

캡스톤 18 조 회의록

프로젝트명	O24Sec	팀 명	앨러리를 찾아서
회의 시간	4 월 30 일 17:00 - 19:00		
장소	국민대학교 미래관 7 층 K 랩		
참석자	김민송, 장우혁	서기	김민송
안건	<ul style="list-style-type: none"> • 현재 진행상황 • 기존 클러스터링 실험 데이터 분석 • 유사도 실험 		
회의 내용	<p>👉 현재 진행 상황</p> <p>1. 유사도 실험</p> <ul style="list-style-type: none"> - 일단 특수문자, 공백 을 제거하여 Word Dictionary 를 만든 상태 - 해당 Dictionary 로 각 패킷 별 tf-idf 벡터까지 만들어 놓았고 이를 통해 클러스터링이나 유사도 평가를 진행 - 클러스터링 부분에 대해서는 하이퍼파라미터를 회사에 어떻게 전해줄지 어려운 부분이 있기 때문에 알고리즘을 통한 유사도를 사용해서 결과를 도출하는 것을 우선적으로 시행함 		
	<p>👉 기존 클러스터링 실험 데이터 분석</p> <p>클러스터링 실험 결과를 DBSCAN 으로 전체적 이미지만 보고 상세 분석이 안되어 있음</p> <p>1. 클러스터링 결과로 만들어 놓았던 CSV 파일 결과 상세분석</p> <ul style="list-style-type: none"> - 비슷한 이벤트들끼리 묶였는지, 탐지명이 같은 이벤트들이 잘 뭉쳐 있는지 <p>2. Byte AE-Chunking 파일 분석</p> <ul style="list-style-type: none"> - 기존에 Byte 로 AE-Chunking 했을 때 Word Dictionary 가 현재 특수문자나 공백을 통한 토큰화보다 단어 수가 적었는데 그 단어들이 더 평가에 효율적일지 판단 <p>3. 하이퍼파라미터 일반화</p> <ul style="list-style-type: none"> - 클러스터링 결과 분석 중 하이퍼파라미터 관련해서 좀 더 일반화 시켜서 내용을 전달할 수 있을지 분석 		
	<p>👉 유사도 실험</p> <p>1. 먼저 Cos 유사도를 기반으로 랜덤 추출 방식 유사도 군집화 진행</p> <ul style="list-style-type: none"> - 객체 나뉜 파일 받은 후 진행 <p>2. 샘플링</p> <ul style="list-style-type: none"> - 한번 실험할 때 시간이 좀 소요되기 때문에 샘플링을 통해 결과를 먼저 도출 		

	<p>3. 효율성</p> <ul style="list-style-type: none"> - 시간 소요 부분이나 메모리 사용량에 있어서 최소화할 수 있는 부분 있는지 찾아보기
프로젝트 진행사항	<p>The Gantt chart displays the following tasks and their durations:</p> <ul style="list-style-type: none"> 인공지능 스터디: Dec 3 - Dec 31 IPS 데이터 분석 및 전처리: Jan 7 - Jan 28 서버 - 클라이언트 객체 분리: Jan 14 - Feb 7 페이로드 암호화 식별: Jan 21 - Feb 14 내부 - 외부 객체 분리: Feb 21 - Mar 7 객체별 유사도 실험: Mar 14 - Mar 28 객체별 클러스터링 기법 실험: Mar 21 - Apr 4 성능평가 및 코드 리팩토링: Mar 28 - Apr 18
서기의 의견	<p>클러스터링 관련해서는 실험 결과가 어느 정도 나와있는 상황이고 유사도 알고리즘에 대해서도 코드는 어느정도 다 짜여진 상황이기 때문에 이후로는 계속 실험을 통해서 최적 결과가 나오는 것을 찾는게 주요한 부분이다 각자 맡은 부분에 있어서 잘하고 있어 협업이 잘되고 있기 때문에 빠른 시간에 마무리 하고 다음 작업으로 넘어 갈 수 있을 것 같다.</p>