Sarcasm Classification

December 6, 2022

1 Sarcasm Classification

repo: https://github.com/andrasreka/AdvancedMLProject

[7]:
```
from transformers import AutoTokenizer, AutoModelForSequenceClassification
import torch
import pandas as pd
import warnings
warnings.filterwarnings('ignore')
tokenizer = AutoTokenizer.from_pretrained("nlptown/
 ↪bert-base-multilingual-uncased-sentiment")
model = AutoModelForSequenceClassification.from_pretrained("nlptown/
 ↪bert-base-multilingual-uncased-sentiment")
df = pd.read_csv('../data/train-balanced-sarcasm.csv', encoding='utf8',␣
 ↪on_bad_lines='skip')
summary1 = pd.read_csv('1iter_summary.csv')
summary2 = pd.read_csv('2iter_summary.csv')
```

2 When Sentiment Analysis fails

[8]:
```
comments = [
    "This is a lovely book. I would recommend it to anyone",
    "This book is just a list of numbers. Useless.",
    "OMG, what a suspenseful read",
    "More interesting than the Bible",
    "This is, without a doubt, a more touching story than twilight",
    "A great read. Captivating. I couldn't put it down anymore, when I have␣
 ↪found out that 0.629 is there",
]

tokens = [tokenizer.encode(comment, return_tensors="pt") for comment in␣
 ↪comments]
results = [model(token) for token in tokens]
classes = [torch.argmax(result.logits) for result in results]
classes
```

[8]:
```
[tensor(4), tensor(0), tensor(4), tensor(4), tensor(4), tensor(4)]
```

3 Sarcasm detection is needed

[9]:
```
from simpletransformers.classification import ClassificationModel
import os
os.environ["TOKENIZERS_PARALLELISM"] = "false"

model = ClassificationModel(
    "roberta", "../models/robertatwitter_reddit/checkpoint-final",␣
 ↪use_cuda=False
)
```

[10]:
```
pred, _ = model.predict(comments)
pred
```

```
  0%|          | 0/6 [00:00<?, ?it/s]
  0%|          | 0/1 [00:00<?, ?it/s]
```

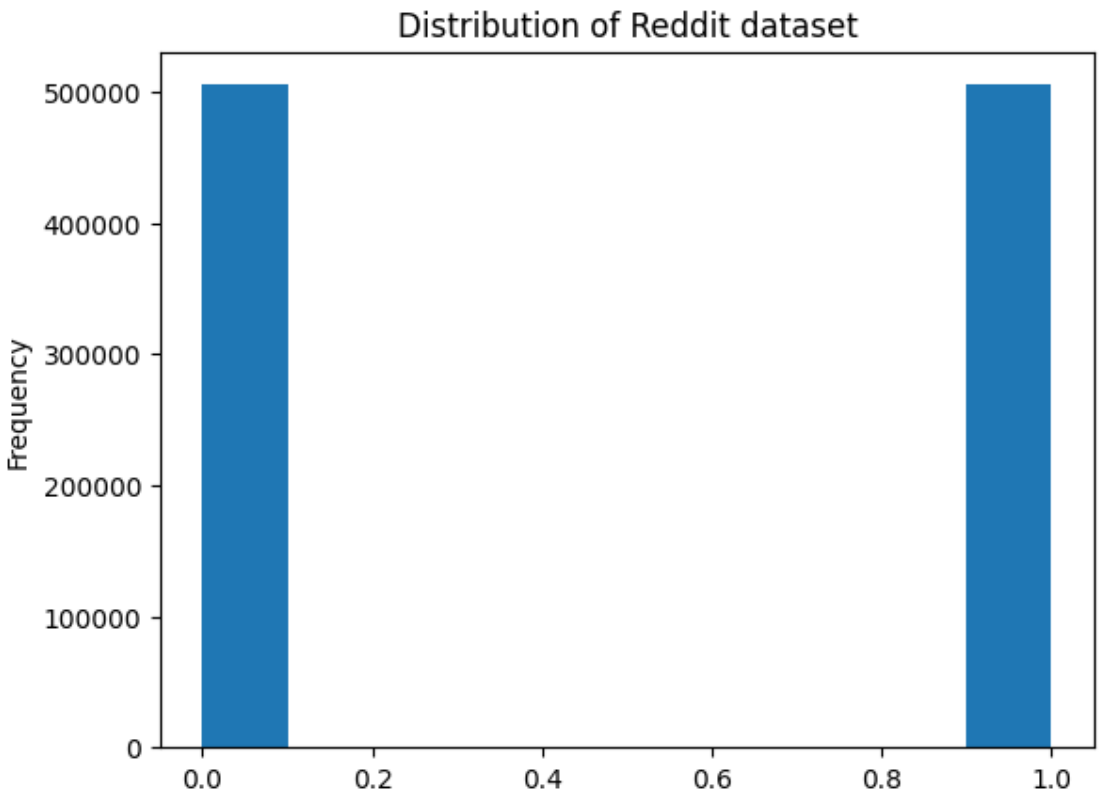
[10]:
```
[0, 0, 1, 1, 0, 0]
```

3.0.1 Dataset

- Twitter (60K data, unbalanced)
- MUSTARD (small data available)
- Reddit (1M data, balanced)

[10]:
```
df['label'].plot(kind='hist', title='Distribution of Reddit dataset')
```

[10]:
```
<AxesSubplot: title={'center': 'Distribution of Reddit dataset'},
ylabel='Frequency'>
```



3.0.2 Model

- pretrained RoBERTa transformer model
- pretrained models used:
  1. https://huggingface.co/cardiffnlp/twitter-roberta-base-irony
  2. https://huggingface.co/jkhan447/sarcasm-detection-RoBerta-base-POS

3.0.3 Hardware Spec

3.0.4 Evaluation

First Iteration

[35]:
```
summary1
```

[35]:

	Pretrained Models	Dataset	Precision	Recall	F1-Score	MCC
0	NoName	MUSTARD	0.675000	0.683544	0.679245	0.243415
1	twitter	MUSTARD	0.862069	0.316456	0.462963	0.301948

Second Iteration

[36]:
```
summary2
```

[36]:

	Pretrained Models	Dataset	Precision	Recall	F1-Score	MCC
0	twitter	twitter	0.550207	0.331998	0.414116	0.173506
1	twitter	reddit	0.677682	0.749624	0.711840	0.505483

3.0.5 Predictions

[12]:
```
reddit_test = [
"Tell her. Nothing good will come from her knowing how he feels!",
"My grandas grapevine That looks like it is growing plums",
"That might actually make me watch the World Cup, I miss Hamr. But he went down␣
 ↪as if shot",
"You can't control your thoughts, because they aren't you. But that means you␣
 ↪CAN control how seriously you take those thoughts. This is incredibly␣
 ↪freeing, especially for somebody with anxiety :) You can watch your thoughts␣
 ↪and choose how you respond to them.",
"OP never said what country. Only Americans on Reddit, right?",
"Apple's $300 book contains 450 photos of Apple products I'm not even mad,␣
 ↪that's amazing!",
"Are you high? Yes",
"When the new Widow is too strong Just pick Winston guys",
"I think a soccer team in Cape Town took ""The Dangerous Darkies"". That is␣
 ↪rather funny and good on them for owning that.",
"How's your life like after discovering Reddit? I've become more enlightened␣
 ↪than the people I surround myself with",
"Best feature in 7.00 yet how do u select to place wards and predict enemy␣
 ↪lanes",
"I need some shows to binge. What do you recommend? I prefer shows like␣
 ↪(Friends, Parks &amp;amp; Rec, The Big Bang Theory, House of Cards, Unbreakable␣
 ↪Kimmy Schmidt). I hope I'm not breaking any sub rules with this question!␣
 ↪Frasier!",
"This intellectual would pour water over fluoroantimonic acid. It's ""add water␣
 ↪to acid"", right?",
"Hate to be that guy but *MG-34 Nah it's cool man",
"One last iconic death... Oh god not Kanye!",
"Snow, forest and mountains. I found Canada. Not enough liberals",
"Take that, flat earthers! ""It was taken using a wide angle lens!!""",
"What do you enjoy that Reddit absolutely shits on? Reposting!",
"he was a fucking white mailman Next they'll be appropriating Iron Man to␣
 ↪Female!",
"Danny Rose is that dude That dude who's the best leftback in the league?",
"It's a neo nazi, I wouldn't be surprised if he was giggling for a week over␣
 ↪killing that person, those assholes are barely human beings. You're pretty␣
 ↪fucking biased ain't you ?",
"Django Unchained. That shootout scene with 2pac/James Brown song is fire",
```

```
"You know that there's this crazy thing called ""not bringing up politics␣
 ↪everywhere,"" right? Everyone's doing it. Maybe you should, too. Maybe if we␣
 ↪just ignore the massive, multi thousand nuke armada that can destroy the␣
 ↪entire earth 5 times over, it'll go away!",
"I'm 24 and still living with my parents. This is what I got for Christmas␣
 ↪Subtle",
"Captain America: Civil War Now Streaming! MEEEEERRY CHRISTMAS!",
"Her responses to this were awesome imo. I never really liked her on the show␣
 ↪but when he's off she seems like a pretty cool person. I hope they make it!␣
 ↪your pronouns confuse me",
"I completely agree with the disgruntled masses. Getting **free** content on a␣
 ↪**free** to play game ***during the holidays*** is excruciatingly unpleasant.
 ↪ You've ruined my winter break. Now, I'll fly back to uni in shambles, fail␣
 ↪my classes, and get kicked out of my program- and it's all your fault. Shame␣
 ↪on you, Digital Extremes. Yeah, because they're releasing this from the␣
 ↪goodness of their little hearts, right?",
"How is trump supposed to help black people? Serious question. What do they␣
 ↪have to lose?"
]

true_values = [
    1,
    0,
    0,
    0,
    1,
    1,
    0,
    1,
    0,
    1,
    0,
    0,
    1,
    0,
    1,
    1,
    1,
    1,
    1,
    0,
    0,
    0,
    1,
    1,
    0,
    0,
```

```
    1,
    1
]

pred, _ = model.predict(reddit_test)
print(true_values, "\n", pred, '\n{} percent of the sentences was correctly␣
 ↪classified'.format(sum([1 if real == pred else 0  for real,pred in␣
 ↪zip(true_values, pred)])/len(true_values)))
```

```
  0%|          | 0/28 [00:00<?, ?it/s]

  0%|          | 0/3 [00:00<?, ?it/s]

[1, 0, 0, 0, 1, 1, 0, 1, 0, 1, 0, 0, 1, 0, 1, 1, 1, 1, 1, 0, 0, 0, 1, 1, 0, 0,
1, 1]
 [1, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 0, 1,
1, 0]
0.7142857142857143 percent of the sentences was correctly classified
```

3.1 References

- S. K. Bharti, R. K. Gupta, P. K. Shukla, W. A. Hatamleh, H. Tarazi, S. J. Nuagah: Multimodal Sarcasm Detection: A Deep Learning Approach, Wireless Communications and Mobile Computing, Article ID 1653696 (2022).
- S. Oprea, W. Magdy: iSarcasm: A Dataset of Intended Sarcasm, arXiv:1911.03123 (2020).
- H. Yaghoobian, H. R. Arabnia, K. Rasheed: Sarcasm Detection: A Comparative Study, arXiv:2107.02276 (2021).
- A. Kumar, V. Anand: Transformers on Sarcasm Detection with Context, In Proceedings of the Second Workshop on Figurative Language Processing, pages 88–92, Online. Association for Computational Linguistics (2020).
- https://www.projectpro.io/article/bert-nlp-model-explained/558

Special thanks for the Romanian Orthodox Church for (unknowingly) providing computational resources

http://museikon.ro