X-PuDu at SemEval-2022 Task 6: Multilingual Learning for English and Arabic Sarcasm Detection

Yaqian Han, Yekun Chai, Shuohuan Wang, Yu Sun
Hongyi Huang, Guanghao Chen, Yitong Xu, Yang Yang

2022-6-21

Motivation

- Sarcasm is omnipresent on the social web.

- Due to its nature, can be highly disruptive of computational systems that harness this data to perform tasks

- It is imperative to devise models for sarcasm detection.

SemEval-2022 Task 6

- This SemEval task requires the identification of sarcasm in either one sentence or sentence pairs in various language settings

Task	Tweets	Languages	Labels
SubTask A	I want to see Drew Lock cry.	English and Arabic	non-sarcastic
SubTask B	Lil Pump is the Nelson Mandela of our generation.	English only	#Satire #Sarcastic
SubTask C	First: Trying to know all this history tonight is gonna kill me. Second: Trying to know all this history is going to be be a challenge.	English and Arabic	First

Methods

- Multilingual Learning

    - SubTask A and C are for the same objective but in different languages

    - Based on multilingual pre- trained models

- Ensemble Learning

    - Based on 10-fold cross-validation for all subtask

Pre-trained Language Models

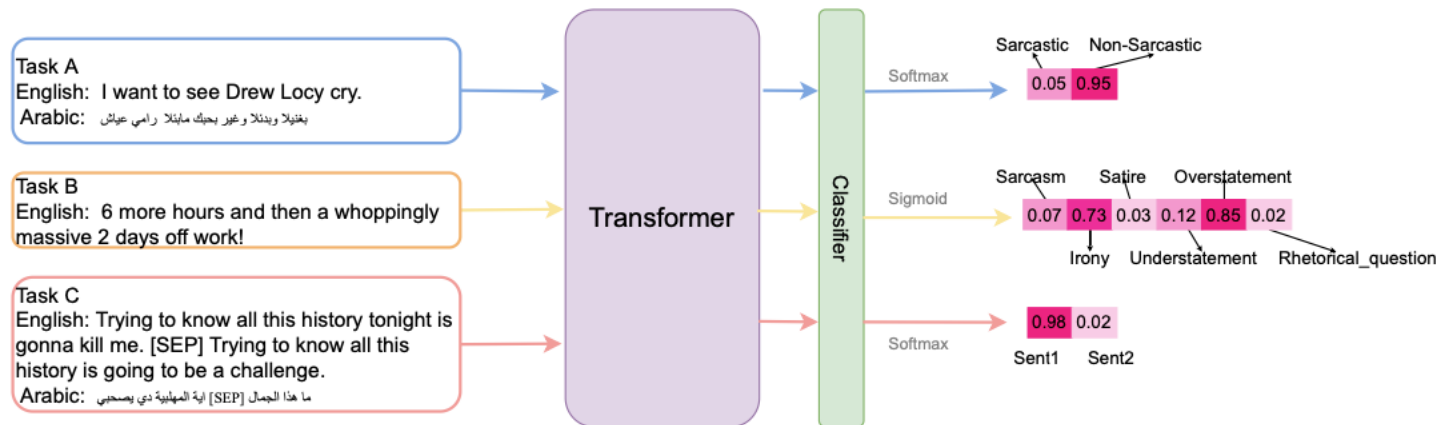- Using ERNIE-M and DeBERTa as ours pre-trained models

Figure 1: Fine-tuning pre-trained models on the *iSarcasmEval* data.

Multilingual Learning
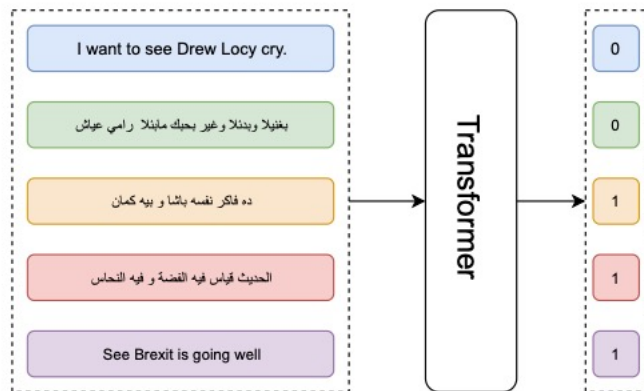
- Multilingual learning for SubTask A and SubTask C

Figure 2: Multilingual learning on Task A. "0/1" indicate the non-sarcastic and sarcastic class.
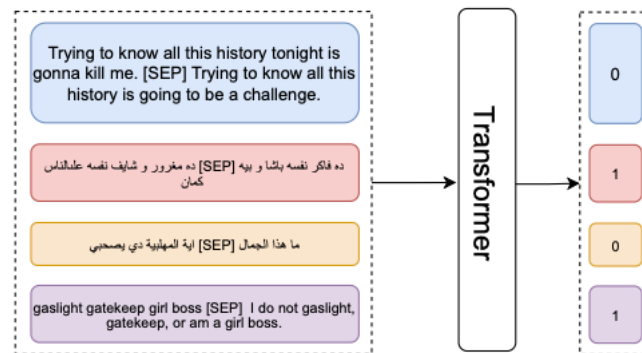

Figure 3: Multilingual learning on Task C. "0/1" indicate the first or second sentence belongs to sarcasm.

Ensemble Learning

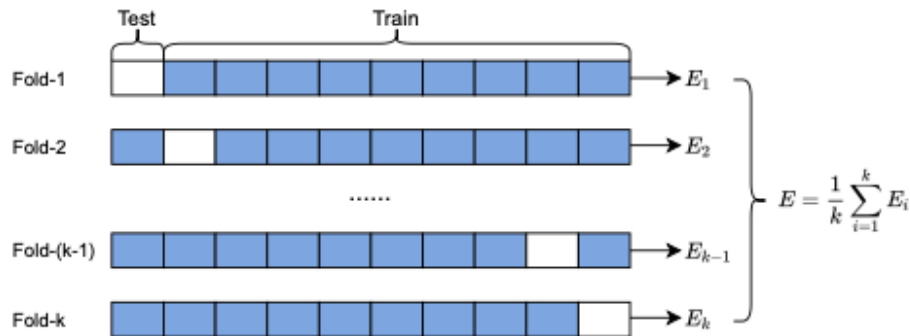- k-fold cross-validation for training to improve the robustness of our model

Figure 4: Illustration of ensemble learning. E_i indicates the prediction of the i-th model on the test set.

Datasets

- ## Evaluation Metrics

Task	#Instances	#Metric
Task A	1400	F1-score
Task B	1400	Macro-F1 score
Task C	200	accuracy

Table 1: Summary of official test set in SemEval-2022 Task6.

- ## Data Details

Class Label	#Instances
sarcastic	867
non-sarcastic	2601
total	3468

Table 2: Satirical and non-satirical categories in training data.

Multi-class Label	#Instances
sarcasm	713
irony	155
satire	25
understatement	10
overstatement	40
rhetorical_question	101

Table 3: Six satirical sub-categories in Task B.

Results

- The multilingual learning can achieve obvious performance gain on SubTask C in Arabic.

- Our best multilingual performance can achieve 2nd in SubTask C (Arabic), due to time constraints, it was not submitted.

Task	Lang	ERNIE-M (multilingual)	ERNIE-M (monolingual)	DeBERTa	Rank
Task A	en	36.75	38.46	**56.91**(∗)	2/43
	ar	40.36	**41.87**(∗)	-	9/32
Task B	en	N/A	-	**7.99**(∗)	5/22
Task C	en	82.50	75.00	**87.00** (∗)	1/16
	ar	**90.50**	84.00(∗)	-	5/13

Table 4: Official test-set performance under various experimental settings. The "ERNIE-M (multilingual)" column indicates the performance of multilingual learning in Task A and C. Scores with asterisk indicate final submitted results. The official evaluation metrics for Task A,B,C are F1-score, macro F1-score, and accuracy, respectively.

Task C - Arabic					
#	User	Entries	Date of Last Entry	Team Name	Accuracy ▲	F-Score ▲
1	lizefeng	2	01/30/22		0.9300 (1)	0.9300 (1)
2	AlamiHamza	1	01/31/22		0.8850 (2)	0.8848 (2)

Conclusion & Future

- Conclusion

  - Proposed multilingual learning method to train the English and Arabic tasks jointly

  - Outranking the monolingual pretrained models on Arabic tasks.

- Future Work

  - We will explore different sarcasm detection approaches under multilingual settings.

THANKS