**NIA 프로젝트 주간보고서(8월 5주차)**

|  |  |  |  |
| --- | --- | --- | --- |
| **팀원** | 김경수, 이세림, 이지상 | | |
| **작성자** | 김경수 | **작성일** | 2021.08.30(월) |

1. **작성자**
2. **주요 진행사항**

|  |  |
| --- | --- |
| **이슈사항** | * **AI-Hub 데이터셋 평가 오류 임시 해결** * 완전히 해결된 것은 아니지만 일부 데이터셋의 평가가 이루어짐 * 초반에 사용하던 데이터셋이 저장된 폴더의 구조와 동일하게 하면서 평가가 가능해짐 * **AI-Hub 데이터셋 일부 파일의 오류** * 이전 발표에서 언급했던 오류가 나는 파일에 대한 오류는 아직까지 확인중에 있으며 관련된 오류에 대한 정보를 찾으며 해결 중에 있음 |

1. **세부 진행사항**

|  |  |
| --- | --- |
| **요 일** | **업무내용** |
| **8월 25일(수)** | * **AI-Hub 데이터셋 평가 오류 해결을 위한 회의(3)**   + AI-Hub 뿐 아니라 기존에 제공되었던 ShanghaiTech 데이터셋에서도 동일한 오류가 나타나 기존에 의심했던 csv파일에는 오류가 없던 것으로 판단   + Train.py 의 input 형식과 evaluate.py의 input 형식을 맞추기 위한 코드 수정 |
| **8월 26일(목)** | * **AI-Hub 데이터셋 평가 오류 해결을 위한 회의(4)**   + 기존에 제공된 ShanghaiTech, Avenue 데이터셋의 폴더 구조와 파일명을 동일하게 바꾸었더니 평가 가능   + 이전에 발표에서 언급했던 오류가 나는 파일을 제외한 일부 파일은 평가에 성공   (1~4번 항목 18개 영상으로 학습, 4개 영상으로 평가)   * + 학습이 되지 않는 파일에 대해서 추가적인 오류 해결을 위한 방안 탐색 |
| **8월 27일(금)** | * **AI-Hub 데이터셋 평가**   + 학습에 대한 오류로는 아래와 같음  |  | | --- | | ValueError: Input contains infinity or a value too large for dtype('float32'). |  * + 위 오류가 나는 원인은 공통적으로 xs, ys = xs / width, ys / height 라는 구문에서 볼 수 있으며, 0 또는 NaN으로 나눗셈을 시도했을 때 뜨는 오류임   + np.seterr(divide=’ignore’, invalid=’ignore’)을 시도했고 동일한 오류 발생   + 나눗셈을 통해 무한한 값이 존재하게 되기 때문에 이는 코드상에서 무한한 값이 나오는 원인을 찾아 코드를 수정해야 할 것으로 보임      * **결과 출력 방식 변경을 위한 코드 분석**   평가를 위한 출력 방식 변경을 위해 evaluate.py, eval.py 등 평가와 관련된 코드를 살펴봤으나 함수를 사용하는 부분이 많고, 동일한 함수가 아닌 미리 짜여 있는 이름만 비슷한 다른 함수를 이용해 코드 수정에 어려움이 많음 |
| **8월 28일(토)** | * **AI-Hub 데이터셋 평가**   + 오류가 나지 않은 파일을 ShanghaiTech와 Avenue 데이터셋의 폴더 구조와 동일하게 만들면 학습과 평가가 가능   + 이는 MPED-RNN 모델을 만들 때 작성한 코드에서 경로의 이름을 나눌 때 동일한 구조로 이루어지지 않으면 생기는 오류로 판단   + 평가 가능한 파일 51개를 동일한 구조로 만들어 39:12 비율로 나눔 * 이상행동 종류 중 평가 가능한 파일이 하나인 경우는 바로 평가로 넣음 * 해당 파일은 평가가 제대로 되지 않는 것이 정상적     <좌 3주 전 평가 결과, 우 이번 학습 평가 결과>   * + 이전 결과에 비해 전체적인 결과가 떨어지는 것을 알 수 있으며 AUROC보다 좀 더 까다로운 측정치인 AUPR이 확연히 떨어졌음을 알 수 있음   + 학습 데이터의 부족(이상행동에 관련된 학습은 하지 않은 채 평가로 바로 넘어간 항목 등)으로 정확도가 떨어지는 것의 문제가 큼 |
| **8월 29일(일)** | * **AI-Hub 데이터셋 평가 및 개선**   + 평가 결과가 좋지 않아 개선하기 위해서 여러가지 방식으로 학습, 평가를 시도   + 기존 epoch 5에서 20, 100으로 다양하게 시도         <순서대로 epoch 5, 20, 100>   * + 결과에서 큰 차이가 보이지는 않았음 |

1. **기타 보고사항**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **차주  진행계획** | * 9월 1주차 진행계획  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **구분** | **8월 30일** | **8월 31일** | **9월 1일** | **9월 2일** | **9월 3일** | | **연구**  **내용** | 일부 데이터셋 학습 오류 해결을 위한 회의(1) | 일부 데이터셋 학습 오류 해결을 위한 회의(2) | 평가 결과 출력 방식 변경을 위한 코드 분석 | 평가 결과 출력 방식 변경을 위한 코드 수정작업(1) | 평가 결과 출력 방식 변경을 위한 코드 수정작업(2) | | **산출물** |  | 학습 오류에 관련된 코드 수정 및 깃허브 커밋 | 수정이 필요할 것 같은 코드 확보 |  | 주간보고서 작성 | |
| **특이사항** | * **일부 파일의 학습 오류**   + 코드 중간 나눗셈 과정에서 INF값이 나와 중간에 학습이 중지되는 현상 발견   + 빠른 방법은 csv파일(또는 json파일)을 수정해보는 것이나 오류가 나는 파일의 양이 상당함 * **평가 지표 수정의 어려움**   + 평가에 대한 정보를 출력하는 기존 함수의 개수가 많으며 각 함수마다 고유의 형태를 지닌 것이 아닌 함수간의 연결성으로 수정의 어려움을 겪고 있음 |