

텍스트 마이닝을 이용한 북한경제 연구

최지영 | 통일연구원 연구위원 | choijy@kinu.or.kr
이혜진 | 서울대 빅데이터시 연구소 연구원 | hyejinlee@bdai.snu.ac.kr

1. 머리말

북한의 국민경제를 분석하기 위한 자료들은 충분하지 않다. 북한 경제통계의 제한성과 신뢰성 문제는 오랜 기간 제기되어 왔고, 이를 보완하기 위한 노력들도 지속되어 왔다. 그간 이루어진 노력들은 크게 두 종류로 구분할 수 있다. 하나는 북한경제에 대한 기존 통계, 즉 GDP 추정치, 인구, 무역, 물가 및 환율 자료 등을 파악하고, 신뢰성을 검증하려는 노력들이다. 선행연구들을 통해 우리는 현재 가용한 북한통계의 한계와 의의를 어느 정도 파악하고 있다. 다른 하나는 새로운 통계와 기법을 발굴하려는 노력들이다. 전자가 국민경제를 설명하는 전통적인 통계와 추정치들을 파악하고자 하는 것이었다면, 후자는 신규 통계들을 발굴함으로써 북한경제를 보다 종합적이고 심층적으로 이해하려는 노력의 일환이라 할 수 있다.

텍스트 마이닝 기법은 북한경제 연구에 새롭게 적용되고 있는 방법 가운데 하나이다. 사실 북한연구의 텍스트 마이닝 기법에 활용되는 텍스트 데이터들은 공식문헌으로부터 추출된 것들로, 새로운 자료라고 볼 수는 없다. 기존 자료에 새로운 분석방법을 적용하는 쪽에 가깝다. 북한 공식문헌은 그동안 다양하게 연구자료로 사용되어 왔다. 그러나 기존의 공식문헌이 연구자의 문헌 독해와 정성 평가 위주로 활용되었다면, 텍스트 마이닝 분석은 텍스트 데이터가 보여주는 정량적 결과에 기초한다. 아직 북한경제 연구에 응용되는 텍스트 마이닝 분석은 초보적인 수준이다. 선행연구가 많다고 보기 어려우며, 다양한 종류의 공식문헌이 사용되었다고 보기 어렵다. 그럼에도 불구하고 텍스트 마이닝 분석은 기존 공식문헌 독해의 한계를

보완하고, 빅데이터를 활용한 다양한 결과를 보여주고 있다.

본고는 지난해에 필자들이 수행한 텍스트 마이닝 분석의 과정을 소개하는 것을 목적으로 한다. 지난해에 발간된 연구는 김정은 위원장 집권 이후 북한의 재정금융 제도 변화를 이해하는 데 중점을 두고 있다.¹⁾ 텍스트 마이닝 분석은 북한경제 연구자와 인공지능 연구자 간의 공동연구로 수행되었다. 이러한 학제 간 융합연구가 앞으로 북한경제 분야에 확대되기를 바라는 마음으로, 연구 과정을 자세하게 소개하고자 한다. 본고의 구성은 다음과 같다. 제2장은 북한경제 연구자의 입장에서 텍스트 마이닝 분석이 북한경제 연구에 갖는 의의와 『김정은 집권 이후 북한의 재정금융 제도 변화』 연구의 텍스트 마이닝 분석 과정을 소개한다. 제3장은 인공지능 연구자의 입장에서 북한경제 연구에 적용 가능한 텍스트 마이닝 분석기법, 분석 결과의 주요 특징을 소개한다. 제4장은 향후 연구과제이다.

II. 북한경제 연구와 텍스트 마이닝

1. 텍스트 마이닝 분석의 북한경제 연구 적용 의의

우리가 노동신문, 경제연구와 같은 북한문헌을 읽는 이유는 북한 당국의 정책방향을 파악하고 제도 변화를 이해하기 위해서이다. 당대회, 당전원회의, 최고인민회의가 개최되고 나면, 노동신문은 각각의 행사에서 논의된 내용을 보도한다. 우리는 결과 보도를 통해 북한 당국이 어떤 산업을 중점적으로 육성하고자 하는지, 코로나19 보건위기에 어떻게 대응하는지, 올해 예산은 어느 부문에서 몇 % 증가하는지를 파악할 수 있다. 그러나 북한문헌에 제시된 내용만으로 정책방향과 제도 변화를 이해하는 데는 한계가 있다. 각각의 행사에서 논의된 내용의 일부만이 보도되고, 구체적인 통계수치는 매우 제한적이기 때문이다. 특정 산업의 실적을 100% 넘게 수행했다는 보도가 나오더라도, 원래 계획 수준이 높았는지, 낮았는지 알 수 없기 때문에 성과를 정확하게 평가하기 어렵다. 특히 북한문헌은 사회주의 체제의 우월성과 최고지도자에 대한 지지를 표현하는 데 지면의 상당 부분을 할애하는 경향이 있다. 경제 현실에 대한 객관적이고 비판적인 평가를 찾기 어렵고, 정책방향에 대한 논쟁을 발견하는 것도 쉽지 않다.

1) 최지영 · 양문수 · 이혜진, 『김정은 집권 이후 북한의 재정금융 제도 변화』, 통일연구원, 2022.

따라서 북한문헌 독해를 통해 새로운 발견을 하기 위해서는 소위 ‘행간’을 읽어 내려는 노력이 필요하다. 양문수는 「북한문헌, 어떻게 읽을 것인가: 『경제연구』의 사례」라는 논문에서 다음의 세 가지 지침을 제시하기도 했다. 1) “...하는 일이 없도록 해야 한다.”와 같은 당위적 표현에서 북한의 현실을 유추할 것, 2) 제목과 본문 내용 간 모순성을 통해서 ‘필자’가 진정 말하고자 하는 부분을 파악할 것, 3) 용어들을 북한적 맥락에서 해석할 것.²⁾ 이러한 지침들을 활용하여 북한문헌을 독해하고, 경제정책에 대한 평가를 내리는 것은 북한경제 연구의 한 축일 수밖에 없다. 어쨌든 북한문헌은 경제정책에 대한 북한당국의 공식적 해설을 포함하고 있기 때문이다.

그러나 주의해야 할 것은 북한문헌에는 사회주의 체제의 우월성에 대한 선전이 상당 부분을 차지하고 있다는 것이다. 새로운 제도의 도입도 사회주의 체제의 우월성을 발현하기 위한 목적으로 포장되는 경우가 있다. 이러한 경우 전통적인 입장과 새로운 입장이 공존함으로써, 위의 지침에서 말하는 일종의 ‘모순성’이 발견되는 경우가 종종 있다. 『경제연구』 2016년 제1호에 실린 「사회주의기업체의 재정관리권」이라는 논문을 보자. 재정관리권은 사회주의기업책임관리제의 핵심으로, 기업소재정의 자율성을 부여한 권한이다. “사회주의 기업체들은 재정관리권을 가지고 확대재생산과 경영활동을 원만히 실현하는 데 필요한 화폐자금을 주동적으로, 창발적으로 마련하며 자체의 실정에 맞게 효과적으로 리용하기 위한 재정활동을 책임지고 조직진행하여야 한다.”와 같은 문장은 ‘주동적’, ‘창발적’, ‘효과적으로 리용’한다는 표현이 들어 있어, 생산단위의 자율성을 강조하는 것으로 보인다. 반면, “기업체의 재정관리권은 국가의 통일적이며 계획적인 지도 밑에 기업체가 주동적으로, 능동적으로 행사하는 권한이다.”라는 문장에는 ‘국가의 통일적이고 계획적인 지도’라는 표현이 포함되어 있다. ‘국가의 통일적이고 계획적인 지도’와 ‘기업소의 주동적, 창발적 자금 활용’ 가운데 우리는 어느 쪽에 중점을 둘 것인가? ‘모순성’을 발견하는 것만으로 충분하지 않을 수 있다. 이러한 상황에서 텍스트 마이닝 분석은 우리에게 또 하나의 기준을 제시해 준다. 공식문헌에 포함된 텍스트 데이터들의 ‘관계’가 ‘정량적’으로 나타나기 때문에, 이를 활용할 수 있다.

이와 같이 텍스트 마이닝 분석은 정성적 평가를 정량적 결과로 보완하는 역할을 한다. 여기서 강조하고 싶은 것은 북한문헌에 대한 독해와 정성적 평가 없이 텍스트 마이닝 분석만으로 우리가 얻는 것은 별로 없을 수 있다는 점이다.

2) 양문수, 「북한 문헌, 어떻게 읽을 것인가: 『경제연구』의 사례」, 『현대북한연구』, 12(2), 2009, pp.7~61.

2. 텍스트 마이닝 분석 적용 사례: 『김정은 집권 이후 북한의 재정금융 제도 변화』 연구

지난해 필자들이 수행한 연구는 김정은 위원장 집권 이후 북한의 재정금융 제도 변화에 대한 것이다. 북한은 김정은 위원장 집권 이후 ‘금융기관 재산제’를 도입했고, 도 단위 지역상업은행의 기능을 활성화하고 있는 것으로 보인다. 사회주의 체제에서는 원래 재정과 금융의 역할이 잘 구분되지 않는다. 중앙은행이 상업은행의 역할을 수행하며, 재정수입의 납부와 재정지출의 집행도 중앙은행을 통해 이루어진다. 이러한 구조는 ‘연성예산제약’을 낳으며, 효율적인 자금배분을 막는다. 이러한 계획 실패를 해소하기 위해, 현실의 사회주의 국가들은 금융기관의 자율성과 독립성을 확대하는 제도개혁을 단행함으로써, 재정과 금융의 역할을 분리하는 개혁을 시도한 바 있다. 북한의 제도 변화도 이러한 사회주의재정금융 개혁의 과정에 있는가? 그렇다면 북한경제에서 재정과 금융의 역할이 어느 정도 분리되고 있는가? 라는 질문을 던질 수 있다. 즉, ‘재정’이라는 단어와 ‘금융’이라는 단어의 관계는 어떻게 나타날 것인가? 지난해의 연구는 이렇게 질문을 던지고 가설을 수립하는 과정이 중요하다.

우선, 가설을 수립하기 이전에 가장 기초적인 단계는 하드카피 또는 PDF 파일 형태의 ‘경제연구’를 기계가 읽어 낼 수 있는 형태로 변환하는 것이다. 이 과정은 매우 노동집약적이며, 예상보다 실수가 많이 발생한다. 하드카피를 한글파일로 입력하거나, PDF를 한글파일로 변환하는 과정은 많은 시간이 소요될 뿐만 아니라 오타가 꽤 발생하기 때문이다. ‘경제연구’ 하드카피를 PDF로 변환할 때, 종이의 질이 좋지 않거나 인쇄체가 선명하지 않으면 원래 글자와 다른 글자로(예를 들어 ‘이’가 ‘어’로) 변환된다. 따라서 사람이 일일이 읽어 보며 오타를 수정하는 과정이 필요했다. 우리는 1991~2020년 『경제연구』에 수록된 재정, 금융, 화폐, 가격 관련 228편의 논문만을 텍스트 데이터로 사용했는데, 자료를 정리하는 데만 약 2~3주가 소요되었다.

다음 단계는 정리된 텍스트 데이터를 ‘전처리’하여 ‘명사’를 추출하는 것이다. 조사나 어미와 같이 큰 의미가 없는 불용어를 삭제하는 과정이다. 우리는 논문을 주제별로 분류(재정, 금융, 화폐, 가격)하고, 유의미한 분석기간을 그룹화하여 명사를 추출하였다. 그리고 빈도수가 높은 명사들을 사용하여 다음과 같이 텍스트 마이닝으로 답하고자 하는 핵심 질문과 이와 연관된 질문들을 제시하고, 이를 토대로 가설을 수립하였다. 아래의 질문들은 연구 초기 단계의 코워킹회에서 제시되었던 질문들인데, 굵게 표시된 부분이 ‘전처리’ 과정을 거쳐 추출된 명사들이다.

핵심 질문(상위가설)

- 재정과 금융의 연관성은 어떻게 변화하였는가?
- 금융에 대한 국가의 통제가 완화되는가?

핵심 질문과 관련된 질문들(하위 가설)

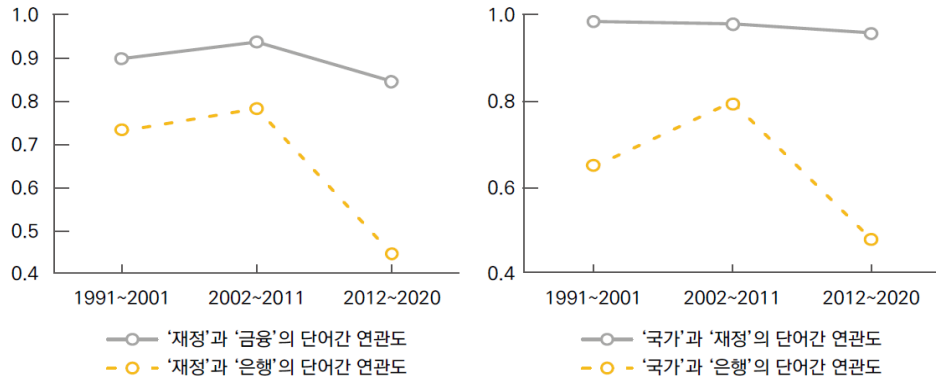
- 저금, 저축, 대부, 송금, 결제와 같은 은행의 기능은 어떻게 변화하는가?
- 계산, 축적, 지불과 같은 화폐의 기능은 어떻게 변화하는가?
- 가격은 어떻게 변화하는가? 가격은 제정되는가, 조절되는가?

다음 단계는 가설을 검증하기에 적합한 텍스트 마이닝 방법을 찾아내는 것이다. 본 연구에서는 통상 텍스트 마이닝 분석에서 기본적으로 적용하는 단순빈도수, 단어-역문서 빈도(TF-IDF), 네트워크 분석 이외에도 LASSO, DOC2VEC 분석을 적용하였다. 이에 대해서는 제3장에서 자세히 소개한다. 지금까지의 과정은 북한경제 연구자가 북한의 재정금융 제도 변화에 대한 가설을 수립하면, 인공지능 연구자가 이를 검증하기 적합한 텍스트 마이닝 기법을 제안·분석하여 정량적 결과를 제시하는 순서로 이루어졌다. 그러나 적합한 정량적 결과가 한 번에 도출되지는 않았다. 예를 들어 상위가설인 가설 1, “재정과 금융의 연관은 약화되었다.”에 대해서는 단어 간 연관도를 측정하는 DOC2VEC 방법을 적용하였는데, ‘재정’과 ‘금융’의 단어 간 연관도가 전 기간 높은 수준(0.8~0.9)으로 도출되었기 때문이다. 연구진은 ‘금융’이라는 단어의 빈도수가 낮고, ‘은행’의 빈도수가 높은 것에 착안하여, ‘금융’이라는 비교단어를 ‘은행’으로 바꾸어 보았다. 결과는 그림과 같이 2012~20년, 즉 김정은 위원장 집권 이후 ‘재정’과 ‘은행’의 단어 간 연관도가 크게 낮아지는 것으로 나타났다. 우리는 이 결과를 토대로 가설 1이 수용되는 것으로 해석하였다.

마찬가지로 가설 2, “국가의 은행(금융)에 대한 통제가 완화되고 있다”도 수용되었다. [그림 1]의 우측 그림을 보면, ‘국가’와 ‘재정’의 단어 간 연관도는 전 기간 높지만, ‘은행’의 연관도는 김정은 위원장 집권 이후 크게 낮아졌기 때문이다.

‘금융’을 ‘은행’이라는 단어로 바꾸는 것과 같은 유사한 조정이 여러 차례 있었다. 전체 연구는 여러 번의 시행착오를 통해 가설을 조정하고, 분석 방법이나 기준을 조정하면서 진행되었다. ‘금융’을 ‘은행’으로 바꾸는 것과 같이 조정이 성공한 경우도 있었지만, 그렇지 않은 경우도 있었다. 예를 들어 ‘현금’, ‘무현금’이라는 단어들의 관계 변화를 통해 현금유통의 기능 활성화 증거를 찾으려고 했지만, 이 시도는 성공하지 못했다.

[그림 1] 가설 1과 가설 2의 DOC2VEC 분석 결과



이와 같이 텍스트 마이닝 분석은 유의미한 정량적 결과를 찾기까지 분석기법과 대상 기간, 단어들을 반복적으로 조정하는 과정이다. 대상기간의 측면에서도 여러 시도가 이루어졌다. 2002년 7.1 조치, 2012년 우리식 경제관리방법, 상업은행법의 제정(2006년)과 개정(2015년) 전후, 대북제재 강화 전후 기간들을 다양하게 적용하였으나, 유의미한 분석 결과는 김정은 위원장 집권 전후로 주로 나타났다. 물론 분석기법과 가설에 따라 2002년 7.1 조치 전후, 대북제재 강화 전후로 차별성이 드러나는 경우도 있었다. 그러나 모든 가설에 적용되지는 않았다.

마지막 단계는 텍스트 마이닝으로 도출한 정량적 결과를 해석하는 과정이다. 우리가 텍스트 마이닝 기법을 북한경제 연구에 얼마나 유의미하게 활용하느냐는 이 마지막 단계에 있다고 생각한다. 중요한 것은 빅데이터가 보여주는 결과를 그 자체로 독해하는 것이 쉽지 않다는 것이다. 여러 분석기법을 통해 공통적으로 충족되는 가설들도 있었지만, 특정한 분석기법을 통해서만 설명되는 경우도 있었다. 텍스트 마이닝의 정량적 결과를 해석하기 위해서는 북한경제의 제도 변화에 대한 이해가 선결될 필요가 있다. 또한 정량적 결과를 해석하는 과정에서 오류나 편향이 발생할 가능성도 없지 않다. 이를 줄이기 위해서는 최대한 다양한 기법을 적용하여 분석을 시도하고, 각각의 기법들이 공통적으로 보여주는 경향성을 찾는 것이 중요하다.

III. 텍스트 마이닝 분석 방법과 결과의 특징

인공지능 연구의 관점에서 바라보는 텍스트 마이닝이란 주어진 문서를 기계상에서 자동으로 처리하고 분석하는 알고리즘을 구현하는 전 과정을 일컫는다. 이때 인간의 언어를 기계가 해석할 수 있는 형태로 변환하기 위해 자연어 처리(Natural Language Processing) 기법이 활용된다. 자연어 처리를 통한 텍스트 마이닝 분석은 크게 전처리, 변수 벡터화, 분석 알고리즘 수행의 세 단계로 이루어지는데, 본 연구에서는 특히 전처리에 집중하였다. 명확하고 유의미한 분석 결과를 도출하기 위해서는 그 기본 재료가 되는 텍스트 데이터의 효과적인 정제가 필수적이기 때문이다.

일반적으로 한국어의 전처리는 매우 어려운 작업으로 여겨진다. 이는 한국어의 언어적 특성에 기반한다. 자연어 처리 연구가 중점적으로 진행되어 온 영어와 달리 한국어에는 조사가 존재한다는 점이 가장 큰 허들이다. 영어의 경우 띄어쓰기를 기준으로 쉽게 단어를 파악할 수 있지만, 한국어는 추가적으로 단어와 결합한 조사나 어미를 인식하고 이를 제거하는 과정을 거쳐야만 비로소 정확하게 단어를 파악할 수 있다. 이외에도 한국어에서는 여러 개의 명사가 합쳐져 새로운 명사를 생성하거나 행동형 어미와 결합하여 동사의 형태를 띠는 경우가 많다. 이러한 성질은 명사 추출 작업의 난도를 증가시킨다. 특히 북한어의 경우에는 한자 단어들의 결합이 빈번하게 일어나기 때문에, 명사를 추출하는 과정에 있어 세심한 주의가 요구되었다. 본 연구에서는 통계적 스코어링 모형 기반의 파이썬 패키지 soynlp를 활용하여 명사 추출 정확도를 높이도록 노력하였다.

추출된 명사들을 분석 단위로 두고, 시기별 주요 키워드 트렌드 파악을 위해 단순 빈도수와 TF-IDF, LASSO 키워드 추출 기법을, 키워드 간 연관 분석을 위해 DOC2VEC과 네트워크 분석 기법을 도입하였다. 단순 빈도수 분석은 단어의 등장 횟수를 기준으로 각 단어들의 중요도를 판단한다. 특정 시기에 주로 등장한 단어들을 바탕으로 당시 주요하게 작용한 주제를 유추하는 방식이다. 그러나 단순히 빈도수가 높다고 해서 해당 단어가 반드시 유의미한 것은 아니다. 예를 들어 '김정은'이라는 단어는 모든 문서에서 빈번하게 등장하므로 단순빈도수는 높겠지만, 그 자체로는 크게 시사하는 바가 없을 것이다. 이렇게 무의미하게 반복되는 단어들을 최대한 배제하기 위해 TF-IDF를 사용하였다. 이 기법은 단어의 단순빈도수(Term Frequency: TF)를 해당 단어가 등장한 문서 수로 나누어(Inverse Document Frequency:

IDF) 좀 더 유의미한 키워드들이 추출되도록 유도한다.

한편, 더욱 차별성 있는 키워드를 도출하기 위해 기계학습 기법 중 하나인 LASSO 분류모델을 적용하는 방식을 고안하였다. 본 연구에서는 문서에서 발생한 단어들을 기준으로 분석의 대상이 되는 문서군을 기타 문서군으로부터 분류하도록 기계를 훈련시켰다. LASSO 분류모델의 구조적 특징으로 인해 기계는 문서 분류 학습 과정에서 분류에 주효하게 작용하는 단어들을 선택하게 된다. 이때 선택한 단어의 주효한 정도가 계수의 형태로 나타나는데, 본 연구에서는 이 계수를 단어가 가지는 중요도로 해석하였다.

단순빈도수, TF-IDF, LASSO 분석이 각 단어 자체가 가지는 주요도를 파악하는 방법론이라면, DOC2VEC과 네트워크 분석은 단어 쌍의 연관도를 도출하는 데 적합한 기법이다. DOC2VEC은 문서 내에 등장하는 단어들을 순차적으로 고려하여 문서와 문서, 문서와 단어, 단어와 단어 간의 문맥적 거리를 학습하는 인공지능망 모델이다. 이렇게 계산된 거리 수치를 활용하면, 사전에 선택된 중요 키워드들과 문맥적으로 유사하다고 판단되는 연관어들을 자동으로 추출할 수 있게 된다. 특히 본 연구에서는 가설을 이루는 단어 쌍의 시기별 연관도 추이를 관찰하여 그 해석을 가설 검증의 근거로 활용할 수 있었다.

마지막으로 네트워크 분석은 다수의 단어 사이의 관계를 동시에 파악하기 위해 사용되었다. 특히 시기별로 문서에서 발생하는 단어들을 각각의 노드(node)로, 동일 문서 발생 횟수를 엣지(edge)로 표현하여 단어 간 다중 관계를 시각화하였다. 이렇게 생성된 네트워크상에서 주요 키워드들이 어떤 단어들과 동시에 발생하였는지 그리고 동시출현의 강도는 어떠한지 효과적으로 관찰할 수 있었다.

텍스트 마이닝 기법은 방대한 양의 문서를 빠르게 처리하여 결괏값을 출력한다. 이때 기계는 주어진 텍스트에 사전에 정해진 기준을 편차 없이 적용하므로, 객관적이며 일관적인 분석을 수행한다는 장점이 있다. 그러나 기계는 단순히 주어진 수학적모형을 최적화할 뿐, 인간처럼 주어진 자료를 바탕으로 추론적 사고를 하지 않는다. 바꿔 말하면, 텍스트 마이닝 기법을 통해 도출된 결과는 기계가 푼 수학 문제의 답안지에 불과하다는 것이다. 이 답지의 해석과 활용은 연구자 본인의 연구 목적과 목표에 의거해야 할 것이다. 본 연구는 이러한 텍스트 마이닝의 성질을 고려하여 검증하고자 하는 가설의 객관적 근거로 삼는 방식을 채택하였다. 이를 통해 텍스트 마이닝의 장점을 활용하면서 동시에 유의미한 연구 결과를 도출할 수 있어 매우 생산적인 학제 간 협업이 가능했다.

IV. 향후 연구과제

텍스트 마이닝 분석을 북한경제 연구에 적용하려는 시도는 아직 초보적인 단계이다. 향후 관련 연구가 활성화되기 위한 과제를 제시하면서 글을 마무리하고자 한다. 첫째, 북한경제 관련 텍스트 데이터의 양을 획기적으로 늘릴 필요가 있다. 우리가 확보한 북한문헌의 텍스트 데이터는 사실 빅데이터라고 하기에는 초라하다. 그런데 기계가 읽어 낼 수 있는 수준으로 텍스트 데이터를 가공하는 것은 매우 노동집약적인 과정이기 때문에, 한 사람의 연구자가 수행하기 어려운 것도 사실이다. 북한문헌을 텍스트 마이닝 분석에 활용할 수 있는 수준으로 가공하여 공유하는 플랫폼이 있다면, 관련 연구 활성화에 기여할 수 있으리라 예상된다.

둘째, 지금까지 북한경제 연구에 활용된 텍스트 마이닝 분석은 북한 당국의 정책적 관심사를 파악하는 데 그치고 있다. 북한문헌의 내용을 정량적으로 해석하는 기존 접근법에서 벗어나, 성장, 무역, 물가, 환율 등 경제지표와의 연관성을 질문하는 방향으로 확장되는 것도 필요하다. 이는 텍스트 마이닝으로 분석한 북한의 경제정책 방향에 근거하여 경제성과를 평가하는 접근법이 될 수 있을 것으로 기대된다.

셋째, 지난해 필자들이 수행한 것과 같은 학제 간 융합연구가 활발해지길 기대한다. 인공지능을 활용하여 텍스트 데이터를 분석하는 연구는 하루가 다르게 새로운 분석기법이 제시되는 분야이다. 따라서 유의미한 분석 결과를 얻기 위해서는 최신의 다양한 분석기법을 활용할 필요가 있고, 이를 위해서는 인공지능 연구자와의 협업이 무엇보다 중요하다.

참고문헌

양문수, 「북한 문헌, 어떻게 읽을 것인가: 『경제연구』의 사례」, 『현대북한연구』, 12(2), 2009, pp.7~61.

최지영 · 양문수 · 이해진, 『김정은 집권 이후 북한의 재정금융 제도 변화』, 통일연구원, 2022.

V. Gupta & G. S. Lehal, “A Survey of Text Mining Techniques and Applications,”
Journal of Engineering Technologies in Web Intelligence, vol. 1, no. 1, 2009, pp.60~76.
GitHub (<https://github.com/lovit/soynlp>, 접속일: 2022. 6. 1).