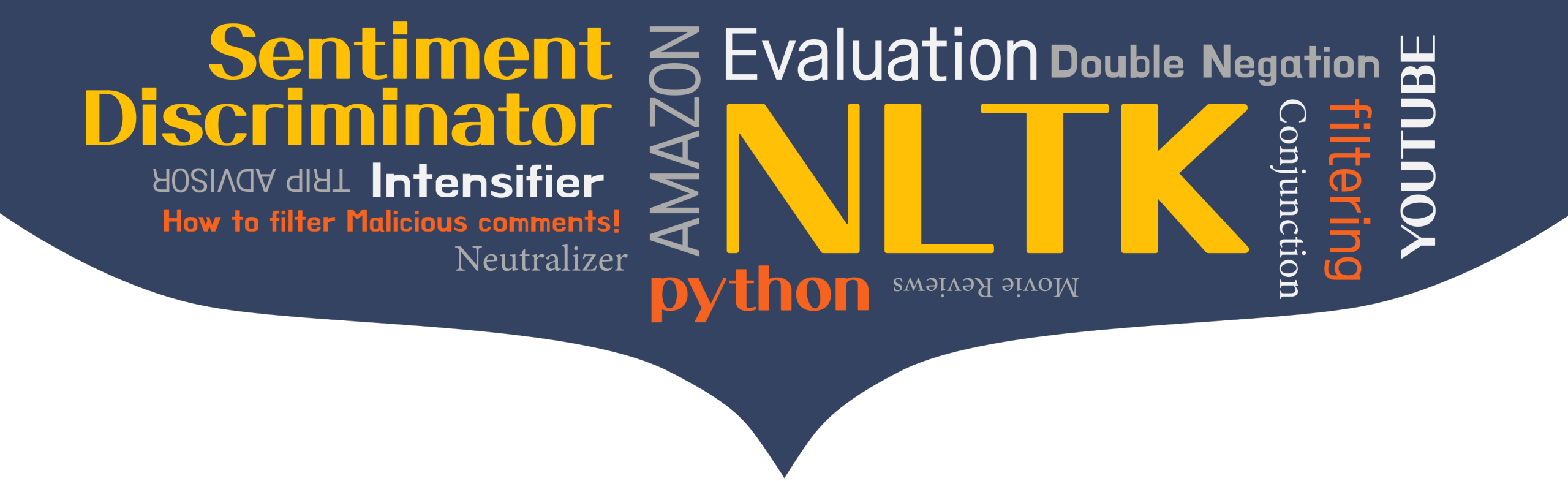


CS372 Team Project

Filtering Malicious Comments with NLTK

Team 23 | 김승호, 명재현, 방형조, 이진원, 최다은

01 | Introduction

Watcha

YouTube

“Comments”

“
lot of malicious comments
→ Rating & Filtering comments with NLTK”

01 | Introduction

Data

1. Collect data with ratings

- e.g. product reviews, book/movie reviews.
- 5 datasets:
  Amazon, Mendeley movie reviews,
  Trip advisor, Book reviews
  -> more than 1 million reviews

2. Youtube Crawling

- For evaluation

External Model & Data

1. vader_lexicon.txt

- SentiWordnet's low accuracy
  (especially for comments)

2. `autocorrect` python package

- Lots of typos in comments
- Spell check each comment

02 | Method

Mode

9 aspects to increase the performance of our results

Intensifiers

Increase score of words that follow intensifiers

e.g. really, very

Neutralizers

Opposite of intensifier

e.g. seldom, sparsely

Uppercase

If all letters are capitalized, amplify score

e.g. I REALLY LIKE THIS

Threshold

Exclusion of words with a score below certain level (for noise removal)

Emphasis

Give an amplified score for the first and last sentences of the review

Conjunction

Put weight on sentences that start with a conjunction.

Exclamation

Magnify score if there are many exclamation marks

e.g. love it!!!

Negative words

If a word follows a negative describer, reverse the sentiment

No/not phrases

Reverse phrases' sentiment if they start with 'not' or 'no'

02 | Method

Mode

9 aspects to increase the performance of our results

Intensifiers

Increase score of words that follow intensifiers

e.g. really, very

Neutralizers

Opposite of intensifier

e.g. seldom, sparsely

Uppercase

If all letters are capitalized, amplify score

e.g. I REALLY LIKE THIS

Threshold

Exclusion of words with a score below certain level (for noise removal)

Emphasis

Give an amplified score for the first and last sentences of the review

Conjunction

Put weight on sentences that start with a conjunction.

Exclamation

Magnify score if there are many exclamation marks

e.g. love it!!!

Negative words

If a word follows a negative describer, reverse the sentiment

No/not phrases

Reverse phrases' sentiment if they start with 'not' or 'no'

02 | Method

Mode

9 aspects to increase the performance of our results

Intensifiers

Increase score of words that follow intensifiers

e.g. really, very

Neutralizers

Opposite of intensifier

e.g. seldom, sparsely

Uppercase

If all letters are capitalized, amplify score

e.g. I REALLY LIKE THIS

Threshold

Exclusion of words with a score below certain level (for noise removal)

Emphasis

Give an amplified score for the first and last sentences of the review

Conjunction

Put weight on sentences that start with a conjunction.

Exclamation

Magnify score if there are many exclamation marks

e.g. love it!!!

Negative words

If a word follows a negative describer, reverse the sentiment

No/not phrases

Reverse phrases' sentiment if they start with 'not' or 'no'

02 | Method

Mode

9 aspects to increase the performance of our results

Intensifiers

Increase score of words that follow intensifiers

e.g. really, very

Neutralizers

Opposite of intensifier

e.g. seldom, sparsely

Uppercase

If all letters are capitalized, amplify score

e.g. I REALLY LIKE THIS

Threshold

Exclusion of words with a score below certain level (for noise removal)

Emphasis

Give an amplified score for the first and last sentences of the review

Conjunction

Put weight on sentences that start with a conjunction.

Exclamation

Magnify score if there are many exclamation marks

e.g. love it!!!

Negative words

If a word follows a negative describer, reverse the sentiment

No/not phrases

Reverse phrases' sentiment if they start with 'not' or 'no'

02 | Method

Scoring Algorithm

(Review Score) = Σ (Sentence Score) * (Sentence Sentiment Importance)

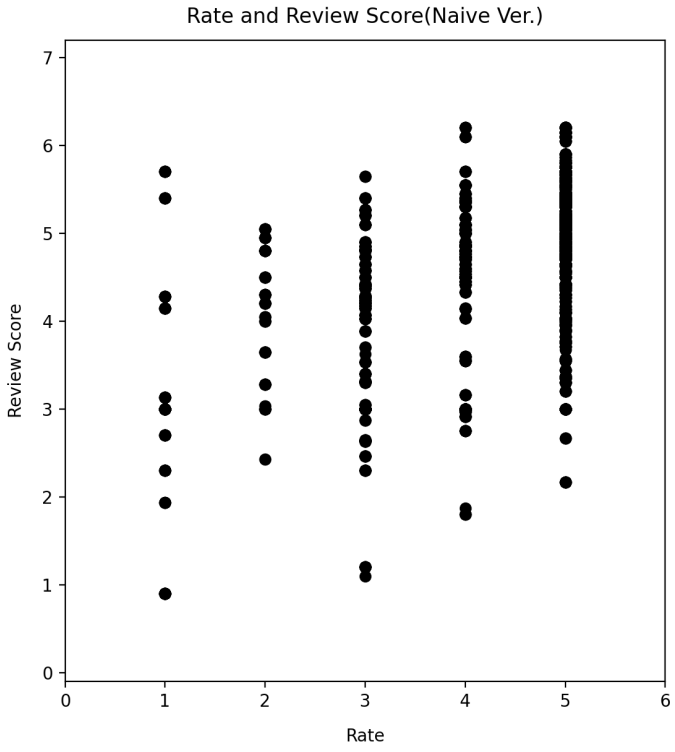
(Sentence Score) = Σ (Word Score) * (Word Sentiment Modifiers)

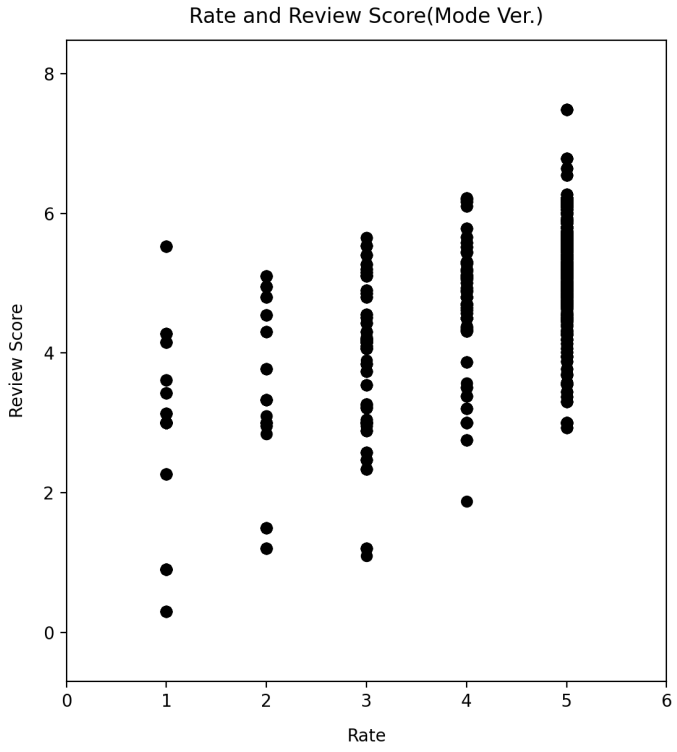
So, what we could discover... **IT PERFORMS WELL OVERALL!**

The main key was
First, amplify the positive/negative sentiments extracted.
Second, reverse the sentiments according to their contexts.

03 | Result

How do these modes affect the results?





03 | Result

Results of individual reviews

```
These run large. I wear a 9.5 exactly in every shoe, but these I bought a 9, and the toe area is pretty wide plus plenty long even with socks.
 I love them but glad I was able to try them on before purchasing.
score_naive:    4.80, score_mode:    5.09, answer:  5.00
```

```
Absolutely love them. Can even wear them without my insoles.
score_naive:    5.70, score_mode:    6.98, answer:  5.00
```

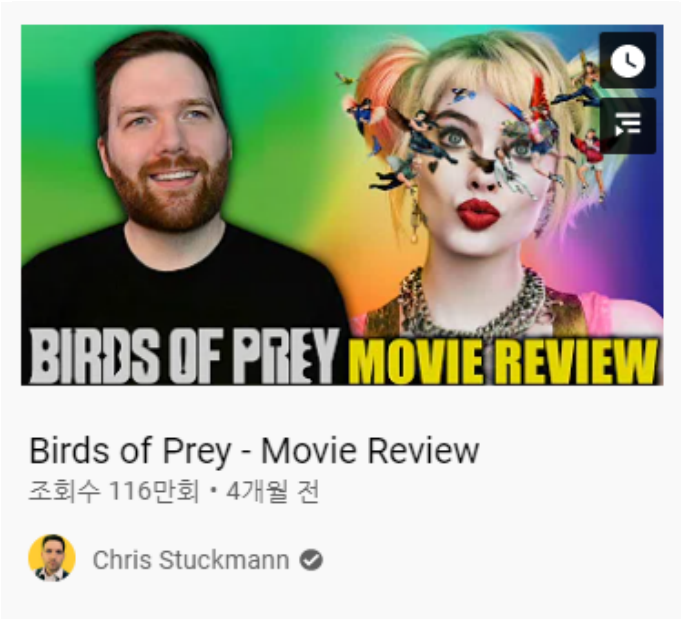
```
Love these! I have 3 pairs...they're so comfortable - I'm on my feet all day & my feet
actually don't hurt at the end of the day when I wear these :)\nI've learned to get these 1/2
size smaller....I'm normally a size 7 and the 6 1/2 fit perfectly
score_naive:    4.13, score_mode:    5.05, answer:  5.00
```

F-score

- **Amazon review dataset**
  True Positive: 0.877 / False Positive: 0.074 /
  False Negative: 0.049 / **F-score: 0.934**
- **TripAdvisor review dataset**
  True Positive: 0.717 / False Positive: 0.110 /
  False Negative: 0.174 / **F-score: 0.835**

04 | Application

Analyzing YouTube comments




```
david@Natrium ~/Documents/cs372/CS372-NLP master python main.py
########## FILTERED OUT 50 REVIEWS WITH THE LOWEST SCORE ##########

Review Text: In Deadpool we side with a fucking killer!!!!!!!!!!
Score Mode:   -1.62

Review Text: This movie is so fucking bad
Score Mode:   -0.50

Review Text: Killed her... husband...whackedd emm!!!
Score Mode:   -0.50

Review Text: KyDogg Batmanfan30 I would rather watch suicide squad 🧢🚫
Score Mode:   -0.50

Review Text: movie was so bad
Score Mode:   -0.50

Review Text: Why would they kill him off before he really even becomes black mask, even just killing him in general was very dumb
Score Mode:   -0.44

Review Text: Saw this movie for the first time last night. What a misandristic piece of bullshit.
I fucking hate modern feminism...
Score Mode:   -0.39

Review Text: Unfunny dead pool
Score Mode:   -0.30

Review Text: If you took a shot Everytime you cringed watching this movie you would be dead from alcohol overdose
Score Mode:   -0.30

Review Text: DeadPool 1 > Dead Pool 2 > Iron Man 3 > Sh*t > Birds of Prey
Score Mode:   -0.30

Review Text: u mean the crossbow killer?
Score Mode:   -0.30

Review Text: I hated this movie.
Score Mode:   -0.20
```

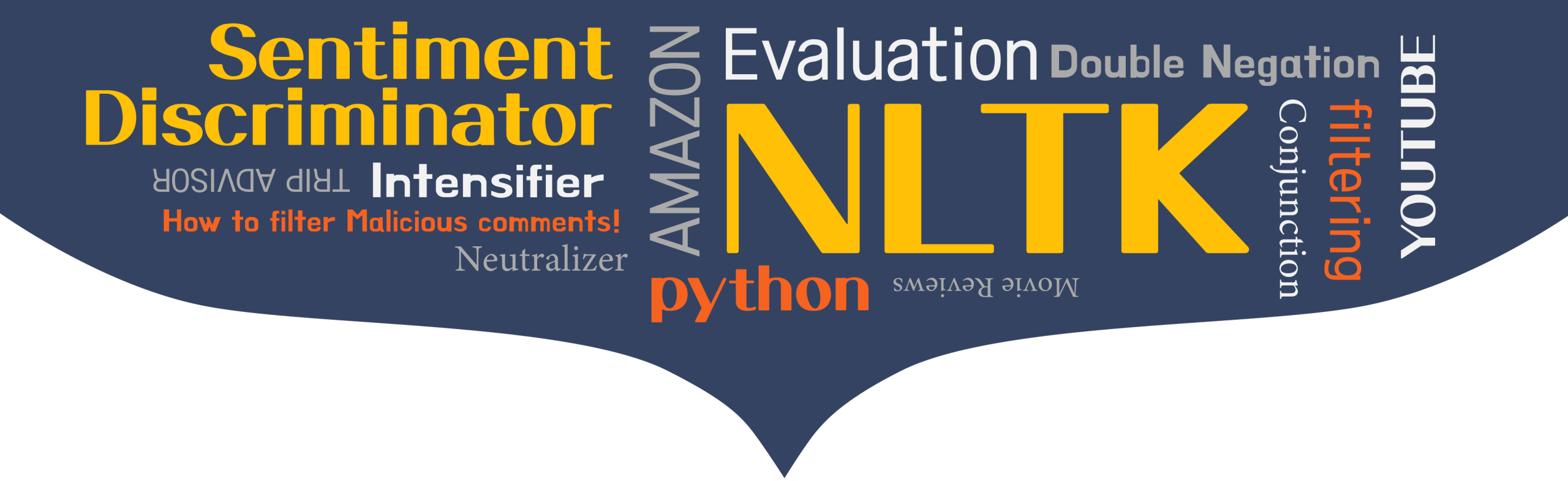
05 | Further Improvement & Suggestion

How to improve quality?

- **Using big data**
  Core of the sentiment discriminator is the scoring of each words.
  → use big data to search for commonly used expressions → score better
- **Use different models depending on the type of comments**
  i.e. different models for comments on movies, news, music videos, mukbang, etc.
- **Method for discrete rating**
  Current implementation: infinite range → 1 to 5 discrete rating (like common reviews)

Where to use this?

- Filtering malicious comments
- YouTube channels can obtain an overall numeric mood of comments
- …


Sentiment Discriminator
TRIP ADVISOR
Intensifier
How to filter Malicious comments!
Neutralizer
AMAZON
Evaluation
Double Negation
NLTK
python
Movie Reviews
Conjunction
filtering
YOUTUBE

THANK YOU :)