**NIA 프로젝트 주간보고서(9월 1주차)**

|  |  |  |  |
| --- | --- | --- | --- |
| **팀원** | 김경수, 이세림, 이지상 | | |
| **작성자** | 이지상 | **작성일** | 2021.09.15(수) |

1. **작성자**
2. **주요 진행사항**

|  |  |
| --- | --- |
| **이슈사항** | * **출력 방식 변경의 문제** * 현재 코드 내에 정의된 predict() 함수나 sklearn에서 제공되는 roc\_auc\_score() 함수를 통해 AUROC, AUPR을 출력으로 사용하고 있음 * 다른 성능 지표를 이용할 수 있는지 확인 * AUROC와 AUPR을 출력하기 위해 사용하는 y\_true와 y\_hat은 각각 이상행동에 대한 해답과 평가를 통해 출력된 결과로 y\_hat을 이용하여 이상행동을 분류하려 함 * 하지만 y\_hat의 값은 모두 리스트의 형태이며, 각 프레임마다 출력되는 값이 1보다 작은 값이기에 이해에 어려움이 있음 * accuracy 출력을 위해 함수 제작과 코드 변경 중에 있음 |

1. **세부 진행사항**

|  |  |
| --- | --- |
| **요 일** | **업무내용** |
| **9월 10일(금)** | * **출력 방식 변경을 위한 회의(1)**      * + roc\_auc\_score 함수에서 이용되는 변수를 조사한 결과 y\_true는 기존에 주어지는 정답값, y\_hat은 예측 값으로 추정   + 각 값은 리스트의 형태로 존재하며 입력값인 csv파일이 리스트의 형태로 변환되기 때문에 출력도 리스트로 됨 |
| **9월 11일(토)** | * **출력 방식 변경을 위한 회의(2)**   + 기존에 사용하고 있는 roc\_auc\_score 함수를 이용한 성능 지표 AUROC, AUPR 말고 활용할 수 있는 다른 성능지표를 조사한 결과 sklearn.metrics에서 지원하는 정확도를 표현하는 acuuracy\_score 함수와 f1\_score 함수를 이용하여 보았으나 해당 오류발생      * 아마 기존에 사용하던 변수 y\_true와 y\_hat의 타입이 함수가 요구하는 타입과 맞지 않아 발생한 것으로 추정 |
| **9월 12일(일)** | * **출력 방식 변경을 위한 회의(3)**   + 원하는 결과 출력 방식과 맞지 않아 계속해서 tbad/combined\_model/evaluate.py파일을 살피고 있음   + AUROC와 AUPR은 y\_true와 y\_hat을 가지고 roc\_auc\_score()함수를 사용해 값을 얻는데, 이때 사용되는 y\_true와 y\_hat은 리스트, AUROC와 AUPR은 실수(또는 지수)로 되어있음   + AUROC와 AUPR이 생성되는 과정은 이해가 가능하나 현재 y\_hat이라는 리스트 형태의 값을 어떻게 loss와 acc로 표현해야 할 지 구상중   + 가장 간단한 방법은 model.evaluate()을 사용하는 것이나 모델 구성을 위해 만들어둔 fusion.py에 model()함수를 구현했지만 model.evaluate()함수에는 출력값도 있어야 해서 사용에 어려움이 있음 |
| **9월 13일(월)** | * **데이터셋 평가 결과에 대한 논의(1)**   + AI-Hub 데이터셋의 일부를 학습된 모델을 가지고 predict()을 했을 때 나오는 값은 아래와 같음      * + 1이 넘는 값도 있지만 극히 일부에 달하며 대부분 0.5 미만의 값을 가지고 있음      * + train함수에는 입력, 출력에 대한 정보가 들어가는데, 없을 경우 None으로 대체됨   + 그러기 때문에 들어가는 정보에 대해선 학습이 가능하나 그에 대한 출력 정보에 대한 학습이 없기 때문에 학습에 부족함을 느낀다 판단   + 다른 딥러닝 학습 자료를 가지고 결과값을 만들어볼 계획도 현재 생각 중 |
| **9월 14일(화)** | * **데이터셋 평가 결과에 대한 논의(2)**   + AI-Hub 데이터셋의 일부를 학습된 모델을 가지고 predict()을 했을 때 나오는 값을 해석하기 위해 모델에서 기존에 사용된 HR-Avenue와 HR-ShanghaiTech 데이터로 해보았으나 같은 형식의 데이터가 출력됨   + 해당 데이터를 해석하기 위해 코드에서 y\_hat의 생성 과정을 역순으로 해석할 필요가 있다고 판단 |

1. **기타 보고사항**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **차주  진행계획** | * 9월 3주차 진행계획  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **구분** | **9월 20일** | **9월 21일** | **9월 22일** | **9월 23일** | **9월 24일** | | **연구**  **내용** | 출력 결과  확인 및  출력 방법 수정 (1) | 출력 결과  확인 및  출력 방법 수정 (2) | 추석 | 추석 | 추석 | | **산출물** |  | Epoch 별  acc, loss |  |  |  | |
| **특이사항** | * **맞지 않는 형식**   + sklearn.metrics에서 기존에 roc\_auc\_score()함수와 average\_precision\_score()함수를 사용했는데, 같은 모듈에 존재함에도 불구하고 accuracy\_score의 경우 형식이 맞지 않아 평가가 중지됨 * **평가 지표 수정의 어려움**   + 평가에 대한 정보를 출력하는 기존 함수의 개수가 많으며 각 함수마다 고유의 형태를 지닌 것이 아닌 함수 간의 연결성으로 수정의 어려움을 겪고 있음 |