

지역 산업 위기 대응을 위한 뉴스 데이터 감성 분석: 광주 자동차 산업 사례를 중심으로

김성진*, 홍종의**, 최성철***

*한국과학기술정보연구원 기술사업화센터

**안동대학교 경영회계학부

***가천대학교 산업경영공학과

e-mail:sungjin.kim@kisti.re.kr

A case study for emotional analysis of news data to deal with regional industrial crisis

Sungjin Kim*, Jong-yi Hong**, Sungchul Choi***

*Technology Commercialization Center, Korea Institute of Science and Technology
Information

**Division of Management and Accounting, Andong National University

***Dept. of Industrial Engineering, Gachon University

요약

지역의 산업생태계는 외부 환경변화에 따른 위기에 직면하는 경우가 발생하는데 이러한 위기에 어떻게 대응하느냐에 따라 다양한 회복 경로가 있을 수 있다. 지역 산업이 강한 회복력을 가지기 위해서는 위기를 조기에 예측하여 그 위기에 대한 대비를 미리 마련하는 것이 매우 중요하다. 따라서 본 연구에서는 지역 산업에 닥치는 위기를 파악하는 수단으로써 뉴스 데이터를 활용하고 감성분석을 통해 지역 산업의 위기를 조기에 예측하고자 하였다. 뉴스 데이터를 기반으로 한 감성분석 결과와 실제 산업단지의 통계지표와의 상관관계 분석 결과 감성분석 결과는 시차를 두고 경제통계지표에 영향을 주는 것으로 확인하였으며 특히 5, 6개월 이후의 경제지표를 선행해서 설명하고 있는 점에서 많은 시사점을 얻을 수 있다.

1. 서론

지역 산업 생태계의 위기는 크게 사회 패러다임 변화 등에 따른 만성적인 스트레스와 단기적인 급성충격으로 구분할 수 있다. 전자의 경우 장기간에 걸쳐 지속되는 쇠퇴로 볼 수 있고 후자의 경우 외부 환경변화에 따른 위기로 파악할 수 있는데 이러한 위기에 어떻게 대응하느냐에 따라 다양한 회복 경로가 있을 수 있다[1]. 지역의 산업생태계가 보유하고 있는 회복력(Resilience)에 따라 얼마나 빠르게 위기 이전으로 산업생태계가 회복되는지가 결정된다고 볼 수 있으며 강한 회복력을 가지기 위해서는 가장 먼저 위기를 조기에 예측하여 그 위기가 단기적인 충격인지 만성적이고 지속적인 산업의 쇠퇴인지를 판단하고 이에 대한 대비를 마련하는 것이 중요하다.

따라서 본 논문에서는 비정형 데이터인 뉴스데이터를 기반으로 지역 산업 관련 뉴스에 대한 감성분석 결과를 통해 지역 산업의 위기를 조기에 예측할 수 있는 보완적 수단을 개발하고자 한다.

2. 관련 문헌 연구

2.1 산업 위기 관련 연구

산업입지 및 개발에 관한 법률로 규정하고 있는 산업단지는 1960년대부터 지정되기 시작하여 현재는 전국적으로 1,196개의 산업단지에 총 99,075개 사가 입주하고 있는 국내의 산업을 견인하는 중추적인 역할을 수행하고 있다[2]. 국가의 경제성장뿐만 아니라 지역 균형발전의 측면에서도 중요한 역할을 담당하고 있는 산업단지는 그 역사가 오래된 만큼 노후화가 많이 진행되었으며, 청년층의 인력 유입 감소, 생산인력의 고령화, 입주업체의 영세화 등 다양한 문제가 발생하고 있는 것 또한 현실이다[3].

산업단지의 쇠퇴를 분석하고 쇠퇴의 정도를 측정하기 위한 연구와 쇠퇴를 극복하기 위한 활성화 정책을 제안한 연구들이 수행된 바 있으며 손영우(2012)는 산업단지의 쇠퇴 형태를 유형별로 정의하고 각 유형별 특성을 제시하였으며[4] 박병호 외(2009)는 고용인구, 가동업체 수, 생산 및 수출액 등을 산업단지의 쇠퇴를 측정하기 위한 지표로 활용하였다[5]. 김성진 외(2019)는 산업단지 위기를 파악하기 위해 여러 경제지표들을 바탕으로 AHP를 활용하여 측정지표를 도출한 후 산업단지 위기지수를 개발하고 실제 국가산업단지에 적용을 하여 그 실효성을 검증하였다. 하지만 이들 연구에서 활용한 쇠퇴

퇴 혹은 위기의 측정은 대부분 정기적으로 발표되는 경제지표를 바탕으로 수행되었기 때문에 갑작스럽게 닥치는 위기에 대한 적시성이 떨어지는 한계가 있으며 이로 인해 위기에 대한 선제적인 대응이 쉽지 않다는 한계가 있다.

2.2 뉴스 데이터를 활용한 예측 모형 연구

텍스트 마이닝 관련 기술들이 발전하면서 언론기사를 분석하여 다양한 분야에 활용하는 연구들이 활발히 수행중에 있다. 송민채와 신경식(2017)은 뉴스 기사를 활용하여 경기심리지수를 개발하였으며 이를 소매판매업지수, 고용률 등의 실물지표와의 비교를 통해 활용성을 검증하였다[7]. 김유신 외(2012)은 뉴스와 주가간에 밀접한 관계가 있다고 가정하고 뉴스 데이터의 감성분석을 통해 주가지수를 예측하는 모형을 제시하였다[8]. 마찬가지로 정지선 외(2015)는 개별 기업에 대한 뉴스 정보의 감성분석을 바탕으로 각 기업에 대한 주가의 변화를 예측하고자 하였다[9]. 상기 연구를 통해 알 수 있듯이 실시간으로 생산되는 뉴스의 빅데이터 분석을 통한 미래에 대한 예측은 충분한 가능성을 가지고 있고 나아가 지역의 산업 생태계 위기 역시 조기에 포착할 수 있는 기회를 가질 수 있다.

3. 산업 관련 뉴스 감성분석

본 연구에서는 대응량 뉴스 데이터를 활용하여 뉴스의 감성 점수에 따른 지역 산업의 위기 예측의 가능성을 점검하고자 한다. 이를 위해 Google 감성분석 API를 사용하여 뉴스 감성분석을 실시하고 월별로 구분하여 점수화한 후 감성 분석 결과와 실제 경제 지표 사이의 연관성을 살펴보았다.

본격적인 감성분석을 실시하기 전 특정 산업 관련 뉴스 데이터 수집이 제대로 되었는지를 확인하기 위해 자동차 산업을 키워드로 한 2018년부터 2019년 8월까지의 일별 뉴스데이터 총 31,479건을 수집한 결과, 자동차 산업 관련 뉴스는 한국 GM의 회생방안 결정(2018.05.10.)이나 자동차부품산업 활력제고 방안(2018.12.18.) 등 큰 이슈가 있을 때 마다 많은 뉴스가 생성되었음을 확인할 수 있다.

1차적인 뉴스 데이터 수집 이후 본격적인 사례분석을 위해 광주 자동차 산업으로 범위를 한정하고 언론사 또한 5대 일간지, 3대 경제지, 광주지방 지역언론으로 제한하였으며, 인사, 동정, 정부정책 등의 뉴스를 제거하는 등 전처리를 시행하여 총 5,803건의 뉴스를 바탕으로 2차 분석을 실시하였다. 구글 NLP API를 활용하여 광주지역 자동차 산업 관련 뉴스에 대한 감성분석을 실시하고 감성 분석 점수를 도출하였다. 감성 분석 결과는 -1과 1 사이의 값을 가지며 -1로 갈수록 부정적인, 1로 갈수록 긍정적인 감성 정보로 판단할 수 있다. 분

석 결과 광주형 일자리 사업과 관련된 긍정적인 뉴스가 많이 보도되었을 때 일별 뉴스의 감성분석 결과가 높은 점수를 획득하였다

도출된 감성분석 결과의 실효성을 검증하기 위해 감성분석 결과와 광주지역의 산업단지인 광주첨단과학산업단지의 경제관련 통계데이터와 이를 기반으로 김성진 외(2019)에서 제안한 산업단지 위기지수 결과와의 비교분석을 수행하였다. 광주첨단과학산업단지의 통계지표와 위기지수의 데이터는 월별로 수집되기 때문에 일별 감성분석 지수를 월별로 산술 평균하여 도출된 점수를 활용하였으며 이들 점수와 월별 지표들의 변화를 비교분석하였다.

추가로 감성분석 결과가 광주첨단과학산업단지의 경제지표와 얼마나 상관관계가 있는지 파악하기 위해 감성분석 데이터와 경제지표의 상관관계를 분석하였다. 추가로 감성분석 결과가 경제지표와의 선행성을 확인하기 위해 이동평균 및 1개월부터 6개월 전까지의 감성분석 결과와의 상관관계 분석을 추가로 수행한 결과는 아래 [표 1]과 같다.

[표 1] 감성분석결과 및 산업단지 경제지표 상관관계

	입주 기업 수	생산 실적	가동 률	수출 실적	고용 현황	위기 지수
senti	-0.31	-0.23	-0.08	-0.10	-0.14	-0.21
senti_rolling_2	-0.01	0.12	0.33	0.23	0.11	0.20
senti_rolling_3	0.23	0.36	0.58	0.36	0.24	0.44
senti_rolling_4	0.38	0.36	0.56	0.39	0.35	0.50
senti_rolling_5	0.49	0.43	0.64	0.41	0.42	0.59
senti_rolling_6	0.58	0.34	0.55	0.41	0.48	0.58
senti_shift_1	0.04	0.05	0.24	0.15	0.11	0.15
senti_shift_2	0.31	0.31	0.50	0.21	0.19	0.36
senti_shift_3	0.39	0.31	0.34	0.31	0.34	0.42
senti_shift_4	0.43	0.40	0.48	0.43	0.40	0.54
senti_shift_5	0.47	0.31	0.46	0.34	0.41	0.49
senti_shift_6	0.59	0.42	0.46	0.42	0.44	0.57

상관관계 결과를 보면 전반적으로 5, 6개월 이전의 감성분

석 결과와 경제지표와 상관관계가 높음을 알 수 있다. 다시 말하면 5개월 이후에 본격적인 감성분석 결과의 영향이 경제지표로 나타난다고 할 수 있고 각 세부지표별로 보면 가장 먼저 입주기업수, 생산실적, 가동률 등이 영향을 받고 수출실적 및 고용현황이 순차적으로 영향을 받는 것으로 보인다.

4. 결론

본 연구는 지역산업의 위기를 조기에 파악하기 위하여 뉴스 데이터를 대상으로 한 감성분석 결과를 활용할 수 있는 가능성을 타진해 보려 하였으며 실제 산업단지의 경제지표와 감성분석 결과와의 상관관계분석을 통해 그 실효성을 입증하였다. 연구 결과 감성분석 결과는 시차를 두고 경제지표에 영향을 주는 것으로 확인하였으며 특히 감성분석 결과는 5, 6개월 이후의 경제지표를 선행해서 설명하고 있는 점에서 많은 시사점이 있을 수 있다. 다만 해당 연구는 가능성을 확인하는 관점에서 수행되었으므로 향후 연구에서는 뉴스데이터의 정교한 노이즈 제거와 특정 산업 관련 뉴스 분류기 구축 등 데이터 정합성에 대한 추가 연구가 필요하며 이를 통하여 감성분석 결과의 예측력과 설명력을 보다 높일 필요가 있다.

참고문헌

- [1] 하수정 외, “지속가능한 발전을 위한 지역 회복력 진단과 활용방안 연구”, 국토연구원, 2014년
- [2] 산업입지연구소, 2018 전국산업단지 현황통계(3/4분기)
- [3] 산업단지 구조고도화사업, 한국산업단지공단 홈페이지, <http://www.kicox.or.kr/home/mwrc/advanc/strctAdvanc.jsp>
- [4] 손영우, 산업단지 쇠퇴 유형화 및 유형별 쇠퇴특성 분석, 충북대학교 석사학위논문, 2012년
- [5] 박병호, 인병철, 김태영, “전국 산업단지 쇠퇴 분석”, 지역연구, 제 25권, 3호, pp.61-73, 2009년
- [6] 김성진, 홍중의, 김한국, “산업단지 위기지수 도출을 위한 방법론 개발”, 한국산학기술학회학술지, 제20권, 10호, pp. 250-258, 2019년
- [7] 송민채, 신경식, “뉴스기사를 이용한 소비자의 경기심리지수 생성”, 지능정보연구, 제23권, 3호, pp. 1-27, 2017년
- [8] 김유신, 김남규, 정승렬, “뉴스와 주가: 빅데이터 감성분석을 통한 지능형 투자의사결정모형”, 지능정보연구, 제18권, 2호, pp. 143-156, 2012년
- [9] 정지선, 김동선, 김중우, “온라인 언급이 기업 성과에 미치는 영향 분석: 뉴스 감성분석을 통한 기업별 주가 예측”, 지능정보연구, 제 21권, 4호, pp. 37-51, 2015년