

# BOK 이슈노트

## 인공지능 언어모형을 이용한 인플레이션 어조지수 개발 및 시사점



한승욱

한국은행 디지털혁신실  
혁신기획팀 과장  
Tel. 02-759-4619  
swhan@bok.or.kr

김태완

한국은행 디지털혁신실  
디지털신기술반 과장  
Tel. 02-759-4862  
kimtwan@bok.or.kr

이현창

한국은행 디지털혁신실  
디지털신기술반 반장  
Tel. 02-759-4869  
hyunchang.yi@bok.or.kr

2022년 10월 17일

뉴스기사, 소셜미디어와 같은 텍스트 데이터는 크게 두 가지 경로를 통해 경제 분석에 유용한 정보를 제공할 수 있다. 첫째, 지정학적 리스크, 이상기후 등 이벤트는 수입물가지수와 같은 공식통계에 앞서 뉴스 속보나 인터넷 커뮤니티에 먼저 반영되며 향후 인플레이션 수준에 대한 정보를 제공한다. 둘째, 거시경제 여건, 기업의 가격설정 행태 등에 대한 언론 보도는 기업의 인플레이션 기대를 변화시킴으로써 의사결정에 영향을 미칠 수 있다. 그러나, 이러한 중요성에도 불구하고, 텍스트 데이터는 계량화가 쉽지 않아 경제 분석에 활용하는 데 어려움이 크다는 평가도 있다(로버트 실러, 2013년 노벨 경제학상 수상).

최근 정보통신 및 인공지능 기술과 함께 텍스트 분석기법이 빠르게 발전하면서 경제 분석에 텍스트를 활용하려는 시도가 점차 늘고 있다. 대표적으로 뉴스기사에 반영된 경제 불확실성을 측정하거나 중앙은행 회의록에 반영된 통화정책 기조를 평가한 연구가 있으며, 각국 중앙은행도 텍스트에 반영된 경제주체의 심리를 파악하여 경기상황을 판단하거나 전망하기 위해 노력하고 있다. 이들 연구는 문서내 단어별 빈도를 계산하는 간단한 방법부터 단어의 문맥적 의미를 파악하는 딥러닝 알고리즘까지 다양한 분석기법을 활용하고 있다. 이 가운데, 최근 특히 주목받는 분석기법으로 인공지능 언어모형이 있다. 자연어는 문맥에 따라 단어의 의미가 달라지기 때문에, 정확도 높은 모형을 훈련하는 데 상당한 시간과 비용이 소요된다. 이에 대안으로 제시된 인공지능 언어

모형은 기존 훈련 결과를 재활용함으로써 개발 시간과 비용은 절감하면서도 문서 분류, 검색 등 과제에서 높은 정확도를 보이고 있다.

본 연구는 인공지능 언어모형을 이용하여 뉴스기사에 나타난 인플레이션 어조(tone)를 측정하고, 인플레이션 전망에 활용할 수 있는지 살펴본다. 먼저, 인공지능 언어모형을 이용하여 개별 문장의 인플레이션 어조를 상승, 중립, 하락으로 분류하고, 문장별 어조를 집계하여 전체 뉴스기사의 인플레이션 어조를 지수로 나타내었다. 다음으로, 어조지수와 소비자물가상승률의 관계를 분석하였다. 분석 결과, 어조지수는 물가상승률의 추세 전환을 1~2분기 선행하고, 전망모형의 예측력도 개선하는 등 인플레이션 전망에 유용한 정보를 제공하는 것으로 판단된다.

텍스트 분석기법이 발전하고 정보 원천으로서 텍스트 데이터에 대한 관심이 높아지면서 각국 중앙은행과 연구기관, 국제기구는 텍스트로부터 유용한 정보를 읽어내기 위해 노력하고 있다. 그러나, 아직 텍스트 데이터는 분석이 까다롭고, 의미있는 정보를 추출하기도 어렵다는 인식이 있다. 본 연구는 인공지능 언어모형을 활용하여 상대적으로 적은 시간과 비용으로 뉴스기사에 포함된 인플레이션 정보를 효율적으로 추출해낼 수 있음을 보였다. 향후 경제·금융 도메인에 특화된 인공지능 언어모형 및 분석 인프라가 갖추어진다면, 경제 분석, 업무효율화 등 과제에서 텍스트 데이터 활용이 크게 증가할 것으로 기대된다.

- 본 자료의 내용은 한국은행의 공식견해가 아니라 집필자 개인의 견해라는 점을 밝힙니다. 따라서 본 자료의 내용을 보도하거나 인용할 경우에는 집필자명을 반드시 명시하여 주시기 바랍니다.
- 본고의 작성 과정에서 큰 도움을 주신 홍원석 디지털혁신실장, 박정필 혁신기획팀장, 안상기 데이터서비스팀장, 행내 세미나에 참석하여 귀중한 의견을 주신 조사국 물가동향팀 임웅지 과장, 경제통계국 통계연구반 서범석 과장을 비롯한 여러 세미나 참석자분들께 감사드립니다. 그리고 원고 심사 과정에서 유익한 의견을 주신 경제연구원 금융통화연구실 박용민 차장, 조사국 전망모형팀 김용건 과장께도 감사드립니다.
- 본 보고서와 관련된 문의나 자료요청은 김태완 과장에게 연락주시기 바랍니다.



한국은행

## I. 머리말

뉴스 기사, 소셜미디어와 같은 텍스트 데이터는 크게 두 가지 경로를 통해 경제 분석에 유용한 정보를 제공할 수 있다. 첫째, 지정학적 리스크, 이상기후 등 이벤트 발생은 수입물가지수와 같은 공식통계에 앞서 뉴스 속보나 인터넷 커뮤니티에 먼저 반영되며 향후 인플레이션 수준에 대한 정보를 제공한다. 둘째, 거시경제 여건 및 기업의 가격설정 행태에 대한 언론 보도는 기업의 인플레이션 기대를 조정·강화함으로써 의사결정에 영향을 미칠 수 있다. 한편, 2013년 노벨 경제학상 수상자인 로버트 실러(Robert J. Shiller) 교수는, 텍스트(내러티브)가 다양한 경로를 통해 우리의 경제적인 의사결정에 영향을 미치지만, 측정이 쉽지 않아 경제 분석 목적으로 활용하는 데 어려움이 크다고 지적한 바 있다.<sup>1)</sup>

최근 텍스트 데이터 분석기법이 발전함에 따라 중앙은행에서도 텍스트 데이터를 경제여건에 대한 경제주체의 심리를 측정하거나 경제전망을 위한 대체 데이터로 활용하기 위해 노력하고 있다. 대표적인 분석기법으로는 문장의 어조를 분석하는 어휘사전(lexicon), 문서의 주요 토픽을 추출하는 토픽 모델링 등이 있다. 최근에는 대규모 학습데이터에 대해 딥러닝 알고리즘을 훈련(지도학습)하여 텍스트의 의미를 보다 정확히 파악하려는 시도가 늘고 있다. 그러나, 어휘사전이나 토픽 모델링과 같은 기존 분석기법은 실행이 용이하다는 장점은 있으나 텍스트로부터 제한적인 정보만 추출 가능

하다는 한계가 있다. 또한, 딥러닝 알고리즘 훈련은 개별 과제마다 대규모 학습데이터를 구축하는 데 많은 시간과 비용이 소요되는 어려움이 있다.

이러한 기존 텍스트 분석기법에 대한 대안으로 제시된 것이 인공지능 언어모형이다. 최근 관심이 높아지고 있는 인공지능 언어모형은 실제 텍스트 분석에 앞서 대규모 텍스트 데이터에 대해 딥러닝 알고리즘을 훈련함으로써 언어에 대해 전반적인 이해도를 높이는 데 초점을 둔다. 이렇게 사전훈련된 언어모형(Pre-trained Language Model, PLM)은 소규모 학습데이터만으로도 개별 과제에 특화된 모형으로 손쉽게 미세조정(fine-tuning)할 수 있다. 이렇게 미세조정된 모형들은 어조 분석, 기계번역, 문서 요약 등 다양한 과제에서 탁월한 성과를 보이고 있다.

본 연구의 목적은 크게 두 가지로 구분할 수 있다. 하나는 효율적인 텍스트 데이터 분석을 위한 도구로서 인공지능 언어모형의 활용 가능성을 검토하는 것이고, 다른 하나는 최근 인플레이션 불확실성이 높아지는 상황에서 인플레이션 전망의 대체 데이터로서 텍스트 데이터의 유용성을 살펴보는 것이다. 이를 위해, 본 연구는 인공지능 언어모형을 이용하여 뉴스기사에 나타난 인플레이션 어조를 측정하고, 이를 지수화하여 소비자물가상승률 추세를 판단하거나 전망모형에 활용할 수 있는지 점검하였다.

본고의 구성은 다음과 같다. 먼저 II절에서 텍스트 데이터를 이용한 기존 연구의 주요 특징을 살펴본다. III절은 인공지능 언어모형을

1) "Trying to understand major economic events by looking only at data on changes in economic aggregates runs the risk of missing the underlying motivations for change. ... But, it is easy to see why economists often fall into this trap: abundant data exists for GDP, wage rates, interest rates, and tax rates, but data on narratives are spotty at best. Economists may be falling into what historian Jerry Z. Muller calls the 'tyranny of metrics'." (Shiller 2020: pp. 75~76)

미세조정하여 뉴스기사에 담긴 인플레이션 어조를 측정하는 프로세스를 설명한다. IV절은 인플레이션 어조지수를 산출하고 활용 가능성을 점검한다. 마지막으로 V절은 본고의 내용을 요약하고 시사점을 정리한다.

## II. 기존 연구

경제 분석에 텍스트 데이터를 활용한 연구는 뉴스기사를 이용한 경제 불확실성 측정이 대표적이다. Baker et al. (2016)은 미국 주요 신문사의 기사를 이용하여 경제 정책 불확실성(Economic Policy Uncertainty, EPU)을 지수화하였다. 이후 여러 국가를 대상으로 경제 불확실성을 측정하고, 경기, 주식시장 예측에 활용한 연구가 이어졌다.<sup>2)</sup>

주요국 중앙은행의 텍스트 분석은 주로 통화정책의 커뮤니케이션 효과를 분석하는 데 초점이 맞추어졌다. 미국 샌프란시스코 연준의 Shapiro and Wilson (2019)과 ECB의 Fraccaroli et al. (2020)은 각각 FOMC의 사록과 국회 청문회 녹취록에 나타난 통화정책 기조를 측정하고 기준금리 전망에 활용할 수 있는지 분석하였다. 우리나라의 경우에는 Park et al. (2019)이 한국은행 금융통화위원회 의사록을 이용하여 유사한 분석을 수행하였다.<sup>3)</sup>

이외에도 중앙은행의 텍스트 분석은 다양한 영역으로 확대되고 있다. 일본은행의 Nakajima et al. (2021)과 이탈리아 중앙은행의 Angelico et al. (2021), 노르웨이 중

앙은행의 Larsen et al. (2019)은 경제주체의 인플레이션 기대 수준을 측정하기 위해 각 경기설문조사와 트위터, 뉴스기사의 텍스트 데이터를 분석하였다. 그리고, 영란은행의 Kalamara et al. (2020)은 경제지표 전망의 정확도를 높이기 위해 뉴스기사를 활용하였다.

대표적인 텍스트 분석기법으로는 어휘사전 기반 분석(lexicon-based analysis) 및 토픽 모델링(topic modeling)이 있다. 어휘사전 기반 분석기법은 각 단어의 어조를 긍정, 부정, 중립 등으로 분류한 어휘사전을 정의한 다음, 각 어조에 해당하는 단어의 상대적 비중으로 문장 또는 문서의 어조를 파악한다. 이 방식은 텍스트의 어조를 손쉽게 측정할 수 있다는 것이 특징이다(Gentzkow et al., 2019). 앞서 소개한 Baker et al. (2016)과 중앙은행의 연구사례는 모두 텍스트에 나타난 어조를 측정하기 위해 어휘사전 기반 분석기법을 이용하였다.

토픽 모델링은 뉴스기사나 회의록과 같은 문서의 주요 토픽 또는 키워드를 추출하는 분석 기법이다. 토픽 추출은 통계적 추론을 통해 이루어진다. 대표적 토픽 모델링 기법인 LDA(Latent Dirichlet Allocation)는 깁스 샘플링(Gibbs sampling)을 이용하여 각 문서를 구성하는 단어와 가장 높은 확률로 관련된 토픽을 추출한다(Blei et al., 2003). 토픽 모델링은 최근 트렌드를 파악하거나 대규모 텍스트 데이터에서 특정 주제와 관련된 문서를 분류하는 데 주로 사용된다. Angelico et al. (2021)과 Larsen et al. (2019)은 인플레이션 관련

2) Baker et al. (2016)의 불확실성 지수와 관련 연구는 [www.policyuncertainty.com](http://www.policyuncertainty.com)에서 확인할 수 있다. 국내에서는 이금희 외(2020)가 주요 언론사의 뉴스기사를 이용하여 불확실성 지수를 산출하고 거시경제 예측에 활용하였다.

3) 중앙은행의 통화정책 커뮤니케이션 효과에 대한 기존 연구는 Hansen and McMahon (2016), Haldane and McMahon (2018), Hansen et al. (2019) 등을 참고하라.

문서를 추출하기 위해 LDA를 이용하였다.

〈표 1〉 주요국 중앙은행의 텍스트 분석 사례

국가	분석기법	분석목적
미국	어휘사전	통화정책 어조 측정
ECB	어휘사전	통화정책 어조 측정
영국	어휘사전, TF-IDF	뉴스심리 측정, GDP, CPI 등 전망
일본	어휘사전, Naïve Bayes	물가심리지수 개발, 인플레이션 기대 측정
이탈리아	어휘사전, LDA	인플레이션 기대 측정
노르웨이	어휘사전, LDA	인플레이션 기대, 정보 경직성 측정
한국	어휘사전	통화정책 어조 측정
	트랜스포머	뉴스심리지수 개발

자료: 각 기관의 working paper 참고

어휘사전, 토픽 모델링 등 분석기법은 별도의 학습데이터 구축이 필요 없고 분석과정이 간단하다는 장점이 있으나, 텍스트의 의미를 정확하게 파악하기 어렵다는 한계가 있다.<sup>4)</sup> 어휘사전 기반 분석기법은 단어들의 단순 조합으로 문장의 어조나 의미를 측정하므로 부정문이나 중의문과 같이 문장과 문장에 쓰인 각 단어의 의미가 일치하지 않는 경우 정확도가 떨어진다.<sup>5)</sup> 또한, 토픽 모델링은 문서의 토픽이 알고리즘에 의해 자동으로 추출되므로, 특정 주제와 관련하여 문서의 어조를 파악하는 목적으로 활용하는 데 어려움이 있다. 이러한 특성으로 인해 앞서 소개한 Angelico et al.(2021), Larsen et al.(2019) 등은 분석 대상 텍스트를 선별하는 전처리 목적으로 토픽 모델링을

사용하였다.

최근에는 각 문장에 라벨(어조)을 지정하여 구축한 학습데이터로 어조분류 모형을 훈련하는 텍스트 분석 사례가 늘고 있다.<sup>6)</sup> 새로 주어진 문장에 대해 통계모형이나 머신러닝 알고리즘이 적절한 라벨을 분류하도록 학습데이터를 이용하여 훈련하는 것이다. 인플레이션 기대 수준을 측정하기 위해 경기설문조사 자료를 이용한 Nakajima et al.(2021)은 전체 응답 가운데 1,500건을 표본으로 추출한 다음 각 응답에 대해 인플레이션 어조(상승·중립·하락)를 라벨로 지정하여 학습데이터를 구축하였다. 어조분류 모형으로는 통계추론 모형의 일종인 나이브 베이즈(Naïve Bayes) 알고리즘을 이용하였다. 뉴스기사에 나타난 경기상황 어조를 지수화한 서범석 외(2022)는 44만여 건의 경제·금융 뉴스기사 문장을 추출한 다음, 각 문장에 경기상황 어조(긍정·부정·중립)를 라벨로 지정하여 학습데이터를 구축하였다. 어조분류 모형으로는 텍스트 데이터에 특화된 딥러닝 알고리즘인 트랜스포머(Transformer)를 이용하였다.

이와 같은 텍스트 어조분류 모형은 대규모 학습데이터를 이용하여 훈련할 경우 분류의 정확도를 높일 수 있으나, 매번 모형을 개발할 때마다 학습데이터 구축에 상당한 시간과 비용이 소요된다는 단점이 있다. 이에 대한 대안으로 최근 주목받는 텍스트 분석기법이 전이학습(transfer learning)을 이용한 인공지능 언어모형이다. 전이학습은 먼저, 대규모 텍스트

4) 본 연구에서 특정 분석기법이 텍스트의 어조 또는 의미를 파악한다는 것은 동 분석기법이 주어진 텍스트에 대해 텍스트의 실제 어조(예, 상승·중립·하락 중 하나)를 매칭시킬 수 있음을 의미한다. 텍스트의 실제 어조는 텍스트를 읽은 연구자가 텍스트에 부여한 라벨(어조)을 의미한다.

5) 예를 들어 "가격 인상의 가능성은 크지 않다."를 어휘사전 기반 분석기법으로 정확하게 분석하기 위해서는 (가격, 인상, 가능성, 없다)와 같이 가능한 모든 단어의 조합을 미리 정의해야 한다는 어려움이 있다.

6) 학습데이터 예시는 III. 인공지능 언어모형을 활용한 인플레이션 어조 분석의 〈표 2〉를 참조하라.

데이터에 대해 인공지능 언어모형을 훈련하여 일반적인 언어구조를 파악하도록 하고, 어조 분류와 같은 특정 과제를 수행할 때는 앞서 훈련된 언어모형을 어조분류 모형으로 미세조정하는 순서로 이루어진다. 이처럼 사전훈련된 언어모형을 활용할 경우 소규모 학습데이터로도 정확도 높은 어조분류 모형을 개발할 수 있다는 장점이 있다. 본 연구는 효율적인 텍스트 분석을 위해 사전훈련된 인공지능 언어모형(이하, 인공지능 언어모형)을 이용하였다.

### III. 인공지능 언어모형을 활용한 인플레이션 어조 분석

#### 1. 인공지능 언어모형

언어모형은 단어의 조합과 순서를 바탕으로 주어진 문장의 구조를 파악함으로써 이어지는 단어 또는 문장을 예측하는 모형을 의미한다. 특히, 인공지능 언어모형(AI language model)은 대규모 텍스트 데이터를 이용하여 수백만 개에서 수천억 개의 파라미터로 구성된 딥러닝 알고리즘을 훈련함으로써 단어의 문맥적 의미를 효과적으로 포착한다.<sup>7)</sup> 텍스트 분석을 위해 고안된 트랜스포머는 문장 내 각 단어들의 다층적 관계를 포착할 수 있는 구조(self-attention)를 여러 단계 중첩하는 방식으로 설계되었다(Vaswani et al., 2017).

트랜스포머에 기반한 언어모형은 각 단어의 문맥적 의미를 파악하므로 기계번역, 질의응답과 같은 다양한 자연어처리 과제에서 높은 정확도를 보이고 있다.

한편, 이러한 복잡한 구조의 모형이 높은 성능을 내기 위해서는 대규모 학습데이터와 이를 처리할 수 있는 고성능 전산자원을 갖추어 언어모형을 훈련해야 하는데, 이 때 상당한 시간과 비용이 소요된다. 이에 대안으로 제시된 것이 전이학습(transfer learning)이다. 전이학습이란 한 과제에 대해 학습한 모형을 다른 과제 수행에 재사용하는 기법을 의미하며, 언어모형 사전훈련과 미세조정 두 단계로 진행된다. 먼저, 사전훈련에서는 위키백과 등 대규모 텍스트 데이터에 대해 자가지도방식(self-supervised learning)<sup>8)</sup>으로 빈칸에 들어갈 단어를 예측하는 것과 같은 일반적인 텍스트 분석 과제(upstream task)를 수행할 언어모형을 훈련한다. 다음 단계에서는 이렇게 사전훈련된 언어모형을 어조분류, 문장요약 등 세부 과제(downstream task)별로 필요한 데이터만 추가로 학습하는 미세조정(fine-tuning)을 통해 고성능의 모형을 개발한다.

이러한 효율성으로 인해 경제·금융 분야에서 인공지능 언어모형을 활용하려는 시도가 활발하다. 인공지능 언어모형을 활용한 대표적인 연구로 통화정책 어조분석(Lee et al., 2021), 추가예측(장은아 외, 2020), 경제·금융 텍스트 감성분석(Araci, 2019) 등이 있다.

7) 예를 들어, 적절히 학습된 언어모형은 '물' 뒤에는 '달리다' 보다 '마시다'가 이어질 가능성이 큰 것으로 평가한다. 언어모형은 크게 단어 간 통계분포를 이용한 통계적 언어모형과 인공신경망 구조를 이용한 인공지능 언어모형으로 구분할 수 있다. 본 연구는 최근 높은 성과를 내며 빠르게 발전하고 있는 인공지능 언어모형을 이용하였다. 이와 대조적으로 전통적인 텍스트 분석기법은 독립적인 단어의 의미를 분석하는 것에 초점을 두었다. 각 단어의 문서별 사용 빈도(Bag of Words, TF-IDF)나 문장 내에서 함께 사용된 단어들의 확률 분포(Word2vec, GloVe) 등을 바탕으로 단어의 고유한 의미를 벡터로 표현하고 이를 기반으로 텍스트를 분석하는 방식이다. 그러나 이러한 방식은 문맥(context)에 따라서 단어의 의미가 달라지는 가능성을 배제하는 한계가 있다.

8) 수작업으로 별도의 라벨을 지정할 필요가 없다는 측면에서 비지도학습(unsupervised learning)과 유사하나 학습할 라벨을 알고리즘이 자체적으로 만들어서 사용하는 과정을 제외하면 지도학습과 유사하다.



그러나, 아직 기존 텍스트 분석기법에 비해 인공지능 언어모형은 구조와 분석방법이 복잡하여 경제분석에 활용하기 어려운 것으로 평가되고 있다(Shapiro et al., 2017). 실제 중앙은행을 비롯한 대부분 경제 분석에서는 아직 기존 텍스트 분석기법을 이용한 연구가 큰 비중을 차지하고 있다. 본 연구는 효율적인 어조분류 모형 개발과 향후 활용성을 고려하여 인공지능 언어모형을 이용하였다.<sup>9)</sup>

## 2. 미세조정

사전훈련된 언어모형을 미세조정하여 뉴스 기사에 나타난 인플레이션 어조를 측정하는 절차는 다음과 같다. 미세조정을 위한 학습데이터는 네이버 뉴스에서 물가 관련 키워드<sup>10)</sup>로 검색하여 수집한 뉴스 기사를 이용하였다. 데이터는 2002년 2월부터 2022년 6월까지 총 188만건(일평균 418건)의 뉴스 기사에서 수집된 6,406만개(일평균 8,653개) 문장이다. 학습데이터는 전체 문장 중 5,000개를 임의로 추출한 다음 아래 <표 2>와 같이 대상품목, 현재 어조, 미래 어조 등 3개 카테고리에 대해 라벨을 지정하여 구축하였다. 대상품목은 식품, 원자재, 부동산 등으로 구분되며, 문장이 어떤 대상에 대한 인플레이션 어조를 포함하는지 나타낸다. 현재(미래) 어조는 대상품목에 대한 현재(미래) 가격변동에 대한 어조를 상승·중립·하락·알수없음으로 구분한 것이다.<sup>11)</sup>

이렇게 준비된 학습데이터는 인공지능 언어모형을 총 3개의 분류모형(대상품목, 현재 어조, 미래 어조)으로 미세조정하는 데 이용된다.

<표 2> 인플레이션 어조분류 학습데이터(예시)

문장	주제 <sup>1)</sup>	어조 <sup>2)</sup>	
		현재	미래
휘발유 가격, 1400원대로 떨어졌다. 다음 주도 하락 전망	8	1	1
이에따라 추가상승 모멘텀인 전기요금 인상이 물건너 가는 것이 아니냐는 의견이 제기돼 관심이다.	5	0	2
오는 18일까지 '포인포', '치크' 등 인기 아동복 브랜드 전 품목을 최대 60% 인하된 가격에 제안한다.	4	1	0
가격은 기존 갤럭시S7엣지보다 조금 비쌀 것으로 예상된다.	9	0	3

주: 1) 1) 4(의류 및 신발), 5(주택, 수도, 전기 및 연료), 8(교통), 9(통신)을 의미, 전체 목록은 <부록1> 참고

2) 0(알 수 없음), 1(하락), 2(중립), 3(상승)

미세조정에는 전체 학습데이터의 80%를 이용하였으며, 나머지 20%에 대해 미세조정된 분류모형의 텍스트 분류 정확도, 재현율, 정밀도, F1-Score를 측정하였다. <표 3>의 성능 측정 결과를 보면, 카테고리별로 70~80% 내외의 정확도를 보였다. 오분류된 문장의 경우 판단 기준에 따라 분류 결과가 맞는 것으로 볼 수 있는 사례<sup>12)</sup>가 상당수 있었다. 특히 주제분류의 정확도가 상대적으로 낮은 것은 하나의 문장이 여러 대상품목을 언급하는 경우가 많은 것에 기인한 것으로 판단된다.<sup>13)</sup>

9) 대표적인 사전훈련 언어모형으로 텍스트 의미 파악에 강점을 가진 BERT와 텍스트 생성에 강점을 가진 GPT가 있다. 본 연구는 뉴스기사의 인플레이션 어조분류를 위해 한국어 버전 BERT(KoBERT)의 경량화 버전인 DistillKoBERT (Park, 2019)를 이용하였다.

10) 물가, 가격, 값, 인플레이션, 디플레이션, 비용, 부담, 요금 등 총 8개의 물가 관련 키워드를 기준으로 검색하였다.

11) 세부 분류기준 및 분류 결과는 <부록 1>을 참조하라.

12) '다만 근원물가 상승세가 주춤하고 연말로 갈수록 유가 상승효과도 작아져 물가가 1%대 중반으로 떨어질 가능성을 열어졌다.'라는 문장에 '하락' 라벨을 지정하였으나, 분류된 라벨은 '중립'이다.

13) '국제유가와 곡물 등 국제 원자재 가격이 폭등하면서 공산품도 전월대비 0.3%가 올랐다.'라는 문장에 '공산품' 라벨을 지정하였으나, 분류된 라벨은 '원자재'이다.

〈표 3〉인플레이션 어조 분류 결과<sup>1)</sup>

(%)

	정확도 <sup>2)</sup>	정밀도 <sup>3)</sup>	재현율 <sup>4)</sup>	F1-점수 <sup>5)</sup>
주제	73.1	72.8	73.2	72.2
현재어조	77.2	75.7	77.2	76.0
미래어조	82.6	81.3	82.6	81.3

주: 1) 정밀도, 재현율, F1-점수는 평균값

$$2) \text{정확도(Accuracy)} = \frac{\text{True Positives} + \text{True Negatives}}{\text{전체 데이터}}$$

$$3) \text{정밀도(Precision)} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

$$4) \text{재현율(Recall)} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

$$5) \text{F1-점수} = 2 \times \frac{\text{정밀도} \times \text{재현율}}{\text{정밀도} + \text{재현율}}$$

평균하여 현재와 미래 각각에 대해 인플레이션 어조지수를 계산하였다.

구체적인 산식은 아래 식—1 및 식—2와 같다. 뉴스기사  $a$ 의 전체 문장 중 상승(중립, 하락) 어조로 분류된 문장수를  $n_a^{\text{상승}}$ ( $n_a^{\text{중립}}$ ,  $n_a^{\text{하락}}$ )라고 하면, 뉴스기사  $a$ 의 인플레이션 어조  $i_a$ 는  $n_a^{\text{상승}}$ 의 상대적 비중으로 측정된다(식—1). 또한, 특정 시점  $t$ 의 인플레이션 어조지수  $I_t$ 는  $t$  시점에 보도된 전체 뉴스기사  $A_t$ 를 대상으로 각 뉴스기사의 인플레이션 어조를 평균하여 계산한다(식—2). 이에 따라, 각 뉴스기사의 인플레이션 어조와 각 시점의 인플레이션 어조지수는 -1(하락)과 1(상승) 사이의 값을 갖는다.

## IV. 인플레이션 어조지수

### 1. 지수 산산

앞 절에서 설명한 분류모형을 이용하여 인플레이션 어조지수(inflation tone-index)를 산출하였다. 먼저 각 뉴스기사에 포함된 문장들의 인플레이션 어조를 현재와 미래 각각에 대해 분류한 다음 상승 어조 문장의 상대적 비중으로 해당 뉴스기사의 현재와 미래 어조를 측정하였다. 다음으로 각 시점에 보도된 전체 뉴스기사를 대상으로 각 뉴스기사의 어조를

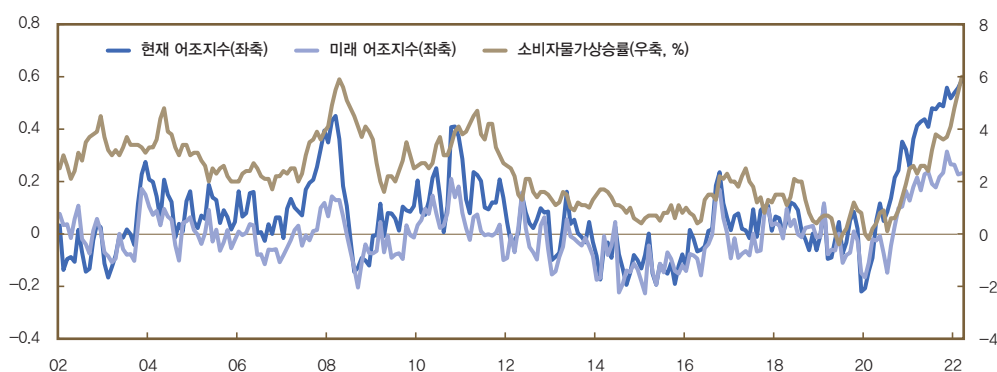
$$i_a = \frac{n_a^{\text{상승}} - n_a^{\text{하락}}}{n_a^{\text{상승}} + n_a^{\text{중립}} + n_a^{\text{하락}}} \quad (\text{식—1})$$

$$I_t = \frac{1}{|A_t|} \sum_{a \in A_t} i_a \quad (\text{식—2})$$

$n_a^{\text{상승}}$ ( $n_a^{\text{중립}}$ ,  $n_a^{\text{하락}}$ )는 뉴스기사  $a$ 의 전체 문장 중 인플레이션 어조가 상승(중립, 하락)인 문장 수,  $|A_t|$ 는  $t$  시점에 작성된 전체 뉴스기사 수

2002년 이후 월별 인플레이션 어조지수는 〈그림 1〉과 같다. 인플레이션 어조지수와 소비자물가상승률(전년동기대비)이 유사한

〈그림 1〉인플레이션 어조지수와 소비자물가상승률



패턴을 보이는 가운데, 대체로 어조지수가 인플레이션을 선행하는 모습을 보이고 있다. 한편, 앞에서 설명한 것과 같이 현재 인플레이션에 대한 어조와 향후 인플레이션에 대한 어조를 구분하였으며, 어조지수와 소비자물가지수를 비교할 수 있도록 대상품목이 부동산, 금융자산 등으로 분류된 뉴스기사는 어조지수 시산 과정에서 제외하였다.<sup>14)</sup>

## 2. 어조지수와 물가지수의 관계

인플레이션 어조지수와 소비자물가지수, 생산자물가지수 등 주요 물가지수가 높은 상관관계를 보이는 가운데, 전반적으로 현재 어조지수가 미래 어조지수에 비해 보다 뚜렷한 상관관계를 보였다(〈표 4〉). 인플레이션 어조지수(현재)는 5개월 후(상관시차 -5) 소비자물가상승률(전년동기대비)과 가장 높은 양의 상관관계(0.57)를 보였다. 한편, 생산자물가지수(0.76), 수출·수입물가지수(0.62, 0.74) 등이 어조지수와 강한 상관관계를 보인 것에 비해 농산물및석유류제외지수는 상대적으로 낮은 계수(0.45)와 긴 상관 시차(-8)을 보였다. 이는 국내 경기의 영향을 크게 받는 서비스 가격에 비해 글로벌 공급충격, 지정학적 리스크의 영향을 크게 받는 원자재, 중간재 관련 언론보도의 비중이 더 크기 때문인 것으로 보인다.

〈표 4〉 어조지수와 물가지수 간 교차상관계수<sup>1)</sup>

물가지수	어조지수(현재)		어조지수(미래)	
	계수	시차 <sup>2)</sup>	계수	시차 <sup>2)</sup>
소비자물가지수	0.57	-5	0.50	-5
(농산물및석유류제외)	0.45	-8	0.36	-8
생산자물가지수	0.76	-3	0.68	-2
수출물가지수(원화)	0.62	-5	0.48	-3
수입물가지수(원화)	0.74	-3	0.63	-2
생활물가지수	0.60	-2	0.53	-2

주: 1) 어조지수와 물가상승률(전년동기대비)의 상관계수가 가장 큰 값을 갖는 시차 기준(02.2~22.6월중)

2) 어조지수가 물가지수에 선행(후행)하는 경우 상관시차가 음수(양수)

다음으로, 인플레이션 어조지수가 소비자물가지수의 변동을 설명하거나 예측할 수 있는 고유한 정보(충격)를 갖는지 점검하기 위해 충격반응함수를 분석하였다. 모형 식별 문제에 강건한 것으로 평가되는 국소투영법(local projection, Jordà, 2005)을 이용하여 아래(식—3)과 같이 어조지수에 대한 인플레이션의 충격반응함수를 추정하였다.<sup>15)</sup> 어조지수는 소비자물가상승률과 강한 상관관계를 보인 현재 어조지수를 이용하였다. 또한, 설문조사를 통해 경제주체의 인플레이션 기대 수준을 측정하는 기대인플레이션율(소비자동향조사)에 대해서도 충격반응함수를 추정하여, 인플레이션 어조지수와 비교하였다.

$$\bar{\pi}_{t,t+h} = \beta_{x,0}^h x_t + \sum_{q=1}^Q \beta_{x,q}^h x_{t-q} + \sum_{p=0}^P \beta_{\pi,p}^h \pi_{t-p} + \epsilon_h \quad (\text{식—3})$$

$\pi_t$ 는 소비자물가상승률(전기대비),  $\bar{\pi}_{t,t+h}$ 는  $t \sim t+h$  평균 소비자물가상승률,  $x_t$ 는 인플레이션 어조지수(현재) 또는 기대인플레이션율,  $Q$ 와  $P$ 는 BIC를 최소화하는 12로 설정

14) 〈부록 1〉의 17개 대상품목 중 부동산, 금융자산을 제외한 15개 대상품목을 포함하였다. 또한, 교통, 보건, 교육 등 대상품목에 따라 나누어 시산한 어조지수(현재)는 해당 품목(지출목적별지수 대분류 기준)별 소비자물가지수와 유사한 움직임을 보였으나, 오락 및 문화, 보건 등 품목의 경우 동조관계가 뚜렷하지 않았다(〈부록 2〉). 한편, 원자재, 부동산 등에 대한 어조지수(현재)는 생산자물가지수, 부동산매매가격지수 등과 동조된 모습을 보였다.

15) Stock and Watson (2016)을 따라 소비자물가상승률의 기간 중 평균을 종속변수로 이용하였다.



분석 결과, 인플레이션 어조지수(현재)가 상승할 경우 소비자물가상승률(전기대비)이 유의하게 상승하였다가 18개월에 걸쳐 점차 감소하는 것으로 나타났다. 이에 반해 기대인플레이션율은 소비자물가상승률에 유의한 영향을 미치지 않는 것으로 추정되었다. 이는 최병재(2012)가 지적한 바와 같이 설문조사에 반영된 경제주체의 인플레이션 기대가 최근 물가상승률의 영향을 크게 받기 때문으로 판단된다.

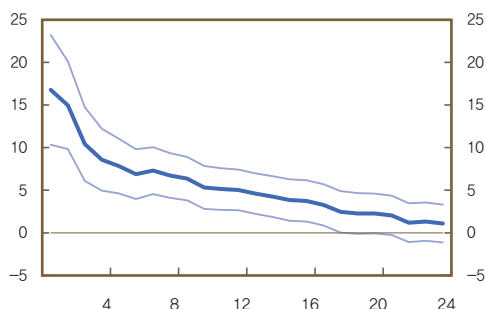
### 3. 변곡점 분석

변곡점 분석(turning point analysis)을 통해 인플레이션 어조지수가 인플레이션 추세를 평가하는 데 도움이 되는지 점검하였다. 물가상승률은 단기 변동성이 크기 때문에 다음 달 또는 다음 분기의 물가상승률을 예측하는 것보다 물가상승률의 추세를 판단하는 것이 더 중요할 수 있다. 본 연구는 인플레이션 현재 어조지수와 소비자물가상승률의 분기 기준 고점(peak)과 저점(trough)을 변곡점으로 식별하고,<sup>16)</sup> 어조지수 고점(저점)이 소비자물가상승률 고점(저점)을 선행 또는 동행하는지 점검하였다.

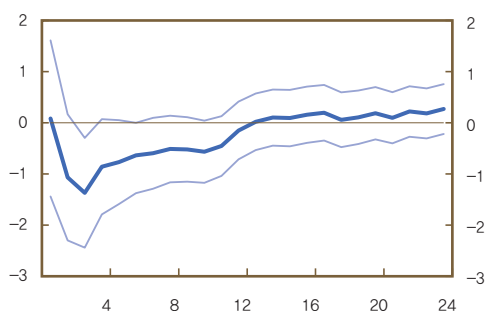
분석 결과, 인플레이션 어조지수의 추세 전환은 소비자물가상승률의 추세 전환을 선행하는 것으로 나타났다(표 5). 2002년 2분기부터 2022년 2분기까지 중 인플레이션 어조지수 변곡점은 총 8건 식별되었으며, 이 가운데 소비자물가상승률 변곡점이 어조지수를 1년 이내 후행한 사례는 7건으로 집계되었다. 대체로 어조지수가 고점(저점)을 지나고 1~2분기 이후 소비자물가상승률도 고점(저점)을 지나는 모습을 보였다. 특히, 어조지수의 고점(저점)을 소비자물가상승률 저점(고점)이 뒤따르는 경우는 없었다. 이는 인플레이션 어조지수가 소비자물가상승률의 추세를 판단하는 데 활용될 수 있음을 시사한다.

〈그림 2〉 충격반응함수 추정 결과

어조지수(현재) → 소비자물가상승률



기대인플레이션율 → 소비자물가상승률



주: 세로축은 당기 어조지수(현재) 및 기대인플레이션율 계수( $\beta_{t,0}^A$ )의 추정값(식-3), 가로축은 월(h), 가는 실선은 95% 신뢰구간

16) 일반적으로 시계열의 고점과 저점은 각각 국소최대값(local maximum)과 국소최소값(local minimum)으로 정의된다(OECD, 2007). 그러나, 단기적 등락이 빈번한 시계열의 경우 고점과 저점이 너무 많이 식별되고 중·장기 추세를 파악하기 어렵다는 문제가 있다. 본 연구는 각 분기의 값이 전·후 6분기(총 13분기) 이내 가장 큰(작은) 경우 고점(저점)으로 정의하였다.

〈표 5〉변곡점 식별 결과

(분기)

	인플레이션 어조지수(현재)	소비자물가 상승률	선행시차
고점	04.1분기	04.3분기	-2
	08.2분기	03.3분기	-1
	11.1분기	11.3분기	-2
	17.1분기	17.3분기	-2
저점	06.4분기	07.1분기	-1
	09.1분기	09.3분기	-2
	15.3분기	15.2분기	+1
	20.2분기	20.2분기	0

주: 선행시차는 어조지수 변곡점이 물가상승률 변곡점을 선행(후행)하는 경우 음수(양수)

〈그림 3〉은 인플레이션 어조지수(현재)와 소비자물가상승률의 변곡점을 표시하였다. 저물가 기조가 장기간 이어진 2011~16년을 제외하고 어조지수 변곡점은 소비자물가상승률 변곡점을 1~2분기 시차를 두고 선행하였다. 그러나, 감염병 유행으로 경제 불확실성이 확대된 2020년의 경우 어조지수와 물가상승률 모두 2020년 2분기 저점을 기록한 이후 급격히 상승하였다.

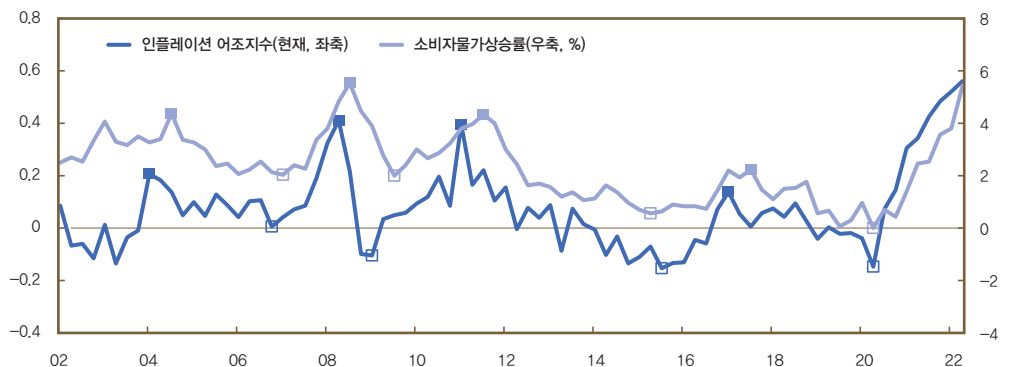
#### 4. 인플레이션 전망모형

앞서 살펴본 충격반응함수 추정모형(식-3)을 확장하여, 인플레이션 전망에 어조지수를 활용할 수 있는지 평가하였다. Stock and Watson(2016)을 따라 모든 변수를 분기화한 다음 아래 (식-4)와 같이 전망모형을 구성하였다. 종속변수는 전기대비 소비자물가상승률의 전망시계( $h$ ) 중 평균값을 이용하였다. 예측 변수로는 인플레이션 어조지수(현재), 기대인플레이션율, 생산자물가지수 등을 포함하였다(〈표 6〉). 또한, 전망모형의 예측력을 비교하기 위한 벤치마크 모형으로 AR모형과 AO모형(Atkeson and Ohanian, 2001)을 함께 사용하였다.

$$\bar{\pi}_{t,t+h} = \sum_{p=0}^P \beta_{\pi,p}^h \pi_{t-p} + \sum_{q=0}^Q \beta_{x,q}^h x_{t-q} + \epsilon_h \quad (\text{식-4})$$

$\pi_t$ 는 소비자물가상승률(전기대비),  $\bar{\pi}_{t,t+h}$ 는  $t \sim t+h$  평균 소비자물가상승률( $h = 2, 4, 8$ ),  $x_t$ 는 예측변수(생산자물가지수, 기대인플레이션율, 인플레이션 어조지수),  $Q$ 와  $P$ 는 모형별 RMSE가 최소화되도록 설정

〈그림 3〉 인플레이션 어조지수(현재) 및 소비자물가상승률 변곡점



주: 어조지수와 물가상승률의 고점(peak)은 속이 채워진 네모, 저점(trough)은 속이 빈 네모로 표시, 각 분기의 값이 전·후 6분기(총 13분기) 이내 가장 큰(작은) 경우 고점(저점)으로 정의

〈표 6〉 전망모형 비교

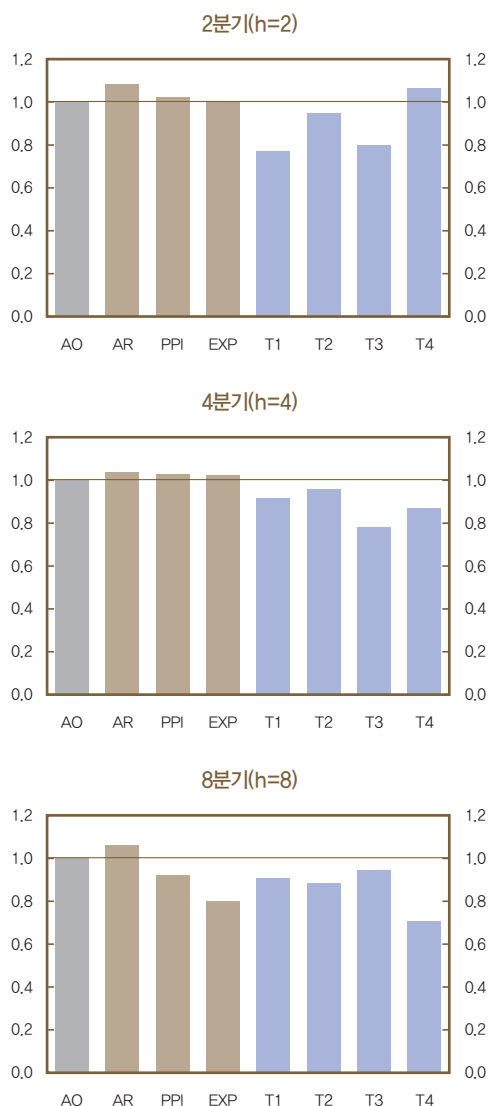
모형	모형구조
AO	$\bar{\pi}_{t,t+h} = \bar{\pi}_{t-4,t} + \epsilon_h$
AR(4)	$\bar{\pi}_{t,t+h} = \sum_{p=0}^4 \beta_{\pi,p}^h \pi_{t-p} + \epsilon_h$
모형	예측변수 <sup>1)</sup>
PPI	생산자물가지수
EXP	기대인플레이션율
T1	어조지수(현재)
T2	어조지수(미래)
T3	어조지수(현재, 원자재 대상)
T4	어조지수(현재, 공업제품 대상)

주: 1) PPI는 생산자물가지수의 전기대비 상승률, 나머지는 각 변수의 전기대비 차분

전망모형별 전망오차는 2013년 1분기부터 2022년 2분기까지의 전망시계별 평균 물가상승률을 WFV(Walk Forward Validation) 방식<sup>17)</sup>으로 예측한 값과 실제 인플레이션과의 차이로 계산하였다. 예측력은 전망오차의 제곱근평균자승오차(Root Mean Square Error, RMSE)<sup>18)</sup>로 계산하여 벤치마크 모형과 비교하였다.

전반적으로 인플레이션 어조지수(현재)를 사용한 모형이 벤치마크 모형보다 전망성도가 좋은 것으로 나타났다. 이는 어조지수가 인플레이션 전망에 유용한 정보를 포함한 것으로 해석된다. 전망시계별로 보면, 어조지수(현재)는 단기시계(2분기)에서, 원자재 및 공업제품 어조지수는 1년 이상 장기시계(4, 8분기)에서 모형의 예측력을 개선하는 효과를 보였다.

〈그림 4〉 전망시계별 전망모형 예측력 비교



주: 전망모형별 RMSE(13.1분기~22.2분기)를 벤치마크(AO) 모형 대비 비율로 표시

17) 각 예측 시점( $t, t+1, \dots, T-h$ )까지의 데이터를 이용하여 예측 시점 이후의 값을 예측하고 실제 값과의 오차를 이용하여 전망모형의 예측력을 평가하는 방식이다.

18)  $RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (\text{전망치}_t - \text{실제치}_t)^2}$

## V. 요약 및 시사점

본 연구는 인공지능 언어모형을 이용하여 뉴스기사의 인플레이션 어조를 측정하고, 인플레이션 전망에 유용한 정보를 제공할 수 있는지 살펴보았다. 인공지능 언어모형을 미세조정하여 개발한 인플레이션 어조분류 모형은 뉴스기사의 인플레이션 어조를 비교적 정확하게 분류하는 것으로 나타났다. 또한, 2002년 이후 뉴스 기사를 대상으로 시산한 인플레이션 어조지수는 여러 물가지수를 선행하는 모습을 보였으며, 특히 소비자물가상승률 추세를 판단하는 데 유용한 것으로 판단된다.

본 연구는 다소 간단한 인플레이션 전망모형을 구성하였으나, 전망모형의 예측력을 개선하고자 한다면 품목별로 세분화된 어조지수와 소비자물가지수 간 관계를 검토하고, 변수에 측에 특화된 머신러닝 알고리즘을 적용할 수 있을 것이다. 또한, GDP 성장률 전망과 같이 다양한 경제활동 관련 데이터가 필요한 경우(이현창 외, 2022), 관련 부문의 어조지수를 개발하여 대체 데이터로 활용할 수 있을 것으로 기대한다.

딥러닝 알고리즘을 적용한 텍스트 분석은 기존 텍스트 분석기법에 비해 텍스트의 어조를 정확하게 측정할 수 있다. 특히, 인공지능 언어모형은 상대적으로 적은 비용과 시간으로도 텍스트 분석의 정확도를 높일 수 있으므로, 여러 전망모형을 위한 대체 데이터 개발, 경제주체 심리 측정 등에 폭넓게 적용될 것으로 기대된다. 그러나, 아직 경제 분석에 인공지능 언어모형을 활용한 모범사례가 적고, 개인 연구자가 인공지능 언어모형을 개발하고 활용하는데 드는 비용을 감당하기 어려워 인공지능 언

어모형을 활용한 사례가 아직 드문 것이 사실이다.

이를 고려할 때, 본 연구는 텍스트 데이터에 인공지능 언어모형을 활용하여 경제 분석에 유용한 정보를 효율적으로 추출하는 모범사례를 제시한 의미가 크다고 하겠다. 이에 더 나아가, 향후 경제·금융 도메인에 특화된 인공지능 언어모형이 개발되고, 학습데이터 구축, 인공지능 언어모형 미세조정 등을 손쉽게 실행할 수 있는 분석 인프라가 구축된다면 텍스트 데이터를 이용한 조사연구 및 업무효율성 향상에 크게 기여할 것으로 기대된다.

한편, 텍스트 데이터는 데이터 특성상 기존 경제·금융 통계와 달리 잡음이 크고, 실제 거시경제 변수와 어떤 메커니즘을 통해 상호 영향을 주고받는지에 대한 이해가 부족하므로 활용에 각별한 주의가 요구된다. 이러한 측면에서 최근 많은 중앙은행이 경제 분석에 텍스트 데이터를 활용하기 위해 노력하고, 관련 연구 결과가 축적되는 것은 향후 텍스트 데이터 활용에 매우 긍정적이라고 할 수 있겠다.

## 〈부록 1〉

## 텍스트 분류기준 및 결과

## 1. 주제 분류

(개, %, 천개, %)

라벨	주제	설명	문장 수	
			학습데이터	전체데이터
1	물가전반	물가(소비자, 생산자물가 등)에 대한 설명	329 (6.6)	1,481 (2.3)
2	식품	식품 및 비주류음료	257 (5.1)	1,307 (2.0)
3	주류	주류 및 담배	205 (4.1)	147 (0.2)
4	의류	의류 및 신발	193 (3.9)	110 (0.2)
5	주택	주택임차료, 수도, 전기, 가스 및 기타연료	230 (4.6)	724 (1.1)
6	가정	가정용품 및 가사 서비스	197 (3.9)	148 (0.2)
7	보건	의료용품 및 장비, 병원 서비스 등	172 (3.4)	448 (0.7)
8	교통	운송장비, 운송 서비스, 개인운송장비 운영	169 (3.4)	1,834 (2.9)
9	통신	우편서비스, 전화 및 팩스 장비/서비스	251 (5.0)	1,245 (1.9)
10	오락	음향, 영상, 사진 및 정보처리 장비 등	234 (4.7)	459 (0.7)
11	교육	유치원, 초·중·고등학교	192 (3.8)	160 (0.3)
12	음식	음식, 숙박 서비스	156 (3.1)	356 (0.6)
13	기타	미용용품 및 미용 서비스, 기타 개인용품 등	178 (3.6)	331 (0.5)
14	원자재	원유, 금, 은, 철광석 등 원자재	170 (3.4)	979 (1.5)
15	부동산	아파트 등 부동산 매매가격	150 (3.0)	2,142 (3.4)
16	금융자산	부동산, 주식, 채권, 코인 등의 금융자산	113 (2.3)	2,700 (4.2)
17	공업제품	디스플레이 패널, 반도체 등 공업제품	122 (2.4)	733 (1.2)
0	알수없음	물가/가격과 관련 없는 내용	1,682 (33.6)	48,758 (76.1)
합 계			5,000 (100)	64,062 (100)

\* 괄호안은 비율을 의미

## 2. 현재어조 분류

(개, %, 천개, %)

라벨	현재어조	설명	문장 수	
			학습데이터	전체데이터
1	하락	가격(물가) 하락	924 (18.5)	2,653 (4.1)
2	중립	가격(물가) 변화 없음, 상승/하락세 완화	298 (6.0)	673 (1.1)
3	상승	가격(물가) 상승	900 (18.0)	4,003 (6.3)
0	알수없음	가격(물가) 변화를 알 수 없는 경우	2,878 (57.5)	56,733 (88.5)
합 계			5,000 (100)	64,062

\* 괄호안은 비율을 의미

## 3. 미래어조 분류

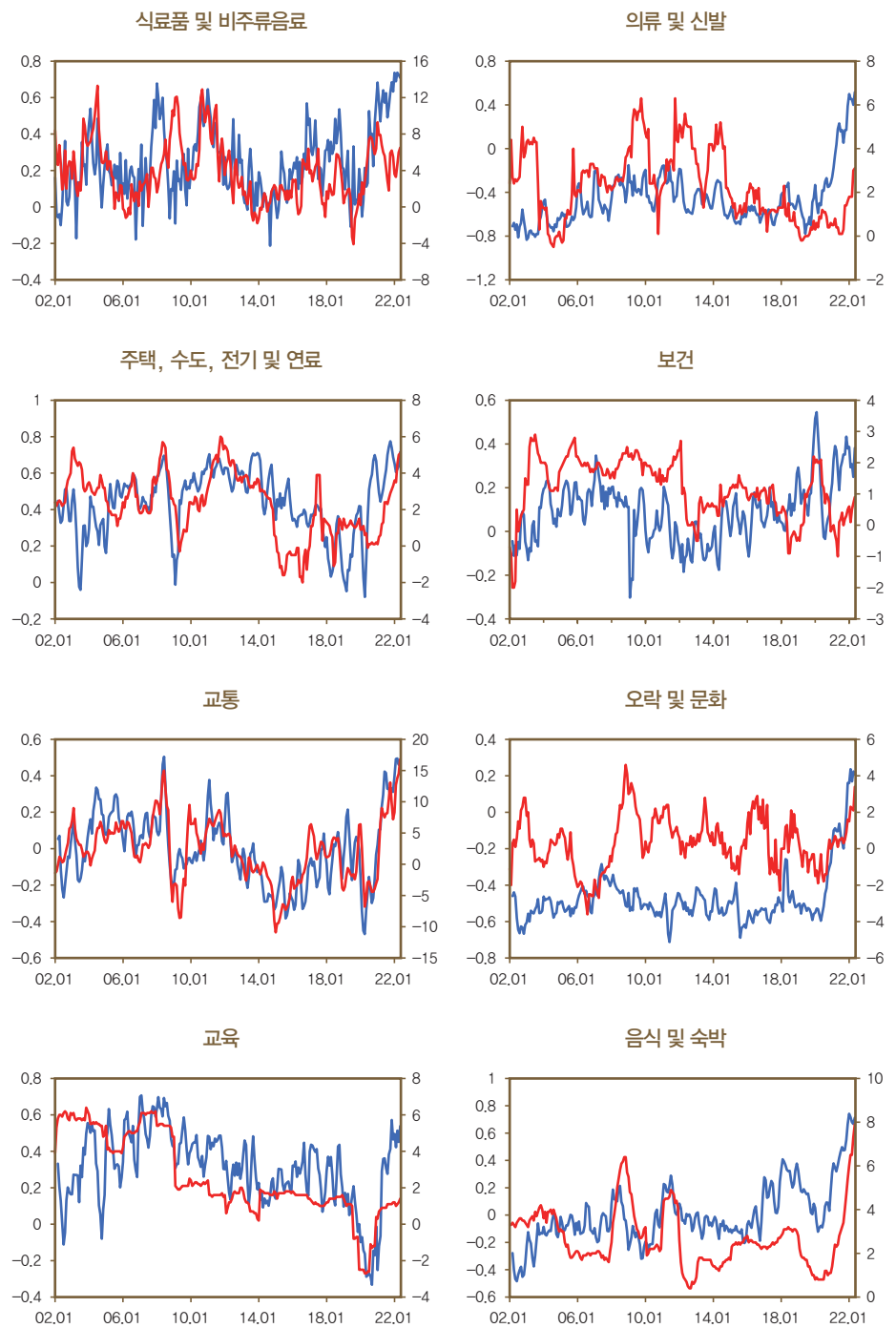
(개, %, 천개, %)

라벨	미래어조	설명	문장 수	
			학습데이터	전체데이터
1	하락	향후 가격(물가) 하락 예상	260 (5.2)	1,272 (2.0)
2	중립	향후 가격(물가) 변화 없음 예상	230 (4.6)	670 (1.0)
3	상승	향후 가격(물가) 상승 예상	379 (7.6)	1,139 (1.8)
0	알수없음	향후 가격(물가) 변화를 알 수 없는 경우	4,131 (82.6)	60,981 (95.2)
합 계			5,000 (100)	64,062 (100)

\* 괄호안은 비율을 의미



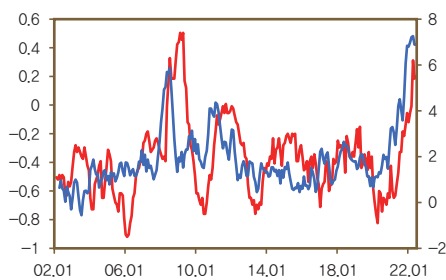
## 〈부록 2〉

부문별 인플레이션 어조지수와 소비자물가상승률<sup>1)</sup>

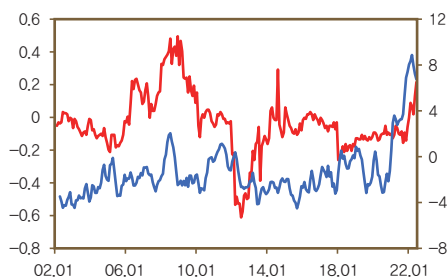
주: 1) 빨간색 실선은 부문별 소비자물가상승률(전년동기대비, 우측, %), 파란색 실선은 인플레이션 어조지수(현재, 좌측)의 3개월 이동평균

## 기타 부문별 어조지수(현재)와 관련 물가지수 상승률<sup>2)</sup>

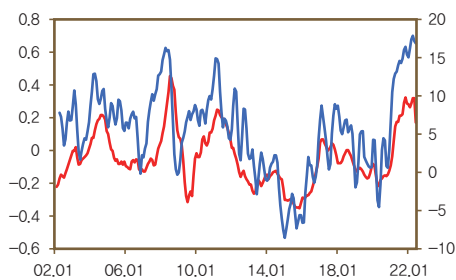
가정용품 및 가사 서비스



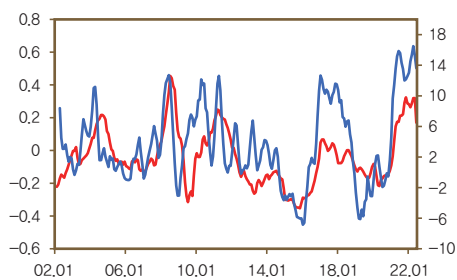
기타 상품 및 서비스



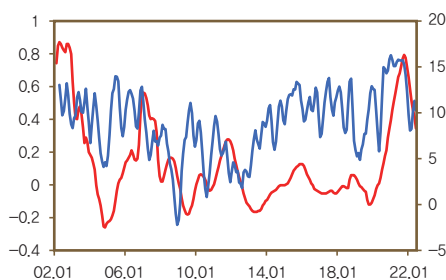
원자재



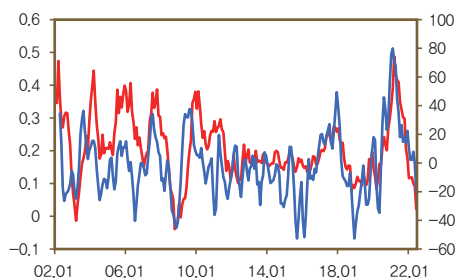
공업제품



부동산



금융상품



주: 2) 원자재와 공업제품은 생산자물가지수, 부동산은 부동산매매가격지수(KB), 금융상품은 KOSPI 지수의 전년동기대비 상승률과 해당 품목에 대한 어조지수와 비교

### 〈참고문헌〉

- 서범석, 이영환, 조형배(2022), “기계학습을 이용한 뉴스심리지수(NSI)의 작성과 활용,” 국민계정리뷰, 2022년 3월호
- 이궁희, 조주희, 조진경(2020), “새로운 우리나라 불확실성 지수의 작성,” 응용통계연구, 33(5), pp. 639-653
- 이현창, 최동규, 김용건, 허정(2022), "디지털 신기술을 이용한 실시간 당분기 경제전망 (GDP nowcasting) 시스템 개발," BOK 이슈노트 제 2022-7
- 장은아, 최희련, 이홍철(2020), “Stock prediction using combination of BERT sentiment Analysis and Macro economic index,” Journal of The Korea Society of Computer and Information, 25(5), pp. 47-56
- 최병재(2012), “일반인 기대인플레이션 지표의 주요 특징과 시사점,” 조사통계월보, 2012년 8월호
- Angelico, C., Marcucci, J., Miccoli, M., and Quarta, F. (2021), “Can we measure inflation expectation using Twitter?” Bank of Italy Temi di Discussione(Working Paper), No. 1318
- Araci, D. (2019), “FinBERT: Financial Sentiment Analysis with Pre-trained Language Models,” arXiv:1908.10063
- Atkeson, A., Ohanian, L. (2001), “Are Philips curves useful for forecasting inflation?” Federal Reserve bank of Minneapolis quarterly review, 25(1), pp. 2-11
- Baker, Scott, Nicholas Bloom, and Steven Davis (2016), “Measuring Economic Policy Uncertainty,” The Quarterly Journal of Economics, 131(4), pp. 1593-1636
- Blei, D., Ng, A., Jordan, M. (2003), “Latent Dirichlet Allocation,” Journal of Machine Learning Research, 3, pp. 993-1022
- Fraccaroli, N., Giovannini, A., Jamet, J. (2020), “Central banks in parliaments: a text analysis of the parliamentary hearings of the Bank of England, the European Central Bank and the Federal Reserve,” ECB Working Paper Series No. 2442
- Gentzkow, M., Kelly, B., and M., Taddy (2019), “Text as Data,” Journal of Economic Literature 57(3), pp. 535-574
- Hansen, Stephen, Michael McMahon, and Matthew Tong (2019), “The long-run information effect of central bank communication,” Journal of Monetary Economics, 108, pp. 185-202
- Hansen, Stephen and Michael McMahon (2016), “Shocking language: Understanding the macroeconomic effects of central bank communication,” Journal of International Economics, 99, pp. 114-133
- Haldane, Andrew and Michael McMahon (2018), “Central Bank Communications and the General Public,” AEA Papers and Proceedings, 108, pp. 578-583

- Jordà, Ò. (2005), “Estimation and Inference of Impulse Responses by Local Projections,” *American Economic Review*, 95(1), pp. 161-182
- Kalamara, E., Turrell, A., Redl, C., Kapetanios, G., and Kapadia S. (2020), “Making text count: economic forecasting using newspaper text,” Bank of England Staff Working Paper No. 865
- Larsen, V., Thorsrud, L., and Zhulanova, J. (2019), “News-driven inflation expectations and information rigidities,” *Norges Bank Working Paper*, No. 5/2019
- Lee, J., Youn, H., Stevens, N., Poon, J., and Han, S. (2021), “FedNLP: An interpretable NLP System to Decode Federal Reserve Communications,” *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 2560-2564
- Nakajima, J., Yamagata, H., Okuda, T., Katsuki, S., and Shinohara, T. (2021), “Extracting Firms’ Short-Term Inflation Expectations from the Economy Watchers Survey Using Text Analysis,” *Bank of Japan Working Paper Series*, No.21-E-12
- OECD (2007), “OECD Glossary of Statistical Terms,” p. 807
- Park, J. (2019), “DistilKoBERT: Distillation of KoBERT,” *GitHub repository* (<https://github.com/monologg/DistilKoBERT>)
- Park, K., Lee, Y., and Kim, S. (2019), “Deciphering Monetary Policy Board Minutes through Text Mining Approach: The Case of Korea,” *BOK Working Paper*, No. 2019-1
- Shapiro, A., Sudhof, M., and Wilson, D. (2017), “Measuring News Sentiment,” *Federal Reserve Bank of San Francisco Working Paper* 2017-01
- Shapiro, A. and Wilson, D. (2019) “Taking the Fed at its Word: A New Approach to Estimating Central Bank Objectives using Text Analysis,” *Federal Reserve Bank of San Francisco Working Paper* 2019-02
- Shiller, R. (2020), “Narrative economics: How Stories Go Viral and Drive Major Economic Events” *Princeton*, pp. 75-76
- Stock, J. and Watson, M. (2016), “Core Inflation and Trend Inflation,” *The Review of Economic and Statistics*, 98(4), pp. 770-784
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, L., and Polosukhin, I. (2017), “Attention is all you need,” *In Advances in Neural Information Processing Systems*, pp. 6000-6010

Copyright © BANK OF KOREA. All Rights Reserved

- 본 자료의 내용을 인용하실 때에는 반드시 “BOK 이슈노트 No.2022-38에서 인용”하였다고 표시하여 주시기 바랍니다.
- 자료 내용에 대하여 질문 또는 의견이 있는 분은 커뮤니케이션국 커뮤니케이션기획팀(02-759-4759)으로 연락하여 주시기 바랍니다.
- 본 자료는 한국은행 홈페이지(<http://www.bok.or.kr>)에서 무료로 다운로드 받으실 수 있습니다.