

# 다변수 LSTM 순환신경망 딥러닝 모형을 이용한 미술품 가격 예측에 관한 실증연구

## An Empirical Study on Prediction of the Art Price using Multivariate Long Short Term Memory Recurrent Neural Network Deep Learning Model

이지민\*, 송정석\*\*

중앙대학교 문화예술경영학과\*, 중앙대학교 경제학부\*\*

Jiin Lee(jiinlee111@gmail.com)\*, Jeongseok Song(jssong@cau.ac.kr)\*\*

### 요약

새로운 미술품 유통방식의 발달로 미술품의 미적 효용을 넘어 투자재로서 바라보는 시각이 활성화되고 있다. 미술품의 가격은 주식이나 채권 등과 달리 객관적 요소와 주관적 요소들이 모두 반영되어 결정되는 이질적 특성이 있기 때문에 가격 예측에 있어서 그 불확실성이 높다. 본 연구에서는 LSTM(장단기 기억) 순환신경망 딥러닝 모형을 활용하여 낙찰총액 순위 1위부터 10위까지의 한국 작가의 회화 작품을 대상으로 작가의 특성, 작품의 물리적 특성, 판매적 특성 등을 입력으로 하여 경매 낙찰가의 예측을 시도하였다. 연구 결과, 모델에 의한 예측 가격과 실제 낙찰 가격의 차이를 설명하는 RMSE 값이 0.064 수준이었으며 작가별로는 이대원 작가의 예측력이 가장 높았고, 이종섭 작가의 예측력이 가장 낮았다. 투자재로서 미술품 시장이 더욱 활성화되고 경매 낙찰 가격의 예측 수요가 높아지면서 본 연구의 결과가 활용될 수 있을 것이다.

■ 중심어 : | 미술품 가격 | 기계학습 | 비모수 추정 | 딥러닝 | 다변수 장단기 기억 순환신경망 |

### Abstract

With the recent development of the art distribution system, interest in art investment is increasing rather than seeing art as an object of aesthetic utility. Unlike stocks and bonds, the price of artworks has a heterogeneous characteristic that is determined by reflecting both objective and subjective factors, so the uncertainty in price prediction is high. In this study, we used LSTM Recurrent Neural Network deep learning model to predict the auction winning price by inputting the artist, physical and sales characteristics of the Korean artist. According to the result, the RMSE value, which explains the difference between the predicted and actual price by model, was 0.064. Painter Lee Dae Won had the highest predictive power, and Lee Joong Seop had the lowest. The results suggest the art market becomes more active as investment goods and demand for auction winning price increases.

■ keyword : | Art Price | Machine Learning | Non-parametric Prediction | Deep Learning | Multivariate LSTM RNN |

\* 이 논문은 2020년도 중앙대학교 연구장학기금 지원에 의한 것임.

접수일자 : 2021년 02월 16일

수정일자 : 2021년 03월 16일

심사완료일 : 2021년 03월 16일

교신저자 : 송정석, e-mail : jssong@cau.ac.kr

## I. 서론

최근 하나의 미술품에 대해 소유권을 나누어 가지는 미술품 공동구매가 활성화되고 있다. 이는 미술품에 지분을 투자하는 크라우드 펀딩의 형식으로 투자자들이 미술품 총 가액의 일부를 원하는 만큼 구매하고 플랫폼을 통해 권리증서를 받는 형태로 진행된다[1]. 이와 같은 새로운 미술품 유통방식은 미술품의 미적 효용을 추구하는 소비재의 역할보다 일정한 수익을 기대하는 투자재로서의 역할에 주목하여 나타나는 현상이라고 볼 수 있다.

투자 목적의 경제적 자산으로서 주목받고 있음에도 불구하고 미술품은 주식이나 채권과 같은 금융자산과는 달리 논리적 패턴을 설정하는데 어려움이 존재한다[2]. 이는 미술품이 가지고 있는 이질적인 특성들 때문이다. 미술품은 비교적 높은 거래비용을 수반하고 매각 결정과 매각 사이에 불가피한 지연이 존재한다. 게다가 대체 가능한 상품으로 투자 다각화하여 위험을 줄일 수 있는 금융자산과는 달리 대체재가 존재하지 않으며 보유하고 있는 기간 또한 길이 유동성이 낮다. 하물며 미술품 가치 결정 요인에는 작품 크기, 작가, 재료 등과 같은 객관적인 요소뿐만 아니라 취향, 유행 등과 같은 주관적인 요소가 복합적으로 작용하기 때문에 객관적인 정보를 통해 분석이 가능한 금융자산과 비교하였을 때 다소 분석이 불투명하다는 단점이 있다[3].

그럼에도 불구하고 미술품 가격 예측 모형을 통한 객관적 지표 산정의 중요성이 대두되고 있다. 이는 국내 미술품 시장에서 중저가의 미술품이 주로 거래되는 온라인 시장의 활성화와 더불어 주식과 같이 유동성을 높은 미술품 공동구매가 새로이 등장함에 따라 구매 진입장벽이 낮아지고 20-30대의 비교적 낮은 연령층의 투자자들이 증가함으로써 미술품에 대한 수요가 증가하고 있다. 이와 같은 시점에서 객관성을 표명할 수 있는 미술품 가격 분석 모형을 통해 정보의 비대칭성의 문제를 가지고 있는 한국미술품 시장에서 구매자로부터 하여금 미술품 가격에 대한 신뢰성을 확립하는 것은 필수적이다.

기존 미술품가격지수 산출 및 미술품 가격 연구에 있어서 작가 특성, 작품 특성, 시장 특성으로 구분한 변수

들의 유의성을 검증하고자 헤도닉 가격 모형이 활용되어왔다. 주식, 건축, 농산물 가격 예측의 많은 연구에서 인공지능을 사용하는 연구가 활발히 진행되고 있음과 동시에 그 성능 또한 입증받아오고 있는 한편, 딥러닝이나 시계열 분석의 인공지능경망을 통한 미술품 가격 예측에 관한 연구는 드물다.

규칙적인 수요, 공급이 이루어지지 않는 유일성과 예술성이라는 이질적 성격의 미술품도 인공지능을 활용하여 그 가격을 예측할 수 있을까? 따라서 본 연구에서는 낙찰총액 순위 10위 이내의 한국 작가의 회화 작품 2045개를 대상으로 경매 데이터를 구조화하고, 딥러닝의 한 종류인 LSTM(Long Short Term Memory)을 이용하여 낙찰가를 예측해보고자 한다. 작가별 예측을 통해 딥러닝을 통한 가격의 예측력이 높은 작가와 그렇지 못한 작가를 파악하고 그 원인을 규명한다.

## II. 선행연구

본 연구는 순환 신경망의 한 종류인 LSTM을 사용하여 미술품의 경매 낙찰가를 예측하고자 한다. 미술품이 가지고 있는 여러 이질적 특성들을 통제하고 미술품 가격지수 및 투자 수익률을 어떻게 분석할 것인가 또 미술품의 투자수익률은 예측 가능한 지에 대한 의문점 제기 및 연구는 국내외에서도 꾸준히 진행되어 오고 있다. 대부분의 연구에서 재화에 잠재되어있는 특성들을 변수로 회귀식을 도출하여 가격을 예측하는 헤도닉 가격 모형을 기반으로 이루어지며 연구마다 그 결과는 상이하게 나타났다.

경제학자인 Baumol은 예측이 어려운 미술품 시장에서 가격 차이를 이용한 미술품 투자에 대해 경고하며 투자 목적의 미술품 시장에 대하여 부정적인 시각을 제시하였다[4]. 또한 당시 국공채의 수익률이 2.5%임에 비해 미술품의 실질적인 수익률은 0.55%로 나타났으며 미술품 가격의 변동 예측하기 어렵다는 점을 강조했다. 반면 Goetzmann은 1716년-1986년, 1850년-1986년, 1900년-1986년으로 기간을 조정하여 주식 및 채권과 미술품 투자수익률을 비교 분석한 결과 1900년부터 1986년의 미술품 투자수익률은 무려 17.5%로 같은

기간 중 4.9%의 주식수익률을 훨씬 능가한 것으로 나타났다[5]. Chanel은 백터자기회귀모형(VAR)을 활용하여 주요 주식지수와 미술품 가격 지수의 인과관계를 분석하였다[3]. 이를 통해 미술품 가격 지수보다 주식지수가 1년 앞선다는 것이 나타났으며 동시에 금융시장의 변화가 미술시장에도 영향을 미친다고 주장하였다. Agnello는 1971년부터 1996에 유찰된 작품을 제외하고 경매에서 실질적인 거래가 이루어진 미국 화가 71명의 작품을 대상으로 분석한 결과, 연간 명목 수익률이 4.2%로 나타났다[6]. 이는 동일 기간의 주식 및 채권 비교하였을 때, 수익률은 낮고 위험은 상대적으로 높은 것으로 나타났다. 동시에 High-end 미술품 투자 수익률은 9.9%로 전체 미술품 투자수익률보다 높은 것으로 나타나 미술품에 투자한다면 구매할 수 있는 최고의 것을 구매하는 것이 유리하다고 밝혔다. 덧붙여 미술품 투자수익률과 주식 채권의 상관 정도는 음의 계수 혹은 낮은 것으로 확인되어 투자 포트폴리오의 다각화가 가능하다고 주장하였다. Campbell은 1980년부터 2006년까지 거래된 작품 데이터를 기반으로 Basquiat에서 Canaletto까지의 서로 다른 장르, 국적을 가진 유명한 작가 100명이 포함된 일반적 미술, 유럽의 18세기까지의 Brueghel부터 Constable 작가를 포함한 명화, 유럽의 인상주의 작가, 근대 작가, 현대 작가로 총 다섯 갈래로 미술 작품을 구분 지어 미술품의 투자 수익률과 표준편차를 토대로 투자 위험성을 도출하였다. 그 결과, 근·현대미술이 각각 7.55%, 9.00%의 높은 수익률을 보였다. 유럽의 인상주의 미술품은 13.12%의 높은 표준편차를 보였다. 이를 바탕으로 다른 투자자산과의 상관관계를 분석하였을 때, 앞서 Agnello의 연구와 동일하게 낮은 것으로 나타났다.[7]

남준우(2011)은 1998년-2008년의 15회 이상의 거래를 한 작가 52명의 한국미술품 경매 데이터를 기반으로 미술품 가격 지수를 산출하고 이를 바탕으로 미술품 투자수익률과 국공채, KOSPI, KOSDAQ 등의 금융자산과 투자수익률을 비교하였다[8]. 미술품 투자수익률은 평균적으로 23.74%의 수익성을 나타내며 가장 높은 투자수익률을 보이지만, 표준편차 또한 가장 높은 것으로 나타났다. 따라서 미술품 투자를 하는 것은 고위험을 동반한 고수익의 시장이라고 분석하였다. 한편

신정원(2019)은 작품의 이미지 분석을 통해 전통동양화, 현대동양화, 변형동양화, 서양화의 네 가지 양식으로 분류한 변수를 추가하여 1988년부터 2018년까지의 46,192점의 회화 작품 데이터를 토대로 단색화지수와 김환기지수를 산출하였다. 단색화지수는 16.0%, 김환기지수는 8.3%로 동일 기간의 각각 5.2%와 3.1%의 투자수익률을 보인 KOSPI와 KOSDAQ과 비교하였을 때, 이를 상회하는 결과가 나타났다[9].

신성호 외 2인(2018)의 연구에서는 쌀, 양파, 대파, 애호박, 시금치 등 농산물 가격 및 기상 데이터, 유가, 전년도 수확량에 대한 17년 동안의 데이터를 대상으로 LSTM 네트워크를 활용하여 가격을 예측하였고 대전, 광주, 서울, 대구, 부산 등의 지역의 실제값과 비교하였을 때 RMSE 값을 약 0.07로 추정하였다[10]. 조유나 외 2인(2017)의 연구에서는 딥러닝을 이용하여 서울시 아파트 4628개에 대해 헤도닉 가격 모형과 딥러닝을 이용하여 주택가격을 예측하였고, 딥러닝 기반 주택가격의 예측 및 부동산 감정 평가로의 확대 적용 가능성을 제시하였다[11]. 이우창 외 3인(2020)의 연구에서는 다층신경망, 순환신경망, LSTM 등의 머신러닝을 이용하여 철광석 가격을 예측하였고, 시차분포와 LSTM의 앙상블 모형이 예측 오차가 가장 낮은 것으로 추정하였다[12].

### III. LSTM의 이론적 배경

1956년 인공지능(Artificial Intelligence:AI)이라는 용어가 탄생한 이래로 기계학습(Machine Learning; ML)에 대한 기존의 연구는 패턴 인식, 분류, 회귀 등 다양한 분야에서 널리 활용되어왔다. 대부분의 선형 문제를 인공지능이 손쉽게 해결할 수 있게 됨에 따라 1974년부터 1980년까지 AI의 1차 겨울이 찾아왔다. 그 후 특정 문제들에 대해 일반적인 규칙을 컴퓨터에게 학습시키고 조건문을 통해 구현함으로써 실행되는 전문가 시스템(Expert System)이 개발되었으나 규칙이 불분명한 것들을 수학적으로 표현하기 어려운 분명한 관계가 존재하였다. 1987년부터 1993년까지 이어진 AI의 2차 겨울에서는 추정모형에 대한 과적합의 기술적인

문제로 인해 그 성능의 의심이 제기되었다. 최근 컴퓨터 하드웨어 및 소프트웨어, 즉 반도체 기술의 집적도 증가, 컴퓨터가 실행할 수 있는 기계어 명령어 집합의 효율화 등의 급격한 발전에 따라 제한된 시·공간복잡도 내에서 많은 계산량을 컴퓨터가 효율적으로 처리할 수 있게 되었다. 이에 따라 딥러닝이 다양한 분야에 융합, 적용되면서 새로운 연구의 지평을 열고 있으며 현재, 학습을 위한 빅데이터 축적의 필요성이 증가하고 있다.

1986년 Rumelhart가 Nature지에 발표한 논문 'Learning Representations by Back-Propagating Errors'에서는 은닉층(Hidden Layer)에 있는 뉴런을 모사한 단위체들의 네트워크를 거꾸로 추정해가면서 학습을 하는 방식을 제안하였다[13]. 이는 네트워크의 연결 사이의 가중치를 반복적으로 조정하면서 실제 출력과 예측 출력의 차이를 최소화하는 방식으로 동작하는 것이다. 순환신경망(Recurrent Neural Network; RNN)은 Rumelhart의 아이디어를 여러 레이어에 대해 증첩적으로 적용하면서 개발되었다. 아래의 그림은 RNN의 구조를 시각화한 것이다.

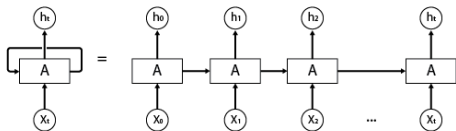


그림 1. RNN의 구조[14]

RNN의 한 종류인 LSTM은 Vanishing Gradient Problem을 해결하기 위해 제안된 것으로 망각 게이트(Forget Gate)를 가지는 것이 구조적으로 가장 큰 특징이다. 기존의 RNN은 Back-Propagating 과정에서 기울기 값이 급격하게 감소하는 문제로 먼 과거에 있었던 일로부터 현재 상태를 학습하는 것이 어려웠다. LSTM은 이러한 문제를 해결할 수 있었기 때문에 문맥을 파악하는 자연어 처리(Natural Language Processing), 음성인식과 같은 기존에 해결하기 어려웠던 문제가 LSTM을 통해 해결되었고, 딥러닝 분야에서 특히 큰 주목을 받게 되었다. LSTM의 구조와 RNN의 구조는 아래의 [그림 2]와 같이 비교해볼 수 있다.

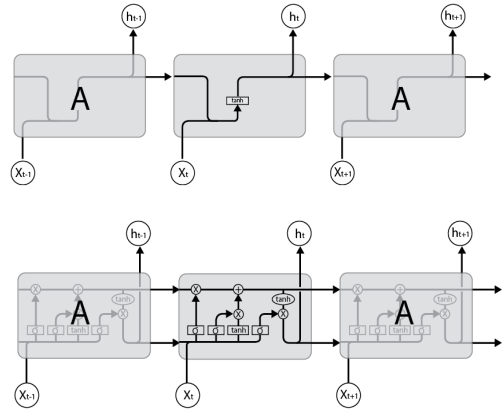


그림 2. RNN과 LSTM 네트워크 Cell의 구조 비교[14]

LSTM 네트워크를 구성하는 Cell의 구조에서 Forget Gate는 과거의 정보를 잊기 위한 게이트로, 값이 0과 1 사이인 Sigmoid 함수를 거쳐 나온 출력이 0이면 모두 잊고, 1이면 모두 기억하는 방식으로 동작 된다. Input Gate는 현재의 정보를 기억하기 위한 게이트로 앞서 Forget Gate에서 사용되었던 Sigmoid와 Hyperbolic Tangent 함수를 거치게 된다. Update Gate는 기존의 Cell State 값에 현재 State에서 업데이트된 값을 더해서 계산된다. 값을 출력하는 Output Gate는 Sigmoid 함수와 Hyperbolic Tangent 함수를 거친 결과가 서로 곱해서 결과가 나오게 된다.

Forget Gate

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

Input Gate

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

Update Gate

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (3)$$

1) Sigmoid 함수는 S자와 유사한 모양의 형태를 가진 함수로, 각 노드에서 임계값을 넘을 때만 출력을 할 수 있도록 하는 활성화 함수의 기능을 한다.

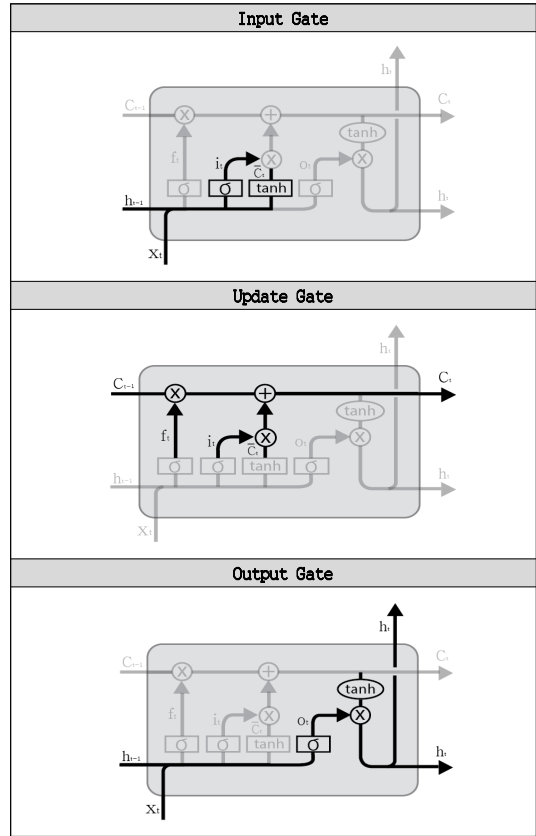
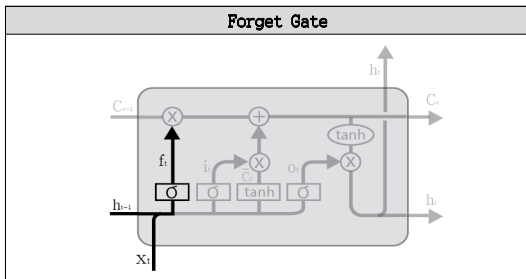
Output Gate

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (4)$$

$$h_t = o_t * \tanh(C_t)$$

LSTM 순환신경망의 각 내부 게이트에서 동작하는 식은 위에 제시한 (1)부터 (4)와 같다. W는 가중치(weight), h는 은닉층(hidden state), b는 편향(bias), x는 입력값, C는 셀의 상태를 의미한다. 미술품 가격 추정모형은 과거의 경매를 통해 획득한 데이터를 바탕으로 앞으로 낙찰될 작품의 가격을 예측하는 것이다. 따라서 과거 정보와 새롭게 입력된 정보를 은닉층에 얼마나 저장할 것인가의 세부 설정들이 알고리즘을 통해 자동으로 업데이트되는 장점이 있어 본 연구에서는 LSTM 순환신경망을 사용하였다. 미술품 가격 추정모형에 널리 활용되고 있는 헤도닉 가격모형은 특성이 추가되었을 때 동일한 비율로 증가하기 때문에 현실적이지 않을 수 있으며 더미 변수의 처리와 해석에 다소 어려움이 존재하지만, LSTM 순환신경망에서는 가중치와 편향을 계속해서 업데이트하기 때문에 이러한 문제에서 비교적 자유로울 수 있다[14]. 실제 LSTM이 미술품 가격 시계열 데이터에 적용하기 위한 방법으로 특성정보에 낙찰가를 낙찰일 순서와 함께 입력( $x_t$ )으로 넣고, 각 게이트의 함수를 통해 출력하여 예측한 값과 실제 값의 비교를 바탕으로 전달할 셀의 상태( $C_t$ )와 은닉층의 값( $h_t$ ), 출력 게이트의 결과( $o_t$ )를 업데이트하게 된다. 이러한 과정이 Batch마다 반복되면서 훈련 데이터 세트 모두에 대해 이루어진다. 검증을 위한 단계에서는 훈련된 셀의 상태와 은닉층 값을 토대로 낙찰가를 제외한 정보를 입력으로 주었을 때 예측된 낙찰가를 얻을 수 있게 된다.

표 1. LSTM 내부 게이트의 구조 및 동작 원리 [14]



#### IV. 연구 설계

기존의 미술품 가격을 다룬 연구에서 가격의 예측을 위해 사용된 변수들은 작가의 생존 여부, 전시 횟수나 수상 경력 등과 같은 화가의 개인적 특성, 작품의 크기, 재료, 재질 등과 같은 작품의 물리적 특성, 경매를 주관하는 회사나 경매 일자 등과 같은 판매적 특성이 있었다. (서병덕 외 1인, 2012) 본 연구의 데이터 세트는 국내 작가의 낙찰총액 1위부터 10위까지인 김환기, 이우환, 박서보, 정상화, 박수근, 천경자, 김창열, 윤형근, 이중섭, 이대원 작가를 대상으로 하였다. 전처리하기 전의 데이터는 작품명, 작가명, 제작연도, 규격, 장르, 기법, 출품처, 거래일, 낙찰가격 등이었다. 학습에 사용하기 위해 가공된 데이터 세트의 변수는 다음의 표와 같다.

표 2. 전처리 후 데이터

변수명		데이터 값
작가 특성	작가	1~10 사이 자연수
	크기	실수
작품의 물리적 특성 (바탕재)	목탄	0/1
	신문지	0/1
	종이	0/1
	직물	0/1
	캔버스	0/1
	하드보드	0/1
	한지	0/1
	도자기	0/1
	세라믹	0/1
	테라코타	0/1
	메소나이트	0/1
	나무	0/1
	낙엽	0/1
	은지	0/1
	작품의 물리적 특성 (재료)	과슈
마커		0/1
매직		0/1
사인펜		0/1
색연필		0/1
수묵		0/1
수묵담채		0/1
수채		0/1
연필		0/1
유화		0/1
잉크		0/1
채색		0/1
펜		0/1
혼합재료		0/1
안료		0/1
그라파이트		0/1
아크릴릭		0/1
콩테		0/1
흑연		0/1
클라주		0/1
먹		0/1
수묵채색		0/1
새김		0/1
크레파스		0/1
에나멜		0/1
오일파스텔	0/1	
판매적 특성 (경매회사)	서울 옥션	0/1
	K 옥션	0/1
	아트데이 옥션	0/1
	칸 옥션	0/1
	아이 옥션	0/1
	에이 옥션	0/1
	마이아트 옥션	0/1
판매적 특성 (경매방식)	온라인	0/1
	오프라인	0/1
	해외	0/1
	거래일	정수
중속변수	낙찰가	정수

사용된 변수의 개수는 총 54개이며 학습 모형에 입력으로 사용되는 데이터 값의 종류가 매우 다양하기에 정규화하여 사용하였다. 알고리즘의 구현 과정에서 min-max scalar를 사용하여 정규화하였으며 그 이는 다음의 식(5)와 같다. 식(5)에서의 x는 학습에 사용하기 위해 가공된 데이터 세트의 모든 변수이다.

$$f(x) = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (5)$$

본 연구에서는 시계열 데이터 분석에 적합한 LSTM 모형을 제시하여 미술품 가격을 활용한 실증분석을 하고자 한다, 따라서 입력변수에 경매 낙찰일을 포함하였고 사용된 모든 데이터는 낙찰일 순서를 기준으로 한 순차 데이터를 사용하였다. 미술품의 특성 데이터인 작품의 물리적 특성, 작가 특성, 판매적 특성을 LSTM 모형의 입력으로 동시에 제시해, 단일 변수가 아닌 다변수로 확장하였기에 그 의미가 있다.

### V. 실증분석 결과

본 연구에서는 Python 프로그래밍 언어를 Jupyter notebook IDE 환경에서 Tensorflow와 Keras 라이브러리를 import함으로써 기계학습이 가능한 환경을 구축하였다. 기계학습의 예측력을 향상시키기 위해서는 초모수(Hyper-parameter) 값의 적절한 설정이 매우 중요하다. 이에 따라 학습 과정 중, 한 번의 반복 동안 전체 데이터 세트를 부분 데이터 세트로 나누어 여러 번 학습하는 방법을 채택하였으며 Batch 크기를 6으로 통제하였다. 또한, 학습 과정에서는 높은 정확성을 보이지만 일반적인 입력에 대해서는 예측력이 떨어지는 과적합(over-fitting) 문제를 방지하기 위해 0.005의 비율로 설정한 Drop-Out 방식의 알고리즘을 구현하였다. Drop-Out Ratio는 데이터의 훈련과정에서 정규화된 모델을 만들기 위해 네트워크의 일부를 무작위로 생략하는 비율로, 생략된 부분은 학습에 영향을 끼치지 않게 된다. Drop-Out Ratio를 조정함으로써 각 네트워크의 가중치가 동조화되는 현상을 막을 수 있다.

표 3. Hyper-parameters 값

Hyper-parameters	값
Batch Size	6
Learning Rate	0.001
Epoch(Number of Iterations)	1000
Drop-Out Ratio	0.005

기계학습의 예측력을 높이기 위한 또 다른 중요한 요소는 최적화(optimizer)이다. 딥러닝 알고리즘 구현에 널리 활용되며, 목적함수의 최솟값을 찾을 수 있도록 관성 개념을 도입하여 기존의 gradient를 포함하고, 더불어 한 번에 세밀하게 학습하면서도 상황에 따라 그 정도를 결정할 수 있는 Adam Optimizer를 사용하였다. 학습 데이터 세트의 반복 횟수인 Epoch는 Cost Function이 최소화될 수 있도록 하는 1000으로 설정하였다. 또한, 기존의 딥러닝 연구들에서 예측한 값이 실제의 값과 차이가 날 때 Cost Function의 그래프에서 얼마나 움직일 것인지를 나타내는 Learning Rate를 0.001로 설정하고 있음을 확인하였고 본 연구에서도 이에 맞추어 설정하였다[9].

예측 모델의 성능인 정확도를 파악하기 위해 실제 낙찰가와 예측된 낙찰가에 대하여 RMSE(Root Mean Squared Error)를 평가 도구로 활용하였다. 실제 낙찰가와 예측된 낙찰가의 차이가 음수일 수도 있기에 단순 오차보다는 RMSE를 사용하는 것이 오차를 정확하게 파악할 수 있다. 본 연구에서 사용된 RMSE의 계산식은 다음과 같다.

$$RMSE = \sqrt{\frac{1}{N} \sum_{k=1}^N (X_{1,k} - X_{2,k})^2} \quad (6)$$

식 (6)에서의  $X_{1,k}$ 와  $X_{2,k}$ 는 각각 실제 낙찰가와 예측 낙찰가를 의미한다.

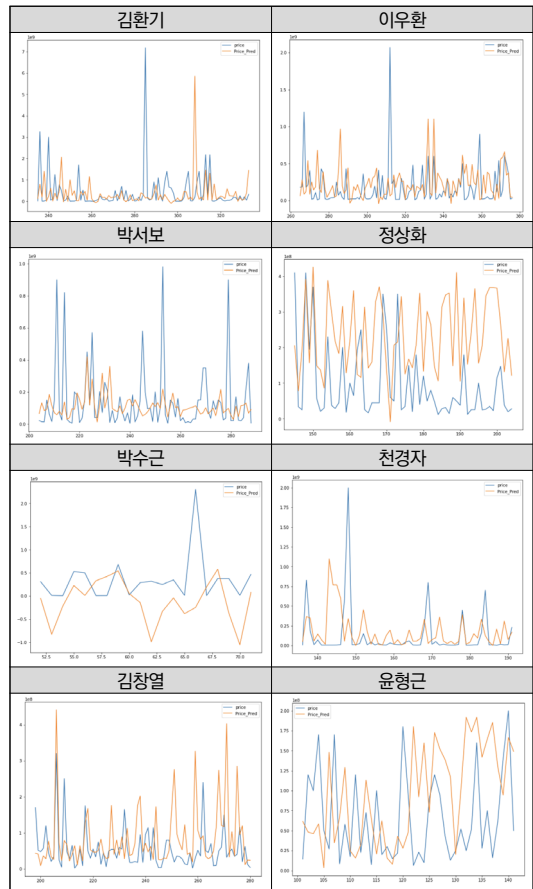
전체 2045개 데이터에서 70%인 1431개를 딥러닝이 학습하는 데에 사용하였고, 30%인 614개를 딥러닝의 성능을 평가하기 위해 테스트하는 데에 사용하였다. 작가별로는 이우환 작가는 총 377개로 가장 많은 데이터를 가지고 있었고, 이종섭 작가는 총 23개로 가장 적은 수의 데이터를 가지고 있었다. 아래의 [표 5]의 파란색 그래프는 실제 경매 낙찰가를 나타내고, 주황색 그래프

는 LSTM을 통해 예측한 낙찰가를 나타내는 척도이다.

표 4. training 및 testing data set의 개수와 variable 수

category	Training	Testing	total	Effective Features
김환기	233	101	334	31
이우환	263	114	377	31
박서보	202	87	289	21
정상화	143	62	205	18
박수근	50	22	72	24
천경자	134	58	192	26
김창열	197	84	281	25
윤형근	99	43	142	14
이종섭	16	7	23	18
이대원	91	39	130	19
전체	1431	614	2045	54

표 5. 작가별 LSTM 예측 결과 그래프



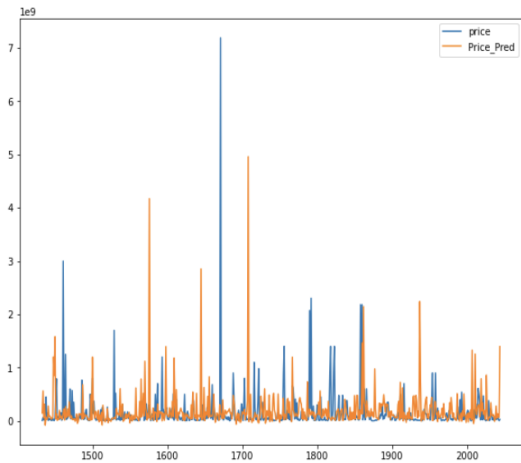
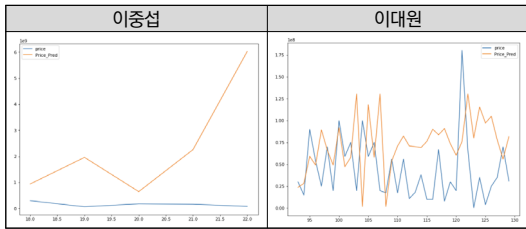


그림 3. 전체 데이터의 LSTM 예측 결과

표 6. 작가별 LSTM 예측 정확도 RMSE값

Artist Name	RMSE Value
김환기	0.129791
이우환	0.160618
박서보	0.186625
정상화	0.171872
박수근	0.343381
천경자	0.183158
김창열	0.181730
윤형근	0.187852
이중섭	0.634627
이대원	0.116104
Total	0.063686

LSTM 순환신경망을 활용한 실증분석 결과 RMSE 값은 약 0.064로 나타났다. 작가별 예측력을 비교하였을 때 이대원 작가는 약 0.116으로 가장 우수한 예측력을 보였으며 이중섭 작가는 약 0.635로 예측력이 가장 낮았다. 본 연구에서 다룬 이대원 작가의 데이터를 살펴보면 전체 130점 회화 작품 중에서 유화가 총 124개

로 압도적인 비율을 보였으며, 캔버스가 총 125점으로 가장 많았다. 따라서 동일한 바탕재와 재료가 사용된 작품들이 많았기 때문에 작품 낙찰가의 예측력이 높은 것으로 추론할 수 있다. 이중섭 작가는 총 23점 작품 중에서 16점의 작품으로 학습을 하고 7점의 작품으로 성능을 검증하였기에 딥러닝의 입력개수가 현저히 부족하여 RMSE 값이 높아 예측력이 낮은 결과가 나왔다고 추론할 수 있다.

## VI. 결론 및 시사점

본 연구는 규칙적인 수요와 공급이 이루어지지 않고 예술적 가치를 평가하는 데에 어려움이 있는 이질적 특징을 가진 미술품에 대하여 LSTM 순환신경망 딥러닝 모형을 활용한 가격 예측모형을 제시하였다. 낙찰총액 순위 10위의 한국 작가의 회화 작품을 대상으로 낙찰가를 예측하고 작가별 분석을 통해 예측력이 높은 작가와 낮은 작가를 확인하고 그 원인을 분석하였다.

작가의 특성변수와 작품의 물리적 특성변수, 판매적 특성변수를 입력변수로 활용하여 Learning Rate, Batch Size, Epoch 수, Drop-Out Value 등의 초모수 값을 설정하고 Adam Optimizer를 활용한 LSTM 순환신경망을 사용하여 미술품의 낙찰가격을 예측한 결과 모형의 RMSE는 약 0.064로 추산되었다. 이대원 작가가 예측력이 가장 높았으며, 반면 이중섭 작가의 예측력이 상대적으로 낮은 것으로 나타났다.

본 연구에서는 딥러닝을 활용하였음에도 불구하고 예측력이 높지는 않아 실제 미술품 경매 환경에 즉각적으로 적용되어 활용되기 어렵다는 한계점이 있지만, 이는 향후 미술품 경매 데이터가 누적될수록 그 예측력이 높아질 것으로 판단된다. 그러나 LSTM 순환신경망을 활용한 미술품 가격 예측모형은 미술품 가격 예측 분야에 널리 적용되지 않은 인공지능을 활용하여 예측해보았다는 점에 그 의미가 있다. 더 나아가 미술품을 미적 효용을 지닌 소비재가 아닌 투자재로서 인식할 때 투자 여부 및 가치 판단에 있어 보다 선명한 기준을 제시하는 도구로 활용될 수 있을 것을 기대한다.



참고 문헌

[1] <https://www.hankyung.com/life/article/2021011435531>, 2020.12.6.

[2] J. Mei and M. Moses, "Art as an investment and the underperformance of masterpieces," *American Economic Review*, Vol.92, No.5, pp.1656-1688, 2002.

[3] O. Chanel, "Is art market behaviour predictable?," *European Economic Review*, Vol.39, pp.519-527, 1995.

[4] W. Baumol, "Unnatural Value: or art investment as floating crap game," *American Economic Review*, Vol.76, pp.10-14, 1986.

[5] W. Goetzmann, "Accounting for Taste: Art and the Financial Markets Over Three Centuries," *American Economic Review*, Vol.83, No.5, pp.1370-1376, 1993.

[6] T. J. Agnello, "Investment Returns and Risk for the Art: Evidence from auctions of American Paintings," *Eastern Economic Journal*, Vol.28, No.4, pp.443-463, 2002.

[7] R. Campbell, "Art as a financial investment," *The Journal of Alternative Investments*, Vol.10, No.4, pp.64-81, 2008.

[8] 남준우, "미술품 가격 결정 요인과 투자 수익률 분석," *경제학연구*, 제59권, 제1호, pp.5-24, 2011.

[9] 신정원, "단색화 열풍은 존재하는가? - 가격지수를 활용한 수익률 비교 분석," *문화경제연구*, 제22권, 제2호, pp.57-79, 2019.

[10] 신성호, 이미경, 송사광, "LSTM 네트워크를 활용한 농산물 가격 예측 모델," *한국콘텐츠학회논문지*, 제18권, 제11호, pp.416-429, 2018.

[11] 조유나, 김수현, 송규원, "헤도닉 가격 모형과 딥러닝을 이용한 주택가격예측 비교: 서울시 아파트를 중심으로," *한국컴퓨터종합학술대회 논문집*, pp.1890-1892, 2017.

[12] 이우창, 김양석, 김정민, 이충권, "머신러닝을 이용한 철광석 가격 예측에 대한 연구," *한국산업정보학회논문지*, 제25권, 제2호, pp.57-52, 2020.

[13] D. Rumelhart, G. Hinton, and R. Williams, "Learning Representations by Back-Propagating Errors," *Nature*, Vol.323, pp.533-536, 1986.

[14] <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>, 2021.1.3.

저자 소개

이지인(Jiin Lee)

준회원



- 2020년 2월 : 중앙대학교 예술대학 한국화전공 (예술학사)
- 2020년 3월 ~ 현재 : 중앙대학교 일반대학원 문화예술경영학과 석박사통합과정

<관심분야> : 미술품 가격 모형, 인공지능

송정석(Jeongseok Song)

정회원



- 2004년 : 미국 미시간주립대학교 (Michigan State University) 경제학 박사
- 2004년 ~ 2006년 : 한국경제연구원
- 2006년 ~ 현재 : 중앙대학교 경영경제대학 경제학부 교수

<관심분야> : 응용 계량경제학, 국제경제학, 미술품 가격 모형