



장려상

딥러닝 기반 감성분석을 이용한 양파수급모형

조수민 · 오정은 · 백종현 · 순병민

딥러닝 기반 감성분석을 이용한 양파수급모형

조수민* 오정은** 백종현*** 순병민****

Key words

감성분석(sentiment analysis), 딥러닝(deep learning), 양파수급모형(onion model)

ABSTRACT

Our study examines how news related to onion affect producer behavior for the harvested area. To do that, we create sentiment index related to Korean onion using LSTM-CNN (Long Sequence Term Memory-Convolution Neural Network). We collect news in June, July, and August before planing onion to make sentiment index. Using these sentiment index during that periods, these sentiment index helps producers plant onion or other agricultural products. We then impose these sentiment index into the harvested area function for onion. Finally, we build the onion supply-demand model to examine the impact of change in sentiment index on supply and demand. To the best of our knowledge, our approach using sentiment index in the agricultural model is a first trial. Therefore, our study can introduce an approach to improve the accuracy of modeling for agriculture.

차례

- | | |
|------------|-------------|
| 1. 서론 | 4. 자료 |
| 2. 선행연구 분석 | 5. 분석 결과 |
| 3. 연구 방법 | 6. 결론 및 시사점 |

* 충남대학교 농업경제학과 학부생, e-mail: sumin_official@naver.com

** 충남대학교 농업경제학과 학부생

*** 충남대학교 컴퓨터공학과 학부생

**** 충남대학교 농업경제학과 조교수, e-mail: soonbm@cnu.ac.kr

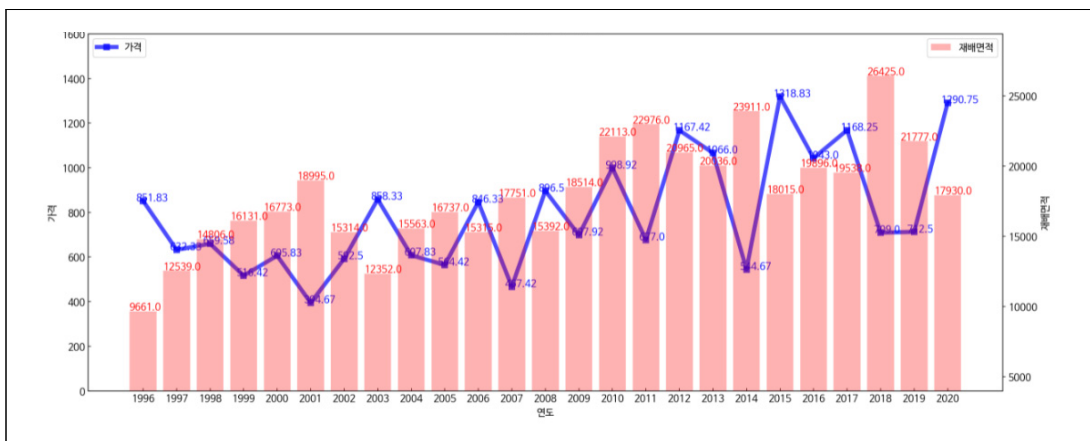
1. 서론

1.1. 연구배경 및 필요성

양파는 배추, 무 등과 함께 국내 5대 민감 채소로 지정된 수급 및 가격 불안정성이 큰 품목이다. 연중 생산이 불가능한 특성상 저장을 통해 수급량을 조절하며, 생산량 변동이 심한 해에는 부분적으로 수출과 수입을 통해 균형을 유지하고 있다.

〈그림 1〉은 1996년부터 2020년까지의 양파 가격과 재배면적 변동추이를 나타낸 그래프이다. 직전 해의 양파 가격이 상승하면 당해 양파 재배면적이 증가하고, 재배면적 증가로 양파 생산량이 증가하면 과잉공급으로 다음 기의 가격이 폭락하기를 반복하며 변동성이 커 전망이 어려운 양상을 보인다. 2018년에는 생산량 예측 실패로 과잉 생산된 양파를 산지 폐기한 규모는 95억 9,900만 원(3만 6,386톤)에 달했다. 농림축산식품부에서는 매년 한국농촌경제연구원과 통계청이 생산량 및 재배면적 전망치를 통해 수급 조절에 나서고 있는데, 농촌경제연구원에서는 양파 재배면적이 18.3% 증가할 것으로 전망한 반면 통계청은 35.2% 증가할 것으로 전망해 농림부의 수급 조절 대책이 차질을 빚은 것이다.

그림 1. 양파 재배면적, 가격 변동 추이



자료: 농산물유통정보(KAMIS), 농업관측통계시스템(OASIS)

양파 생산량 파악을 위해서는 재배면적과 단수 정보가 중요하나, 대체로 단수보다는 재배면적이 생산량에 더 큰 영향을 주는 것으로 알려져 있다(이승인, 2020). 따라서 재배면적 의사결정은 가격 안정과 생산량 전망의 중요한 요인이며, 이를 반영하기 위해 농업관측본부에서 매달 관측 월보를 통해 재배의향면적 조사를 발간하고 있다. 하지만 월 단위 조사로 실시간 정보 수집이 어려우며 비용 문제로 인해 표본농가와 지역 모니터 조사가 이루어진다는 한계로 재배면적 의사결정의 불확실성 문제를 완전히 해결하지 못하고 있다. 본 연구에서는 양파 수급모형에 농가의 재배면적 의사결정을 보다 정밀하게 반영하기 위해 감성분석을 활용한 감성지수를 재배면적 함수의 설명변수로 도입하고, 양파 가격 및 생산량 전망을 정교화하고자 한다.

1.2. 연구 목적

본 연구는 텍스트 마이닝 기법의 하나인 감성분석을 활용해 양파 관련 감성지수를 생성하고, 양파 재배면적을 함수의 설명변수로 도입해 양파 관련 감성 변화가 양파 수급에 미치는 영향을 분석한다. 양파 파종기(8월 말)에 근접할수록 감성지수가 양파 재배면적 결정에 미치는 영향이 크다는 가정 하에 6월, 7월, 8월, 6·7월, 7·8월, 6·7·8월로 기간을 나누어 재배면적 함수에 감성지수를 포함하여 추정한다. 포함된 감성지수에 따른 각 모형의 추정값과 실제값을 비교하기 위해 평균 제곱근 오차(RMSE)를 계산, 가장 RMSE가 작은 함수를 활용하여 양파 수급모형을 구축한 뒤, 감성지수 등락에 따른 시나리오 분석을 실시하여 감성지수 변화가 양파 재배면적에 미치는 영향을 분석한다. 본 연구는 양파 수급 모형을 정교화함으로써 농가와 정책 담당자의 합리적인 의사결정에 도움을 주고자 하며, 더 나아가 감성분석과 같은 빅데이터 분석의 농업 분야에서의 활용 가능성을 보이고자 한다.

2. 선행연구 분석

양파 재배면적과 관련된 국내 연구로 류재현(2020)은 OLS 분석을 통해 농업관측본부가 제공하는 재배면적 전망치가 농민의 재배면적 조절에 유의한 영향을 미친다는 사실을 보였지만, 농가 특성을 설명변수에 사용하지 않은 자료 제약의 문제로 양파 도매가격과 금년도 후기 마늘 재배면적이 비례하는 결과가 도출되었다는 한계가 있다. 하지희(2019)는 양파와 마늘 가격 예측을 위해 이분산성과 월별 더미, 물량자료를 반영하여 ARMGARCH(36개월)이 가장 적합한 시계열 모형임을 추정하였다. 연구결과, 마늘 수입량은 마늘 가격에 유의한 영향을 미쳤지만, 양파의 경우 양파 가격에 영향을 미치는 적합한 외생변수를 모색하지 못한 한계가 있다.

최근 스마트폰의 보급과 SNS의 활성화로 수많은 정보들이 생성되었고, 비정형 데이터 처리 및 활용의 중요성이 커지고 있는 흐름 속에서 금융, 부동산 등 다양한 분야에서 빅데이터를 활용하여 기존과는 다른 새로운 분석 방법과 연구결과를 제시하는 시도가 많이 이루어지고 있다. 차은정(2016)은 뉴스 정보를 수집, 감성사전(SentiWordNet)에 대입하여 뉴스가 가진 감성값을 확인하고 감성정보가 실제로 주식 시장에 미치는 영향을 SVM과 로짓분석을 이용해 추정하였고, 그 결과 S&P500의 예측 정확도가 4% 향상된다는 결론을 도출했다. 하지만 하나의 언론사 뉴스만을 이용해 다소 국부적이며, 여러 다양한 외부적 요소들을 포함하지 않아 복잡한 주식 시장을 반영하기 힘들다는 한계가 있다. 박재수(2019)는 주요 온라인 신문기사에서 추출한 아파트 매매와 관련된 긍정 감정지수와 아파트 매매가격과의 연관 관계를 알아보기 위해 교차상관 및 인과관계 분석을 시행하였고 서울시와 모든 권역에서 긍정 감성지수는 3개월 이내의 선행성을 보이며 소형 아파트 매매가격과 유의미한 교차 상관관계를 보임을 확인하였다. 하지만 2008년 이후의 자료들만을 대상으로 해 보합세와 상승기만 포함하였고, 심리지표인 감성지수만을 포함했기 때문에 경제변수를 포함시키지 못했다는 한계점이 있다.

또한 농업 분야에서도 빅데이터 활용의 중요성이 증가함에 따라 비정형 데이터와 기존에 사용되어온 경제변수 간의 관계를 분석하는 연구가 이루어지고 있다. 강산하(2017)는 웹 크

롤링 방식을 통해 텍스트 마이닝을 시행해 부정기사 수와 '양파 가격' 검색 빈도는 재배면적과 반비례하며 긍/부정 단어에 의해 양파 가격이 영향을 받음을 시사하였다. 하지만, 심도있는 모형에 적용한 것이 아닌 단순 상관 관계분석을 시행하였다는 한계점이 있다. 나형철(2018)은 방송 뉴스 및 소비자 패널 데이터 등과 같은 비정형 자료들을 변수로 도입하여 양파 소비와 일정 시차를 두고 양의 상관관계가 있음을 확인하였다. 하지만 수요 예측 모형과 예측력을 제시하지 하지 않아 양파 관련 정보가 양파 소비에 미친 영향을 심도있게 분석하지 못하였으며, 양파 가격이 폭락하였던 2014년의 자료만을 사용했기 때문에 결과 검증을 위해 추가적인 연구가 필요하다는 한계가 있다.

양파 생산량 및 가격 전망의 정확도 향상을 위해서는 여러 경제적 요인과 더불어 양파 재배면적 의사결정에 영향을 미칠 수 있는 새로운 요인들을 반영한 확장된 연구가 필요하나 이와 관련한 연구는 아직 부족한 실정이다. 본 연구에서는 타 분야에서 비교적 활발하게 활용되고 있는 비정형 데이터인 감성지수를 농업 분야에 도입해 재배면적 함수를 도출하고, 수급 균형모형을 통해 생산량 및 가격 전망과 시나리오 분석을 시행하였다는 차별점이 있다.

3. 연구 방법

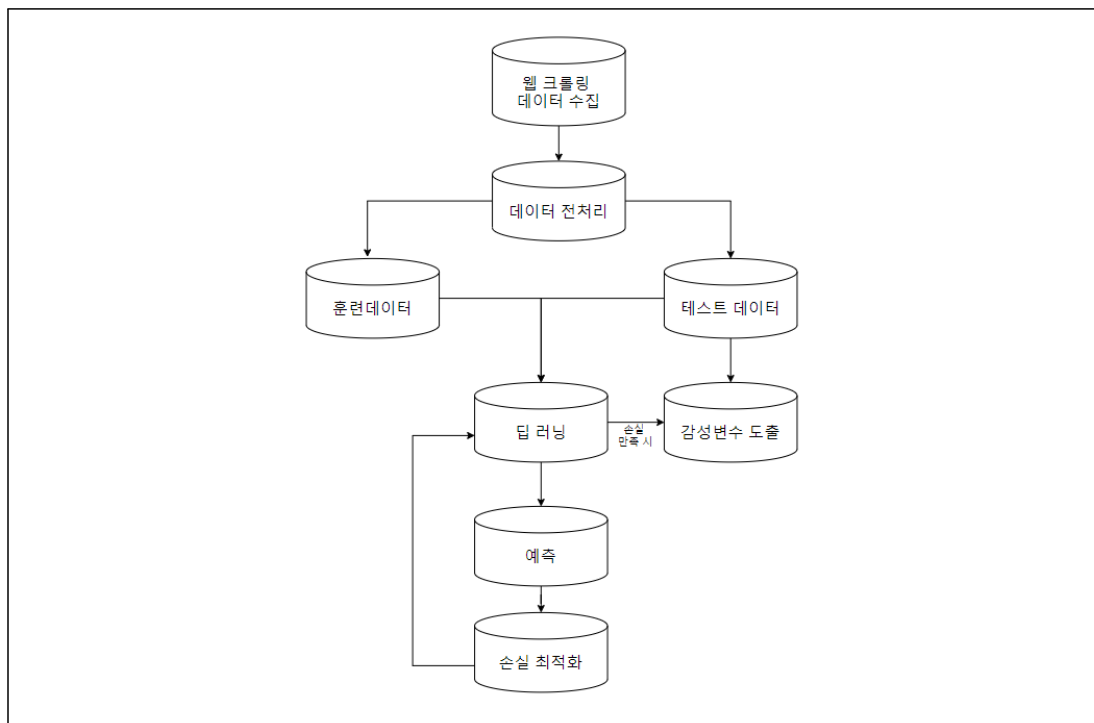
3.1. 감성분석

기존 선행연구에서는 군집 알고리즘을 이용한 감성사전 구축 방식을 활용하고, 웹 크롤링 기술로 문서 데이터를 수집하였다. 수집한 데이터를 바탕으로 TF-IDF(Term Frequency-Inverse Document Frequency)라는 문서별 단어 빈도수를 통하여 가격 문서, 재배면적 문서 등을 파악하여 빈도수를 계산한 뒤 각 빈도수 마다 유사도를 측정했다. 유사도 값들을 바탕으로 K-means, DBSCAN의 클러스터링 알고리즘을 적용했다. 적용한 알고리즘을 바탕으로 군집을 형성하여 긍정, 부정, 무정의 비율 값을 측정하여 감성사전을 구축하는 비지도 학습 방식이다.

본 연구에서는 기존 연구와의 차별성을 두기 위해 입력층, 은닉층, 결과층으로 나누어 각 층을 거쳐서 학습하는 지도학습 방식인 신경망 기반으로 감성지수를 도출한다. 입력층은 데이터를 넣는 역할을 하며, 은닉층은 가중치 벡터와 편향 벡터를 통해 층을 통과하면서 가중치와 편향을 학습하는 역할, 결과층은 학습된 가중치와 편향 벡터를 데이터에 적용하여 결과를 출력하는 역할을 한다. 신경망을 통해 나온 결과는 0~1 사이의 실수 형태로 적용하여 각 기사마다 감성지수를 도출한다.

감성분석은 <그림 2>와 같은 방식으로 감성의 수치화된 값을 얻는다. 인터넷 기사 데이터를 수집하는 웹 크롤링 방식으로 감성지수 데이터를 수집하고, 데이터 수집 기간은 연구 범위에 맞게 1996년부터 2020년까지 양파와 관련된 기사들로 선정하기로 한다. 수집한 기사들을 바탕으로 제목과 내용을 따로 저장하여 각각 데이터들의 신경망이 인식하지 못하는 문자나 다른 특수기호를 삭제하는 전처리 과정을 거친다.

그림 2. 감성지수 도출 과정

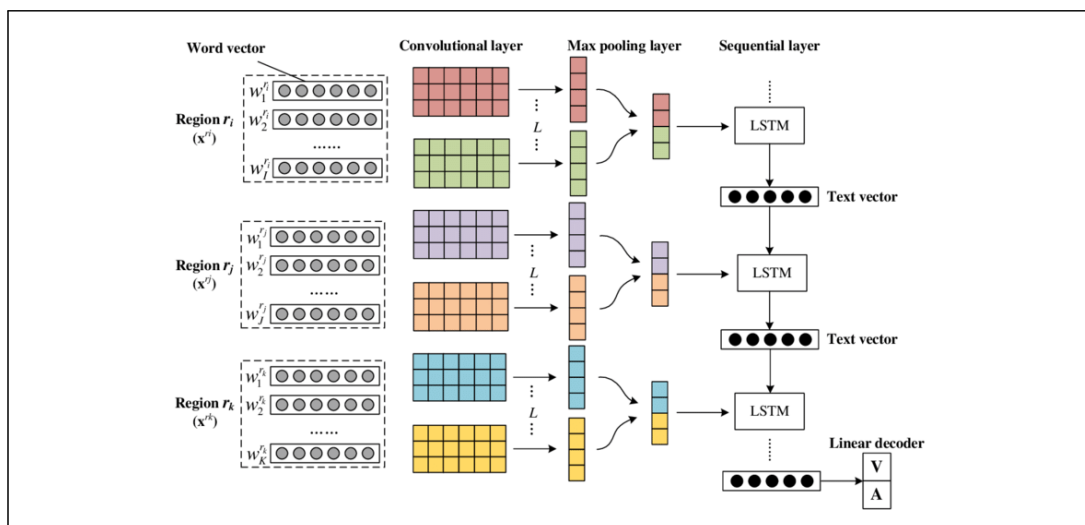


자료: 저자 작성

전처리 과정을 거친 제목과 기사는 신경망이 학습할 수 있게 숫자 형태의 데이터로 변환되며, 이를 위해 단어사전을 구축한다. 단어사전은 데이터에 포함된 단어를 기반으로 구축되며, 해당하는 단어의 키워드를 통해 문자열 형태의 데이터를 실수화한 숫자로 변환한다. 이때 신경망 훈련을 위해 전체 데이터를 훈련 데이터와 테스트 데이터로 분류하게 되는데 훈련 데이터는 모델의 가중치와 편향을 학습하는 역할을 하고, 테스트 데이터는 모델이 훈련 데이터의 과적합 현상을 막는 역할을 한다.

모델은 자연어 처리 딥러닝에서 많이 사용하는 LSTM-CNN(Long Sequence Term Memory-Convolution Neural Network) 모델을 이용한다. LSTM은 장기 기억 메모리 셀 계층으로, 모델의 학습시간이 길수록 처음에 학습했던 내용을 기억하지 못하는 경우가 잦지만, LSTM은 처음에 학습했던 내용까지 셀에 저장하여 가중치가 모든 학습의 경우를 반영하여 학습한다는 장점이 있다. CNN은 합성곱 신경망을 뜻하며, <그림 3>과 같이 각 계층마다 필터를 적용하여 범위를 축약한 값을 나타내는 형태이다. 신경망 모델을 구축하였으면 학습률과 훈련 횟수 등 세밀한 매개변수를 통해서 학습한 결과를 예측하며 예측한 값과 실제 값의 오차를 최적화하여 모델을 반복적으로 학습한다. 모델 학습 중 오차 범위 안에 들어오면 학습을 중지하고 모델을 통하여 감성지수 값을 도출한다.

그림 3. LSTM-CNN 모델



자료: Jin 외(2016)

3.2. 감성지수를 도입한 재배면적 함수의 추정

농가의 양파 재배면적 결정에 양파의 직전 해 실제 가격(P_{Ot-1}), 대체 작목인 마늘의 직전 해 실제 가격(P_{Gt-1}), 당해의 양파 농가 경영비(C_t), 당해의 데이터 감성분석으로 도출된 양파 관련 감성지수(SI_t)와 기술 발전을 대표하는 대리변수(proxy variable)인 추세변수($trend$)가 영향을 주는 것으로 가정했다. 이때 양파 기대가격, 마늘 기대가격은 직전 해의 실제 가격이 당해의 기대가격이라는 단순기대($P_t^* = P_{t-1}$)를 가정한다.

농가는 6월에는 주로 마늘 수확작업과 모내기 등에 노동력을 집중시키며, 여름작물 파종이 어느 정도 마무리된 다음인 8월 말에 양파 재배면적을 결정한다. 이는 9월 중/하순에 이루어지는 양파 파종을 위한 종자와 모판 등의 준비를 보름 정도 앞둔 시점이다(이중웅, 1995). 따라서 월별로 분석된 감성지수(SI)는 양파의 파종기를 기점으로 주된 의사결정이 이루어지는 8월에 근접할수록 양파 재배면적 의사결정에 더 유의미한 영향을 미칠 것으로 가정해 <표 1>과 같이 재배면적 함수에 반영되었다. 모형 1, 2, 3에서는 각각 6월(SI_6), 7월(SI_7), 8월(SI_8)의 감성지수가 포함되었으며, 모형 4에서 6·7월의 감성지수(SI_{67}), 모형 5에서 7·8월의 감성지수(SI_{78}), 모형 6에서 7·8·9월의 감성지수(SI_{678})가 포함되었다. 이렇게 월별 데이터가 둘 이상 포함된 변수는 파종기(8월 말)에 가까울수록 가중치를 두어 가까운 기간의 데이터가 재배면적 결정에 더 큰 영향을 미치는 점을 반영했다.

표 1. 도입된 감성지수 기간별 재배면적 함수

모형	다중회귀분석식
모형 1	(1.1) $A_t = \beta_0 + \beta_1 P_{Ot-1} + \beta_2 P_{Gt-1} + \beta_3 A_{t-1} + \beta_4 C_t + \beta_5 SI_6 + \beta_6 trend + e_t$
모형 2	(1.2) $A_t = \beta_0 + \beta_1 P_{Ot-1} + \beta_2 P_{Gt-1} + \beta_3 A_{t-1} + \beta_4 C_t + \beta_5 SI_7 + \beta_6 trend + e_t$
모형 3	(1.3) $A_t = \beta_0 + \beta_1 P_{Ot-1} + \beta_2 P_{Gt-1} + \beta_3 A_{t-1} + \beta_4 C_t + \beta_5 SI_8 + \beta_6 trend + e_t$
모형 4	(1.4) $A_t = \beta_0 + \beta_1 P_{Ot-1} + \beta_2 P_{Gt-1} + \beta_3 A_{t-1} + \beta_4 C_t + \beta_5 SI_{67} + \beta_6 trend + e_t$
모형 5	(1.5) $A_t = \beta_0 + \beta_1 P_{Ot-1} + \beta_2 P_{Gt-1} + \beta_3 A_{t-1} + \beta_4 C_t + \beta_5 SI_{78} + \beta_6 trend + e_t$
모형 6	(1.6) $A_t = \beta_0 + \beta_1 P_{Ot-1} + \beta_2 P_{Gt-1} + \beta_3 A_{t-1} + \beta_4 C_t + \beta_5 SI_{678} + \beta_6 trend + e_t$

주: $SI_{67} = SI_6 \times 0.4 + SI_7 \times 0.6$, $SI_{78} = SI_7 \times 0.4 + SI_8 \times 0.6$, $SI_{78} = SI_6 \times 0.1 + SI_7 \times 0.3 + SI_8 \times 0.6$

3.3. 시장 균형

본 연구에서는 양파 수급모형을 작성하기 위해서 <표 2>에 제시된 행태방정식을 활용했다. 일반적으로 단수함수의 추정에는 기상 여건(평균온도, 최저온도, 강수량, 일조시간, 풍속 등)과 생육 정보, 기술변수의 함수로 이루어지나, 모형의 단순화를 위해 식 (2)와 같이 기술 발전을 대표하는 대리변수(proxy variable)로 하여 추세변수(*trend*)만을 사용했다. 구축된 단수함수에서는 다른 내생변수를 포함하지 않아 수급 균형 모형 시스템에 영향을 주지 않는다. 단년 생 작물의 생산량은 식 (3)과 같이 재배면적과 단수의 곱으로 도출되며, 계산된 총생산량과 이 전재고량, 수입량의 합이 당해의 총공급량이 된다. 소비함수는 식 (5)와 같이 당해의 양파 가격, 마늘 가격을 포함했고 1인당 실질 GDP를 통해 1인당 소비량을 계산했다. 총소비량은 계산된 1인당 소비량에 해당 년도의 인구수를 곱함으로써 도출되며, 총수요량은 식 (6)과 같이 총소비량과 기말재고량, 수출량의 합이다.

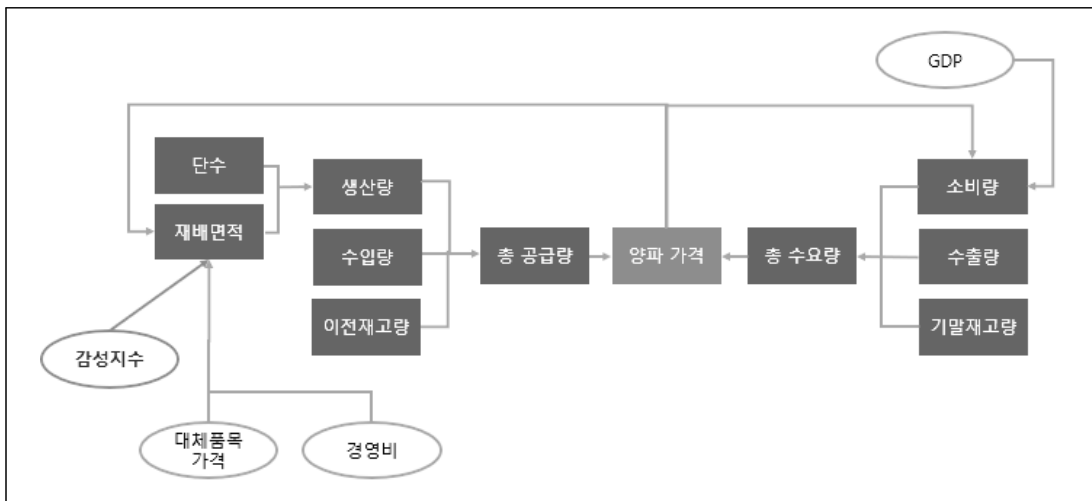
이렇게 구성된 균형 모형은 양파에 관한 단일품목 모형으로 마늘 도매가격, 이전재고량, 수입량, 기말재고량, 수출량, 1인당 GDP, 인구수는 모형 내에서 결정되지 않는 외생변수로 가정한다. 미래 전망을 위한 시나리오 분석 시에는 마늘 도매가격, 이전재고량, 수입량, 기말재고량, 수출량, 1인당 명목 GDP, GDP 디플레이터는 전년도 5년 이동평균으로 계산하고, 인구는 통계청의 장래인구추계 자료를 활용하기로 한다.

표 2. 수급 균형 모형 도출을 위한 함수

구분		행태방정식
공급	단수함수	(2) $YD = f(trend)$
	총생산함수	(3) $YD_t \times A_t$
	총공급량	(4) 총생산 + 이전재고량 + 수입량
수요	소비함수	(5) $C = f(P_{ot}, P_{gt}, GDP_t)$
	총수요량	(6) 총소비 + 기말재고량 + 수출량
시장청산		(7) 총공급(S) = 총수요(D)

〈그림 4〉는 양파의 수급 구조를 나타낸다. 총공급량(s)과 총수요량(D)이 일치할 때, 즉 초과공급($S - D$) 또는 초과수요($D - S$)가 0이 될 때 시장청산(market clearing)이 이루어지고 양파의 시장균형가격(equilibrium price)이 도출된다.

그림 4. 양파의 수급 구조



자료: 저자 작성

시장균형가격은 식 (8)처럼 P_{i-1} 이 높게 설정되었다면 공급은 증가하고 수요는 감소해 초과공급($S - D$)은 양의 값을 취하고, 조정계수(equilibrator) 앞의 음의 부호 때문에 P_i 는 P_{i-1} 보다 낮아져 초과공급($S - D$)이 0이 될 때까지의 반복(iteration)과정을 통해 도출된다.

$$(8) P_i = P_{i-1} - \delta(S_i - D_i)$$

i : 반복횟수, P : price, S : supply, D :demand, δ : equilibrator

4. 자료

4.1. 정형데이터

재배면적 함수 분석 자료로 농산물유통정보(KAMIS)에서 제공하는 양파, 마늘 도매가격 자료와 통계청에서 제공하는 양파 재배면적, 농업경영종합정보시스템에서 제공하는 농축산물 소득자료집의 농가 경영비를 활용했다. 양파 가격은 양파 상품 도매가격(원/kg)을 사용하였고, 마늘 가격은 난지 마늘 상품 도매가격(원/kg)을 사용하였다. 단수는 농업관측통계시스템 자료를 활용했고, 수요함수 추정을 위해서는 농업관측본부에서 매년 발간되는 농업전망 자료집에서 확인된 양파의 1인당 소비량, 농산물유통정보(KAMIS)에서 제공하는 양파·마늘 도매가격 자료와 통계청의 1인당 GDP를 활용했다. 도매가격 자료는 재배면적함수와 소비함수에서 각각 통계청에서 제공하는 생산자물가지수, 소비자물가지수를 통해 실질가격으로 디플레이트했다. 활용된 자료의 기초통계량은 <표 3>에 제시되어 있다.

표 3. 분석모형 정형 데이터 기초통계량

변수(단위)	기간	평균	표준편차
양파 재배면적(ha)	1997~2020년	18,324	3522
양파 단수(kg/10a)	1997~2020년	6267.13	585.12
양파 도매가격(원/1kg)	1997년 4월~2020년 3월	752.76	348.28
마늘 도매가격(원/1kg)	1997년 6월~2020년 5월	3122.076	1583.55
양파 농가 경영비(원)	1997~2020년	1,055,424	513473.41
1인당 GDP(만 원)	1997~2020년	2440.63	862.05
1인당 소비량(kg)	1997~2020년	24.4	4.61

자료: 농업관측통계시스템(OASIS), 농산물유통정보(KAMIS), 농업경영종합정보시스템, 통계청, 농업전망 자료집

4.2. 비정형데이터

재배면적 함수에서 변수로 활용된 양파 관련 감성지수는 0에서 1 사이의 값을 가지며 소비자의 경기 전망 판단 정도를 분석한 경기심리지수와 유사한 역할을 한다(송민채, 2017). 언론

사는 국내 점유율 1위 포털 사이트인 네이버 뉴스에서 제공하는 기사로 설정하였으며, 기간은 1992년 1월부터 2020년 12월까지로 수집했으나, 1992년부터 1995년까지 기사는 데이터의 수가 많지 않고 직접적인 영향을 끼친 기사가 존재하지 않아 모델이 학습할 때 제외하였다. 기사 수집은 일차적으로 네이버 포털 사이트에서 양파와 관련된 모든 기사를 수집하였고, 2차적으로는 본 연구와 상관관계가 낮은 기사는 삭제하였다.

표 4. 감성지수 도출에 활용된 기사의 수

데이터	6월 기사의 수	7월 기사의 수	8월 기사의 수
1997 ~ 1999년도 기사	32개	25개	32개
2000 ~ 2004년도 기사	56개	41개	39개
2005 ~ 2009년도 기사	82개	76개	78개
2010 ~ 2014년도 기사	89개	95개	57개
2015 ~ 2020년도 기사	162개	117개	121개
합계	421개	354개	317개
총합계	1,092개		

자료: 1997~2020년 네이버 포털 양파 관련 기사

5. 분석 결과

5.1. 감성지수

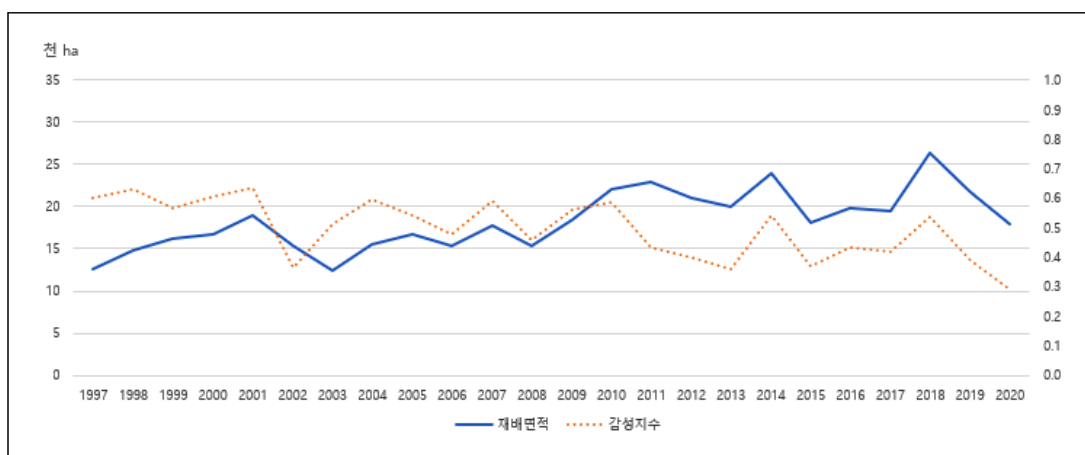
감성지수는 비정형 데이터로 일정한 규칙이 없고, 특정 사건에 단기적으로 영향을 받기 쉽다는 특성이 있으므로 양파의 파종 시기(8월 말)에 근접한 6, 7, 8월 감성지수를 채택하였다. 도출한 감성지수에 대한 기초통계량은 아래의 <표 5>와 같다.

표 5. 감성지수 기초통계량

변수(단위)	기간	평균	표준편차
감성지수(6월)	1997~2020년	0.47	0.11
감성지수(7월)	1997~2020년	0.50	0.13
감성지수(8월)	1997~2020년	0.50	0.12

아래 <그림 6>은 6월 감성지수 데이터에 0.1, 7월 감성지수 데이터에 0.3, 8월 감성지수 데이터에 0.6의 가중치를 준 자료와 재배면적의 추이를 비교한 그래프이다. 감성지수가 상승할 때 재배면적에 긍정적 영향을 미치는 것으로, 하락할 때 재배면적에 부정적 영향을 미치는 것으로 해석할 수 있으며, 감성지수와 재배면적의 추세는 선행 또는 후행하지 않고 동행하는 행태를 보인다. 감성지수가 가장 하락한 값을 보인 연도는 2002년과 2020년으로 2002년의 감성지수는 한-칠레 FTA 관련 기사 자료가 많아 자료가 많아, 양파 공급 과잉을 예상해 낮게 나타났음을 알 수 있다. 2020년의 경우 재배면적에 직접적인 영향을 끼치는 기사의 수는 적고 COVID-19의 영향으로 인해 유통 과정의 축소 및 소비자 소비 위축 심리가 반영된 기사가 많아 당해 감성지수가 하락하였고, 이에 따라 재배면적 또한 2019년도에 비해 감소한 것으로 나타났다.

그림 6. 감성지수와 재배면적 변동 추이



자료: 저자 작성(감성지수), 통계청(재배면적)

5.2. 양파 재배면적 반응함수 추정

<표 6>과 <표 7>은 감성지수를 도입한 log-log 형태의 재배면적 반응함수의 OLS(Ordinary Least Square) 추정결과이다. <표 6>의 모형 1, 모형 2, 모형 3은 각각 6월(SE_6), 7월(SE_7),

8월(SE_8)의 감성지수를 도입한 모형을 의미한다. 표에서 확인할 수 있는 바와 같이, 각 설명 변수 계수 값의 부호와 크기, 통계적 유의성은 다소 차이가 있다. 세 모형 모두 전년도 재배면적은 1보다 작은 계수 값을 가지며, 1% 유의수준에서 통계적으로 유의한 것으로 나타났다. 감성지수가 1% 증가했을 때, 모형 1, 2, 3은 각각 재배면적이 8%, 15%, 51% 증가하는 것으로 나타나 파종기에 근접할수록 감성지수가 재배면적 의사결정에 미치는 영향이 큰 것으로 보인다. 단, 감성지수는 모형 3에서는 1% 유의수준에서 통계적으로 유의한 것으로 나타났으나, 모형 1, 2의 감성지수 변수 유의성이 떨어지는 한계가 있다.

표 6. 재배면적 반응함수 추정 결과(모형 1, 2, 3)

변수	모형 1	모형 2	모형 3
intercept	1.636 (2.91)	0.45 (2.851)	-0.716 (2.162)
전년도 재배면적	0.822 (0.253)***	0.92 (0.247)***	0.766 (0.183)***
trend	0.358 (0.279)	0.276 (0.264)	0.054 (0.208)
경영비	-0.259 (0.143)*	-0.23 (0.134)	0.162 (0.113)
전기 양파 가격/농가물가지수	0.486 (0.137)***	0.481 (0.126)***	0.118 (0.12)
전기 마늘 가격/농가물가지수	-0.081 (0.092)	-0.066 (0.085)	0.025 (0.074)
감성지수 (모형 1: SI_6 , 모형 2: SI_7 , 모형 3: SI_8)	0.083 (0.097)	0.149 (0.086)	0.51 (0.126)***
R-squared	0.7827	0.8073	0.872

주: () 표준오차를 뜻함. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.1$

〈표 7〉의 모형 4, 모형 5, 모형 6은 각각 6월 데이터에 0.4, 7월 데이터에 0.6의 가중치를 둔 감성지수(SE_{67}), 7월 데이터에 0.4, 8월 데이터에 0.6의 가중치를 둔 감성지수(SE_{78}), 6월 데이터에 0.1, 7월 데이터에 0.3, 8월 데이터에 0.6의 가중치를 둔 감성지수(SE_{678})를 포함한 모형이다. 표에서 확인할 수 있는 바와 같이 감성지수는 모형 4에서 10% 유의수준, 모형 5, 6에서는 1% 유의수준에서 통계적으로 유의한 것으로 나타났으며 감성지수가 1% 증가하면 농가들은 당해 양파재배면적을 각각 21%, 50%, 55% 증가시키고 있는 것으로 나타난다. 따라

서 파종기에 근접할수록 농가들은 양파와 관련된 정보에 민감하게 반응하는 것으로 분석된다.

표 7. 재배면적 반응함수 추정 결과(모형 4, 5, 6)

변수	모형 4	모형 5	모형 6
intercept	0.2 (2.9)	-1.442 (2.336)	-1.513 (2.313)
전년도 재배면적	0.954 (0.251)***	0.954 (0.19)***	0.956 (0.188)***
trend	0.29 (0.261)	0.115 (0.216)	0.117 (0.214)
생산비	-0.238 (0.133)*	-0.002 (0.124)	0.007 (0.123)
전기 양파 가격/농가물가지수	0.507 (0.127)***	0.295 (0.108)**	0.29 (0.107)**
전기 마늘 가격/농가물가지수	-0.08 (0.085)	-0.005 (0.07)	-0.009 (0.069)
감성지수 (모형 3: SI_{67} , 모형 4: SI_{78} , 모형 5: SI_{678})	0.21 (0.114)*	0.503 (0.131)***	0.545 (0.139)***
R-squared	0.8101	0.8782	0.8809

주: () 표준오차를 뜻함. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.1$

아래의 <표 8>은 감성지수가 포함된 모형과 감성지수를 모형에 포함하지 않았을 때의 재배면적 추정 정확도를 비교하기 위해 추가한 log-log 형태로 추정된 다중회귀분석식이다.

표 8. 감성지수를 포함하지 않은 재배면적 함수

모형	다중회귀분석식
모형 7	(9) $A_t = \beta_0 + \beta_1 P_{A-1} + \beta_2 P_{A-1} + \beta_3 A_{t-1} + \beta_4 C_t + \beta_6 trend + e_t$

추정값과 실제값을 비교하는 척도로서 평균 제곱근 오차(RMSE)를 이용했으며, 모형 1~7의 RMSE는 다음 <표 9>와 같다. 감성지수를 포함하지 않은 모형 7의 RMSE가 가장 높게 나타났고, 파종기인 8월 말에서 상대적으로 먼 시기인 6월을 포함한 모형 1, 모형 4 순서대로 RMSE가 크게 나타났다. 하지만, 6, 7, 8월 감성지수를 포함한 모형 6의 RMSE가 7, 8월 감성지수만을 포함한 모형 5의 RMSE보다 더 낮은 값을 갖는 것으로 나타났다. 이는 파종기와 시

차가 거의 없는 기사만이 아닌 최대 3개월 전의 기사도 또한 낮은 가중치(0.1)를 부여해 함께 고려했을 때, 추정된 모형의 정확도가 올라감을 알 수 있다.

표 9. 모형별 RMSE 값 비교

	모형 1	모형 2	모형 3	모형 4	모형 5	모형 6	모형 7
RMSE	1,661	1,553	1,192	1,519	1,133	1,107	1,724

5.3. 수급 균형 모형 구축

양파 시장에 대한 수급 균형 모형 구성을 위해 식 (1.6)과 식 (2)~식 (7)의 방정식을 이용했다. 본 연구의 균형 모형에서 마늘 도매가격, 이전재고량, 수입량, 기말재고량, 수출량, 1인당 GDP, 인구수는 모형 내에서 결정되지 않는 외생변수로 가정한다. 재배면적 함수는 RMSE가 가장 낮았던 <표 7>의 모형 6 추정결과를 사용했으며, 단수함수의 추정 결과는 <표 10>, 소비함수의 추정 결과는 <표 11>에 정리되어 있다.

단수함수 추정결과, 기술진보에 따라 단수가 양(+)의 방향으로 증가하는 것으로 나타났다. 소비함수 추정결과, 소득의 대리변수인 1인당 GDP의 계수가 양(+)의 방향으로, 양파 가격이 음(-)의 방향으로 경제 이론과 부합하는 것으로 추정되었으며, 유의성도 각각 1%와 10%에서 유의한 것으로 높게 나타났다. 대체가격인 마늘 가격은 통계적으로 유의하게 추정되지 않았으며, 이는 소비자에게는 같은 양념채소더라도 마늘이 양파의 대체 품목이 되기 어렵기 때문으로 해석된다.

표 10. 단수함수 추정결과

변수	Coefficient
intercept	5121 (412.03)***
trend	43.23 (15.04)***

주: () 표준오차를 뜻함. *** p<0.01 ** p<0.05 * p<0.1

표 11. 소비함수 추정결과

변수	Coefficient
intercept	12.5368 (2.34)***
1인당 GDP	0.0063 (0.0008)***
양파 가격/소비자물가지수	-0.0079 (0.0021)***
마늘 가격/소비자물가지수	0.0006 (0.0004)

주: () 표준오차를 뜻함. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.1$

소비에 대한 자체가격 탄력성, 소득 탄력성은 <표 12>와 같다. 자체가격 탄력성은 -0.29로 나타났으며, 이는 양파 가격이 1% 상승할 때 1인당 소비량은 0.29% 감소한다는 의미로 가격 변화에 대한 소비량 변화가 비탄력적임을 알 수 있다.

표 12. 수요의 자체가격 탄력성, 소득 탄력성

구분	탄성치 계산값
자체가격 탄력성	-0.28581
소득 탄력성	0.82409

5.4. 베이스라인

추정된 모형 6의 재배면적 함수, 단수함수, 소비함수를 <표 5>의 계산식을 사용해 총공급 함수, 총수요함수로 계산하고 식 (8)과 같은 방식으로 초과공급($S - D$) 혹은 초과수요($D - S$)를 줄이며 반복계산을 통해 시장 균형에 이를 수 있다. 모형에서 결정된 2021~2025년 전망치는 <표 13>에 정리되어 있다. 2021년 베이스라인은 한국농촌경제연구원의 농업전망 2021에서 제시하는 중장기 수급 전망 자료를 참고하여 결정되었다. 감성지수는 직전 3년 감성지수의 평균값이 2021~2025년에도 같다는 가정 하에 6월의 감성지수가 0.35, 7월이 0.48, 8월이 0.35로 계산되었다.

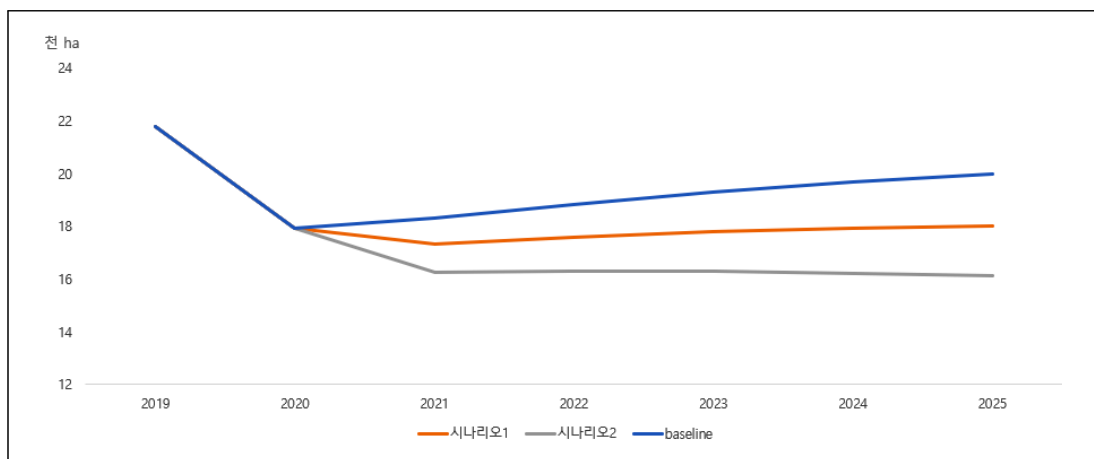
표 13. 양파 베이스라인(2021~2025년)

변수	2021	2022	2023	2024	2025
재배면적(ha)	18,336	18,843	19,292	19,673	19,986
생산량(천 톤)	1,247	1,282	1,314	1,340	1,362
소비량(천 톤)	1,318	1,359	1,314	1,340	1,394
도매가격(원/kg)	1,361	1,282	1,257	1,185	1,117
6월 감성지수	0.35	0.35	0.35	0.35	0.35
7월 감성지수	0.48	0.48	0.48	0.48	0.48
8월 감성지수	0.35	0.35	0.35	0.35	0.35

5.5. 시나리오 분석

시나리오 1은 2021년 7월의 감성지수가 0.1점 하락했을 때, 시나리오 2는 2021년 8월의 감성지수가 0.1점 하락했을 때의 상황을 가정한다. 이때 <그림 7>은 2021~2025년의 재배면적 전망을 나타내는 그래프이다. 재배면적은 시간이 지남에 따라 증가하는 양상을 보이며, 감성지수가 상승하면 재배면적도 증가하고 감성지수가 하락하면 재배면적도 감소한다.

그림 7. 양파 재배면적 시나리오 분석



2021년의 결과를 비교하면, 시나리오 1에서는 베이스라인에 비해 재배면적이 5.51% 하락했으며, 생산량 역시 5.51% 하락, 1인당 소비량은 5.21% 하락하고, 도매가격은 12.74% 증가하였다. 시나리오 2에서는 베이스라인에 비해 재배면적과 생산량이 11.3% 감소하였으며, 1인당 소비량은 10.70% 감소하고, 도매가격은 26.12% 증가하였다. 시나리오 분석 결과는 감성지수 가중치에 따라 달라질 수 있다(부록 참고).

표 14. 시나리오 분석 결과(2021~2025년)

	2021	2022	2023	2024	2025
재배면적(ha)					
베이스라인	18,336	18,843	19,292	19,673	19,986
시나리오 1	17,326	17,588	17,793	17,932	18,007
% 변화	-5.51%	-6.66%	-7.77%	-8.85%	-9.90%
시나리오 2	16,264	16,310	16,304	16,243	16,129
% 변화	-11.30%	-13.44%	-15.49%	-17.44%	-19.30%
생산량(천 톤)					
베이스라인	1,247	1,282	1,314	1,340	1,362
시나리오 1	1,179	1,197	1,212	1,222	1,227
% 변화	-5.51%	-6.66%	-7.77%	-8.85%	-9.90%
시나리오 2	1,106	1,110	1,110	1,107	1,099
% 변화	-11.30%	-13.44%	-15.49%	-17.44%	-19.30%
소비량(천 톤)					
베이스라인	1,318	1,359	1,314	1,340	1,394
시나리오 1	1,249	1,273	1,269	1,275	1,290
% 변화	-5.21%	-6.28%	-7.64%	-8.51%	-9.47%
시나리오 2	1,177	1,186	1,166	1,160	1,162
% 변화	-10.70%	-12.69%	-14.86%	-16.76%	-18.45%
도매가격(원/kg)					
베이스라인	1,361	1,282	1,257	1,185	1,117
시나리오 1	1,534	1,498	1,515	1,485	1,458
% 변화	12.74%	16.84%	20.55%	25.31%	30.52%
시나리오 2	1,716	1,718	1,772	1,776	1,781
% 변화	26.12%	34.00%	40.95%	49.86%	59.49%

6. 결론 및 시사점

본 연구는 감성분석을 통해 얻어지는 비정형 데이터인 양파의 관련 감성지수를 재배면적 반응함수의 변수로 활용함으로써 양파 가격 전망의 정확도를 높이고, 시나리오 분석을 통해 감성지수가 재배면적에 미치는 영향을 제시해 양파 재배 농민 및 정책 수립자들에게 보다 정확한 데이터를 제공하고자 수행되었다. 이와 같은 연구목적을 달성하기 위하여 본 연구에서는 국가통계포털과 KAMIS, 농축산물 소득자료집을 통해 양파 재배면적, 양파 단수, 양파 도매가격, 마늘 도매가격, 양파농가 경영비, 1인당 GDP, 1인당 소비량 자료를 수집하였다. 이후 웹 크롤링 방식으로 기사를 수집하고, LSTM-CNN 모델을 이용해 재배면적 함수에 활용할 감성지수를 도출하였다. 감성지수를 월별(6, 7, 8월), 가중치를 둔 기간별(6·7월, 7·8월, 6·7·8월)로 분류, OLS(Ordinary Least Square)로 추정한 양파 재배면적 분석모형에 도입하였다. 각 모형별 평균 제곱근 오차(RMSE)를 이용해 가장 적합한 모형을 선정한 후, 추정된 재배면적 함수를 활용하여 양파 수급예측 모형을 설계하였으며 시나리오 분석을 통해 양파 감성지수가 재배면적에 미치는 영향을 분석하였다.

본 연구에 대한 분석 결과를 요약하면 다음과 같다. 첫째, 양파 재배면적 함수에 감성지수를 도입하고, 파종기에 가까울수록 감성지수의 추정된 계수가 통계적으로 유의해진다는 점을 확인했다. RMSE를 통해 비교한 결과 6월, 7월, 8월에 각각 0.1, 0.3, 0.6의 가중치를 둔 모형이 가장 설명력이 높게 나타났으며, 이를 감성지수가 포함되지 않은 모형과 비교했을 때 감성지수를 변수로 포함한 경우에 실측값과의 괴리가 감소하는 것으로 밝혀졌다. 둘째, 양파 단수 함수와 소비함수를 추정하여 총공급함수와 총수요함수를 구축하고, 감성지수의 움직임에 따른 시장균형가격의 움직임을 살핀 뒤 감성지수의 변화에 따른 2021~2025년 재배면적 전망치를 살펴보았다. 그 결과, 7월의 감성지수가 0.1점 감소(시나리오 1)했을 때 2021년에 베이스라인에 비해 재배면적이 5.51% 감소하였으며, 8월의 감성지수가 0.1점 감소(시나리오 2)했을 때는 베이스라인에 비해 11.30% 감소하는 모습을 보였다.

이러한 분석 결과를 토대로 한 시사점은 다음과 같다. 본 연구에서는 부동산, 금융과 같은 경제학 분야에서 활용되고 있으나 농업 분야에서는 활용이 미진했던 뉴스의 감성분석을 통한

감성지수를 재배면적 함수에 설명변수로 추가하고, 비정형 데이터를 활용하여 재배면적 추정 및 가격 전망의 정교화가 가능함을 보였다. 이는 농업 부문 전망모형에서 기존에 활용되던 여러 경제적 변수 외에도 시장 동향 및 전망 예측에 감성분석과 같은 빅데이터 분석이 활용될 수 있다는 점을 환기하는 결과이다.

본 연구는 다음과 같은 몇 가지 한계를 지닌다. 첫째로, 가격자료 확보의 어려움으로 모든 데이터를 1996년부터 2020년까지 수집하여, 상대적으로 적은 기간의 자료가 활용되었다. 또, 품종이 조생종과 중만생종으로 나뉘는 양파와 생산지 기후에 따라 난지형과 한지형으로 나뉘는 마늘의 특성을 고려하지 못하고 양파는 두 품종을 한꺼번에, 마늘은 난지형 자료만 사용했다. 보다 정확한 추정을 위해서는 본 연구에서 사용된 자료보다 더 이전의 과거 자료를 확보해 데이터의 길이를 늘려야 하며, 확보하고 활용할 수 있어야 하며, 양파와 마늘의 품종이 다를 수 고려하여 두 품종에 따라 다른 변수로 포함하는 과정이 필요하다고 판단된다.

둘째, 수급 균형 모형의 단순화를 위하여 재배면적 함수 외에 다른 함수에서 설명변수가 적게 활용되었고, 수입수요 함수, 재고량 함수 등의 함수가 포함되지 않았다. 특히 양파 단수 함수에는 기술 발전의 대리변수인 추세변수(*trend*)만이 사용되어 모형 적합도가 낮은 문제가 있다. 선행연구에서 연구된 양파 단수 결정함수와 같이 기상변수, 생육 정보 등을 함수에 포함시켜 균형 모형 분석에 반영한다면 균형 모형의 체계화가 가능할 것으로 보인다.

그럼에도 불구하고 본 연구는 양파 관련 뉴스 감성지수를 산출하여 재배면적에 영향을 미치는지 살펴보고, 이를 통해 수급전망을 실시했다는 점에서 연구적 의의가 있다. 감성지수 산출을 통해 시간과 비용이 많이 드는 재배의향 설문조사를 대체할 수 있을 뿐만 아니라 최근 뉴스 감성지수 산출이 가능하여 즉각적인 수급대응에 용이할 것이다. 마지막으로 빅데이터를 활용한 분석기법을 농업 수급모형에 활용할 수 있는 기초연구로 활용될 것으로 기대한다.

참고문헌

- 강산하, 오승원, 강혜정. 2017. “농업분야 빅데이터의 활용 사례와 소셜 네트워크 데이터 분석을 통한 농산물 가격 분석.” 「한국농식품정책학회 2017년 하계학술대회 논문집」 한국농식품정책학회.
- 나형철, 오은화, 유도일, 조완섭, 이지스 나스리디노프, 박성호, 조용빈, 류관희. 2018. “정형 및 비정형 빅데이터를 이용한 양파 소비 예측.” 「한국콘텐츠학회논문지」 제18권 제11호. 한국콘텐츠학회.
- 류재현, 이진아, 안병일. 2020. “농업관측 정보가 농간의 재배면적 조절에 미치는 영향: 양파와 마늘을 중심으로.” 제61권 제1호. 한국농촌계획학회 「농업경제연구」.
- 박재수, 이재수. 2019. “아파트 매매가격과 부동산 온라인 뉴스의 교차상관관계와 인과관계 분석: 온라인 뉴스 기사의 비정형 빅데이터를 활용한 감성분석 기법의 적용.” 「國土計劃」 제54권 제1호. 대한 국토·도시계획학회.
- 서홍석, 순병민, 김충현, 김문희. 2019. 「농업부문 전망모형 KREI-KASMO 2019 운용·개발 연구」. 한국 농촌경제연구원.
- 송민채, 신경식. 2017. “뉴스기사를 이용한 소비자의 경기심리지수 생성.” 「지능정보연구」 제23권 3호. 한국지능정보시스템학회.
- 이승인, 손찬수, 이해림. 2020. “기상 및 영농투입재 비용이 양파 단수에 미치는 영향: 패널회귀분석과 베이지안네트워크모형을 중심으로.” 「농촌경제」 제43권 제2호. p. 3, 한국농촌경제연구원 보고서.
- 이중용. 1995. “마늘, 양파 주산지 육성을 위한 기본방향.” 「농촌경제」 제18권 제4호, 한국농촌경제연구원 보고서.
- 차은정, 홍태호. 2016. “뉴스 감성분석과 SVM을 이용한 S&P 500 주가지수 예측.” 「한국경영정보학회 학술대회논문집」 제6호. 한국경영정보학회.
- 하지희, 서상택, 김선웅. 2019. “양파와 마늘가격 예측모형의 예측력 고도화 방안.” 「농촌계획」 제25권 제4호. 한국농촌계획학회 「농업경제연구」.
- 한국농촌경제연구원. 2021. 『농업전망 2021: 코로나19 이후 농업·농촌의 변화와 미래』. pp. 442-455. 한국농촌경제연구원 보고서.
- Jin, Z, Y Yang, Y Liu - Neural Computing and Applications, 2019 - Springer Stock closing price prediction based on sentiment analysis and LSTM.
- Saura, JR , P Palos-Sanchez, A Grilo - Sustainability, 2019 Detecting indicators for startup business success: Sentiment analysis using text data mining.
- Tang, D, F Wei, N Yang, M Zhou, T Liu... - Proceedings of the 52nd ..., 2014 Learning sentiment-specific word embedding for twitter sentiment classification.
- Yenter, A, A Verma - 2017 IEEE 8th Annual Ubiquitous ..., 2017 Deep CNN-LSTM with combined kernels from multiple branches for IMDB review sentiment analysis.
- Ynkun Ma. and Haiyun Peng and Erik Cambria. 2018. “Targeted Aspect-Based Sentiment Analysis via Embedding Commonsense Knowledge into an Attentive LSTM.”

부록

부록 1. 양파 재배면적, 생산량, 1인당 소비량, 도매가격 전망치 표(시나리오 2 가중치 변경)

	2021	2022	2023	2024	2025
재배면적(ha)					
베이스라인	18,336	18,843	19,292	19,673	19,986
$SI_6 \times 0.3 + SI_7 \times 0.3 + SI_8 \times 0.4$	16,981	17,118	17,201	17,224	17,190
% 변화	-7.39%	-9.15%	-10.84%	-12.45%	-13.99%
$SI_6 \times 0.6 + SI_7 \times 0.3 + SI_8 \times 0.1$	17,891	18,173	18,399	18,562	18,664
% 변화	-2.43%	-3.56%	-4.63%	-5.65%	-6.62%
생산량(천 톤)					
베이스라인	1,247	1,282	1,314	1,340	1,362
$SI_6 \times 0.3 + SI_7 \times 0.3 + SI_8 \times 0.4$	1,155	1,165	1,171	1,173	1,172
% 변화	-7.39%	-9.15%	-10.84%	-12.45%	-13.99%
$SI_6 \times 0.6 + SI_7 \times 0.3 + SI_8 \times 0.1$	1,217	1,237	1,253	1,265	1,272
% 변화	-2.43%	-3.56%	-4.63%	-5.65%	-6.62%
소비량(천 톤)					
베이스라인	1,318	1,359	1,314	1,340	1,394
$SI_6 \times 0.3 + SI_7 \times 0.3 + SI_8 \times 0.4$	1,226	1,241	1,227	1,227	1,234
% 변화	-6.99%	-8.64%	-10.40%	-11.97%	-13.38%
$SI_6 \times 0.6 + SI_7 \times 0.3 + SI_8 \times 0.1$	1,288	1,313	1,308	1,318	1,335
% 변화	-2.30%	-3.36%	-4.44%	-5.43%	-6.33%
도매가격(원/kg)					
베이스라인	1,361	1,282	1,257	1,185	1,117
$SI_6 \times 0.3 + SI_7 \times 0.3 + SI_8 \times 0.4$	1,593	1,579	1,617	1,607	1,599
% 변화	17.08%	23.15%	28.66%	35.60%	43.13%
$SI_6 \times 0.6 + SI_7 \times 0.3 + SI_8 \times 0.1$	1,437	1,398	1,410	1,376	1,344
% 변화	5.61%	8.99%	12.24%	16.15%	20.40%