Homework 1A Task 24

Federico Raschiatore 2071132

Abstract

This homework aims to address the critical aspect of input data preparation for leveraging a language generative model (LLM). The manner in which input data are formatted and presented significantly impacts the performance and effectiveness of an LLM. In particular, this study focuses on transforming input data originating from the SENTIPOLC dataset into a format suitable for comprehension by an LLM.

1 Introduction

SENTIPOLC (SENTIment POLarity Classification) is a task presented at Evalita 2016, the fourth evaluation campaign of Natural Language Processing and Speech tools for Italian

The main goal of SENTIPOLC is sentiment classification at message level on Italian tweets.

The task is divided into three sub-tasks with an increasing level of complexity.

- Subjectivity Classification: Given a message, decide whether the message is subjective or objective.
- Polarity Classification: Given a message, decide whether the message is of positive, negative, neutral or mixed sentiment (i.e. conveying both a positive and negative sentiment).
- Irony Detection: Given a message, decide whether the message is ironic or not

2 Input Shape

The inputs have the following shape:

```
"idtwitter","subj","opos","oneg",
"iro","lpos","lneg", "top", "text"
```

- "idtwitter" is always a number composed by 18 figures
- "subj" is a binary label (1 if the sentence is subjective, 0 if the sentence is objective)
- "opos" and "oneg" are two binary labels that depending on the the combination could have different meanings:
  - 0, 0 : Neutral sentiment
  - 0, 1 : Negative sentiment
  - 1, 0 : Positive sentiment
  - 1, 1 : Mist sentiment
- "iro" is a binary label (1 if the sentence express irony, 0 otherwise)
- "lpos" and "lneg" express the polarity of the sentence
- "text" is the sentence itself and have a variable length

3 Reframe Method

To reframe the dataset I first of all splitted in three sections: "id_twitter", "labels", "text".

Once the file was splitted I build a function for each subtask in order to create the specific records for each of them. The structure of the generic record is the following:

Listing 1: Generic record

```
{
    "id":       int,
    "text":     str,
    "choices":  list[str],
    "label":    int
}
```

Each field have the following meaning:

- id is an incremental number starting from 0
- text is the corpus of the sentence

- "choices" is the list of the possible classification that could be done for the sentence and depends on the task
- "label" is a number that represent the index of the correct classification of the sentence referring to the precedent list

For the first task the choices list is simply:

["Soggettivo", "Oggettivo"]

For the second task the choices list is:

["Neutrale" , "Negativo", "Positivo", "Misto"]

For the third task the choices list is:

['Non Ironico', 'Ironico']

4 Prompts

For each task I created 3 prompts and tested them for 100 samples using 2 different techniques. The main problem for the evaluation is the lack of models trained on an Italian vocabulary, so the first technique is based on the translation of the compiled prompt from italian to english and than i used a BERT model for the classification. Instead in the second techniques i used a BERT multilingual model trained on multiple languages. Both the applied methods are not perfect in fact didn't reach an high accuracy but they are good enough to understand which prompt works better.

4.1 Task 1

- "Considerando la frase {text} questa frase è (Oggettiva) o (Soggettiva) ?"
- "Considerando l'affermazione {text}, trasmette una prospettiva (Oggettiva) o (Soggettiva) ?"
- "Considerando l'affermazione {text}, riflette sentimenti (Oggettiva) o (Soggettiva) ? "

The results with the first method are the following:

Prompt	Accuracy
Prompt 1	40%
Prompt 2	28%
Prompt 3	36%

Table 1: Accuracy for task 1

The results with the second method are the following:

Prompt	Accuracy
Prompt 1	43%
Prompt 2	59%
Prompt 3	55%

Table 2: Accuracy for task 1

4.2 Task 2

- "Data la frase '{text}' questa esprime un sentimento (Neutro), (Negativo), (Positivo) o (Misto)?"
- "Considerando l'affermazione '{text}', questa esprime un sentimento (Neutro), (Negativo), (Positivo) o (Misto)?"
- "Considerando l'affermazione '{text}', riflette un sentimento (Neutro), (Negativo), (Positivo) o (Misto)?"

The results with the fist method are the following:

Prompt	Accuracy
Prompt 1	23%
Prompt 2	28%
Prompt 3	27%

Table 3: Accuracy for task 2

The results with the second method are the following:

Prompt	Accuracy
Prompt 1	16%
Prompt 2	12%
Prompt 3	16%

Table 4: Accuracy for task 2

4.3 Task 3

- "Considerando la frase '{text}' questa è (Non ironico) o (Ironico)?"
- "Considerando l'affermazione '{text}', è (Non Ironico) o (Ironico)?"
- "Considerando l'affermazione '{text}', riflette il contesto (Non Ironico) o (Ironico)?"

The results with the first method are the following:

The results with the second method are the following:

Prompt	Accuracy
Prompt 1	47%
Prompt 2	48%
Prompt 3	56%

Table 5: Accuracy for task 3

Prompt	Accuracy
Prompt 1	69%
Prompt 2	68%
Prompt 3	59%

Table 6: Accuracy for task 3

5 Instruction to run the code

To run the code is sufficient execute all the cells.
Be sure that a GPU device is available.