

감정 데이터를 활용한 딥러닝 기반 돈육가격 예측 연구

정재민*¹, 정경창**², 고은영***³, 이의종*²

Deep Learning based Pork Price Prediction using Sentimental Analysis

Jaemin Jeong*¹, Kyungchang Jeong**², Eunyoung Ko***³, and Euijong Lee*²

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학지원사업(2019-0-1183)과 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2021R1G1A101097111)

요 약

해마다 육류 소비는 증가하고 있고 돼지고기는 많이 소비되는 육류 종류이다. 따라서 돈육 가격을 예측하는 것으로 선제적 돈육 수급 조절 등 이해관계자들에게 이점을 제공한다. 축산 시장에서 축산물의 가격은 다양한 요인에 영향을 받으므로 다양한 예측 변인을 고려해야 한다. 따라서 본 연구에서는 여러 정보가 담겨 있는 뉴스 데이터를 활용하여 돈육 도매가격을 예측하는 모델을 제안하고자 한다. 감정분석기를 활용하여 뉴스 문장들의 감정을 분석하고 분석한 감정을 이용하여 감정 점수를 측정했다. 측정된 감정 점수를 활용하여 딥러닝 모델을 통해 도매가격을 예측했다. 시계열 딥러닝 모델 학습 방식 중 다대다 방식을 이용하여 학습날짜와 예측 날짜 일수를 조정하며 성능 비교평가를 했다. 전체적으로 돈육 도매가격으로만 학습하였을 때보다 감정 점수를 활용하여 예측한 모델이 더 낮은 오차로 예측함을 확인할 수 있었다.

Abstract

Meat consumption has increased every year, and pork is the most consumed type of meat. Therefore, predicting pork prices provides advantages to various stakeholders such as preemptive control of the supply and demand of pork. However, livestock prices are affected by various factors; thus, considering various variables has to be required. In this study, we proposed a model to predict pork prices using news data that contains various information about the pork market. A sentimental score was applied to analyze the atmosphere of the pork market, and the score was used to predict pork prices. Deep learning models are applied to predict the pork price using the sentimental score. We collected experimental data from 4 years of wholesale pork prices, and pork-related news. The results demonstrated the efficiency of the proposed method to predict pork prices using the sentimental score.

Keywords

price prediction, text mining, sentiment analysis, deep learning

* 충북대학교 소프트웨어학부(*² 교신저자)
- ORCID¹: <https://orcid.org/0000-0003-2229-3898>
- ORCID²: <https://orcid.org/0000-0002-7308-7392>
** 충북대학교 컴퓨터과학과 석·박사통합과정
- ORCID: <https://orcid.org/0000-0003-3659-4923>
*** 도드람 협동조합 실장
- ORCID: <https://orcid.org/0000-0002-8686-1762>

· Received: Sep. 04, 2023, Revised: Nov. 22, 2023, Accepted: Nov. 25, 2023
· Corresponding Author: Euijong Lee
School of Computer Science, Chungbuk National University, Republic of Korea
Tel.: +82-43-261-3133, Email: kongjigae@cbnu.ac.kr

I. 서 론

해마다 육류 소비는 증가하고 있고 돼지고기는 많이 소비되는 육류 종류이다[1]. 따라서 돈육 가격을 예측하는 것으로 물가안정 및 정부, 농민, 소비자와 같은 여러 이해관계자에게 이점을 줄 수 있다.

시장에서 가격은 수요와 공급으로 결정되고[2], 수요와 공급은 시장참여자들의 반응으로 결정된다. 축산 시장에서 시장참여자들의 반응은 기온, 질병, 수입품목 등 다양한 외부 요인에 따라 달라진다[3]-[5]. 예를 들어, 정부는 질병이 넓은 지역으로 확산하면 방역을 위해 축산물 이동 금지 명령을 내려 축산물 공급을 통제한다. 농부들은 이와 같은 통제 명령과 질병으로 인한 축산물 피해를 염려하여 선제적으로 돼지고기 물량을 선 출하하여 공급 물량을 늘리게 된다. 소비자는 축산물 품질에 대한 의심으로 소비량을 줄이게 된다. 이런 종합적인 반응으로 공급은 증가하는 반면 수요는 감소하게 되어 가격이 하락한다. 이처럼 정확한 가격 예측을 위해서는 시장참여자들의 반응을 고려하여 예측할 필요가 있다.

여러 매체 중 뉴스 기사는 질병, 수입 정책, 소비 심리 등 기타 사건에 대해 시장참여자들이 빠르고 편리하게 접근할 수 있는 신뢰성 있는 매체 중 하나이다[6]. 따라서 본 연구에서는 뉴스 기사를 활용하여 시장참여자들의 반응을 고려한 돈육 도매가격 예측 모델을 제안하고자 한다.

텍스트 마이닝[7]이란 비정형 텍스트를 정형화하여 의미 있는 패턴과 새로운 정보를 찾아내는 프로세스이다. 축산 시장뿐만 아니라 다양한 시장 분야에서 텍스트 마이닝을 활용하여 가격 예측하는 연구들이 선행됐다[8]-[12]. 하지만 축산 시장에서는 주로 정보의 빈도나 정형 데이터를 이용한 가격 예측으로 표면적인 데이터로만 활용했다는 점에서 한계가 있었다. 본 연구에서는 뉴스 기사의 내용을 분석하여 내용적인 측면을 고려하여 가격을 예측하고자 한다.

가격 예측 모델로는 딥러닝 모델을 활용하였다. 딥러닝 모델 중 시계열 예측 모델로 잘 알려진 LSTM(Long Short Term Memory)[13], GRU(Gated Recurrent Unit)[14]를 성능 비교하여 성능이 높은 딥러닝 모델을 돈육 가격 예측 모델로 제안하고자 한다.

이를 위한 본 논문의 구성은 다음과 같다. 2장에서는 텍스트 마이닝 기법을 이용한 가격 예측 연구에 대해 살펴보고 본 연구의 차별성을 소개한다. 3장은 돈육 도매가격 예측 시스템을 소개하는 장으로 본 연구에서 제안한 시스템 구조도, 데이터 습득 및 전처리, 감정분석, 모델링에 대해 소개한다. 4장 실험 설계, 데이터 구성 및 전처리, 모델 구조 및 학습, 실험 결과를 제시하며 5장에서 결론을 맺는다.

II. 관련 연구

본 장에서는 선행되었던 텍스트 마이닝[7]을 이용한 가격 예측 연구를 소개한다. Haider Maqsood 외[8]의 연구에서는 2012년부터 2016년까지 유명한 사건 중 일부의 영향을 탐색하여 미국, 홍콩, 터키, 파키스탄의 주식 가격 예측 모델을 제안하였다. 해당 기간 내에 일어난 중요 사건을 탐색하고 이 사건들을 영향에 따라 국가별로 지역적인 사건과 세계적인 사건으로 분류하였다. 해당하는 사건에 관련된 트위터를 수집하고 수집한 트위터를 SentiWordNet[15]를 활용하여 감정 분석하고 분석된 감정의 비율을 통해 감정 점수를 측정한다. 이렇게 측정한 감정 점수를 시가, 최고가, 최저가, 거래량과 함께 활용하여 증가를 예측하였다.

Wu, Binrong 외 [9]의 연구에서는 원유 가격의 고유한 특성, 원유 시장 요인과 외생적 요인에 영향을 받는다는 것을 고려하여 다양한 요인들을 고려할 수 있는 변수들이 필요성을 발견하였다. 빅데이터 기술 등장으로 온라인 데이터가 이러한 조건에 충족하고 구글 트렌드 검색어와 온라인 뉴스를 통해 원유 가격을 예측하는 연구를 하였다. 온라인 뉴스 특성상 양이 방대하고 불필요한 내용이 많아 CNN(Convolutional Neural Network)[16] 기술을 이용하여 내용을 선별하는 작업을 하였다. 선별된 뉴스 기사와 구글 트렌드 검색어를 이용하여 원유 가격을 예측하는 모델을 제안하였다.

Chuluunsaikhan Tserenpurev 외 [10]의 연구에서는 일별 뉴스 기사 수와 돈육 가격의 변동성에 상관성이 있다는 것을 관측하였고 뉴스 기사를 활용하여 가격 예측 모델을 제안하였다.

먼저 뉴스 기사를 LDA(Latent Dirichlet Allocation)[17] 기법을 활용하여 주제들을 선별하고 관련성 있는 주제를 기사에 할당한다. LDA 주제 선별 결과 수입, 질병, 농장, 시장, 정부, 가격 등 돈육 가격에 영향을 줄 수 있는 주제들을 선별할 수 있었다. 이 중 가장 유사성이 높은 주제를 뉴스 기사 주제로 선정한 후 주제에 있는 단어들을 TF-IDF(Term Frequency-Inverse Document Frequency)를 이용하여 LSTM[13] 모델에 인풋 데이터(Input data)로 활용한다. 마지막으로 최신 통계 기술 (ARIMAX, Ridge), 머신러닝, 딥러닝 기술을 활용하여 성능평가를 하였다. 성능평가 결과 LSTM을 사용하여 제안한 모델이 가장 높은 정확성을 보였다.

텍스트 마이닝을 이용하여 예측하는 다수의 연구들은 선행됐으나 키워드 빈도수에 기반을 둔 표면적인 데이터만을 이용하였다는 점에 한계점이 있었다. 따라서 본 연구에서는 기사의 내용을 표면적인 정보가 아니라 내용을 파악하기 위해 감정 점수라

는 파생 변수를 만들어 가격 예측에 활용했다.

III. 돈육 도매가격 예측 시스템

3.1 시스템 구조

축산 시장은 기온, 질병, 수입품목 등 여러 외부 요인에 의해 영향을 받는다. 이러한 영향은 시장참여자의 반응에 영향을 주게 되고 가격은 변동성을 갖게 된다. 따라서 본 논문에서는 시장참여자들의 반응을 가격 예측에 활용하기 위해 뉴스 기사 내용을 감정 분석기를 이용하여 감정을 평가하고 평가된 감정으로 파생 변수인 감정 점수를 만들어 돈육 도매가격 예측 시스템을 제안하고자 한다.

그림 1은 본 논문에서 제안하는 감정분석을 이용한 돈육가격 예측 시스템 구조를 나타낸다. 해당 시스템의 전체적인 구조는 데이터 습득 및 전처리, 감정 분석, 모델링으로 구성된다.

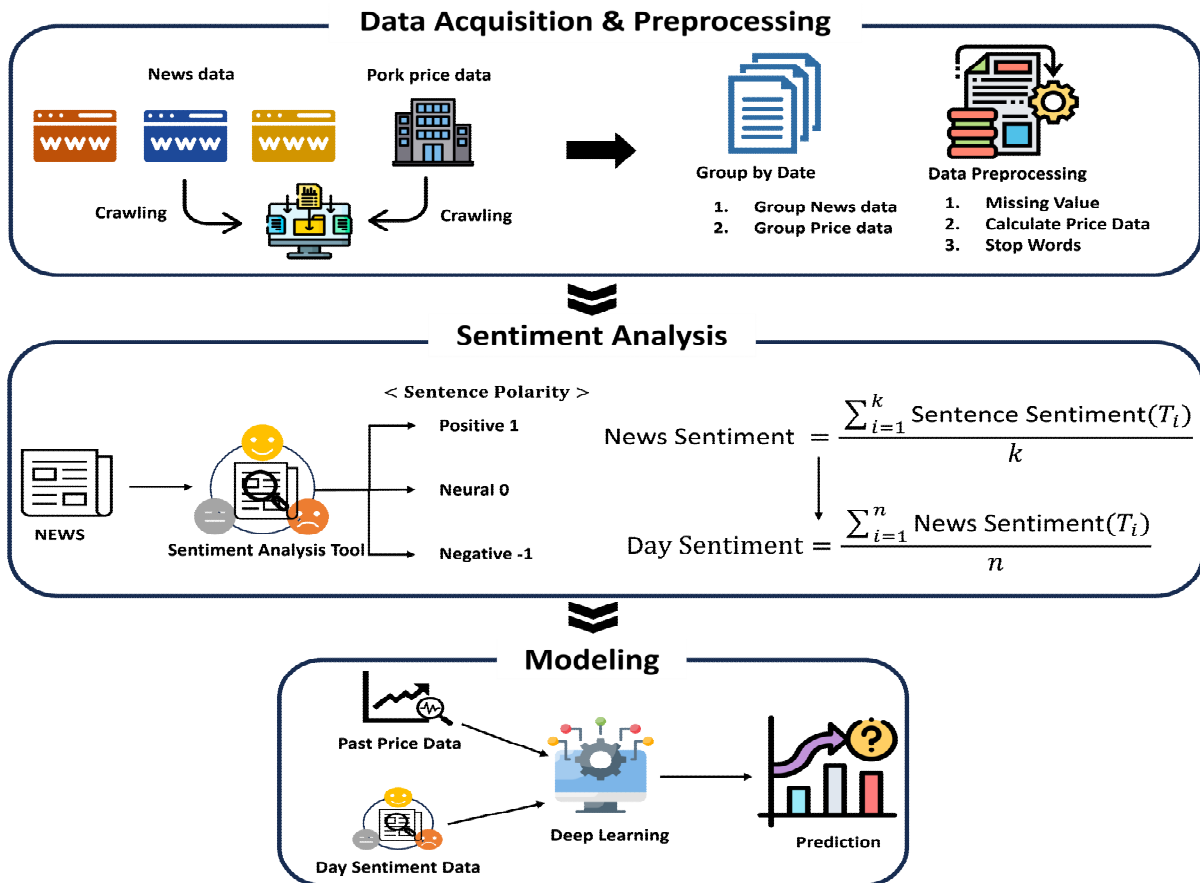


그림 1. 감정분석을 이용한 돈육 가격 예측 시스템 구조
Fig. 1. Structure of pork price prediction system using emotional analysis

3.2 데이터 수집 및 전처리

양돈 시장과 관련 있는 뉴스 정보를 수집하기 위해 양돈 전문 뉴스 웹사이트 [18]-[20]를 선정하였다. 파이썬 라이브러리 Beautiful Soup[21]과 프레임워크 Selenium[22]을 사용하여 각 웹사이트에서 양돈 관련 기사 내용만 선별하여 크롤링하였다. 가격 데이터는 축산물 품질 평가원[23]에서 도매가격 데이터를 수집했다.

일별 가격 예측을 위해 뉴스 데이터와 돈육 가격 데이터를 같은 일별로 그룹화했다. 일별로 결측값 있는 데이터들은 이전일 값으로 결측값을 변경했다. 돈육 도매가격 데이터는 일별로 평균값을 계산하여 날짜별 돈육 도매가격을 계산했다. 뉴스 기사 데이터에서 특수문자, 이름, 구두점, 개행문자 등 감정분석에 불필요한 불용어들은 제거했다.

3.3 감정분석

뉴스 기사의 긍정, 부정적인 내용에 따라 시장참여자들의 반응은 달라진다. 따라서 뉴스 기사의 내용에 대한 정량적인 감정 지표 측정이 필요하다.

감정 지표를 측정하기 위해 수집된 뉴스 기사 내용을 감정 분석기를 활용하여 감정 점수를 평가했다. 감정 분석[24]이란 기계 학습과 자연어 처리(NLP)를 사용하여 텍스트 감정(긍정적, 부정적, 중립적)을 자동으로 분석하는 텍스트 마이닝 기법이다. 감정 분석기로는 네이버에서 개발한 CLOVA Sentiment API[25]를 이용하여 감정 분석했다. CLOVA Sentiment API는 BERT(Bidirectional Encoder Representations from Transformers)[26] 모델 기반으로 다양한 연구 분야에서 텍스트 감정분석에 활용되고 있다[27-28]. 뉴스 기사를 텍스트 형태로 API를 통해 데이터를 보내면 네이버 CLOVA Sentiment 감정 분석기에서 문장 단위로 감정을 분리하여 표 1과 같이 긍정, 부정, 중립 감정을 반환하고 반환된 감정을 긍정은 1, 부정은 -1, 중립은 0으로 바꾸어 감정 점수를 계산할 수 있게 숫자 형태로 바꿨다. 문장 단위로 반환된 감정을 식 (1)처럼 문장별 감정 합과 문장 수로 나누어 평균으로 계산하여 뉴스 감정 점수를 계

산하였다. 이후 뉴스 감정 점수를 식 (2)을 이용하여 일자별 평균으로 계산하여 일자별 뉴스 감정 점수를 구했다.

$$\text{News Sentiment} = \frac{\sum_{i=1}^k \text{Sentiment}(T_i)}{k} \quad (1)$$

$$\text{Day Sentiment} = \frac{\sum_{i=1}^n \text{News Sentiment}(T_i)}{n} \quad (2)$$

- Sentiment={Positive :1, Neutral :0, Negative :-1}
- Sentiment= 특성 문장의 감정 점수
- News Sentiment=뉴스 기사 감정 점수
- Day Sentiment=일별 감정 점수
- T_i =i 번째 감정 점수
- k=감정 평가된 특성 문장 수
- n=뉴스 기사 수

3.4 모델링

돈육 도매가격은 주기적인 계절성 변화와 이전의 일련의 사건과 밀접하게 연관되어 있다. 따라서 시계열적 특성을 고려할 수 있는 시계열 딥러닝 모델 LSTM을 시스템 모델로 선정했다.

시계열 모델 예측 방식은 일대다, 다대일 등 다양한 방식이 있지만 그림 2와 같은 다대다[29] 방식을 활용했다. 다대다 방식은 시퀀스 데이터가 입력됐을 때 시퀀스 데이터가 나오도록 학습하는 방식으로 여러 일수의 날짜들을 학습하여 여러 일수의 미래 날짜를 예측하는 방식으로 모델링 했다.

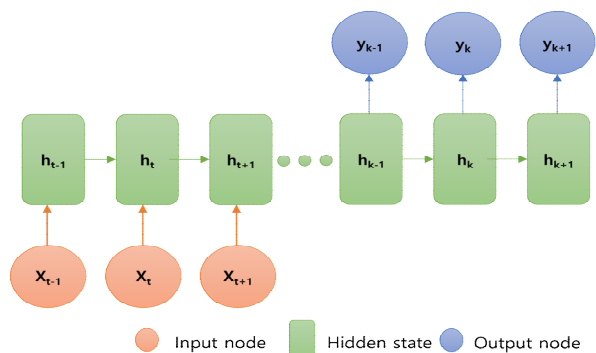


그림 2. 다대다 방식
Fig. 2. Many to many

IV. 실험 설계 및 성능평가

4.1 실험 설계

본 장에서는 감정 점수가 돈육 도매가격 예측에 영향을 줄 수 있는 지표인지 확인하기 위한 실험을 기술한다. 돈육 도매가격만 가지고 윈도우 사이즈를 활용하여 학습한 모델과 돈육 도매가격과 감정 점수를 함께 학습한 모델을 성능 비교하는 방식을 통해 감정 점수의 유효성을 검증했다. 또한 학습 시 훈련 데이터 세트의 윈도우 사이즈와 미래의 예측일 수를 조정해가면서 모델 간의 성능을 비교 분석했다.

4.2 데이터 구성 및 전처리

가격 데이터는 축산물품질평가원[23]에서 크롤링한 도매가격 데이터를 활용했다. 표 2에 기술한 내용처럼 가격 데이터는 2018년 1월 2일부터 2021년 12월 31일까지 1,004개의 일별 도매가격 데이터를 수집했다. 다만, 주말이나 공휴일과 같이 도매시장이 개장하지 않은 날에는 거래가 이루어지지 않아 결측치가 있었다. 안정적인 일별 가격 예측을 위해 결측치가 없는 모든 일 단위 도매가격이 필요했고

결측치가 있는 일별 도매가격은 직전일 도매가격으로 변경했다.

표 2. 수집 데이터 설명

Table 2. Details of collected data

	Source	Content	Amount	Date
Before	Pigtimes	News	1,375	2018.01.02. ~ 2021.12.31
	Chuksan news	News	1,896	
	Handonnews	News	1,017	
	Ekapepia	Wholesale price	1,004	
After	Day sentiment	Sentiment	1,460	
	Day wholesale price	wholesale price	1,460	

총 4,288개 뉴스 기사를 3.3장에서 기술한 바와 같이 일별 감정 점수로 측정했다. 하나의 날짜에 여러 개의 뉴스 기사가 게시된 경우, 해당 날짜의 뉴스 감정 점수들의 평균값을 계산하여 일별 감정 점수를 계산했다. 뉴스 기사가 게시되지 않은 날짜에는 이전 날짜의 감정 점수 값을 사용하여 대체했다. 측정된 감정 점수는 같은 일별로 도매가격 데이터와 하나의 데이터 세트로 합쳤다.

표 1. 뉴스 문장별 감정분석 결과 예시

Table 1. Examples of sentiment analysis results

Content	sentiment	confidence
이로써 지난 14년 이후 돼지 값은 연평균 4천500원 이상을 기록하며 호황을 이어갔다.	Positive	Negative: 0.0009 Positive: 0.9984 Neutral: 0.0005
돼지 출하두수 증가에도 소비가 늘면서 돈가 강세를 뒷받침한 결과다.	Positive	Negative: 0.3647 Positive: 0.3904 Neutral: 0.2447
농가 55% MSY 18두 미만한돈 하락 시 경영 악화 우려이유 후 육성에 관심 집중을양돈농가 중 절반이 넘는 농가가 MSY 18두 미만을 기록하고 있는 것으로 나타났다	Negative	Negative: 0.9986, Positive: 0.0005 Neutral: 0.0008
한국은행은 현재까지 제주로 반입된 다른 지역 한돈물량이 도내 도축물량의 1% 미만으로 그 영향이 미미한 상황이지만 향후 반입 물량이 확대될 경우 지역 양돈농가 채산성 악화 가능성이 있다고 전망했다.	Negative	Negative: 0.9988 Positive: 0.0004 Neutral: 0.0007
MSY의 경우 두당 섭취량 증가와 출하일령 감소로 19.1두를 기록, 연중 가장 높게 나타났으며, PSY도 21.6두로 나타났다.	Neutral	Negative: 0.2541 Positive: 0.0650 Neutral: 0.6807

일별 도매가격 데이터와 감정 점수 데이터의 기초 통계 정보는 표 3과 같다. 일별 도매가격 최댓값은 6,649원이고, 최솟값은 2,021원이다. 평균적으로 4,446.34원에서 도매가격이 형성된다는 것을 알 수 있다. 일별 감정 점수는 중앙값이 -0.05로 0을 기준으로 감정이 편향되지 않고 긍정, 부정으로 고르게 분포하고 있음을 확인할 수 있었다.

표 3. 도매가격과 감정 점수 기초 통계량
Table 3. Statistics on price and sentiment score

	Source	Day wholesale price	Day sentiment
	Content	wholesale price	Sentiment
Basic statistics	Amount	1,460	1,460
	Mean	4446.34	-0.1114
	Std	790.71	0.26
	Min	2021.00	-1.00
	Median	4628.50	-0.05
	Max	6649.00	1.00

4.3 모델 구조 및 학습

본 연구에서 활용한 시계열 딥러닝 모델은 LSTM으로 자세한 모델 구조는 표 4와 같다. 돈육 가격 데이터와 감정 점수를 인풋 사이즈가 2차원 형태로 모델에 입력한다. 그림 3과 같이 과거 w 일치 데이터를 학습하고 미래의 k 일치 데이터를 예측하는 방식으로 윈도우 사이즈를 나누어 학습했다.

훈련데이터와 검증 데이터를 8:2로 나누어 20%를 검증 데이터 세트로 사용했다. 에포크(Epoch)는

100, 배치 크기는 100으로 설정하여 학습했다.

같은 수의 레이어 층과 활성화 함수들을 사용하여 GRU 모델에 대해서도 똑같이 학습하여 두 모델의 성능을 비교평가 하는 방식으로 실험을 진행했다.

표 4. LSTM 모델 구조
Table 4. LSTM model structure

Type	Content
Input size	2
LSTM layer	1
Hidden layer size	10
Activation function	Tanh[30]
Optimizer	Adam optimizer[31]
Loss function	MeanSquaredError[32]
Output size	k(Predict day)

4.4 실험 결과

돈육 도매가격 예측 모델을 구축하는 데 있어 감정 점수만 활용한 경우, 도매가격 데이터만 활용한 경우, 감정 점수와 도매가격을 함께 활용한 경우의 예측 결과를 비교 분석했다.

앞 장에서 언급한 바와 같이 윈도우 사이즈로 나누어 학습했다. 최적의 학습일 수를 파악하기 위해 과거 학습 데이터 윈도우 크기를 2일부터 28일까지, 예측 기간을 2일부터 7일까지 다양하게 조합하여 모든 경우의 수에 대한 성능평가를 했다. 학습 윈도우 사이즈는 10일에서 20일 사이, 예측 일수는 2일에서 4일 사이에서 학습한 모델들이 성능이 높았다.

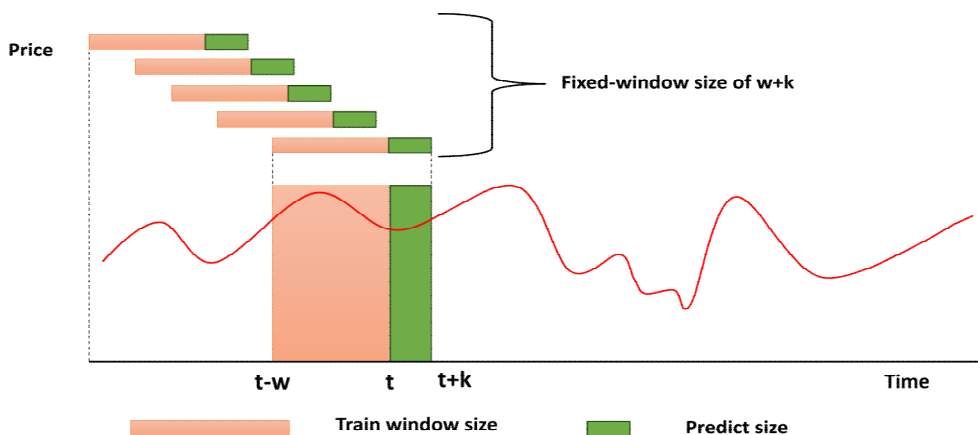


그림 3. Window size를 활용하여 학습 후 예측
Fig. 3. Using window size to predict after learning

위 언급한 세 가지 경우 중 도매가격과 함께 감정 점수를 사용하여 학습한 예측 모델이 다른 두 경우 (도매가격으로만 학습한 모델, 또는 감정 점수로만 학습한 모델)보다 정확한 예측 성능을 보였다. 표 5는 도매가격과 함께 감정 점수를 활용한 예측 모델에서 가장 성능이 높은 10가지를 제시한다.

그림 4는 세 경우 중 LSTM과 GRU 모델을 활용했을 때 평균 MAE 값이다. 감정 점수만 이용하였을 때 평균 MAE 값은 LSTM 743.79, GRU 757.72로 가장 오차율이 컸고, 그 뒤 도매가격으로만 예측하였을 때가 LSTM 375.99, GRU 403.07 도매가격과 감정 점수를 함께 예측하였을 때 LSTM 245.17, GRU 386.60으로 가장 정확한 성능을 보였다. 위의 결과를 바탕으로 각 단일 변수로만 학습하여 예측한 결과보다 감정 점수를 고려한 도매가격 추세 예측이 세 경우 가장 정확한 결과를 얻었다. 따라서 감정 점수가 도매가격의 추세 예측에 유의미한 보조 지표라는 것을 확인하였다.

가장 정확한 예측 모델은 LSTM으로 학습 기간 13일, 예측 기간 3일을 예측한 모델이다. 회귀모델 성능 평가지표인 MAE로 정확도를 확인해보면 216.87이다. 그림 5는 LSTM 모델, GRU 모델 예측값과 실제 값의 추세를 그린 그래프이다. 급격하게 값이 상승하거나 하락하면 예측값과 실제값의 차이가 벌어지는 경향을 보였으나 하락과 상승 추세는 전 기

간에 걸쳐 효과적으로 예측했다. 동 기간 모델별 오차값 차이는 그림 6, 그림 7 와 같다. 전반적으로 메인모델로 활용한 LSTM 모델 예측이 같은 시계열 딥러닝 모델 GRU보다 더 실제 가격과 일치하는 추세를 나타냈다.

돈육 도매가격 예측 모델에 감정 점수를 통합한 접근법이 도매가격 또는 감정 점수 단일 데이터만을 사용한 예측보다 예측 정확도를 향상하는 데 긍정적인 영향을 미친다는 결과를 확인할 수 있었다.

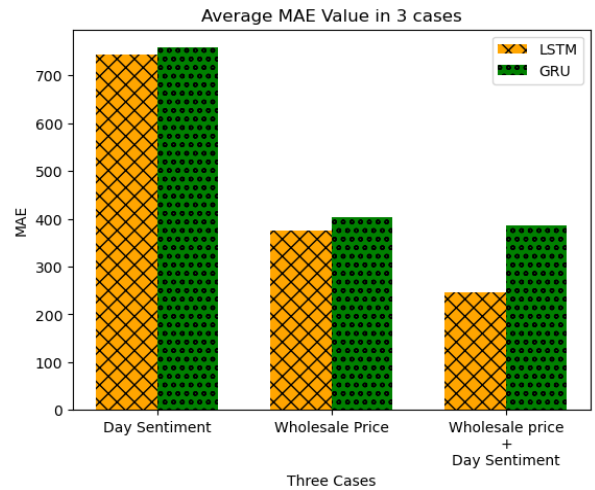


그림 4. 세 경우 중 각 LSTM, GRU 모델 MAE 평균값 비교

Fig. 4. Comparison of MAE between LSTM and GRU models in three cases

표 5. 가격 예측 오차 결과

Table 5. Results for pork price prediction error

Case	Window size		Errors											
	Train size	Predict size	Day sentiment				Wholesale price				Wholesale price+Day sentiment			
			LSTM		GRU		LSTM		GRU		LSTM		GRU	
			MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE
1	13	3	738.63	20.23	789.18	20.18	361.38	10.21	387.56	11.24	216.87	5.92	293.43	5.93
2	19	4	774.29	20.12	778.76	20.25	379.98	10.95	282.80	8.02	229.44	6.01	494.51	12.56
3	4	2	719.96	19.65	731.57	20.08	257.46	7.24	574.45	13.56	233.92	6.60	347.55	10.08
4	10	4	722.25	19.75	705.03	19.35	378.45	10.99	395.04	11.44	235.97	6.42	388.87	11.26
5	12	2	708.04	19.35	749.39	20.43	356.97	10.41	344.89	9.69	236.27	6.76	302.40	8.61
6	16	2	717.08	19.58	817.92	20.12	320.84	9.19	592.37	14.20	254.22	7.02	210.33	5.83
7	19	3	794.23	20.26	720.63	19.88	235.35	6.54	626.23	16.09	257.04	7.17	563.68	12.94
8	13	2	736.05	20.23	725.83	19.95	283.25	8.03	299.68	8.48	259.26	7.24	569.61	13.45
9	14	2	750.11	20.95	824.78	20.49	902.63	20.27	269.85	7.65	260.80	7.54	262.07	7.61
10	12	3	777.27	19.85	734.18	20.13	283.59	8.05	257.84	7.35	267.97	7.55	430.62	10.2
		Mean	743.79	19.99	757.72	20.08	375.99	10.18	403.07	10.77	245.17	6.82	386.30	9.84

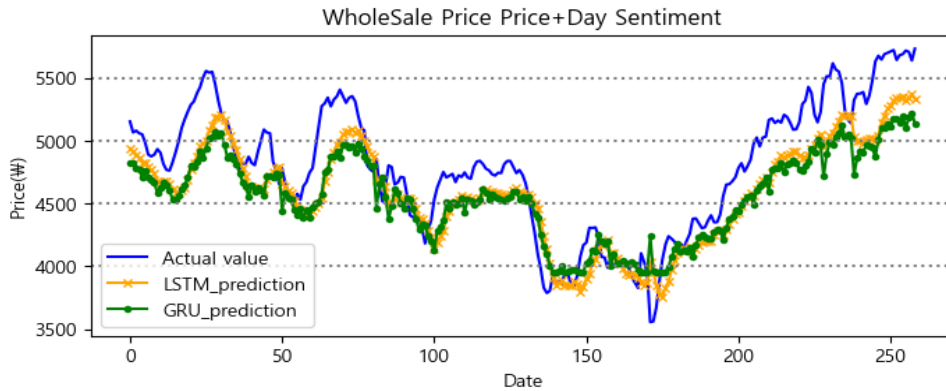


그림 5. 가격 예측 그래프
Fig. 5. Price prediction graph

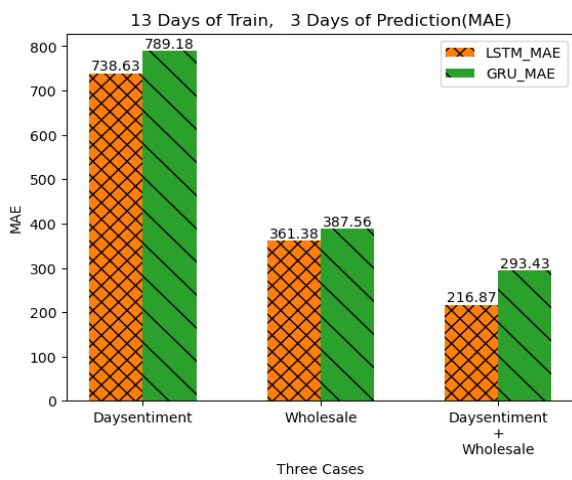


그림 6. MAE 결과(13일치 학습, 3일 예측 오차값 그래프)
Fig. 6. Results of MAE
(13 days of train, 3 days of prediction)

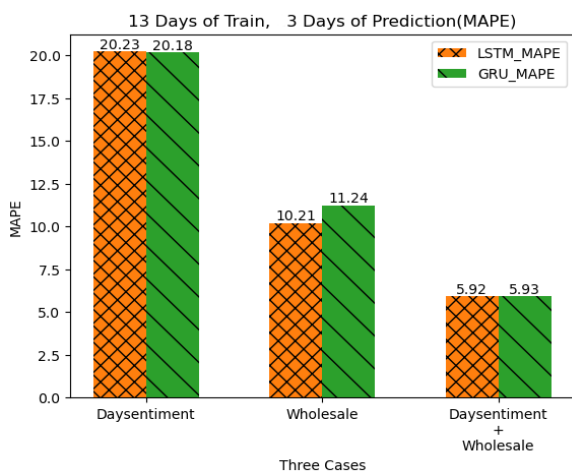


그림 7. MAPE 결과(13일치 학습, 3일 예측 오차값 그래프)
Fig. 7. Result of MAPE
(13 days of train, 3 days of prediction)

V. 결론 및 향후 과제

가격은 시장에서 수요와 공급으로 결정된다. 특히 축산 시장은 기온, 질병, 수입품목 등 여러 외부 요인에 의해 수요와 공급이 지속해서 변화하고 이는 가격 변동성으로 이어진다.

따라서 본 논문에서 정확한 가격 예측을 위해 뉴스 데이터의 감정을 분석하여 과거의 가격 데이터 뿐만 아니라 현 상황까지 고려한 돈육 가격 예측 모델을 제안하였다. 뉴스 데이터를 네이버 CLOVA 감정 분석기로 분류한 결과 높은 수준의 문장별 감정을 분석할 수 있었다. 이렇게 분류한 감정을 이용하여 일별 감정 점수를 계산하였고 돈육 도매가격 예측 모델에 함께 활용하였다.

실험은 다음과 같이 3가지 경우로 분류하여 성능 테스트를 진행하였다. 과거 도매가격 데이터만 이용하여 도매가격 예측, 감정 점수 데이터만 이용하여 도매가격 예측, 과거 도매가격과 감정 점수를 함께 활용하여 도매가격 예측 총 3가지 경우로 나누어 실험하였고 각각의 단일 변수만 활용했을 때 보다 도매가격과 감정 점수를 함께 이용하여 예측하였을 때 정확한 예측값을 얻을 수 있었다. 과거의 도매가격만을 활용하여 미래의 도매가격을 예측하는 것보다 감정 점수를 함께 활용하여 도매가격을 예측하는 것이 돈육가격의 주기성 변화뿐만 아니라 현 축산 업황까지 고려하여 정확한 추세 예측을 할 수 있었다.

본 논문에서는 전통적인 과거의 가격 패턴으로 미래의 가격을 예측하는 것에서 더 나아가 뉴스 데이터의 감정을 분석하여 현 추산 시장의 상황까지 고려한 가격 예측이라는 점에서 의의를 찾을 수 있다. 그러나 돈육 가격 특성상 계절성 변화가 있고 뉴스 감정 점수를 일괄적으로 일별 평균으로 계산하였다는 점에서 한계가 있었다. 따라서 향후 연구에서는 이러한 한계점을 극복하기 위한 연구를 본 논문의 향후 과제로 한다.

References

- [1] M. K. Jeong, H. J. Kim, and H. W. Lee, "Consumer Behavior for Meat Consumption and Tasks to Respond to Its Changes", *KREI*, pp. 1-240, Oct. 2020.
- [2] K. Sevinch, H. Bahrom, and S. Sirojiddin, "Theory of Supply and Demand. Market Equilibrium", *Pedagogical Sciences and Teaching Methods*, Vol. 2, No. 18, pp. 253-257, Dec. 2022.
- [3] H. N. Kim and I.-C. Choi, "The Economic Impact of Government Policy on Market Prices of Low-Fat Pork in South Korea: A Quasi-Experimental Hedonic Price Approach", *Sustainability*, Vol. 10, No. 3 Mar. 2018. <https://doi.org/10.3390/su10030892>.
- [4] D. Vandone, M. Peri, L. Baldi, and A. Tanda, "The impact of energy and agriculture prices on the stock performance of the water industry", *Water Resources and Economics*, Vol. 23, pp. 14-27, Jul. 2018. <https://doi.org/10.1016/j.wre.2018.02.002>.
- [5] T. N. Vu, C. M. Ho, T. C. Nguyen, and D. H. Vo, "The Determinants of Risk Transmission between Oil and Agricultural Prices: An IPVAR Approach", *Agriculture*, Vol. 10, No. 4, pp. 120, Apr. 2020. <https://doi.org/10.3390/agriculture10040120>.
- [6] M. İ. Y. Kaya and M. E. Karşilgil, "Stock price prediction using financial news articles", 2010 2nd IEEE International Conference on Information and Financial Engineering, Chongqing, pp. 478-482, Sep. 2010. <https://doi.org/10.1109/ICIFE.2010.5609404>.
- [7] A. Hotho, A. Nürnberger, and G. Paass, "A brief survey of text mining", *Journal for Language Technology and Computational Linguistics*, Vol. 20, No. 1, pp. 19-62, Jul. 2005. <http://dx.doi.org/10.21248/jlcl.20.2005.68>.
- [8] H. Maqsood, et al., "A local and global event sentiment based efficient stock exchange forecasting using deep learning", *International Journal of Information Management*, Vol. 50, pp. 432-451, Feb. 2020. <https://doi.org/10.1016/j.ijinfomgt.2019.07.011>.
- [9] B. Wu, L. Wang, S.-X. Lv, and Y.-R. Zeng, "Effective crude oil price forecasting using new text-based and big-data-driven model", *Measurement*, Vol. 168, pp. 108468, Jan. 2021. <https://doi.org/10.1016/j.measurement.2020.108468>.
- [10] T. Chuluunsaikhan, G.-A. Ryu, K.-H. Yoo, H. C. Rah, and A. Nasridinov, "Incorporating deep learning and news topic modeling for forecasting pork prices: the case of South Korea", *Agriculture*, Vol. 10, No. 11, pp. 513, Oct. 2020. <https://doi.org/10.3390/agriculture10110513>.
- [11] M. N. Ashtiani and B. Raahmei, "News-based intelligent prediction of financial markets using text mining and machine learning: A systematic literature review", *Expert Systems with Applications*, Vol. 217, pp. 119509, May 2023. <https://doi.org/10.1016/j.eswa.2023.119509>.
- [12] Y. Ding, P. Wu, J. Zhao, and L. Zhou, "Forecasting product sales using text mining: a case study in new energy vehicle", *Electron Commer Res*, May 2023. <https://doi.org/10.1007/s10660-023-09701-9>.
- [13] R. C. Staudemeyer and E. R. Morris, "Understanding LSTM—a tutorial into long short-term memory recurrent neural networks", *arXiv preprint arXiv:1909.09586*, Sep. 2019. <https://doi.org/10.48550/arXiv.1909.09586>.

- [14] B. Garcia, "Hands-On Selenium WebDriver with Java", O'Reilly Media, 2022. <https://www.oreilly.com/catalog/errata.csp?isbn=9781098110000>.
- [15] S. Baccianella, A. Esuli, and F. Sebastiani, "Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining", Proc. of the Seventh International Conference on Language Resources and Evaluation (LREC'10), Valletta, Malta, Vol. 10, No. 2010, May 2010.
- [16] L. Alzubaidi, et al., "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions", Journal of Big Data, Vol. 8, No. 53, pp. 1-74, Mar. 2021. <https://doi.org/10.1186/s40537-021-00444-8>.
- [17] S. Borgeaud, et al., "Improving language models by retrieving from trillions of tokens", In International conference on machine learning, pp. 2206-2240, Dec. 2021. <https://doi.org/10.48550/arXiv.2112.04426>.
- [18] Pigtimes, www.pigtimes.co.kr/news/articleList.html?sc_section_code=S1N7&view_type=sm [accessed: May 19, 2023]
- [19] Chuksannews, www.chuksannews.co.kr/news/section.html?sec_no=83 [accessed: May 19, 2023]
- [20] Handonnews, www.handonnews.kr/news/article_list_all.html [accessed: May 19, 2023]
- [21] L. Richardson, "Beautiful soup documentation", 2007.
- [22] B. Garcia, "Hands-On Selenium WebDriver with Java", O'Reilly Media Inc, Mar. 2022.
- [23] Ekapepia, <https://www.ekapepia.com/index.do> [accessed: May 19, 2023]
- [24] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges", Artificial Intelligence Review, Vol. 55, No. 7, pp. 5731-5780, Sep. 2022. <https://doi.org/10.1007/s10462-022-10144-1>.
- [25] Clova Sentiment, Naver Cloud Platform, <https://www.ncloud.com/product/aiService/clovaSentiment> [accessed: May 21, 2023]
- [26] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding", arXiv, preprint arXiv:1810.04805, May 2019. <https://doi.org/10.48550/arXiv.1810.04805>.
- [27] H. J. Choi, W. Khern-am-nuai, K. Han, and H. So, "Economic Implications of Rating-Sentiment Inconsistency in Online Reviews and the Role of User-Generated Photos", pp. 1760, Aug. 2023. https://aisel.aisnet.org/amcis2023/sig_dsa/sig_dsa/12.
- [28] J. Hur and J. Yang, "South Korean newspaper coverage of Yemeni refugees: analysis of topics and sentiments using machine learning techniques", Asian Journal of Communication, pp. 1-16, Sep. 2023. <https://doi.org/10.1080/01292986.2023.2257230>.
- [29] G. Shen, Q. Tan, H. Zhang, P. Zeng, and J. Xu, "Deep learning with gated recurrent unit networks for financial sequence predictions", Procedia computer science, Vol. 131, pp 895-903, Apr. 2018. <https://doi.org/10.1016/j.procs.2018.04.298>.
- [30] S. R. Dubey, S. K. Singh, and B. B. Chaudhuri, "Activation functions in deep learning: A comprehensive survey and benchmark", Neurocomputing, Vol. 503, pp. 92-108, Jun. 2022. <https://doi.org/10.1016/j.neucom.2022.06.111>.
- [31] K. K. Chandriah and R. V. Naraganahalli, "RNN/LSTM with modified Adam optimizer in deep learning approach for automobile spare parts demand forecasting", Multimedia Tools and Applications, Vol. 80, No. 17, pp. 26145-26159, Jul. 2021. <https://doi.org/10.1007/s11042-021-10913-0>.
- [32] T. O. Hodson, T. M. Over, and S. S. Foks, "Mean squared error, deconstructed", Journal of Advances in Modeling earth Systems, Vol. 13, No. 12, pp. 1942-2466, Nov. 2021. <https://doi.org/10.1029/2021MS002681>.

저자소개

정 재 민 (Jaemin Jeong)



2018년 3월 ~ 현재 : 충북대학교
소프트웨어학과 학사과정
관심분야 : 데이터 마이닝,
빅데이터 분석, AI 모델링

정 경 창 (Kyungchang Jeong)



2022년 2월 : 충북대학교
소프트웨어학과(공학사)
2022년 3월 ~ 현재 : 충북대학교
컴퓨터과학과 석박사통합과정
관심분야 : 빅데이터 분석, 데이터
기반 시스템

고 은 영 (Eunyoung Ko)



2000년 2월 : 삼육대학교
응용동물학과 졸업(축산학사)
2014년 2월 : 강원대학교 수의학과
수의병리학(석사)
2018년 2월 : 충남대 축산학과
축산식육 대학원(축산식육박사)
1999년 10월 ~ 현재 :

도드람양돈농협 실장, 품질관리본부 CQO
관심분야 : AutofomIII, 돈육 품질, 돈육 맛분석, 돈가 예측

이 의 종 (Euijong Lee)



2012년 2월 : 고려대학교
컴퓨터정보학과(이학사)
2018년 8월 : 고려대학교
컴퓨터공학과(공학박사)
2020년 9월 ~ 현재 : 충북대학교
소프트웨어학부 조교수
관심분야 : 소프트웨어 공학,
사물인터넷, 빅데이터 분석, 인공지능 응용