딥러닝 기반의 기상정보를 반영한 배추 가격 예측

채명수* · 정성관*§

*한국과학기술원(KAIST) IT융합연구소

Price Forecasting of a Chinese Cabbage with Meteorological Information using Deep Learning Technique

Myungsu Chae* · Sungkwan Jung*§

*KAIST Institute for IT Convergence

E-mail : {mschae89, skjung§}@itc.kaist.ac.kr

요 약

농산물에 대한 가격을 정확하게 예측하는 것은 정부와 지방자치단체, 농업관련 주체들에게 중요한 정보가 된다. 또한 농산물의 생산 및 출하는 기상상황에 큰 영향을 받는다. 본 논문에서는 기상상황에 따른 가격변동성이 큰 배추에 대해 최근 각광받는 딥러닝 기술을 적용하여 가격 예측 모형을 제안하였다. 기존의 배추 가격 예측 모형과 예측 성능을 비교하였고 그 우수성을 확인하였다.

ABSTRACT

It is important to predict price of agricultural products accurately to government, local government, bodies in charge of agriculture. Production and shipping of agricultural products are affected by weather condition significantly. In this research, prediction model of a Chinese cabbage which is highly sensitive to weather condition is proposed using deep learning technique. After performance of proposed model and a model of previous research is compared, superiority of proposed model is proved.

키워드

농산물, 가격예측, 기상정보, 기계학습, 딥러닝

Ⅰ. 서 론

농산물의 가격을 정확하게 예측하는 것은 정부를 비롯한 지방자치단체와 농업관련자들에게 매우 중요하다. 하지만 가격의 급격한 변동인 가격파동은 작황의 변화에 따라서 그 변동 폭이 더욱 심화되기 때문에 가격변화에 대한 예측을 더욱 어렵게 만드는 경향이 있다[1]. 농산물의 계절성 및 저장성에 따라 다양한 유형의 가격변화가 나타나는데, 대부분의 농산물의 경우 수요량의 변동은 적으나 의생적인 기후, 자연재해 등의 영향으로 생산량이 불안정하고 낮은 저장성으로 인해 재고관리에 어려움이 있어 잦은 가격파동을 유발한다[2, 3]. 이러한 가격의 불안정으로 인한 파동은 가격폭락으로 인해 미판매 방치하거나 자원의 비효율적 이용을 야기하여 농가소득의 안정성 측면에 큰 피해를 입히거나 반대로 가격폭등으로 인한 소비자 물가불안정의 원인이 된다[4].

본 연구에서는 이러한 가격변동에 대해 정확한 예측을 하기 위한 예측모형을 제안하였다. 특히, 야외의 노지에서 재배되는 채소류의 경우 기상상황에 따른 생산량의 변동이 심하기 때문에 기상정보를 함께 고려한 농산물 가격 예측모형을 제안하였다. 특히, 기상정보의 영향을 강하게 받는 것으로 기존 연구[5]에서 확인된 배추의 가격 예측모형을 수립하여 그 성능을 기존 예측모형과 비교하여 우수성을 검증하였다. 예측모형은 최근 예측 및 분류문제에서 우수한 성능을 보여 각광받는 기계학습(machine learning) 기법인 딥 러닝(deep learning)을 적용하였다.

Ⅱ. 관련 연구

농산물 가격에 대한 예측모형 연구는 국내외를 막론하고 다양하다. 해외 사례로는 인도 코임바토르 시장의 2009년부터 2011년까지 토마토 가격정보에 대해 역전파 뉴럴 네트워크(backpropagation neural network)를 적용하여 예측한 연구가 있다[6].

국내 사례로는 가격의 변동성이 매우 큰 배추와 무, 양파, 마늘을 대상으로 1987년부터 2004년까지 가락시장 월별 가격 예측 모형을 ARIMA 기법을 적용하여 수행한 연구가 있다[7]. 또한, 청과물 14종의 가격 계절성을 고려하여 1996년부터 2005년까지의 가락시작 평균가격을 GARCH와 VAR 모형을 적용하여 예측한 연구가 있다[8]. 최근에는 배추에 대하여 1996년부터 2016년까지의 가격정보를 토대로 2017년 배추가격을 다중 뉴럴 네트워크(multi-layer neural network: Multi-layer NN)를 적용하여 예측한 연구가 있다[9].

다양한 연구가 있지만, 기존 연구에서는 가격 변동성이 높은 것으로 확인된 배추의 가격 예측 모형에 대한 연구가 부족한 실정이고 최근 예측 및 분류문제에 우수한 성능을 보여 각광받는 딥러닝 기법을 적용한 연구가 부족하다. 그리하여 본 연구에서는 배추 가격의 예측모형을 딥 러닝 기법 중 시계열 예측에 적합한 순환 뉴럴 네트워크(recurrent neural network: RNN)를 적용하여 수립하고 실험을 통해 그 성능을 검증하였다.

Ⅲ. 제안하는 방법

본 연구에서는 배추 가격 예측모형 수립에 순환 뉴럴 네트워크 기법을 적용하였다. 이 기법은 기계학습 기법들 중 뉴럴 네트워크 기법의 한 갈래이다. 뉴럴 네트워크 기법은 동물의 중추신경계 중 특히 뇌의 신경망에서 영감을 얻은 통계학적 학습 알고리즘이다[9]. 딥 러닝은 뉴럴 네트워크의 은닉층(hidden layer)이 깊은 뉴럴 네트워크를 말하며 최근 컴퓨터 비전(computer vision) 분야에서 딥 러닝의 예측 성능이 매우 우수하다는 것을 입증하는 연구들이 많이 있다. 본 연구에서 적용한 순환 뉴럴 네트워크가 일반적인 뉴럴 네트워크와 다른 점은 각각의 입출력이 서로 독립적이지 않고 순서를 가진다는 점이다. 따라서, 모든 입력 데이터에 대해 동일한 연산을 수행하지만, 연산 시점에서의 출력 데이터는 이전의 모든 연산에 영향을 받는다는 특성이 있다. 이러한 특성으로 인해 영상의 순서를 매기거나 주식가격의 예측, 파형 생성, 다국어 기계번역 등의 분야에 널리 적용되고 있다.

순환 뉴럴 네트워크 등장 이후에 그래디언트 소실 문제(gradient vanishing problem)를 해결한 LSTM(long-shrot term memory)와 GRU(gated recurrent unit) 등이 개발되었으나 본 연구에서는 뉴럴 네트워크의 입출력에 순서를 반영하는 시도 측면에서 순환 뉴럴 네트워크를 적용하였다.

순환 뉴럴 네트워크 수립에 필요한 매개변수인 학습율(learning rate)과 은닉층의 수, 그리고 에폭(epoch)의 수는 실험적으로 결정하였다.

Ⅳ. 실험 및 결과

본 연구에서는 한국농수산식품유통공사에서 제공하는 농산물 유통정보 서비스에서 배추가격 데이터를 수집하였다. 수집기간은 1996년 1월부터 2016년 말까지이고, 수집지역은 서울지역이며, 수집 데이터는 도매시장의 월초, 월중, 월말 평균가격이다. 예측모형에 함께 고려한 기상정보는 기상청 국가기후데이터센터에서 수집하였으며 기존 연구에서 배추의 가격에 영향력이 큰 기온과 강수량 정보를 활용하였다.

예측모형의 학습 및 검증은 k-겹 교차검증(k-fold cross validation) 기법으로 하였고, 학습은 21기간 중 20기간으로, 검증은 나머지 1기간으로 진행하였다. k-겹 교차검증을 적용하여 학습 데이터에의 과적합(overfitting) 문제를 피하였다.

표 1. 배추 가격 예측모형의 실험결과

	RNN	Multi-layer NN[9]
RMSE	1.69%	6.42%

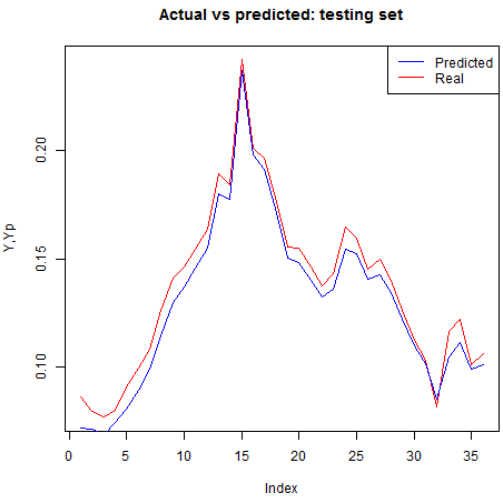

그림 1. 2015년도 배추 가격 예측값과 실제값 비교 그래프

예측모형의 성능지표로는 시계열 예측에 보편적으로 사용되는 평균제곱근오차(root mean square error: RMSE)를 사용하였고, 실험 성능의 우수성을 검증하기 위해 다중 뉴럴 네트워크를 적용했던 기존 연구[9]와 예측 성능을 비교하였다. 표 1을 통해 RMSE 기준으로 순환 뉴럴 네트워크가 기존의 다중 뉴럴 네트워크보다 훨씬 우수한 성능을 보이는 것을 알 수 있다. 이를 통해, 순환 뉴럴 네트워크의 시계열에서의 강점을 확인할 수 있다. 그림 1은 검증 데이터를 2015년도로 설정하였을 때의 순환 뉴럴 네트워크 기반의 배추 가격 예측값과 실제값을 그래프로 나타낸 것이다.

Ⅴ. 결 론

본 연구에서는 농산물 가격 변동이 매우 큰 배추를 대상으로 기상정보를 반영한 순환 뉴럴 네트워크 기반의 가격 예측모형을 수립하여 기존 연구 대비 성능의 우수성을 검증하였다. 그 결과, 순환 뉴럴 네트워크 기반의 가격 예측모형이 기존의 다중 뉴럴 네트워크 기반의 가격 예측모형 대비 매우 뛰어난 예측성능을 보이는 것을 확인하였다. 향후에는 앞서 언급했던 LSTM과 GRU 등의 적용과 함께 실험적으로 선택했던 매개변수들의 최적화에 대한 추가적인 연구가 필요하다.

ACKNOWLEDGMENT

이 논문은 2017년도 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (2015-0-00215, GS1 표준 기반의 균형생산·투명유통·안전소비를 위한 농·축산 클라우드 및 응용서비스 개발)

참고문헌

[1] S, Kyo, T. Kim, and J. Lee. “Reliability Analysis for Price Forecasting of Chinese Cabbage.“ Journal of the Korean Society of Agricultural Engineers, Vol. 50, No. 3, pp.71-79, 2008.

[2] H. Kim, H. T. Kim, and T. H. Lee, Price Theory for Agricultural Products, Park Young Sa, 2003.

[3] S. Yoon, and J. Noh, The Price Fluctuations of the Major Agricultural Products Analysis and the Introduction of the Futures Market, Vol. 2, pp.53-82, 1994.

[4] B. S. Lee, H. W. Jung, and C. K. Choi, “Acreage Decision Process of Radish and Chinese Cabbage Under Asymmetric Response to Price Changes,” Agricultural Policy Research, Vol. 23, No. 1, pp.79-93, 1996.

[5] H. Kim, and Y. Kim, “A Study on the Dynamic Relations between Price of Agricultural Wholesales Markets : Focused on price of Garak Wholesale Market and Gangseo Wholesale Market,“ Journal of Distribution Research, Vol. 17, No. 5, pp.59-83, 2012.

[6] N. Hemageetha, and G. M. Nasira, “Radial basis function model for vegetable price prediction,“ Pattern Recognition, Informatics and Mobile Engineering (PRIME), 2013 International Conference on. IEEE, 2013.

[7] B. S. Kim, “A Comparison on Forecasting Performance of the Application Models for Forecasting of Vegetable Prices,“ The Korean Journal of Agricultural Economics, 2005.

[8] Y. S. Lee, B. O. Choi, and S. B. Shim, “A Time Series Analysis on Prices of Fruits and Vegetables,“ R537, Korea Rural Economic Institute, 2006.

[9] M. Chae, S. Kim, C. Lee, Y. Lee, and S. Jung, “Price Forecasting of Agricultural Product with Meteorological Information using Multi-layer Neural Networks”, The Korean Institute of Communications and Information Sciences Summer Conference, pp. 878-879, 2017.