

SemEval 2024 Task 10: Emotion Discovery and Reasoning its Flip in Conversation (EDiReF)

Substitute the ↑ title ↑ with your project's title, or with Assignment 1 / 2

↓ Keep only one of the following three labels / leave empty for assignments: ↓

NLP Course Project

Daniel Bernardi, Daniele Santini, Hiari Pizzini Cavagna and Muhammad Saleem Ghulam

Master's Degree in Artificial Intelligence, University of Bologna

{ daniel.bernardi, daniele.santini2, hiari.pizzinicavagna, muhammad.ghulam }@studio.unibo.it

DO NOT MODIFY THIS TEMPLATE - EXCEPT, OF COURSE FOR TITLE, SUBTITLE AND AUTHORS. IN THE FINAL VERSION, IN THE L^AT_EX SOURCE REMOVE THE `guidelines` OPTION FROM `\usepackage[guidelines]{nlpreport}`.

Abstract

With this work we challenged the SemEval task about Emotion Discovery and Reasoning its flip in Conversations. The task requires to create a system that, based on a conversation between two or more parties, is able to classify the emotion of each sentence and if it causes an emotion flip in the conversation. To challenge this problem we fine-tuned a BERT model, with two classification heads on top, in two configurations: in the first instance we froze the embedding layer for the training loop, in the second instance we finetuned the whole model. We trained it on a dataset in english language, which had an unbalanced distribution on the trigger labels.

The abstract is very brief summary of your report. Try to keep it no longer than 15-20 lines at most. Write your objective, your approach, and your main observations (what are the findings that make this report worthwhile reading?)

NOTICE: THIS REPORT'S LENGTH MUST RESPECT THE FOLLOWING PAGE LIMITS:

- **ASSIGNMENT: 2 PAGES**
- **NLP PROJECT OR PROJECT WORK: 8 PAGES**
- **COMBINED NLP PROJECT + PW: 12 PAGES**

PLUS LINKS, REFERENCES AND APPENDICES. THIS MEANS THAT YOU CANNOT FILL ALL SECTIONS TO MAXIMUM LENGTH. IT ALSO MEANS THAT, QUITE POSSIBLY, YOU WILL HAVE TO LEAVE OUT OF THE REPORT PART OF THE WORK YOU HAVE DONE OR OBSERVATIONS YOU HAVE. THIS IS NORMAL: THE REPORT SHOULD EMPHASIZE WHAT IS MOST SIGNIFICANT, NOTEWORTHY, AND REFER TO THE NOTEBOOK FOR ANYTHING ELSE. FOR ANY OTHER ASPECT OF YOUR WORK THAT YOU WOULD LIKE TO EMPHASIZE BUT CANNOT EXPLAIN HERE FOR LACK OF SPACE, FEEL FREE TO ADD COMMENTS IN THE NOTEBOOK. INTERESTING TEXT EXAMPLES THAT EXCEED THE MAXIMUM LENGTH OF THE REPORT CAN BE PLACED IN A DEDICATED APPENDIX AFTER THE REFERENCES.

1 Introduction

Here a brief description of the two tasks that need to be tackled.

The first task is Emotion Recognition in Conversation (ERC), and it's defined as follows: given a dialogue, ERC aims to assign an emotion to each utterance from a predefined set of possible emotions. Seven emotions were defined for this task: *neutral*, *joy*, *sadness*, *surprise*, *disgust*, *anger* and *fear*.

The second task is Emotion Flip Reasoning and it's defined as follows: given a dialogue, EFR aims to identify the trigger utterance(s) for an emotion-flip in a multi-party conversation dialogue. The task of recognizing an emotion-flip is not trivial. Let's suppose that two people are involved in the conversation, S_1 and S_2 , and that we're analyzing an utterance pronounced by S_2 . To grasp the eventful presence of an emotion-flip there is need to comprehend and link the emotion of the previous

and future sentence of S_1 .

Our approach to solve the task is to fine-tune a BERT model with two classification heads (to output both an emotion and a boolean value for the trigger). The context of the dialogue is included in the input embeddings by adding one or more neighbouring sentences.

MAX 1 COLUMN FOR ASSIGNMENT REPORTS / 2 COLUMNS FOR PROJECT OR PW / 3 FOR COMBINED REPORTS.

The Introduction is an executive summary, which you can think of as an extended abstract. Start by writing a brief description of the problem you are tackling and why it is important. (Skip it if this is an assignment report).

Then give a short overview of known/standard-/possible approaches to that problems, if any, and what are their advantages/limitations.

After that, discuss your approach, and motivate why you follow that approach. If you are drawing inspiration from an existing model, study, paper, textbook example, challenge, . . . , be sure to add all the necessary references (Chowdhery et al., 2022; Lorenzo et al., 2022; Antici et al., 2021; Nakov et al., 2021; Röttger et al., 2022; Lippi and Torroni, 2016).¹

Next, give a brief summary of your experimental setup: how many experiments did you run on which dataset. Last, make a list of the main results or take-home lessons from your work.

HERE AND EVERYWHERE ELSE: ALWAYS KEEP IN MIND THAT, CRUCIALLY, WHATEVER TEXT/CODE/-FIGURES/IDEAS/... YOU TAKE FROM ELSEWHERE MUST BE CLEARLY IDENTIFIED AND PROPERLY REFERENCED IN THE REPORT.

2 Background

MAX 2 COLUMNS (3 FOR COMBINED REPORTS). DO NOT INCLUDE SECTION IF NO BACKGROUND NECESSARY. OMIT SECTION IN ASSIGNMENT REPORTS.

The Background section is where you briefly provide whatever background information on the domain or challenge you're addressing and/or on the techniques/approaches you're using, that (1) you think is necessary for the reader to understand your work and design choices, and (2) is not something that has been explained to you during the NLP course (to be clear: do NOT repeat explanations of things seen in class, we already know that stuff). If you adapt paragraphs from articles, books,

online resources, etc: be sure to clarify which parts are yours and which ones aren't.

3 System description

MAX 1 COLUMN FOR ASSIGNMENT REPORTS / 4 COLUMNS FOR PROJECT OR PW / 6 FOR COMBINED REPORTS.

Describe the system or systems you have implemented (architectures, pipelines, etc), and used to run your experiments. If you reuse parts of code written by others, be sure to make very clear your original contribution in terms of

- architecture: is the architecture your design or did you take it from somewhere else
- coding: which parts of code are original or heavily adapted? adapted from existing sources? taken from external sources with minimal adaptations?

It is a good idea to add figures to illustrate your pipeline and/or architecture(s) (see Figure 1)

4 Data

MAX 2 COLUMNS / 3 FOR COMBINED REPORTS. OMIT SECTION IN ASSIGNMENT REPORTS.

Provide a brief description of your data including some statistics and pointers (references to articles/URLs) to be used to obtain the data. Describe any pre-processing work you did. Links to datasets must be placed later in Section 8.

5 Experimental setup and results

MAX 1 COLUMN FOR ASSIGNMENT REPORTS / 3 COLUMNS FOR PROJECT OR PW / 5 FOR COMBINED REPORTS.

Describe how you set up your experiments: which architectures/configurations you used, which hyper-parameters and what methods used to set them, which optimizers, metrics, etc.

Then, **use tables** to summarize your findings (numerical results) in validation and test. If you don't have experience with tables in \LaTeX , you might want to use \LaTeX table generator to quickly create a table template.

6 Discussion

MAX 1.5 COLUMNS FOR ASSIGNMENT REPORTS / 3 COLUMNS FOR PROJECT / 4 FOR COMBINED REPORTS. ADDITIONAL EXAMPLES COULD BE PLACED IN AN APPENDIX AFTER THE REFERENCES IF THEY DO NOT FIT HERE.

¹Add only what is relevant.

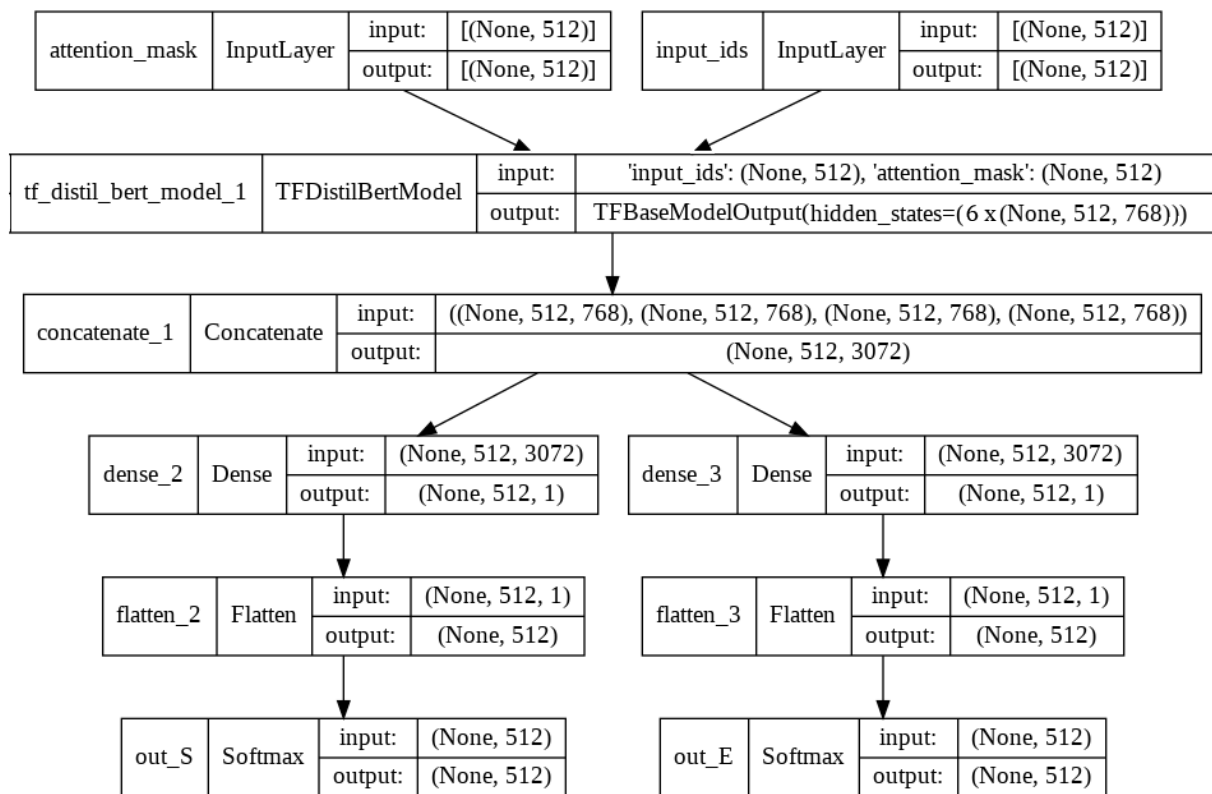


Figure 1: Model architecture

Here you should make your analysis of the results you obtained in your experiments. Your discussion should be structured in two parts:

- discussion of quantitative results (based on the metrics you have identified earlier; compare with baselines);
- error analysis: show some examples of odd-/wrong/unwanted outputs; reason about why you are getting those results, elaborate on what could/should be changed in future developments of this work.

7 Conclusion

MAX 1 COLUMN.

In one or two paragraphs, recap your work and main results. What did you observe? Did all go according to expectations? Was there anything surprising or worthwhile mentioning? After that, discuss the main limitations of the solution you have implemented, and indicate promising directions for future improvement.

8 Links to external resources

THIS SECTION IS OPTIONAL

Insert here:

- a link to your GitHub or any other public repo where one can find your code (only if you did not submit your code on Virtuale);
- a link to your dataset (only for non-standard projects or project works).

DO NOT INSERT CODE IN THIS REPORT

References

Francesco Antici, Luca Bolognini, Matteo Antonio In-ajetovic, Bogdan Ivasiuk, Andrea Galassi, and Federico Ruggeri. 2021. SubjectivITA: An Italian corpus for subjectivity detection in newspapers. In *CLEF*, volume 12880 of *Lecture Notes in Computer Science*, pages 40–52. Springer.

Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, Parker Schuh, Kensen Shi, Sasha Tsvyashchenko, Joshua Maynez, Abhishek Rao, Parker Barnes, Yi Tay, Noam Shazeer, Vinodkumar Prabhakaran, Emily Reif, Nan Du, Ben

Hutchinson, Reiner Pope, James Bradbury, Jacob Austin, Michael Isard, Guy Gur-Ari, Pengcheng Yin, Toju Duke, Anselm Levskaya, Sanjay Ghemawat, Sunipa Dev, Henryk Michalewski, Xavier Garcia, Vedant Misra, Kevin Robinson, Liam Fedus, Denny Zhou, Daphne Ippolito, David Luan, Hyeontaek Lim, Barret Zoph, Alexander Spiridonov, Ryan Sepassi, David Dohan, Shivani Agrawal, Mark Omernick, Andrew M. Dai, Thanumalayan Sankaranarayanan Pillai, Marie Pellat, Aitor Lewkowycz, Erica Moreira, Rewon Child, Oleksandr Polozov, Katherine Lee, Zongwei Zhou, Xuezhi Wang, Brennan Saeta, Mark Diaz, Orhan Firat, Michele Catasta, Jason Wei, Kathy Meier-Hellstern, Douglas Eck, Jeff Dean, Slav Petrov, and Noah Fiedel. 2022. PaLM: Scaling language modeling with pathways. *CoRR*, abs/2204.02311.

Marco Lippi and Paolo Torroni. 2016. Argumentation mining: State of the art and emerging trends. *ACM Transactions on Internet Technology*, 16(2):10:1–10:25.

Abelardo Carlos Martinez Lorenzo, Marco Maru, and Roberto Navigli. 2022. Fully-semantic parsing and generation: the BabelNet meaning representation. In *ACL (1)*, pages 1727–1741. Association for Computational Linguistics.

Preslav Nakov, David P. A. Corney, Maram Hasanain, Firoj Alam, Tamer Elsayed, Alberto Barrón-Cedeño, Paolo Papotti, Shaden Shaar, and Giovanni Da San Martino. 2021. Automated fact-checking for assisting human fact-checkers. In *IJCAI*, pages 4551–4558. ijcai.org.

Paul Röttger, Bertie Vidgen, Dirk Hovy, and Janet B. Pierrehumbert. 2022. Two contrasting data annotation paradigms for subjective NLP tasks. In *NAACL-HLT*, pages 175–190. Association for Computational Linguistics.