Patronizing and Condescending Language Detection in English

Catherine Qingyun Wang
Vicky Xinyi Xiang
Paula Cortés Lemos

Task description

What is PCL?

- Perez-Almendros, et al. (2020):
  - “often unconscious but harmful and discriminative”
  - Language that reveals a superior attitude or harmful assumptions (especially about marginalized groups)
  -
- PCL detection
  - Similar to toxicity detection, but not overtly inflammatory

“Don’t worry, I know this is a mistake you usually make, we all make it sometimes, but I am bringing you a solution.”

“Norberto Quisumbing Jr. of the Norkis Group of Companies has a challenge for families who can spare some of what they have: why not adopt poor families and help them break the cycle of poverty?”

Primary Task

- Binary classification
- PCL detection in English news articles
  - *Don't Patronize Me!* Dataset

- Challenges:
  - PCL can be hard for human judges to distinguish, and often subjective
  - PCL is context-dependent, and often contains similar tokens and phrases to objective reporting

"I've also dressed up as a homeless man for a shoot once , but decided to stay like that for a few more hours after experiencing being ignored and feeling invisible. It was an eye opener."

"In our society minorities can seem invisible at times due to language barrier and discrimination."

Data

Training set: 9423 items

Dev set: 1048 items

Eval set: 3831 items

- Binary classification problem
- Original dataset is labeled by human annotators on a scale of 0-4, with 4 being the most condescending
- As per the SemEval task instructions, we treat 0 and 1 as negative samples, and 2-4 as positive samples
- Each split contains 10% positive samples

Preprocessing

- Tokenization
- Remove punctuation and emojis, but keep stop words
- Insert start/stop padding tokens (and create attention masks to differentiate padding vs. non-padding)

```
Tokenized:  ['still', 'able', 'walk', ',', 'said', 'jackson', 'therapy', ',', 'tr
avel', 'st', 'ann', 'bay', 'per', 'week', 'proving', 'costly', 'using', 'walker',
'ho', '##bble', 'around', 'first', ',', 'treatment', 'working', 'wonders', ',',
'due', 'financial', 'difficulties', 'disco', '##nti', '##nu', '##e', 'therapy',
'need', 'financial', 'assistance', ',', 'unable', 'work', 'farm']
Token IDs:  [2145, 2583, 3328, 1010, 2056, 4027, 7242, 1010, 3604, 2358, 5754, 30
16, 2566, 2733, 13946, 17047, 2478, 5232, 7570, 11362, 2105, 2034, 1010, 3949, 25
51, 16278, 1010, 2349, 3361, 8190, 12532, 16778, 11231, 2063, 7242, 2342, 3361, 5
375, 1010, 4039, 2147, 3888]
```

Adaptation Task

- PCL detection in Mandarin social media forums
  - *CCPC* Dataset
  - First PCL dataset in Chinese, inspired by the *Don't Patronize Me* dataset
  - 5000 entries in a single dataset (we divided it 8-1-1 for train, dev, and eval)

问题是他们贫穷和苦难是因为他们自己吗，苦难本身就是社会不公的反 应。
The question is, are their poverty and suffering caused by themselves? Suffering itself is a reflection of social injustice.

不是因为单亲，是成长经历，缺爱的缺温暖的单不单亲他都敏感都关 闭都防备
It's not because of single parents, it's because of the upbringing, lack of love and warmth. Whether single (parents) or not, he is sensitive, closed, and defensive.

Adaptation Task Data

- Even though it is sampled with a similar method to the English dataset, there are key differences:
  - Sampled from social media, not news; more casual speech; often shorter phrases or comments/replies that may require prior context that is not available
  - Cultural & social differences mean that the vulnerable demographics being discussed are usually different
  - The positive samples seem to be more overtly inflammatory than in the English dataset; this might have to do with differences in annotation or social media vs. newswire

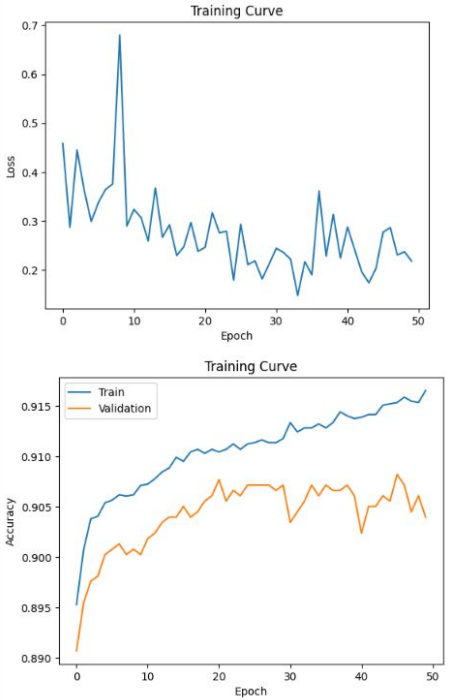
严惩农名工，还欠薪者一个公道
Severely punish migrant workers and give justice to those who owe wages

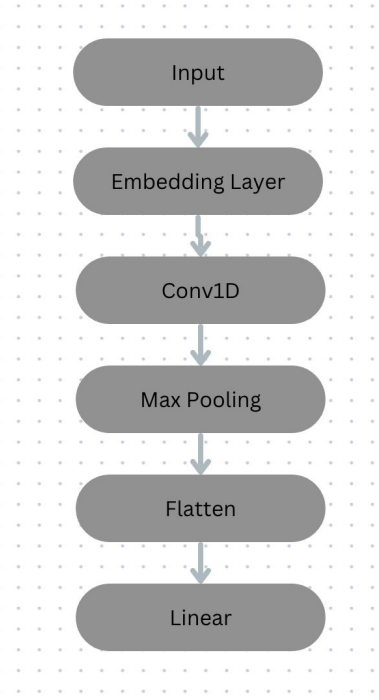
舍不得孩子就 带走呗
If you can't bear to leave your child, take him with you.

High-level approach

First approach: CNN with static embeddings

- Represent words with GloVe embeddings
- Input sequences represented with average of word embeddings
  - Problem: doesn't capture word-order and contextual information





Second approach: BERT model

- Dynamic embeddings and transformer-based models are better at handling context-dependent information
- We choose BERT for its ability to handle bidirectional context (relationships between words) and long-distance dependencies

- Goal: fine-tune a pre-trained BERT model on task-specific data
- Specific model: **BERT-base-uncased**, pre-trained in English with 12 layers, 110M parameters

Training

- Optimizer: AdamW
- Linear learning rate scheduling
- 80-20 train/validation split
- Initially trained for 4 epochs (BERT model authors recommend 2-4 for transfer learning)
- After 2 epochs, training loss continuously dropped with no change in validation accuracy; switch to 2 epochs to avoid overfitting

Adaptation: two approaches

- Translate Chinese data to English, then evaluate using our BERT model trained on the English data
- Fine-tune a Chinese BERT model directly on the Chinese data

Translation Approach

- Used Google Cloud Translate API to batch translate data
- Pros: Allows us to directly adapt our existing model to the new dataset without additional training
- Cons: MT output lacks emotional nuance for affect recognition task; different domain, topics, and culture makes it hard to adapt

text	labels
It's really tasteless. I'll unfollow you and don't need to send you any gifts.	0
Why don't you do something practical? Isn't it a shame to be a "good man" just l	0
In reality, there are many more people who are not from single-parent families a	0
"I was trying to get more attention and I accidentally got everyone's attention."	0
When I think of the recommendation to read People's Daily comments for civil s	0
The premise of discussing whether bad laws are also laws and whether bad law	0
I recall that two years ago when I was working on a project, I had a 10-yuan lun	0
Not to mention a 14-year-old child, the epidemic has lasted for such a long time	0
Does the People's Daily recruit highly educated people just to write these disgus	0
I also had my phone stolen by someone from Xinjiang. She reached into my poc	0
Are your words very worldly-wise? Haha. The truth is that there is no worldly-wis	0
Beware of foreign forces trying to get their salaries.	0
If you don't know the suffering, you won't believe in gods or Buddhas.	0
It's OK to use someone else's bank card and put the house and car in someone	0
It is suggested to update the selection criteria for the minimum living allowance,	0
This kind of marriage. In fact, it doesn't matter whether you take your children wi	1
I don't believe lies either!	0
Paying attention to homeless people starts with me	0
It can be broken, but it is difficult. You must heal yourself before getting married.	0
Many families fake divorce in order to pay less down payment for buying a hous	0
Some people don't want to give money to disadvantaged groups	0
Liu Xin was originally the person Chen Shifeng wanted to kill [Yunbei]	0
No, to be honest, I have always felt that children from single-parent families are	0
Your son is so filial, your father doesn't even want you anymore, yet you still wa	0
So you are your own God and pay yourself your own salary	0
I shouldn't advise you. If there is any problem in the future, I will blame it on you	0
Then don't think it's easy, okay?	0
asked several doctors offline, but they didn't give me any advice, and told me to	0

Training a new model

- ‘bert-base-chinese’: vocab size of 21128, 12 hidden layers

- RoFormer Chinese
  - Transformer with **rotary position embedding**
  - Encodes absolute positional information; literature review demonstrates improved performance on *longer* texts in Chinese
  - Vocab size of 50000; tokenizer uses Jieba

```
 Original:  你在说什么啊，医院食堂你都不让你穿白大褂进，得脱了挂外面
Tokenized:  ['你', '在', '说', '什么', '啊', '，', '医院', '食堂', '你', '都', '不让', '你', '穿', '白', '大', '褂', '进', '，', '得', '脱', '了', '挂', '外面']
Token IDs:  [381, 1101, 4656, 8377, 956, 5661, 11541, 34388, 381, 5034, 7045, 381, 3653, 3396, 1230, 4554, 4921, 5661, 1729, 4069, 266, 2029, 14435]
```

Results

Initial Results

Model	Precision	Recall	F1 Score
Linear NN	0.129	0.038	0.058
Conv1D NN (with undersampling)	0.084	0.343	0.135
BERT (4 epochs)	0.589	0.306	0.403
BERT (2 epochs)	0.565	0.495	0.528

Results - Primary Task

Precision	Recall	F1 Score
0.65789	0.47619	0.55248

devset

Precision	Recall	F1 Score
0.58297	0.43217	0.49637

evaltest

Results - Adaptation Task

Precision	Recall	F1 Score
0.59829	0.56451	0.58091

devset

Precision	Recall	F1 Score
0.53508	0.55707	0.54586

evaltest

Issues and solutions

Adaptation Task: Translation

- Challenges
  - A lot of nuance can be lost in translation
  - What is/isn't considered PCL depends on fine connotations.
  - Several translated samples would not be easy to label as PCL by a human annotator
  - Two approaches → 1) translated text and 2) directly working with the Chinese data..

I am a child with a family. Hehehe	*Civil servants can only make a living in Beijing*
Sorry, this disease is all about luck	*I'm afraid that it won't reach the people who really need it.*

Adaptation Task: Translation

- Challenges
  - A lot of nuance can be lost in translation
  - What is/isn't considered PCL depends on fine connotations.
  - Several translated samples would not be easy to label by PCL by a human annotator.
  - Two approaches → 1) translated text and 2) directly working with the Chinese data.

I am a child with a family. Hehehe

Civil servants can only make a living in Beijing

Sorry, this disease is all about luck

I'm afraid that it won't reach the people who really need it.

Adaptation Task: Translation

- Challenges
  - A lot of nuance can be lost in translation
  - What is/isn't considered PCL depends on fine connotations.
  - Several translated samples would not be easy to label by PCL by a human annotator.
  - Two approaches → 1) translated text and 2) directly working with the Chinese data.

Further reading: Beddiar, Djamila Romaissa, Md Saroar Jahan, and Mourad Oussalah. "Data expansion using back translation and paraphrasing for hate speech detection." *Online Social Networks and Media* 24 (2021): 100153.

→ Successful use of back-translation to generate adequate new data for hate speech detection → suggests hate speech is preserved after translation

Revised System: Data Augmentation

- Random Deletion
- Synonym Replacement

We experimented with several different combinations (only one data augmentation technique at a time, less artificial sentences, etc). → No improvement (possible reason: class imbalance in all sets).

Results better with fewer artificial samples
→ foregoing data augmentation altogether

Sample sentence: "least hungry **traumatised** refugees sought refuge bangladesh since october"

Artificial sentence: "least hungry **shock** refugees sought refuge bangladesh since october" [*traumatised* was replaced by its synonym *shock*]

Revised System: Stopwords Removal

- NLTK English stopwords
- Originally removed them, as recommended by Haddi et al, 2013: “results show the importance of removing stopwords in achieving higher accuracy in sentiment classification”. Several similar classification tasks remove them.
- Further research showed *subtle* tasks can benefit from keeping them → Saif et al “results show that using pre-compiled lists of stopwords negatively impacts the performance of Twitter sentiment classification approaches.”
- Kept them → results improved.

```
70   about
71   against
72   between
73   into
74   through
75   during
76   before
```

Further reading: Haddi, Emma & Liu, Xiaohui & Shi, Yong. (2013). The Role of Text Pre-processing in Sentiment Analysis. Procedia Computer Science. 17. 26–32. 10.1016/j.procs.2013.05.005.

Hassan Saif, Miriam Fernandez, Yulan He, and Harith Alani. 2014. On Stopwords, Filtering and Data Sparsity for Sentiment Analysis of Twitter. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 810–817, Reykjavik, Iceland. European Language Resources Association (ELRA).