

뉴스 빅데이터 분석을 활용한 가뭄지수 재생산 Reproduction of drought index using news big data analysis

정진홍*, 박동혁**, 안재현***

Jung Jin Hong, Park Dong Hyeok, Ahn Jae Hyun

요 지

가뭄은 강수, 증발산, 대기온도, 토양수분 등 다양한 수문기상학적 인자들이 복합적으로 작용하여 발생되기 때문에 가뭄의 정확한 사상을 분석하는 것은 매우 어렵다. 또한 어떤 요인을 중심으로 고려하느냐에 따라 가뭄은 다양한 시각으로 정의되고 있다. 일정기간 평균 강수량보다 적은 강수로 인해 건조한 날이 지속되는 것, 즉 기상요소를 중심으로 가뭄을 정의하는 것을 기상학적 가뭄이라 하며, 작물의 생육에 필요한 수분을 중심으로 고려하는 것을 농업적 가뭄이라 한다. 또한 하천유량, 댐 저수량 등 전반적인 수자원 공급원의 부족을 수문학적 가뭄이라 한다. 이와 같이 다양하게 나타는 가뭄의 발생특성을 정량적으로 해석하기 위해 다양한 가뭄지수가 개발되어 왔다. 그러나 현재까지 개발된 가뭄지수들은 공통적으로 정형데이터를 활용하여 산정한다. 하지만 최근에는 비정형데이터를 활용하여 지수(Index)를 산정하거나, 재난관리에 적용하는 등 비정형 데이터의 활용이 급증하고 있다. 따라서 본 연구에서는 비정형 데이터(뉴스 데이터)를 활용하여 가뭄지수를 산정하고 기존의 가뭄지수들과의 상관성 분석을 실시 한 뒤, 지수결합을 통해 가뭄사상 분석의 새로운 방안을 제시하고자 하였다.

본 연구의 공간적범위는 2014~2015 충남서북부가뭄 지역 중 가장 큰 피해를 입었던 보령지역으로 선정하였으며 시간적범위는 2013~2016년으로 설정하였다.

비정형 데이터의 구축은 크롤링(Crawling)을 활용하여 네이버 뉴스의 기사를 수집하였으며 자료의 신뢰성을 위해 URL이 동일한 중복기사 및 ‘보령’, ‘가뭄’ 단어가 없는 기사는 제거하였다. 구축된 데이터를 기반으로 월별 빈도를 산출하고 표준점수(Z-score)로 환산하여 가뭄지수를 산정하였다. 산정된 가뭄지수가 어떤 가뭄의 유형(기상학적, 농업적, 수문학적)을 보이는지 확인하기 위해 기존의 가뭄지수들과 상관성분석을 실시하였으며, 가장 높은 상관성을 보이는 가뭄지수와 결함을 통해 새로운 가뭄 사상을 분석하였다. 본 연구에서 진행한 가뭄사상 분석은 향후 가뭄만이 아니라 다양한 재난분야에서 비정형 데이터를 활용한 분석의 기초자료로 활용될 수 있을 것이다.

핵심용어 : 가뭄지수, 비정형 데이터, 빅데이터, 크롤링, 가뭄사상

감사의 글

“이 논문은 행정안전부 재난안전 취약핵심역량 도약기술개발 사업의 지원을 받아 수행된 연구임(2020-MOIS33-006)”

* 정희원 · 서경대학교 대학원 도시기반방재안전공학과 석사과정 · E-mail : wlsghd821@naver.com

** 정희원 · (주)라온티앤씨 R&D센터 책임연구원 · E-mail : smilehyuki@naver.com

*** 정희원 · 서경대학교 토목건축공학과 교수 · E-mail : wrr21@naver.com