

Threshold 자동계산 코드추가

목표

- 현재 모든 모델이 학습 후 default 값(1)으로 설정됨
- 학습 완료 후 자동으로 Threshold 를 구하는 기능을 추가하고자 함

수정사항 및 차이점 확인

Model Name	수정 파일경로	데이터 경로	데이터 분리	test 데이터 적용
ANNOY	src\api_server\apis\multiproc\proc_annoy.py	InfluxDB	앞의 5 일 데이터를 Test 데이터로 사용	Test_data 에 window_size 를 적용시켜서 사용(vectorize)
VAE	src\api_server\apis\multiproc\proc_vae.py	InfluxDB	앞의 5 일 데이터를 Test 데이터로 사용	Test_data 그대로 사용
ICA	src\api_server\apis\multiproc\proc_ica.py	InfluxDB	앞의 5 일 데이터를 Test 데이터로 사용	Test_data 그대로 사용
AAKR	src\api_server\apis\routes\aaokr.py	request.data	전체데이터를 8:2 비율로 나눠서 사용	Test_data 그대로 사용

기존 방식 확인, 변경 및 추가사항

- ANNOY 기존 방식, 변경사항

Model_training ([api_server/apis/multiproc/proc/annoy.py](#) 파일)

influx 에서 데이터 조회 -> 전체 데이터 concat 후 scaler fit -> 각 pkl 파일 scaler 적용
-> 조정된 데이터로 모델 학습 -> 모델 및 스케일러 저장

influx 에서 데이터 조회 -> train_test 데이터 split -> train 데이터로만 scaler fit ->

-> train, test 데이터 scaler 적용 -> train 데이터로 모델 학습 -> 모델 및 스케일러 저장

-> 학습된 모델로 test 진행 -> test_data 크기 조정 -> threshold 계산 -> threshold csv 파일로 저장

- VAE 기존 방식, 변경사항

Model_training ([api_server/apis/multiproc/proc/vae.py 파일](#))

influx 에서 데이터 조회 -> 전체 데이터 concat 후 scaler fit -> 각.pkl 파일 scaler 적용
-> 조정된 데이터로 모델 학습 -> 모델 및 스케일러 저장

influx 에서 데이터 조회 -> train_test 데이터 split -> train 데이터로만 scaler fit ->
-> train, test 데이터 scaler 적용 -> train 데이터로 모델 학습 -> 모델 및 스케일러 저장
-> 학습된 모델로 test 진행 -> threshold 계산 -> threshold csv 파일로 저장

- ICA 기존 방식, 변경사항

Model_training ([api_server/apis/multiproc/proc/ica.py 파일](#))

influx 에서 데이터 조회 -> 전체 데이터 concat 후 scaler fit -> 각.pkl 파일 scaler 적용
-> 조정된 데이터로 모델 학습 -> 모델 및 스케일러 저장

influx 에서 데이터 조회 -> train_test 데이터 split -> train 데이터로만 scaler fit ->
-> train, test 데이터 scaler 적용 -> train 데이터로 모델 학습 -> 모델 및 스케일러 저장
-> 학습된 모델로 test 진행 -> threshold 계산 -> threshold csv 파일로 저장

- AAKR 기존 방식, 변경사항

Model_training ([api_server/apis/multiproc/proc/aakr.py 파일](#))

Request 에서 데이터 확보 -> 모델저장

Request 에서 데이터 확보 -> train_test 데이터 split -> train 데이터로 모델저장
-> 학습된 모델로 test 진행 -> threshold 계산 -> threshold csv 파일로 저장

- 추가사항

threshold 계산함수가 들어있는 .py 생성 ([src\tag_setting\threshold_calculator.py](#))

기존방식 (잔차 평균 + 6 x 잔차 표준편차), 잔차의 min 값, 잔차의 max 값,
raw_data 와의 이동평균 잔차값, 잔차 상위 95%값의 5 가지 방식을 사용함

추후 계획

- threshold 계산 방법 수정
- 현재는 csv 파일로 결과가 저장되지만 추후에는 model 객체에 threshold 를 추가하여 화면에서 선택할 수 있는 기능 추가 예정