

뉴스 추출 증감사전을 통한 뉴스감정 물가지수와 실제 물가지수 비교

황재성, 이나경, 김문규, 박동현*

세종대학교

jshwang@sju.ac.kr, lnk2427@sju.ac.kr, mgkim@sju.ac.kr, parkdh@sejong.ac.kr

Comparison of the Consumer Price Index (CPI) and the News Emotion Price Index (NEPI) through News Extraction increase and decrease dictionary

JaeSung Hwang, NaKyung Lee, Mungyu Kim, Donghyeon Park
Sejong University

요약

물가지수는 화폐의 구매력 측정수단, 경기동향 판단지표, 명목가치를 실질가치로 환산하는데 필요한 지수이다. 이러한 지수는 월 또는 년 주기로 물가의 변화를 빠르게 반영하지 못한다. 이러한 한계점을 해결하기 위하여 자연어처리 기술을 이용하여 실시간 뉴스 기사를 추출하고 키워드 레벨의 분석을 한다면 일 단위, 시간단위의 변동성을 확인할 수 있다. 이를 위해 본 논문에서는 뉴스로부터 증감 단어사전을 생성하였고, 증감 단어사전과 2021년 한 해 동안의 뉴스 기사를 비교하여 기존의 물가지수를 보완한 뉴스감정 물가지수를 제안하였다. 그 결과 실제 물가지수와 논문의 뉴스감정 물가지수의 추이가 유사한 것을 알 수 있었다.

I. 서론

기존의 물가지수(Consumer Price Index: CPI) [1]는 화폐의 구매력 측정수단, 경기동향 판단지표, 명목가치를 실질가치로 환산하는데 필요한 지수로 국민 경제의 전 부문에 걸쳐 큰 영향을 끼치기 때문에 경제주체들의 관심사다[2]. 정부는 물가를 바탕으로 정책을 세우기 때문에[3] 물가의 변화를 빠르게 반영하는 것이 좋다. 하지만 이러한 물가지수는 월 또는 년 단위로 물가의 변화를 빠르게 보여주지 못한다는 단점이 있다. 본 논문은 이 한계점을 보완하고자 증감사전을 통한 뉴스감정 물가지수(News Emotion Price Index: NEPI)를 제안한다. 이를 위해 우리는 2만 개의 물가 관련 키워드가 들어간 경제 기사를 이용해 48개 단어로 이루어진 증감 단어사전을 생성하였다. 또한, 2021년 한 해 동안 생산된 경제 관련 기사와 증감 단어사전을 비교하였고 이를 통해 일 단위로 상승 또는 하락을 예측했다. 우리는 기존 물가지수(CPI)[1]와 뉴스기사의 신속성을 활용한 뉴스감정 물가지수(NEPI)의 상관관계 파악하고, 기존 물가지수의 한계점을 보완하여 더 신속하게 물가의 변화를 예측할 수 있다.

II. 본론

II.1 연구 방법 및 정의

감성사전[4]을 통해서 글의 내용이 긍정인지 부정인지 예측하는 연구가 많이 이루어지고

있다[5]. 하지만, 물가상승에 대한 뉴스기사에 긍정과 관련된 키워드가 많다고 해서 ‘이 뉴스기사는 물가상승을 예측한다.’라고 말할 수는 없을 것이다. 따라서, 물가와 관련된 기사를 통해 상승 또는 하락을 예측하기 위해서는 보다 더 객관적인 정보를 이용하는 것이 옳다고 판단했다.

이를 위해서 상승(하락)과 유사도가 일정 수준 이상인 단어들을 추출하여 증감 단어사전을 생성하였고, 증감 단어사전과 뉴스 기사를 비교하여 상승(하락) 관련 단어 개수를 세어주었으며, 해당 기사에 상승 관련 단어 개수와 하락 관련 단어 개수를 세어서 상승 기사인지 하락 기사인지를 판단한다. 이후, 기존 물가지수(CPI) 계산법에 착안하여 뉴스감정 물가지수(NEPI)를 수식을 사용하여 월 별로 값을 도출하고 그래프로 시각화 하였다.

II.2 데이터셋

연구가 진행되는 날로부터 가장 가까운 날들에 생산된 뉴스기사 중 뉴스기반통계검색 서비스[6]를 통해 물가와 관련 있는 키워드가 들어간 기사 2만 개를 빅카인즈 사이트[8]를 통해 수집했고 FastText[7]를 이용하여 단어를 임베딩하였다. “상승”(“하락”) 과 유사도가 0.8 이상인 단어들을 가져와서 리스트에 넣어주고, 리스트 내의 각각의 단어들과 0.7 이상의 유사도를 가지는 단어들을 뽑아서 1차적으로 단어사전을 생성했다. 이렇게 각각 상승 관련 단어 임베딩 벡터와 하락 관련 임베딩 벡터를

*박동현: Corresponding Author (parkdh@sejong.ac.kr)

뽑아 비지도학습을 이용해 각 임베딩 벡터를 군집화 하였고, 군집화 된 임베딩 벡터들을 상승 관련 단어와 하락 관련 단어로 구분했다.

하지만, “상승” 이라는 단어의 임베딩과 “하락”이라는 단어 임베딩의 유사도가 0.9 이상이였기 때문에 중복되는 단어가 많다는 문제가 발생하였다. 이 문제를 해결하기 위해, 논문 참여자 3명이 각각 주식(상승:1, 하락:2, 무관:3)을 달아줬고, 다수결의 원칙에 따라서 각 단어들에 label을 달아 상승 25개, 하락 23개인 증감 단어사전을 생성했다.

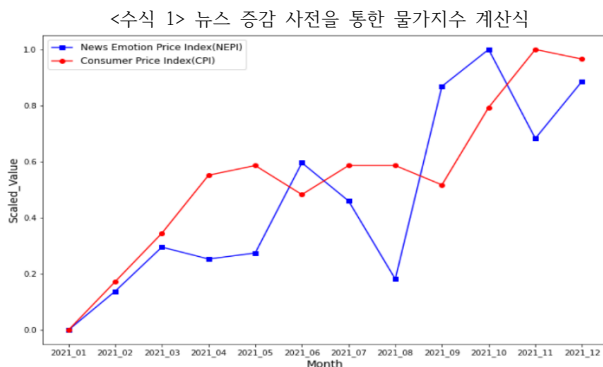
단어	UP/ DOWN	Annotator 1	Annotator 2	Annotator 3	최종
급등	UP	1	1	1	1
급상승	UP	1	1	1	1
상승	UP	1	1	1	1
상승세	UP	1	1	1	1
거듭	-	1	3	3	3
폭락	DOWN	2	2	2	2
밀바닥	DOWN	2	2	2	2
바닥	DOWN	2	2	2	2
하락	DOWN	2	2	2	2
가까스로	-	2	3	3	3

<그림 1> 뉴스 증감 단어사전 예시

II.3 뉴스감정 물가지수(NEPI) 제안

<그림 1>의 예시와 같이 생성된 증감 단어사전과 2021년 한 해 동안 생성된 (물가 관련 단어가 포함된) 뉴스기사 224,323개를 비교해서 각 기사에 상승 관련 단어가 몇 개 들어있는지, 하락 관련 단어가 몇 개 들어있는지를 계산하였다. 이때 상승 관련 단어가 더 많다면 그 기사는 상승 label을 달아주고 하락 관련 단어가 더 많다면 그 기사는 하락 label을 달아주었다. Labeling을 한 후, <수식 1>의 식을 이용해서 일 단위로 점수를 예측한다.

$$NEPI = \frac{SUM(상승 예측 기사) \times (+1) + SUM(하락 예측 기사) \times (-1)}{SUM(전체 기사)}$$



<그림 2> 기존의 물가 지수(CPI)와 증감사전 통한 뉴스감정 물가지수(NEPI) 비교

II.4 실험 결과

<그림 2>의 실제 사용되는 소비자 물가

지표(CPI)[1]와 증감 단어사전을 통한 뉴스감정 물가지수(NEPI) 그래프를 보면 (예외인 구간도 있지만) 유사한 추이를 보여주고 있다. 피어슨 상관계수가 0.734로 유사한 추이를 보여주고 있다는 내용을 뒷받침한다.

III. 결론 및 한계점

III.1 결론

물가에 대한 여러 수치적 근거들을 이용하여 물가지수를 만들어서 여러 가지 정책에 활용하는 방안은 많은 국민에게 도움이 된다. 우리는 경제 뉴스에 대한 자연어처리를 이용하여 이러한 수치 통계적인 지수를 보완하고자 하였다.

이를 위해 물가 관련 경제 기사를 토대로 증감사전을 생성하였고 이 증감사전을 기준으로 1년 치의 뉴스 기사를 분석하였다. 기존 물가지수와 논문의 증감사전을 통해 생성한 물가지수는 상관계수 계산 결과 유의미한 관계가 있음을 확인할 수 있었다. 이와 비슷하게, 다른 많은 분야에서도 자연어처리를 이용해서 수치적인 표현의 한계를 개선하는 방안이 연구된다면 유용할 것이라고 기대한다.

III.2 한계점

우리는 감성사전보다 객관적인 증감 단어사전을 생성하여 뉴스기사의 객관적 정보들을 추출하여 물가의 상승 또는 하락을 예측하고자 하였다. 하지만 여러 단어가 합쳐져 다른 의미를 생성하는 경우, 의미가 함축되어 나타나는 경우 등의 변수를 고려하지 못했다.

논문에서 제시하는 단어사전은 뉴스 기사의 일부만 보고 예측한다는 아쉬움이 있다. 하지만 기존 수치지수의 한계를 보완하는 용도로는 충분히 그 쓰임의 활용도가 있을 것으로 기대한다. 또한 각 단어를 보는 것뿐만 아니라 두 단어 이상 혹은 문장 단위 등 범위를 더 넓게 보면서 사전을 보완한다면 보다 더 유의미한 증감 단어사전을 만들 수 있을 것이라 기대한다.

IV. Acknowledgement

이 논문은 2022년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(NRF-2022R1F1A1069639)

참 고 문 헌

- [1] 소비자물가지표 통계청 :
<http://kostat.go.kr/incomeNcpi/cpi/index.action>
- [2] 한국은행-경제교실-경제지표해설 :
http://www.csr.co.kr/pds_data/11_2_0.htm
- [3] CPI 소비자물가지수 :
<https://kostat.go.kr/incomeNcpi/cpi/index.action>
- [4] KNU 감성사전 :
<http://dilab.kunsan.ac.kr/knusl.html>
- [5] 한국어 감성사전 :
<https://github.com/park1200656/KnuSentiLex>
- [6] 뉴스기반통계검색서비스:
<https://data.kostat.go.kr/social/keyword/index.do>
- [7] FastText :
<https://fasttext.cc/docs/en/support.html>
- [8] 빅카인즈 : <https://www.bigkinds.or.kr/>
- [9] 송민채, 신경식 뉴스기사를 이용한 소비자의
경기심리지수 생성, 지능정보연구 = Journal of
intelligence and information systems v.23 no.3 , 2017년,
[https://scienceon.kisti.re.kr/srch/selectPORSrchArticle](https://scienceon.kisti.re.kr/srch/selectPORSrchArticle.do?cn=JAKO201730475989205)
.do?cn=JAKO201730475989205
- [10] 이창용, 진효원, 문호석 텍스트마이닝을 이용한
북한 동향 핵심단어 사전 구축 및 분석, 한국경영과학회
2016년 춘계공동학술대회논문집,
[https://www.dbpia.co.kr/journal/articleDetail?nodeId=](https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE07168571)
NODE07168571