

# 이머징 이슈 발굴을 위한 대안적 방법론 및 데이터 활용 연구

연구 기관 : 연세대학교

연구책임자 : 송 민

연세대학교 산학협력단



# 제 | 출 | 문

국회미래연구원 원장 귀하

본 보고서를 “이머징 이슈 발굴을 위한 대안적 방법론 및 데이터 활용 연구”의 최종보고서로 제출합니다.

2021년 9월 28일

연구기관명 : 연세대학교 산학협력단

연구책임자 : 송 민(교수)

연구 원 : 서수민(석사과정)

김여은(석사과정)

최연정(석사과정)

홍기봉(석사과정)

고영수(석사과정)

---

# 목 차

## 제1장 서론: 연구 배경 및 필요성 ..... 1

### 제1절 연구 배경 및 필요성

## 제1장 연구 질문과 연구 방법론 및 선행연구 리뷰 ..... 13

### 제1절 연구 질문과 연구 방법론

1. 연구질문: 기계학습 예측 모델을 사용한 이머징 이슈 도출 방법론 ... 13
2. Term Burstiness 개념을 활용한 Emerging Issue 예측 모델 ..... 13

### 제2절 선행연구 리뷰

1. 텍스트 데이터 기반 이머징 이슈 발굴 관련 연구 ..... 16
2. Term Burstiness 관련 연구 ..... 16

## 제3장 연구 과정 및 결과 ..... 19

### 제1절 연구 과정

1. 자료 수집 및 전처리 ..... 19
2. Term Burstiness 및 normalized log Term Frequency 계산 ..... 19
3. 이머징 키워드 예측 모델 학습 및 예측 결과 도출 ..... 19

### 제2절 연구 결과

1. 분야별 이머징 이슈 키워드 분석 ..... 21
2. 학제간 비교 ..... 34

---

**제4장 연구 결과의 시사점 및 결론 ..... 40**

제1절 연구 결과의 시사점

제2절 결론

## 표 목 차

[표 1] 모델에 사용한 변수 .....	20
[표 2] Computer Science 분야 키워드 .....	21
[표 3] Social Science 분야 키워드 .....	24
[표 4] Economics 분야 키워드 .....	27
[표 5] Arts and Humanities 분야 키워드 .....	32
[표 6] Burstiness 상승 여부에 대한 모델 성능 비교 .....	34

## 그림 목 차

[그림 1] MACD 관련 수식 .....	15
[그림 2] Term Burstiness 수식 .....	15
[그림 3] Computer Science와 Social Science 분야의 키워드 시계열 추이 비교	35
[그림 4] 각 분야별 Term Burstiness 바이올린 플롯 .....	36
[그림 5] Computer Science Feature Importance .....	37
[그림 6] Social Science Feature Importance .....	48

## 요 약

### I 장. 서론: 연구 배경 및 필요성

#### 1 연구 배경 및 필요성

사회적 불확실성 증가와, 대규모 텍스트데이터의 축적 및 텍스트마이닝 기술의 발전으로 자동화된 이머징 이슈 발굴 방법론의 필요성이 대두되었다는 연구 배경 및 필요성을 다룬다.

### II 장. 연구 질문과 연구 방법론 및 선행 연구 리뷰

#### 1 연구 질문과 연구 방법론

키워드의 현재 영향력 뿐 아니라, 미래의 예측 영향력까지 반영한 기계 학습 기반 이머징 이슈 발굴 방법에 대한 연구 질문을 제시한다.

방법론의 핵심이 되는 Term Burstiness의 개념에 대해 설명하고, 이를 활용한 이머징 이슈의 조작적 정의를 제시한다.

#### 2 선행연구 리뷰

텍스트 데이터 기반 이머징 이슈 발굴 방법과 관련된 선행 연구와, Term Burstiness와 관련된 선행 연구를 리뷰한다.



## III장. 연구 질문과 연구 방법론 및 선행 연구 리뷰

### 1 연구 과정

데이터 수집 및 전처리, Term Burstiness 및 normalized log Term Frequency 계산, 이머징 키워드 예측 모델 및 예측 결과 도출의 과정을 설명한다.

### 2 연구 결과

Computer Science, Social Science, Economics, Arts and Humanities 네 가지 학술 문헌의 이머징 키워드 분석 결과와, 학제간 비교 결과를 다룬다.

## IV장. 연구결과의 시사점 및 결론

### 1 연구결과의 시사점

학제에서 공통적으로 등장한 키워드와 학제간 차이를 중심으로 연구 결과의 시사점을 분석한다.

### 2 결론

Term Burstiness 기반 이머징 이슈 도출 방법론 제안에 대한 결론을 제시한다.

# 제 1 장

## 서론: 연구 배경 및 필요성

제1절 연구 배경 및 필요성

## 1 연구 배경 및 필요성

사회 변화 속도가 빠른 4차 산업혁명 시대에 접어들면서 미래에 대한 불확실성이 증가하였다. 이에 대응하여 미래를 더 잘 이해하고 대비할 필요성도 떠오르고 있다. 그렇기에 현재 명확하게 드러나지는 않지만 중장기적으로 사회에 큰 영향을 미칠 이슈인 이머징 이슈(Emerging Issue)를 발굴하는 것이 중요한 연구 분야로 인식되고 있다.

하지만 이머징 이슈는 약한 신호(weak signal)라는 특성 상 탐색에 어려움이 있다. 기존의 탐색 방법은 전문가가 개입하여 이머징 이슈를 발견하여 미래 전략을 수립하는 것이었다. (유순덕 외, 2014) 그러나 폭발적으로 늘어나는 정보 속에서 기존의 방식 대신 자동적으로 처리할 수 있는 방법이 필요하게 되었다. 폭발적으로 증가하는 비정형 데이터의 텍스트를 텍스트 마이닝 및 자연어 처리 기술을 사용해 데이터 간의 암묵적인 관계를 발견한다면 효율적으로 이머징 이슈 탐지가 가능해진다. 여전히 전문가의 판단도 필요하지만, 텍스트 마이닝 기술을 활용한다면 대량의 데이터 중 이머징 이슈 후보를 좁힐 수 있으며 더 나아가 전문가 편향을 완화할 수 있다.

## 제2장

### 연구 질문과 연구 방법론 및 선행연구 리뷰

---

제1절 연구 질문과 연구 방법론

제2절 선행연구 리뷰

## 1 연구 질문: 기계학습 예측 모델을 사용한 이머징 이슈 도출 방법론

본 연구에서는 기계학습 기반의 예측 모델을 사용하여 이머징 키워드를 도출한다. 현재 약한 신호를 드러내기 시작 해서 점점 그 신호가 커질 것으로 예상되는 ‘이머징 이슈’의 개념을 조작적으로 정의하고, 이를 반영할 수 있는 기계학습 모형을 설계하는 것이다. 현재 그 이슈의 영향력 뿐 아니라 미래의 영향력을 예측하고 반영하여 이머징 이슈를 도출한다는 점에서 기존의 연구들과 차별성을 지닌다.

이머징 이슈 예측 모델링을 위해서는 이슈의 영향력을 측정할 수 있는 지표가 필요하다. 이를 위해, 장기적인 추세에 비해 단기적으로 특정 단어가 얼마나 활발하게 문서에서 등장하고 있는지를 나타내는 Term Burstiness 개념을 사용한다.

또한 본 연구는 키워드를 기반으로 이머징 이슈를 도출한다는 특징을 가진다. 선제적으로 키워드를 유형화하기 보다는, 키워드 자체의 언급빈도에 집중하여 영향력을 측정한다는 점에서 예상하지 못한 새로운 주제의 출현을 발견하기 유연하다는 장점을 지닌다.

## 2 Term Burstiness 개념을 활용한 Emerging Issue 예측 모델

### 1. 연구 방법 개요

본 연구에서는 Term Burstiness 개념을 활용하여 이머징 이슈 키워드를 도출하는 방법론을 설명하고, Computer Science, Social Science, Economics,

Arts and Humanities의 학술 문헌에 해당 방법론을 적용한 분석 결과를 제시한다. Term Burstiness는 장기적인 추세에 비해 단기적으로 특정 단어가 얼마나 빈번하게 나타나는지를 측정하는 지표로서, 본 연구에서는 단어의 영향력을 측정하기 위해 사용하였다. 연구는 자료수집 및 전처리, Term Burstiness 및 normalized log Term Frequency 계산, 두 가지 지표를 사용한 이머징 키워드 예측 모델 학습 및 예측 결과 도출 순으로 진행하였다.

## 2. Term Burstiness

Term Burstiness는 장기적인 추세에 비해 단기적으로 특정 단어가 얼마나 자주 언급되는지를 측정하는 지표로서, 연구에 따라 다양한 수식을 사용한다. 본 연구에서는 He and Parker(2010), E.Tattershall(2020)에서 제시한 수식을 활용하였다. 해당 수식은 주가 예측에서 자주 활용되는 MACD(Moving Average Convergence Divergence) 개념을 응용하여 Term Burstiness를 정의한다. MACD는 Long term Moving Average와 Short term Moving Average의 차이를 가지고 장기이동평균에 비해 단기이동평균이 높게 나타나는 경우 측정 변수가 bursty하다고 판단한다. 이를 term에 적용하는 경우, 장기적인 등장빈도 이동평균에 비해 단기적인 등장빈도 이동평균이 높을수록 해당 키워드의 term burstiness가 높다고 본다. 보다 구체적으로는, Long term MA와 Short term MA의 차이인 MACD선과 MACD선에 이동평균을 취한 signal 선의 차이(=histogram)를 활용할 경우 장기적인 추세와 단기적인 추세가 교차하는 지점을 빠르게 예측할 수 있어서 이를 통해 수식을 정의한다. 따라서 본 연구에서 사용한 Term Burstiness 수식은 아래와 같다. <그림1>은 MACD 관련 수식이다.  $n_1$ 은 Long term MA의 기준 기간( $n_1$ 년),  $n_2$ 는 short term MA의 기준 기간( $n_2$ 년),  $n_3$ 는 signal 선의 MA 기준 기간( $n_3$ 년)을 의미한다. EMA는 exponential moving average의 약자이다. <그림2>는 MACD를 통해 도출한 signal을 사용해서 Term Burstiness를 정의한 수식이다.  $w$ ,  $t$ 는 각각 단어와 특정 시점을 의미하는 것으로  $t$ 시점의 단어  $w$ 에 대한 Term Burstiness를 나타낸다.  $p(w, t)$ 는  $t$ 시점  $w$ 의 빈도수를 의미하고, 식의 분모는 normalize를 위해 추가되었다.

$$\begin{aligned} \text{MACD}[n_1, n_2] &= \text{EMA}[n_1] - \text{EMA}[n_2] \\ \text{Signal}[n_1, n_2, n_3] &= \text{EMA}[n_3](\text{MACD}[n_1, n_2]) \\ \text{Histogram}[n_1, n_2, n_3] &= \text{MACD}[n_1, n_2] - \text{Signal}[n_1, n_2, n_3] \end{aligned}$$

[그림 1] MACD 관련 수식, E.Tattershall(2020)

$$\text{Burstiness}[n_1, n_2, n_3](p(w, t)) = \frac{\text{histogram}[n_1, n_2, n_3](p(w, t))}{\sqrt{\max(p(w, t))}}$$

[그림 2] Term Burstiness 수식, E.Tattershall(2020)

### 3. 이머징 이슈의 조작적 정의

본 연구에서는 Term Burstiness와 normalized log Term Frequency를 활용하여 이머징 키워드를 정의한다. 이머징 키워드는 현재 과거에 비해 영향력이 커진 동시에 미래에도 그 영향력이 증가될 것으로 예측되는 키워드라고 할 수 있다. 따라서 (1) 현재 Term Burstiness가 0보다 커서 과거 대비 영향력이 커진 단어 중, (2) 미래에 Term Burstiness가 증가하며 (3) normalized log Term Frequency가 커질 것으로 예측되는 단어를 이머징 키워드로 정의한다. (3)번 조건의 normalized log Term Frequency의 경우 전체 단어 언급 빈도 중 해당 단어 언급 빈도의 비율에 log를 취한 것이다. Term Burstiness의 경우 과거의 추세 대비 상승 정도를 의미하는 상대적인 지표이므로, 실제로 절대적인 언급 빈도 역시 상승하는지를 측정하기 위해 해당 조건을 추가하였다.

## 1 텍스트 데이터 기반 이머징 이슈 발굴 관련 연구

대규모 텍스트 데이터의 축적과, 텍스트마이닝 기술의 발전을 바탕으로 텍스트 데이터를 사용한 이머징 이슈 발굴 방법론에 대한 연구가 이루어지고 있다. 대표적으로 이머징 이슈 인덱스 생성과 관련된 연구와, 토픽모델링을 활용한 연구들이 있다.

먼저, 이머징 이슈를 정의할 수 있는 지표 개발 관련 연구로, Hiltunen(2008)은 weak signal이 strong signal로 이동하는 경로를 설명하는 3가지 축을 제시한다. signal은 시그널의 수 또는 가시성, issue는 뉴스 기사 등 사건의 수, interpretation은 정보 수용자들이 인식하는 주관적 중요도를 의미한다. Yoon(2012)는 이 개념을 활용하여, 객관적 지표인 signal과 issue를 측정할 수 있는 지표를 생성하고 weak signal을 탐지하는 모형을 제안한다. Yi-Ning Tu(2012)는 학술문헌에 NI(novelty index)와 PVI(published volume index)를 적용하여 weak signal을 판단하는 연구를 진행하였다.

토픽모델링 관련 연구는, 시간에 따른 문서 집합에 LDA와 같은 토픽 모델링을 적용하여 시계열적 토픽의 변화를 살펴본다. 임치현(2018)은 '4차 산업혁명'과 관련된 문서 집합에 LDA를 사용하여 중요 주제를 추출하였다. 김현정(2015)은 항공산업 관련 문서의 단어를 co-occurrence 기반으로 군집화하여 토픽을 추출하고, 시계열적으로 중요한 주제가 어떻게 변화하는지 살펴본다. Sun(2017)은 교통 분야에 시계열 토픽 모델링을 적용하여 hot topic과 cold topic을 구분하는 연구를 진행하였다.

## 2 Term Burstiness 관련 연구

시계열적인 단어 빈도 기반의 Term Burstiness는 장기적인 추세에 비해 단기적으로 언급되는 키워드나 토픽을 추출하기 위한 관련 연구들이 이루어지고 있으며, 연구 별로 상이한 Term Burstiness 계산식을 제시하기도 한다.

Kleinberg(2002)는 단어의 언급 빈도는 폭발적인 시기(bursty)와 그렇지 않은 시기(non-bursty)의 상태를 반복한다고 본다. 이러한 가정을 바탕으로 Term Burstiness를



통계적인 방식으로 정의하고, email 데이터를 통해 단어들의 시기별 burstiness를 측정한다. 이후 문서의 생성 주기가 실시간성을 가지는 온라인 뉴스나, 소셜네트워크서비스(SNS)에 Term Burstiness 개념을 적용하여 현재 가장 활발하게 언급되고 있는 hot-topic을 추출하는 연구들(Takahashi (2012), Kou (2002))이 진행되었다.

He and Parker(2010)는 문서의 생성 주기에 따라 Term Burstiness 계산식이 상이해야한다고 본다. 이에 뉴스나 SNS 데이터에 비해 불연속적으로 생산되는 학술문헌에 적용하기 위해 MACD(Moving Average Convergence Divergence)에 근거한 Term Burstiness 개념을 제시한다. 더 나아가 E.Tattershall(2019)는 MACD 기반 Term Burstiness를 컴퓨터공학 분야 학술 문헌에 적용하여, 미래의 bursty term을 예측하고 이를 기반으로 키워드를 클러스터링하는 연구를 제시하였다.

## 제3장

### 연구 과정 및 결과

---

제1절 연구 과정

제2절 연구 결과

## 1 자료 수집 및 전처리

학술정보 데이터베이스 SCOPUS API를 사용하여, Computer Science, Social Science, Economics, Arts and Humanities 4가지 분야의 1991년부터 2020년까지 총 30년의 영문 학술 문헌 제목과 초록을 수집하였다. 분야별로 평균적으로 약 1,500,000 건의 문서를 수집하였다. 데이터 전처리 과정에서 제목과 초록을 합친 것을 하나의 문서로 보고, 불용어를 제거 하고, unigram, bigram, trigram을 사용하고, 전체 5회 미만 등장한 단어를 제거하였다.

## 2 Term Burstiness 및 normalized log Term Frequency 계산

Term Burstiness와 normalized log Term Frequency의 계산을 위해 전처리된 키워드의 연도별 언급빈도를 구하고, 이를 전체 단어의 언급빈도 합으로 나눈다. 연도별로 문헌의 건수가 다르기 때문에 필요한 normalization 과정이다. 이후 [그림2]에서 소개된 식을 적용하여 각 단어의 연도별 Term Burstiness 및 normalized log Term Frequency를 구할 수 있다.

## 3 이머징 키워드 예측 모델 학습 및 예측 결과 도출

단어의 언급빈도와 관련된 시계열적 특성을 독립변수로 하여, 미래에 Term Burstiness 및 normalized log Term Frequency가 증가할 단어를 예측하는 모델을 학습한다. 이 때, 기계학습 모델로는 Random Forest, LightGBM, XGBoost를 비교 실험하였고, (n1, n2, n3)으로는 선행연구를 참고하여 (6, 12, 3), (7, 9, 5), (4, 8, 5), (3, 6, 2), 그리고 몇 년 뒤의 상승 여부를 예측할

것인지와 관련하여 2, 3, 4, 5, 6년에 대한 실험을 진행하였다. 4가지 분야 모두 LightGBM, (7, 9, 5), 5년 후에 대한 예측 성능이 가장 높게 나와 해당 모델을 최종 선택하여 분석을 진행하였다. 모델에 사용된 변수는 다음과 같다.

[표 1] 모델에 사용한 변수

단어	log_tf
long_ma_0 ~ long_ma_10	현재 시점 t부터 t-10까지의 9년 기준 장기 이동 평균.  예) 2020년 데이터셋의 경우, long_ma_0은 2020년의 장기 이동 평균, long_ma_10은 2010년의 장기 이동 평균
hist_0 ~ hist_10	현재 시점 t부터 t-10까지의 히스토그램. 히스토그램은 MACD와 signal의 차이를 의미.  예) 2020년 데이터셋의 경우 hist_0은 2010년의 히스토그램, hist_10은 2020년의 히스토그램
signal_0 ~ signal_10	현재 시점 t부터 t-10까지의 signal. (7, 9, 5) 하에서 signal은 MACD의 5년 기준 이동 평균을 의미.  예) 2020년 데이터셋의 경우 signal_0은 2010년의 signal 이고 hist_10은 2020년의 signal
significance	최근 3년 기준 TF의 이동 평균 값  예) 2020년 데이터셋의 경우, 2020년과 2019년, 2018년의 TF를 평균낸 값
prevalence	현재 시점의 normalized TF
scaled_std	3년 기준 이동 평균의 표준 편차를 scaling으로 나눈 값
max	3년 기준 이동 평균의 최댓값
min	3년 기준 이동 평균의 최솟값
scaling	3년 기준 이동 평균의 최댓값에 루트를 취한 값

## 1 분야별 이머징 이슈 키워드 분석

## 1. Computer Science

## 가) 키워드

[표 2] Computer Science 분야 키워드

세부 분야	키워드
인공지능	humanrobot interaction / neutrosophic / extreme learning machine / recurrent neural network / deep belief network / attention / autoencoders / adversarial attack / fine tuning / siamese network / dqn
네트워크	massive mimo system / interferometric / network function virtualization / 5g network / multi access edge computing
분산시스템	distributed ledger / blockchain network / cryptocurrency
사물인터넷	iot data / cyber physical system / robotics / unmanned aerial vehicle / iiot / industrial internet
기타	additive manufacturing / energy efficiency / energy harvesting / chemometrics / aerospace

## 나) 키워드 설명

## i. neutrosophic

- (예시 1) Abdel-Basset (2020)은 professional selection 분야에 서, 채용과정에 존재하는 불확실성 및 모호함을 반영하기 위해 hybrid neutrosophic analytic network process 프레임 워크를 제시하였다.

- (예시 2) Abdel-Basset and Ali (2020)는 프로젝트에 자원을 할당 하는 과정에서 발생할 수 있는 불확실성을 반영하는 fuzzy scheduling 및 time-cost tradeoff 모델을 제안한다.

- 뉴트로소픽 (neutrosophic) 이론은 참과 거짓으로 분류될 수 없고

중간의 모호한 값을 가지는 대상에 대한 연구이다. 실제 상황에서 발생할 수 있는 불확실성을 모델링 과정에 반영하기 위한 뉴트로소픽 이론 적용 연구들이 등장하고 있다.

## ii. energy harvesting

- (예시 1) Pan (2020)는 무선통신 기술에서의 에너지 하베스팅 성능을 높일 수 있는 intelligence reflecting surface (IRS)에 대한 연구를 다룬다.

- (예시 2) Khan (2020)은 6G 무선통신 시대의 비전을 다루면서 intelligence energy harvesting에 대한 연구 필요성을 제시한다.

- 에너지 하베스팅 (energy harvesting)은 버려지는 에너지를 수확하여 재활용하는 기술을 의미한다. 자원 절약과 환경 문제에 대한 중요성이 커지면서 에너지 하베스팅에 대한 연구도 활발해지는 추세이다. 컴퓨터 공학 분야에서는 IRS 등과 같은 에너지 절약 관련 연구들이 이루어지고 있다.

## iii. cyber physical system

- (예시 1) Panicucci (2020)는 사이버 물리 시스템 하에서 산업 장비들의 잔여수명 (remaining useful life, RUL)을 예측하는 cloud-to-edge 기반 방법론을 제시한다. 이를 통해 사전 예측을 통해 장비를 유지보수하여 사이버 물리 시스템을 최적화할 수 있다고 제안한다.

- (예시 2) Ionescu (2020)은 사이버 물리 시스템과 같이 다양한 특징을 가진 상이한 요소로 구성된 체계에서는 다목적 최적화 절차의 효율적인 구현이 필요하다고 지적한다. 이를 위해 Extended Prediction self-adaptive Control(EPSAC) 방법론을 제시한다.

- 사이버 물리 시스템 (cyber physical system)은 다수의 임베디드 시스템과 통신망 그리고 현실세계인 물리 시스템이 서로 밀접하게 연계되어서 사용자가 필요로 하는 서비스를 제공한다. (박정민, 2013) IoT(Internet of Things) 기술의 발달에 따라 사이버 물리 시스템에 대한 연구도 증가하는 추세를 보인다.

#### iv. humanrobot interaction

- (예시 1) Barchard (2020)은 로봇의 사회적 지능을 강조한다. 이에 인간과 로봇의 특징을 융합하여 믿음 (belief), 욕망 (desire) 및 심리학적 행동, 인지, 감정 개념을 포함하는 20개의 Perceived Social Intelligence (PSI) 지표를 제시한다. 이를 통해 로봇의 인식을 평가하는 실험을 진행하였다.

- (예시 2) Iyer (2020)은 걷기 재활을 위한 로봇에 있어서, 로봇과 인간 사이의 상호작용이 중요하다는 것에 초점을 맞춘다. 로봇의 상이한 보행 보조 방식에 따라 인간이 어떻게 반응하는지에 대한 실험을 진행하였다.

- 코로나-19 국면 속에서 언택트 (un-tact)가 사회적 트렌드로 자리 잡고, 헬스케어 등의 분야에서 로봇의 활용이 활발해지면서 로봇과 사람 사이의 상호작용에 대한 연구의 중요성도 커지고 있다. 로봇의 사람의 사회적, 물리적 상호작용에 대한 연구가 이루어지고 있다.

#### v. extreme learning machine

- (예시 1) Wei (2020)은 kernel extreme learning machine (KELM)을 사용하여 학생들의 기업가적 정신을 예측할 수 있는 인공지능 모델을 제안하였다. 다른 모델에 비해 harris hwak's optimizer를 사용한 KELM이 우수한 성능을 보이는 것으로 나타났다.

- (예시 2) Mohanty (2020)은 kernel extreme learning machine (KELM)을 사용한 유방암 조기 진단 모델을 제시하였다. 유방 x-ray 사진을 바탕으로 정상/비정상, 악성/양성을 판단한다.

- extreme learning machine은 기존 인공신경망에서 사용하는 gradient descent (경사하강법)을 사용하지 않음으로써, 학습 속도와 모델의 일반화 성능을 높였다고 알려진 알고리즘이다. (안태천, 2015) 딥러닝의 발전과 함께, extreme learning machine을 비롯하여 기존 방식의 한계를 개선할 수 있는 연구들이 진행되고 있다.

## 2. Social Sciences

### 가) 키워드

[표 3] Social Science 분야 키워드

세부 분야	키워드
환경 문제	environmental impact / unsustainable / renewable energy / energy consumption / carbon / wastewater / emission / nitrogen / precipitation / spatiotemporal / bioactive / electric vehicle / climate change / ecosystem service
통계방법론	significant / correlation / modeling / statistical / structural equation modeling / estimation / dataset / sampling / 0.05
인공지능	deep learning / learning algorithm / parameter / neural network cnn / convolutional neural network / random forest / machine learning / data driven / optimization / database / LSTM
일반과학	digital technology / blockchain technology / vr(virtual reality) / smart city / hotspot / fourth industrial revolution / whatsapp / industry 4.0 / Instagram / thing iot
일반사회 /트렌드	mediating role / trump / misinformation / fake / china / 2030 agenda / terrorist attack / health crisis

### 나) 키워드 설명

#### i. Ecosystem service

- (예시 1) Brown & Fagerholm (2015)는 Ecosystem service를 위한 대중 참여 GIS (Public Participation GIS) 연구를 진행했다. 대중 참여 GIS는 다양한 ecosystem services의 특질을 파악하기 위해 지난 10년 동안 꾸준히 진화했다. 대중 참여 GIS 연구를 위해서, 매핑된 ecosystem services의 유형, 공간에 대한 매핑 방법론, 참여자들에 대한 샘플링 방식 등이 연구되어왔다.

- (예시 2) McPhearson et al. (2015)는 도시 생태계 서비스 (Urban Ecosystem services)가 보다 지속 가능한 도시로의 전환을 모색하는 계획, 관리 및 거버넌스 관행을 연결하고 도시 시스템의 회복력을 구축하는 데 중요한 역할을 할 것을 제안했다. 특히 지구 환경 변화에 지속적으로 대응하는 도시 시스템의 동적 특성을 고려할 때, 살기 좋고 지속 가능한 도시를 보장하기 위해 도시가 생태계 서비스의 탄력적인 공급 보호를 우선시해야 한다고 주장했다.



- 자연 생태계에 대한 인간의 의존성의 특징을 살펴보고, 이를 강조하기 위해 Ecosystem service라는 개념이 환경, 자원경제 관련 사회과학 연구에 점차 널리 활용되고 있다.

## ii. health crisis

- (예시 1) Beery (2020)는 COVID-19가 교육 분야에 있어 광범위한 영향력을 끼쳤는데, 특히 많은 수의 교육 기관들이 온라인 형태의 수업을 제공하는 데 기여했다고 밝힌다. 해당 연구에서는 환경/야외 교육과 관련해서 COVID-19가 야기한 영향에 대해, 참여적 위기 관리를 검토한다.

- (예시 2) Lim H.Y. (2020)는 철학자 Alasdair MacIntyre가 제시한 철학을 통해, 공중 보건을 연구하고 더 나아가 liberal individualism과 균형을 맞추기 위해 윤리 교육을 강조하는 연구를 진행했다. 이의 목적은 현대 사회에서 또 다른 유행성, 범대중적 health crisis를 예방하기 위함이다.

- COVID-19 확산으로 인해 점화된 이슈이다. 현재 전 세계적으로 퍼져서 쉽게 해결되지 않는 COVID-19의 사회적 영향력을 이해하고, 유사한 전염병에 의해 촉발될 수 있는 사회적 위협을 사전에 차단하기 위한 사회과학 연구가 활발히 진행되고 있다.

## iii. misinformation

- (예시 1) Morales-i-Gras, J. (2020)는 SNS 데이터를 활용한 선거 결과 예측 연구가 이루어진 이래로, 관련 연구자들은 misinformation으로 연구 주제를 좁혀나가고 있음을 밝힌다. 여기서의 misinformation이란 단순히 의도적인 가짜 뉴스만을 의미하지는 않고, 특정 집단에서 발생하는 선택적인 정보 수용까지 포함한다. 이를 위해 연구자는 2019년 스페인 general election 캠페인 기간 중 트위터에서 발생한 사용자들의 debate를 연구하였다.

- (예시 2) Greenspan & Loftus (2020)는 misinformation feedback에 대한 실험 연구를 진행했다. 실험에 따르면 misinformation feedback은 상당한 신뢰 변화를 일으켰다. 초기 식별 단계에서, 실험 이후 보고한 것보다 더 자신감이 있었다는 잘못된 피드백을 받은 참가자들은 실험 초

반에 보였던 자신감보다 더 큰 자신감을 갖고 있었다고 후에 밝혔다.

- 최근 몇 년간 다양한 매체를 통한 정보공급이 기하급수적으로 증가했고, 이에 따라 잘못된 정보의 확산에 대한 관심이 높아지고 있다. 이를 연구하기 위해 개인 수준에서의 잘못된 정보 수용과 더불어, 집단/사회 수준에서 잘못된 정보의 확산에 기여하는 요소를 연구하는 사례가 증가하고 있다.

#### iv. LSTM (Long Short-Term Memory)

- (예시 1) Davis et al. (2020)의 연구에서 제안하는 EVT(Extreme Value Theory)-LSTM 모델은 운송 네트워크에서의 시간 데이터를 대상으로, 딥러닝 기반 이상 탐지 모델을 구현했다. 연구자들은 해당 모델을 기존의 전통적 통계 기반 모델링, 기계 학습 모델링 그리고 다른 딥러닝 기반 모델들과 성능을 비교했다.

- (예시 2) Lu et al. (2021)은 다변량 분석을 기반으로 한 LSTM 네트워크를 통해 시간의 흐름에 따른 교통 상황 변화량 예측 연구를 진행했다. 예측 시점을 기준으로 이전 교통 상황, 날씨, 강우량 등을 변수로 활용했다. 해당 연구의 결과는 LSTM 네트워크를 기반으로 한 일변량 (univariate) 분석, CNN 기반의 다변량 분석의 결과와 그 성능을 비교했다.

- 사회과학 연구주제에서 다양한 예측 모델을 설계할 때, 인공 신경망 기반의 예측 모델을 최근에 많이 활용하고 있다. 그 중에서도 시계열적 특성을 담은 데이터를 기반으로 예측 모델링을 진행할 때 Long Short-Term Memory 알고리즘을 활용한다.

#### v. virtual reality

- (예시 1) Ha et al. (2020)은 Oculus 모범 사례 가이드(OBPG)에 초점을 맞춰 가상현실 디자인의 관행이 표준화되는 방식을 분석한다. 저자는 OBPG가 가상현실(VR)에 대한 사용자 몰입을 강조하고 유지하는 표준 및 관행을 만드는 역할을 하여, 기술의 약점을 완화하고 마케팅과 픽션에 대한 가상현실의 이상적 표현에 더 잘 순응하는 목적을 달성한다고 주장한다.

- (예시 2) LaRocco (2020)에 따르면, 가상현실을 위한 몰입형 헤드

마운트 디스플레이(HMD)의 이미지는 시각적 적응을 위한 유일한 소스를 제공한다. 따라서 HMD를 보는 동안 휘도가 거의 순간적으로 크게 증가하면 시각적인 불편함이 발생할 수 있다. 따라서 해당 연구에서는 이러한 불편함을 유발하기 위해 필요한 휘도 변화를 실험하고 그 결과를 밝혔다.

- 가상현실이 사회과학 연구에 있어 점차 널리 활용되고 있다. 가상현실은 그 자체로서 연구의 대상이 될 뿐만 아니라, 기존에는 진행하기 어려웠던 실험을 가능케하는 ‘도구’로서 사회과학 연구자들에게 색다른 의미를 지닌다. 즉 가상현실은 사회적 변화를 야기하는 현상으로서 연구될 뿐만 아니라, 새로운 차원의 실험 연구를 진행하는 데 도움을 주는 기술로서도 주목받고 있다.

### 3. Economics

#### 가) 키워드

[표 4] Economics 분야 키워드

세부 분야	키워드
경제 용어	pegged exchange rate / embedded option / import substitution / domestic financial market / poverty reduction / market result / sanction / global value chain / central bank / public debt / startup
국가	kazakhstan / eu / islamic / eur asian / indonesia stock / asian / croatia / eur asian economic / brazil / russian oil / india / trump / east asian country / colombia / brics
규제/법규	legal regulation / law enforcement / macroprudential
통계 기법	spss / science spss / social science spss / amos / random forest / chi square / linear regression / regression / convenience sampling / structural equation modeling / equation modeling
사회 용어	educational institution / company business / nobel laureate / education system / populism / school principle / higher education / residential property / intellectual property / higher education institution / inclusive

#### 나) 키워드 설명

##### i. poverty reduction

- (예시 1) Berdegue et al. (2015)는 칠레, 콜롬비아, 멕시코의 농촌

지역의 증가하는 도시화의 영향을 조사하고 농촌-도시 영역 내에 중소 도시의 존재가 경제 성장을 촉진하고 빈곤과 소득 불평등을 감소시키는지 여부를 조사한다. 위 연구 결과에 따르면 칠레와 콜롬비아의 경우 이러한 도시 중심지가 깊은 농촌 지역에 비해 경제 성장률과 농촌 지역의 빈곤 감소에 더 큰 기여를 할 수 있다는 것을 암시한다.

- (예시 2) Sehrawat et al. (2018)은 Autoregressive Distributed Lag (ARDL) 한계 시험 절차를 사용하여 1970년부터 2015년까지 인도의 재정 개발, 경제 성장과 소득 불평등이 빈곤에 미치는 영향을 조사한다. 이들의 연구 결과는 금융 개발, 경제 성장, 불평등과 빈곤 사이의 강력한 장기적 관계를 보여준다. 결과는 소득 불평등과 인플레이션이 빈곤을 악화시키는 반면, 재정 발전과 경제 성장은 인도의 빈곤 감소에 도움이 된다는 것을 보여준다.
- (예시 3) Wu et al. (2020)은 해당 논문에서 기업가 정신 (entrepreneurship)이 빈곤 완화의 중요한 수단이며 기업가적 기회 (entrepreneurial opportunity)가 기업가 활동 (entrepreneurial activity)의 중심에 있음을 시사한다. 그러나 현존하는 다른 연구들은 기업가 활동과 빈곤 감소에서 기업가적 기회의 역할에 거의 관심을 기울이지 않았다. 이 논문은 기업가적 기회 유형과 농민 기업가의 기업가적 성과 사이의 관계를 탐구한다. 그리고 기업가적 기회의 역할이 개선된 기업가적 성과를 달성할 때 더 복잡하고 미묘한 차이가 있다는 것을 보여줌으로써 농민 기업가 정신과 빈곤 완화에 대한 연구에 기여한다.
- 국가의 교외 지역의 도시화나 재정적 · 경제적 발전이나 성장, 기업 활동의 기회 등이 빈곤 감소 (poverty reduction)에 어떻게 영향을 미치는지에 대한 맥락에서 해당 키워드가 사용되었다.

ii. global value chain

- (예시 1) Fernández (2015)에 따르면 Global Value Chain (GVC) 접근법은 글로벌 경제 네트워크의 기업의 현재 전략 분석 뿐만 아니라 초국가적 기관이 추진하는 경제 개발 정책에도 점점 더 유의미한 도구가 되었다. 저자는 논문에서 이러한 접근법의 초

국가적 제도화가 개발도상국과 그 행위자들의 전략을 가능하게 하는 도구를 구성하기보다는 자본의 초국가적 부분에 의해 지배되는 네트워크에 대한 부차적이고 배타적인 통합의 패턴을 합법화하는 데 기여해왔다고 주장한다. 또한, 저자는 이는 GVC가 세계 정치적 네트워크에 의해 구현된 전략을 위한 신자유주의 장치로 변환하는 것뿐만 아니라 세계 정치적 네트워크가 이론적 접근 방식을 통합할 때 GVC의 이론적 및 방법론적 말뭉치의 한계의 비(非)비판적인 동화를 통해서도 가능하다고 언급한다.

- (예시 2) Imbruno (2019)는 논문에서 기업들이 외국 상품을 구입할 때 대규모 매물 비용에 직면할 수 있다는 점을 고려하여 Trade Policy Uncertainty (TPU)의 감소에 대한 수입의 조정을 경험적으로 탐구한다. 논문의 주요 결과는 TPU의 감소가 더 높은 품질과 관련된 더 다양한 외국 상품에 대한 접근을 허용한다는 것을 암시한다. 마지막으로, 저자는 해당 논문을 통해 서로 다른 소유권을 가진 기업들뿐만 아니라 서로 다른 최종 용도를 가진 제품들 간에 다차원적 TPU 효과를 문서화하여 글로벌 가치 사슬(global chain value)의 맥락에 대한 흥미로운 통찰력을 드러낸다.
- 세계 경제에서 global value chain의 역할이나 그 영향이 어떠한지에 대한 맥락을 설명할 때 해당 키워드가 사용되었다. (cf. global value chain: 제품 및 서비스의 생산과 공급, 분배나 판매 후 활동이 여러 지역에 걸쳐서 조정되는 경우 이러한 것들에 관련된 사람들이나 역할 및 활동을 의미한다.)

### iii. brics

- (예시 1) Fungáčová et al. (2015)는 해당 논문에서 World Bank Global Findex 데이터베이스의 2011년 데이터를 사용하여 다른 BRICS 국가와의 비교를 포함하여 중국의 금융적 수용성(financial inclusion)을 분석한다. 저자들은 연구를 통해 중국에서 다른 BRICS보다 공식 계정(formal account)과 공식 저축(formal savings)을 더 많이 사용함으로써 높은 수준의 금융적 수용성이 나타난다는 것을 발견한다.

- (예시 2) Ortobelli et al. (2018)은 해당 논문에서 브라질, 러시아, 인도, 중국, 남아프리카 공화국 등 5개국에 적용되는 고전적인 포트폴리오 전략에 초점을 맞추고 있다. 저자들에 따르면, 이 5개국은 소위 BRICS 그룹을 형성한다. 특히 저자들은 회사채와 국채 시장을 조사하여 이러한 시장이 만족시킬 수 없고 위험을 회피하는 투자자들에게 수익성 있는 투자가 될 수 있는지 여부를 평가한다.
- BRICS에 해당하는 나라들을 대상으로 하거나 BRICS에 해당하는 나라들과 다른 나라의 금융경제적 관점을 비교·분석할 때 해당 키워드가 사용되었다. (cf. BRICS는 브라질 (Brazil), 러시아 (Russia), 인도 (India), 중화인민공화국 (China), 남아프리카 공화국 (South Africa)을 통칭하는 말로, 골드만삭스가 처음으로 사용하기 시작했다.)

#### iv. macroprudential

- (예시 1) Claessens (2015)에 따르면 macroprudential 정책은 시장 실패와 외부 효과에 의해 동기부여가 되어야하지만 이는 식별하기 어려울 수 있다. 또한 macroprudential 정책은 조정 문제를 제기하면서 통화 및 microprudential과 같은 다른 다양한 정책들과 상호작용할 수 있다. 저자는 논문에서 국가들 중에서도 특히 신흥 시장을 가진 국가들은 이러한 방식들을 사용해 왔다는 점을 언급하고 있고, 논문의 분석 결과들은 이러한 방식들 중 몇 가지가 경기 순응성과 위기 위험 (crisis risk)을 감소시킨다는 것을 보여준다.
- (예시 2) Dräger et al. (2020)는 논문에서 금융 부문이 유럽 통화 동맹 (European Monetary Union) 내에서의 거시 경제 불균형의 출현을 악화시켰을 수 있는 경로를 보여준다. 저자들에 따르면 이러한 효과는 macroprudential 정책들에 의해서 완화될 수 있는데 특히 은행의 대출 기준을 덜 순응적으로 만들도록 하는 정책들이 통화 동맹의 두 지역 모두에서 생산을 안정화하는 데 효과적임이 입증된다.
- 경제의 불균형이나 위기에 macroprudential 정책들이 어떻게 그

영향력을 발휘하고 부정적인 영향을 완화시킬 수 있는지 설명하는 맥락에서 해당 키워드가 등장하였다.

v. populism

- (예시 1) Perkins (2019)에 따르면 경제적 불평등은 공포에 기반한 포퓰리즘을 생산함으로써 기후 변화를 다루기 위한 정치적 여지를 감소시킨다. 저자는 모든 사회 구성원들의 안전, 사회적 지위, 그리고 생계가 보장될 때에만 자발적이고 민주적인 결정들이 기후 변화를 전환시키고 그 영향을 상당히 완화시킬 수 있다고 주장한다.
- (예시 2) Cha (2020)에 따르면, 세계 테러와의 전쟁의 대실패와 2008년의 글로벌 금융 위기를 포함하여 21세기의 첫 10년 동안 자유주의 세계 질서의 이중 위기 이후 우리는 대서양 전역에서 포퓰리즘의 폭발과 함께 미국 헤게모니와 초국가적 순간의 결합된 위기를 목격하고 있다. 이러한 맥락에서 저자의 연구는 미국이 어떻게 자유주의적 주간 질서 (liberal interstate order)와 세계화를 설계하고 유지해왔는지 분석할 뿐만 아니라 오늘날 패권국이 주도적으로 범국가적 프로젝트를 파괴하기 시작하는 방식을 분석한다.
- 단순히 정치적 맥락에서 뿐만 아니라 경제적 맥락에서 포퓰리즘이 어떠한 일이나 사건의 결과가 되거나 포퓰리즘 자체가 주체가 되어서 세계 경제와 정치에 어떻게 영향을 끼치는지 분석하는 맥락에서 해당 키워드가 사용되었다.

4. Arts and Humanities

가) 키워드

[표 5] Arts and Humanities 분야 키워드

세부 분야	키워드
환경 문제	glaciation / volcanic ash / co2 emission / human immunodeficiency virus / marine isotope / american medical / spectroscopy / boundary integral
심리학	comprehension strategy / personal interaction / signal processing / decisionmaking / younger adult / psychosomatic / psychiatric disorder / psychosomatic medicine / countertransference
언어학적	assisted language learning / translation training / translation competence / language learning / computer assisted language / language therapist
문학	horror genre / musicological / hypermedia / african traditional / life beyond / audiovisual medium / antiquity publication / contractualism / phantasm
사회	debt crisis / postcolonial nation / free trade agreement / intellectual property / misogynistic / material cultural / immigrant / extirpation / poverty / ethical obligation / familial relationship

나) 키워드 설명

i. psychosomatic

- (예시 1) Abitov (2015)는 심신증 및 신경증성장애를 겪는 환자들을 위한 일관된 심리적 부적응 모델 부재로 모델 개발의 의의를 설명한다.
- (예시 2) Freudenreich et al. (2015)는 자원봉사 전문가를 이용한 그룹 프로세스가 임상에게 심신의학의 중요한 출판물의 관리 가능한 선택권을 제공하기 위해 어떻게 활용될 수 있는지 소개한다.
- 심신증과 관련된 심리학적 모델 연구의 의의와 많아진 심신의학과 관련된 출판물 중 중요한 것을 어떻게 선택해야 할지 방안을 제시하는 연구에 psychosomatic이라는 단어가 쓰였다.

ii. extirpation

- (예시 1) Rawlence et al. (2015)에 따르면 메가답테스 펭귄의 멸종은 인공적인 힘이 빠른 생물지리학적 변화를 일으킬 수 있다는



점을 뒷받침하는 증거다.

- (예시 2) Schmitt and Lupo (2018)에 따르면 시원하고 습기가 많은 환경이 건조한 기후로 대체되면서 저지대 및 고지대에 있는 생물들은 분류학적으로 풍요로웠지만 중간 지대에 있는 동물들이 멸종했다.
- 생물 멸종과 관련된 연구에서 extirpation이라는 단어가 자주 쓰였다.

### iii. psychiatric disorder

- (예시 1) Fettich et al. (2015)는 정신질환 중 간헐적 폭발 장애 (IED)에 대한 효과적인 과정과 감정 조절에 대한 연구가 없다는 점을 짚으면서 연구의 의의를 밝힌다.
- (예시 2) Kompus et al. (2015)는 정신 질환을 예측하는 데 있어서 AVH의 가치를 조사하는 추가적인 종적인 연구가 필요하다는 점을 강조한다.
- 정신 질환 관련 연구에서 psychiatric disorder이라는 단어가 자주 쓰인다.

### iv. ethical obligation

- (예시 1) Spearey (2016)는 연극 공연의 맥락에서 사용되는 어떤 종류의 간청이 반응을 이끌어 내고 윤리적 행동을 조장하며 종종 윤리적 의무가 어디에 있고 어디까지 확장되는지 결정하는 근접성과 거리에 대한 이해를 재교정할 수 있는지 또는 방법을 조사한다.
- (예시 2) Mintz and Wasserman (2020)에 따르면 Kittay의 규범적 진료 이론이 보호자와 보호받는 사람 사이에 존재해야 할 윤리적 의무를 포착하는지는 여전히 의문으로 남아 있다.
- 연구를 함에 있어서 윤리적 의무를 고려한다는 맥락에서 ethical obligation이라는 단어가 쓰인다.

### v. phantasm

- (예시 1) Salin (2015)은 에로틱한 텍스트가 ‘이상한 나라의 앨리스’에서 병이 아니라 환상을 반영하는 논리적 텍스트가 되도록 인코딩 되었는지 연구한다.
- (예시 2) Oniščik (2015)에 따르면 토마스 아퀴나스의 지식 이론 중 지식의 대상에 대한 질문을 강조하면 안다는 것은 상상력과 이해하기 쉬운 종족으로 대표된다.
- 문학 및 이론가의 이론을 이해하기 위해 phantasm이라는 단어가 쓰인다.

## 2 학제 간 비교

### 1. 분야별 성능 비교

각 분야의 성능 분석으로는 AUC (Area Under the ROC Curve)를 사용하였다. ROC (Receiver Operating Characteristic)는 모든 임계값 (threshold)에서 이진 분류 모델의 성능을 보여주는 그래프이고 AUC는 ROC 곡선 아래 영역을 의미한다. AUC의 값이 클수록 클래스를 구별하는 모델의 성능이 뛰어나다는 것을 의미한다.

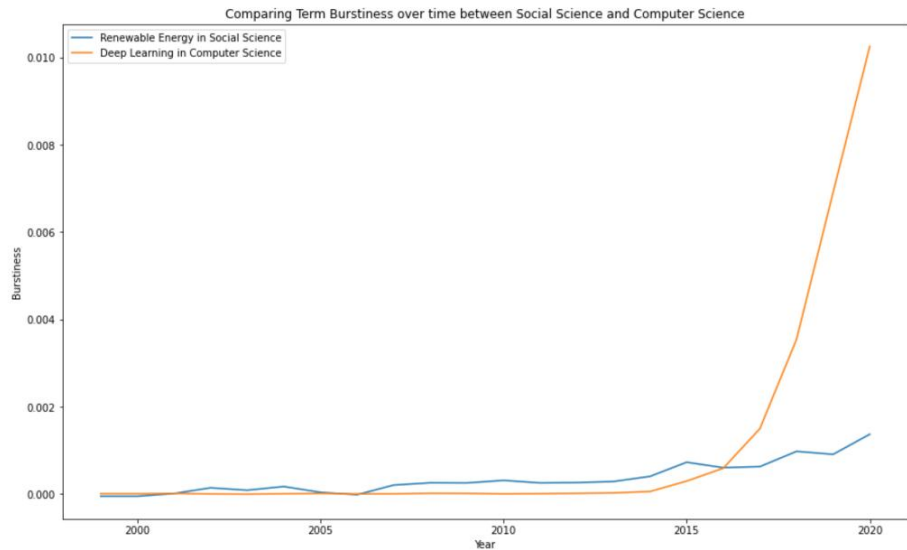
[표 6] Burstiness 상승 여부에 대한 모델 성능 비교

분야	AUC
Computer Science	0.88
Social Science	0.94
Economics	0.86
Arts and Humanities	0.77

각 분야의 AUC 값들을 비교해봤을 때, 기술 분야와 비기술 분야의 모델 예측 성능과 결과에 있어서 차이가 존재함을 알 수 있다. 결과적으로 도출된 이머징 이슈 키워드를 보면, 비기술 분야의 경우 기술 분야에 비해 이머징 이슈를 구별하기 어려운 특징이 나타났다. (기술 분야 AUC 평균: 0.88, 비기술 분야 AUC 평균: 0.86) 이어지는 파트에서 분야에 따라 결과가 상이하게 나타나는 이유를 분석해보았다.

## 2. 시계열 추이

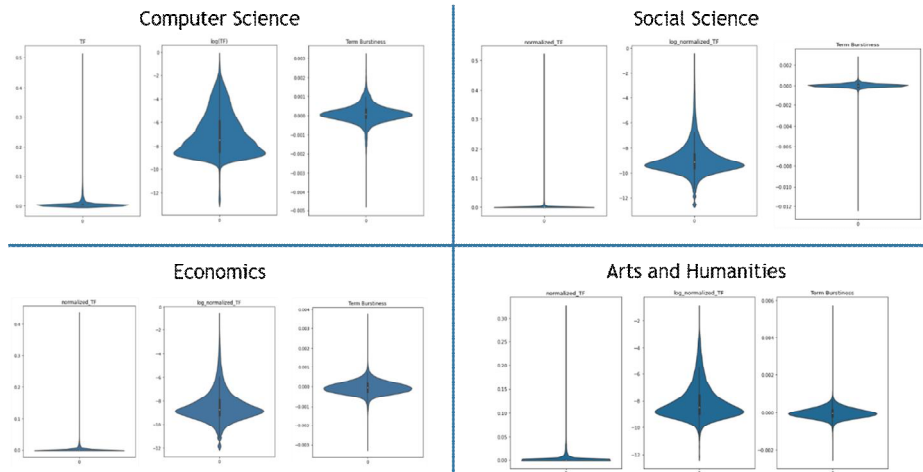
가) Computer Science 분야의 키워드 ‘deep learning’ (주황색)과 Social Science 분야의 키워드 ‘renewable energy’ (파란색)의 burstiness 시계열 추이 비교



[그림 3] Computer Science와 Social Science 분야의 키워드 시계열 추이 비교

- i. 기술 분야의 경우, 새로운 기술이 등장하면 이와 관련된 키워드가 급격히 증가하는 등 burstiness가 주제의 중요도와 관련하여 크게 변동함을 알 수 있다.
- ii. 반면 비기술 분야인 사회 과학의 경우, 중요하게 다루어 지는 토픽은 실업, 국제 관계 등 시대에 따라 변동 정도가 작고 burstiness도 크게 변동하지 않는 것을 확인할 수 있다.
- iii. 인문, 사회 과학 등의 비기술 분야의 경우 기술 분야에 비해 이머징 이슈를 구별하기 어렵다.

### 3. 바이올린 플롯



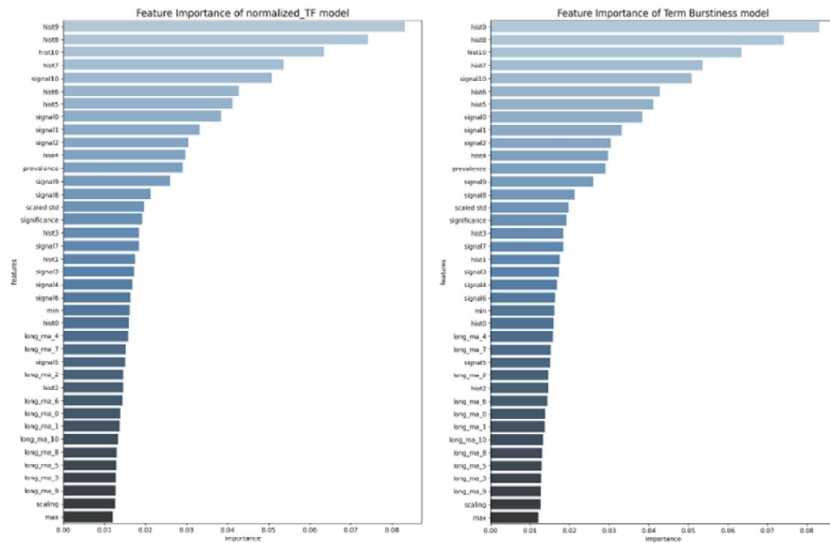
[그림 4] 각 분야별 Term Burstiness 바이올린 플롯

비기술 분야의 Term Burstiness는 기술 분야의 Term Burstiness에 비해 분산이 작다. 이를 통해 단어 간 Term Burstiness 차이가 비교적 작다는 것을 알 수 있다. 또한 Term Burstiness에 로그 정규화를 취했을 때 비기술 분야 Term Burstiness의 중앙값이 기술 분야 Term Burstiness의 중앙값보다 작은 것을 볼 수 있다. 중앙값이 데이터의 이상치를 고려한 평균이라고 생각했을 때 Term Burstiness의 평균이 기술 분야에서 더 높은 것을 알 수 있다.

#### 4. 이머징 이슈 결정 요인 분석

##### 가) Computer Science

### Computer Science Feature Importance



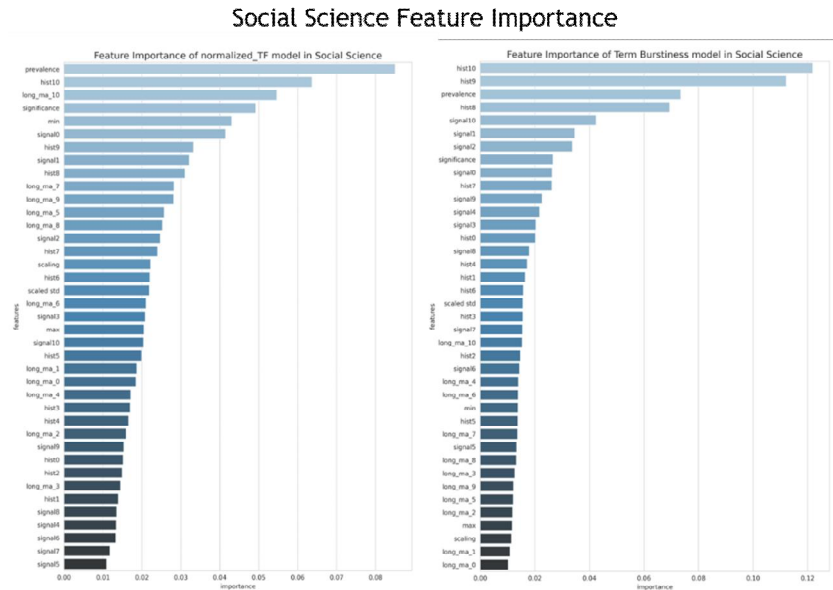
<그림 14> normalized TF 상승 여부 예측 모형의 feature importance

<그림 15> Term Burstiness 상승 여부 예측 모형의 feature importance

### [그림 5] Computer Science Feature Importance

- i. RandomForest, XGBoost, LGBM은 모두 종속변수 결정에 중요한 영향을 미치는 변수를 분석할 수 있다는 장점을 지닌다.
- ii. Term Frequency 상승 여부 예측 모형과 Burstiness 상승 여부 예측 모형 모두에서 최근의 히스토그램 지표가 중요한 역할을 하는 것으로 나타났다.
- iii. Long-term moving average는 영향력이 낮은 변수로 나타났다.
- iv. 히스토그램은 장기 이동평균과 단기 이동평균의 차이를 의미하므로, 장기적인 추세와 단기적인 추세의 상이한 방향과 정도가 미래의 영향력 상승 여부를 예측하는 중요한 변수라고 볼 수 있다.

나) Social Science



[그림 6] Social Science Feature Importance

- i. RandomForest, XGBoost, LGBM은 모두 종속변수 결정에 중요한 영향을 미치는 변수를 분석할 수 있다는 장점을 지닌다.
- ii. Term Frequency 상승 여부 예측 모형과 Burstiness 상승 여부 예측 모형 모두에서 prevalence와 hist 9, 10이 높게 나타났다. 이는 예측에 있어서 예측 시점 기준 최근 데이터가 중요하다는 것을 시사한다.
- iii. Long-term moving average가 Term Frequency 예측 모델에서는 높게 나타나는 반면, signal은 Term Burstiness 예측 모델에서 높게 나타나는 것을 확인할 수 있다.

## 제 4장

### 연구 결과의 시사점 및 결론

---

제1절 연구 결과의 시사점 및 결론

## 1 연구 결과의 시사점

### 1. 학제 간 공통적으로 등장한 키워드

본 연구에서는 Computer Science, Social Science, Economics, Arts and Humanities의 네 가지 분야의 학술문헌을 통한 이머징 키워드 분석을 실시했다. 학제에 공통적으로 등장한 키워드들을 통해 이머징 이슈의 전반적인 흐름을 보다 잘 이해할 수 있을 것으로 기대된다.

첫째, 이번 연구 결과에서 주목할 만한 점은 4차 산업혁명과 관련된 키워드들이 다수 등장했다는 점이다. 특히 Computer Science와 Social Science 분야의 경우 그 특징이 두드러졌다. Computer Science 분야에서는 인공지능과 밀접한 연관이 있는 human robot interaction (인간로봇 상호작용)이나 recurrent neural network (순환신경망 모델), 분산 시스템과 관련된 blockchain network나 cryptocurrency, 그리고 전반적인 과학 기술과 관련된 5g network와 energy harvesting, IoT data, unmanned aerial vehicle (무인 항공기) 등의 키워드들이 등장하였다. 또한 Social Science 분야에서도 마찬가지로 인공지능과 관련하여 deep learning이나 machine learning, convolutional neural network (합성곱 신경망), LSTM (Long Short-Term Memory; 장단기 메모리)가 등장하였고 기타 과학 기술과 관련한 blockchain technology, VR (Virtual Reality), smart city, fourth industrial revolution, industry 4.0, thing IoT 등이 등장하였다. 해당 키워드들은 모두 4차 산업혁명의 바탕이 되는 과학 기술이거나 그 기술의 결과물로, 매우 밀접한 관련성을 보인다.

둘째, 환경문제와 관련된 키워드도 학제간 공통적으로 등장하였다. Computer Science의 경우 enenergy harvesting, Social Science의 경우 ecosystem service, Arts and Humanities의 경우 co2 emission가 이머징 키워드로 나타났다.



셋째, Social Science 분야의 misinformation과 Economics 분야의 populism은 선거와 관련된 부정적인 측면을 시사한다. 인터넷 발달에 따른 잘못된 정보 또는 선별된 정보의 수용과, 이와 무관하지 않은 포퓰리즘이 미치는 경제적, 정치적 악영향의 연관성으로 해석할 가능성이 있다.

## 2. 학제간 차이점: 기술분야와 비기술 분야를 중심으로

학제간 차이와 관련해서는 기술분야와 비기술 분야의 차이가 특히 두드러졌다. Computer Science에 비해 Social Science, Economics, Arts and Humanities 분야의 이머징 키워드들은 새롭게 등장한 단어라기보다는 이전부터 꾸준히 등장하던 단어의 성격이 강하다.

이는 키워드 시계열 추이나 Term Burstiness의 바이올린 플롯을 통해서도 확인할 수 있었다. 실제로 Computer Science 분야의 'deep learning'이라는 키워드는 대략 2017년을 기점으로 term burstiness가 폭발적으로 증가하였지만 Social Science 분야의 'renewable energy' 키워드의 경우 시기별로 term burstiness에 큰 차이가 없음을 확인하였다. 즉, 비기술 분야인 인문이나 사회과학 등의 분야의 경우 기술 분야에 비해 이머징 이슈를 구별하기 어려운 특징이 나타났다.

또한, 각 분야별 Term Burstiness의 바이올린 플롯들을 확인해보았을 때 비기술 분야의 Term Burstiness가 분산이 더 작았는데 이는 단어 간 Term Burstiness 차이가 비교적 작다는 것을 의미한다. 비슷한 맥락으로 로그 정규화를 취한 Term Burst연구 결과의 시사점 iness의 중앙값은 기술 분야에서 더 큰 값을 보였는데, 중앙값이 이상치를 고려한 평균이라는 것을 가정했을 때 기술 분야의 Term Burstiness의 평균이 더 높다는 것을 확인할 수 있었다.

## 2 결론

이처럼 이번 연구에서는 학술정보 데이터베이스인 Scopus에서 총 4개의 분야에서 30년의 영문 학술 문헌 제목과 초록을 수집하여 텍스트 마이닝 기술을 통해 이머징 이슈를 도출하였고 각 분야 전문가들과의 협업을 통해 그 중 유의미한 이머징 키워드들을 선정하였다. 그리고 각 분야별 이머징 키워드들의 의미와 맥락을 살펴보고 기술/비기술 분야의 이머징 이슈들이 어떠한 특징과 차이를 보이는지 확인해보았다.

## 참고 문헌

---

### 1. 문헌 자료

## 참고 문헌

NATIONAL ASSEMBLY FUTURES INSTITUTE

### 1 문헌 자료

임치현, 이창현(2018), 「4차 산업혁명의 주요 융합 R&D 이슈 파악: 텍스트 마이닝을 통한 접근」, 융합연구리뷰, 4(11), 융합연구정책센터, 3-34

김현정, 조남옥, 신경식(2014) 「항공산업 미래 유망분야 선정을 위한 텍스트 마이닝 기반의 트렌드 분석」, 한국정보시스템학회학술대회논문집, 한국지능시스템학회, 194-202

유순덕, 이민수, 신선영(2014), 「미래예측 방법론」, 주간기술동향 통권, 1646, 정보통신산업진흥원, 1-13

안태천(2015), 「Extreme Learning Machine 기반 퍼지 패턴 분류기 설계」, 한국지능시스템학회 논문지, 25(5), 한국지능시스템학회, 509-514

Tattershall, E., Nenadic, G. & Stevens, R. D. (2020). "Detecting bursty terms in computer science research" *Scientometrics*, 122, 681-699. <https://doi.org/10.1007/s11192-019-03307-5>

Sun, L. , Yin, Y.(2017). "Discovering themes and trends in transportation research using topic modeling", *Transportation Research Part C: Emerging Technologies*, 77, 49-66, <https://doi.org/10.1016/j.trc.2017.01.013>.

Hiltunen, E.(2008), "The future sign and its three dimensions", *Futures*, 40(3), 247-260, <https://doi.org/10.1016/j.futures.2007.08.021>.

Tu, I. and Seng, J.L.(2012) "Indices of novelty for emerging topic detection", *Information Processing and Management*, 48(2), 303-325,

<https://doi.org/10.1016/j.ipm.2011.07.006>

Yoon J.(2012), “Detecting weak signals for long-term business opportunities using text mining of Web news”, *Expert Systems with Applications*, 39(16), 12543-12550, <https://doi.org/10.1016/j.eswa.2012.04.059>.

Kleinberg, J.(2003), “Bursty and Hierarchical Structure in Streams.” *Data Mining and Knowledge Discovery*, 7, 373-397, <https://doi.org/10.1023/A:1024940629314>

Takahashi, Y. & Utsuro, T. & Yoshioka, M. & Kando, N. & Fukuhara, T. & Nakagawa, H. & Kiyota, Y. (2012). “Applying a Burst Model to Detect Bursty Topics in a Topic Model.” *Advances in Natural Language Processing*, 239-249 [https://doi.org/10.1007/978-3-642-33983-7\\_24](https://doi.org/10.1007/978-3-642-33983-7_24).

He, D., Parker, D. S. (2010). “Topic dynamics: an alternative model of bursts in streams of topics” *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, 443-452, <https://doi.org/10.1145/1835804.1835862>

Abdel-Basset, M., Gamal, A., Son, L. H & Smarandache, F. (2020). “A Bipolar Neutrosophic Multi Criteria Decision Making Framework for Professional Selection” *Applied Sciences*, 10, 1202, <https://doi.org/10.3390/app10041202>

Mathioudakis, M. & Koudas, N.(2010), “TwitterMonitor: trend detection over the twitter stream” *In Proceedings of the 2010 ACM SIGMOD International Conference on Management of data (SIGMOD '10)*. 1155-1158. <https://doi.org/10.1145/1807167.1807306>

Abdel-Basset, M., Ali, M. & Atef, A. (2020). “Resource levelling problem in construction projects under neutrosophic environment” *The Journal of Supercomputing*, 76, 964-988, <https://doi.org/10.1007/s11227-019-03055-6>

Pan C. et al., (2020). "Intelligent Reflecting Surface Aided MIMO Broadcasting for Simultaneous Wireless Information and Power Transfer," *IEEE Journal on Selected Areas in Communications*, 38, 1719-1734, <https://doi.org/10.1109/JSAC.2020.3000802>

Khan, L. U., Yaqoob, I., Imran, M., Han, Z. & Hong, C. S. (2020). "6G Wireless Systems: A Vision, Architectural Elements, and Future Directions," *IEEE Journal on Selected Areas in Communications*, 8, 147029-147044, <https://doi.org/10.1109/ACCESS.2020.3015289>.

Panicucci S, Nikolakis N, Cerquitelli T, Ventura F, Proto S, Macii E, Makris S, Bowden D, Becker P, O'Mahony N, Morabito L, Napione C, Marguglio A, Coppo G, Andolina S. (2020). "A Cloud-to-Edge Approach to Support Predictive Analytics in Robotics Industry. Electronics." *Robotics Industry. Electronics*, 9(3), 492, <https://doi.org/10.3390/electronics9030492>

Ionescu, C., Alfredo Cajo Diaz, R., Zhao, S., Ghita, M., Ghita, M. & Copot D. (2020). "A Low Computational Cost, Prioritized, Multi-Objective Optimization Procedure for Predictive Control Towards Cyber Physical Systems," *IEEE Access*, 8, 128152-128166, <https://doi.org/10.1109/ACCESS.2020.3007924>.

Barchard, K. A., Lapping-Carr, L., Shane Westfall, R., Fink-Armold, A., Banisetty, S. B., & Feil-Seifer, D. (2020). "Measuring the Perceived Social Intelligence of Robots." *Human-Robot Interactacion*, 9, 1-29, <https://doi.org/10.1145/3415139>

Iyer, S. S., Joseph, J. V., & Vashista, V. (2020). "Evolving Toward Subject-Specific Gait Rehabilitation Through Single-Joint Resistive Force Interventions." *Front Neurorobot*, 14, 15, <https://doi.org/10.3389/fnbot.2020.00015>

Wei, Y. et al. (2020). "Predicting Entrepreneurial Intention of Students: An Extreme Learning Machine With Gaussian Barebone Harris Hawks Optimizer," *IEEE Access*, 8, 76841-76855, <https://doi.org/10.1109/ACCESS.2020.2982796>

Mohant, F., Rup, S., Dash, B. (2020). "Automated diagnosis of breast cancer using parameter optimized kernel extreme learning machine", *Biomedical Signal Processing and Control*, 62, 102108, <https://doi.org/10.1016/j.bspc.2020.102108>.

Brown, G. & Fagerholm, N. (2015). "Empirical PPGIS/PGIS mapping of ecosystem services: A review and evaluation". *Ecosystem Services*, 13, 119-133. <https://doi.org/10.1016/j.ecoser.2014.10.007>

McPhearson, T., Andersson, E., Elmqvist, T., Frantzeskaki, N. (2015). "Resilience of and through urban ecosystem services". *Ecosystem Services*, 12, 152-156. <https://doi.org/10.1016/j.ecoser.2014.07.012>

Beery T. (2020), "What We Can Learn from Environmental and Outdoor Education during COVID-19: A Lesson in Participatory Risk Management." *Sustainability*, 12(21), 9096. <https://doi.org/10.3390/su12219096>

Lim H. Y. (2020). "Ethics Education for Successful Infectious Disease Control of COVID-19." *Asian bioethics review*, 12(2), 1-9. Advance online publication. <https://doi.org/10.1007/s41649-020-00124-4>

Morales-i-Gras, J. (2020). "Cognitive Biases in Link Sharing Behavior and How to Get Rid of Them: Evidence from the 2019 Spanish General Election Twitter Conversation." *Social Media + Society*. 6. 205630512092845. <https://doi.org/10.1177/2056305120928458>.

Greenspan, R. L., & Loftus, E. F. (2020). "Eyewitness confidence malleability: Misinformation as post-identification feedback." *Law and human behavior*, 44(3), 194-208. <https://doi.org/10.1037/lhb0000369>

Davis, N., Raina, G., Jagannathan, K. (2020). "A framework for end-to-end deep learning-based anomaly detection in transportation networks."

*Transportation Research Interdisciplinary Perspectives*, 5.  
<https://doi.org/10.1016/j.trip.2020.100112>.

Lu, S., Zhang, Q., Chen G., Seng, D. (2021). “A combined method for short-term traffic flow prediction based on recurrent neural network.” *Alexandria Engineering Journal*, 60(1), 87-94, <https://doi.org/10.1016/j.aej.2020.06.008>.

Ha, H., Kwak, Y., Kim, H., Seo, Y. (2020). “Discomfort luminance level of head-mounted displays depending on the adapting luminance.” *Color Research and Application*, 45: 622-631. <https://doi.org/10.1002/col.22509>

LaRocco, Michael. (2020). “Developing the ‘best practices’ of virtual reality design: industry standards at the frontier of emerging media.” *Journal of Visual Culture*. 19. 96-111. <https://doi.org/10.1177/1470412920906255>.

Abitov, I. (2015). “Model of Psychological Disadaptation at Psychosomatic and Neurotic Disorders.” *Review of European Studies*, 7, 136-140. <https://doi.org/10.5539/res.v7n1p136>.

Freudenreich, O., Huffman, J. C., Sharpe, M., Beach, S. R., Celano, C. M., Chwastiak, L. A., Cohen, M. A., Dickerman, A., Fitz-Gerald, M. J., Kontos, N., Mittal, L., Nejad, S. H., Niazi, S., Novak, M., Philbrick, K., Rasimas, J. J., Shim, J., Simpson, S. A., Walker, A., Walker, J., ... Stern, T. A. (2015). “Updates in Psychosomatic Medicine: 2014.” *Psychosomatics*, 56(5), 445-459. <https://doi.org/10.1016/j.psych.2015.04.001>

Rawlence, N.J., Perry, G., Smith, I., Scofield, R., Tennyson, A., Matisoo-Smith, E., Boessenkool, S., Austin, J., & Waters, J. (2015). “Radiocarbon-dating and ancient DNA reveal rapid replacement of extinct prehistoric penguins.” *Quaternary Science Reviews*, 112, 59-65.

Schmitt, D. N., & Lupo, K. D. (2018). “On early-Holocene moisture and small-mammal histories in the Bonneville basin, Western United States.” *The Holocene*, 28(3), 492-498. <https://doi.org/10.1177/0959683617729453>

Fettich, K. C., McCloskey, M. S., Look, A. E., & Coccaro, E. F. (2015). "Emotion regulation deficits in intermittent explosive disorder." *Aggressive behavior*, 41(1), 25-33. <https://doi.org/10.1002/ab.21566>

Kompus, K., Løberg, E. M., Posserud, M. B., & Lundervold, A. J. (2015). "Prevalence of auditory hallucinations in Norwegian adolescents: results from a population-based study." *Scandinavian journal of psychology*, 56(4), 391-396. <https://doi.org/10.1111/sjop.12219>

Spearey, S. (2016) "Fostering receptivity: cultural translation, ethical solicitation, and the navigation of distance in J.T. Rogers'" *The Overwhelming*, Safundi, 17:2, 231-248, <https://doi.org/10.1080/17533171.2016.1172800>

Mintz, K. and Wasserman, D. (2020), "Caring for People with Disabilities: An Ethics of Respect." *Hastings Center Report*, 50: 44-45. <https://doi.org/10.1002/hast.1084><sup>[SEP]</sup>

Salin, S. (2015). "Concept of corporeal and erotical semiotic in Gilles Deleuze's logic of sense. Deleuze's way to approach a «grammar of phantasms» in Lewis Carroll's work." *Studii si Cercetari Filologice*, Seria Limbi Romanice. 1. 59-72.

Oniščik M. (2015). "Empiricism of Thomas Aquinas: The Question of the Object of Knowledge." *Problemos*, 88(88), 66-79. <https://doi.org/10.15388/Problemos.2015.88.8479>

Berdegúe, J. A., Carriazo, F., Jara, B., Modrego, F., & Soloaga, I. (2015). "Cities, Territories, and Inclusive Growth: Unraveling Urban-Rural Linkages in Chile, Colombia, and Mexico." *World Development*, 73, 56-71. [https://doi.org/https://doi.org/10.1016/j.worlddev.2014.12.013](https://doi.org/10.1016/j.worlddev.2014.12.013)

Sehrawat, M., & Giri, A. K. (2018). "The impact of financial development, economic growth, income inequality on poverty: evidence from India." *Empirical Economics*, 55(4), 1585-1602.



<https://doi.org/10.1007/s00181-017-1321-7>

Wu, A., Song, D., & Yang, Y. (2020). "Untangling the effects of entrepreneurial opportunity on the performance of peasant entrepreneurship: the moderating roles of entrepreneurial effort and regional poverty level." *Entrepreneurship & Regional Development*, 32(1-2), 112-133. <https://doi.org/10.1080/08985626.2019.1640479>

Fernández, V.R. (2015). "Global Value Chains in Global Political Networks." *Review of Radical Political Economics*, 47, 209 - 230.

Imbruno, M. (2019). "Importing under trade policy uncertainty: Evidence from China." *Journal of Comparative Economics*, 47(4), 806-826. <https://doi.org/https://doi.org/10.1016/j.jce.2019.06.004>

Fungáčová, Z., & Weill, L. (2015). "Understanding financial inclusion in China." *China Economic Review*, 34, 196-206. <https://doi.org/https://doi.org/10.1016/j.chieco.2014.12.004>

Ortobelli, S., Petronio, F., & Vitali, S. (2018). "Price and market risk reduction for bond portfolio selection in BRICS markets." *Investment Management and Financial Innovations*, 15, 120-131. [https://doi.org/10.21511/imfi.15\(1\).2018.11](https://doi.org/10.21511/imfi.15(1).2018.11)

Claessens, S. (2015). "An Overview of Macroprudential Policy Tools." *Annual Review of Financial Economics*, 7(1), 397-422. <https://doi.org/10.1146/annurev-financial-111914-041807>

Dräger, L., & Proaño, C. (2020). "CROSS-BORDER BANKING AND MACROPRUDENTIAL POLICIES IN ASYMMETRIC MONETARY UNIONS." *Macroeconomic Dynamics*, 24(2), 255-290. doi:10.1017/S1365100518000214

Perkins, P. E. (2019). "Climate justice, commons, and degrowth." *Ecological Economics*, 160, 183-190. <https://doi.org/https://doi.org/10.1016/j.ecolecon.2019.02.005>

Cha, T. (2020). "Is anybody still a globalist? Rereading the trajectory of US grand strategy and the end of the transnational moment." *Globalizations*, 17(1), 60-76. <https://doi.org/10.1080/14747731.2019.1611011>