유튜브 비속어 필터링

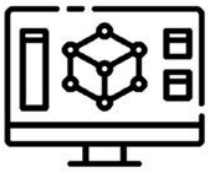
~~빠~~--

10기 이민주 11기 심은선 12기 김주호 12기 이승현

YouTube

Index

투빅스 제 9회 컨퍼런스
삐--



주제 선정 배경 데이터 모델 결과

주제 선정 배경

YouTube 주제 선정 배경

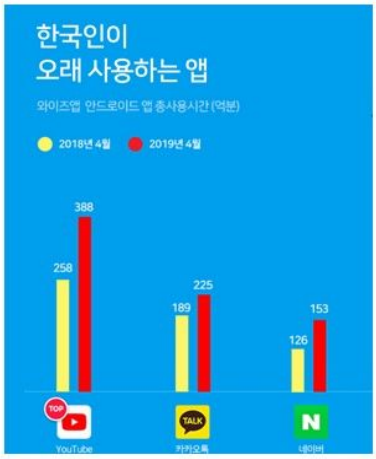
- Youtube 영향력과 문제점
  - 유튜브 사용시간이 증가하지만,
    비속어에 대한 규정사항이 부족해 아동이 그대로 배울 위험

최신기사

"유튜브, 국내 전 연령대에서 가장 오래 사용"

꼬마펭귄이 나와 "XX새끼" 4살배기 울린 '유튜브 키즈'

"다 같이 욕하고 따라해요" 저질 인터넷 방송에 물든 초등학생들



- 한국어 음성 기반 비속어 필터링의 필요성

1 한글 자소 정렬 기법을 이용한 적응적 비속어 비속어 필터링 시스템
윤태진 | 부산대학교 | 2011 | 국내석사
원문보기 목차검색조회

2 이미지 학습 기반 텍스트 필터링 개선 연구 :실시간 비속어 탐지 기법을 위한 사전 연구
유주연 | 성균관대학교 일반대학원 | 2019 | 국내석사
원문보기 목차검색조회

RISS 인기논문
3 게임 채팅 경험 개선을 위한 비속어 우회 이모티콘 개발
고원준 | 홍익대학교 대학원 | 2014 | 국내석사
원문보기 목차검색조회

4 변형 한글 규칙어 실시간 필터링 제재 시스템 개선 연구
김찬우 | 인천대학교 일반대학원 | 2019 | 국내석사
원문보기 목차검색조회

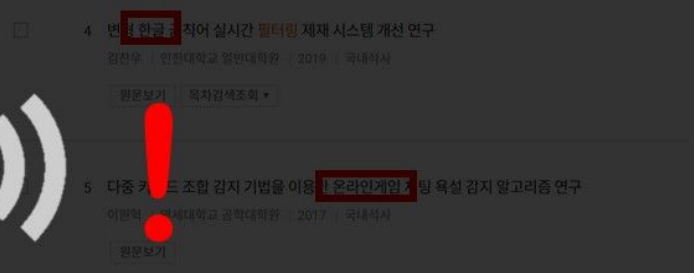
5 다중 키워드 조합 감지 기법을 이용한 온라인게임 채팅 욕설 감지 알고리즘 연구
이원혁 | 연세대학교 공학대학원 | 2017 | 국내석사
원문보기

한국어 비속어 필터링은
텍스트 기반 선행 연구가 많지만
음성 기반 연구가 없음

YouTube 주제 선정 배경

투빅스 제 9회 컨퍼런스
삐--

- 한국어 음성 기반 비속어 필터링의 필요성



한국어 비속어 필터링은
텍스트 기반 선행 연구가 많지만
음성 기반 연구가 없음

YouTube

투빅스 제 9회 컨퍼런스
삐--

데이터

YouTube 데이터 - 수집 투빅스 제 9회 컨퍼런스 뻬--

유튜브 영상 크롤링 문장 단위 분할 단어 단위 음성 분할

YouTube 데이터 - 수집 투빅스 제 9회 컨퍼런스 뻐--

• 문장 단위 분할

영상 자막의 문장 별 시간을 이용해 음성 분할 (28,000개 +)

단어 단위 labeling에 필요한 비속어 음성의 위치를 쉽게 알 수 있음

※ Tool: VoyagerX - Vrew

YouTube

데이터 - 수집

투빅스 제 9회 컨퍼런스
뻬--

• 유튜브 영상 크롤링

먹방, 토크 방송 등 장르에서 비속어가 많은 영상 (700개 +)

※ Library: Pytube

YouTube 데이터 - 수집

투빅스 제 9회 컨퍼런스
삐--

원본 음성 속 sparse한 비속어

28,000+ 문장 중 비속어가 있는 문장 1,200+

• 문장 단위 분할

영상 자막의 문장 별 시간을 이용해 음성 분할 (28,000개 +)

단어 단위 labeling에 필요한 비속어 음성의 위치를 쉽게 알 수 있음

※ Tool: VoyagerX - Vrew

→ 학습이 용이한 데이터 생성 필요

YouTube 데이터 - 수집

투빅스 제 9회 컨퍼런스
뻐--

• 단어 단위 음성 분할

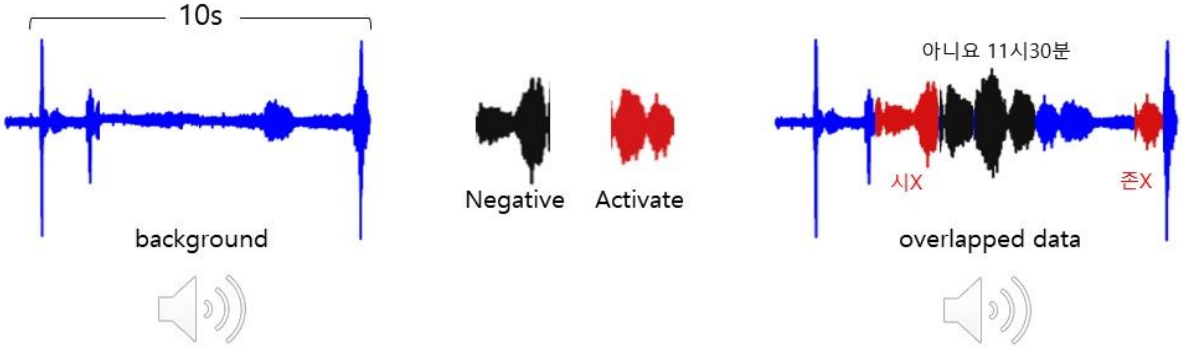
Activate : 시X, 존X, 새X, 병X 등 필터링 대상 비속어 - 1초 단위 (250개 +)

Negative : 감사합니다, 야, 별풍선, 양각 등 일상어 - 1초 단위 (700개 +)

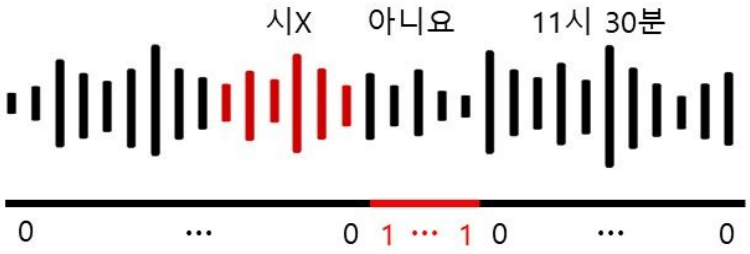
Background : 음악, 먹방 소리 / 백색소음 등 배경음 - 5초이상 (500개 +)

※ Tool: Audacity

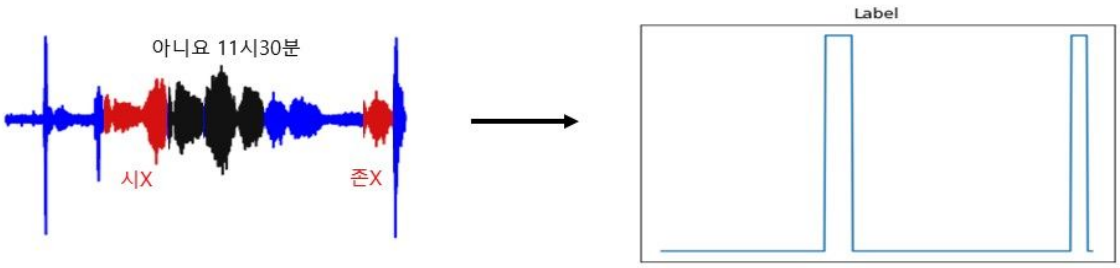
- Features
  - 10초 background에 activate와 negative 랜덤으로 삽입
    ※ 0~3개 랜덤 activates, 0~5개 랜덤 negatives, -2~5 dB 변화
  - 동일한 길이의 dataset 생성 & Imbalanced data 문제 해결



- Target labels
  - Activate = 1 & Background, Negative = 0
  - Imbalanced label 문제 해결: 전체 음성의 길이와 비속어 음성 길이 비례하게 1 label 설정
  - 비속어를 모두 듣고 판단: 비속어가 끝나는 시점에 1 부여

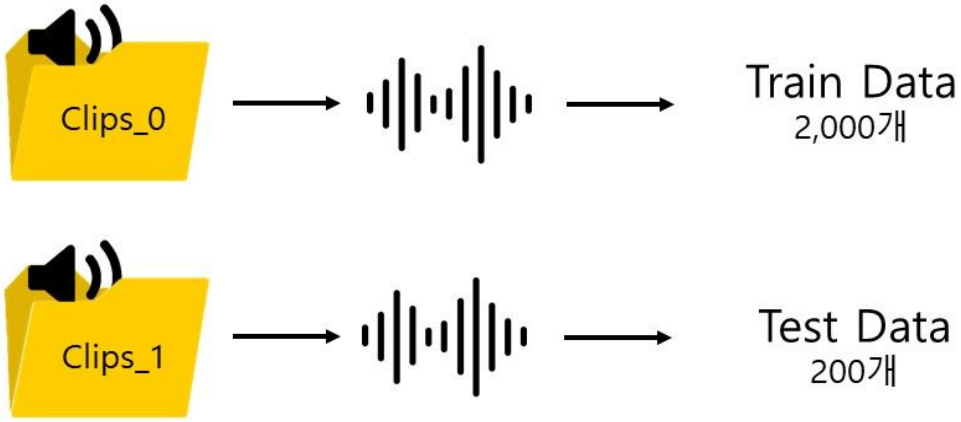


- Target labels
  - Activate = 1 & Background, Negative = 0
  - Imbalanced label 문제 해결: 전체 음성의 길이와 비속어 음성 길이 비례하게 1 label 설정
  - 비속어를 모두 듣고 판단: 비속어가 끝나는 시점에 1 부여

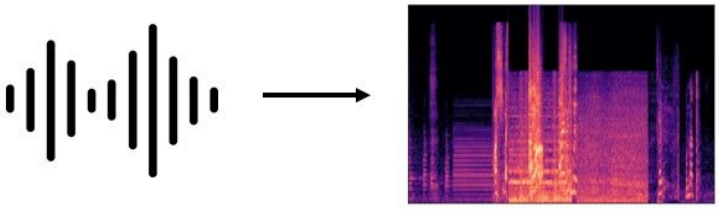


- Dataset

  독립된 clip으로 train/test data 생성



- Data preprocessing
  - Flipping: 음성 vector를 반전시켜 1개의 음성을 2개로 augmentation
  - Mel-spectrogram: 음성 vector를 Log scale mel-spectrogram으로 변환
    params: n_fft(윈도우 크기) 200, hop_length(겹침 크기) 80
  - Normalization: -80~0 scale → 0~1 scale



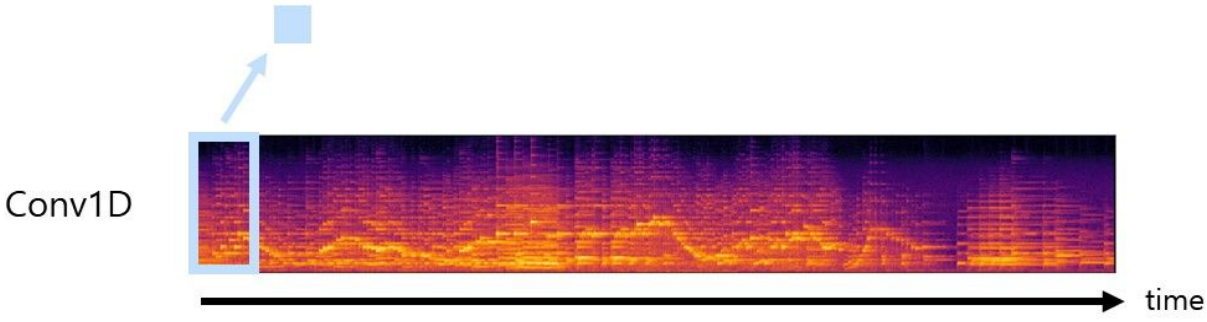
Mel-spectrogram 변환 결과

모델

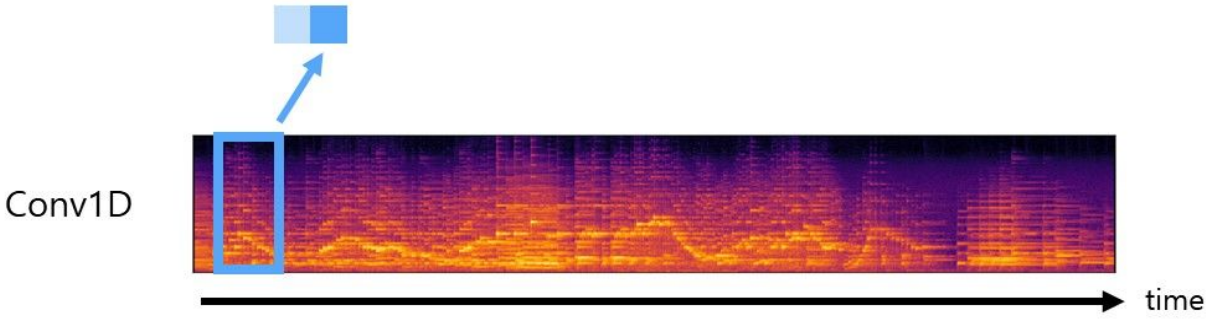
- Model
  - 시계열의 음성데이터를 처리하기에 적합한 모델
  - Conv1D + GRU

    convolution으로 mel-spectrogram의 길이를 줄이고, RNN 계열의 모델 GRU로 학습

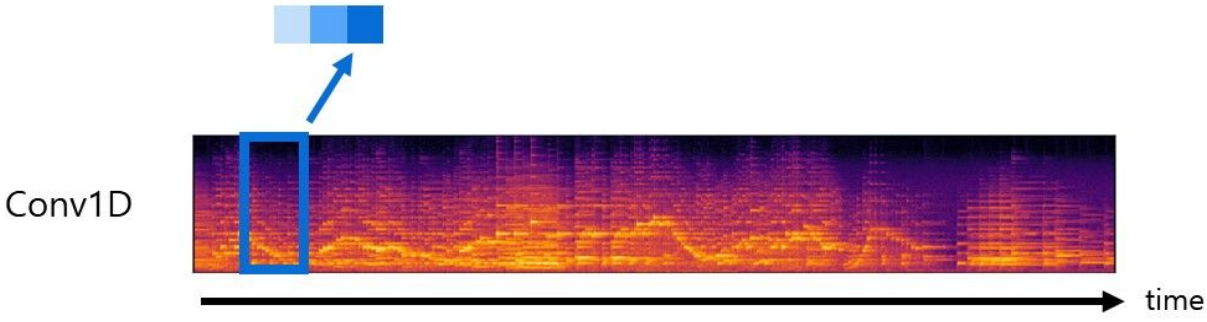
- Conv1D + GRU



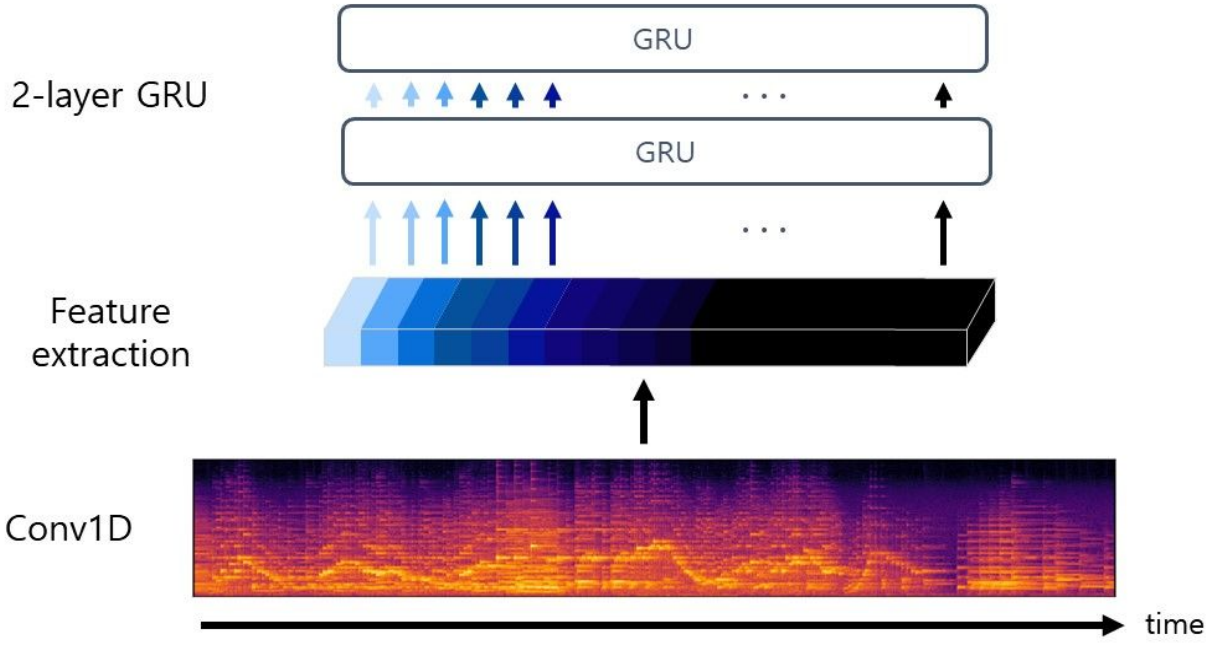
- Conv1D + GRU



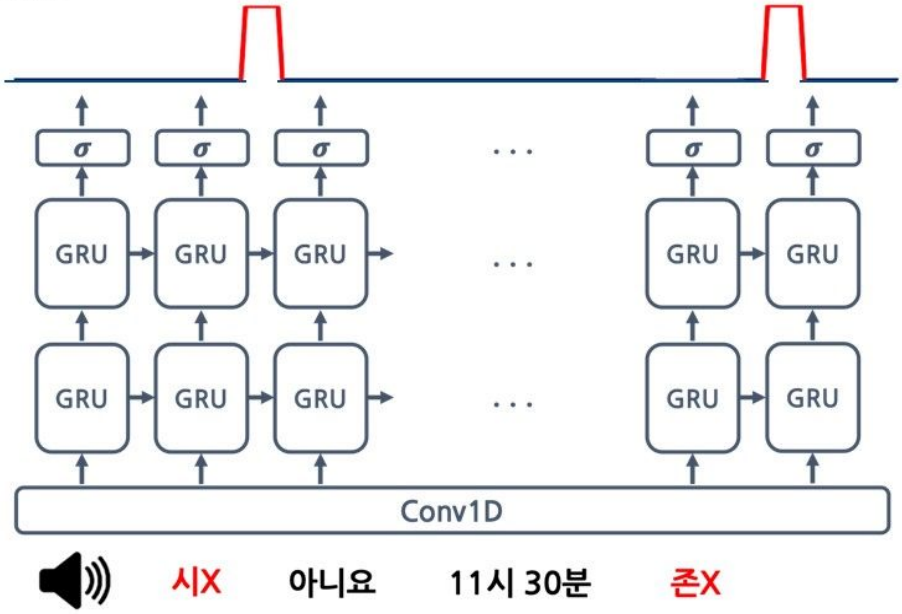
- Conv1D + GRU



- Conv1D + GRU

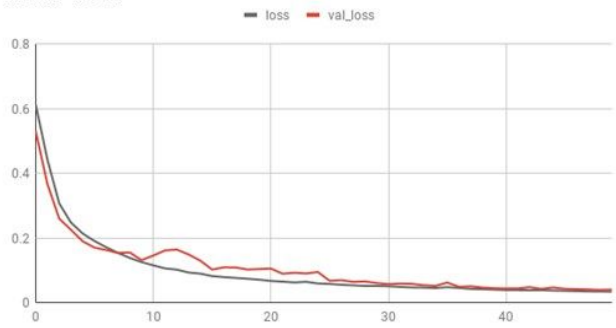


- Conv1D + GRU

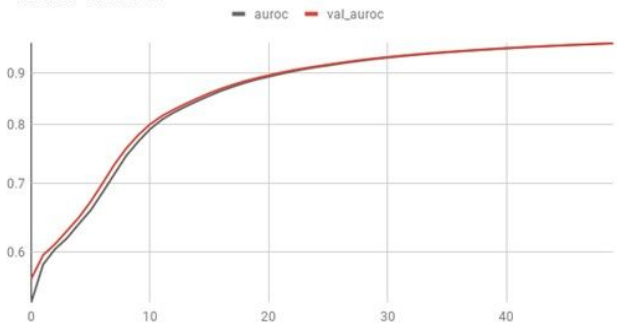


- Train result
  - Loss: binary crossentropy
  - Metric: Imbalanced label의 이유로 auroc 사용
    ※ Test auroc: 0.98, accuracy: 0.97

Train loss

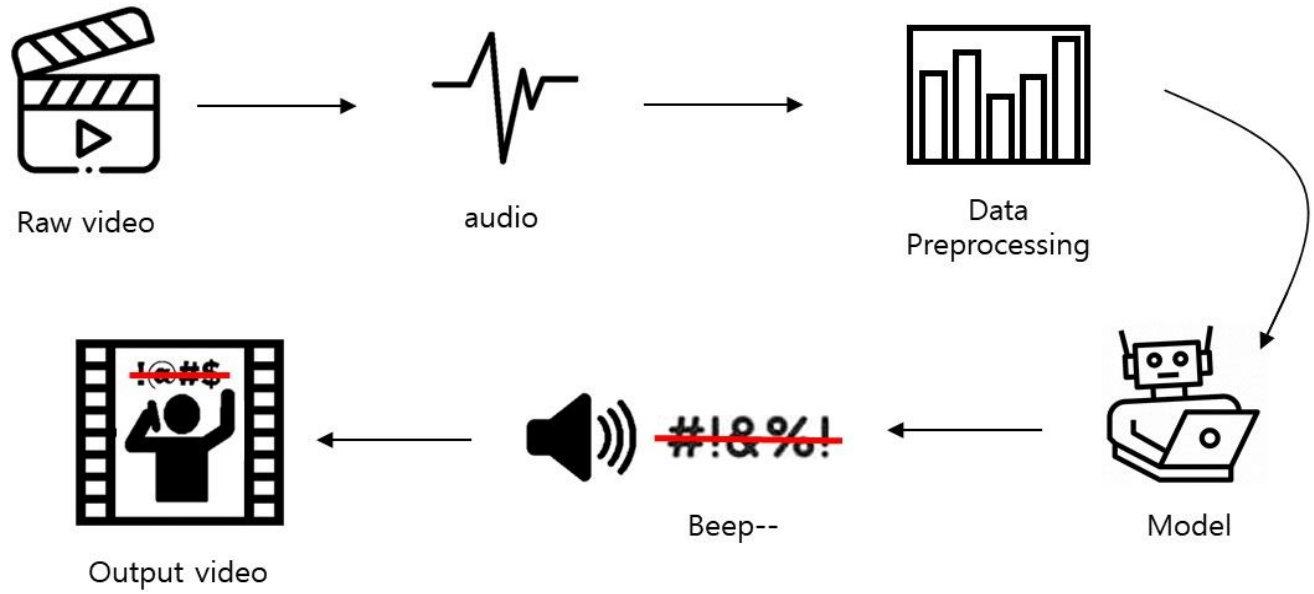


Train auroc



결과

- Inference Flow



YouTube 결과

투빅스 제 9회 컨퍼런스
삐--

- DEMO 1 - 음성

Without Beep

With Beep

YouTube 결과

투빅스 제 9회 컨퍼런스
삐--

- DEMO 2 - 방송

YouTube 결과

투빅스 제 9회 컨퍼런스
뻐--

- DEMO 3 - 영화

YouTube

마무리 – 의의 및 한계

투빅스 제 9회 컨퍼런스
뻬--

- 의의

- 한국어 비속어 필터링 구현

- 한계

- 새로 만들어지는 비속어 필터링 불가능

- 여성 화자 데이터 부족

→ 데이터 추가 수집 및 학습 필요

결과

투빅스 제 9회 컨퍼런스
뻐--

DEMO

Real Time

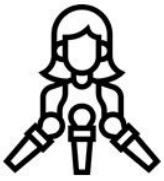
YouTube 마무리 - 한계

투빅스 제 9회 컨퍼런스
뻬--

- 여성데이터 부족

- 비속어 데이터가 남성에 편중되어 있음
- 여성 화자의 비속어는 잘 탐지 못하는 경향

→ 데이터 추가로 극복 가능



- 비속어 한정성

- 학습한 비속어만 필터링 가능
- 새로운 비속어, 혹은 잘 쓰지않는 비속어는 필터링 곤란

→ 꾸준한 데이터 추가 및 업데이트로 극복 가능

YouTube 마무리 - 활용방안

투빅스 제 9회 컨퍼런스
뻬--

• 텔레마케터

- 감정노동자들을 폭언 및 비속어로부터 보호 가능
- 근무 환경 개선 및 이직률 하락 기대

• 음성채팅

- 게임, 채팅 관련 필터링 사각지대인 음성 채팅에서 비속어 필터링 가능
- 욕설 지양 문화 조성 및 어린이들 모방 차단 기대

마무리

투빅스 제 9회 컨퍼런스
삐--

Q & A

숙명여자대학교
통계학과 16
이민주

건국대학교
응용통계학과 17
심은선

국민대학교
빅데이터경영통계 17
김주호

서울시립대학교
컴퓨터과학부 16
이승현

YouTube

삐--

투빅스 제 9회 컨퍼런스
삐--

Thank you

https://github.com/LEEMINJOO/Beeeep--