

전자입찰에서 딥러닝을 이용한 입찰 가격예측

황대현* · 배영철**

The Prediction of Bidding Price using Deep Learning in the Electronic Bidding

Dae-Hyeon Hwang* · Young-Chul Bae**

요 약

입찰프로그램은 민/관으로부터 고지되는 입찰 정보의 수집과 누적된 입찰결과의 통계적 분석방법을 사용하고 있지만 복수예가 추첨을 통한 낙찰방식으로 정확한 낙찰가를 예측하는 것은 쉽지 않다. 따라서 본 논문은 MLP, RNN 등의 방법을 이용하여 전자입찰 사이트인 전기넷에서 취득한 2015년 1월부터 2019년 8월까지 전기공사 낙찰현황 데이터의 정확도 등을 분석하고, 이를 통해 낙찰 하한가에 가장 근접하고 1순위 금액 사이의 금액을 예측하여 낙찰에 필요한 입찰금액을 예측하기 위한 기법을 제안한다.

ABSTRACT

The bidding program uses statistical analysis method of the collected bidding information and the accumulated bidding results from the public/private sector; however, it is not easy to predict the accurate bidding price by winning the bid method through multiple lottery. Therefore, this paper analyzes the accuracy of the current state data of the electric construction bid from January 2015 to August 2019 acquired from the electric net, which is an electronic bidding site. We use MLP and RNN method, and proposes a technique to predict the bidding amount necessary for the winning bid by predicting the amount between the first and the lowest bidder.

키워드

Deep learning, Electronic bidding, Machine Learning, MLP, RNN
전자 입찰, 딥러닝, 머신 러닝, MLP, RNN

1. 서 론

현재 우리나라에서는 입찰방법 간소화 및 입찰부정을 방지하기 위해 나라장터(G2B 시스템)라는 전자입찰시스템을 도입하여 입찰을 전자적으로 처리하여 낙찰자를 선정하는 제도 즉 전자입찰이 이루어지고 있

으며 가장 많이 사용되는 전자입찰 방식은 복수 예가 추첨[1]을 통한 낙찰방식이 사용되고 있다. 기존의 여러 전자입찰업체에서 사용하고 있는 방법으로 민/관으로부터 고지되는 입찰 정보의 수집과 누적된 입찰 결과의 구간을 통계적으로 이용한 예측방법으로 고빈도 낙찰구간 선정법, 경쟁률 낮은 구간 선정법, 두 가

*전남대학교 전기 및 반도체공학과(vervain0916@gmail.com)

**교신저자 : 전남대학교 전기·전자통신·컴퓨터공학부
·접 수 일 : 2019. 11. 20
·수정완료일 : 2020. 01. 02
·게재확정일 : 2020. 02. 15

· Received : Nov. 20, 2019, Revised : Jan. 02, 2020, Accepted : Feb. 15, 2020

· Corresponding Author : Young-Chul Bae

Division of Electrical · Electronic communication and Computer Eng. Chonnam National University,

Email : ycbae@chonnam.ac.kr

지를 복잡한 선정방법¹⁾, 전문가에 의해 예측된 입찰 금액 분포도 이용방법²⁾, 예정가격 사정률의 구간확률을 이용한 방법³⁾ 등이 있다. 이는 입찰결과로부터 낙찰금액의 분포도, 빈도수 등 복수의 통계 분석 방법을 적용해 입찰금액을 예측한다. 하지만 발주처에서 사용되는 낙찰가 추첨방식²⁾(난수 체계)에 대한 정보가 부족하다. 이로 인해 낙찰가의 정확한 추측이 어렵고 정확한 입찰가를 예측하기가 쉽지 않은 상황이다.

따라서 본 논문은 발주처로부터 주어지는 낙찰현황 데이터를 이용하여 딥러닝 기법인 다층퍼셉트론(MLP : Multi Layer Perceptron), 재귀 신경망(RNN : Recurrent Neural Networks) 알고리즘을 사용해 데이터를 수치적 방법으로 분석하여 입찰가격을 예측하는 모델을 모색하고 낙찰 하한가에 근접하고 1순위 낙찰 금액보다 낮은 금액 예측을 통하여 낙찰에 필요한 입찰(투찰)금액을 예측하는 기법을 제안한다.

II. 신경망

2.1 다층퍼셉트론(MLP)

MLP³⁾[4-6]는 퍼셉트론으로 이루어진 층(layer) 여러 개를 순차적으로 붙여놓은 형태로 정방향(feed forward) 인공신경망이라고도 불린다.

MLP는 그림 1과 같이 한 개의 입력층(Input layer), 한 개 이상의 은닉층(Hidden layer), 그리고 한 개의 출력층(Output layer)으로 구성된다. 각각의 층은 처리요소인 노드(Node)들로 구성되며 이들 노드는 연결 가중치(Weight)가 부여된 선으로 연결된다. 개별 노드(Node)는 이전 단계의 출력값을 다시 입력값으로 받아 활성화 함수를 거쳐 출력값을 산출한다. 수학적으로 식(1)과 같이 쓸 수 있다.

$$y = \varphi\left(\sum_{i=1}^n w_i x_i + b\right) = \varphi(W^T X + b) \quad (1)$$

W는 가중치, X는 입력, b는 bias, φ 는 활성화 함수이다. 그림 1은 다층퍼셉트론의 일반적인 구조를 나타낸 것이다.

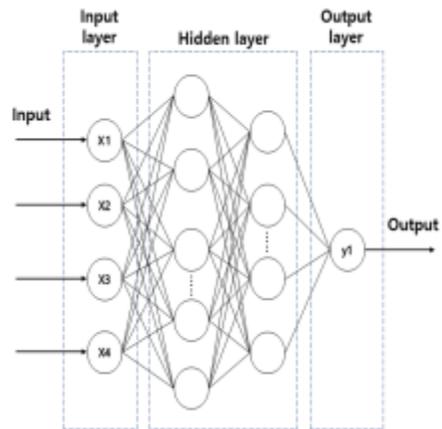


그림 1. 다층퍼셉트론(MLP)의 구조
Fig. 1 Structure of multi layer perceptron

2.2 재귀 신경망(RNN)

RNN[7-10]은 재귀 신경망 또는 순환신경망으로 불리며 히든 노드가 방향을 가진 엣지로 연결된 순환구조를 이루는(directed cycle) 인공신경망의 한 종류다.

재귀 신경망은 기본적으로 은닉층 한 개와 출력층으로 구성된다. RNN 구조에서 일어나는 계산식은 식(2), 식(3), 식(4)과 같다.

$$f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f) \quad (2)$$

$$h_t = f(U x_t + W h_{t-1}) \quad (3)$$

$$y_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o) \quad (4)$$

x_t 와 h_t 는 각각 시점 t에서의 입력값이다. h_t 는 시점 t-1의 값과 시점 t의 입력값으로 계산되는 부분이다. y_t 는 출력, W, U, V는 파라미터, σ_g , f는 활성화 함수, b는 bias이다.

그림 2는 재귀 신경망의 구조를 나타낸 것이다. RNN은 외부입력(x_t)과 자신의 이전 상태(h_{t-1})를 입력받아 자신의 상태를 갱신한다.

1) <http://bid114.tistory.com/2>

2) <http://bid114.tistory.com/1>

3) <http://users.ics.aalto.fi/ahonkela/dippa/node41.html>

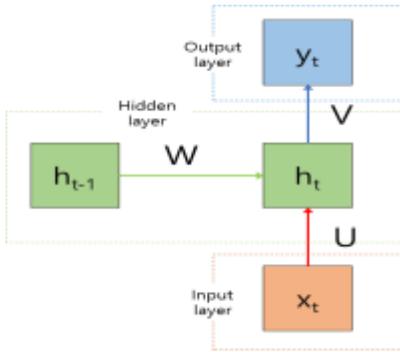


그림 2. 재귀 신경망(RNN)의 구조
Fig. 2 Basic structure of recurrent neural network

RNN의 학습에는 기본적으로 오차 역전파 알고리즘을 사용한다. MLP에서는 오차 정보가 출력층에서 입력층 방향으로 가면서 전달되지만, RNN에서는 현재에서 과거 시점으로 시간을 거스르며 오차 정보가 전달되어 간다.

2.3 데이터 처리 및 분석

본 입찰 프로그램은 기존 전자입찰⁴⁾의 낙찰 데이터를 분석하기 위해 전자입찰업체인 전기넷 사이트⁶⁾에서 2015년 1월부터 2019년 8월까지의 전기공사 낙찰현황 데이터 중 데이터가 빠진 부분을 삭제하고 총 87,319건의 데이터를 4개의 입력 특성(기초금액, 추정가격, 예가 범위, 낙찰하한율), 1개의 출력특성(낙찰금액)으로 사용하고, 훈련 세트 70% 테스트 세트 30%로 나누어 분석하였고 딥러닝 알고리즘인 MLP, RNN 알고리즘을 이용하여 분석 예측하여 최적의 낙찰가를 추정한다.

표 1은 전기넷 사이트에서 취득한 기존의 낙찰현황 데이터의 변수 목록을 표로 나타낸 것이다.

표 1. 취득한 데이터 변수
Table 1. Acquisition data variable

No.	Variable name	Variable type
1	Construction name	categorical
2	Ordering organization	categorical
3	Regional restrictions	categorical
4	Basic amount	continuous
5	Estimated Price	continuous
6	Lower bid rate	continuous
7	Bid date	continuous
8	1 winning companies	categorical
9	Estimated Price Rate	continuous
10	Price range	categorical
11	Successful bid amount	continuous
12	Basic contrast bid rate	continuous

표 2는 입찰공고에서 주어지는 4가지 변수 기초금액, 추정가격, 예가 범위, 낙찰 하한율을 독립변수로 설정하고, 낙찰금액을 종속변수로 설정하여 데이터를 전처리한 것을 나타낸 것이다.

표 2. 분석용 데이터 전처리
Table 2. Data preprocessing for analysis

Basic amount	Estimated Price	Price range	Lower bid rate	Successful bid amount
53338000	48489090	3	87.745	46697977
47729000	43389999	3	87.745	42352592
45534735	41395213	2	87.745	39879900
71548699	65044271	2	87.745	62575931
587224534	533840485	2	86.745	506631115
99394671	90358791	2	87.745	86688300
29376480	26705890	3	87.745	25624690
26902000	24456363	3	87.745	23472520
~	~	~	~	~
201597000	183269999	3	87.745	175975610
113586000	103259999	3	87.745	99627864
68719000	62471818	3	87.745	60246220
40000000	36363636	2	87.745	35267000
301158000	273779999	3	87.745	261742550

4) KONEPS, <http://www.g2b.go.kr>

5) http://www.info21c.net/v4/bbsdata/chobo/info21c_bid_study.pdf, pp. 4.

6) <http://www.jungi.net/nbid>

표 2의 특성 데이터를 Standardization(표준정규분포 표준화)으로 표준화하였다. 표준분포 표준화란 데이터 값에서 평균을 빼고 표준편차로 나눈 것으로 평균이 0, 표준편차가 1인 정규분포를 말한다.

MLP 알고리즘을 이용한 예측과 RNN 알고리즘을 이용한 예측을 진행하였고 그 결과를 그래프로 나타내었다.

그림 3은 4개의 입력특성(기초금액, 추정가격, 예가범위, 낙찰하한율), 1개의 출력특성(낙찰금액)으로 데이터를 MLP 알고리즘으로 이용하여 예측한 예측값과 관측값(test data 30%)을 나타낸 그래프이다.

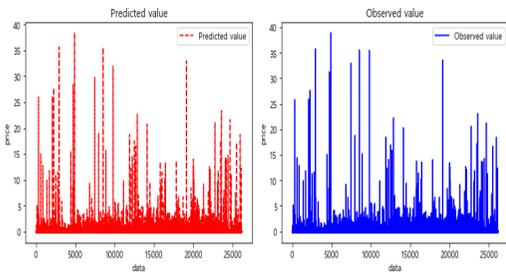


그림 3. MLP에서 예측값과 관측값 그래프
Fig. 3 Predicted value and Observation value graph in MLP (dot-line : Predicted value, line : Observation value)

그림 4는 그림 3과 그림 4의 예측값과 관측값을 합쳐서 나타낸 그래프이다. 그래프를 보면 오차로 인한 차이를 볼 수 있다.

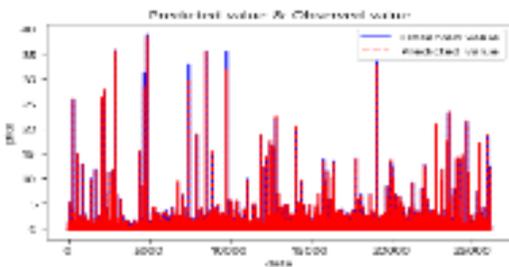


그림 4. MLP에서 예측값과 관측값 조합 그래프
Fig. 4 Predicted value and Observation value combined graph in MLP (dot-line : Predicted value, line : Observation value)

그림 5는 4개 입력특성, 1개의 출력특성 데이터를 RNN 알고리즘을 이용하여 예측한 예측값과 관측값 (test data 30%)을 나타낸 그래프이다.

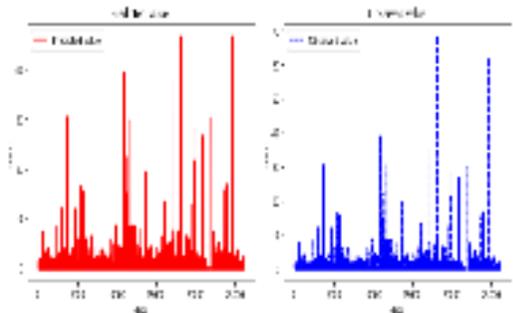


그림 5. RNN에서 예측값과 관측값 그래프
Fig. 5 Predicted value and Observation value graph in RNN (line : Predicted value, dot-line : Observation value)

그림 6은 그림 5의 예측값과 관측값을 합쳐서 나타낸 그래프이다. 그래프를 보면 오차로 인한 차이를 볼 수 있다.

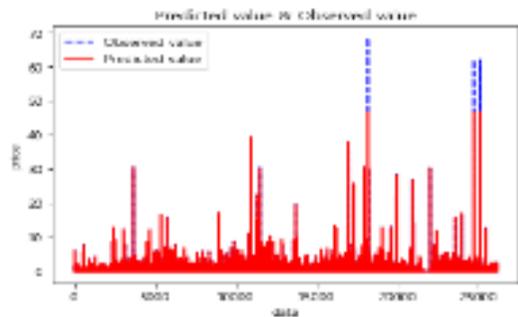


그림 6. RNN에서 예측값과 관측값 조합 그래프
Fig. 6 Predicted value and Observation value combined graph in RNN (line : Predicted value, dot-line : Observation value)

모델 평가를 위해 제공된 평균 제공 오차(MSE)와 평균 제공근 오차(RMSE), 평균 절대 오차(MAE), 평균 절대 비율 오차(MAPE)를 각각 식(6), 식(7), 식(8), 식(9)으로 표시하고, 이를 이용하였다(7)(8)(9).

7) <https://ldgeao99.wordpress.com/2017/04/01/>

8) <https://otexts.com/fppkr/accuracy.html>

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (6)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (7)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |f_i - y_i| = \frac{1}{n} \sum_{i=1}^n |e_i| \quad (8)$$

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{A_t - F_t}{A_t} \right| \quad (9)$$

MSE, RMSE에서 y_i 는 실제 관측값, \hat{y}_i 는 예측값이다. MAE에서 f_i 는 실제값, y_i 는 예측값, e_i 는 오차이다. MAPE에서 A_t 는 실제값, F_t 는 예측값이다.

표 3. MLP와 RNN의 회귀모델 평가

Table 3. Regression model evaluation of MLP and RNN

	MLP	RNN
MSE	0.0021667685	0.034356266
RMSE	0.04654856	0.18535444
MAE	0.018996092	0.0042970236
MAPE	57.113445	3.804691

표 3과 같이 MSE, RMSE, MAE, MAPE가 나왔으며 이 중 RMSE 값이 적을수록 정확도가 높다. RNN보다 MLP에서 RMSE가 더 작게 나왔다. 이를 통해 MLP의 정확도가 더 높은 것을 확인할 수 있다. 정확도가 더 높게 나타난 MLP 알고리즘을 이용하여 2019년 9월 데이터를 1,660건을 실제 데이터로 하여 예측하였다. 그림 7은 MLP에서 예측값과 실제값을 나타낸 그래프이다.

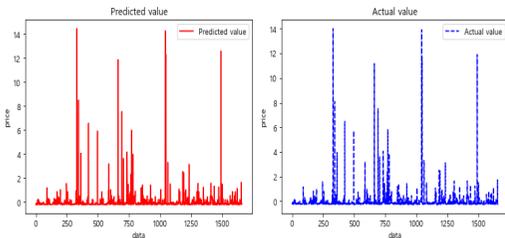


그림 7. MLP에서 예측값과 실제값 그래프
Fig. 7 Predicted value graph in MLP
(line : Predicted value, dot-line : Actual value)

그림 8은 MLP에서 예측값과 실제값을 합쳐서 나타낸 그래프이다. 그래프처럼 오차를 볼 수 있다.

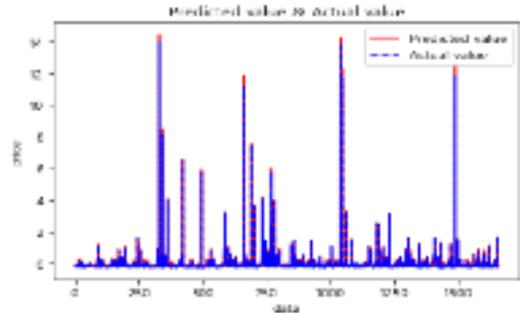


그림 8. MLP에서 예측값과 관측값 조합 그래프
Fig. 8 Predicted value and Actual value combined graph in MLP
(line : Predicted value, dot-line : Actual value)

III. 결론

본 논문에서는 입력특성 4개, 출력특성 1개의 87319건의 수치 데이터를 훈련 및 테스트 데이터를 70% : 30%로 나누어 MLP와 RNN 예측기법으로 분석하고 입찰가격을 예측하였다. MLP와 RNN의 MSE, RMSE, MAE, MAPE의 값은 표 3과 같이 나왔다. 데이터의 오차값이 적을수록 정확도가 높아진다. MLP와 RNN의 그래프 및 평가내용을 보면 MLP에서 MSE, RMSE 오차가 더 적게 나왔고, RNN에서 MAE, MAPE 오차가 더 적게 나왔다. RMSE를 기준으로 봤을 때 MLP의 정확도가 더 높음을 알 수 있었고, MLP 알고리즘으로 2019년 9월 낙찰자료 1,660건의 데이터를 실제 데이터로 하여 훈련 시키고 예측한 결과는 그림 8과 같이 나왔다. 예측값과 실제값이 일치하지는 않지만 유사한 형태의 그래프로 나타남을 확인했다. 예측값 중에는 1순위 낙찰금액과 낙찰 하한을 사이의 값에 든 경우도 존재했다. 전자입찰업체들이 여러 사업체에서 대리 받아 투찰을 하는 것으로 봤을 때 순위 안에 들어간 예측값이 몇 건 되지 않더라도 낙찰의 가능성이 충분히 있는 것으로 보인다.

낙찰 확률을 높이기 위해서 오차를 더욱 줄여야 하

며, 오차를 줄이기 방법으로 역전파 알고리즘 등을 추가로 구성하고, 입찰공고문에서 제공하는 변수만을 이용하여 예측하였으나 공고문에 나오지 않아도 입찰과 관련된 변수를 좀 더 추가한 연구 및 minmax 정규화를 이용한 방법과의 비교 등을 통한 추가적인 연구의 필요성을 느꼈다.

References

- [1] M. Kang, "The effect of multiple reserve price generation range and section setting method on estimated price in bidding and successful bidding system," Master's Thesis, Korea University Graduate School of Management and Information, 2014, pp. 11-15.
- [2] S. Kang, "Analysis Method for Tender Us ing Probability Distribution of Bid Price," Korea Patent, no. 1009441170000, Feb. 18, 2010.
- [3] W. Park, "Bid consulting system for using distribution chart of expert's successful bid prediction and method thereof," Korea Patent, no. 1014482900000, Sept. 30, 2014.
- [4] J. Yoon, "Effectiveness Analysis of Credit Card Default Risk with Deep Learning Neural Network," *Journal of Money & Finance*, vol. 33, Mar. 2019, pp. 161-162.
- [5] J. Choi, "Voiced-Unvoiced-Silence Detection Algorithm using Perceptron Neural Network," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 6, no. 2, 2011, pp. 238.
- [6] J. Choi, "Performance Comparison of Machine Learning in the Prediction for Amount of Power Market," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 14, no. 5, Oct. 2019, pp. 944.
- [7] G. Lee, *Artificial Intelligence: From Turing Test to Deep Learning*. Paju: Life and Power press, 2018, pp. 317-320.
- [8] J. Park, "Study on the Prediction of High and Low Rate of Financial Time Series Using RNN," Master's Thesis, Sookmyung Women's University, Feb. 2019, pp. 9-10.
- [9] Y. Lee and P. Moon, "A Comparison and Analysis of Deep Learning Framework," *J. of the Korea Institute of Electronic*

Communication Sciences, vol. 12, no. 1, Feb. 2017, pp. 117.

- [10] S. Chung and Y. Chung, "Sound Event Detection based on Deep Neural Networks," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 14, no. 2, Apr. 2019, pp. 391-394.

저자 소개



황대현(Dae-Hyeon Hwang)

2010년 전남대학교 응용화학공학부 (공학학사)

2018년~현재 전남대학교 대학원 전기 및 반도체 공학과 석사과정

※ 관심분야 : Python, Tensorflow, Deep learning, Machine learning.



배영철(Young-Chul Bae)

1984년 광운대학교 전기공학과 (공학사)

1986년 광운대학교 대학원 전기공학과 (공학석사)

1997년 광운대학교 대학원 전기공학과(공학박사)

1986년~1991년 한국전력공사

1991년~1997년 산업기술정보원 책임연구원

1997년~현재 전남대학교 전기·전자통신·컴퓨터공학부 교수

2002년~2002년 Brigham Young University 방문교수

2011년~2011년 University of Utah 방문교수

※ 관심분야 : Chaos Control and Chaos Robot, Robot control etc.