신경회로망을 이용한 한우의 산지 가격 예측

Prediction of Cow Price Using Neural Network

조성인* 배영민*
정회원 정회원
S.I. Cho Y.M. Bae
E-mail : sicho@plaza.snu.ac.kr

1. 서론

UR의 타결 이후 상대적으로 열세에 있는 우리 나라 농축산물의 국제경쟁력을 강화시키기 위해서는 농산물의 생산성과 품질을 향상시키는 것 이외에도 수급 문제도 적절히 개선시켜 나가야 할 것이다. 수급 문제를 해결하기 위해서는 우선은 농축산물의 가격의 변동 및 예측에 대하여 고찰할 필요가 있다. 특히, 쌀과 소고기는 농축산물 중에서 큰 비중을 차지하고 있으므로 그것들에 대한 가격의 변동에 대한 연구가 필요하다.

시계열 자료는 일정한 시간 간격(time step)을 두고 모아진 자료이며, 이미 주어진 시계열 자료에 의해서 미래의 시계열 값을 예측하는 것을 시계열 분석(time-series analysis)이라 한다. 일반적으로 시계열 분석은 식 1 과 같은 시계열 모형을 개발하는 것이다.

$$\hat{Z}_t = f(Z_{t-1}, Z_{t-2}, Z_{t-3}, \cdots)$$

Z_t : Time－series value at time t

$\hat{Z}_t$: Predicted value at time t

f() : Time－series model (Eq. 1)

함수 f는 예측하고자 하는 시점 이전의 시계열 값을 가지고 만들어진 회귀 모형이 되며, 일반적으로 시계열 모형으로 가장 많이 사용되는 방법으로는 Box-Jenkins에 의해서 정리된 통계적 방법인 ARIMA모형이 있다(김, 1993). 이 예측 모형은 그 분야에 지식이 없는 사람이 접근하기 어렵고 예측 모형의 결정에 많은 시간이 들어 간다. 또한 시간과 시계열 자료간에 비선형성이나 잡음(noise)이 존재할 경우에는 그 예측 성능이 현저히 떨어지거나 모형의 적합이 이루어지는 않을 수도 있다. 따라서, 근래에는 시계열 분석에 통계적 방법 대신에 신경회로망을 이용하려는 연구가 진행되었다(Jhee등, 1994). 시계열 분석에 널리 이용되는 신경회로망은 구조는 다층 퍼셉트론(Multi-Layer Perceptron, MLP)으로, 이는 시스템의 입력과 출력간의 관계가 비선형적일 경우에도 모형화가 가능하고, 입력값에 섞인 잡음의 영향도 적게 받으며, 비선형 자기회귀분석(Nonlinear Autoregression)의 특징도 있다. 하지만, 입력층에서의 입력값의 개수나 은닉층에서의 노드 개수를 결정하는 일반적인 방법이 없다.

본 연구에서는, ARIMA모형과 신경회로망기법에 의해서 한우의 산지 가격에 대한 시계열 분석을 수행하였다. 그리고, 각각의 개발된 모형에 대한 성능을 비교하였다.

* : 서울대학교 농업생명과학대학 농공학과 농업기계전공

2. 재료 및 방법

2.1 한우 산지 가격 시계열 자료

본 연구에서의 시계열 분석 대상 자료는 1978년 1월 부터 1996년 8월까지의 한우(♀, 400kg) 산지 가격의 월별 자료이다. 이 자료의 출처는 축협 중앙회 조사부에서 발간하는 '축산물 가격 및 수급자료'이다

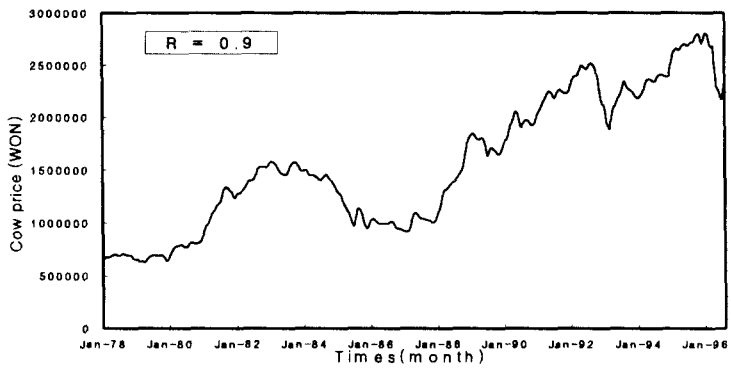

Fig. 1 The trend of cow price (Jan-78 ~ Aug-96)

한우 산지 가격의 시계열 자료는 그림 1과 같이 어느 정도의 변동(fluctuation)은 존재하지만 시간이 지남에 따라서 가격이 증가하는 추세를 나타내고 있다. 시간과 산지 가격과의 상관 계수(R)의 값은 0.9이다.

이 시계열자료 중에서 1978년 1월부터 1994년 8월까지의 자료는 시계열의 예측 모형을 개발(calibration)하는데 이용하였고 나머지는 24개월은 개발된 모형의 검증하는 데 이용하였다.

2.2 통계적 방법에 의한 시계열 예측

일반적으로 시계열분석에 이용되는 방법은 Box-Jenkins에 의해서 정리된 ARMA(Autoregressive Moving Average)모형이다.

$$Z_t = \phi_1 Z_{t-1} + \phi_1 Z_{t-2} + \cdots + \phi_p Z_{t-p+e_t} - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}$$

Z_t : Time-series value at time t

$e_t, e_{t-1}, \cdots, e_{t_q}$: White noises (Eq. 2)

$\phi_{t-1}, \phi_{t-2}, \cdots, \phi_{t-p}, \theta_{t-1}, \theta_{t-2}, \cdots, \theta_{t-p}$: Coefficients

ARMA(p, q) 모형은 식 2와 같은 식으로 나타나며, 상용화된 통계 팩키지를 이용하여 분석하였다. 분석순서는 '모형 식별', '모형 추정', '진단적 검사'의 세 단계로 이루어졌다. 모형 식별의 단계에서는 ARMA모형의 차수인 p, q를 결정하였으며, 모형 추정의 단계에서는 식별된 모형을 가지고 검증 자료를 이용하여 예측값을 계산하였다. 진단적 검사의 단계에서는

예측값들의 잔차 분석 등을 통하여 모형의 타당성을 조사하였다. 본 연구에서는 이(1995)의 연구 결과를 참고하여 시계열 분석을 수행하였다.

2.3 신경회로망에 의한 시계열 예측

신경회로망에 의한 시계열 예측은 주로 그림 2와 같은 다층 퍼셉트론(Multi-layer Perceptron, MLP)의 구조를 가진다. 입력값은 예측 시점의 이전 몇 단계의 시계열 자료값이 되고 출력값은 예측 시점의 시계열 자료값이 된다.

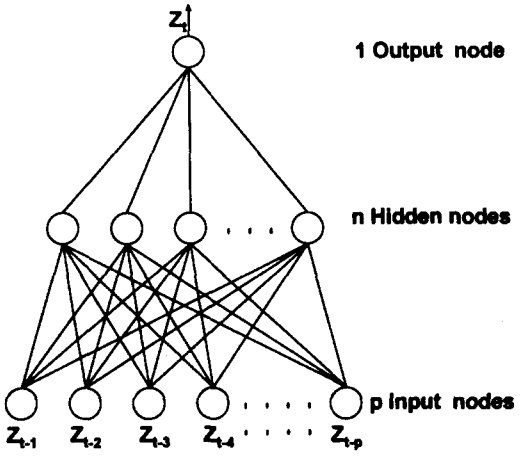

Fig. 2. Structure of multi-layer perceptron for Time-series analysis(지, 1992)

본 연구에서는 예측하고자 하는 시점에서 두 시차 이전의 값부터 입력 변수의 개수를 증가시켜가면서 학습을 시켰다. 또한, 은닉층의 노드 개수를 증가시켜가면서 학습을 시켰다.

2.4 시계열 모형간의 비교

통계적 모형과 신경회로망의 예측 모형의 성능을 비교하기 위한 기준으로서 평균절대백분위오차(Mean Absolute Percentage Error : MAPE)를 이용하였다. 이 MAPE는 식 3과 같이 정의된다.

$$\text{MAPE} = \frac{\sum_{n=1}^{N} \frac{|Z_n - t_n|}{Z_n} \times 100}{N} \qquad \text{(Eq. 3)}$$

Z_i : Time－series value at time step t

t_i : Predicted value at time step t

N : Number of sample

3. 결과 및 고찰

3.1 통계적 방법에 의한 소 산지 가격 예측

한우의 산지 가격은 그림 1 에서와 같이 시간이 지남에 따라 그 값이 증가(R = 0.9)하는 추세를 보이고 있다. 이것은 일반적인 시계열 분석에서는 정상성(stationary)이 결여되었다는 것을 의미한다. 따라서, 이러한 정상성의 결여를 제거하기 위해서 이웃하는 시계열 값사이의 차($\Delta Z = Z_t - Z_{t-1}$)를 시계열 분석에 이용하였다. ΔZ의 시계열 추세는 그림 3 과 같다.

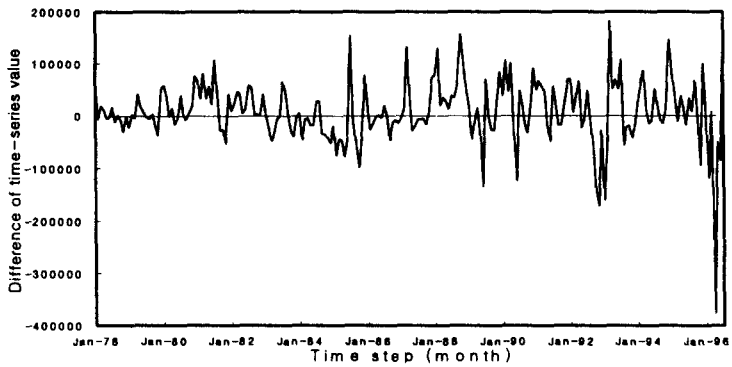

Fig. 3 Plot of difference of time-series values

그림3에서의 ΔZ은 변동은 원 시계열 데이터에서 나타났던 증가 추세가 제거되었으며, 이 데이터를 가지고 시계열 분석을 수행하였다(이, 1995 참조). 분석된 모형은 ARIMA(1, 1, 1)(0, 0, 1)12 모형으로 나타났으며 이 모형에 의한 검증 데이터의 검증 결과는 그림 4와 같다. MAPE의 값은 8.67%였다.

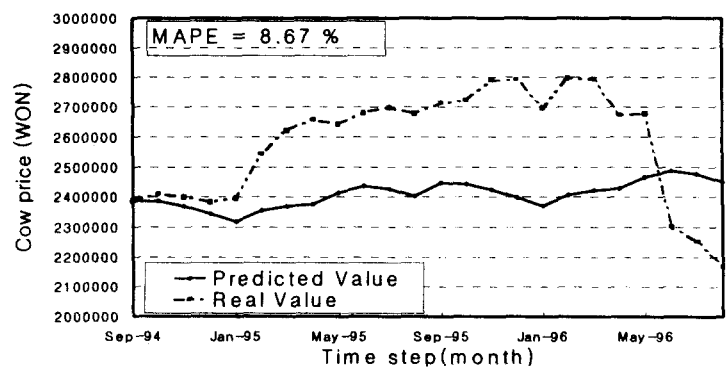

Fig. 4 Result of prediction of cow price using ARIMA model

3.2 신경회로망에 의한 소 산지 가격 예측

그림 5는 신경회로망의 입력 변수의 개수에 따른 학습용 데이터의 MAPE와 검증용 데이터의 MAPE를 도시한 것이다. 본 신경회로망의 학습에서는 학습률의 값은 0.05로 하였으며, 모멘텀은 0, 학습횟수는 100000번으로 하였다.

학습용 데이터의 경우 입력변수의 개수가 증가함에 따라, MAPE가 감소하는 경향이 나

타났다. 이는 입력 변수가 증가함에 따라 입력 변수에 포함되어 있는 잡음까지도 모형화하는 현상(overfitting)이 나타나는 것으로 판단된다. 하지만, 검증용 자료의 MAPE는 입력 변수의 개수가 증가함에 따라 어느 정도 감소하다가 다시 증가하는 형태를 나타내고 있다. 입력 변수가 4개보다 적을 때는 학습용 데이터에 존재하는 시계열 값의 예측에 필요한 정보를 완전하게 적합시키는 못하다가(underfitting), 5개 이상일 때는 학습용 데이터의 값에 존재하는 잡음까지도 적합시키는 현상(overfitting)이 나타난 것으로 판단된다. 따라서 입력 변수의 개수가 4개 일때 한우 산지 가격의 시계열 값을 가장 적절히 적합시키는 것으로 판단된다.

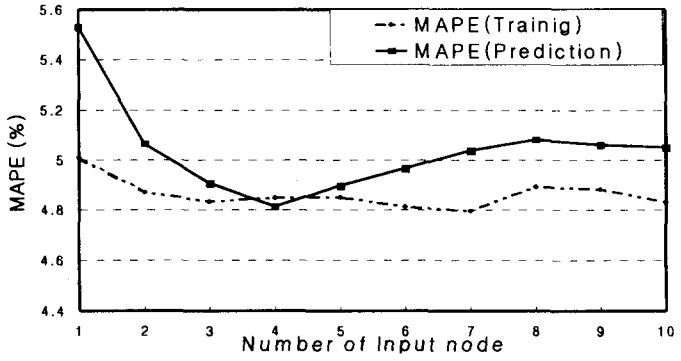

Fig. 5 Plot of MAPE vs number of input nodes.

그림 6은 입력 변수가 4개일 경우에 은닉층에서의 노도의 개수에 따른 MAPE의 변화를 도시한 것이다.

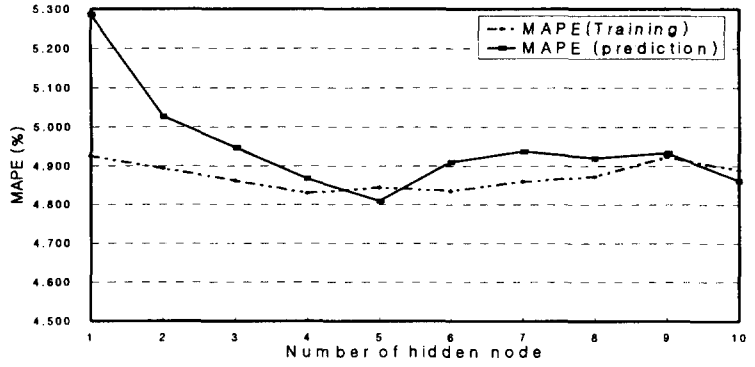

Fig. 6 Plot of MAPE vs number of hidden node at 4 input node

은닉노드의 수가 5개 일때 검증용 데이터의 MAPE가 최소가 되는 것으로 나타났다. 따라서, 입력 변수의 수를 4개 은닉노드의 수를 5개로 하여 검증용 데이터의 예측값을 구하였다 (그림 7).

* : 서울대학교 농업생명과학대학 농공학과 농업기계전공

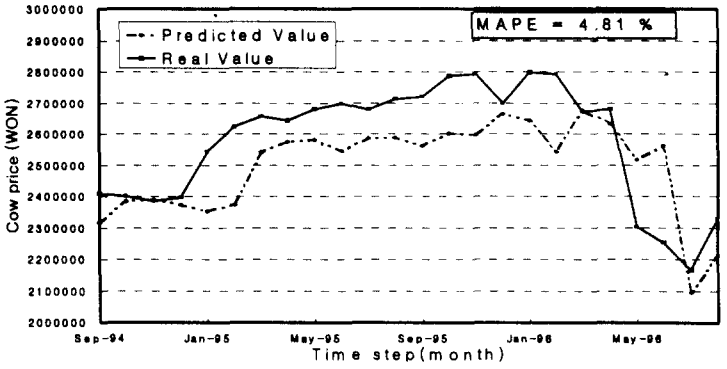

Fig. 7 Result of prediction using neural network(prediction)

MAPE가 4.81%로 통계적 방법인 ARMA모형에 의한 결과보다 우수한 결과를 나타내었다. 따라서, 한우의 산지 가격의 예측을 위한 시계열 분석법으로서는 통계적 방법인 ARMA보다다 신경회로망이 우수한 것으로 판단된다.

4. 결론

한우(♀, 400kg)의 산지 가격을 예측하는 시계열 모형을 개발하기 위해서 통계적 방법인 ARMA모형과 신경회로망기법을 이용하였으며 개발된 두 모형의 성능을 비교하였다. 각 모형의 성능을 비교하기 위해서 평균절대백분위오차(MAPE)를 이용하였다.

ARMA모형의 분석에서는 시계열 자료의 추세를 제거하기 위해서 차분법을 수행한 후에 시계열 분석을 수행하였다. 분석된 모형은 ARIMA(1, 1, 1)(0, 0, 1)12모형이며, 이 모형에 의한 MAPE는 8.61% 였다.

신경회로망에 의한 시계열 분석에서는 예측 시점의 두 시점 이전의 값부터 입력 변수의 개수를 증가시켜가며 학습을 시켰고, 입력 변수 4개일 경우에 검증용 데이터의 MAPE가 가장 우수하였다. 또한 입력 변수 4개일 경우의 은닉층의 노드 개수를 증가시켜가며 학습을 수행하였다. 은닉층의 노드 개수가 5개일 때 검증용 데이터의 MAPE가 가장 좋았다. 따라서, 입력 변수 4개 은닉층의 노드개수 5개일 때, MAPE가 4.9%로 가장 좋은 예측 성능을 나타냈으며, 이는 통계적 방법인 ARMA모형의 분석에 의한 예측 결과보다 우수하였다.

5. 참고 문헌

· 지원철, 1992, 경영을 위한 신경망 응용, 정보과학회지, 제 10 권 제 2 호

· 이진원, 1995, 한우의 산지가격에 관한 시계열 분석, 식품유통연구 Vol.12(1) pp. 137-147

· 김원경, 1993, 시계열 분석, 경문사, 서울

· Jhee, W. C., Michael J. Shaw, 1994, Minimally-structured Neural Networks for Times Series Forcasting, Artificial Intelligence Technical Report UIUC-BI-AI-DSS-94-04