

인천지역 AWS 기상데이터를 활용한 강우시계열 자료의 전처리 기법 개발

장동민*, 정창후*, 김원수*, 박성원*, 임중현**, ***
*한국과학기술정보연구원 데이터기반문제해결연구단
**인하대학교 사회인프라공학과
***한국건설기술연구원 건설시험인증본부
e-mail:jhum@kict.re.kr

Application of Pre-processing Method on the Time Series Data of Rainfall Using AWS Weather Data in Incheon City

Dongmin Jang*, Chang-Hoo Jeong*, Wonsu Kim*, Sung Won Park*, Junghyun Eom**, ***
*Dept. of Data-centric Problem Solving Research, Korea Institute of Science and Technology Information
**Dept. of Civil Engineering, Inha University
***Construction Test & Certification Center, Korea Institute of Civil Engineering and Building Technology

요약

기상청에서는 지역별 상세관측자료를 자동기상관측장비(AWS: Automatic Weather System)를 통해서 생산하고 있다. 생산되는 주요자료로는 강수관련자료(강수여부 및 15분, 60분, 6시간, 12시간 누적 강수량)와 함께 기온, 풍향, 풍속, 습도, 해면기압 등이 있으며, 1분 간격으로 취득되기 때문에 다양한 강우분석에 활용이 가능하다. 하지만 이와 같은 AWS의 시계열 취득 자료는 다양한 원인에 따라서 결측값이 발생하게 되며, 대부분 단발성이 아닌 지속적인 결측이 발생하여 후처리 및 다양한 분석 시에 큰 어려움을 야기한다. 따라서 본 연구에서는 인천지역 8개 지점의 최근 8년간(2010년 ~ 2017년) AWS 자료를 취득하여 자료군에서 발생하는 결측치의 발생을 파악하고 데이터 정제하는 전처리 과정을 수행하였다. 데이터 정제 단계는 시간대별로 생성된 자료군에서 인천지역 8개 지점에 대한 자료를 추출하는 과정, 추출한 데이터를 시계열자료로 재배치하는 과정, 그리고 각 항목별로 결측치가 연속적으로 발생된 빈도수의 모니터링으로 구성된다. 이처럼 재구성 및 정제된 데이터군은 강우 시계열 분석을 위한 feature engineering 단계를 거치게 되는데 이때는 전통적인 시계열 분석 시에 활용되는 STL(Seasonal-Trend decomposition using Loess) 기법을 활용하였다. 즉 우리나라의 강우발생패턴이 여름철에 집중호우기간에 대부분 집중되는 특성을 계절성과 추이분석을 통해 분석이 가능하다. 이후 8개 지점에 대한 과거 강우 시계열 분석결과를 바탕으로 향후 단기 강우 예측모형에 대한 적용성을 확보하고자 한다.

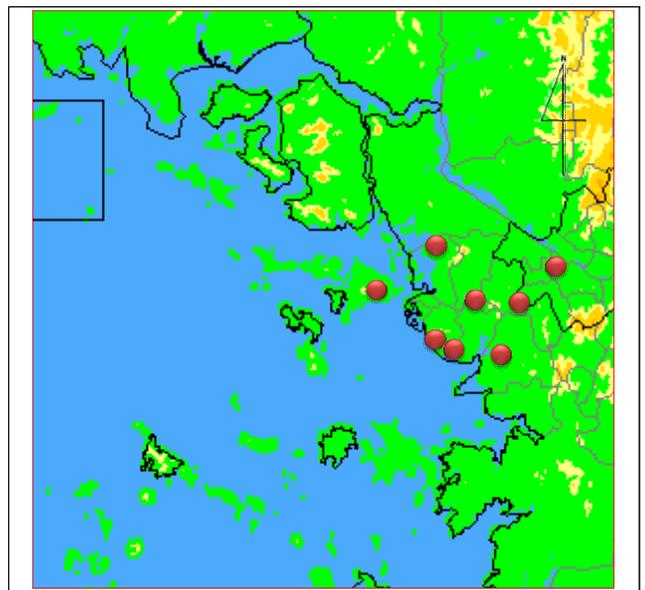
keyword : AWS; 데이터 정제; 강우 시계열분석; STL; 단기 강우 예측모형

1. 서론

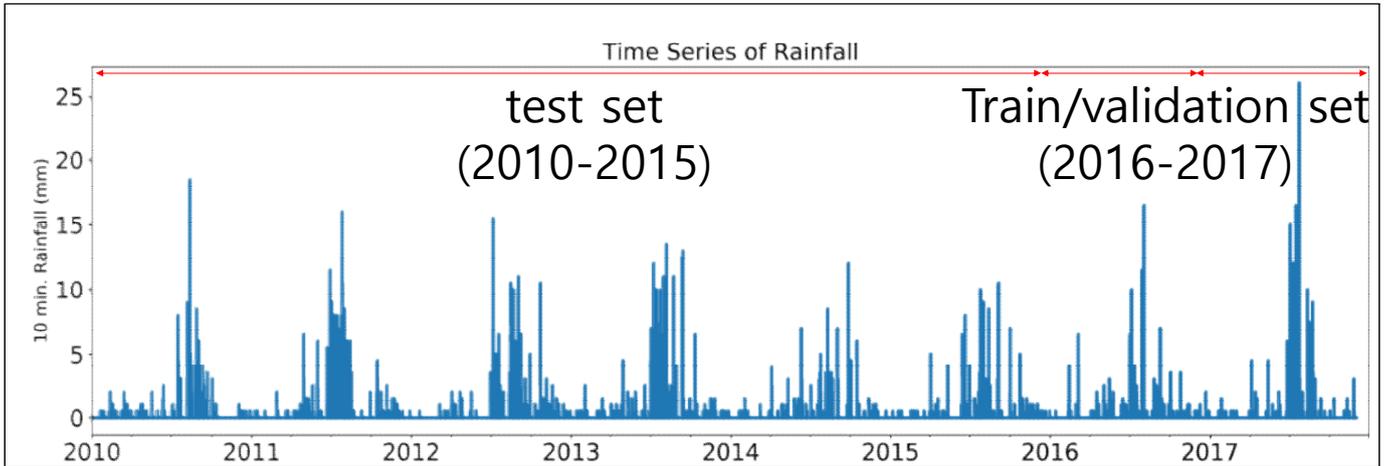
도시지역의 집중호우발생에 대한 관심은 침수피해 발생의 빈도와 강도가 증가함에 따라 더욱 심화되고 있다. 우리나라는 매년 반복적으로 5월 ~ 8월까지 집중호우가 발생하며 이에 대한 내수침수피해가 발생하고 있다. 이에 따라 다양한 강우예측과 이를 기반으로 하는 침수예측기법을 개발 중이나 한계가 있다. 본 연구에서는 전통적인 시계열 분석기법을 활용하여 AWS 자료에 적용하고자 하며 이를 통해 계절성 분석 등의 다양한 기법을 활용하여 예측성능을 확보하고자 한다.

2. 적용대상 및 분석방법

본 연구에서는 인천광역시 내륙지역 위주의 AWS 취득지점 8개소로부터 2010년~2017년까지의 강우자료를 이용하였다.



[그림 1] 적용대상(인천광역시 8개소)



[그림 2] 인천지역 강우자료의 set 구분

AWS는 분당 취득자료이며, 본 연구에서는 인천지역의 강우 특성을 분석하고 단기예측을 위한 계절성 분석을 위해서 강서(404), 공촌동(511), 연수(512), 영종도(543), 시흥(565), 금곡(570), 송도(631), 부평(649)의 8개소로부터 취득된 자료를 이용하였다.

3. 분석결과 및 향후연구

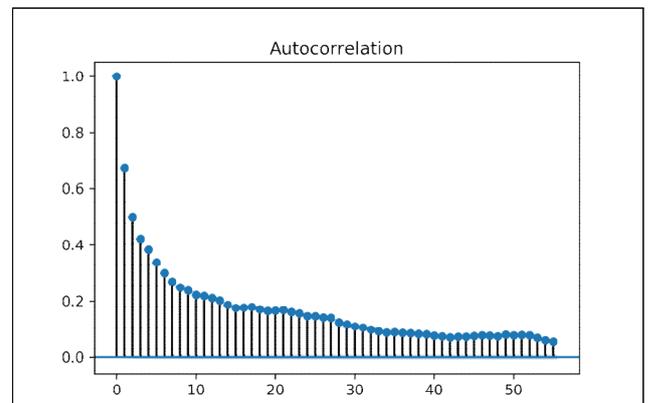
우리나라의 계절적 특성대로 강우발생은 여름철에 집중적으로 발생하였으며 이에 따라 시계열 분석을 실시하였다. 파이선으로 seasonality analysis를 수행하였으며 잔차를 최소화 하는 배열을 적용하였다. 또한 2010년부터 2015년까지의 강우자료를 테스트셋으로 2016년도 자료를 트레인셋으로 구성하였으며 마지막으로 2017년 강우자료를 검증자료로 활용하였다(그림 2). 이와 같은 자료의 분류는 전형적인 러닝기법의 분류방법이며 학습자료를 활용하여 모형의 정확도를 향상시킬 수 있는 가장 적절한 방법이다.

자기상관함수(auto-correlation function)와 부분자기상관함수(partial auto-correlation function)의 적용결과는 그림 3~그림 4에 표시하였으며 이를 통해 모형의 예측정확도를 위한 매개변수 조정을 수행하였다.

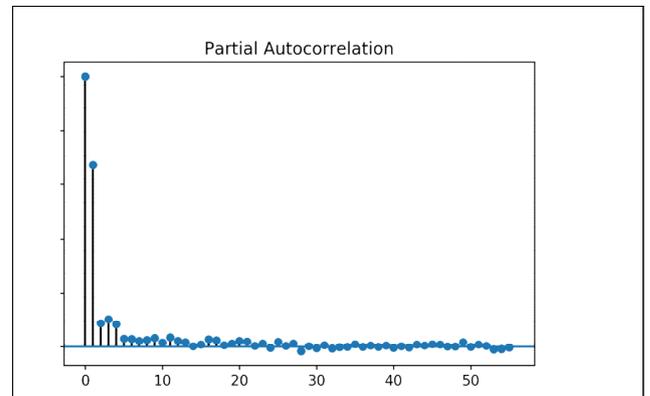
이와 같은 전처리기법을 활용하여 전통적인 시계열 분석방법 중 하나인 ARIMA 모형의 적용성을 확보하였으며 이를 머신러닝기법이나 딥러닝기법의 적용결과와 비교하여 예측 성능개선의 근거를 마련할 수 있을것으로 사료된다.

감사의 글

This research was supported by Ministry of Science, ICT, Republic of Korea (Project No. K-20-L01-C06-S01).



[그림 3] auto-correlation function 적용결과



[그림 4] Partial auto-correlation function 적용결과