Chinchunmei at SemEval-2025 Task 11: Boosting the emotion perception ability of large language models using Contrastive Reasoning Calibration and Self-refinement

Anonymous ACL submission

Abstract

This document is a supplement to the general instructions for *ACL authors. It contains instructions for using the LaTeX style files for ACL conferences. The document itself conforms to its own specifications, and is therefore an example of what your manuscript should look like. These instructions should be used both for papers submitted for review and for final versions of accepted papers.

1 Introduction

Text-Based Emotion Detection (TBED) has long been a prominent research area in NLP, with widespread applications in social media analysis (Kuamri and Babu, 2017; Salam and Gupta, 2018; Cassab and Kurdy, 2020), mental health treatment (Kusal et al., 2021; Krommyda et al., 2021), and dialogue systems (Liu et al., 2022; Ide and Kawahara, 2022; Hu et al., 2021). Depending on how emotions are defined, TBED can be broadly categorized into two approaches: (1). Classification-based methods, where emotions are categorized into discrete labels (Ekman and Friesen, 1969; Plutchik, 1982). (2). Scoring-based methods, where emotions are treated as interrelated entities with varying intensity levels (Russell and Mehrabian, 1977).

However, due to the nuanced and complex nature of emotional expression, TBED faces several key challenges (Al Maruf et al., 2024): (1). The distinction between different emotions is often subtle, and emotions are often conveyed implicitly—through metaphors or situational cues rather than explicit words. (2). Cultural and linguistic differences influence emotion perception. These challenges make TBED difficult to rely solely on predefined lexicons. A robust TBED system must integrate cultural context, linguistic diversity, background knowledge, and advanced semantic understanding.

For SemEval-2025, Task 11, titled Bridging the Gap in Text-Based Emotion Detection, introduces a multilingual benchmark covering 28 languages. The competition consists of task A (classification) and task B (intensity prediction). The goal is to identify the speaker's perceived emotion in a given sentence. Emotion categories follow Ekman's framework(Ekman and Friesen, 1969), encompassing six basic emotions: anger, fear, sadness, joy, disgust, and surprise. Task B further introduces four intensity levels for each emotion. This competition setup encapsulates both primary TBED methodologies while incorporating challenges in multilingual and fine-grained emotion recognition.

To address both tracks while supporting emotion prediction across 28 languages, we give up the traditional encoder-based classification frameworks and adopt the generative large language model (LLM). This decision is driven by its robust multi-task integration capabilities and strong support for cross-linguistic applications.

We first designed a standard prediction task (ST), training the LLM to predict each sample's emotion label and intensity level directly. Meanwhile, to enhance the model's ability to distinguish nuanced emotional traits in text, we employ Contrastive Reasoning Calibration (CRC) (Li et al., 2024) techniques and refine it to suit this competition. This method allows the model to adjust its predictions through direct sample comparisons. Additionally, to further improve the prediction accuracy, we introduced a Self-refinement process, enabling the model to review and refine its outputs.

Our contributions are as follows:

- We extended the previous CRC technique from 1V1 comparison to 1V3 comparison (CRC4). Incorporating more samples in each comparison enhanced the model's robustness and discriminative ability.
- We introduced an error correction training task to improve the model's understanding of emotion labels. We repurposed the error correction

template for Self-refinement, enabling it to refine predictions across ST, CRC, and CRC4 outputs, improving accuracy for certain labels.

- Based on the performance of different task fusion strategies across languages and labels, we selected the optimal approach for each label as our final submission. In the leaderboard, our approach achieved Track A top 10 in 16 languages and 12th in English; Track B top 10 across all languages and 7th in English.

2 Methodology

3 Engines

To produce a PDF file, pdfLaTeX is strongly recommended (over original LaTeX plus dvips+ps2pdf or dvipdf). The style file `acl.sty` can also be used with luaLaTeX and XeLaTeX, which are especially suitable for text in non-Latin scripts. The file `acl_lualatex.tex` in this repository provides an example of how to use `acl.sty` with either luaLaTeX or XeLaTeX.

4 Preamble

The first line of the file must be

```
\documentclass[11pt]{article}
```

To load the style file in the review version:

```
\usepackage[review]{acl}
```

For the final version, omit the `review` option:

```
\usepackage{acl}
```

To use Times Roman, put the following in the preamble:

```
\usepackage{times}
```

(Alternatives like txfonts or newtx are also acceptable.)

Please see the LaTeX source of this document for comments on other packages that may be useful.

Set the title and author using `\title` and `\author`. Within the author list, format multiple authors using `\and` and `\And` and `\AND`; please see the LaTeX source for examples.

By default, the box containing the title and author names is set to the minimum of 5 cm. If you need more space, include the following in the preamble:

```
\setlength\titlebox{<dim>}
```

where `<dim>` is replaced with a length. Do not set this length smaller than 5 cm.

Command	Output	Command	Output
`{\"a}`	ä	`{\c c}`	ç
`{\^e}`	ê	`{\u g}`	ğ
`{\`i}`	ì	`{\l}`	ł
`{\.I}`	İ	`{\~n}`	ñ
`{\o}`	ø	`{\H o}`	ő
`{\'u}`	ú	`{\v r}`	ř
`{\aa}`	å	`{\ss}`	ß

Table 1: Example commands for accented characters, to be used in, *e.g.*, BibTeX entries.

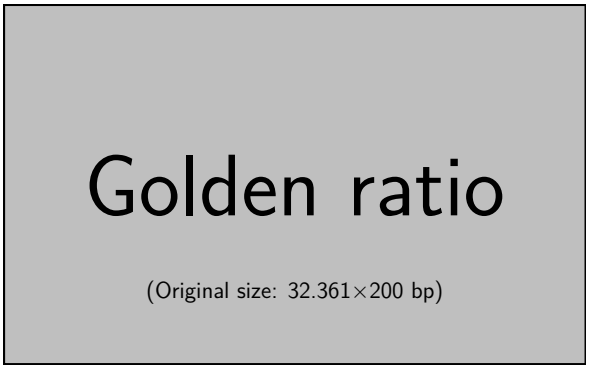

Figure 1: A figure with a caption that runs for more than one line. Example image is usually available through the `mwe` package without even mentioning it in the preamble.

5 Document Body

5.1 Footnotes

Footnotes are inserted with the `\footnote` command.[1]

5.2 Tables and figures

See Table 1 for an example of a table and its caption.
Do not override the default caption sizes.

As much as possible, fonts in figures should conform to the document fonts. See Figure 1 for an example of a figure and its caption.

Using the `graphicx` package graphics files can be included within figure environment at an appropriate point within the text. The `graphicx` package supports various optional arguments to control the appearance of the figure. You must include it explicitly in the LaTeX preamble (after the `\documentclass` declaration and before `\begin{document}`) using `\usepackage{graphicx}`.

[1] This is a footnote.

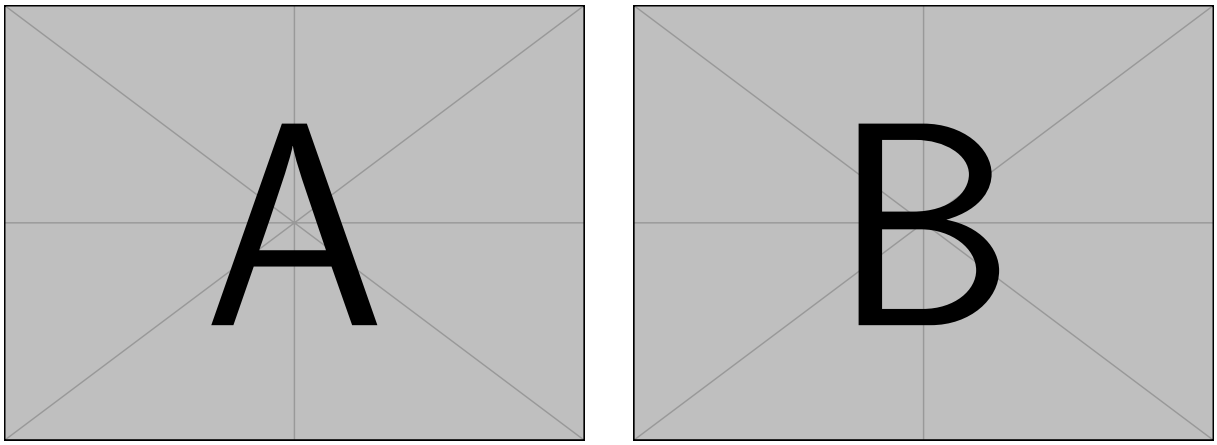

Figure 2: A minimal working example to demonstrate how to place two images side-by-side.

5.3 Hyperlinks

Users of older versions of LaTeX may encounter the following error during compilation:

> `\pdfendlink` ended up in different nesting level than `\pdfstartlink`.

This happens when pdfLaTeX is used and a citation splits across a page boundary. The best way to fix this is to upgrade LaTeX to 2018-12-01 or later.

5.4 Citations

Table 2 shows the syntax supported by the style files. We encourage you to use the natbib styles. You can use the command `\citet` (cite in text) to get "author (year)" citations, like this citation to a paper by Gusfield (1997). You can use the command `\citep` (cite in parentheses) to get "(author, year)" citations (Gusfield, 1997). You can use the command `\citealp` (alternative cite without parentheses) to get "author, year" citations, which is useful for using citations within parentheses (e.g. Gusfield, 1997).

A possessive citation can be made with the command `\citeposs`. This is not a standard natbib command, so it is generally not compatible with other style files.

5.5 References

The LaTeX and BibTeX style files provided roughly follow the American Psychological Association format. If your own bib file is named `custom.bib`, then placing the following before any appendices in your LaTeX file will generate the references section for you:

```
\bibliography{custom}
```

You can obtain the complete ACL Anthology as a BibTeX file from `https://aclweb.org/anthology/anthology.bib.gz`. To include both the Anthology and your own .bib file, use the following instead of the above.

```
\bibliography{anthology,custom}
```

Please see Section 6 for information on preparing BibTeX files.

5.6 Equations

An example equation is shown below:

$$A = \pi r^2 \tag{1}$$

Labels for equation numbers, sections, subsections, figures and tables are all defined with the `\label{label}` command and cross references to them are made with the `\ref{label}` command.

This an example cross-reference to Equation 1.

5.7 Appendices

Use `\appendix` before any appendix section to switch the section numbering over to letters. See Appendix A for an example.

6 BibTeX Files

Unicode cannot be used in BibTeX entries, and some ways of typing special characters can disrupt BibTeX's alphabetization. The recommended way of typing special characters is shown in Table 1.

Please ensure that BibTeX records contain DOIs or URLs when possible, and for all the ACL materials that you reference. Use the `doi` field for DOIs and the `url` field for URLs. If a BibTeX entry has a URL or DOI field, the paper title in the references section will appear as a hyperlink to the paper, using the hyperref LaTeX package.

Output	natbib command	ACL only command
(Gusfield, 1997)	\citep	
Gusfield, 1997	\citealp	
Gusfield (1997)	\citet	
(1997)	\citeyearpar	
Gusfield's (1997)		\citeposs

Table 2: Citation commands supported by the style file. The style is based on the natbib package and supports all natbib citation commands. It also supports commands defined in previous ACL style files for compatibility.

Limitations

Since December 2023, a "Limitations" section has been required for all papers submitted to ACL Rolling Review (ARR). This section should be placed at the end of the paper, before the references. The "Limitations" section (along with, optionally, a section for ethical considerations) may be up to one page and will not count toward the final page limit. Note that these files may be used by venues that do not rely on ARR so it is recommended to verify the requirement of a "Limitations" section and other criteria with the venue in question.

Acknowledgments

This document has been adapted by Steven Bethard, Ryan Cotterell and Rui Yan from the instructions for earlier ACL and NAACL proceedings, including those for ACL 2019 by Douwe Kiela and Ivan Vulić, NAACL 2019 by Stephanie Lukin and Alla Roskovskaya, ACL 2018 by Shay Cohen, Kevin Gimpel, and Wei Lu, NAACL 2018 by Margaret Mitchell and Stephanie Lukin, BibTEX suggestions for (NA)ACL 2017/2018 from Jason Eisner, ACL 2017 by Dan Gildea and Min-Yen Kan, NAACL 2017 by Margaret Mitchell, ACL 2012 by Maggie Li and Michael White, ACL 2010 by Jing-Shin Chang and Philipp Koehn, ACL 2008 by Johanna D. Moore, Simone Teufel, James Allan, and Sadaoki Furui, ACL 2005 by Hwee Tou Ng and Kemal Oflazer, ACL 2002 by Eugene Charniak and Dekang Lin, and earlier ACL and EACL formats written by several people, including John Chen, Henry S. Thompson and Donald Walker. Additional elements were taken from the formatting instructions of the *International Joint Conference on Artificial Intelligence* and the *Conference on Computer Vision and Pattern Recognition*.

References

Abdullah Al Maruf, Fahima Khanam, Md Mahmudul Haque, Zakaria Masud Jiyad, Firoj Mridha, and Zeyar Aung. 2024. Challenges and opportunities of text-based emotion detection: A survey. *IEEE Access*.

Rie Kubota Ando and Tong Zhang. 2005. A framework for learning predictive structures from multiple tasks and unlabeled data. *Journal of Machine Learning Research*, 6:1817–1853.

Galen Andrew and Jianfeng Gao. 2007. Scalable training of L1-regularized log-linear models. In *Proceedings of the 24th International Conference on Machine Learning*, pages 33–40.

S Cassab and Mohamad-Bassam Kurdy. 2020. Ontology-based emotion detection in arabic social media. *International Journal of Engineering Research & Technology (IJERT)*, 9(08):1991–2013.

Paul Ekman and Wallace V Friesen. 1969. The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *semiotica*, 1(1):49–98.

Dan Gusfield. 1997. *Algorithms on Strings, Trees and Sequences*. Cambridge University Press, Cambridge, UK.

Dou Hu, Lingwei Wei, and Xiaoyong Huai. 2021. Dialoguecrn: Contextual reasoning networks for emotion recognition in conversations. *arXiv preprint arXiv:2106.01978*.

Tatsuya Ide and Daisuke Kawahara. 2022. Building a dialogue corpus annotated with expressed and experienced emotions. *arXiv preprint arXiv:2205.11867*.

Maria Krommyda, Anastasios Rigos, Kostas Bouklas, and Angelos Amditis. 2021. An experimental analysis of data annotation methodologies for emotion detection in short text posted on social media. In *Informatics*, volume 8, page 19. MDPI.

Santoshi Kuamri and C Narendra Babu. 2017. Real time analysis of social media data to understand people emotions towards national parties. In *2017 8th international conference on computing, communication and networking technologies (ICCCNT)*, pages 1–6. IEEE.

Sheetal Kusal, Shruti Patil, Ketan Kotecha, Rajanikanth Aluvalu, and Vijayakumar Varadarajan. 2021. Ai based emotion detection for textual big data: Techniques and contribution. *Big Data and Cognitive Computing*, 5(3):43.

Tian Li, Nicolay Rusnachenko, and Huizhi Liang. 2024. Chinchunmei at wassa 2024 empathy and personality shared task: Boosting llm's prediction with role-play augmentation and contrastive reasoning calibration. In *Proceedings of the 14th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis*, pages 385–392.

Yuchen Liu, Jinming Zhao, Jingwen Hu, Ruichen Li, and Qin Jin. 2022. Dialogueein: Emotion interaction network for dialogue affective analysis. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 684–693.

R Plutchik. 1982. A psycho evolutionary theory of emotions. *Social Science Information*.

Mohammad Sadegh Rasooli and Joel R. Tetreault. 2015. Yara parser: A fast and accurate dependency parser. *Computing Research Repository*, arXiv:1503.06733. Version 2.

James A Russell and Albert Mehrabian. 1977. Evidence for a three-factor theory of emotions. *Journal of research in Personality*, 11(3):273–294.

Shaikh Abdul Salam and Rajkumar Gupta. 2018. Emotion detection and recognition from text using machine learning. *Int. J. Comput. Sci. Eng*, 6(6):341–345.

A Example Appendix

This is an appendix.