

ARMA모형과 머신러닝기법을 활용한 조피볼락 가격 예측

우경원 · 신용민[†]부경대학교(학생), [†]부경대학교(교수)

Prediction of Rockfish Price Using ARMA Model and Machine Learning Technique

Kyeong-Won WOO · Yong-Min SHIN[†]Pukyong National University(student) · [†]Pukyong National University(professor)

Abstract

The aquaculture industry is being proposed as a way to cope with the declining fishery self-sufficiency rate. However, a problem arises that aquatic products must be sold in a short period of time once produced due to their unique perishability. In an aspect of this, it is very important to predict the price of seafood in advance. A number of econometric models have been developed for this purpose. Recently, artificial neural network models of machine learning techniques are being developed instead of these econometric models. However, there are not many studies using such artificial neural network models in the domestic fishery industry. In this study, MLP (Multi Layer Perceptron) and LSTM (Long Short-Term Memory) among artificial neural network models were compared with ARMA. As a result, MLP showed superior predictive power compared to other models, and this difference was statistically significant as a result of Diebold and Mariano Test, indicating that there is a possibility of introducing artificial neural networks into the domestic fishery industry.

Key words : ARMA(Auto regressive moving average), ANN(Artificial neural network), MLP(Multi layer perceptron), LSTM(Long short-Term memory), Rockfish

I. 서론

국내 1인당 수산물 소비량은 지속적으로 증가하고 있는 반면에 수산물 자급률은 점점 감소하고 있다(E-Country Indicator, 2020). 1인당 수산물 소비량은 2012년 53.8kg에서 2019년 69.8kg까지 증가하였으나, 수산물 자급률은 2012년 76.5%에서 2019년 68.4%로 줄어들어 시장안정을 위해서는 국내 수산물 공급능력의 확보가 중요한 과제가 되고 있다. 그 대응방안으로 양식업의 확대가

추진되고 있는데, 2014년 약 155만 톤이었던 국내 양식 생산량은 2021년 약 240만 톤까지 증가하였다(KOSIS, 2022). 그러나 이러한 양식 생산량의 증가는 대부분 해조류 양식에서 기인되며, 어류 양식 생산량은 정체되어 있다.

양식업은 생산량의 변동성과 품질의 가변성으로 인해 생산자들이 안정적 공급능력 확보에 어려움을 겪고 있다. 이에 대응하기 위해 정부는 생산자들에게 관련 정보를 제공하는 수산업 관측 사업을 시행하고 있으며, 이에 수산물 가격 예측

[†] Corresponding author : 051-629-5313, ymshin@pknu.ac.kr

* 이 논문은 부경대학교 자율창의학술연구비(2021년)에 의하여 연구되었음.

이 주요한 연구 대상이 되고 있다. 양식수산물은 입식이나 시설 설치 후 출하까지 걸리는 기간 즉, 생산의 회임기간이 길며, 생산 이후에도 폐사나 품질 저하 가능성이 높아 장기보관에 한계가 있다. 이로 인해 시장 수급 상황에 따른 탄력적 대응이 어려우며 가격변동성이 높다. 또한 최근에는 연어 등 수입 수산물의 증가로 인해 국내 양식업계의 수지 악화가 가중되고 있다. 이에 생산자들에게 시장의 상황과 보다 정확한 가격 정보를 제공하는 것이 중요해졌다. 이러한 점에서 수산물 가격을 예측하기 위한 여러 계량경제학적 방법이 활용되고 있다.

대표적인 선행연구를 소개하면, 우선 국내 어류양식 대표품목인 넙치의 경우 Ock et al.(2007)에서 ARIMA 모형과 VAR 모형을 통해 분석하였고, Nam and Jung(2017)에서 ARMA 모형을 이용하여 분석하였다. 그 외의 어종에서는 Nam et al.(2012)에서 다중회귀모형, ARIMA모형, VAR모형을 이용하여 굴 가격을 분석하였으며, Han and Park(2018)에서 ARIMA 모형, SARIMA모형, Holt-Winters 지수평활법을 이용하여 오징어 가격을 예측하였고, Kim and Choi(2021)에서 지수평활법, ARIMA 모형으로 김의 위판가격을 분석하였다. 또한 Kim and Nam(2016)에서 ARMA, GARCH, VAR모형을 이용, Jung and Nam(2021)에서 ARMA, ARMAX, VAR모형을 이용하여 분석, 비교하는 등 고등어 및 냉동고등어 어종에서도 가격 예측을 위한 여러 시계열분석이 이루어진 바 있다.

수산업과 유사한 농업에서도 가격 예측에 관한 연구가 많이 이루어지고 있다. Ahn et al.(2002)은 양파, 고추 등 양념채소를 분석하였으며, Kim et al.(2016)은 인삼가격을 ARM, ARXM, ARMG, ARXMG모형을 이용하여 분석하였다.

이처럼 많은 계량경제학 모형들을 이용하여 가격을 예측하고, 우수한 모형을 검증하기 위한 연구가 이루어져 왔다. 최근에는 계량경제학 모형이 아닌 머신러닝을 이용한 분석이 진행되고 있

다. 머신러닝 혹은 딥러닝 기법은 인간의 신경을 모방하여 컴퓨터 스스로 훈련하여 예측하게 만드는 방법으로 인공신경망(ANN, Artificial Neural Network), 순환신경망(RNN, Recurrent Neural Network) 등 다양한 모형이 개발되고 있다. 이러한 신경망 모형은 Bae and Yu(2017), Yoon and Shin(2020) 등에서 부동산가격을 예측하는 데에 자주 이용되고 있으며, 축산업에도 Yang and Park(1999), Kim et al.(2021)에서 돼지고기를 대상으로 신경망 기법을 이용한 가격 예측 연구가 진행되고 있다. 그러나 수산물의 경우 아직 신경망 기법이 많이 이루어지지 않고 있는데, Hwang et al.(2012)에서 고등어의 위판가격을 인공신경망을 이용하여 예측하였고, Song(2021)에서 양식 넙치의 가격을 계량모형과 신경망 모형을 이용하여 비교하는 연구를 진행하였다.

본 연구에서는 어류양식의 대표품목인 조피볼락의 대상으로 가격 예측을 하고자 한다. 조피볼락은 국내 어류양식 중 넙치 다음으로 많이 생산되는 해면어류양식의 대표품목이나 넙치에 비해 관련 연구가 많지 않아 분석대상으로 하였다.

분석 모형은 기존 계량경제학 모형인 ARMA 모형을 비롯하여 머신러닝기법의 인공신경망(ANN, Artificial Neural Network)모형 중 하나인 다층퍼셉트론(MLP, Multi Layer Perceptron)모형과, 순환신경망(RNN, Recurrent Neural Network)모형 중의 하나인 LSTM(Long Short-Term Memory)모형을 이용하고자 한다.

신경망기법은 단위근 검정이나 시차 검정 등 다양한 사전·사후검정이 필요한 계량모형과 달리 컴퓨터가 주어진 데이터를 스스로 학습하여 예측하므로 간편하다는 장점이 있으나, 아직까지 국내 수산업 연구에서 많이 진행되지 않고 있다. 신경망기법은 다양한 모형이 있으나, 그 중 다층퍼셉트론에 비해 처리 범위가 넓은 다층퍼셉트론과 기존 순환신경망을 개선하여 현재도 많이 사용되는 LSTM을 분석에 사용하였다. 이후 평균제곱근오차(Root Mean Square Error, RMSE), 평균절

대오차(Mean Absolute Error, MAE), 평균절대비오차(Mean Absolute Percentage Error, MAPE)를 도출하여 각 모형 간 DM검정으로 기존 계량 모형에 비해 예측력이 어느 정도 되는지 비교하고자 한다.

논문의 구성은 II장에서 각 모형에 사용되는 모형을 서술한 후, III장에서 본 연구를 위한 기초자료에 관한 설명과 분석을 실시한다. 마지막으로 IV장에서 각 모형의 결과와 이들 간 통계적 차이를 검정한다. 최종적으로 이러한 결과를 이용하여 수산물에서의 신경망 모형의 활용 가능성을 제시해보고자 한다.

II. 분석 방법

1. ARMA Model

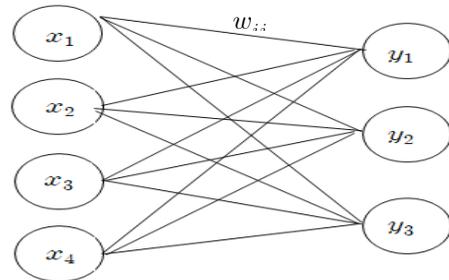
ARMA(Auto Regressive Moving Average)는 AR(Auto Regressive)모형과 MA(Moving Average) 모형이 결합된 모형으로, 종속변수인 특정 시계열값이 p 기 이전의 값과 q 기 이전의 충격에 의해 설명되는 모형이다(Box and Jenkins, 1976). 따라서 AR모형의 시차 p 와 MA모형의 시차 q 를 결정하여야 하는데, 이는 AIC(Akaike Information Criterion) 정보요인이나, HQ(Hannan-Quinn), SC(Schwarz Criterion) 정보요인을 이용하여 결정한다(Chris, 2008). 구체적인 ARMA 모형의 식은 다음과 같다.

$$Y_t = \delta + \rho Y_{t-1} + \rho Y_{t-1} + \dots + \rho Y_{t-p} + \beta \epsilon_{t-1} + \beta \epsilon_{t-2} + \dots + \beta \epsilon_{t-q} + \epsilon_t \quad \dots (1)$$

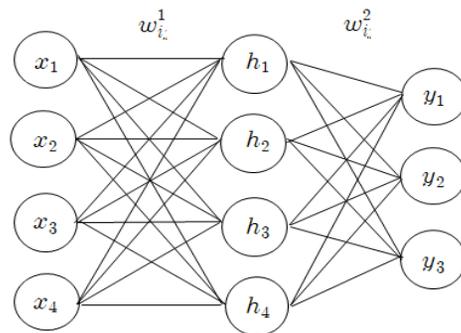
2. 다층 퍼셉트론 인공신경망모형(MLP, Multi Layer Perceptron)

인공신경망모형은 인간의 신경망을 모방하여 W. McCulloch and W. Pitts(1943)에서 논리, 산술, 기호연산 기능을 컴퓨터가 구현할 수 있도록 제시한 것이다. 인공신경망은 일종의 알고리즘인

퍼셉트론을 통해 반복적으로 학습하게 된다(Kim, 2021). 퍼셉트론이 단층이나 다층이나에 따라 단층 퍼셉트론 모형과 다층 퍼셉트론 모형으로 나뉜다. 단층 퍼셉트론은 [Fig. 1]과 같은데, 가중치 w 에 대해 입력층 x 가 활성화함수를 이용하여 출력층 y 를 도출한다. 이때 활성화함수는 시그모이드($\frac{1}{1+e^{-x}}$)함수나 Tanh(하이퍼볼릭 탄젠트) ($\frac{e^x - e^{-x}}{e^x + e^{-x}}$) 등이 이용된다.



[Fig. 1] Single Layer Perceptron.



[Fig. 2] Multi Layer Perceptron.

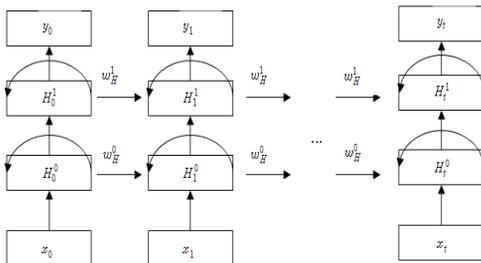
단층 퍼셉트론모형은 선형 모형이기 때문에 XOR(eXclusive OR) 즉, 입력값이 모두 1이면 1, 한쪽만 1이면 0이 되는 문제를 해결할 수 없다. 이는 입력값의 벡터가 (1, 0), (0, 1)인 경우와 (0, 0), (1, 1)을 분리시킬 수 없기 때문이다. 이를 해결하기 위해 은닉층(Hidden Layer)을 추가한 다층 퍼셉트론 모형이 개발되었다. 다층 퍼셉트론 모

형은 [Fig. 2]와 같은데, 은닉층은 입력층에서 받은 데이터를 활성화시켜 출력층으로 보낸다(Kim, 2021).

이때 각 층을 이동할 때 끊임없이 가중치를 업데이트하게 되는데, 이러한 업데이트 과정방식은 순전파와 역전파로 나뉘어진다. 순전파는 입력층에서 은닉층, 은닉층에서 출력층으로 이동하면서 가중치를 업데이트하는 과정이고, 역전파는 반대로 출력층에서부터 입력층으로 이동하며 가중치를 업데이트하는 과정이다.

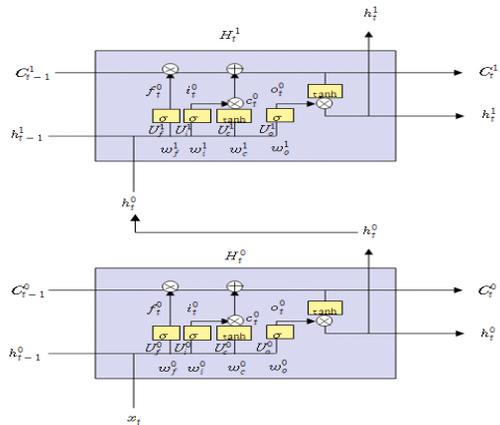
3. LSTM(Long Short - Term Memory)

LSTM모형은 순환신경망(RNN, Recurrent Neural Network)의 한 종류이다. 순환신경망이란 [Fig. 3]과 같이 개별의 다층 퍼셉트론에서 순전파된 데이터가 수평으로도 전달되는 형태이다.



[Fig. 3] Recurrent Neural Network.

이러한 순환신경망은 반복될수록 장기 데이터가 소실되는 문제가 발생한다. 따라서 이를 극복하기 위해 만들어진 모형이 LSTM인데, S. Hochreiter and J. Schmidhuber(1997)에서 개발되었다. LSTM 셀은 [Fig. 4]와 같이 각각의 셀(H_i)들이 수직으로 쌓여있다. 각각의 은닉상태벡터(h^i)는 단기 기억 정보를 처리하고, 셀상태벡터(C^i)는 장기 기억 정보를 처리한다. σ 는 시그모이드 함수를 의미하며, f, i, o, α 는 각각 forget gate vector / input·update gate vector / output gate vector / cell input gate vector이다.



[Fig. 4] LSTM Cell.

4. 예측력 비교 및 검증

본 연구에서는 ARMA, MLP, LSTM모형을 각각 분석한 후 우선 이들의 예측력 지표인 RMSE, MAE, MAPE를 도출한다.

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (y_t - f_t)^2} \dots\dots\dots (2)$$

$$MAE = \frac{1}{T} \sum_{t=1}^T |y_t - f_t| \dots\dots\dots (3)$$

$$MAPE = \frac{1}{T} \sum_{t=1}^T \left| \frac{y_t - f_t}{y_t} \right| \times 100 \dots\dots\dots (4)$$

y_t : 실제값, f_t : 예측값

이후 각 지표를 비교하여, 실제로 모형간 설명력의 차이가 있는지에 대해 DM검정(Diebold and Mariano Test)을 실시한다. DM통계량은 다음과 같다(Jung and Nam, 2021).

$$DM = \left[\frac{1}{\widehat{V}(\bar{d})} \right]^{\frac{1}{2}} \times \bar{d}$$

$$\widehat{V}(\bar{d}) = \frac{1}{T} \left[\hat{\gamma}(0) + \frac{2}{t} \sum_{k=1}^{t-1} ((T-k) \hat{\gamma}(k)) \right] \dots\dots\dots (5)$$

$\widehat{V}(\bar{d})$: \bar{d} 의 분산, d : 예측오차 차이,
 \bar{d} : 예측오차 차이의 평균, $\hat{\gamma}(k)$: k시차간 자기공분산

DM검정의 귀무가설과 대립가설은 다음과 같다.

$$\begin{aligned} H_0 : E[d_t] &= 0 \\ H_1 : E[d_t] &> 0 \end{aligned} \dots\dots\dots (6)$$

Ⅲ. 분석 자료

1. 기초 자료

본 연구에서 사용한 조피볼락의 가격 자료는 한국해양수산개발원 수산업관측센터에서 제공하는 하남수산물시장의 500~600g 조피볼락 도매가격을 2006년 9월부터 2022년 1월까지 185개의 월별자료를 통계청 품목별 생산자물가지수를 이용하여 실질화하였다. 동 자료를 이용한 이유는 500~600g이 다른 중량대(예, 400~500g)에 비해 출하비중이 더 높기 때문이다. 또한 도매가격을 이용한 이유는 월별 산지가격 자료에 결측치가 많아 분석에 어려움이 있었으며, 도매가격이 생산자는 물론 가공업자, 최종판매자 등의 수요자뿐만 아니라 생산에게도 영향을 주기 때문이다.

신경망 모형을 이용한 시계열분석은 데이터를 크게 훈련(Training), 검증(Validation), 테스트(Test)의 세 가지 형태로 나눈다. 훈련데이터는 컴퓨터가 스스로 학습하는 데에 이용되고, 검증데이터는 모형의 성능을 평가하는 데 이용된다. 마지막으로 테스트데이터는 훈련과 검증 이후에 실제로 예측하는 것에 이용된다.

본 연구에서는 185개의 자료 중 첫 60%의 자료를 훈련으로, 이후 20%를 검증으로, 마지막 20%를 테스트로 설정하였는데, 즉 112개의 자료가 훈련, 37개의 자료가 검증, 36개의 자료가 테스트로 사용되었다. 이러한 이유로 ARMA 분석 역시 전체 자료 중 36개의 자료를 예측에 사용하였다.

<Table 1>의 기초통계량을 살펴보면, 평균가격은 약 13,331원이며, 최솟값은 2020년 4월의



Source : KMI Fishery Outlook Center

[Fig. 5] Hanam Wholesaleprice of Rockfish.

5,614원, 최댓값은 2021년 10월의 26,704원으로 나타났다. 도매가격의 추이를 살펴보면 2021년의 가격이 비교적 높은 수준을 가지는 것으로 나타났는데, 이는 코로나19와 같은 충격 발생이 입식에 영향을 주어, 생산까지의 시차가 반영되어 나타난 결과로 추정된다(KOSTAT, 2022).

<Table 1> Hanam Wholesale Price Basic Statistics

| (Unit : Time, won) | | | | | |
|--------------------|-----|----------|----------|--------|----------|
| Var | Obs | Mean | Std. Dev | Min | Max |
| Price | 185 | 13330.89 | 4152.49 | 5614.4 | 26703.67 |

2. 단위근 검정

ARMA모형을 사용하기 위해서는 우선 해당 시계열 변수의 안정성을 검정하여야 한다. 안정적 시계열이란 자료의 평균, 분산 등의 분포가 시간에 따라 변화하지 않는 것을 말하며, 불안정적 시계열은 시간에 따라 변화하는 것을 말한다. 시계열 자료를 이용한 분석은 안정적인 자료를 이용하여야 자기상관이 발생하지 않는다. 따라서 불안정한 자료를 이용한다면 자료를 차분하면 불안정성이 해결될 수 있다. 이러한 안정성을 검정하는 방법은 단위근 검정이며, 그중 ADF검정(Augmented Dickey Fuller Test), PP검정(Phillips-Perron Test)을 수행하였다. 각 검정의 귀무가설은 “단위근이 존재한다”이다.

<Table 2> Unit Root Test

| Variable / t-stat | ADF | PP |
|-------------------|------------------------|------------------------|
| Wholesale Price | -3.36642 (0.0134**) | -3.38639 (0.0127**) |

Note 1 : () means p-value.

ADF검정과 PP검정 모두 5% 유의수준에서 귀무가설이 기각되어 단위근이 존재하지 않는 것으로 나타났다. 만약 단위근이 존재한다면 차분을 하여 자료를 안정적으로 만들어야 하므로, ARMA 대신 차분과정이 포함된 ARIMA모형을 사용하여야 한다. 그러나 원자료가 안정적으로 나타나 ARMA모형을 사용하였다.

3. 시차 검정

ARMA모형의 적절한 AR 시차인 p, MA 시차인 q를 선정하기 위해 AIC(Akaike Information

Criterion), SC(Schwarz Criterion), HQ(Hannan Quinn) 정보요인 기준을 이용하였다. 이때 정보요인 값이 낮을수록 상대적으로 적합하다.

AIC 요인에서는 ARMA(5, 5)가 적합하게 나왔으나, SC와 HQ 요인에서는 ARMA(1, 1)이 적합하게 나타났으므로, 본 연구에서는 ARMA(1, 1)을 선정하였다.

IV. 분석 결과

1. ARMA모형

정보요인에서 선정된 ARMA(1, 1)의 결과는 <Table 4>에 제시되어 있다. 상수항과 AR(1), MA(1) 모두 유의성이 높게 나타났으며, 결정계수 역시 비교적 높게 나타났다.

ARMA모형 추정 후 오차항에 대하여 자기상관검정과 이분산 검정을 수행하여야 한다.

<Table 3> ARMA Time Lag Test Result

| | p/q | 0 | 1 | 2 | 3 | 4 | 5 |
|-----|-----|----------|----------|----------|----------|----------|----------|
| AIC | 0 | 19.33362 | 18.49177 | 18.21961 | 18.04716 | 18.03568 | 18.00562 |
| | 1 | 17.97346 | 17.93429 | 17.94726 | 17.96065 | 17.95491 | 17.96814 |
| | 2 | 17.93996 | 17.9471 | 17.95223 | 17.96451 | 17.96781 | 17.98158 |
| | 3 | 17.94702 | 17.94815 | 17.96135 | 17.97778 | 17.98028 | 17.99311 |
| | 4 | 17.95976 | 17.96148 | 17.97146 | 17.89381 | 17.90812 | 17.98331 |
| | 5 | 17.96713 | 17.97041 | 17.97959 | 17.96434 | 17.9256 | 17.90164 |
| SC | 0 | 19.35379 | 18.53209 | 18.28009 | 18.12780 | 18.13648 | 18.12658 |
| | 1 | 18.01379 | 17.99477 | 18.02791 | 18.06145 | 18.07587 | 18.10927 |
| | 2 | 18.00044 | 18.02774 | 18.05303 | 18.08547 | 18.10894 | 18.14286 |
| | 3 | 18.02766 | 18.04895 | 18.08232 | 18.1189 | 18.14156 | 18.17455 |
| | 4 | 18.06057 | 18.08244 | 18.11258 | 18.05509 | 18.08957 | 18.18492 |
| | 5 | 18.08809 | 18.11153 | 18.14088 | 18.14578 | 18.12721 | 18.12341 |
| HQ | 0 | 19.34182 | 18.50815 | 18.24418 | 18.07992 | 18.07663 | 18.05477 |
| | 1 | 17.98985 | 17.95886 | 17.98003 | 18.0016 | 18.00405 | 18.02548 |
| | 2 | 17.96453 | 17.97987 | 17.99318 | 18.01365 | 18.02515 | 18.0471 |
| | 3 | 17.97978 | 17.9891 | 18.0105 | 18.03511 | 18.04581 | 18.06683 |
| | 4 | 18.00072 | 18.01062 | 18.02879 | 17.95934 | 17.98184 | 18.06522 |
| | 5 | 18.01628 | 18.02774 | 18.04512 | 18.03805 | 18.00751 | 17.99174 |

<Table 4> ARMA(1, 1) Result

| Variable | Coefficient | Std.Error | t-value | R-sq | F-value |
|----------|-------------|-----------|-------------------------|----------|-------------------------|
| C | 13547.88 | 961.0054 | 14.09761 (0.0000***) | 0.762205 | 233.9869 (0.0000***) |
| AR(1) | 0.801770 | 0.056449 | 14.20351 (0.0000***) | | |
| MA(1) | 0.279214 | 0.091325 | 3.057347 (0.0027***) | | |

Note 1 : () means p-value.

Note 2 : * p<0.1, ** p<0.05, *** p<0.0.

<Table 5> Autocorrelation and Heteroscedasticity Test Result

| Model | Breusch - Godfrey LM Test | White Test | ARCH Test |
|------------|---------------------------|----------------------|---------------------|
| ARMA(1, 1) | 1.166542 (0.5581) | 8.115588 (0.5225) | 0.03101 (0.8602) |

Note 1 : () means p-value.

Note 2 : The null hypothesis of the autocorrelation test is <there is no autocorrelation>.

Note 3 : The null hypothesis of the heteroscedasticity test is <the heteroskedasticity variance does not exist>.

이는 오차항에 자기상관과 이분산이 존재한다면 회귀계수 추정량의 분산이 커져 효율성이 낮아져서 기타 통계량들이 잘못 나타나기 때문이다 (Jung and Nam, 2021). 따라서 자기상관을 검정하기 위해 Breusch-Godfrey LM 검정을 수행하였고, 이분산을 검정하기 위해 White 검정과 ARCH 검정을 수행하였다. 이들 검정의 귀무가설은 각각 “자기상관이 존재하지 않는다”와 “이분산이 존재하지 않는다”이다.

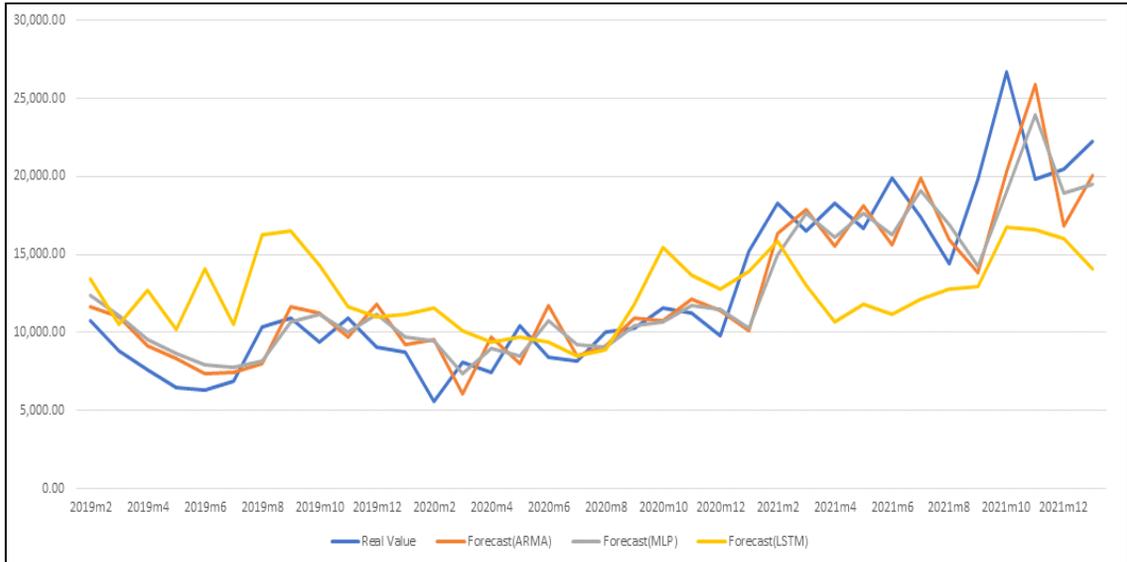
그 결과는 <Table 5>에 나타나 있는데, 모두 통계적 유의성이 존재하지 않아 귀무가설이 기각되지 않는다. 이는 귀무가설에 따라 ARMA(1, 1) 모형에 자기상관과 이분산이 존재하지 않는 것으로 나타났다.

2. NLP 인공신경망모형

신경망모형을 추정하기 위해서는 하이퍼파라미터를 결정해야 한다. 하이퍼파라미터는 신경망의 학습 과정에서 스스로 조정되는 파라미터가 아닌 사용자가 직접 결정해야 하는 모수이다. 이러한

하이퍼파라미터를 선정하는 방법은 따로 정해져 있지 않으며, 사용자가 이를 적절히 조정해가며 모델링을 수행한 후 손실함수값이 낮게 나타나는 모수를 이용하는 것이 일반적이다. 다층 퍼셉트론 모형에서 이는 은닉유닛의 크기와 시간지연, 에포크(epoch)가 있다. 본 연구에서는 NLP 인공신경망을 추정하기 위해 MATLAB의 Deep Learning Toolbox의 nnstart 명령어를 이용하였다.

해당 모형에서 은닉유닛의 크기는 10개에서 다른 크기에 비해 손실함수가 낮게 나타났다. 은닉유닛이 10개, 20개, 100개일 때의 손실함수값은 각각 1,838, 2,085, 1,500으로 100개에서 가장 낮은 손실함수값을 보였다. 그러나 이 경우 연산속도가 오래 걸리는 문제가 있었으며, 또 학습데이터의 RMSE값은 은닉유닛이 10개일 때 가장 낮았기 때문에 은닉유닛을 10개로 설정하였다. 에포크는 1,000으로 설정하였다. 또한 훈련에 사용되는 역전파 방식은 David J. C. and Mackay(1992)와 Foresee et al.(1997)에서 사용된 베이스 정규화 역전파 방식을 이용하였다.



[Fig. 6] RMSE Comparison between Models.

3. LSTM모형

LSTM모형에서도 NLP와 마찬가지로 하이퍼파라미터를 결정하여야 한다. LSTM에서의 하이퍼파라미터는 NLP 모형과 동일하게 은닉유닛의 크기, 에포크가 있으며, 그 외에도 한 번의 학습 데이터 크기인 배치(batch)의 크기, 비용함수 그래프에서 한 번에 이동하는 거리인 학습률(Learning Rate) 등이 있다.

은닉유닛이 20개, 50개, 100개, 200개일 때의 검증 RMSE는 각각 0.46, 0.65, 0.91, 0.36이었다. 따라서 은닉유닛의 크기는 NLP와 다르게 검증 RMSE가 가장 낮은 200개로 설정하였다. 에포크는 1,000회로 설정하였다. 배치의 크기는 50으로 설정하였는데, 배치의 크기가 20, 30, 50개일 때의 검증 RMSE는 각각 0.8, 0.45, 0.39로 나타났다. 배치의 크기가 클수록 모델의 속도가 빨라지나 정확도가 낮아지는 단점이 있다(Ko et al., 2021). 따라서 배치 크기를 50개 이상으로 설정하는 것은 연산속도에서 비효율적이기 때문에 배치를 50개로 하였다. 학습률은 클수록 최소화하는 매개변수를 지나칠 수 있고, 반대로 작아질수록

수행 속도가 느려진다는 단점이 존재한다(Kim et al., 2021). 이러한 학습률의 적정값을 설정하는 것은 어렵기 때문에 초기 학습률을 설정 후 특정 에포크에 따라 감소(Drop)시키는 방법이 수행된다. 따라서 본 연구에서는 초기 학습률을 0.005로 설정하고, 에포크 500회당 학습률을 0.01배 감소하도록 하였다.

마지막으로 LSTM의 옵티마이저(Optimizer)는 최근 가장 많이 사용되는 Adam을 이용하였다. Adam은 딥러닝 초기에 사용되던 기울기 하강 옵티마이저에서 개선된 방식이다(Kim et al., 2021).

4. 예측력 비교

각 모형의 예측 결과와 RMSE, MAE, MAPE는 [Fig. 6]과 <Table 6>에 제시되어 있다.

RMSE, MAE, MAPE 모두 다층 퍼셉트론 인공 신경망모형이 가장 낮게 나타났으며, 기존 계량 모형인 ARMA는 이와 비슷한 값이 도출되었으며, LSTM모형은 가장 높게 나타났다.

특히 MAPE는 일반적으로 $0\% < MAPE < 10\%$ 인 경우 매우 정확한 예측, $10\% < MAPE < 20\%$

ARMA모형과 머신러닝기법을 활용한 조피블락 가격 예측

<Table 6> Forecast results and predictive power indicators

| Time | Real Value | Forecast(ARMA) | Forecast(MLP) | Forecast(LSTM) |
|---------|------------|----------------|---------------|----------------|
| 2019m2 | 10,746 | 11,682 | 12,356 | 13,431 |
| 2019m3 | 8,839 | 11,041 | 11,051 | 10,507 |
| 2019m4 | 7,569 | 9,158 | 9,549 | 12,719 |
| 2019m5 | 6,496 | 8,311 | 8,624 | 10,209 |
| 2019m6 | 6,281 | 7,387 | 7,894 | 14,112 |
| 2019m7 | 6,868 | 7,412 | 7,754 | 10,485 |
| 2019m8 | 10,337 | 8,040 | 8,142 | 16,267 |
| 2019m9 | 10,886 | 11,614 | 10,718 | 16,503 |
| 2019m10 | 9,426 | 11,211 | 11,166 | 14,313 |
| 2019m11 | 10,930 | 9,745 | 9,997 | 11,647 |
| 2019m12 | 9,085 | 11,779 | 11,202 | 10,982 |
| 2020m1 | 8,748 | 9,217 | 9,735 | 11,129 |
| 2020m2 | 5,614 | 9,568 | 9,480 | 11,600 |
| 2020m3 | 8,059 | 6,083 | 7,331 | 10,101 |
| 2020m4 | 7,410 | 9,699 | 8,973 | 9,415 |
| 2020m5 | 10,415 | 7,988 | 8,513 | 9,682 |
| 2020m6 | 8,399 | 11,714 | 10,781 | 9,386 |
| 2020m7 | 8,200 | 8,494 | 9,221 | 8,526 |
| 2020m8 | 10,011 | 9,178 | 9,075 | 8,929 |
| 2020m9 | 10,280 | 10,944 | 10,456 | 11,803 |
| 2020m10 | 11,552 | 10,743 | 10,672 | 15,466 |
| 2020m11 | 11,234 | 12,173 | 11,721 | 13,699 |
| 2020m12 | 9,798 | 11,430 | 11,454 | 12,773 |
| 2021m1 | 15,237 | 10,086 | 10,288 | 13,893 |
| 2021m2 | 18,247 | 16,341 | 14,957 | 15,845 |
| 2021m3 | 16,503 | 17,848 | 17,635 | 13,035 |
| 2021m4 | 18,288 | 15,542 | 16,092 | 10,703 |
| 2021m5 | 16,654 | 18,115 | 17,670 | 11,786 |
| 2021m6 | 19,942 | 15,631 | 16,227 | 11,171 |
| 2021m7 | 17,402 | 19,878 | 19,080 | 12,143 |
| 2021m8 | 14,419 | 15,946 | 16,892 | 12,771 |
| 2021m9 | 19,845 | 13,820 | 14,224 | 12,923 |
| 2021m10 | 26,704 | 20,279 | 18,999 | 16,786 |
| 2021m11 | 19,803 | 25,890 | 23,936 | 16,564 |
| 2021m12 | 20,447 | 16,864 | 18,965 | 16,053 |
| 2022m1 | 22,234 | 20,079 | 19,497 | 14,090 |
| RMSE | - | 2,783 | 2,632 | 4,579 |
| MAE | - | 2,269 | 2,062 | 3,732 |
| MAPE | - | 19 | 17 | 32 |

인 경우 비교적 정확한 예측, 20% < MAPE < 50%인 경우 비교적 합리적 예측, MAPE ≥ 50%인 경우 부정확한 예측으로 판단된다(Choi, 2016).

ARIMA와 MLP는 모두 MAPE가 10%~20% 사이이므로 비교적 정확한 예측으로 나타났으나, LSTM은 약 32.5%로 비교적 합리적 예측으로 나

타났다.

이러한 모형 간 설명력의 차이가 통계적으로도 유의한 것인지를 검정하기 위해 DM검정을 실시하였다.

DM검정에서 기준이 되는 모형은 RMSE, MAE, MAPE 모두 가장 낮았던 MLP모형으로 선정하였고, 구체적인 DM검정 결과는 <Table 7>에 제시되어 있는데, DM검정의 결과 MLP모형과 ARMA, MLP모형과 LSTM 모두 예측력이 동일하다는 귀무가설이 기각되어 MLP모형이 더 우수한 것으로 나타났다.

<Table 7> DM Test Result

| Model | DM Stat |
|-------|---------------------|
| MLP | - |
| ARMA | 2.234 (0.0255**) |
| LSTM | 2.934 (0.0033**) |

Note 1 : () means p-value.

Note 2 : * p<0.1, ** p<0.05, *** p<0.0.

V. 결론

본 연구는 수산물 가격 예측을 위한 신경망모형의 활용가능성에 대한 것이다. 이를 위해 2006년 8월부터 2022년 1월까지의 대표 해면어류 양식종인 조피볼락을 대상으로 가격 예측 분석을 시행하였다.

신경망모형과의 비교 계량모형으로 ARMA모형을 선정하였다. ARMA모형은 지난 시차의 가격만으로 예측하는 단변량 시계열 모형인데, 본 연구에서 사용한 신경망 모형이 과거의 가격을 이용하여 미래를 예측하는 모형이기 때문에, VAR모형과 같은 다변량 시계열이 아닌 단변량 시계열 모형을 선택하였다.

분석 결과, 조피볼락 도매가격의 기초통계량과 월별 추이와 ARMA모형 분석을 위한 단위근 검정인 ADF, PP검정을 수행하여 단위근이 존재하

지 않는다는 결과를 얻었다. 적절한 시차를 선정하기 위해 정보요인인 AIC, SC, HQ 기준을 이용하여 적정시차가 ARMA(1, 1)이라는 결과를 얻었다. 최종적으로 ARMA(1, 1), MLP, LSTM 모형을 이용한 분석 결과, RMSE, MAE, MAPE 모두 MLP모형이 가장 낮게 나타났고, ARMA(1, 1) 모형이 이와 큰 차이를 보이지 않았으며, LSTM 모형이 비교적 높게 나타났다.

이후 이들 모형 간 예측력의 차이가 있는지를 검정하기 위해 DM검정을 수행한 결과, MLP가 다른 모형들에 비해 통계적으로 유의하게 설명력이 우수한 것으로 나타났다. 한 가지 유의해야 하는 것은 신경망모형의 학습과정에서 주어진 학습 자료를 표본으로 나누어 Epoch 수만큼 학습하게 되는데, 이로 인해 동일한 하이퍼파라미터를 두고도 조금씩 다른 예측력을 가진다는 것이다. 물론 이러한 예측력의 차이가 크지는 않지만, 본 연구 결과에서 나타난 MLP와 ARMA모형은 경우에 따라 ARMA모형이 더 우수하게 나타날 수 있다는 것이다.

그럼에도 불구하고 신경망 모형은 단위근 검정, 시차 검정, 이분산 및 자기상관 검정, 그 외 VAR과 같은 다변량 모형에서 수행해야 하는 그랜저 인과성 검정 등을 수행할 필요가 없어 보다 간편하게 예측할 수 있다는 장점이 있다.

따라서 향후 수산물 가격 예측에 이러한 신경망모형을 적극적으로 사용할 수 있을 것으로 기대된다. 신경망모형 역시 계량모형만큼 다양한 모형이 존재한다. 따라서 본 연구에서 사용한 MLP, LSTM 외에도 다양한 모형을 이용하여 비교하면, 향후 수산물 가격 예측의 정확도를 높이는 데 충분히 활용 가능할 것으로 보인다.

양식업 경영체의 합리적 의사결정에 도움을 주기 위한 정보제공기능으로서 가격 예측기법의 고도화가 중요한 과제가 되고 있다. 이런 점에서 신경망모형의 활용은 수산물 가격 예측의 편의성과 다양성 증대라는 점에서 의미가 있는 것이다.

References

- Ahn BI, Kim SY and Kim BR(2002). Trends and Causes of Spice Vegetable Price Instability in Korea, *The Korean Journal of Agricultural Economics*, 43(1), 103~121.
- Bae SW and Yu JS(2017). Predicting the Real Estate Price Index Using Deep Learning, *Korea Real Estate Review*, 27(3), 71~86.
- Box GEP and Jenkins GM(1976). *Time Series Analysis: Forecasting and Control*, Halden-Day.
- Choi BO and Choi IC(2007). Monthly price Forecasting of Fruit-type Vegetables Using Time-Series Analyses, *Journal of Rural Development*, 30(1), 129~148, <http://doi.org/10.36464/jrd.2007.30.1.012>.
- Choi JS(2016). Evaluation of Estimation and Forecast Accuracy on Retail Meat Prices by Seasonal Time-Series Models, *Korean Journal of Food Marketing Economics*, 33(1), 1~31.
- Chris B(2008). *Introductory Econometrics for Finance*, 2nd ed., Cambridga University Press.
- David JC and MacKay(1992). Bayesian interpolation, *Neural Computation*, 4(3), 415~447.
- E-Country Indicator(2020). “Seafood Consumption (per person per year) and Self-Sufficiency Rate”, <http://index.go.kr> on March.
- Foresee FD and Martin TH(1997). Gauss-Newton approximation to Bayesian learning, *Proceedings of the International Joint Conference on Neural Networks*.
- Han DJ and Park CH(2018). Comparative Analysis on the Forecasting Power of Squid Retail Price Forecasting Models by Processing Types Considering Seasonality, *The Journal of Fisheries and Marine Sciences Education*, 30(4), 1471~1483, <https://doi.org/10.13000/JFMSE.2018.08.30.4.1471>.
- Hochreiter, S. and Schmidhuber, J(1997). Long Short-Term Memory, *Neural Computation*, 9(8), 1735~1780.
- Hwang KS, Choi JW and Oh TY(2012). Forecasting common mackerel auction price by artificial neural network in Busan Cooperative Fish Market before introducing TAC system in Korea, *Journal of the Korean Society of Fisheries and Ocean Technology*, 48(1), 72~81.
- Jung MG and Nam JO(2021). A Comparison of Predictive Power among Forecasting Models of Monthly Frozen Mackerel Consumer Price Models, *The Journal of Fisheries Business Administration*, 52(4), 13~28, <https://doi.org/10.12939/FBA.2021.52.4.013>.
- Kim DH, Kim SH and Yu CJ(2021). A Study on Pork Price Prediction Using LSTM, *Korean Journal of Agricultural Management and Policy*, 48(4), 593~612, <http://dx.doi.org/10.30805/KJAMP.2021.48.4.593>.
- Kim SW, Yoon BS and B.Wade Brorsen.(2016). Forecasting Performance of Times Series Models for Korean Ginseng, *Korean Journal of Agricultural Management and Policy*, 43(2), 287~304.
- Kim TH and Choi SD(2021). Comparison of Forecast Models for Purple Laver Producer Price, Jeollanam-do - Exponential Smoothing Technique and ARIMA Model, *The Journal of Fisheries and Marine Sciences Education*, 33(4), 831~483, <https://doi.org/10.13000/JFMSE.2021.8.33.4.831>.
- Kim TH and Nam JO(2016). A Comparative Analysis on the Forecasting Power among Frozen Mackerel Retail Price Determination Models Using DM test, *The Journal of Korean Island*, 28(3), 59~81.
- Kim UJ(2021). *Concepts and Applications of Deep Learning*, MyrradSpace., 58~337.
- KMI Fishery Outlook Center(2022). “Wholesale Price of Rockfish”, <https://www.foc.re.kr> on March.
- Ko KS, Kim YW, Byeon SH and Lee SJ(2021). LSTM Based Prediction of Ocean Mixed Layer Temperature Using Meteorological Data, *Korean Journal of Remote Sensing*, 37(3), 603~614, <https://doi.org/10.7780/kjrs.2021.37.3.19>
- KOSIS(2022). “Fishery Production Trend Survey”, <http://kosis.kr> on March.
- KOSIS(2022). “Producer Price Index”, <http://kosis.kr> on March.
- KOSTAT(2022). *Result of Fish Farming Trend in 2021*, 1~44.
- Nam JO, Noh SG and Park EY(2012). A Study on Forecasting Oyster Price with Time Series Analysis, *Ocean Policy Research*, 27(1), 65~94.
- Nam JO and Jung MJ(2017). Comparison on Forecasting Power and Monthly Producer Price

- Forecast of Jeju Olive Flounder, Ocean Policy Research, 32(2), 1~21.
- Ock YS, Kim ST and Ko BH(2007). A Study on the Price Fluctuation and Forecasting of Aquacultural Flatfish in Korea, The Journal of Fisheries Business Administration, 38(2), 41~62.
- Song YN(2021). A study on predicting the producer price of farmed Olive flounder in Jeju Island using artificial neural network models, Unpublished master's thesis, Pukyong University.
- Warren S, McCulloch and W Pitts(1943). A logical calculus of the ideas immanent in nervous activity, Bulletin of mathematical biophysics, 5, 115~143.
- Yang SR and Park YS(1999). Forecasting Livestock Prices Using Neural Networks, The Korean Journal of Agricultural Economics, 40(1), 105~127.
- Yoon MS and Shin SY(2020). Prediction of the Apartment Sales Price Index Using Quantitative Time Series and Artificial Neural Network - Focused on Seoul and 6 Metropolitan Cities -, Journal of Korea Community Development Society, 45(2), 9~19.
-
- Received : 30 March, 2022
 - Revised : 02 May, 2022
 - Accepted : 06 May, 2022