

# 항로표지 고장예측 서비스를 위한 기계학습 모델 연구

김 환\* · 정수환\*\* · † 임성수

\*,\*\*충남대학교 컴퓨터공학과 석박사통합과정, † 충남대학교 컴퓨터공학과 교수

**요 약** : 다양한 소스에서 수집되고 연동되는 항로표지 상태 데이터에서의 이상탐지는 항로표지의 고장예측에 있어서 중요한 역할을 한다. 이 연구에서는 항로표지 고장예측 서비스를 위해 상태 데이터를 모델링하고 분석할 수 있는 기계학습 모델의 연구 방법을 소개한다.

**핵심용어** : 항로표지, 고장예측, 시계열분석, 이상탐지

## 서론

- 연구개발 필요성
  - 항로표지의 상태 및 주변 정보를 연계하는 빅데이터 모델링 필요
  - 항로표지를 위한 빅데이터 분석 알고리즘 및 서비스 개발 필요
- 연구개발 목적
  - 항로표지 데이터 이해: 데이터 획득, 탐색, 전처리, 특징추출 등
  - 항로표지 고장예측: 항로표지 데이터 변화를 통한 이상징후 탐지
- 진행 상황
  - 항로표지 상태 및 주변 정보의 연관 관계를 고려한 그래프 모델링 기법 연구
  - 시계열 데이터의 이상탐지를 위한 기계학습 모델 연구 및 샘플 데이터 분석

항로표지 고장예측 서비스를 위한 기계학습 모델 연구 Page 01

## 항로표지 상태 데이터

- 항로표지의 센서를 통해 시간에 따라 변화하는 관측값 수집
- 데이터 전처리 과정을 거쳐서 다변수 시계열 데이터로 변환
- 시계열 데이터의 패턴을 학습하고 모니터링하여 이상탐지
  - 기존 패턴과 다른 데이터를 이상치 분류
  - 이상징후 탐지 결과를 고장 예측에 활용

항로표지 고장예측 서비스를 위한 기계학습 모델 연구 Page 02

## 항로표지 상태 데이터 샘플

- 항로표지 상태 데이터의 이상탐지를 위한 샘플 데이터 활용
- 샘플 데이터: 월별 500만~700만 건
  - 2021년 1월: 6,029,648건
  - 2021년 2월: 5,465,276건
  - 2021년 3월: 5,861,851건
- 데이터 속성
  - 모든 속성값이 결측치인 경우 제거 → 62개 중 18개 속성만 사용
  - MMAF\_CODE(해양관측소코드), MMSI(해상이동업무식별번호), 측정 시각, 위치, 센서 on/off 상태, 각종 전류 및 전압 상태 등으로 구성

	2017	2018	2019	2020	2021
1월	X	O	O	O	O
2월	X	O	O	O	O
3월	X	O	O	O	O
4월	X	O	O	O	O
5월	X	O	O	O	O
6월	X	O	O	O	O
7월	X	O	O	O	O
8월	X	O	O	O	O
9월	O	O	O	O	X
10월	O	O	O	O	X
11월	O	O	O	O	X
12월	O	O	O	O	X

항로표지 고장예측 서비스를 위한 기계학습 모델 연구 Page 03

## 항로표지 상태 데이터 샘플

- 메모리 이슈로 인하여 전체 샘플 데이터 중 일부를 활용하여 분석
- 항로표지 각 장비의 전압, 전류 결렬 사용
- 예제:

항로표지 고장예측 서비스를 위한 기계학습 모델 연구 Page 04

† 교신저자 : 정희원, sunsu@cnu.ac.kr  
\* hwan.kim@ocnu.ac.kr  
\*\* integerhwan@gmail.com

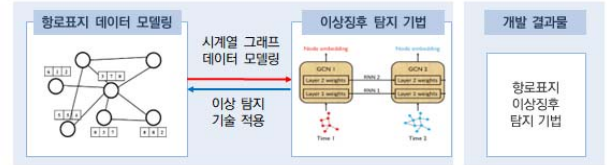
## 기계학습과 문제해결

- 기계학습: 컴퓨터가 데이터를 학습하고 경험을 통해 개선하여 예측 및 분석
  - 문제를 데이터로 표현하고, 기계학습을 통해 의사 결정, 서비스 개발 등에 활용
- 예제: 지도학습
  - 훈련 데이터:  $(x_i, y_i)_{i=1, \dots, n}$  ( $n$ : # 데이터 포인트 수,  $x_i \in \mathbb{R}^p$ ,  $p$ : # 속성 수)
  - 기계학습 모델:  $f_\theta: \mathbb{R}^p \rightarrow \mathbb{R}$  ( $\theta$ : 모수)
  - 손실함수:  $L(f_\theta(x), y) = \sum_i (f_\theta(x_i) - y_i)^2$
  - 최적화:  $\hat{\theta} = \underset{\theta}{\operatorname{argmin}} L(f_\theta(x), y)$
  - 예측:  $\hat{y}_{\text{new}} = f_\theta(x_{\text{new}})$
  - 테스트:  $L(\hat{y}_{\text{new}}, y_{\text{new}})$



## 향후 연구 계획

- 예지보전 알고리즘 개발 및 테스트
- 알고리즘 및 기계학습 성능 고도화
- 추가 항로표지 데이터 및 기상 데이터 활용
- 이상치 탐지 결과 검증 및 분석



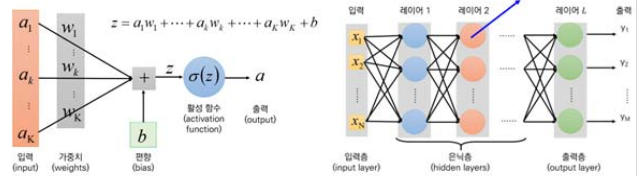
## 표현학습의 중요성

- 최소표현: 표현하고자 하는 카테고리마다 차원을 부여한 벡터 표현
  - 관측 데이터는 고차원인데 반해 실제 학습에 영향을 미치는 차원은 소수
- 분산표현: 의미를 여러 차원에 분산적으로 표현한 벡터
  - 비슷한 카테고리들이 비슷한 벡터 표현을 가지도록 학습 가능
  - 차원 축소 또는 다양체 학습과 연관 (예제: PCA, spectral embedding 등)



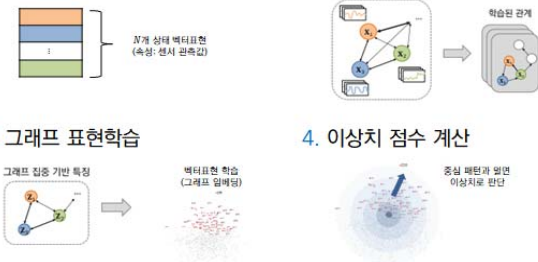
## 딥러닝과 표현학습

- 딥러닝: 인공신경망 모델을 통한 기계학습
  - 모델  $f$ : 입력과 출력 사이에 여러 층으로 연결된 구조, 각 층의 노드는 전파자 역할
  - 모수  $\theta$ :  $W$  &  $b$ , 이전 층에서 다음 층의 노드에 대한 벡터 표현을 연산하는데 사용
  - 학습: 최적의 모수를 학습하기 위한 수치적 최적화 기법 활용



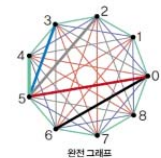
## 이상탐지를 위한 표현학습

- 상태 임베딩:  $k$ 개 상태 벡터 표현 (속성·센서 관측값)
- 그래프 구조학습: 학습된 관계
- 그래프 표현학습: 그래프 집중 기반 특징, 벡터표현 학습 (그래프 임베딩)
- 이상치 점수 계산: 중심 패턴과 벗어난 이상치로 판단



## 이상탐지를 위한 표현학습

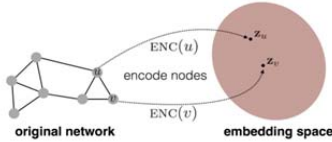
- 상태 임베딩
  - 각 상태에 대해 하나의 벡터표현을 학습하는 것을 목표로 한다.
  - 각 상태에 대한 속성은 특정 시간, 조건에서의 센서 관측값들이다.
- 그래프 구조학습
  - 패턴 학습을 위한 관심 상태들 간의 관계를 그래프로 모델링한다.
  - 관심 상태들 간의 관계를 그래프 구조로 나타낸다.
  - 유사도 기반으로 그래프를 생성할 수 있다.
  - 완전-연결 그래프를 통해 생성할 수 있다.



## 이상탐지를 위한 표현학습

### 3. 그래프 표현학습

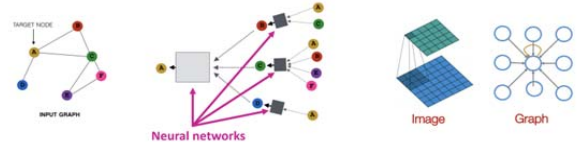
- 각 상태를 하나의 벡터표현으로 임베딩(또는 매핑)한다.
- 좋은 임베딩 혹은 벡터표현이란 기존의 그래프 구조를 복원 가능한 경우이다.
- 벡터 공간 상에서의 표현으로부터 기존 그래프 구조를 잘 복원하는지를 손실함수로 하는 학습을 수행할 수 있다. → 그래프 딥러닝 등을 활용한 표현 학습



## 이상탐지를 위한 표현학습

### 3. 그래프 표현학습

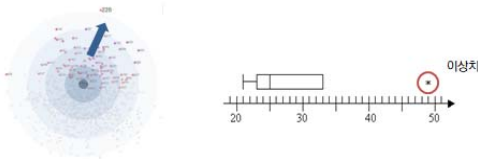
- 예제: 메시지 전달을 활용한 그래프 신경망(GNN)
- $\mathbf{h}_u^{(k+1)} \leftarrow \text{UPDATE}^{(k)}(\mathbf{h}_u^{(k)}, \text{AGGREGATE}^{(k)}(\{\mathbf{h}_v^{(k)}, \forall v \in N(u)\}))$ ,  $\forall u \in V$
- CNN과 유사: CNN은 이미지 패치 기반, GNN은 그래프 인접성 기반 업데이트 활용



## 이상탐지를 위한 표현학습

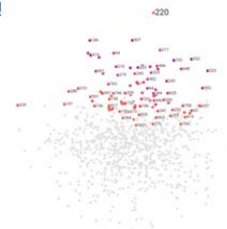
### 4. 이상치 점수 계산

- 유사한 상태 노드들은 유사한 벡터 표현을 가지게 된다.
- 벡터 표현이 중심에서 멀어질수록 이상치 점수를 높게 부여할 수 있다.
- 상태 노드들의 패턴이 여러개인 경우 가우시안 혼합 모델(GMM) 등을 통해 각 분포의 중심에서 멀리 떨어진 정도로 이상치 점수를 높게 부여할 수 있다.



## 실험 결과

- 모델 학습 결과 시각화: 차원 축소 기법을 통한 2차원 시각화
- 예제: 2018년 10월 31일 XXX청 항로표지 데이터
  - 중심에서 멀리 떨어진 포인트 확인 가능: 220번
  - 이상 패턴일 확률 높다고 유추할 수 있음
  - 실제 데이터 확인 결과: AIS 전류값이 전체 훈련 데이터에서 최댓값을 기록하며 크게 벗어난 경우였음



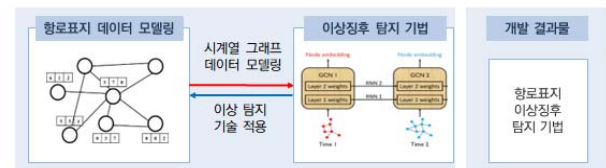
## 실험 결과

- 데이터 전처리 과정에서 정제를 많이 할 경우 패턴이 단순해질 수 있음
- 항로표지의 조건에 따른 패턴이 달라질 수 있음
- 현재까지는 그래프 딥러닝 + 표현학습이라는 기계학습 모델을 통한 이상탐지의 가능성 확인
  - 일관성 있는 모델 학습 방법 고안 필요
  - 더욱 다양한 사례 탐구를 통한 패턴 확인
  - 패턴이 둘 이상인 경우가 존재할지 고려
  - 훈련에 필요한 적절한 샘플 사이즈 결정



## 향후 연구 계획

- 예지보전 알고리즘 개발 및 테스트
- 알고리즘 및 기계학습 성능 고도화
- 추가 항로표지 데이터 및 기상 데이터 활용
- 이상치 탐지 결과 검증 및 분석



## 사 사

이 논문은 2022년 해양수산부 재원으로 해양수산과학기술진흥원의 지원을 받아 수행된 연구임(해양 디지털 항로표지 정보협력시스템 개발(2/5) (20210650))