Author's affiliation logo

12th ICAEIC-2024
ICT-Advanced Engineering Society

An Engineering Approach to Datasets and Models for Language Sensitivity

Huh Eun

Hansung University, Seoul, Korea

e2huh321@gmail.com

Outlines

1. What is language sensitivity?
2. Dataset and Dictionary construction
3. Sentiment Analysis Model
4. Language Sensitivity Analysis Model
5. Alternative Expressions Proposal
6. Result
7. Conclusion

1. What is Language Sensitivity?

Language Sensitivity:

- When using language, you must sensitively consider various situations.
- There is a growing trend that words with such discriminatory meanings should be replaced with other words.
- But there was no dataset or model to distinguish language sensitivity.
- An engineering approach about language sensitivity is needed.

2. Dataset and Dictionary construction (1)

- Based on the 2019 National Human Rights Commission statistics 107p, we set up 6 categories
- Multi-labeling was performed
- Train not to misjudge homophonic words

Table 1. Example of Labeling

No.	Sentence	Disability and Medical history	Gender and Family	Social status	Origin	Profanity	Other	Language sensitivity_LOW	Language sensitivity_HIGH
1	나는 새 유모차를 샀어	0	1	0	0	0	0	1	0
2	깜깜이 환자가 증가하고 있다	1	0	0	0	0	0	1	0
3	불우이웃을 위해 반팔티를 기부했다	1	0	1	0	0	0	1	0

2. Dataset and Dictionary construction (2)

Table 2. Example of Dictionary [Gender and Family]

Index	From	To
ㄱ	가정부	가사관리사
	결손가족	가족
	결손가정	가정
	김 여사	운전 미숙자
	김여사	운전 미숙자
	고아원	아동복지시설
	경력단절여성	고용중단여성
	광부	광산 노동자
	과부	故 000 씨 배우자

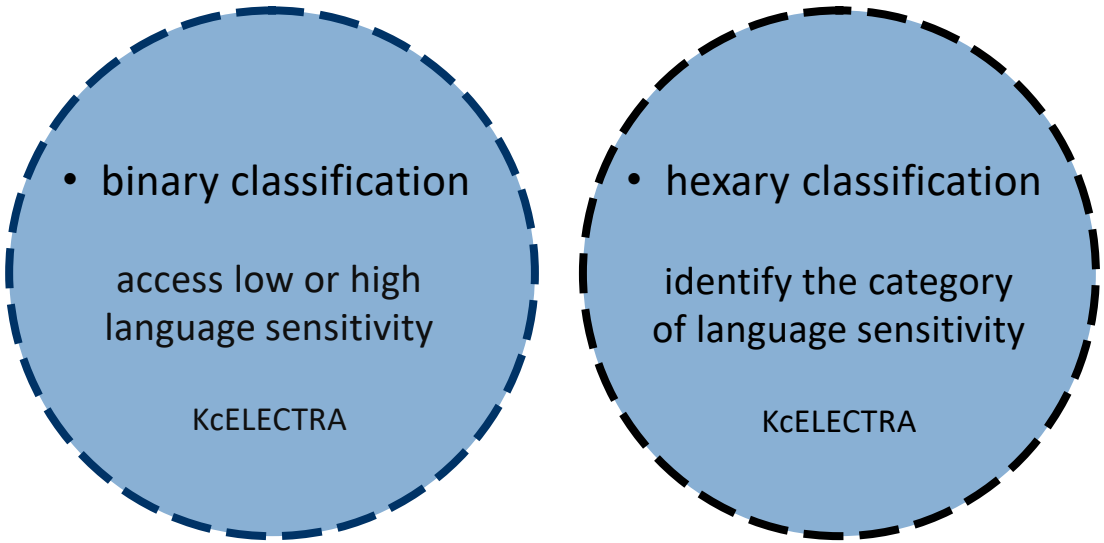
- Alternative words
- Semi-automated dictionary update

3. Sentiment Analysis Model

train data	fine-tuned model	evaluation
Naver movie reviews online shopping mall reviews	Kobert	accuracy 0.8947 F1-score 0.8902

4. Language Sensitivity Analysis Model

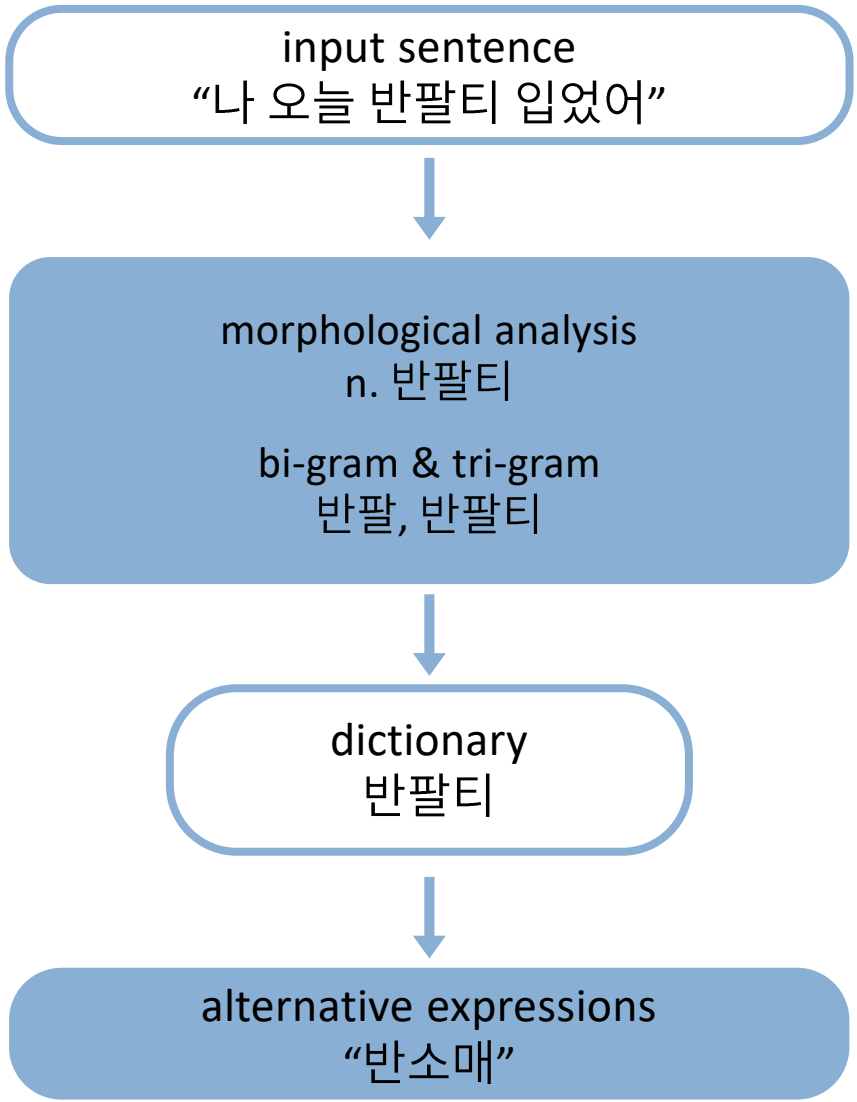
<Language Sensitivity Model>



<example>

input
"나 오늘 반팔티 입었어"
↓
model
↓
result
low language sensitivity
Category of disability

5. Alternative Expressions Proposal



6. Result (1)

Model Evaluation

- Sentiment Analysis model based on KOBERT
- Language Sensitivity Analysis models based on KCELECTRA

Table 3. Classification Model Evaluation Metrics

Evaluation Metrics	KOBERT for binary classification	KCELECTRA for binary classification	KCELECTRA for binary classification
Accuracy	0.9532	0.9852	0.9812
Precision	0.9643	0.9655	0.9433
Recall	0.9424	0.9865	0.9943
F1	0.9535	0.9763	0.9688

6. Result (2)

Output Results

- Terms such as “독거노인” , “장애우 가정” and “취약 계층” were identified as discriminatory towards disabilities and social status
- Alternatives for words “독거노인” and “장애우 가정” are “홀몸 어르신” and “장애인 가정” respectively

Table 4. Language Sensitivity Analysis Model Result

Sentence	Disability	Gender	Social Status	Origin	Profanity	Other
관악농업협동조합은 독거노인 및 장애우 가정, 취약 계층을 대상으로 설날맞이 사랑의 쌀 전달식을 개최하였다.	0.7778	0.0001	0.9998	0.00002	0.00002	0.2292

6. Result (3)

Summarization - Limitation

1. Building word-based datasets

The study's language sensitivity range is too limited, resulting in overly high accuracy.

2. Traditional language sensitivity studies are based on specific keywords

Datasets that can determine language sensitivity in context can make progress with the LLM model.

7. Conclusion

- It can be a good tool for correcting artificial intelligence like ChatGPT that uses expressions, which do not take into account language sensitivity.

- It is first attempted an engineering approach to language sensitivity, a humanistic element.

- For companies or public institutions, unnecessary costs can be reduced and emotional connections with consumers can be maintained to generate more stable profits in the long run.

Thank you !!!