

3 주차 업무 보고서

작성자

장 두 혁

작성일자

2021.07.15

금일 업무 내용

[MIMIC3 데이터 활용한 수혈량 예측연구]

- 1) 이성빈 T 와 MIMIC 데이터 엑셀파일 마다 수혈량 feature 에 맞는 예측모델 디자인하기.
 - 환자아이디 필터링 1 : 진단정보로 필터링하기(DIAGNOSES_ICD, CPT, DRGCODE)
 - 각 코드마다 분류 기준이 달라서 서로 호환하기가 좋다.
 - 환자아이디 필터링 2 : 처치(수혈량)정보로 필터링하기(INPUTEVENT_MV, INPUTEVENT_CV, PROCEDUREEVENTS_MV, PROCEDURE_ICD)
 - PROCEDURE ICD 는 수혈 이진분류 라벨링만됨
 - INPUTEVENTS_CV, INPUTEVENTS_MV 는 수혈량 기록됨.
 - 필터링 1, 2 통해서 4 그룹나눠 :
 - (수술 O, 수혈 O), (수술 O, 수혈 X), : 수술과 pipeline
 - (수술 X, 수혈 O), (수술 X, 수혈 X), : 내과 pipeline

>> 방향 변경 surgery 항목에 대해서 datetime 이 없어서, 수술을 필터링 기준보단 하나의 feature 로 넣기로 결정
- 2) 파이썬 코드로 csv 파일 load 및 통합작업 실시 – csv 파일크기를 고려하지 못해서, 연구실 PC 의 작업의 어려움. 대학원 연구실 램 64GB 로 진행 중

추후 할 일

1. 모든 환자에 대해서 time series 기준을 어떻게 잡을 것인가 고려 중
2. 데이터 통합을 위해 key 의 기준을 각 데이터에서 어떻게 선별하지 고려
3. 대학원 컴퓨터 진행하다가 안될 시에 파이썬보다 MySQL 로 convert CSV to DB 로 진행하여 쿼리문을 통해 Feature Selection 진행해야함