Irony Detection

Pierre EPRON, Maxime RENARD, Shu ZHANG

Reminder

- Goal: to detect irony
- Using a perspective oriented dataset [1]
- Using Sequence classification [2][3] and Causal language modeling [4] models

Results

Batch Size	0			1			Macro			MCC
	P	R	F1	P	R	F1	P	R	F1	
BCE	0.819	**0.778**	0.798	0.686	0.739	0.712	0.753	0.758	0.755	0.511
MCC	**0.856**	0.776	**0.814**	**0.701**	**0.800**	**0.747**	**0.778**	**0.788**	**0.780**	**0.567**

Table: Results for Binary Cross Entropy (BCE) and Matthew Coefficient Correlation (MCC) loss using twitter-roberta-large. Batch size: 64. Starting learning rate: 1e-5. Optimizer: AdamW

Issues with scores

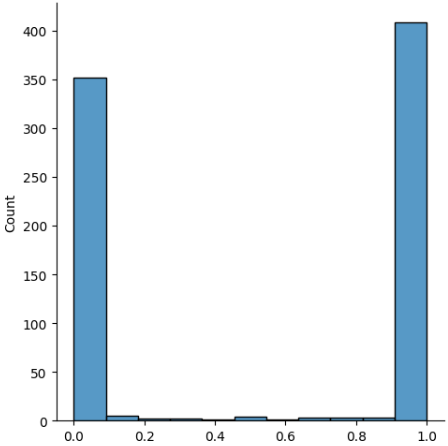

Figure: Distribution of classification scores for MCC loss

Issues with scores : BCE

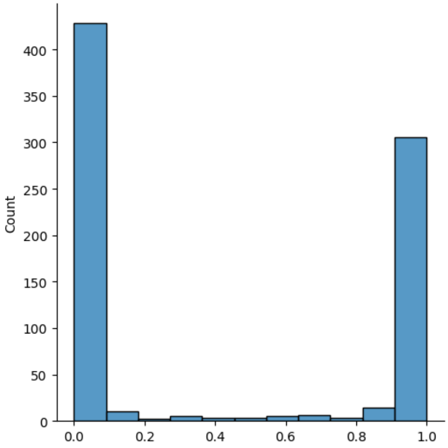

Figure: Distribution of classification scores for BCE loss

Issues with scores

- We used validation loss and a patience of 5 to prevent overfitting
- At least BCE should not behave like this.

HF Space

huggingface.co/spaces/MR17u/tweeteval-irony-mcc

HF Dataset

huggingface.co/datasets/MR17u/tweeteval-irony-mcc

Examples

index	Entry	Label	Score
0	Dentist with cavities	irony	0.9999974966049194
1	devil reading bible	irony	0.999997615814209
2	What I do is for your own good.	non_irony	0.999996542930603
3	The child of a formula one driver not being able to drive.	irony	0.9999977350234985
4	A pair of star-cross'd lovers take their life.	non_irony	0.9999922513961792
5	You use my money to help others, you are so generous	irony	0.9999977350234985
6	You are so beautiful today, like a witch	non_irony	0.9999969005584717
7	You treat me like Snow White's stepmother treats Snow White	non_irony	0.9999970197677612
8	hate	non_irony	0.9999958276748657
9	The weasel greets to the chicken	non_irony	0.9999961853027344

Figure: Common Examples

Lexical

index	Entry	Label	Score
0	you look so beautiful like princess	non_irony	0.9999366998672485
1	you look so beautiful like a princess	non_irony	0.9996936321258545
2	you look so beautiful like the princess	irony	0.9999954700469971
3	you look so beautiful as princess	irony	0.9999972581863403
4	you look so beautiful as a princess	irony	0.999997615814209
5	you look so beautiful as the princess	irony	0.999997615814209
6	you are so beautiful like princess	non_irony	0.9999963045120239
7	you are so beautiful like a princess	non_irony	0.9999961853027344
8	you are so beautiful like the princess	non_irony	0.999140739440918
9	you are so beautiful as princess	irony	0.999995231628418
10	you are so beautiful as a princess	irony	0.9999973773956299
11	you are so beautiful as the princess	irony	0.9999948740005493
12	you are so beautiful like shit	non_irony	0.9999957084655762
13	you are so beautiful like a shit	non_irony	0.9999971389770508
14	you are so beautiful like the shit	non_irony	0.9999971389770508
15	you are so beautiful as shit	non_irony	0.9999969005584717
16	you are so beautiful as a shit	non_irony	0.9999938011169434
17	you are so beautiful as the shit	non_irony	0.9999971389770508
18	you look so beautiful like the shit	non_irony	0.9999972581863403

Figure: Different Lexical Examples

Different synonyms, antonym, preposition, phrase give different could not always but sometimes impact the results .

Punctuation

index	Entry	Label	Score
0	Cat greet mouse	non_irony	0.9999253749847412
1	Cat greets mice	irony	0.9999850988388062
2	Cat greets mice	non_irony	0.9998910427093506
3	Cat greets mice	irony	0.9999850988388062
4	Cat /greets mice	irony	0.999997615814209
5	Cat/ greets mice	non_irony	0.9999551773071289
6	Cat+ greets mice	irony	0.9999961853027344
7	cat greets mice	irony	0.9999932050704956
8	cat greets mice	non_irony	0.9999966621398926
9	cat greets mice	irony	0.9999912977218628
10	Cat greet mouse	non_irony	0.9999253749847412
11	Cat greet 'mouse'	non_irony	0.9999961853027344
12	you are pretty	non_irony	0.9999960660934448
13	you are 'pretty'	irony	0.9999972581863403

Figure: Different Punctuation Examples

Different punctuation marks in different positions will affect the results.
Note: [N.17] 2 spaces between 'cat' and 'greets', [N.18] 2 spaces between 'greets' and 'mice', [N.19] extract space after 'mice'

Plural & Letter Case

index	Entry	Label	Score
0	Cat greet mouse	non_irony	0.9999253749847412
1	Cat greets mouse	non_irony	0.998047947883606
2	Cats greets mouse	non_irony	0.9999957084655762
3	Cat greet mice	irony	0.9999960660934448
4	Cats greets mice	irony	0.9999963045120239
5	cats greets mice	irony	0.9999959468841553
6	CATS greets mice	irony	0.9999973773956299
7	Cat greets mice	irony	0.9999973773956299
8	cat greets mice	irony	0.9999974966049194
9	CAT GREETS MICE	non_irony	0.9999966621398926
10	CAT GREETS MOUSE	non_irony	0.9999959468841553

Figure: Different Plurals and Letter Cases

The plural form of nouns has a greater impact on the results than the verbs, and capitalizing the first letter of the sentence has little impact on the overall results. However, if the sentence contains multiple capital letters, the results will be affected.

Emojis

index	Entry	Label	Score
0	you look pretty	non_irony	0.999914288520813
1	you look pretty 🤢	irony	0.9997101426124573
2	you look pretty🤢	non_irony	0.9987372756004333
3	you look pretty ❤️	non_irony	0.9999973773956299
4	you look pretty ♥	non_irony	0.9999973773956299
5	you look pretty 😊	non_irony	0.9999973773956299
6	you look pretty 😍	non_irony	0.9999973773956299
7	you are so beautiful as the bitch	irony	0.9999967813491821
8	you are so beautiful as the bitch 🌈	non_irony	0.9999887943267822
9	you are so beautiful as the bitch ❤️	non_irony	0.9999973773956299
10	you are so beautiful as the bitch 🤢	irony	0.999996542930603

Figure: Some Examples with Emojis

There should be a space between the emoji and the previous token. No space will produce different results. Maybe emojis are divided into positive and negative, which will have different effects on different sentences.

Hypothesis

index	Entry	Label	Score
0	Beijing is the capital of China	irony	0.9999977350234985
1	capital	non_irony	0.999982476234436
2	Beijing	irony	0.999996542930603
3	Peking	irony	0.9999945163726807
4	China	irony	0.9999948740005493
5	Chine	irony	0.9999966621398926
6	Pairs is the capital of France	irony	0.9999977350234985
7	America is safer than China	irony	0.9999974966049194
8	China is large	irony	0.9999974966049194
9	China occupies the third largest area in the world	irony	0.9999973773956299
10	Trump is a nice guy	irony	0.999997615814209
11	trump	irony	0.9084716439247131
12	Macron never lies everybody know that	irony	0.9999974966049194
13	Macron	irony	0.9999960660934448
14	New York	irony	0.9999960660934448
15	Washington	non_irony	0.9999954700469971
16	Washington D.C.	non_irony	0.7705661654472351
17	Washington D.C. capital	irony	0.910758912563324
18	Washington is the capital of the United States	irony	0.9999974966049194
19	Washington is the capital of the USA	irony	0.999997615814209

Figure: Different Examples with Different tokens

We assume that if a single token is ironic, then all the sentences containing this token will be ironic, we are not sure because we can't prove it.

Old results

Batch Size	0			1			Macro			MCC
	P	R	F1	P	R	F1	P	R	F1	
close	0.721	0.126	0.214	0.319	**0.894**	0.470	0.520	0.509	0.343	0.028
open	**0.751**	**0.275**	**0.402**	**0.439**	0.862	**0.581**	**0.595**	**0.568**	**0.492**	**0.161**

Table: Results for tweeteval test set for open and close generation on llama2-7b-chat-hf

Parameter-Efficient Fine-Tuning (PEFT)

- The main idea is to finetune models using by updating only few additional paramaters.
- At least 11 differents methods to do it.
- We tried **Prompt tuning** [5] which is a "simplification" on another method: **Prefix tuning** [6]

Parameter-Efficient Fine-Tuning (PEFT)

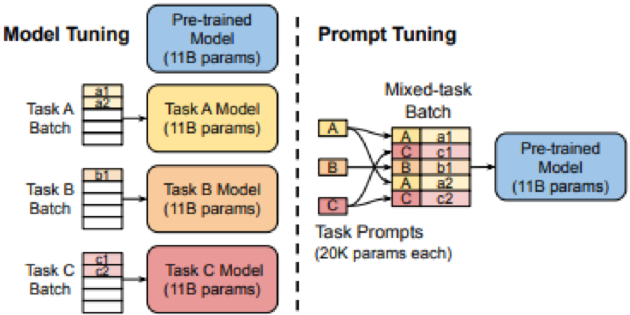

Figure: Difference between classic fine-tuning and prompt-tuning

LLama and Peft

```
['no ironic',
 'no ironic',
 'iris',
 'no ironic',
 'ironic',
 'ironic',
 'no ironic',
 'no ironic',
 'ironic',
 'no iris no iris',
 'no ironic',
 'ironic',
 'ironic',
 'no ironic',
 'no ironic',
 'ironic',
 'ironic',
 'iris',
 'iris',
 'ironic',
```

Conclusion

- Fixing our PEFT experiment
- Use it on a space to see if even if it have the same behavior

References I

[1] Simona Frenda et al. "EPIC: Multi-Perspective Annotation of a Corpus of Irony". In: *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Toronto, Canada: Association for Computational Linguistics, July 2023, pp. 13844–13857. DOI: 10.18653/v1/2023.acl-long.774. URL: https://aclanthology.org/2023.acl-long.774.

[2] Yinhan Liu et al. "RoBERTa: A Robustly Optimized BERT Pretraining Approach". In: *ArXiv* abs/1907.11692 (2019). URL: https://api.semanticscholar.org/CorpusID:198953378.

[3] Francesco Barbieri et al. "TweetEval: Unified Benchmark and Comparative Evaluation for Tweet Classification". In: *Findings of the Association for Computational Linguistics: EMNLP 2020*. Online: Association for Computational Linguistics, Nov. 2020, pp. 1644–1650. DOI: 10.18653/v1/2020.findings-emnlp.148. URL: https://aclanthology.org/2020.findings-emnlp.148.

References II

[4] Hugo Touvron et al. "Llama 2: Open Foundation and Fine-Tuned Chat Models". In: *ArXiv* abs/2307.09288 (2023). URL: https://api.semanticscholar.org/CorpusID:259950998.

[5] Brian Lester, Rami Al-Rfou, and Noah Constant. "The Power of Scale for Parameter-Efficient Prompt Tuning". In: *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. Online and Punta Cana, Dominican Republic: Association for Computational Linguistics, Nov. 2021, pp. 3045–3059. DOI: 10.18653/v1/2021.emnlp-main.243. URL: https://aclanthology.org/2021.emnlp-main.243.

[6] Xiang Lisa Li and Percy Liang. "Prefix-Tuning: Optimizing Continuous Prompts for Generation". In: *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Online: Association for Computational Linguistics, Aug. 2021, pp. 4582–4597. DOI: 10.18653/v1/2021.acl-long.353. URL: https://aclanthology.org/2021.acl-long.353.

Thanks

Thanks for watching!