Homework 1 A

Yusupha Juwara
Sapienza University of Rome
juwara.1936515@studenti.uniroma1.it

Abstract

This report walks through two tasks: task 0 (EmotivITA) and subtask A of task 1 (HODI). It provides a brief explanation of the two tasks, their input and output formats, how the data are formated, the prompts and the motivation for those specific prompt choices. It also explains how to run the scripts succesfully.

1 Task 0 – EmoITA

1.1 Overview

EmotivITA (Gafà et al., 2023) is a task of emotion regression that aims to advance research in emotion detection within the Italian language. It is based on EmoITA , the Italian version of the EmoBank dataset, annotated based on the dimensional model of emotions, where each of the three dimensions (Valence, Arousal, and Dominance) is a real-value associated with a sentence, allowing for nuanced comparisons. Note that in this task, however, we are asked to convert the real-values to categorical labels.

1.2 Input Format of the dataset

The dataset is in a **.csv** format with five fields: **id** (int), **text** (string), **V** (float), **A** (flaot), and **D** (float).

An example of the dataset with the headers and id 1 and 11 is:

```
id,text,V,A,D
1,Auguriamo a voi e alla vostra
↪  famiglia un nuovo anno pieno di
↪  gioia e amore.,5,3,4.25
11,"Ci sono ristoranti, piscine e
↪  spogliatoi.",4,3.33,3.67
```

1.3 Mapping Float to Categorical

The values of each dimension is a real-value. But since we are explicitly required to use **categorical** labels, we have to map them.

One simple option is to split the range (0-5) **equally** between categories. With this choice, I found that the values are skewed toward the 'Alta' and 'Molto Alta' as shown below.

```
{'Bassa': 1, 'Media': 799, 'Alta': 14307,
↪  'Molto Alta': 8893}
```

This is problematic since the model can **always** predict the majority class and get a high score.

To mitigate this problem, another option is to split in this non-equal way:

```
[0.0,3.2) -> "Bassa"
[2.5, 3.8) -> "Media"
[4.0, 5.0] -> "Alta"

# frequencies
{'Bassa': 5315, 'Media': 11396, 'Alta':
↪  7289}
```

1.4 Reframing the Dataset

The input data is reformated in this way:

- Use pandas to read the **.csv** file, get the 4 fields (see the Input Format 1.2), convert the string of floats to floats, and map the floats to categorical labels (see the Categorical Mapping 1.3)

- Shuffle the choices/labels so that the model learns not just their ordering, but their actual semantic meaning. Note that there is a paper on this, that models sometimes memorize the ordering of the labels, and that if the ordering is changed, there may be some (drastic) drop in accuracy.

- Create a jsonl entry for dimension in their respective files, add the 'dimension' field to the jsonl entries so that the model can differentiate between the dimensions (said by one of the TAs on Google Classroom). E.g.,

```
# Output format for each jsonl entry
{'text': "...", 'choices': [...],
↪  'label': int, 'dimension': '...'}
```

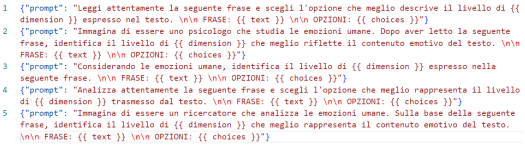
1.5 Run the Code

To run the script for the **Development set.csv**, **cd** to the Home directory 'HM1_A-1936515' and run:

```
python ".\EmotivITA\scripts.py"
↪ --shuffle_labels --verbose
↪ --map_option=0
```

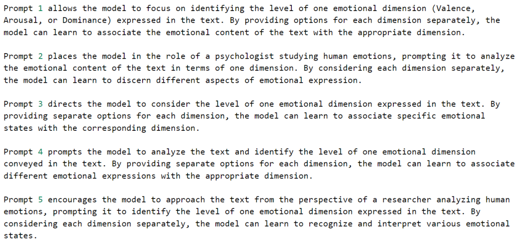
For the **Test set - Gold labels.csv**, do the same but add the 'test' arg. Type 'help' after the file name to see what these args are.

1.6 Prompts



```
{"prompt": "Leggi attentamente la seguente frase e scegli l'opzione che meglio descrive il livello di {{ dimension }} espresso nel testo. \n\n FRASE: {{ text }} \n\n OPZIONI: {{ choices }}"}
{"prompt": "Immagina di essere uno psicologo che studia le emozioni umane. Dopo aver letto la seguente frase, identifica il livello di {{ dimension }} che meglio riflette il contenuto emotivo del testo. \n\n FRASE: {{ text }} \n\n OPZIONI: {{ choices }}"}
{"prompt": "Considerando le emozioni umane, identifica il livello di {{ dimension }} espresso nella seguente frase. \n\n FRASE: {{ text }} \n\n OPZIONI: {{ choices }}"}
{"prompt": "Analizza attentamente la seguente frase e scegli l'opzione che meglio rappresenta il livello di {{ dimension }} trasmesso dal testo. \n\n FRASE: {{ text }} \n\n OPZIONI: {{ choices }}"}
{"prompt": "Immagina di essere un ricercatore che analizza le emozioni umane. Sulla base della seguente frase, identifica il livello di {{ dimension }} che meglio rappresenta il contenuto emotivo del testo. \n\n FRASE: {{ text }} \n\n OPZIONI: {{ choices }}"}
```

Figure 1: Prompts of task 0



Prompt 1 allows the model to focus on identifying the level of one emotional dimension (Valence, Arousal, or Dominance) expressed in the text. By providing options for each dimension separately, the model can learn to associate the emotional content of the text with the appropriate dimension.

Prompt 2 places the model in the role of a psychologist studying human emotions, prompting it to analyze the emotional content of the text in terms of one dimension. By considering each dimension separately, the model can learn to discern different aspects of emotional expression.

Prompt 3 directs the model to consider the level of one emotional dimension expressed in the text. By providing separate options for each dimension, the model can learn to associate specific emotional states with the corresponding dimension.

Prompt 4 prompts the model to analyze the text and identify the level of one emotional dimension conveyed in the text. By providing separate options for each dimension, the model can learn to associate different emotional expressions with the appropriate dimension.

Prompt 5 encourages the model to approach the text from the perspective of a researcher analyzing human emotions, prompting it to identify the level of one emotional dimension expressed in the text. By considering each dimension separately, the model can learn to recognize and interpret various emotional states.

Figure 2: Motivation of the prompt choices for task 0

2 Task 1 – HODI

2.1 Overview

HODI (Homotransphobia Detection in Italian) is a shared task introduced at EVALITA 2023 that focuses on automatically identifying homotransphobic content in Italian text (Nozza et al., 2023). The HODI dataset includes messages directed at Trans, and the goal is to develop sytems like LLMs for detecting homotransphobic language in Italian.

HODI consists of two subtasks. **Subtask A** (binary classification) – classifies a given text as homotransphobic or not. This part is what concerns this section. **Subtask B** (explainability) – identifies which parts of the text are homotransphobic.

2.2 Input Format of the dataset

The dataset is in a **.tsv** format with 3 fields: **id** (int), **text** (string), **homotransphobic** (int). The homotransphobic field defines if a text is homotransphobia (1) or not (0).

An example of the dataset with the headers and id 10 is:

```
id,text,homotransphobic
10          @user_ab io ho una faccia
↪ proprio da frocia          0
```

2.3 Reframing the Dataset

The input data is reformated in a similar way as already mentioned in the **EvaITA** section. The difference is that this file is a **.tsv** (tab separated) and the choices are already in categorical format.

- Use pandas to read the **tsv** file, get the fields, shuffle the choices/labels just as explain in the **EvaITA** section, and create a jsonl entry that looks like this:

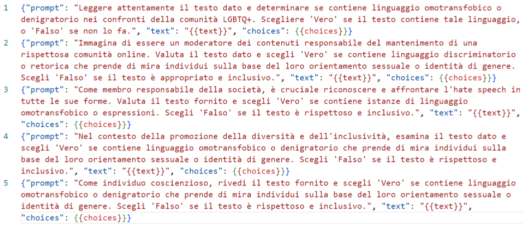
```
# Output format for each jsonl entry
{"text": "...", "choices": ["Vero",
↪ "Falso"], "label": int}
```

2.4 Run the Code

To run the script, **cd** to the Home directory 'HM1_A-1936515' and run:

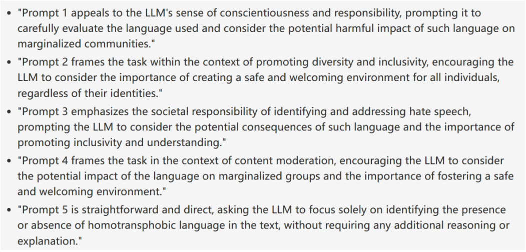
```
python "./HODI_2023/scripts.py"
↪ --shuffle_labels --verbose
```

2.5 Prompts



```
{"prompt": "Leggere attentamente il testo dato e determinare se contiene linguaggio omotransfobico o denigratorio nei confronti della comunità LGBTQ+. Scegliere 'Vero' se il testo contiene tale linguaggio, o 'Falso' se non lo fa.", "text": "{{text}}", "choices": {{choices}}}
{"prompt": "Immagina di essere un moderatore dei contenuti responsabile del mantenimento di una rispettosa comunità online. Valuta il testo dato e scegli 'Vero' se contiene linguaggio discriminatorio o retorica che prende di mira individui sulla base del loro orientamento sessuale o identità di genere. Scegli 'Falso' se il testo è appropriato e inclusivo.", "text": "{{text}}", "choices": {{choices}}}
{"prompt": "Come membro responsabile della società, è cruciale riconoscere e affrontare l'hate speech in tutte le sue forme. Valuta il testo fornito e scegli 'Vero' se contiene istanze di linguaggio omotransfobico o espressioni. Scegli 'Falso' se il testo è rispettoso e inclusivo.", "text": "{{text}}", "choices": {{choices}}}
{"prompt": "Nel contesto della promozione della diversità e dell'inclusività, esamina il testo dato e scegli 'Vero' se contiene linguaggio omotransfobico o denigratorio che prende di mira individui sulla base del loro orientamento sessuale o identità di genere. Scegli 'Falso' se il testo è rispettoso e inclusivo.", "text": "{{text}}", "choices": {{choices}}}
{"prompt": "Come individuo coscienzioso, rivedi il testo fornito e scegli 'Vero' se contiene linguaggio omotransfobico o denigratorio che prende di mira individui sulla base del loro orientamento sessuale o identità di genere. Scegli 'Falso' se il testo è rispettoso e inclusivo.", "text": "{{text}}", "choices": {{choices}}}
```

Figure 3: Prompts of task 1, subtask A



- "Prompt 1 appeals to the LLM's sense of conscientiousness and responsibility, prompting it to carefully evaluate the language used and consider the potential harmful impact of such language on marginalized communities."
- "Prompt 2 frames the task within the context of promoting diversity and inclusivity, encouraging the LLM to consider the importance of creating a safe and welcoming environment for all individuals, regardless of their identities."
- "Prompt 3 emphasizes the societal responsibility of identifying and addressing hate speech, prompting the LLM to consider the potential consequences of such language and the importance of promoting inclusivity and understanding."
- "Prompt 4 frames the task in the context of content moderation, encouraging the LLM to consider the potential impact of the language on marginalized groups and the importance of fostering a safe and welcoming environment."
- "Prompt 5 is straightforward and direct, asking the LLM to focus solely on identifying the presence or absence of homotransphobic language in the text, without requiring any additional reasoning or explanation."

Figure 4: Motivation of the prompt choices for task 1, subtask A

References

Giovanni Gafà, Francesco Cutugno, and Marco Venuti. 2023. Emotivita at evalita2023: Overview of the dimensional and multidimensional emotion analysis task. In *Proceedings of the Eighth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2023)*. CEUR.org.

Debora Nozza, Alessandra Teresa Cignarella, Greta Damo, Tommaso Caselli, and Viviana Patti. 2023. HODI at EVALITA 2023: Overview of the Homotransphobia Detection in Italian Task. In *Proceedings of the Eighth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2023)*, Parma, Italy. CEUR.org.