
CYBERAFFITI

인터넷방송 유해성 산출을 통한
필터링 시스템

3조

남궁찬 박재은 장민교
조민수 정우태 채수영

목차



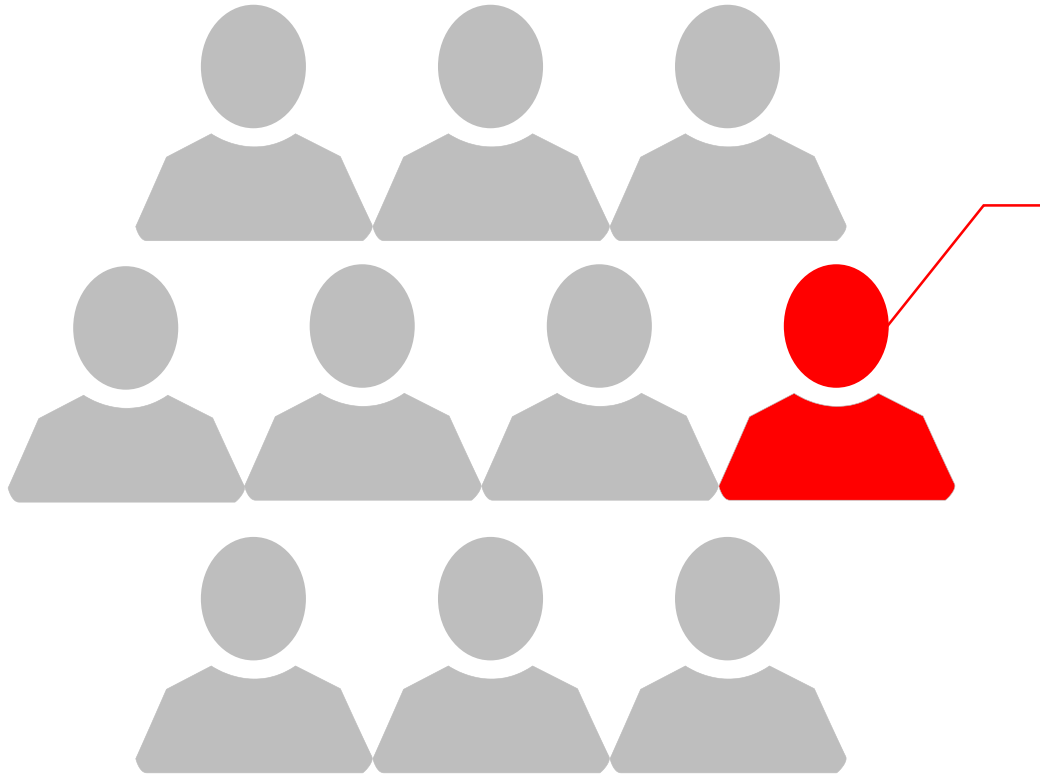
01.필요성 02.차별점 03.진행 과정
05.시연 06.QnA



경고

본 발표자료에는 다수의 욕설이
포함되어 있습니다

필요성


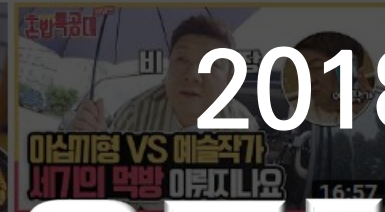


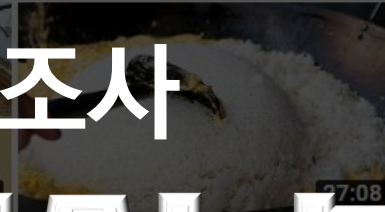
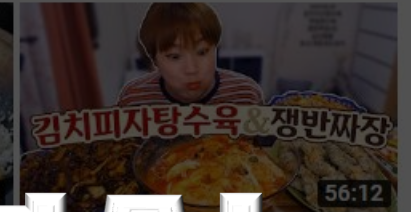





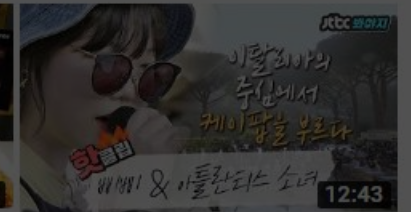


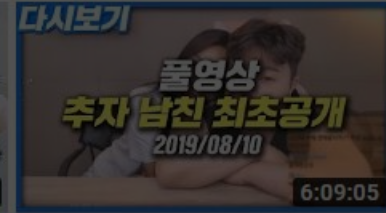





**10대 청소년
10명 중 1명
인터넷 방송
시청**

필요성

**10대 청소년 하루 평균
인터넷 방송 시청시간
2시간**

2018 방송통신위원회 조사 총 27개 유해 영상 분석

 <p>오동통면 질릴 때까지 면 먹방~! 190807/Mukbang, eating...</p> <p>입짧은햇님 조회수 33만회</p>	 <p>아침미행 VS 예술작가 새민 먹방 아티스트</p> <p>식스틴 조회수 19만회</p>	 <p>초간판면 초간판면</p> <p>초간판면 조회수 13만회</p>	 <p>김치피자탕수육 & 쟁반짜장</p> <p>김치피자탕수육 & 쟁반짜장!! 조회수 29만회</p>	 <p>이탈리아의 수원에서 케이팝을 부르다</p> <p>이탈리아의 수원에서 케이팝을 부르다 조회수 133만회</p>	 <p>멜론뮤</p> <p>멜론뮤 조회수 47만회</p>
 <p>선공개] 비밀 연애 커플의 사연?! 액셀 개발하신 분들이 ...</p> <p>tvN 조회수 5.2만회</p>	 <p>엄마코끼리가 잠든 아기코끼리 깨우는 방법 (대박~) ...</p> <p>SBS TV동물농장x애니멀봐 조회수 863만회</p>	 <p>울진 기성면 갯바위 투망 이게 사실이야</p> <p>이호fan art 조회수 39만회</p>	 <p>김치크림떡볶이 새우튀김 먹방 Mukbang</p> <p>FRAN 조회수 210만회</p>	 <p>이탈리아의 수원에서 케이팝을 부르다</p> <p>이탈리아의 수원에서 케이팝을 부르다 조회수 133만회</p>	 <p>블랙헤드 없애고 안 생기게 하는 법!</p> <p>블랙헤드 없애고 안 생기게 하는 법 클렌징 오일 (광고 없음) 조회수 38만회</p>
 <p>여름엔 열무보리비빔밥!!! 190807/Mukbang, eating...</p> <p>입짧은햇님 조회수 10만회</p>	 <p>트렌 장추자 생방송 다시보기 #19.08.10 [추자 남친 최초...</p> <p>장추자 조회수 29만회</p>	 <p>당신의 시간을 절약해 줄 엄청난 난 부엌 꿀팁 30가지</p> <p>5 분 Tricks 조회수 91만회</p>	 <p>백종원의 백종원 레시피! 돼지고기 등뼈! 불맛 짬뽕라면</p> <p>백종원의 요리비책 조회수 173만회</p>	 <p>부부용으로 강력 추천 듀오탑 메탈</p> <p>품질과 구성까지, 모든 것이 놀라운 르노 마스터 캠핑카 (듀... 조회수 119만회</p>	 <p>블랙헤드 없애고 안 생기게 하는 법!</p> <p>블랙헤드 없애고 안 생기게 하는 법 클렌징 오일 (광고 없음) 조회수 38만회</p>

2018 방송통신위원회 조사


총 27개 유해 영상 분석




오동통면 질릴 때까지
면 먹방~! 190807/Mukbang, eating...
입짧은햇님
조회수 33만회



아스피린 VS 예술작가
세인의 먹방 아티스트
16:57




초간단면
6:01




김치피자탕수육 & 쟁반짜장
56:12


1034개



[선공개] 비밀 연애 커플의 사연?! 액셀 개발하신 분들이 ...
tvN
조회수 5.2만회 · 4시간 전




[멜론차트] 2019년 8월 1주차 2일자 실시간 TOP 33 | 광고...
멜론뮤직
조회수 47만회 · 1주 전



애니멀왕
4:54




치크림떡볶이 새우튀김 먹 Mukbang
AN
조회수 210만회 · 3개월 전




핫클럽
12:43


유해 내용 발견




여름엔 열무보리비빔밥!!!
190807/Mukbang, eating...
입짧은햇님
조회수 10만회 · 5일 전




블랙헤드 없애고 안 생기게 하는 법!
Yoo True
조회수 38만회 · 1주 전




트젠 장추자 생방송 다시보기 #19.08.10 [추자 남친 최초...
장추자
조회수 29만회 · 스트리밍 시간: 2일 전



당신의 시간을 절약해 줄 엄청난 부엌 꿀팁 30가지
5 분 Tricks
조회수 91만회 · 1개월 전



[백종원의 백종원 레시피] 돼지고기 등뼈! 불맛 짬뽕라면
백종원의 요리비책
조회수 173만회 · 2주 전



부부용으로 강력 추천 듀오탑 메탈
최초공개
27:37

필요성



보검정모 7천명이와서 영화관
터졌습니다 영화관측에는진...

조회수 131만회 · 3개월 전

평균 1편당 38회

1분당 3.9회의

유해 내용 발견

- 2018 방송통신위원회 -

필요성

YouTube

afreecaTV



**청소년들 TOP3 플랫폼에서
욕설, 성인영상 등
유해한 영상 및 채팅이 난무하는 상황**

필요성 : 기존 플랫폼 필터링

YouTube 신고 기반

afreecaTV 채팅방 필터링

twitch 채팅방 필터링

차별점 01. 욕설 필터링 개선

[기존 플랫폼 구현방식]

● 비속어 사전 기반 필터링

예) 시발, 시11발, 시111발

차별점 01. 욕설 필터링 개선

[기존 플랫폼 구현방식]

● 비속어 사전 기반 필터링

예) 시발, 시11발, 시111발

차별점 01. 욕설 필터링 개선

[기존 플랫폼 구현방식]

● 비속어 사전 기반 필터링

예) 시발, ~~시11발~~, ~~시111발~~

차별점 01. 욕설 필터링 개선



팬 머리카락.(holl2725) :

♡지가 바닥인데 누가 너를 영화출연시켜 —
—



두릅두릅두릅(hsk9015) :

블루무새 ㅅㅂ



패르고(thkksw) :

마나 ㅈㄴ 안채우고가잖아 집가고나서



세번째기회(meme1212) :

너검마가 니 이딴생각으로 글싸지르는거보면
통곡할듯ㅋ



팬 다단계란(terry6976) :

광고 ㅅㅂ



kcs4909(kcs4909) :

세에에에에에에에에에에엑스



팬 [K] 무상(duswns913) :

빨리켜 ㅂㅅ아



처처처(dudcks1999) :

아 노래 ㅈ같네

차별점 01. 욕설 필터링 개선

♡지가 바닥인데 누가 너를 영화 출연시켜

세에에에에에에에에에에엑스

아 노래 지같네

광고스비


빨리켜 브스아

너검마가 니 이딴생각으로 글싸지른거보면
통곡할듯ㅋ

차별점 01. 욕설 필터링 개선


 팬 머리카락.(holl2725) :

10분동안 필터링된 채팅은 1개뿐

 두릅두릅두릅(hsk9015) :
블루무새 ㅅㅂ

- BJ 철구 -

 팬 다단계란(terry6976) :

 kcs4909(kcs4909) :
세에에에에에에에에에에엑스


 패르고(thkksw) :

마나 ㄸ 안채우고가잖아 집가고나서

 팬 [K] 무상(duswns913) :
빨리켜 ㅂㅅ아

 세번째기회(meme1212) :

너검마가 니 이딴생각으로 글싸지르는거보면
통곡할듯ㅋ

 처처처(dudcks1999) :
아 노래 ㄸ갈네

차별점 01. 욕설 필터링 개선

Recall

기존 비속어
사전 기반

56%

모델
기반


90%

차별점 02. 유해방송 차단규제



데이터 수집 이유1

문어체 ≠ 구어체



위키백과
우리 모두의 백과사전

대문
사용자 모임
요즘 화제
최근 바뀜
모든 문서 보기
임의 문서로
도움말
기부

도구

여기를 가리키는 문서
가리키는 글의 최근 바뀜

로그인하지 않음 토론 기여 계정 만들기 로그아웃

문서 토론 읽기 편집 역사 보기 위키백과 검색

이 문서의 과거 편집 L

자연어 처리

위키백과, 우리 모두의 백과사전.


이 문서의 내용은 출처가 분명하지 않습니다.
이 문서를 편집하여, 신뢰할 수 있는 출처를 표기해 주세요.
검증되지 않은 내용은 삭제될 수도 있습니다. 내용에
대한 의견은 토론 문서에서 나누어 주세요. (2010년 10월)

자연어 처리(自然語處理) 또는 자연 언어 처리(自然言語處理)는 인간의 언어 현상을 컴퓨터와 같은 기계를 이용해서 모사 할수 있도록 연구하고 이를 구현하는 인공지능의 주요 분야 중 하나다. 자연 언어 처리는 연구 대상이 언어 이기 때문에 당연하게도 언어 자체를 연구하는 언어학과 언어 현상의 내적 기재를 탐구하는 언어 인지 과학과 연관이 깊다. 구현을 위해 수학적 통계적 도구를 많이 활용하며 특히 기계학습 도구를 많이 사용하는 대표적인 분야이다. 정보검색, QA 시스템, 문서 자동 분류, 신문기사 클러스터링, 대화형 Agent 등 다양한 응용이 이루어 지고 있다.

베인 꿈살 ㅋㅋㅋㅋ

핀 낙천(shinyas) : 스웨인 서폿 아니었나

해니겨 우비비서갯쿠(iammeteor2) : 베인 오열



우비의아망(winds4171) 별풍선 3개 선물!

해니겨 우비일벌맨듀(kye8899(2)) : 우비의아망(winds4171)님 별풍선 3개 선물 나이따~!

우비의아망(winds4171(2)) : 우비가 약자열시는 잘하는 타입이었구나.. 상 약자여서 몰랐어

핀 낙천(shinyas) : 왜케 썸

데이터 수집 이유2

유해도가 측정된 데이터의 부재

크롤링

데이터 수집
약 150만

**데이터
Set
voting**

투표로 유해도 판별



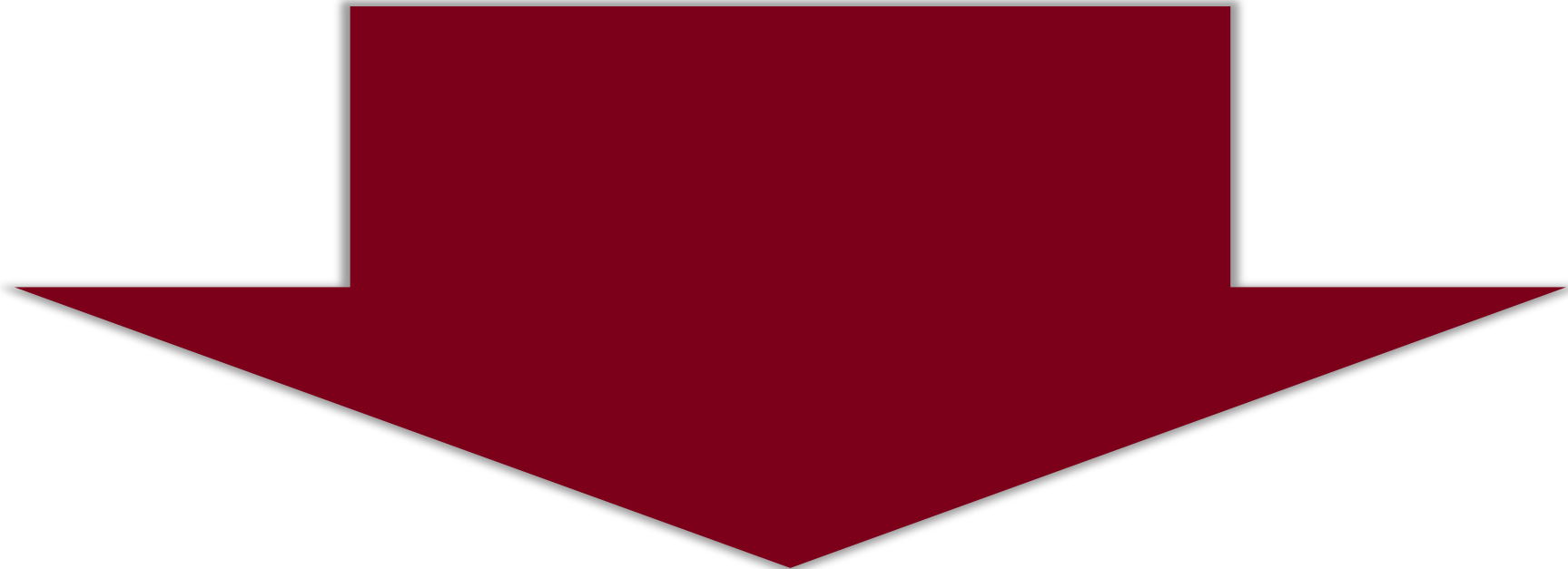
레이블링 약 10만 개

데이터 Set

Inbalanced data

유해데이터 < 정상데이터

데이터 Set



Data Augmentation

Wei and Zou(2019) Wei, J., Zou, K. 2019. EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks. arXiv:1901.11196.

데이터 전처리

형태소 분석 Okt

1 tmp.pos('양기모띠 ㅋㅋㅋ 시1발')

executed in 19ms, finished 12:34 16 2019-08-27

[('양', 'Adverb'),
('기', 'Modifier'),
('모', 'Modifier'),
('ㅋㅋㅋ', 'KoreanParticle'),
('시', 'Noun'),
('1', 'Number'),
('발', 'Noun')]

([('썩뻘' , 'Noun') , [('씨빠' , 'Noun')]])

데이터 전처리

형태소 분석 Okt

한계

1. 일부 채팅을 의미 단위로 토큰화하지 못함

2. 유사 토큰을 인식하지 못함

1 tmp pos('앙기무띠 ㅋㅋㅋ 시1박')

(('앙', 'Adverb'),
(('기', 'Modifier')

(('ㅋㅋㅋ', 'KoreanParticle'),
(('시', 'Noun'),
(('1', 'Number'),
(('박', 'Noun'))]

([('씩뽀' , 'Noun') , [('씨빠' , 'Noun')])

데이터 전처리

NOGADA Okt

**(1. 일부 채팅을 의미단위로
토큰화하지 못함)**

데이터 전처리

NOGADA Okt

**(1. 일부 채팅을 의미단위로
토큰화하지 못함)**

```
1 nKT.prep2(['앙기모띠 ㅋㅋㅋ 시1발'], '1')[0][0]
```

executed in 86ms, finished 12:34 18 2019-08-27

['앙기모띠', '시발']

약 3% 성능이 향상

3%

데이터 전처리

BPE Byte Pair Encoding
(2. 유사 토큰 인식하지 못함)

데이터 전처리

BPE Byte Pair Encoding (2. 유사 토큰 인식하지 못함)

```
1 bp_tokenize(['싹뻥', '씨빠'])
```

executed in 18ms, finished 12:45:57 2019-08-27

100%|

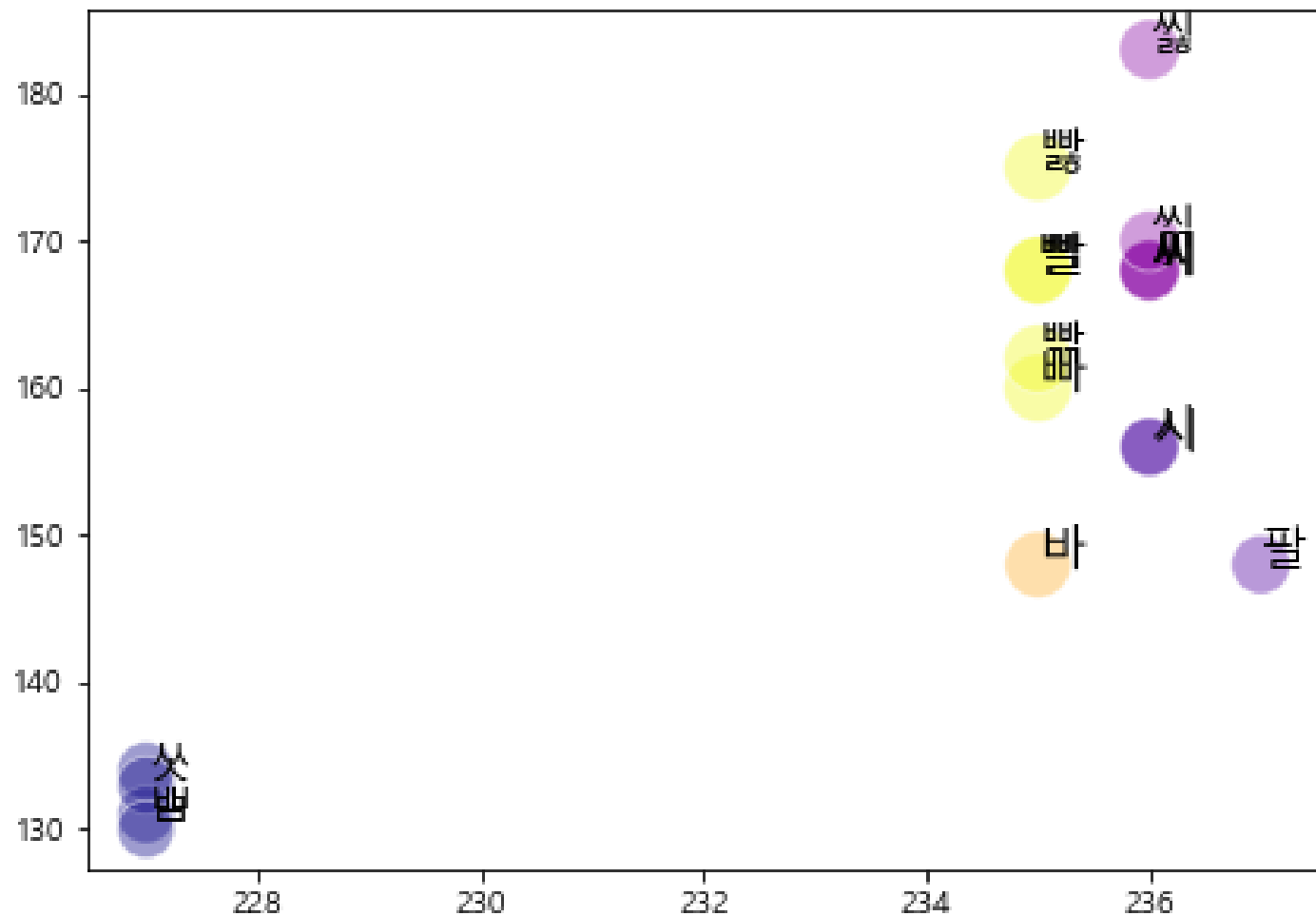
```
array([[ '236', '148', '170', '235', '185', '162'],  
       [ '236', '148', '168', '235', '185', '160']], dtype='<U3')
```

노가다와 성능 유사

일반화가 좋음

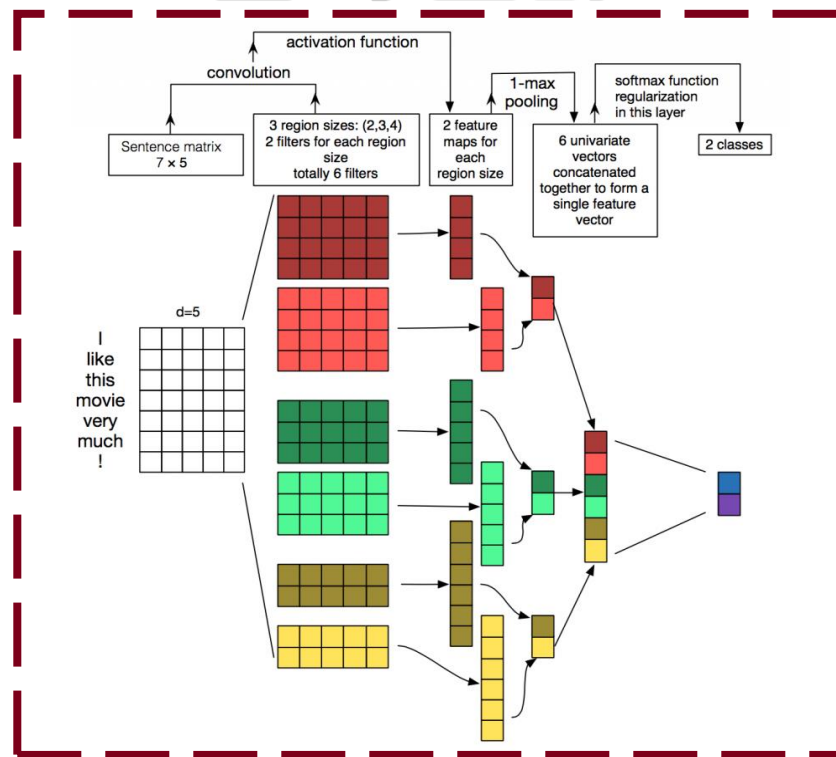
데이터 전처리

썩썩썩썩



적용 모델

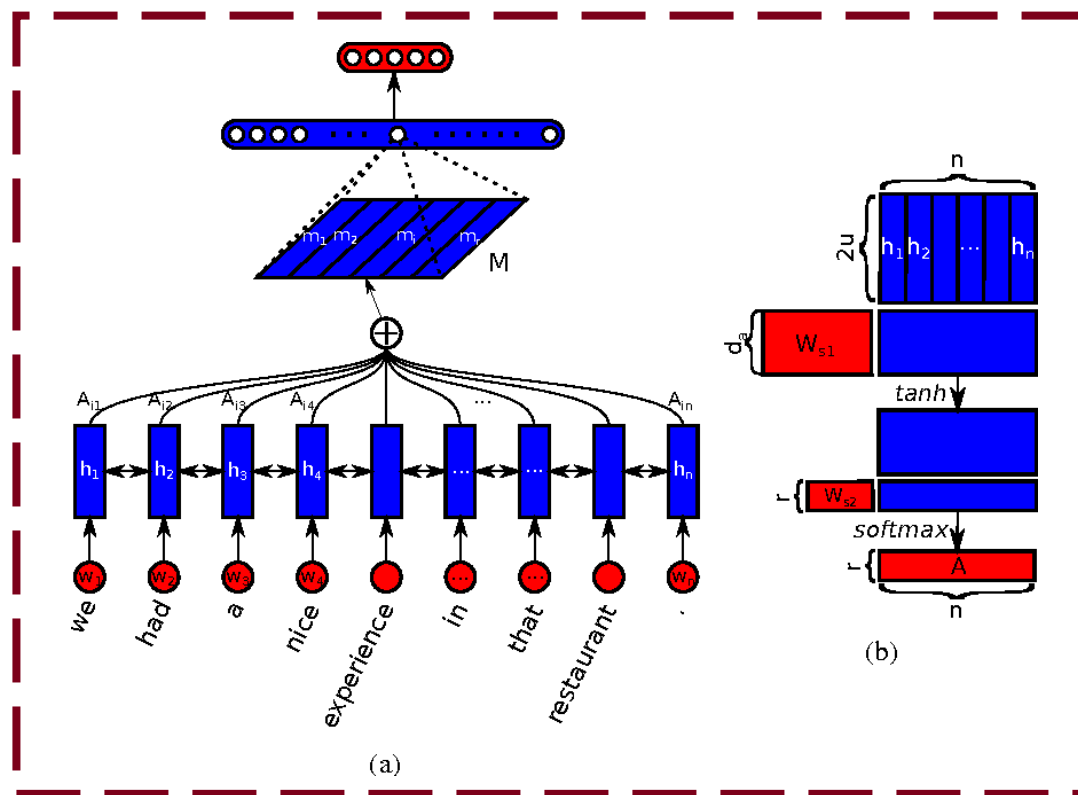
CONV1d 문맥 고려



유해 토큰을 제대로 인식하는지 검증 할 수 없는 한계

적용 모델

Bi-LSTM Self-Attention



적용 모델

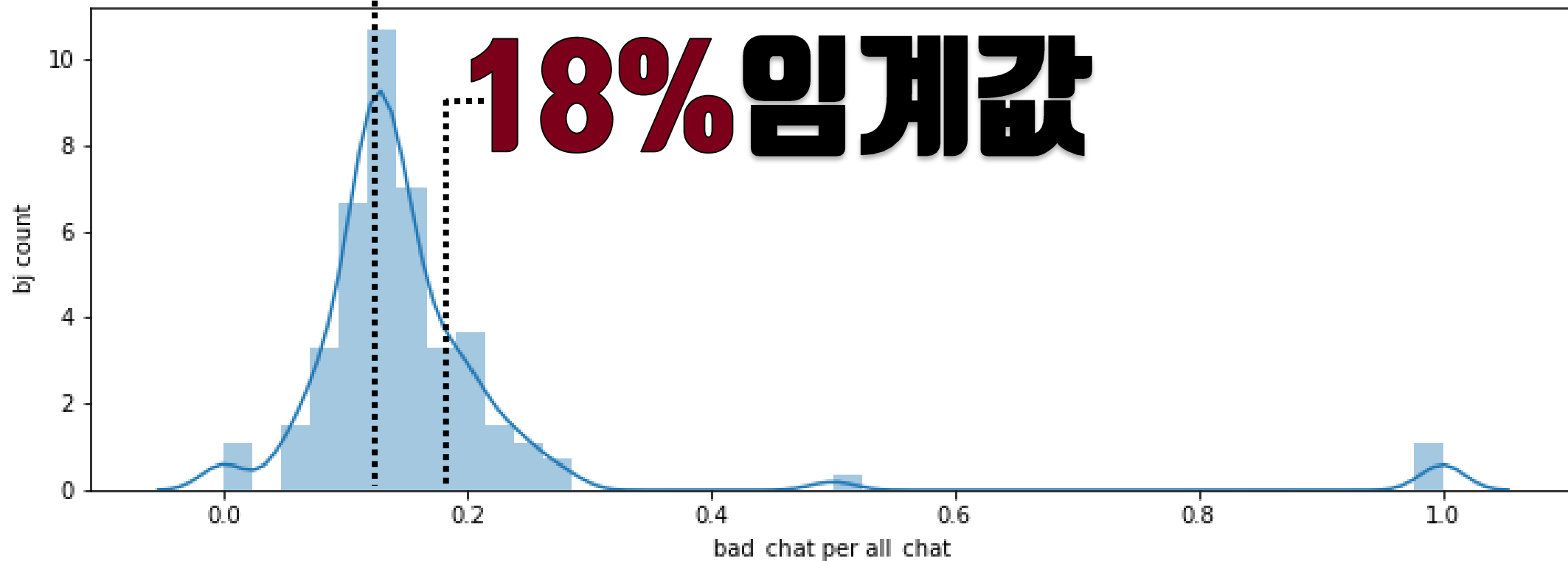
Bi-LSTM Self-Attention

_____조_금_모_자_라_지_만_착_한_친_구_야_
_미_10_뒤_진_넘_아_이_러_면_욕_못_잡_겠_지_?
하_하_이_착_한_친_구_는_개_뽕_개_병_신_새_끼_
_____착_한_척_하_는_나_의_씩_뽕_씩_깍_약_
_씩_독_이_는_프_로_그_램_이_야_뽕_다_겔_아_

유해도 산출

13% 중간값

18% 임계값



시연 및 데모 페이지

<https://chrome.google.com/webstore/search/Cyberaffiti>

<http://150.95.153.18:8000/>

QnA

경청해 주셔서 감사합니다

