IUST at ClimateActivism 2024: Towards Optimal Stance Detection: A Systematic Study of Architectural Choices and Data Cleaning Techniques

Ghazaleh Mahmoudi, Sauleh Eetemadi

February 2024

Overview

- Climate change is a pressing issue affecting ecosystems, economies, and communities globally.
- Understanding public perspectives on climate change is crucial.
- Leveraging NLP techniques to analyze public Stance Toward Twitter offers real-time insights.
- The ClimateActivism 2024 Shared Task focusing on Stance and Hate Event Detection.

Data

Data Preprocessing

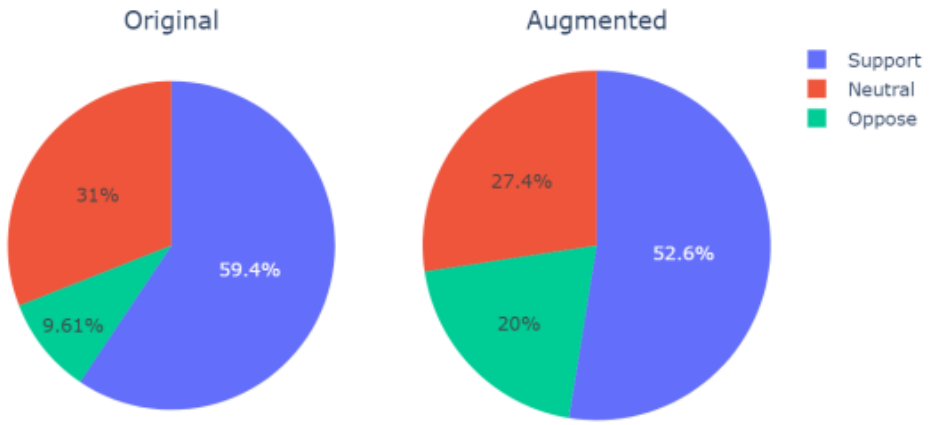
- Original Tweet
- Removing URL
- Removing Username
- Removing URL and Username
- Removing URL and Username and split hashtag
- Removing URL and Username and split hashtag and lower case
- Complete Cleaning

Data

Data Augmentation

To address data imbalance challenge we two different methods to generate additional data.

- Substitution
  - use synonym substitution as an augmentation method.
- Round-trip translation
  - translate the English texts to German and then back, to generate extra data.



Methodology

- proposed a model with four modules
    - conducted multiple experiments to find the best parameters for each module.
- The Optuna library was used to select the optimal model configuration based on the Macro F1-score.
- The search space for each module was defined to find the most suitable values.

Methodology

Parameter	Search Space
Embedding	BERT, RoBERTa, BERTweet, XLM-RoBERTa
Classifier	CNN, FNN
Optimizer	Adam, AdamW, RMSPropb, SGD
Loss Function	Weighted Cross Entropy, Focal

Architecture search space

Methodology

Hyperparameter	Search Space
Dropout	[0.1: 0.5]
Learning Rate	[0.00001: 0.01]
Batch Size	[4, 8]
Focal Gamma	[1, 2, 3, 4, 5]

Hyperparameters search space

Result

	Parameter	Value
Model Configuration	Epoch	8
	Batch Size	4
	Dropout	0.5
	Learning Rate	0.007903
	Learning Rate Scheduler	Linear Scheduler with Warmup
Hyperparameters	Embedding	BERTweet
	Classifier	CNN
	Optimizer	SGD
	Loss Function	Weighted Cross Entropy

Best model configuration and hyperparameters.

Result

Rank	Team Name	Codalab Username	Accuracy	Precision	Recall	F1-score
1	ARC-NLP (Kaya et al., 2024)	kagankaya1	**74.90**	**78.48**	**72.26**	**74.83**
2	HAMiSoN-Generative (Fraile-Hernandez and Peñas, 2024)	JesusFraile	74.78	78.27	72.23	74.79
3	IUST (Mahmoudi and Eetemadi, 2024)	gh_mhdi	73.11	78.63	71.45	74.47
4	HAMiSoN-MTL (Rodriguez-Garcia and Centeno, 2024)	Raquel	74.33	77.02	72.42	74.02
5	AAST-NLP (El-Sayed and Nasr, 2024b)	AhmedElSayed	74.39	79.31	70.78	73.98
6	MasonPerplexity (Gangul et al., 2024)	Sadiya_Puspo	73.69	77.80	70.90	73.73
7	-	kojiro000	73.43	77.44	70.89	73.58
8	-	refaat1731	72.22	77.49	70.06	73.15
9	HAMiSoN-baselines (Montesinos and Rodrigo, 2024)	julioremo	74.01	78.17	70.36	73.13
10	-	Nikhil_7280	71.90	76.62	68.13	70.81
11	-	swatirajwal	67.86	70.83	70.05	70.26
12	Bryndza (Suppa et al., 2024)	mareksuppa	71.19	68.72	71.23	69.33
13	NLPDame (Christodoulou, 2024)	christiechris	66.52	71.16	67.94	69.30
14	byteSizedLLM	mdp0999	65.24	72.55	66.85	69.10
15	CUET_Binary_Hackers (Farsi et al., 2024)	SalmanFarsi	66.13	69.08	66.91	67.94
16	Z-AGI Labs (Narayan and Biswal, 2024)	mrutyunjay_research	69.08	79.26	62.94	63.72
17	Team +1	pakapro	32.71	32.66	31.51	28.98
18	-	ankitha11	0.38	1.32	0.16	0.29
19	pokemons	md_kashif_20	0.00	0.00	0.00	0.00

Result

Cleaning	F1-Score
C1	73.98±0.0012
C2	73.92±0.0017
C3	74.35±0.0015†
C4	74.11±0.0029†
C5	73.76±0.0014
C6	73.72±0.0009
C7	72.42±0.0020

Experiment with Best Model Configuration and hyperparameter for data cleaning impact. † indicates significance ($p < 0.005$) comparing to C1.

Result

Classifier	Cleaning	F1-Score
CNN	C3	74.35±0.0015†
	C4	74.11±0.0029†
FNN	C3	73.73±0.0058
	C4	73.91±0.0045

Classifier impact . † indicates significance ($p < 0.005$) comparing to FNN.

Conclusion

- This work involved a systematic exploration of model architecture and data cleaning methods.
- We find that the optimal configuration combining BERTweet and CNN with Weighted Cross Entropy and SGD, along with data augmentation.
- We demonstrate that a combination of CNN and Encoder-only models such as BERTweet outperforms FNNs. Moreover, by utilizing data augmentation, we are able to overcome the challenge of data imbalance.
- Our best system achieves 74.47% F1- Score on the unseen test set
- outperforming the baseline by 19.97%
- ranked 3th among 19 participants.

Thank you

gh_mahmoodi@comp.iust.ac.ir

sauleh@comp.iust.ac.ir