

03

물환경을 이해하는 도구, 인간사회 탐사(Human Social Sensing)



이주성
서울과학기술대학교 환경공학과
/ 환경모델링연구실 박사과정생
juseonglee90@gmail.com



김경민
서울과학기술대학교 환경기술연구소
/ 연구원
kyun9minkim@gmail.com



김정진
서울과학기술대학교 환경기술연구소
/ 연구교수
kimjj82@seoultech.ac.kr



정한석
서울과학기술대학교 환경공학과
/ 교수
hanjeong@seoultech.ac.kr

1. 머리말

“Knowing is easy, understanding is hard.”

이는 것과 이해하는 것은 다르다는 간단하면서도 어려운 말은 물환경시스템을 이해하기 위한 우리의 노력을 요약하고 있다. 우리는 물환경을 더 잘 알고 또 이해하기 위해 다양한 데이터를 수집하고 있다. 현재 우리나라에서 운영 중인 물환경측정망은 8종류로, 수질, 생물, 비점오염물질, 퇴적물, 그리고 총량 등을 포함하고 있으며, 그중 수질측정망은 54개의 분석항목을 가진다. 이는 다각도로 물환경을 이해하고자 하는 노력의 일환이다.

물환경을 이해(Understanding)하는 것은 물환경을 관측하고 관측을 통해 수집한 데이터를 들여다보는 과정에서 물환경의 다양한 가능성을 해석(Translation)하는 것을 의미한다. 관측 데이터의 해석은 주로 관측값의 중심과 산포 등 통계적 의미를 찾음으로써 물환경의 현재 상태와 정도를 이해하는 과정으로 정의할 수 있으며, 이는 정형데이터의 효과적 활용과 관련이 깊다. 하지만 관측을 통해 수집한 정형데이터는 특정 시점의 정보만을 포함하기에, 동적이고 불확실한 물환경의 변화과정을 대표하기 어렵다.

물환경은 계속해서 변화하며, 이는 곧 물환경의 (우리가 말하는) 현재와 미래 상태를 실제로 이해하는 것이 매우 어렵다는 것을 의미한다. 다시 말해, 물환경에 대해 정형화된 데이터를 분석하는 것만으로는 물환경을 충분히 이해하기 어려우며, 진정한 물환경에 대한 이해는 물환경의 변화과정을 대변해주는 다양한 정보가 함께 해석되었을 때 가능함을 의미한다. 이는 물리화학적 대리변수 외에 물환경의 변화를 야기하는 그 밖의 다른

변수들의 탐구 필요성을 제기하며, 같은 맥락에서 물환경의 구성요소로서 인간사회를 포함하고자 하는 연구가 이루어져 왔다(Jeong 등, 2020; Sivapalan 등, 2012).

물환경과 관련된 인간의 태도, 의견, 감정과 같은 정보는 정형데이터의 해석만으로 취득하기 어렵다. 인간 관련 정보는 인간 스스로와 인간 사이(사회)에서 만들어진 데이터를 해석함으로써 얻을 수 있기 때문이다. 최근, 인간과 사회에서 만들어지는 데이터를 해석하기 위한 수단으로 인간사회탐사(Human Social Sensing)가 소개된 바 있다(Galesic 등, 2021). 인간사회탐사는 인간의 행동, 의견, 태도 등을 분석하여 인간과 사회현상을 이해하고 설명하는 데 중점을 둔 방법론이다. 인간사회탐사는 앞에서 제시된 정형데이터 이용의 한계를 보완하고 물환경의 실제적 이해에 도움을 줄 수 있다. 따라서 본고에서는 인간사회를 이해하기 위한 도구로 제안된 인간사회탐사를 소개하고, 물환경 이해의 도구로써 인간사회탐사의 활용 예를 제시하고자 한다.

2. 인간사회탐사

Galesic 등 (2021)은 빅데이터 환경과 비약적으로 발전하는 데이터 분석기술을 활용하는 ‘인간사회탐사’를 제안한 바 있다. 기존에는 인간과 사회를 이해하기 위해 설문 조사라는 전통적인 방식을 이용하여 대중들의 신념, 가치, 태도 등을 파악하였다(Olsson 등, 2021). 설문 조사는 전체 집단을 이해하는 것을 목적으로 하나의 표본 집단에 관련질문을 통해 정보를 수집하는 방법이다. 주로 설문지나 인터뷰를 활용하여 응답자의 의견을 수집하게 되며, 구조화된 질문을 통해 일정한 형식으로 된 응답을 얻게 된다. 하지만 응답자는 주

어진 질문에 대해 제한된 선택지 내에서 답변하는 역할에 그치게 되며, 응답자의 주관적 의견이나 감정을 자세히 파악하기 어렵다는 한계가 존재한다(Wang 등, 2019). 그리고, 응답자들이 설문 조사의 질문을 다르게 해석할 여지 또한 존재한다(Enns 등, 2017). 결정적으로 현대사회에서 인간의 가치, 태도 행동 등은 빠르게 변하고 있음에도 불구하고 많은 시간과 비용이 요구되는 설문 조사는 현대사회의 동적 변화에 부응하기 어렵다(Tucker 등, 2018).

현대사회에서 끊임없이 생성되고 있는 비정형데이터는 인간사회탐사의 새로운 수단으로 활용될 수 있다. 컴퓨팅 및 네트워크 기술의 발전은 사람들이 시간과 장소에 구애받지 않고 콘텐츠를 생성하고 공유하는 환경을 만들었다(DataReportal, 2023). 개인들은 다양한 소셜미디어 플랫폼에서 자기 생각과 일상의 경험을 텍스트, 사진, 영상 등 다양한 형태로 기록하고 공유한다. 그리고 이 과정에서 생성된 비정형데이터는 현실 세계의 다양한 의견과 감정을 반영한다(Rydning 등, 2018). 이와 같은 비정형데이터는 실제 대중들의 경험과 피드백을 알아채는 데 유용하게 사용될 수 있으며 인간의 주관적인 측면을 더 정확하게 이해할 수 있게 한다.

인간사회탐사는 실시간으로 방대한 데이터를 분석하므로 빠르게 변화하는 인간사회에 대한 즉각적인 이해를 가능하게 하고, 대량의 응답자 확보를 통해 다양한 의견과 행동을 파악할 수 있게 한다. 더욱이 통계적 분석이 주를 이뤘던 기존의 분석과 달리 데이터 시각화, 텍스트 분석, 기계학습, 자연어 처리와 같은 고급 분석기법들이 활용되며, 더 자세하게는 텍스트 감정 분석, 토픽 모델링, 예측분석 등이 사용될 수 있다. 비정형데이터의 분석은 기존에는 알 수 없었던 사회적 변화,

문제, 행동 패턴 등에 대한 풍부한 정보를 제공하며, 정형데이터만으로 파악하기 어려운 현대사회의 다양하고 동적인 특성을 반영하여 사회현상에 대한 깊은 이해를 제공한다(Lepping 2018).

3. 인간사회탐사 활용 물환경에 대한 인간 인식변화 분석

인간사회탐사를 신문기사에 적용하여 물환경에 대한 인간의 인식변화를 분석하였다. 신문기사는 당시의 사회적 이슈와 의견을 반영하기 때문에 특정 시기에 대중의 사회적 가치변화를 추적하기 위해 가장 흔히 사용되는 언론매체이다(Lee와 Kwon, 2020). 따라서 인간사회탐사 기법의 하나인 텍스트마이닝을 이용하여 물환경과 관련한 과거 신문기사(1960~2020년) 키워드를 분석하고 대중의 사회 환경적 가치변화를 탐사하였다. 텍스트마이닝은 비정형텍스트 데이터를 수집, 전처리, 분석하여 데이터 내 패턴을 찾거나 시각화함으로써 새로운 정보를 추출하는 기법이다. 텍스트마이닝은 방대한 양의 데이터를 분석할 수 있는 도구이지만, 객관적 결과 도출 과정과 정확성에 대한 의문이 제기될 수 있다는 한계점이 존재한다(Kononova 등, 2021). 따라서, 개인이 직접 데이터를 검토하고 정해진 기준에 따라 분류하는 방식인 수동코딩(Manual Coding)을 텍스트마이닝의 검증방법으로 활용하였다.

수동코딩과 기계학습을 통해 물환경과 관련한 신문기사를 분류하고 그 결과를 비교하였다. 신문기사는 기사 주제에 따라 농촌용수, 도시 상하수도, 자연재해 및 피해, 물정책, 댐, 물환경 7개로 분류되었다. 수동코딩은 사람(휴먼코더)이 직접 신문기사를 읽고 신문기사 이면의 어조를 판별하여 기사를 분류하는 방식으로 이루어졌고, 기

계학습은 자연어 처리를 위해 한글 형태소분석기인 KoNLP MeCab(동전한뉘)을 이용해 신문 기사를 형태소 단위로 구분하여 데이터 전처리를 진행하였다. 이 과정에서 불필요한 숫자, 두 칸 이상의 공백, 문장 부호 등과, 언론사, 고유명사 등 중요하지 않은 단어는 제외되었다. 데이터 분석을 위한 기계학습 알고리즘으로는 Logistic Regression(LR)이 선정되었다. LR은 회귀를 사용하여 데이터가 어떤 범주에 속할 확률을 0에서 1 사이의 값으로 예측하여 그 확률에 따른 가능성이 더 큰 범주에 속하는 결과를 바탕으로 입력값들을 분류해주는 지도 학습방법으로, 주로 스팸메일 분류에 많이 사용된다. 기계학습 모형의 훈련은 각각 주제별 신문기사에서 상위 100개의 명사 키워드를 추출하고, 추출된 키워드를 학습데이터로 이용하여 훈련시키는 방식으로 이루어졌다. 학습 데이터에 대한 모형의 훈련 정확도는 1.0으로 훈련이 매우 잘된 것으로 나타났으며, 실제 기사의 분류 정확도는 0.86으로 해당 모형이 인간사회탐사에 활용 가능한 수준의 정보를 추출함을 확인했다.

텍스트마이닝 결과를 토픽 모델링하여 시간에 따른 물환경 관련 주제 비중의 변화를 살펴보고 우리나라 물환경 관련 주요정책 변화와 비교분석하였다. 토픽 모델링은 문서 내 비정형데이터에서 핵심단어에 따라 주제를 분류해주는 방법으로, 잠재디리클레 할당(Latent Dirichlet Allocation, LDA)은 확률적 생성모형으로써 토픽 모델링의 대표적인 알고리즘 중 하나이다(Blei 등, 2003). 토픽 트렌드 분석을 위해 7개의 주제를 LDA분석하였고, 도출된 결과를 포괄하는 용어를 각 주제에 명명하였다. 이어서 우리나라 물환경 관련 주요정책 변화를 정리하였다(그림 1).

정책은 인간 활동을 대리하는 중요한 요소 중 하나로, 사회 경제 수준과 사회적 가치에 따라 정

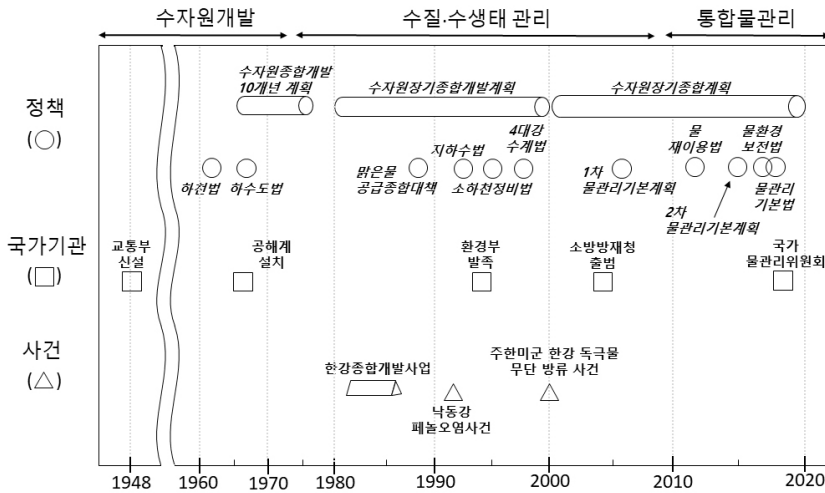


그림 1. 물환경 관련 주요정책 및 계획의 변천

책 기초가 달라지고, 이는 곧 물환경의 반응으로 나타난다. 우리나라의 물환경 관련 정책은 시대별 주안점에 따라 경제개발 단계, 수질관리 단계, 통합물관리 단계로 나눌 수 있다. 1960~1990년은 경제개발 단계로, 경제발전을 위한 안정적인 수자원 공급, 자연재해 관리를 위한 이수과 치수에 중점을 둔 정책들이 시행되었다. 이후 1990년대 여러 수질 사고로 물환경에 대한 국민적 관심 증대는 물환경 정책을 수질관리 단계로 이끌었다. 이 과정에서 정부 주도의 국민 편익과 수생태 전반을 위한 “깨끗한 물” 확보가 시작되었다. 그리고 2018년 제정된 물관리기본법을 기반으로 하여 지속가능한 발전을 위한 새로운 물관리 패러다임의 시작을 알리며 통합물관리 단계로 변화하였다.

토픽 모델링을 통해 얻은 물환경 주제 비중 변화와 물환경 관련 주요정책 변화를 비교분석한 결과, 둘은 유사한 경향을 띠는 것으로 나타났다(그림 2). 토픽 트렌드를 살펴보면 경제개발 단계에는 농촌용수와 자연재해 및 피해와 관련된 주제의 비중이 지배적이다. 1960년대에는 전후복구와 홍수 예방을 위한 사회기반시설 확보에 중점을 둔

정책들이 시행되었기 때문으로 보인다. 이 시기에는 물환경의 사회적 가치가 경제개발의 가치보다 상대적으로 낮았다. 하지만 1989년 수돗물 오염 파동, 1991년 낙동강 페놀유출사건 등 여러 건의 수질사고로 인해 전국적으로 환경, 특히 수질에 대한 국민적 관심이 커지게 되면서 수질관리 단계에 접어들게 되었으며, 이에 맞춰 수질 관련 주제의 비중은 점차 증가하는 경향을 보인다. 반면, 농촌용수와 자연재해 관련 문제들이 댐 건설 및 사회기반시설 확보 등으로 해소되며 관련된 주제의 비중은 점차 줄어드는 경향을 보였다. 이후 4대강 정비사업으로 인해 댐 관련 주제의 비중이 증가하였다. 물환경, 물정책 및 도시 상하수도는 꾸준히 관심을 받는 주제이며, 정책이나 관리체계의 변화에 따라 빈도가 달라졌다.

4. 마치면서

본고에서는 변화하는 물환경을 더 잘 이해하기 위한 방법으로 인간사회탐사를 소개하고 비정형 데이터를 이용해 물환경에 대한 대중의 인식변화

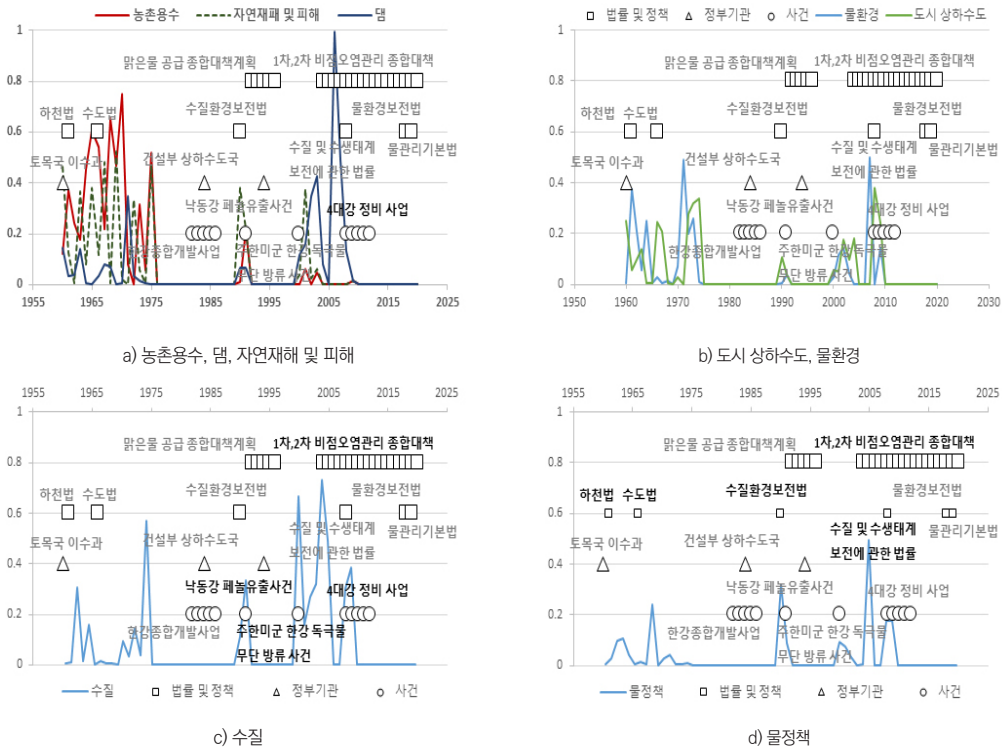


그림 2. 물환경 관련 주제 비중의 변화

를 분석한 사례를 다루었다. 빅데이터 환경과 데이터 분석기술 발전을 기반으로 제안된 인간사회탐사는 인간과 사회현상을 이해하고 설명하는 데 중점을 둔 방법론으로 현대사회의 변화와 전통적인 분석방법의 차이를 극복할 수 있음을 시사했다. 수동코딩과 기계학습을 통한 비정형데이터 분석결과, 기계학습 모형은 적절한 분류성능을 보였으며 토픽 모델링 결과 역시 물환경정책 변화와 유사한 경향을 보이며 비정형데이터 분석을 통한 인간사회탐사의 가능성과 비정형데이터 분석의 필요성을 제시하였다. 이처럼 인간사회탐사는 물환경과 인간사회의 긴밀한 상호작용을 더 깊게 이해하기 위한 도구로 사용 가능함을 확인하였다. 추후 텍스트가 아닌 다른 형태의 비정형데이터 분석 연구와 비정형데이터와 정형데이터의 유기적

분석을 통해 인간을 포함한 물환경시스템의 복잡한 상호작용을 모형화하고 이해하는 연구 등이 제안될 수 있다. 이를 통해 인간사회탐사는 복잡한 물환경을 이해하는 새로운 도구로 지속 가능한 미래 물관리 및 물환경정책 수립 방향 설정에 도움을 줄 수 있을 것으로 기대된다.

참고문헌

1. Blei, D., Ng, A. and Jordan, M. 2003. Latent Dirichlet Allocation, Journal of Machine Learning Research, 3, 993–1022.
2. DataReportal, 2023. Digital 2023: October Global Statshot Report: Number of internet and social media users worldwide as of October 2023.

+
특집 | 물환경을 이해하는 도구, 인간사회탐사(Human Social Sensing)

- <https://www.statista.com/statistics/617136/digital-population-worldwide>.
3. Enns, P. K., Lagodny, J. and Schuldt, J. P. 2017. Understanding the 2016 US presidential polls: the importance of hidden Trump supporters. *Stat. Politics Policy* 8, 41–63.
 4. Galesic, M., Bruine de Bruin, W., Dalege, J., Feld, S., Kreuter, F., Olsson, H., Prelec, D., Stein, D. and van der Does, T. 2021. Human social sensing is an untapped resource for computational social science. *Nature* 595, 214–222. <https://doi.org/10.1038/s41586-021-03649-2>.
 5. Jeong H., Bhattarai, R., Adamowski, J. and Yu D. 2020. Insights from socio-hydrological modeling to design sustainable wastewater reuse strategies for agriculture at the watershed scale. *Agricultural Water Management*, Vol. 231, 105983. <https://doi.org/10.1016/j.agwat.2019.105983>.
 6. Kononova, O., He, T., Huo, H., Trewartha, A., Oliverti, E. and Ceder, G. 2021. Opportunities and challenges of text mining in materials research. *Iscience*, 24(3), 102155
 7. Lee, D. and Kwon, H. 2020. Keyword analysis of the mass media's news articles on maker education in South Korea. *International Journal of Technology and Design Education*, 1–21.
 8. Lepping, J. 2018. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery.
 9. Olsson, H., Bruine de Bruin, W., Galesic, M. and Prelec, D. 2021. Election polling is not dead: a Bayesian bootstrap method yields accurate forecasts. Preprint at <https://doi.org/10.31219/osf.io/nqcg5>.
 10. Rydning, J., Reinsel, D. and Gantz, J. 2018. *The digitization of the world from edge to core*. Framingham: International Data Corporation, 16.
 11. Sivapalan, M., Savenije, H. H. and Blöschl, G. 2012. Socio-hydrology: A new science of people and water. *Hydrol. Process*, 26(8), 1270–1276
 12. Tucker, J. et al. 2018. Social media, political polarization, and political disinformation: a review of the scientific literature. SSRN <https://doi.org/10.2139/ssrn.3144139>.
 13. Wang, D., Szymanski, B. K., Abdelzaher, T., Ji, H. and Kaplan, L. 2019. The age of social sensing. *Computer* 52, 36–45.

사사

이 학술기사는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(NRF-2021R1C1C1004179)을 받아 수행된 기초연구 사업입니다. 이에 감사드립니다.