

K-means를 활용한 항로표지 센서 데이터 군집화

† 김두환 · 성상하* · 최형림**

† 동아대학교 스마트물류연구센터 선임연구원, *동아대학교 경영정보학과 박사과정, **동아대학교 스마트물류연구센터 교수

요약 : 해양에 설치된 항로표지는 선박의 안전한 항해를 위해 위치 정보를 제공하고, 항로표지에 부착된 센서를 통해 다양한 해양 정보를 수집하고 있다. 하지만 항로표지는 육지와 멀리 떨어진 해상이라는 특수한 작업환경으로 인해 항로표지 유지보수를 위한 많은 시간과 비용이 발생하게 된다. 현재 항로표지에 부착된 센서를 통해 다양한 정보를 수집하고 있지만, 정상 데이터와 비정상 데이터를 구분할 수 있는 정보가 없어 고장진단에 어려움이 있다. 따라서 본 연구에서는 항로표지 센서 고장진단을 위해 머신러닝 비지도학습 중 하나인 K-means 알고리즘을 활용하여 정상 데이터와 비정상 데이터로 군집화하였으며, 분류가 잘 되는 것을 확인할 수 있었다. 향후 연구방향으로는 2개의 클러스터로 구분된 데이터가 실제로 정상 데이터인지, 비정상 데이터인지에 대한 비교·분석이 필요하다.

핵심용어 : 항로표지, 센서, 고장진단, 머신러닝

I. 서론

- ☑ 항로표지 상태 데이터는 항로표지에 부착된 센서를 통해 수집하고 있으며 기압, 시정, 염분, 온·습도, 유향·유속, 파고·파향, 풍향·풍속 등 다양한 해상 정보를 수집하고 있음
- ☑ 항로표지는 육지와 멀리 떨어진 해상이라는 특수한 작업환경으로 인해 항로표지 유지보수를 위해 많은 시간과 비용이 발생함
- ☑ 해양측위정보원에서는 항로표지 센서를 통해 데이터를 수집하고 있으나, 고장유무를 파악할 수 있는 변수는 없음
- ☑ 본 연구에서는 정상 데이터와 비정상 데이터로 구분하기 위해 K-means 알고리즘을 사용하여 수집된 항로표지 센서 데이터를 군집화함

II. K-means를 활용한 항로표지 센서 데이터 군집

- ☑ 항로표지 센서 데이터
 - 해양측위정보원에서는 우리나라 해안의 항로표지에 부착된 센서를 통해 풍향, 풍속, 온도, 습도, 기압, 시정, 염분, 유향, 유속, 수온, 파향, 파고, 전압 등을 수집하고 있음

구분	수집 데이터
기압 센서	500 ~ 1100hPa / 800 ~ 1060hPa / 600 ~ 1060hPa / 900 ~ 1100hPa
시정 센서	강수량, 이슬비, 비/눈, 눈, 안개, 해무, 연무(안개, 황사), 안개 등
염분 센서	전도도, 수온, 염도
온습도 센서	온도, 습도
유향유속 센서	유향, 유속, 수온
파고파향 센서	파고, 파향, 수온
풍향풍속 센서	풍속, 풍향

II. K-means를 활용한 항로표지 센서 데이터 군집

- ☑ K-means 알고리즘
 - 머신러닝 비지도학습 중 하나인 K-means 알고리즘은 데이터를 K개의 군집(Cluster)으로 묶는 알고리즘임
 - 각 군집의 평균(mean)을 활용하여 K개의 군집으로 묶음
 - K는 묶을 군집의 개수
 - means는 평균 : 각 클러스터의 중심과 데이터들의 평균 거리

$$X = C_1 \cup C_2 \dots \cup C_K, \quad C_i \cap C_j = \phi$$

$$\operatorname{argmin}_C \sum_{i=1}^K \sum_{x_j \in C_i} \|x_j - c_i\|^2$$

II. K-means를 활용한 항로표지 센서 데이터 군집

- ☑ 항로표지 센서 데이터 군집 프로세스
 - 해양측위정보원에서 제공하는 항로표지 센서 데이터를 고장과 정상 데이터로 군집하기 위해 다음과 같은 프로세스로 진행함
 - 먼저 항로표지 센서 데이터를 읽어, 데이터 전처리 및 분석을 통해 데이터 속성을 파악함
 - 필요한 일부 데이터를 추출하여 K-means 알고리즘에 적용하여 데이터를 K값을 2로 하여 데이터를 군집함



† 교신저자 : kdhblack@dau.ac.kr

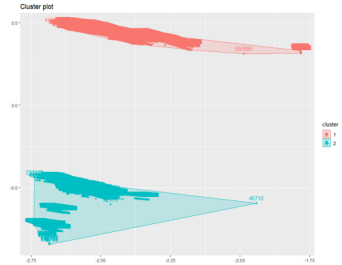
II. K-means를 활용한 항로표지 센서 데이터 군집

- ☑ 항로표지 센서 데이터 분석 및 전처리
 - 해양측위정보원에서 제공하는 항로표지 센서 데이터는 2017~2021년 5년치의 데이터가 있음
 - 관측소별로 수집되는 데이터에 차이가 있으며, 총 62개의 컬럼 중 17~19개의 컬럼은 전혀 수집되고 있지 않음
 - 데이터 수집이 우수한 2021년도 데이터 중에서 데이터 결측치가 가장 적은 C 관측소의 데이터를 추출하여 K-means 알고리즘에 활용함

II. K-means를 활용한 항로표지 센서 데이터 군집

- ☑ K-means를 활용한 항로표지 센서 데이터 군집 결과
 - K-means 알고리즘 적용 결과 2개의 군집으로 분류가 잘 되는 것을 확인할 수 있음

Cluster	1	2
nrow	739,647	461,042



II. K-means를 활용한 항로표지 센서 데이터 군집

- ☑ 항로표지 센서 데이터 분석 및 전처리
 - 데이터가 1개도 없는 컬럼과 분석에 불필요한 컬럼(경도, 위도, 설치일시 등)을 제외하였음
 - 최종적으로 1,200,689행 x 13열의 데이터를 사용함
 - 또한 데이터 값을 똑같은 수준으로 맞추기 위해서 0~1사이로 정규화(normalization) 하여 사용하였음

III. 결론

- ☑ 본 연구에서는 고장과 정상 데이터로 구분하기 위해 K-means 알고리즘을 사용하여 수집된 항로표지 센서 데이터를 군집화함
- ☑ 향후 연구 방향으로는 2개의 클러스터로 구분된 데이터가 실제로 정상 데이터인지, 비정상 데이터인지 확인과 항로표지 전체 데이터에 K-means 알고리즘 적용이 필요함

ACKNOWLEDGMENT

이 논문은 2022년 해양수산부 재원으로 해양수산과학기술진흥원의 지원을 받아 수행된 연구임(스마트항로표지 현장시설 고도화, 20210636)