Text Based Cross-Lingual Emotion Detection

Cristian Daniel Paduraru

1. Introduction

Emotion Detection (ED) task:

- for a given text sample (sentence, paragraph, sequence of messages), determine which emotions is the author experiencing or conveying
- the set of possible emotions is predefined

1. Introduction

- Emotion Detection has received a lot of attention in the recent years

1. Introduction

- Emotion Detection has received a lot of attention in the recent years
- Largest gains in performance are thanks to the rise of pre-trained Large Language Models (LLMs)

2. Task and Data

Cross-Lingual Emotion Detection:

2. Task and Data

Cross-Lingual Emotion Detection:

- texts from a set of **source languages** are annotated for ED
- the goal is to obtain an emotion detector for a new **target language** which has no annotated data for this task

2. Task and Data

The data used in this project comes from Task 11 of SemEval 2025.

2. Task and Data

The data used in this project comes from Task 11 of SemEval-2025.

The set of emotions to be recognised are: anger, disgust, fear, joy, sadness and surprise

Designed setup:

- Use English (eng)*, German (deu) and Spanish (esp) as **source languages**
- Test on Romanian (ron), Ukrainian (ukr) and Hindi (hin) as **target languages**

*English texts are not labeled for **disgust**

3. Approach & Results

3. Approach & Results

- the centerpiece of the approach is a Cross-Lingual LLM capable of processing all the source and target languages

3. Approach & Results

- the centerpiece of the approach is a Cross-Lingual LLM capable of processing all the source and target languages
- assuming that the LLM provides language-agnostic representations for the input texts, then a detector trained on the source languages should generalize to the target languages as well

3. Approach & Results

a. LEALLA-large: linear probing (lp)

Source languages	Val acc.	Target language		
		ron	ukr	hin
eng	53.49	41.87	18.12	27.16
deu	46.37	45.89	20.89	26.56
esp	59.23	42.89	**26.50**	**44.43**
eng, deu, esp	52.94	**46.97**	23.58	32.68

Table 1: Results on the dev set of target languages for the linear classifiers trained on embeddings from the LEALLA-large model.

3. Approach & Results

b. LEALLA-large: linear probing and fine-tuning (LP-FT)

Source languages	Val acc.	Target language		
		ron	ukr	hin
eng, deu, esp	70.93	53.72	36.32	53.17

Table 2: Results of the finetuned LEALLA-large model on the dev set of target languages. The model checkpoint is selected based on the validation macro F_1 score.

3. Approach & Results

c. QWEN2.5 7B: linear probing

Source languages	Val acc.	Target language		
		ron	ukr	hin
eng	46.01	36.69	14.53	19.35
deu	40.44	43.94	15.82	22.60
esp	55.58	44.10	**16.79**	21.39
eng, deu, esp	52.03	**45.21**	16.67	**23.46**
eng*	41.04	37.26	**16.85**	16.78
deu*	37.85	31.48	11.69	19.08
esp*	52.66	28.81	15.52	18.90
eng, deu, esp*	50.80	**37.35**	15.16	**21.43**

Table 3: Results on the dev set of target languages for the linear classifiers trained on embeddings from the QWEN2.5 model. The mark * indicates results for the sklearn implementation of logistic regression.

3. Approach & Results

c. QWEN2.5 7B: PCA & linear probing

#Principal components	Val acc.	Target language		
		ron	ukr	hin
64	44.05	27.38	15.53	**23.93**
128	45.17	29.09	16.17	20.96
256	45.97	34.15	**17.64**	19.94
3584	52.03	**45.21**	16.67	23.46

Table 4: Results on the dev set of target languages for the linear classifiers trained on embeddings from the QWEN2.5 model (from all source languages) after dimensionality reduction with PCA.

4. Discussion - Distribution Shift

4. Discussion - Distribution Shift

Model selection

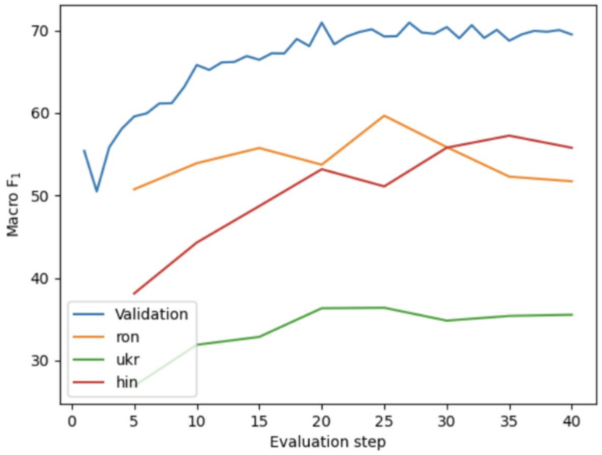
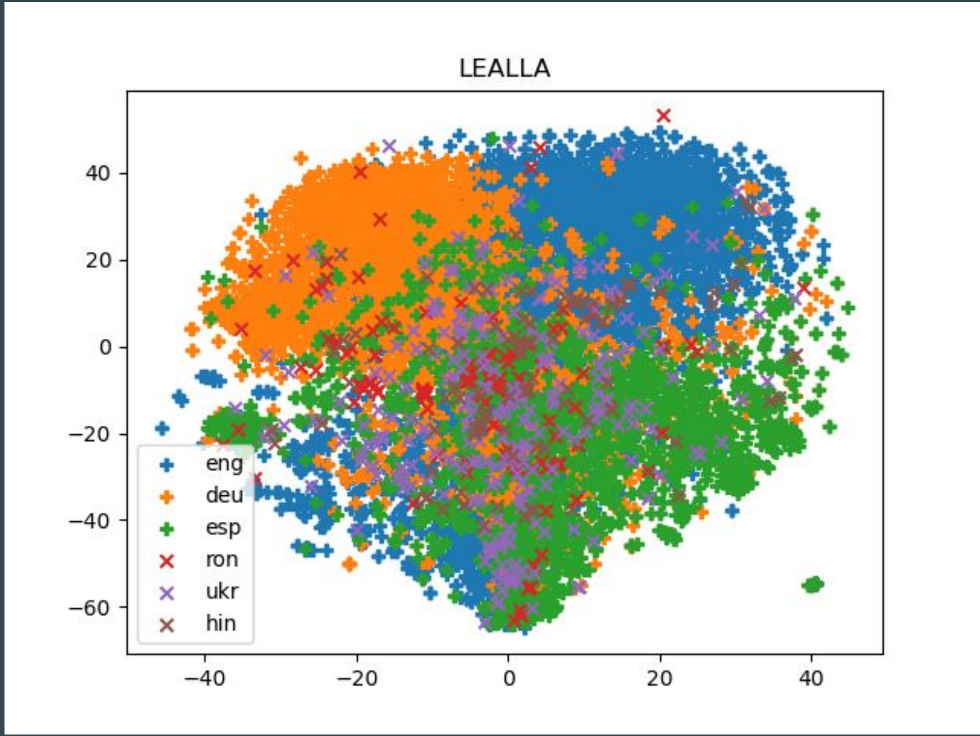


Figure 1: Macro F_1 scores on the validation set of source languages and the dev set of target languages when fine-tuning the LEALLA-large model on all the source languages.

4. Discussion - Distribution Shift

Embedding space

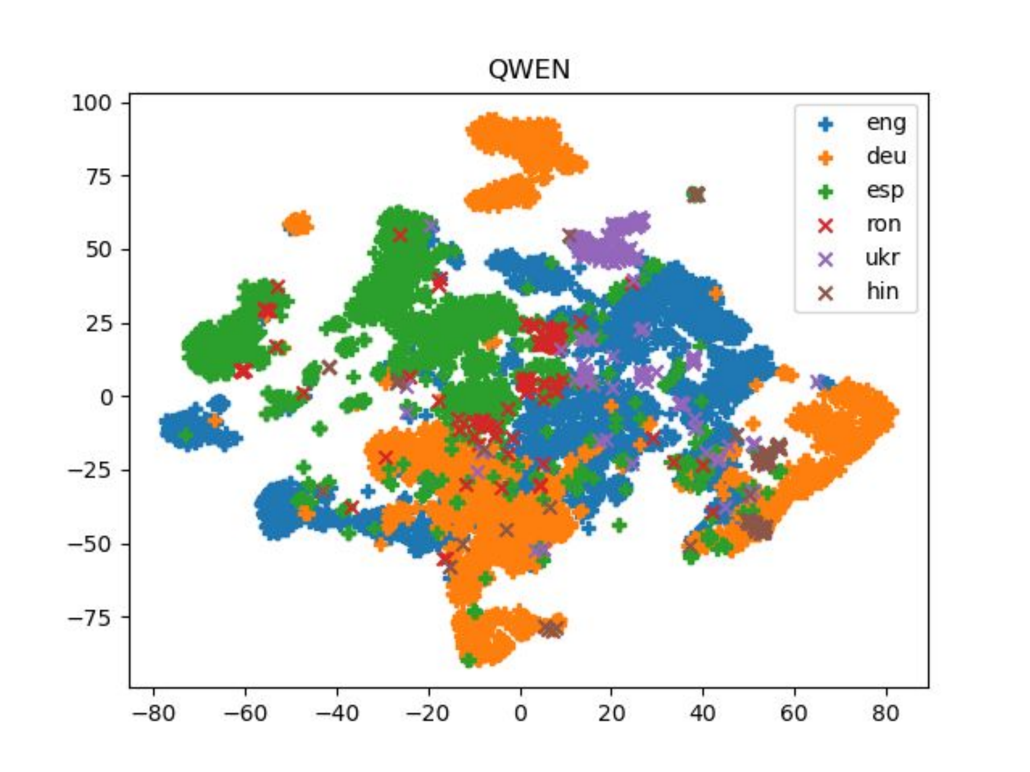
t-SNE visualization



4. Discussion - Distribution Shift

Embedding space

t-SNE visualization



4. Discussion - Distribution Shift

Topics covered in the data

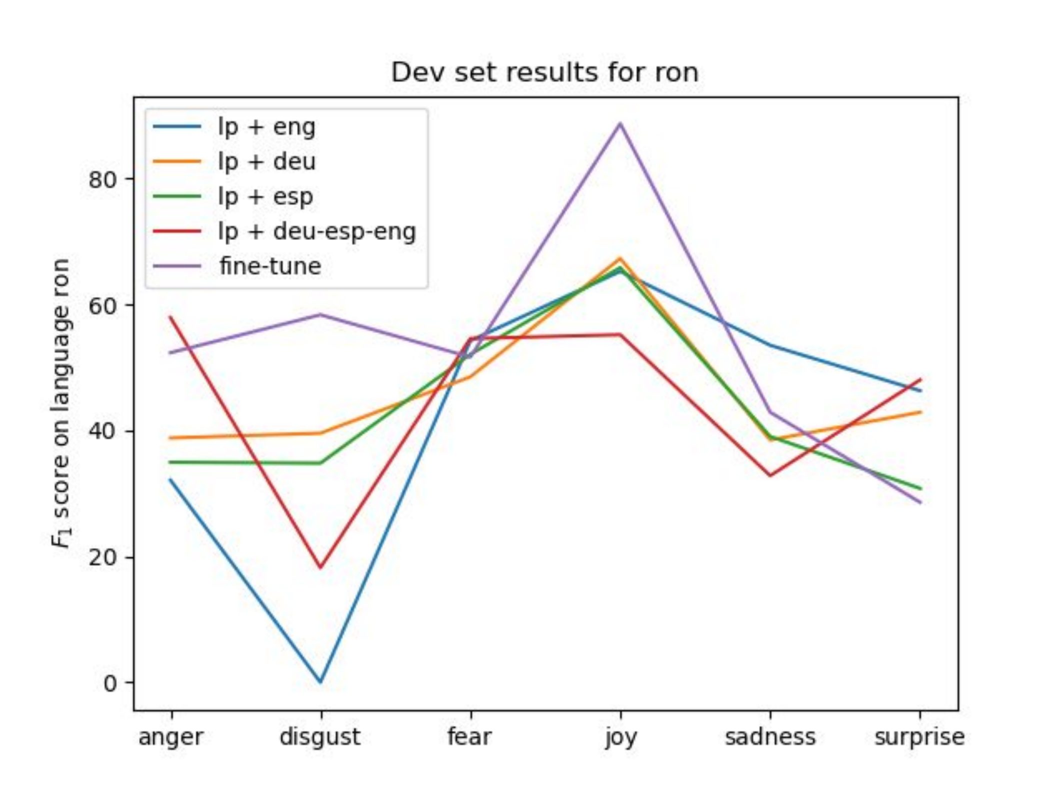
- I manually looked at the samples from the Romanian and English datasets and concluded that the texts were most likely retrieved from news websites and social media
- Romanian texts were mostly about political debates or the recent COVID-19 pandemic
- the English ones where a lot more diverse, covering stories and experiences of people

4. Discussion - Per emotion performance

For each language I observed a trend where certain emotions are better recognized than the others, irrespective of the source languages used for training.

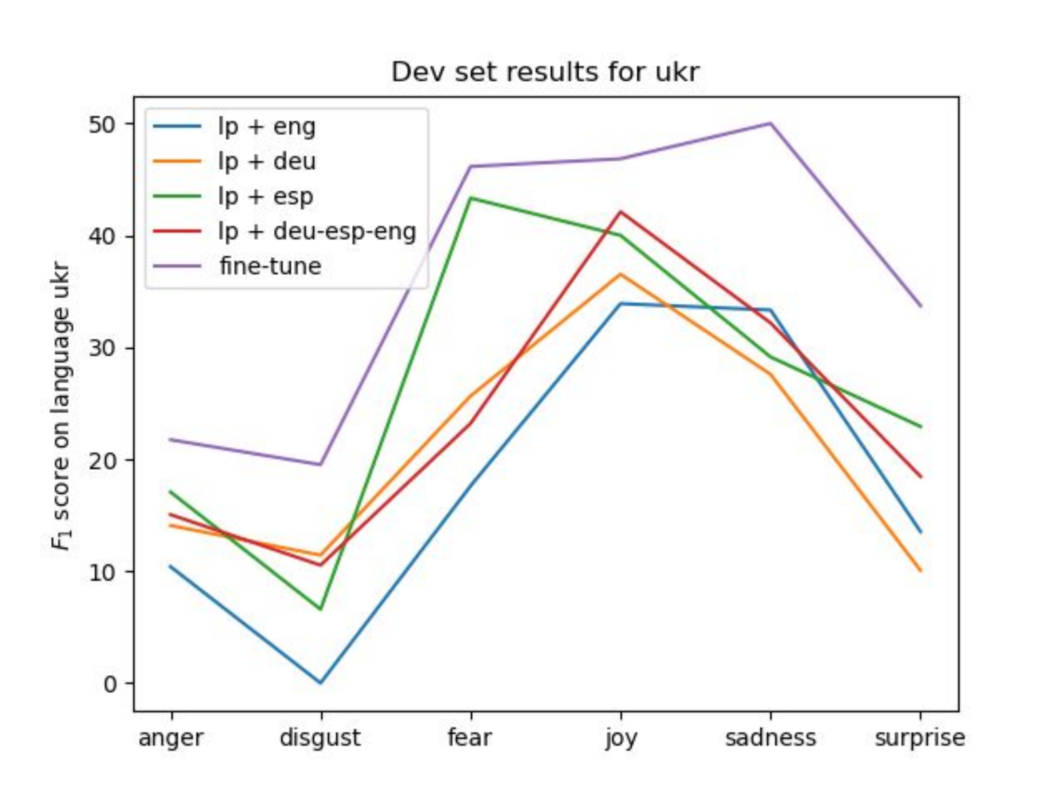
4. Discussion - Per emotion performance

LEALLA encoder



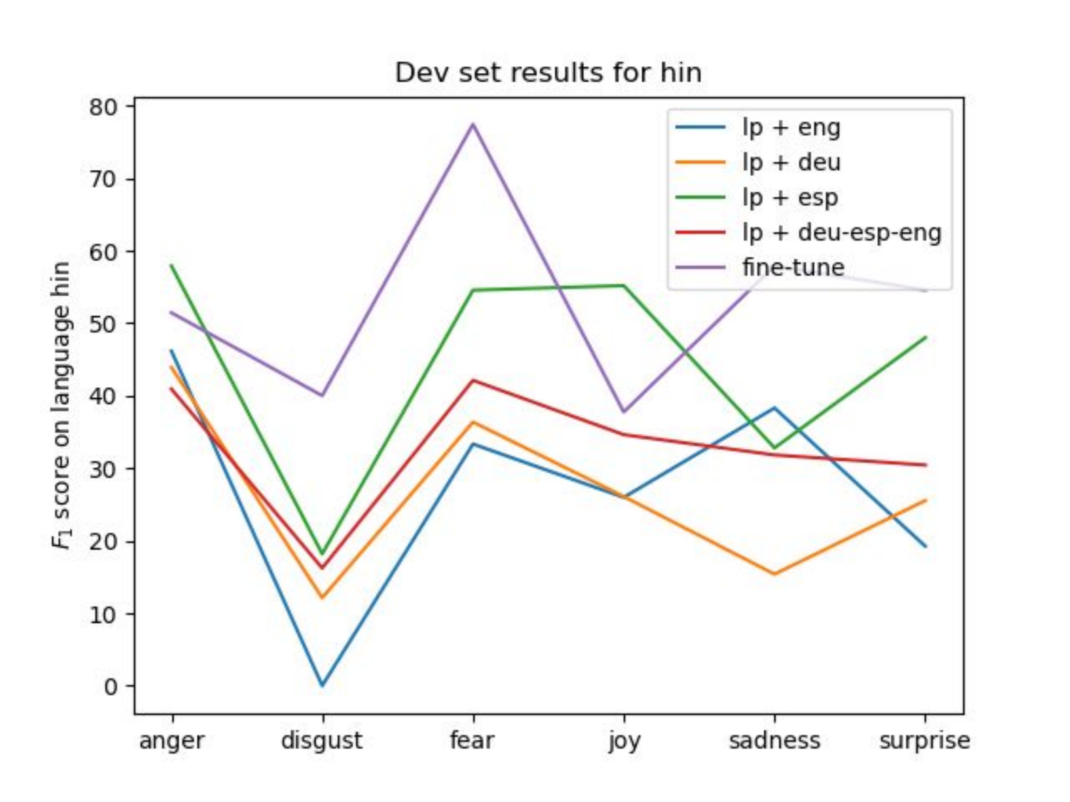
4. Discussion - Per emotion performance

LEALLA encoder



4. Discussion - Per emotion performance

LEALLA encoder



5. Discussion - More data is not always better

5. Limitations

5. Limitations

- the LLM

5. Limitations

- the LLM
- resources (data & GPU)

6. Conclusions

6. Conclusions

- the generalization capabilities of the approach is heavily impacted by the similarity of the source and target language

6. Conclusions

- the generalization capabilities of the approach is heavily impacted by the similarity of the source and target language
- model selection based on source language validation data is not always optimal

7. Future work

- cover a larger number of target languages and looking for trends based on the studied language families and architectures

7. Future work

- cover a larger number of target languages and looking for trends based on the studied language families and architectures
- translating the annotated texts from the source languages into the target one to escape the requirement of having a good Cross-Lingual LLM

7. Future work

- cover a larger number of target languages and looking for trends based on the studied language families and architectures
- translating the annotated texts from the source languages into the target one to escape the requirement of having a good Cross-Lingual LLM
- using cross-lingual data augmentations and specialized algorithms to train detectors which are invariant to changes in the language of texts