 $S(t)$ 에서 데이터셋 $S(t+1)$ 을 제외시킨 잔여 데이터셋으로 직전 학습모델 $M(t)$ 를 학습시켜 새로운 학습모델 $M(t+1)$ 을 생성시키고, 다음으로, 직전 학습모델 $M(t+1)$ 의 조정 가중치 $W(t+1)$ 를 통해 새로운 학습모델 $M(t+2)$ 를 만들고, 데이터셋 $S(t+1)$ 에서 데이터셋 $S(t+2)$ 를 제외시킨 잔여 데이터셋으로 새로운 학습모델 $M(t+2)$ 의 전이학습을 추가로 수행하는 것을 반복하는 것을 특징으로 하는 계층적 분석 및 전이학습을 이용한 부동산 경매 낙찰가율 예측시스템.

청구항 2

청구항 1에 있어서,

상기 데이터가공 및 변수생성부에서,

독립변수는 낙찰가율이며,

종속변수는 지형, 법면, 도로접합, 토지거래허가, 개발제한구역 또는 지역지구변수갯수 중 적어도 하나를 포함하는 것을 특징으로 하는 계층적 분석 및 전이학습을 이용한 부동산 경매 낙찰가율 예측시스템.

청구항 3

청구항 1에 있어서,

상기 데이터가공 및 변수생성부에서,

로그화를 진행하는 종속변수는 감정가, 감정경과일자, 건물면적, 공고후조회수, 국민은행시세, 기계비중, 대표이외건물감정비율, 대표이외토지감정비율, 실거래가최근3개월, 이전매매가, 임대보증금비율, 제시외비중, 총세대수, 최근3개월낙찰가율, 최저가, 토지면적 또는 공시지가 중 적어도 어느 하나인 것을 특징으로 하는 계층적 분석 및 전이학습을 이용한 부동산 경매 낙찰가율 예측시스템.

청구항 4

청구항 1에 있어서,

상기 데이터 분류부에서 지역별 분류는 기 설정된 지역 단위로 분류되는 것을 특징으로 하는 계층적 분석 및 전이학습을 이용한 부동산 경매 낙찰가율 예측시스템.

청구항 5

청구항 1에 있어서,

상기 데이터 분류부에서 용도별 분류는 아파트용도, 연립용도, 다가구용도, 공장용도, 상가용도, 오피스텔용도,

전답및과수원용도, 임야용도, 대지용도 및 공원기타용도로 분류되는 것을 특징으로 하는 계층적 분석 및 전이학습을 이용한 부동산 경매 낙찰가율 예측시스템.

청구항 6

청구항 1에 있어서,

상기 지역계층화 분류부는

경매물건 데이터를 광의의 지역계층($t:t$ 는 양의 정수)에서 협의의 지역계층($t+1$)의 순서로 구분하며,

상기 협의의 지역은 상기 광의의 지역에 포함되는 지역이며,

구분되는 지역계층 개수는 기 설정되는 것을 특징으로 하는 계층적 분석 및 전이학습을 이용한 부동산 경매 낙찰가율 예측시스템.

청구항 7

청구항 6에 있어서,

상기 구분되는 지역계층 개수가 4개인 때에는,

지역계층1은 대한민국 전국의 지역, 지역계층2는 남북을 통합한 도 단위 지역, 지역계층3은 전국의 특별시, 광역시와 도 단위 지역, 지역계층4는 특별시, 광역시를 제외한 시 단위 지역 및 구, 군단위 지역으로 구분되는 것을 특징으로 하는 계층적 분석 및 전이학습을 이용한 부동산 경매 낙찰가율 예측시스템.

청구항 8

삭제

청구항 9

청구항 1에 있어서,

상기 학습데이터 전처리부는

데이터 전처리시에 다음의 수학적 식 1을 사용하여 학습데이터에 있는 이상치 데이터를 제거하는 것을 특징으로 하는 계층적 분석 및 전이학습을 이용한 부동산 경매 낙찰가율 예측시스템.

[수학적 식 1]

$$Range = [Q_1 - 1.5(Q_3 - Q_1), Q_3 + 1.5(Q_3 - Q_1)]$$

(여기서, Range는 Tukey fences를 시행했을 때 정상범위로 정의하는 범위를 나타내며, Q1은 전체 데이터를 낙찰가율의 크기에 따라 순서대로 정렬하였을 때 작은 쪽의 사분위 수를 나타내며, Q3는 전체 데이터를 낙찰가율의 크기에 따라 순서대로 정렬하였을 때 큰 쪽의 사분위 수를 나타낸다.)

청구항 10

청구항 1에 있어서,

상기 학습데이터 전처리부는

학습데이터의 일 변수에 값이 미기재된 비율이 학습데이터 개수와 대비하여 기 설정된 비율 이상이면 상기 변수를 학습에서 제외시키는 것을 특징으로 하는 계층적 분석 및 전이학습을 이용한 부동산 경매 낙찰가율 예측시스템.

청구항 11

삭제

청구항 12

청구항 1에 있어서,

상기 학습모델은 손실함수(loss function)를 최소화하도록 학습되는 것을 특징으로 하는 계층적 분석 및 전이학습을 이용한 부동산 경매 낙찰가율 예측시스템.

청구항 13

청구항 12에 있어서,

상기 손실함수는 MSE loss(Mean Squared Error loss)로서, 다음의 수학식 2로 나타나는 것을 특징으로 하는 계층적 분석 및 전이학습을 이용한 부동산 경매 낙찰가율 예측시스템.

[수학식 2]

$$MSE\ Loss = \frac{1}{n} \sum_i^n (y_i - \hat{y}_i)$$

(여기서, n은 학습 데이터 샘플의 수를 나타내며, y_i 는 i번째 학습 데이터 샘플의 실제 값을 나타내며, \hat{y}_i 는 i번째 학습 데이터 샘플의 관찰치(예측치)를 나타낸다.)

청구항 14

청구항 1에 있어서,

상기 지역계층화 분류부에서 구분된 지역계층 개수가 4개인 L(4)인 때에는,

최초에는 데이터셋 S(1)에 초기 가중치 W(1)를 무작위로 부여하여 학습시켜 학습모델 M(1)을 생성하고, 데이터셋 S(1)에서 데이터셋 S(2)를 제외시킨 잔여 데이터셋으로 직전 학습모델 M(1)을 학습시켜 새로운 학습모델 M(2)을 생성시키고,

다음으로, 직전 학습모델 M(2)의 조정 가중치 W(2)를 통해 새로운 학습모델 M(3)을 만들고, 데이터셋 S(2)에서 데이터셋 S(3)을 제외시킨 잔여 데이터셋으로 학습모델 M(3)의 전이학습을 추가로 수행하며,

다음으로, 직전 학습모델 M(3)의 조정 가중치 W(3)를 통해 새로운 학습모델 M(4)을 만들고, 데이터셋 S(3)에서 데이터셋 S(4)을 제외시킨 잔여 데이터셋으로 학습모델 M(4)의 전이학습을 추가로 수행하는 것을 특징으로 하는 계층적 분석 및 전이학습을 이용한 부동산 경매 낙찰가율 예측시스템.

청구항 15

제어서버 및 부동산 경매물건 데이터베이스가 네트워크로 연결되고, 컴퓨터를 포함하는 연산처리수단에 의해 경매물건 데이터를 딥러닝하여 낙찰가율을 예측하도록 수행되는 부동산 경매 낙찰가율 예측방법으로서,

제어서버의 데이터가공 및 변수생성부가 경매물건 데이터를 가공하고, 변수를 생성하는 S100 단계; 제어서버의 데이터 분류부가 가공된 경매물건 데이터를 지역별 및 용도별로 분류하는 S200 단계; 제어서버의 지역계층화 분류부가 분류된 경매물건 데이터를 지역의 광협을 기준으로 복수개의 지역계층지역계층으로 분류하는 S300 단계; 및 제어서버의 낙찰가율 예측부가 계층화된 지역의 경매물건 데이터를 딥러닝 및 전이학습하여 학습모델을 생성하고 경매물건의 낙찰가율을 예측하는 S400 단계를 포함하며,

S400 단계의 상기 낙찰가율 예측부는 학습데이터 전처리부가 모델의 학습을 위해 데이터 전처리를 수행하는 S410 단계 및 딥러닝 및 전이학습 수행부가 전처리된 데이터셋을 이용하여 딥러닝 및 전이학습을 수행하는 S420 단계를 포함하며,

S420 단계의 상기 딥러닝 및 전이학습 수행부는 상기 지역계층화 분류부의 지역계층 L(t:t는 기 설정된 양의 정수)에 대응되는 데이터셋 S(t)을 구비하며, 최초에는 데이터셋 S(t)에 초기 가중치 W(t)를 무작위로 부여하여 학습시켜 학습모델 M(t)을 생성하고, 데이터셋 S(t)에서 데이터셋 S(t+1)을 제외시킨 잔여 데이터셋으로 직전 학습모델(t)을 학습시켜 새로운 학습모델 M(t+1)을 생성시키고, 다음으로, 직전 학습모델 M(t+1)의 조정 가중치 W(t+1)를 통해 새로운 학습모델 M(t+2)을 만들고, 데이터셋 S(t+1)에서 데이터셋 S(t+2)를 제외시킨 잔여 데이터셋으로 새로운 학습모델 M(t+2)의 전이학습을 추가로 수행하는 것을 반복하는 것을 특징으로 하는 계층적 분

석 및 전이학습을 이용한 부동산 경매 낙찰가율 예측방법.

청구항 16

삭제

청구항 17

삭제

청구항 18

하드웨어와 결합되어, 청구항 15에 기재된 계층적 분석 및 전이학습을 이용한 부동산 경매 낙찰가율 예측방법을 컴퓨터에 의해 실행시키기 위하여 컴퓨터가 판독 가능한 기록매체에 저장된 컴퓨터 프로그램.

발명의 설명

기술 분야

[0001] 본 발명은 부동산 경매 낙찰가율 예측시스템 및 예측방법에 관한 것이다. 구체적으로는 계층적 분석 및 전이학습을 이용한 부동산 경매 낙찰가율 예측시스템 및 예측방법에 관한 것이다.

배경 기술

[0002] 부동산 경매는 경매 물건에 대한 감정가를 기반으로 경매가 진행되고 경매 입찰에 의해 최종 낙찰가가 결정되게 된다. 이때 감정가 대비 낙찰가의 비율을 '낙찰가율'이라고 한다.

[0003] 부동산 경매 물건에 대한 다양한 데이터와 변수가 존재하는데, 이러한 다양한 변수를 낙찰가율 예측에 사용함에 있어서 딥러닝 모델을 사용해 변수들을 학습하고 가중치를 결정할 수 있을 것이다.

[0004] 머신러닝 모델의 정밀한 예측을 위해, 학습이 수행될 때에는 훈련 데이터가 필요하므로, 훈련 데이터의 개수를 일정 수준 이상으로 확보하는 것이 중요하다.

[0005] 대한민국의 경우, 2005년~2020년까지 진행된 부동산 경매 낙찰 물건은 약 90만건이 있고, 경매 물건은 동산을 제외하고 59가지 용도로 구분하는데, 부동산의 용도별로 특징이 상이하고 평균적인 낙찰가율도 상이하다.

[0007] 종래 부동산 영역에서 물건의 가격이나 매매 지수 예측이 다양한 양상으로 시도되었지만 다음과 같은 문제점이 있었다.

[0008] 첫째, 데이터 수가 충분히 크고 매매가 예측력이 높은 한정된 지역, 한정된 용도에 대한 분석만이 주로 수행되는 문제점이 있었다. 전국적인 분석이 아닌 한정된 지역, 특히 수도권 또는 단일 지역자치단체에 대한 분석이 주를 이루었다. 부동산은 지역적 특성이 강한 분야인만큼 전국적 분석이 잘 이루어지지 않고 있다. 부동산 경매 물건의 용도의 경우, 아파트와 같은 주된 종목에 대한 분석이 주로 수행되었다. 공장, 임야 등의 용도의 경우, 상대적으로 경매물건 데이터 수가 적으며, 물건별로 가격차가 큰 종목에 대한 분석이 미비한 실정이었다.

[0010] 둘째, 다수의 지역에 대한 분석의 경우 세부 단위 지역들끼리의 군집화를 이용하려는 시도가 있었으나 매우 단순한 형태로 이루어지는 문제점이 있었다. 예를 들어, 군집화의 기준이 되는 변수가 단일 변수의 형태로 수행되었다. 또한, 복수가 아닌 단일횟수 군집화가 수행되었고, 군집화가 한번 이루어지고 나면 유사성이 높은 군집끼리 다시 군집으로 묶지 않는 실정이었다.

[0012] 셋째, 거리적 유사성을 기반으로 개별 물건을 평가하려는 시도가 있었으나, 지역별 또는 용도별로 유사성 측정이 세분화되지 않고 일률적인 방법을 적용하는 문제점이 있었다. 지역별 및/또는 용도별 고려가 필요한 이유는 예를 들면, 수도권 아파트의 경우 유사한 물건이 매우 근소한 거리 내에 있을 가능성이 높으나, 지방 상가의 경우 유사한 물건이 근소한 거리 내에 있을 가능성이 상대적으로 낮기 때문이다.

선행기술문헌

특허문헌

[0013] (특허문헌 0001) (문헌 1) 한국등록특허공보 제10-2130104호 (2020.06.29)

발명의 내용

해결하려는 과제

- [0014] 본 발명에 따른 계층적 분석 및 전이학습을 이용한 부동산 경매 낙찰가율 예측시스템 및 예측방법은 다음과 같은 해결과제를 가진다.
- [0015] 첫째, 지역별 데이터의 규모가 불균등하더라도, 빅데이터에 기반한 분석 알고리즘을 적용하여 정확한 분석이 가능하도록 한다.
- [0016] 둘째, 데이터의 수가 적어 훈련에 어려움이 큰 지역과 용도별 경매 데이터의 경우에도 정확한 분석이 가능하도록 한다.
- [0017] 셋째, 데이터 규모를 기준으로 지역별, 용도별 군집화를 수행하여 데이터 맞춤형 분석이 가능하도록 한다.
- [0018] 본 발명의 해결과제는 이상에서 언급한 것들에 한정되지 않으며, 언급되지 아니한 다른 해결과제들은 아래의 기재로부터 당업자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

- [0019] 본 발명은 제어서버 및 부동산 경매물건 데이터베이스가 네트워크로 연결되고, 컴퓨터를 포함하는 연산처리수단에 의해 경매물건 데이터를 딥러닝하여 낙찰가율을 예측하도록 수행되는 부동산 경매 낙찰가율 예측시스템으로서, 상기 제어서버는 경매물건 데이터를 가공하고, 변수를 생성하는 데이터가공 및 변수생성부; 가공된 경매물건 데이터를 지역별 및 용도별로 분류하는 데이터 분류부; 분류된 경매물건 데이터를 지역의 광협을 기준으로 복수개의 지역계층으로 분류하는 지역계층화 분류부; 및 지역계층화된 경매물건 데이터를 딥러닝 및 전이학습하여 학습모델을 생성하고 경매물건의 낙찰가율을 예측하는 낙찰가율 예측부를 포함한다.
- [0020] 본 발명에 있어서, 상기 데이터가공 및 변수생성부에서, 독립변수는 낙찰가율이며, 종속변수는 지형, 법면, 도로접합, 토지거래허가, 개발제한구역 또는 지역지구변수값 중 적어도 하나를 포함한다.
- [0021] 본 발명에 있어서, 상기 데이터가공 및 변수생성부에서, 로그화를 진행하는 종속변수는 감정가, 감정경과일자, 건물면적, 공고후조회수, 국민은행시세, 기계비중, 대표이외건물감정비율, 대표이외토지감정비율, 실거래가최근 3개월, 이전매매가, 임대보증금비율, 제시외비중, 총세대수, 최근3개월낙찰가율, 최저가, 토지면적 또는 공시지가 중 적어도 어느 하나인 것이 가능하다.
- [0022] 본 발명에 있어서, 상기 데이터 분류부에서 지역별 분류는 기 설정된 지역 단위로 분류될 수 있다.
- [0023] 본 발명에 있어서, 상기 데이터 분류부에서 용도별 분류는 아파트용도, 연립용도, 다가구용도, 공장용도, 상가용도, 오피스텔용도, 전담및과수원용도, 임야용도, 대지용도 및 공원기타용도로 분류될 수 있다.
- [0024] 본 발명에 있어서, 상기 지역계층화 분류부는 경매물건 데이터를 광의의 지역계층($t:t$ 는 양의 정수)에서 협의의 지역계층($t+1$)의 순서로 구분하며, 상기 협의의 지역은 상기 광의의 지역에 포함되는 지역이며, 구분되는 지역계층 개수는 기 설정될 수 있다.
- [0025] 본 발명에 있어서, 상기 상기 지역계층화 분류부에서 구분된 지역계층 개수가 4개인 때에는, 지역계층1은 대한민국 전국의 지역, 지역계층2는 남북을 통합한 도 단위 지역, 지역계층3은 전국의 특별시, 광역시와 도 단위 지역, 지역계층4는 특별시, 광역시를 제외한 시 단위 지역 및 구, 군단위 지역으로 구분될 수 있다.
- [0026] 본 발명에 있어서, 상기 낙찰가율 예측부는 학습모델의 학습을 위해 데이터 전처리를 수행하는 학습데이터 전처리부 및 전처리된 데이터셋을 이용하여 딥러닝 및 전이학습을 수행하는 딥러닝 및 전이학습 수행부를 포함할 수 있다.
- [0027] 본 발명에 있어서, 상기 학습데이터 전처리부는 데이터 전처리시에 다음의 수학식 1을 사용하여 학습데이터에 있는 이상치 데이터를 제거할 수 있다.

[0028] [수학식 1]

$$Range = [Q_1 - 1.5(Q_3 - Q_1), Q_3 + 1.5(Q_3 - Q_1)]$$

[0029]

[0030] (여기서, Range는 Tukey fences를 시행했을 때 정상범위로 정의하는 범위를 나타내며, Q1은 전체 데이터를 낙찰가율의 크기에 따라 순서대로 정렬하였을 때 작은 쪽의 사분위 수를 나타내며, Q3는 전체 데이터를 낙찰가율의 크기에 따라 순서대로 정렬하였을 때 큰 쪽의 사분위 수를 나타낸다.)

[0031] 본 발명에 있어서, 상기 학습데이터 전처리부는 학습데이터의 일 변수에 값이 미기재된 비율이 학습데이터 개수와 대비하여 기 설정된 비율 이상이면 상기 변수를 학습에서 제외시킬 수 있다.

[0032] 본 발명에 있어서, 상기 딥러닝 및 전이학습 수행부는 상기 지역계층화 분류부의 지역계층 $L(t:t$ 는 기 설정된 양의 정수)에 대응되는 데이터셋 $S(t)$ 을 구비하며, 최초에는 데이터셋 $S(t)$ 에 초기 가중치 $W(t)$ 를 무작위로 부여하여 학습시켜 학습모델 $M(t)$ 를 생성하고, 데이터셋 $S(t)$ 에서 데이터셋 $S(t+1)$ 을 제외시킨 잔여 데이터셋으로 직전 학습모델(t)를 학습시켜 새로운 학습모델 $M(t+1)$ 을 생성시키고, 다음으로, 직전 학습모델 $M(t+1)$ 의 조정 가중치 $W(t+1)$ 를 통해 새로운 학습모델 $M(t+2)$ 를 만들고, 데이터셋 $S(t+1)$ 에서 데이터셋 $S(t+2)$ 을 제외시킨 잔여 데이터셋으로 새로운 학습모델 $M(t+2)$ 의 전이학습을 추가로 수행하는 것을 반복할 수 있다.

[0033] 본 발명에 있어서, 상기 학습모델은 손실함수(loss function)를 최소화하도록 학습될 수 있다.

[0034] 본 발명에 있어서, 상기 손실함수는 MSE loss(Mean Squared Error loss)로서, 다음의 수학식 2로 나타낼 수 있다.

[0035] [수학식 2]

$$MSE\ Loss = \frac{1}{n} \sum_i^n (y_i - \hat{y}_i)$$

[0036]

[0037] (여기서, n 은 학습 데이터 샘플의 수를 나타내며, y_i 는 i 번째 학습 데이터 샘플의 실제 값을 나타내며, \hat{y}_i 는 i 번째 학습 데이터 샘플의 관찰치(예측치)를 나타낸다.)

[0038] 본 발명에 있어서, 상기 지역계층화 분류부에서 구분된 지역계층 개수가 4개인 $L(4)$ 인 때에는, 최초에는 데이터셋 $S(1)$ 에 초기 가중치 $W(1)$ 를 무작위로 부여하여 학습시켜 학습모델 $M(1)$ 을 생성하고, 데이터셋 $S(1)$ 에서 데이터셋 $S(2)$ 을 제외시킨 잔여 데이터셋으로 직전 학습모델 $M(1)$ 을 학습시켜 새로운 학습모델 $M(2)$ 을 생성시키고, 다음으로, 직전 학습모델 $M(2)$ 의 조정 가중치 $W(2)$ 를 통해 새로운 학습모델 $M(3)$ 를 만들고, 데이터셋 $S(2)$ 에서 데이터셋 $S(3)$ 을 제외시킨 잔여 데이터셋으로 학습모델 $M(3)$ 의 전이학습을 추가로 수행하며, 다음으로, 직전 학습모델 $M(3)$ 의 조정 가중치 $W(3)$ 를 통해 새로운 학습모델 $M(4)$ 를 만들고, 데이터셋 $S(3)$ 에서 데이터셋 $S(4)$ 을 제외시킨 잔여 데이터셋으로 학습모델 $M(4)$ 의 전이학습을 추가로 수행할 수 있다.

[0040] 본 발명은 제어서버 및 부동산 경매물건 데이터베이스가 네트워크로 연결되고, 컴퓨터를 포함하는 연산처리수단에 의해 경매물건 데이터를 딥러닝하여 낙찰가율을 예측하도록 수행되는 부동산 경매 낙찰가율 예측방법으로서, 제어서버의 데이터가공 및 변수생성부가 경매물건 데이터를 가공하고, 변수를 생성하는 S100 단계;

[0041] 제어서버의 데이터 분류부가 가공된 경매물건 데이터를 지역별 및 용도별로 분류하는 S200 단계; 제어서버의 지역계층화 분류부가 분류된 경매물건 데이터를 지역의 광협을 기준으로 복수개의 지역계층/지역계층으로 분류하는 S300 단계; 및 제어서버의 낙찰가율 예측부가 계층화된 지역의 경매물건 데이터를 딥러닝 및 전이학습하여 학습모델을 생성하고 경매물건의 낙찰가율을 예측하는 S400 단계를 포함한다.

[0042] 본 발명에 있어서, S400 단계의 상기 낙찰가율 예측부는 학습데이터 전처리부가 모델의 학습을 위해 데이터 전처리를 수행하는 S410 단계 및 딥러닝 및 전이학습 수행부가 전처리된 데이터셋을 이용하여 딥러닝 및 전이학습을 수행하는 S420 단계를 포함한다.

[0043] 본 발명에 있어서, S420 단계의 상기 딥러닝 및 전이학습 수행부는 상기 지역계층화 분류부의 지역계층 $L(t:t$ 는 기 설정된 양의 정수)에 대응되는 데이터셋 $S(t)$ 을 구비하며, 최초에는 데이터셋 $S(t)$ 에 초기 가중치 $W(t)$ 를 무작위로 부여하여 학습시켜 학습모델 $M(t)$ 를 생성하고, 데이터셋 $S(t)$ 에서 데이터셋 $S(t+1)$ 을 제외시킨 잔여 데

이터넷으로 직전 학습모델(t)를 학습시켜 새로운 학습모델 M(t+1)을 생성시키고, 다음으로, 직전 학습모델 M(t+1)의 조정 가중치 W(t+1)를 통해 새로운 학습모델 M(t+2)를 만들고, 데이터셋 S(t+1)에서 데이터셋 S(t+2)를 제외시킨 잔여 데이터셋으로 새로운 학습모델 M(t+2)의 전이학습을 추가로 수행하는 것을 반복할 수 있다.

[0045] 본 발명은 하드웨어와 결합되어, 본 발명에 따른 계층적 분석 및 전이학습을 이용한 부동산 경매 낙찰가율 예측 방법을 컴퓨터에 의해 실행시키기 위하여 컴퓨터가 판독 가능한 기록매체에 저장된 컴퓨터 프로그램으로 구현될 수 있다.

발명의 효과

[0046] 본 발명에 따른 계층적 분석 및 전이학습을 이용한 부동산 경매 낙찰가율 예측시스템 및 예측방법은 다음과 같은 효과를 가진다.

[0047] 첫째, 부동산 분석이 한정된 지역이 아닌 전국적으로 이루어지도록 함으로써, 지역별 데이터의 규모가 불균등하더라도, 빅데이터에 기반한 분석 알고리즘을 적용하여 정확한 분석이 가능한 효과가 있다.

[0048] 둘째, 전이학습을 통해, 데이터의 수가 적어 훈련에 어려움이 큰 지역과 용도별 경매 데이터의 경우에도 정확한 분석이 가능한 효과가 있다.

[0049] 셋째, 데이터 규모를 기준으로 지역별, 용도별 군집화를 수행하여 데이터 맞춤형 분석이 가능한 효과가 있다.

[0050] 본 발명의 효과는 이상에서 언급된 것들에 한정되지 않으며, 언급되지 아니한 다른 효과들은 아래의 기재로부터 당업자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

[0051] 도 1은 본 발명에 따른 부동산 경매 낙찰가율 예측시스템의 구성도이다.

도 2는 로그화 결과의 예시로서, 도 2a는 로그화 진행 전의 원 데이터이며, 도 2b는 로그화를 진행한 후의 로그 데이터 결과이다.

도 3은 제로 데이터(zero data) 비율에 따른 변수의 개수를 나타내는 그래프이다.

도 4는 일 실시예에 따른 데이터셋을 활용하여 지역계층별 딥러닝을 하는 과정을 나타낸다.

도 5는 일 실시예에 따른 데이터셋을 활용하여 타겟 지역을 딥러닝하는 과정을 나타낸다.

도 6은 본 발명에 따른 부동산 경매 낙찰가율 예측방법의 순서도이다.

발명을 실시하기 위한 구체적인 내용

[0052] 이하, 첨부한 도면을 참조하여, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 본 발명의 실시예를 설명한다. 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자가 용이하게 이해할 수 있는 바와 같이, 후술하는 실시예는 본 발명의 개념과 범위를 벗어나지 않는 한도 내에서 다양한 형태로 변형될 수 있다. 가능한 한 동일하거나 유사한 부분은 도면에서 동일한 도면부호를 사용하여 나타낸다.

[0053] 본 명세서에서 사용되는 전문용어는 단지 특정 실시예를 언급하기 위한 것이며, 본 발명을 한정하는 것을 의도하지는 않는다. 여기서 사용되는 단수 형태들은 문구들이 이와 명백히 반대의 의미를 나타내지 않는 한 복수 형태들도 포함한다.

[0054] 본 명세서에서 사용되는 "포함하는"의 의미는 특정 특성, 영역, 정수, 단계, 동작, 요소 및/또는 성분을 구체화하며, 다른 특정 특성, 영역, 정수, 단계, 동작, 요소, 성분 및/또는 군의 존재나 부가를 제외시키는 것은 아니다.

[0055] 본 명세서에서 사용되는 기술용어 및 과학용어를 포함하는 모든 용어들은 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자가 일반적으로 이해하는 의미와 동일한 의미를 가진다. 사전에 정의된 용어들은 관련 기술문헌과 현재 개시된 내용에 부합하는 의미를 가지는 것으로 추가 해석되고, 정의되지 않는 한 이상적이거나 매우 공식적인 의미로 해석되지 않는다.

[0057] 딥러닝(Deep Learning)은 여러 비선형 변환기법의 조합을 통해 높은 수준의 추상화(abstractions, 다량의 데이터나 복잡한 자료들 속에서 핵심적인 내용 또는 기능을 요약하는 작업)를 시도하는 기계 학습 알고리즘의 집합

을 의미한다.

- [0058] 가중치(weight)는 입력신호가 결과 출력에 주는 영향도를 조절하는 매개변수를 의미한다.
- [0059] 전이학습(Transfer Learning)은 특정한 Task(Classification, Detection, Segmentation 등)에 대하여 학습된 딥러닝 모델을, 다른 Task로 Transfer(전이)하여 해당 모델을 사후적으로 학습하는 개념을 포괄한다. 특정 분야에서 학습된 신경망의 일부 능력을 유사하거나 전혀 새로운 분야에서 사용되는 신경망의 학습에 이용하는 것을 의미한다. 따라서 기존의 만들어진 모델을 사용하여 새로운 모델을 만들때 학습을 빠르게 하여, 예측을 더 높이는 방법에 해당된다.
- [0061] 이하에서는 도면을 참고하여 본 발명을 설명하고자 한다. 참고로, 도면은 본 발명의 특징을 설명하기 위하여, 일부 과장되게 표현될 수도 있다. 이 경우, 본 명세서의 전 취지에 비추어 해석되는 것이 바람직하다.
- [0063] 본 발명은 계층적 분석 및 전이학습을 이용한 부동산 경매 낙찰가율 예측시스템에 관한 것이다.
- [0064] 본 발명은 제어서버(10) 및 부동산 경매물건 데이터베이스(20)가 네트워크로 연결되고, 컴퓨터를 포함하는 연산 처리수단에 의해 경매물건 데이터를 딥러닝하여 낙찰가율을 예측하도록 수행되는 부동산 경매 낙찰가율 예측시스템이다.
- [0065] 본 발명에 따른 네트워크는 유선 및 무선 네트워크를 포함한다.
- [0067] 도 1은 본 발명에 따른 부동산 경매 낙찰가율 예측시스템의 구성도이다.
- [0068] 본 발명에 따른 제어서버(10)는 경매물건 데이터를 가공하고, 변수를 생성하는 데이터가공 및 변수생성부(100); 가공된 경매물건 데이터를 지역별 및 용도별로 분류하는 데이터 분류부(200); 분류된 경매물건 데이터를 지역의 광협을 기준으로 복수개의 지역계층으로 분류하는 지역계층화 분류부(300); 및 지역계층화된 경매물건 데이터를 딥러닝하여 학습모델을 생성하고 경매물건의 낙찰가율을 예측하는 낙찰가율 예측부(400)를 포함한다.
- [0070] 본 발명에 따른 데이터가공 및 변수생성부(100)는 경매물건 데이터를 가공하고, 변수를 생성할 수 있다.
- [0071] 데이터가공 및 변수생성부(100)에서, 독립변수는 낙찰가율이며, 종속변수는 지형, 법면, 도로접합, 토지거래허가, 개발제한구역 또는 지역지구변수갯수 중 적어도 하나를 포함한다.
- [0072] 본 발명에서는 각 경매물건에 대한 감정평가서 요약 문자열 데이터를 변수화하여 물건에 대한 낙찰가율을 예측하는 변수로 사용할 수 있다. 이때 사용되는 정보는 '지형' / '법면' / '도로접합' / '토지거래허가' / '개발제한구역' / '지역지구변수갯수' 등을 포함한다. 문자열 데이터의 '난방방식' 변수는 '가스' / '유류' / '지역' / '천장' / '개별' / '중앙' / '채래식'으로 분류하여 더미변수로 변환할 수 있다. 문자열 데이터의 '지역' 변수는 '시군구' / '시도' / '시도_남북통합' / '일반시'를 더미변수로 변환할 수 있다. 또한, '도로접합' / '법면' / '지형' / '용도' 변수들을 더미변수로 변환할 수 있다.
- [0073] 아래 표 1은 변수목록의 예시이다.

표 1

No	변수	No	변수
1	낙찰가율	33	맹지
2	시도	34	일부맹지
3	시구군	35	건물만
4	읍면동	36	토지만
5	용도	37	대지권미등기
6	경매구분	38	대지권없음
7	감정가	39	토지별도등기
8	최저가	40	재매각
9	최근3개월낙찰가율	41	위반건축물
10	일괄소재지갯수	42	대항력임차인
11	대표이외토지감정비율	43	총층수
12	대표이외건물감정비율	44	층수
13	입찰보증금	45	보존일자
14	임대보증금비율	46	난방방식
15	임차인수	47	총세대수
16	변경횟수	48	대형건설사
17	감정경과일자	49	이전매매가
18	1금융채권자	50	국민은행시세
19	병합중복	51	실거래가최근3개월
20	건물면적	52	관심물건수
21	토지면적	53	공고후조회수
22	제시외비중	54	용도그룹A
23	기계비중	55	공시지가
24	유치권	56	납부상태
25	법정지상권	57	납부일자
26	지분경매	58	최종상태
27	분묘기지권	59	지형
28	선순위전세권	60	법면
29	선순위임차권	61	도로접합
30	선순위가처분	62	토지거래허가
31	선순위가등기	63	개발제한구역
32	일부지분경매	64	지역지구변수갯수

[0074]

[0076]

변수들 중 이상치 또는 값의 편향 때문에 데이터의 그래프가 정규성을 띄지 못하는 변수들이 존재할 수 있다. 해당 변수들을 로그화 하였을 때 데이터가 정규성을 갖게 되고 학습이 쉬운 데이터가 될 수 있다. 도 2는 로그화 결과의 예시로서, 도 2a는 로그화 진행 전의 원 데이터이며, 도 2b는 로그화를 진행한 후의 로그 데이터 결과이다. 도 2를 통해, 매우 편향되어 학습하기 어려웠던 데이터가 로그화를 통해 정규성을 갖게 되는 것을 확인할 수 있다.

[0077]

본 발명에 따른 데이터가공 및 변수생성부(100)에서, 로그화를 진행하는 종속변수는 감정가, 감정경과일자, 건물면적, 공고후조회수, 국민은행시세, 기계비중, 대표이외건물감정비율, 대표이외토지감정비율, 실거래가최근3개월, 이전매매가, 임대보증금비율, 제시외비중, 총세대수, 최근3개월낙찰가율, 최저가, 토지면적 또는 공시지가 중 적어도 어느 하나인 것이 가능하다.

[0079]

본 발명에 따른 데이터 분류부(200)는 가공된 경매물건 데이터를 지역별 및 용도별로 분류할 수 있다.

[0080]

본 발명에서 지역별 분류는 기 설정된 지역 단위로 분류되는 것을 의미한다.

[0081]

일 실시예로서, 부동산 경매 물건 주소지 기준 전국의 250개 시구군으로 데이터를 분류할 수 있다.

[0082]

기존의 부동산 경매의 경우 59개의 용도 구분을 사용한다. 하지만 경매 데이터를 59개를 묶을 경우 각각의 용도별 데이터의 수가 충분하지 않아 학습을 수행할 때 충분히 학습되지 못하는 결과를 낸다.

[0083]

이에, 본 발명에 따른 상기 데이터 분류부(200)에서 용도별 분류는 아파트용도, 연립용도, 다가구용도, 공장용도, 상가용도, 오피스텔용도, 전답및과수원용도, 임야용도, 대지용도 및 공원기타용도로 분류될 수 있다.

[0084]

본 발명에서는 위와 같이 유사한 용도를 그룹으로 집결하여 표 2의 분류표를 생성하였다.

표 2

용도 분류표	
1	아파트 주상복합(아파트) 아파트(생활주택)
2	다세대 연립 다세대(생활주택) 연립(생활주택)
3	주택 다가구주택 근린주택
4	공장 공장용지 제조업소
5	상가 목욕시설근린상가 목욕시설상가 목욕시설아파트상가
	목욕시설주상복합 근린상가 아파트형공장(상가) 아파트형공장
	점포 시장 아파트상가 주상복합(상가) 오피스텔(상가) 기숙사(고시원)
6	오피스텔(주거용) 오피스텔(업무용) 사무실 빌딩
7	전 답 과수원
8	임야
9	대지 잡종지
10	공원 광천지 구거 기타토지 도로 목장용지 묘지 사적지 수도용지 양어장용지 염전 유원지 유지 제방 종교용지 주유소용지 주차장용지 창고용지 철도용지 체육용지 하천 학교용지

[0085]

[0087]

본 발명에 따른 지역계층화 분류부(300)는 분류된 경매물건 데이터를 지역의 광협을 기준으로 복수개의 지역계층으로 분류할 수 있다.

[0088]

본 발명에 따른 지역계층화 분류부(300)는 경매물건 데이터를 광의의 지역계층($t:t$ 는 기 설정된 양의 정수)에서 협의의 지역계층($t+1$)의 순서로 구분하며, 상기 협의의 지역은 상기 광의의 지역에 포함되는 지역이며, 구분되는 지역계층 개수는 기 설정될 수 있다.

[0090]

대한민국의 행정구역은 현재 1특별시, 6광역시, 1특별자치시, 8도, 1특별자치도로 편제되어 있다. 3단계 행정체계로 도/특별시/광역시 > 시/군/구 > 읍/면/동으로 구분된다. 다만, 서울특별시와 광역시를 제외한 인구 50만 이상 대도시에는 시 아래에 행정구를 둘 수 있다. 이들 중 도와 동급의 행정구역은 광역시이며, 특별시, 특별자치시, 특별자치도도 존재한다.

[0092]

대한민국의 경우, 일 실시예로서 지역계층 개수는 4개로 설정되어, 다음 표 3과 같이 구분될 수 있다.

표 3

지역계층1	대한민국 전국의 지역
지역계층2	도 단위 지역(남도와 북도를 통합한 지역)
지역계층3	특별시, 광역시와 도 단위 지역(남도와 북도로 구분된 지역)
지역계층4	전국의 시(특별시, 광역시를 제외)단위 지역 및 구, 군단위 지역

[0095]

이를, 전술한 현재의 행정구역을 반영하면 다음 표 4와 같이 구분될 수도 있다.

표 4

지역계층1	전국
지역계층2	경기도, 강원도, 충청도, 경상도, 전라도, 제주도
지역계층3	서울특별시 부산광역시, 인천광역시, 대구광역시, 대전광역시, 광주광역시, 울산광역시 충청남도, 충청북도, 경상남도, 경상북도, 전라남도, 전라북도
지역계층4	수원시, 화성시, 평택시 등등

[0097]

여기서, 지역별 및 용도별 구분의 필요에 따라, 경기도를 경기남도와 경기북도로 임의로 구분하는 것이 가능하다. 또한, 연관성이 높은 지역을 하나의 단위로 묶는 것도 가능하다. 예를 들어, 전라북도의 무주군, 진안군,

장수군을 1개의 단위로 삼을 수도 있다.

- [0099] 본 발명에 따른 낙찰가율 예측부(400)는 지역계층화된 경매물건 데이터를 딥러닝 및 전이학습하여 학습모델을 생성하고 경매물건의 낙찰가율을 예측할 수 있다.
- [0100] 본 발명에 따른 낙찰가율 예측부(400)는 학습모델의 학습을 위해 데이터 전처리를 수행하는 학습데이터 전처리부(410) 및 전처리된 데이터셋을 이용하여 딥러닝 및 전이학습을 수행하는 딥러닝 및 전이학습 수행부(420)를 포함한다.
- [0102] 본 발명은 학습모델의 학습을 원활히 하기 위해 학습 데이터에 대해 전처리를 수행한다. 학습 데이터의 낙찰가율을 기준으로 학습을 방해하는 이상치를 갖는 데이터들이 존재하기 때문에 Tukey Fences Method를 사용하여 학습 데이터에 있는 이상치 데이터들을 제거한다.
- [0103] 본 발명에 따른 학습데이터 전처리부(410)는 데이터 전처리시에 Tukey Fences의 범위를 정하는 다음의 수학적 식 1을 사용하여 학습데이터에 있는 이상치 데이터를 제거할 수 있다.

수학적 식 1

$$Range = [Q_1 - 1.5(Q_3 - Q_1), Q_3 + 1.5(Q_3 - Q_1)]$$

- [0104]
- [0105] 여기서, Range는 Tukey fences를 시행했을 때 정상범위로 정의하는 범위를 나타내며, Q1은 전체 데이터를 낙찰가율의 크기에 따라 순서대로 정렬하였을 때 작은 쪽의 사분위 수(lower quartiles)를 나타내며, Q3는 전체 데이터를 낙찰가율의 크기에 따라 순서대로 정렬하였을 때 큰 쪽의 사분위 수(upper quartiles)를 나타낸다.
- [0107] 본 발명에 따른 학습데이터 전처리부(410)는 학습데이터의 일 변수에 값이 미기재된 비율이 학습데이터 개수와 대비하여 기 설정된 비율 이상이면 상기 변수를 학습에서 제외시킬 수 있다.
- [0109] 도 3은 제로 데이터(zero data) 비율에 따른 변수의 개수를 나타내는 그래프이다. 도 3에 도시된 바와 같이, 변수들 중 데이터에서 0 값을 갖는 비율이 높은 변수들이 많다. 0 값의 비율이 높은 변수의 경우 정보는 거의 담고 있지 않지만, 데이터의 차원수를 늘리기 때문에 학습하는데 방해가 된다. 따라서 0 값을 일정 비율 이상 갖고 있는 변수를 제거할 수 있다.
- [0110] 일 실시예로서, 더미화한 데이터에서 0이 차지하는 비율에 따라 카운트했을 때 0이 차지하는 비율이 95%보다 작은 변수는 96개밖에 되지 않기 때문에 95% 이상이 0으로 이루어진 변수들은 제거할 수 있다.
- [0112] 본 발명에 따른 딥러닝 및 전이학습 수행부(420)는 상기 지역계층화 분류부(300)의 지역계층 L(t:t는 기 설정된 양의 정수)에 대응되는 데이터셋 S(t)을 구비하며,
- [0113] 최초에는 데이터셋 S(t)에 초기 가중치 W(t)를 무작위로 부여하여 학습시켜 학습모델 M(t)를 생성하고, 데이터셋 S(t)에서 데이터셋 S(t+1)을 제외시킨 잔여 데이터셋으로 직전 학습모델(t)를 학습시켜 새로운 학습모델 M(t+1)을 생성시키고,
- [0114] 다음으로, 직전 학습모델 M(t+1)의 조정 가중치 W(t+1)를 통해 새로운 학습모델 M(t+2)를 만들고, 데이터셋 S(t+1)에서 데이터셋 S(t+2)를 제외시킨 잔여 데이터셋으로 새로운 학습모델 M(t+2)의 전이학습을 추가로 수행하는 것을 반복할 수 있다.
- [0116] 도 4는 일 실시예에 따른 데이터셋을 활용하여 지역계층별 딥러닝을 하는 과정을 나타낸다.
- [0117] L(t)는 지역계층을 나타내며, t는 양의 정수로서, 1부터 T까지 존재한다. 지역계층 L(T)의 데이터는 타겟 데이터셋으로 따로 구분한다.
- [0118] S(t)는 지역계층 L(t)에 대한 데이터셋을 나타내며, 이전 모델에서 전이된 학습모델 가중치(weight)를 W(t)로 나타낸다.
- [0119] W(t+1)는 데이터셋 S(t)에서 데이터셋 S(t+1)을 제외시킨 잔여 데이터셋으로 추가학습(전이학습)시킨 모델(t+1)의 조정 가중치를 나타낸다.

[0120] 한편, 최초 가중치 $W(1)$ 는 무작위(random)한 값이 부여될 수 있다.

[0122] 본 발명에 따른 학습모델은 손실 함수(loss function)를 최소화하도록 학습되며, 사용된 손실함수는 MSE loss(Mean Squared Error loss)로서, 다음의 수학적 식 2로 나타낼 수 있다.

수학적 식 2

$$MSE\ Loss = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)$$

[0123]

[0124] 여기서, n 은 학습 데이터 샘플의 수를 나타내며, y_i 는 i 번째 학습 데이터 샘플의 실제 값을 나타내며, \hat{y}_i 는 i 번째 학습 데이터 샘플의 관찰치(예측치)를 나타낸다.

[0126] 본 발명에 따른 전이학습을 L(4)의 사례를 예를 들어, 보다 상세히 설명하면 다음과 같다.

[0127] 먼저, 지역계층화 분류부에서 구분된 지역계층 개수가 4개인 L(4)인 때, 최초에는 초기 가중치 $W(1)$ 을 랜덤으로 부여하여 학습모델 M(1)을 생성하고, 데이터셋 S(1)에서 데이터셋 S(2)을 제외시킨 잔여 데이터셋으로 직전 학습모델 M(1)을 학습시켜 새로운 학습모델 M(2)를 생성한다.

[0128] 다음으로, 직전 학습모델 M(2)의 조정 가중치 $W(2)$ 를 통해 새로운 학습모델 M(3)를 만들고, 데이터셋 S(2)에서 데이터셋 S(3)을 제외시킨 잔여 데이터셋으로 학습모델 M(3)의 전이학습을 추가로 수행한다.

[0129] 다음으로, 직전 학습모델 M(3)의 조정 가중치 $W(3)$ 를 통해 새로운 학습모델 M(4)를 만들고, 데이터셋 S(3)에서 데이터셋 S(4)을 제외시킨 잔여 데이터셋으로 학습모델 M(4)의 전이학습을 추가로 수행한다.

[0131] 이하에서는, 일 실시예에 따른 전이학습 방법을 하나의 예를 들어 설명하기로 한다.

[0132] 일 실시예의 타겟 지역은 전라북도 임실군이며, 용도는 아파트이다.

[0133] 계층별 지역은 다음 표 5와 같다.

표 5

[0134]

지역계층1(전국)	S(1) : 전국
지역계층2(시도_남북통합)	S(2) : 전라도
지역계층3(시도)	S(3) : 전라북도
지역계층4(시군군)	S(4) : 임실군

[0136] 먼저, 초기 가중치 $W(1)$ 을 랜덤으로 부여하여 학습모델 M(1)을 생성하고, ‘전국’에 관한 데이터셋 S(1)에서 ‘전라도’에 관한 데이터셋 S(2)을 제외시킨 잔여 데이터셋으로 학습모델 M(1)의 전이학습을 추가로 수행하여, 새로운 학습모델 M(2)를 생성한다.

[0137] 다음으로, 직전 학습모델 M(2)의 조정 가중치 $W(2)$ 를 통해 새로운 학습모델 M(3)를 만들고, ‘전라도’에 관한 데이터셋 S(2)에서 ‘전라북도’에 관한 데이터셋 S(3)을 제외시킨 잔여 데이터셋으로 학습모델 M(3)의 전이학습을 추가로 수행한다.

[0138] 다음으로, 직전 학습모델 M(3)의 조정 가중치 $W(3)$ 를 통해 새로운 학습모델 M(4)를 만들고, ‘전라북도’에 관한 데이터셋 S(3)에서 ‘임실군’에 관한 데이터셋 S(4)을 제외시킨 잔여 데이터셋으로 학습모델 M(4)의 전이학습을 추가로 수행한다.

[0140] 이하에서는 본 발명에 따른 전이학습모델을 타 모델과 대비하고자 한다.

[0141] 먼저, 데이터 분할의 경우, 2010~2020년 대한민국의 법원경매 데이터 중 2010~2019년 데이터를 훈련 데이터, 2020년 데이터를 테스트 데이터로 사용하였다.

[0142] 다음으로, 비교모델은 다음 표 6과 같이 구분될 수 있다.

표 6

제1 모델	본 발명에 따른 전이학습모델
제2 모델	가장 작은 지역단위인 시군구를 기준으로, 해당 시군구 내의 경매 데이터만을 사용해 훈련한 모델
제3 모델	가장 큰 지역단위인 전국을 기준으로, 전국 모든 물건의 경매 데이터를 사용해 훈련한 모델

다음으로, 결과 측정의 측도는 다음 수학적 식 3과 같은 평균 절대 백분율 오차(MAPE: Mean Absolute Percentage Error) 및 중앙 절대 비율 오차(MdAPE: Median Absolute Percentage Error)를 활용하였다.

수학적 식 3

$$MAPE = \frac{100}{n} \sum_{t=1}^n \left| \frac{Prediction_t - Actual_t}{Actual_t} \right|$$

여기서, MAPE는 t번째 샘플의 실제값을 Actual(t)로 두고, 예측값을 Prediction(t)로 두어, 둘 사이의 차의 절대값을 Actual(t)로 나누고 100을 곱해 백분율로 만들어, 모든 t에 대해 합한 값을 n으로 나눠 평균값을 구한 것이다.

수학적 식 4

$$MdAPE = median \left(\left| \frac{Prediction_t - Actual_t}{Actual_t} \right| \right) * 100$$

여기서, MdAPE는 t번째 샘플의 실제값을 Actual(t)로 두고, 예측값을 Prediction(t)로 두어, 둘 사이의 차의 절대값을 Actual(t)로 나누고 100을 곱해 백분율로 만들어, 모든 t에 대해 중앙값을 구한 것이다.

상기 수학적 식 3에 따른 모델 별 결과값은 다음 표 7과 같이 산출되었고, 상기 수학적 식 4에 따른 모델 별 결과값은 다음 표 8과 같이 산출되었다.

표 7

• MAPE(%)

용도분류	지역	데이터 수	Model1	Model2	Model3
1	전라_전북_임실군_임실군	39	9.03	18.38	15.51
1	서울_서울_종로구_종로구	55	10.74	19.16	11.40
1	서울_서울_도봉구_도봉구	143	5.43	9.87	6.58
1	강원_강원_정선군_정선군	657	7.00	9.70	10.58
1	부산_부산_연제구_연제구	707	5.49	6.98	7.32
1	전라_전남_보성군_보성군	1,122	4.67	5.51	7.04
8	부산_부산_연제구_연제구	40	8.92	50.63	37.89
8	부산_부산_기장군_기장군	52	12.60	51.14	22.58
8	경상_경북_봉화군_봉화군	230	7.39	25.50	10.37
8	서울_서울_광진구_광진구	544	5.67	10.14	6.35
8	전라_전남_목포시_목포시	1,088	8.03	8.88	9.87

표 8

• MdAPE(%)

용도분류	지역	데이터 수	Model1	Model2	Model3
1	전라_전북_임실군_임실군	39	8.71	15.23	11.93
1	서울_서울_종로구_종로구	55	9.74	16.87	12.89
1	서울_서울_도봉구_도봉구	143	4.37	9.14	6.73
1	강원_강원_정선군_정선군	657	5.89	7.51	8.44
1	부산_부산_연제구_연제구	707	4.38	5.01	6.21
1	전라_전남_보성군_보성군	1,122	4.22	4.88	4.81
8	부산_부산_연제구_연제구	40	16.15	35.17	28.91
8	부산_부산_기장군_기장군	52	10.58	21.35	16.45
8	경상_경북_봉화군_봉화군	230	6.54	18.34	8.77
8	서울_서울_광진구_광진구	544	4.43	9.25	5.44
8	전라_전남_목포시_목포시	1,088	6.42	6.88	8.46

[0156]

[0158]

상기 용도분류 1번은 일반적으로 부동산 경매 데이터의 수가 많고, 학습하기 쉬운 아파트, 주상복합(아파트), 아파트(생활주택) 등에 관한 것이다.

[0159]

상기 용도분류 8번은 상대적으로 부동산 경매 데이터의 수가 적고, 비교대상이 없어 학습이 어려운 임야 등에 관한 것이다.

[0160]

제2 모델의 경우, 해당 물건이 포함된 가장 작은 지역단위인 시군군 내의 부동산 경매 데이터를 가지고 학습하기 때문에, 가장 비슷한 물건들에 대해서 학습이 수행된다. 하지만 부동산 경매 데이터의 수가 적으면 학습이 제대로 되지 않는 단점이 있다.

[0161]

제3 모델의 경우, 전국의 모든 물건에 대해서 학습하기 때문에 해당 물건이 속한 지역단위의 데이터가 적어도 학습을 할 수는 있지만, 해당 물건이 속한 지역의 특성이 잘 반영되지 못하는 단점이 있다.

[0162]

각 모델의 결과를 보면, 제2 모델의 낙찰가율 예측능력은 데이터 개수가 적을수록 매우 떨어지는 것을 확인할 수 있다. 제3 모델도 지역적인 특성을 반영하지 못해 예측 오차가 크게 나타난다.

[0163]

반면에, 데이터가 부족한 용도분류 1번의 임실군 사례를 보았을 때 제1 모델의 경우 시군군만 학습하는 제2 모델과 전국으로 학습하는 제3 모델과 비교했을 때 큰 폭으로 오차를 줄이는 것을 확인할 수 있다.

[0164]

나아가, 데이터의 수가 충분한 용도분류 1번의 보성군 사례를 보았을 때도, 제1 모델이 제2 모델, 제3 모델과 비교했을 때 예측 성능을 개선 시키는 것을 확인할 수 있다.

[0166]

한편, 본 발명은 부동산 경매 낙찰가율 예측방법으로 구현될 수 있다. 예측방법은 전술한 예측시스템과 발명의 카테고리가 상이하나, 실질적인 기술구성은 동일하다. 이에, 공통되는 사항의 설명은 생략하고, 예측방법의 요지를 위주로 설명하고자 한다.

[0168]

도 6은 본 발명에 따른 부동산 경매 낙찰가율 예측방법의 순서도이다.

[0169]

도 6에 도시된 바와 같이, 본 발명은 제어서버(10) 및 부동산 경매물건 데이터베이스(20)가 네트워크로 연결되고, 컴퓨터를 포함하는 연산처리수단에 의해 경매물건 데이터를 딥러닝하여 낙찰가율을 예측하도록 수행되는 부동산 경매 낙찰가율 예측방법으로서, 제어서버(10)의 데이터가공 및 변수생성부(100)가 경매물건 데이터를 가공하고, 변수를 생성하는 S100 단계; 제어서버(10)의 데이터 분류부(200)가 가공된 경매물건 데이터를 지역별 및 용도별로 분류하는 S200 단계; 제어서버(10)의 지역계층화 분류부(300)가 분류된 경매물건 데이터를 지역의 광협을 기준으로 복수개의 지역계층지역계층으로 분류하는 S300 단계; 및 제어서버(10)의 낙찰가율 예측부(400)가 계층화된 지역의 경매물건 데이터를 딥러닝 및 전이학습하여 학습모델을 생성하고 경매물건의 낙찰가율을 예측하는 S400 단계를 포함한다.

[0171]

본 발명에 따른 S400 단계의 낙찰가율 예측부(400)는 학습데이터 전처리부(410)가 모델의 학습을 위해 데이터 전처리를 수행하는 S410 단계 및 딥러닝 및 전이학습 수행부(420)가 전처리된 데이터셋을 이용하여 딥러닝 및

전이학습을 수행하는 S420 단계를 포함한다.

[0172] 본 발명에 따른 S420 단계의 상기 딥러닝 및 전이학습 수행부(420)는 지역계층화 분류부(300)의 지역계층 $L(t:t)$ 는 기 설정된 양의 정수에 대응되는 데이터셋 $S(t)$ 을 구비하며, 최초에는 데이터셋 $S(t)$ 에 초기 가중치 $W(t)$ 를 무작위로 부여하여 학습시켜 학습모델 $M(t)$ 를 생성하고, 데이터셋 $S(t)$ 에서 데이터셋 $S(t+1)$ 을 제외시킨 잔여 데이터셋으로 직전 학습모델(t)을 학습시켜 새로운 학습모델 $M(t+1)$ 을 생성시키고, 다음으로, 직전 학습모델 $M(t+1)$ 의 조정 가중치 $W(t+1)$ 를 통해 새로운 학습모델 $M(t+2)$ 를 만들고, 데이터셋 $S(t+1)$ 에서 데이터셋 $S(t+2)$ 을 제외시킨 잔여 데이터셋으로 새로운 학습모델 $M(t+2)$ 의 전이학습을 추가로 수행하는 것을 반복할 수 있다.

[0174] 한편, 본 발명은 컴퓨터 프로그램으로 구현될 수도 있다. 구체적으로 본 발명은 하드웨어와 결합되어, 본 발명에 따른 계층적 분석 및 전이학습을 이용한 부동산 경매 낙찰가율 예측방법을 컴퓨터에 의해 실행시키기 위하여 컴퓨터가 판독 가능한 기록매체에 저장된 컴퓨터 프로그램으로 구현될 수 있다.

[0176] 본 발명의 실시예에 따른 방법들은 다양한 컴퓨터수단을 통하여 판독 가능한 프로그램 형태로 구현되어 컴퓨터로 판독 가능한 기록매체에 기록될 수 있다. 여기서, 기록매체는 프로그램 명령, 데이터 파일, 데이터구조 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), CDROM, DVD와 같은 광 기록매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치를 포함한다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어를 포함할 수 있다. 이러한 하드웨어 장치는 본 발명의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

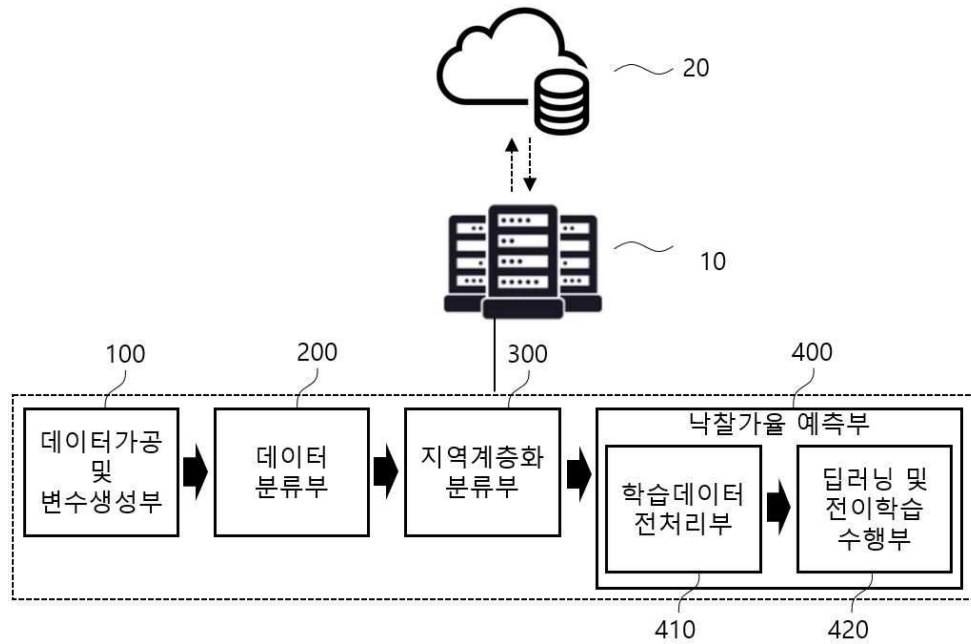
[0178] 본 명세서에서 설명되는 실시예와 첨부된 도면은 본 발명에 포함되는 기술적 사상의 일부를 예시적으로 설명하는 것에 불과하다. 따라서, 본 명세서에 개시된 실시예들은 본 발명의 기술적 사상을 한정하기 위한 것이 아니라 설명하기 위한 것이므로, 이러한 실시예에 의하여 본 발명의 기술 사상의 범위가 한정되는 것은 아님은 자명하다. 본 발명의 명세서 및 도면에 포함된 기술적 사상의 범위 내에서 당업자가 용이하게 유추할 수 있는 변형예와 구체적인 실시 예는 모두 본 발명의 권리범위에 포함되는 것으로 해석되어야 할 것이다.

부호의 설명

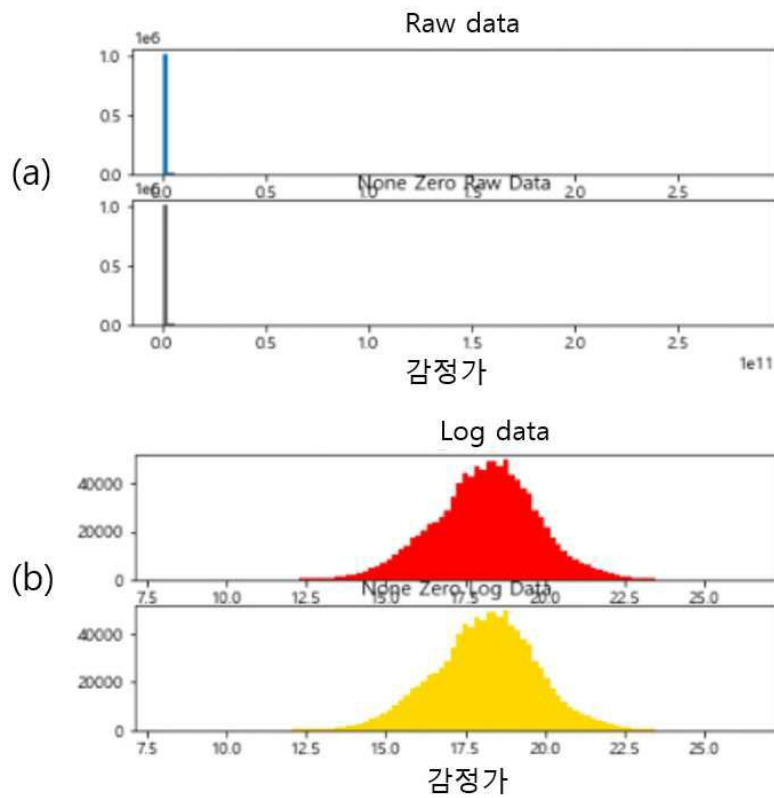
[0179] 10 : 제어서버
20 : 데이터베이스
100 : 데이터가공 및 변수생성부
200 : 데이터 분류부
300 : 지역계층화 분류부
400 : 낙찰가율 예측부
410 : 학습데이터 전처리부
420 : 딥러닝 및 전이학습 수행부

도면

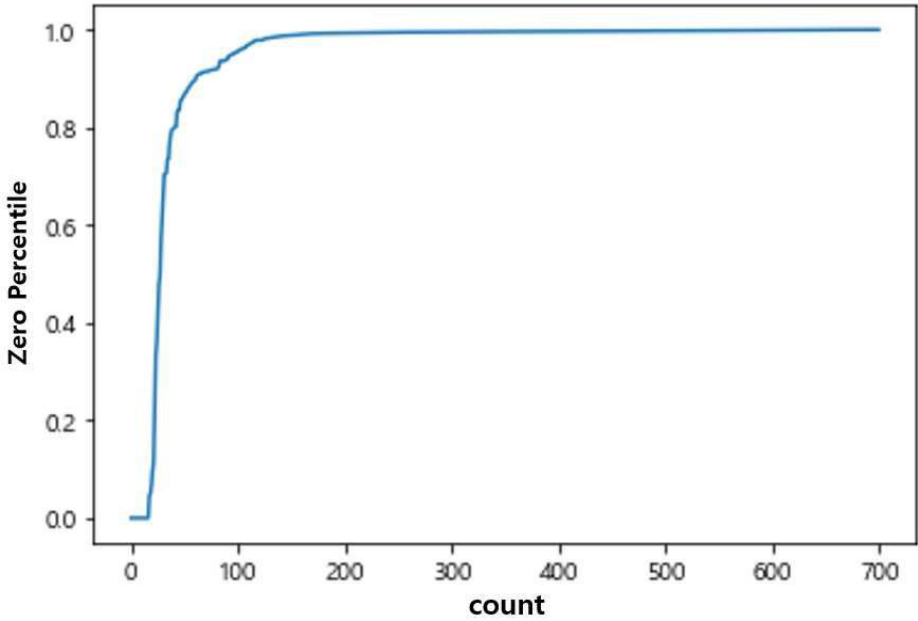
도면1



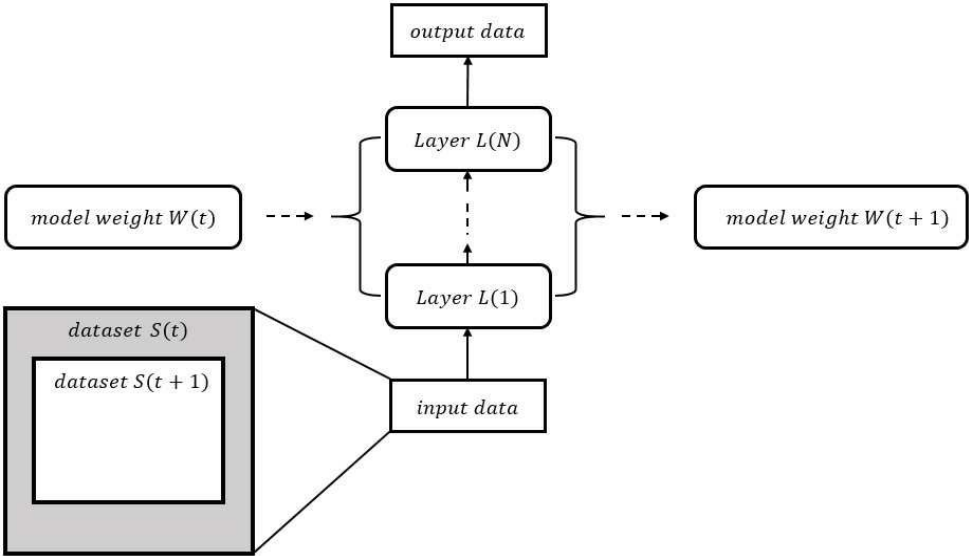
도면2



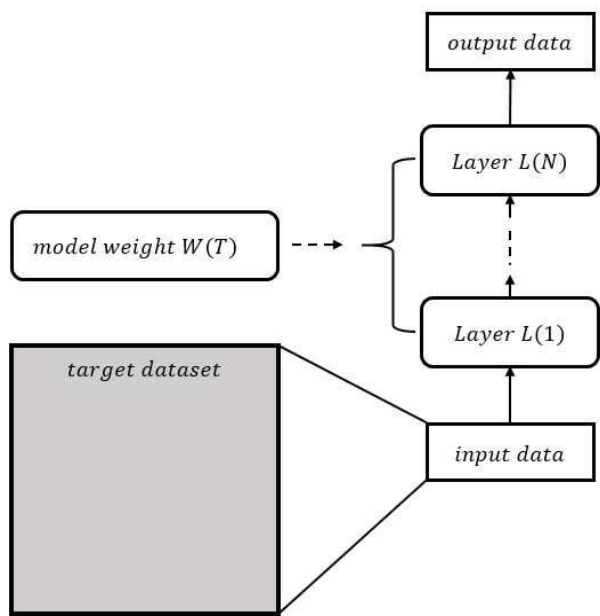
도면3



도면4



도면5



도면6

