

항로표지 고장진단 및 예측기술 개발 연구

김 환* · † 임성수

*충남대학교 컴퓨터공학과 석박사통합과정, † 충남대학교 컴퓨터공학과 교수

요 약 : 다양한 소스로부터 수집되고 연동되는 데이터를 모델링하는 기술로 그래프 데이터베이스를 활용한 분석 기법이 각광받고 있다. 이 연구에서는 항로표지에서 관측되는 상태 및 주변 정보를 모델링하고, 고장진단 및 예측에 적용할 수 있는 기계학습 기법을 소개한다.

핵심용어 : 그래프 데이터베이스, 그래프 신경망, 시계열 분석, 패턴 인식, 이상 탐지

서론

- 연구개발 필요성
 - 항로표지의 상태 및 주변 정보를 연계하는 빅데이터 모델링 필요
 - 항로표지를 위한 빅데이터 분석 알고리즘 및 서비스 개발 필요
- 연구개발 목적
 - 항로표지 이상탐지: 상태를 모니터링하여 이상 징후 조기 발견
 - 항로표지 예지보전: 고장 시점을 예측하여 사전 점검 및 교체
- 진행 상황
 - 항로표지 상태 및 주변 정보의 연관 관계를 고려한 그래프 모델링 기법 연구
 - 시계열 그래프 데이터의 이상탐지를 위한 예측 알고리즘 동향 파악 및 개발

항로표지 고장진단 및 예측기술 개발 연구 Page 01

항로표지 센서 데이터

- 항로표지의 센서를 통해 시간에 따라 변화하는 관측값 수집
- 센서들의 관측 시점을 맞춰서 다변수 시계열 데이터로 변환
- 결측치가 발생한 경우 대체 (imputation) 혹은 이상치 분류
- 이상치에 대한 기준이 한정적이며 사전 징후 예측이 필요
→ 시계열 데이터 이상 탐지 기법 연구



항로표지 고장진단 및 예측기술 개발 연구 Page 02

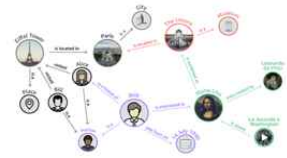
시계열 데이터 이상탐지

- 센서 간의 물리적 혹은 특징에 따른 유사성 및 인과 관계 존재
- 기존의 시계열 데이터 분석은 변수 간 관계를 반영하기 어려움
- 해결 방안: 다변수 시계열 데이터를 그래프 데이터로 모델링
- 최근 여러 연구를 통해 높은 성능 입증 → 본 연구에 적용
 - H. Zhao et al., Multivariate time-series anomaly detection via graph attention network, ICDM 2020
 - A. Deng et al., Graph neural network-based anomaly detection in multivariate time series, AAAI 2021
 - Z. Chen et al., Learning graph structures with transformer for multivariate time series anomaly detection in IoT, IEEE IoT 2021
 - L. Wenkai et al., Stacking VAE with graph neural networks for effective and interpretable time series anomaly detection, Preprint

항로표지 고장진단 및 예측기술 개발 연구 Page 03

그래프 데이터 모델링

- 기존의 관계형 데이터베이스는 행과 열로 이루어진 테이블 형태
- 그래프 데이터베이스는 객체 간의 관계를 직관적인 형태로 표현
- 객체는 노드, 객체 간의 관계는 링크, 노드와 링크의 속성을 포함
- 적용 사례
 - 소셜 네트워크: 추천, 마케팅
 - 단백질 상호작용: 신약 개발
 - 접촉자 네트워크: 질병 확산 예측
 - 범죄자 네트워크: 핵심 인물 추적
 - 상거래 네트워크: 이상 거래 탐지
 - 제어 흐름 네트워크: 코드 품질 탐지
 - ...

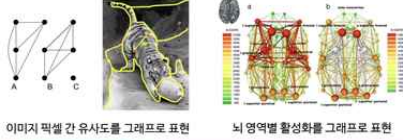


항로표지 고장진단 및 예측기술 개발 연구 Page 04

† 교신저자 : 정희원, sungsu@cnu.ac.kr
* hwan.kim@o.cnu.ac.kr

그래프 데이터 모델링

- 각 센서를 객체로 인식하여 노드로 정의 후 링크 및 속성을 추가
- 의존성(dependency) 모델링: 관측치가 유사한 패턴, 물리적 혹은 특성에 따른 유사도가 존재하는 센서 노드를 링크로 연결
- 센서 간의 관계를 고려한 이상탐지가 가능할 것으로 기대됨
- 알려져 있는 의미적 관계가 없더라도 객체 정의 및 그래프 모델링을 통해 데이터 기반의 새로운 통찰을 얻는 사례들이 많이 있음



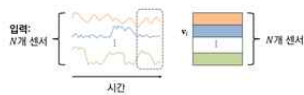
그래프 스트림 데이터

- 기존의 그래프 데이터 분석 기법은 주로 정적 그래프에 적용됨
- 시간에 따라 변화하는 그래프에서 패턴을 인식하기 어려움
- 그래프 스트림: 시간 t 에 따른 그래프 G_t 가 순차적으로 등장
- 최근 그래프 스트림 데이터에 대한 분석의 수요가 증가
- 그래프 신경망 (graph neural network) 적용을 시작으로 그래프 데이터를 활용한 기계학습 모델 개발의 성공 시대가 열림

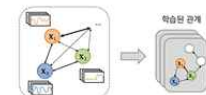


그래프 스트림 이상탐지 기법 소개

1. 센서 임베딩



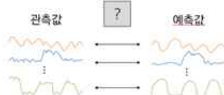
2. 그래프 구조 학습



3. 그래프 신경망 기반 예측

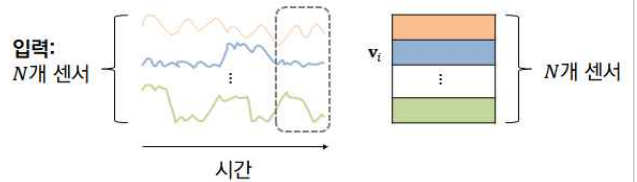


4. 성능 평가



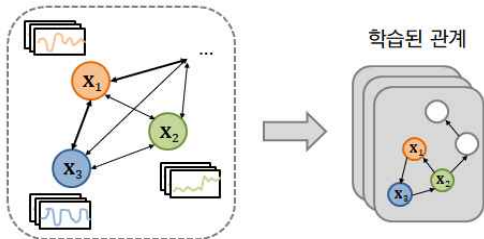
센서 임베딩

- 각 센서의 고유 특성을 반영하는 벡터 표현을 학습한다.
- 센서 $i \in \{1, 2, \dots, N\}$ 에 대해 벡터 표현 $\mathbf{v}_i \in \mathbb{R}^d$ 를 구한다.
- 비슷한 경향을 보이는 센서들은 비슷한 벡터 표현을 가지도록 한다.



그래프 구조 학습

- 센서 임베딩 간의 유사도, 상관계수, 인과 관계 등을 고려하여 센서 간의 연관 관계를 그래프 데이터베이스로 표현한다.



그래프 구조 학습

- 센서 i 와 관련 있는 후보 센서들 C_i 를 식별한다.
 $C_i \subseteq \{1, 2, \dots, N\} \setminus \{i\}$
- 후보 센서 $j \in C_i$ 에 대해, 센서 i 와의 유사도 e_{ji} 를 계산한다.

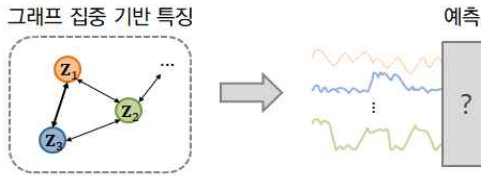
$$e_{ji} = \frac{\mathbf{v}_j^\top \mathbf{v}_i}{\|\mathbf{v}_i\| \cdot \|\mathbf{v}_j\|}$$

- 각 센서와 가장 유사한 k 개의 센서가 연결되는 그래프를 구축하기 위해 인접행렬 \mathbf{A} 를 정의한다.

$$A_{ji} = \mathbb{1}\{j \in \text{TopK}(e_{ki}; k \in C_i)\}$$

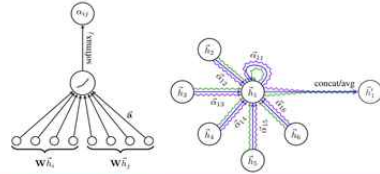
그래프 신경망 기반 예측

- 과거 관측값을 통한 예상 행동과 벗어나는 센서를 식별한다.
- ML 모델은 과거 관측값으로 현재 관측값을 예측한다.
 - Input: $\mathbf{x}^{(t)} := [\mathbf{s}^{(t-w)}, \mathbf{s}^{(t-w+1)}, \dots, \mathbf{s}^{(t-1)}] \in \mathbb{R}^{N \times w}$
 - Output: $\mathbf{s}^{(t)} \in \mathbb{R}^N$



그래프 신경망 기반 예측

- 그래프 집중 신경망 (Graph Attention Network)
 - 노드 i 의 특징 벡터 학습: $\mathbf{z}_i^{(t)} = \text{ReLU}(\alpha_{i,i} \mathbf{W} \mathbf{x}_i^{(t)} + \sum_{j \in N(i)} \alpha_{i,j} \mathbf{W} \mathbf{x}_j^{(t)})$
 - 어텐션 가중치 $\alpha_{i,j}$: 훈련 가능한 가중치
 - 어텐션 가중치를 통해 주변 정보 취합



그래프 신경망 기반 예측

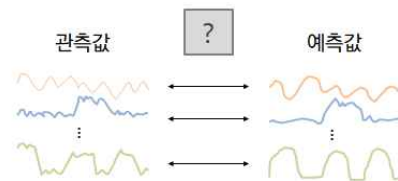
- 완전-연결 (Fully-Connected) 신경망
 - 그래프 집중 기반 임베딩: $\{\mathbf{z}_1^{(t)}, \dots, \mathbf{z}_N^{(t)}\}$
 - 예측값 $\hat{\mathbf{s}}^{(t)}$ 은 센서 임베딩과 그래프 집중 기반 특징 벡터의 element-wise 곱에 대해 신경망 fully-connected layers f_θ 를 거친 결과로 얻어진다.

$$\hat{\mathbf{s}}^{(t)} = f_\theta([\mathbf{v}_1 \circ \mathbf{z}_1^{(t)}, \dots, \mathbf{v}_N \circ \mathbf{z}_N^{(t)}])$$
 - 손실 함수 L_{MSE} 가 최소가 되도록 신경망의 모수를 학습한다.

$$L_{\text{MSE}} = \frac{1}{T_{\text{train}} - w} \sum_{t=w+1}^{T_{\text{train}}} \|\hat{\mathbf{s}}^{(t)} - \mathbf{s}^{(t)}\|_2^2$$

성능 평가

- 정상 및 비정상에 대한 이진 분류 (binary classification)
- Precision (정밀도): $TP / (TP + FP)$
- Recall (재현율): $TP / (TP + FN)$
- F1-score: 정밀도와 재현율의 기하 평균



관련 연구 결과

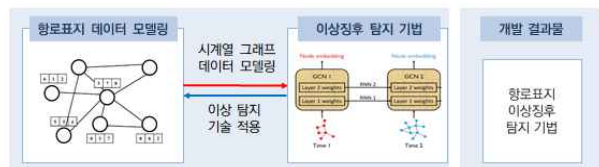
- 비교 ML 기법
 - PCA (주성분 분석), KNN (k-최근접 이웃), Autoencoder (오토인코더)
 - LSTM-VAE: 신경망과 변분적 오토인코더를 결합하여 복원 오차 계산
 - MAD-GAN: 정상 데이터에 대한 GAN 학습 후 이상 점수 계산
 - 복원 오차 또는 이상 점수에 따라 정상 또는 비정상으로 분류

- 시계열 데이터셋
 - WADI: 수처리 공정 데이터셋, 정상 및 비정상 행위 기록
 - #Features: 112
 - #Train: ~105K
 - #Test: ~17K
 - 이상치: ~6%

	정밀도	재현율	F1 점수
PCA	0.40	0.06	0.10
KNN	0.08	0.08	0.08
AE	0.34	0.34	0.34
LSTM-VAE	0.88	0.14	0.25
MAD-GAN	0.41	0.34	0.37
GAT 기반	0.98	0.40	0.57

결론

- 항로표지 고장진단 및 예측을 위한 데이터 모델링 기법 연구
- 시계열 그래프 데이터에서의 이상징후 탐지 기법 동향 조사
- 향후 목표: 항로표지 데이터 이상징후 탐지, 예측 기법 개발



이 논문은 2021년 해양수산부 재원으로 해양수산과학기술진흥원의 지원을 받아 수행된 연구임(해양 디지털 항로표지 정보협력시스템 개발(1/5) (20210650))