HITSZ-HLT at WASSA-2024 Shared Task 2: Language-agnostic Multi-task Learning for Explainability of Cross-lingual Emotion Detection

Feng Xiong[1,3] **Jun Wang**[1,3] **Geng Tu**[1,3] **Ruifeng Xu**[1,2,3]*
[1]Harbin Institute of Technology, Shenzhen, China
[2]Peng Cheng Laboratory, Shenzhen, China
[3]Guangdong Provincial Key Laboratory of Novel Security Intelligence Technologies
23s151006@stu.hit.edu.cn, xuruifeng@hit.edu.cn

Abstract

This paper describes the system developed by the HITSZ-HLT team for WASSA-2024 Shared Task 2, which addresses two closely linked subtasks: Cross-lingual Emotion Detection and Binary Trigger Word Detection in tweets. The main goal of Shared Task 2 is to simultaneously identify the emotions expressed and detect the trigger words across multiple languages. To achieve this, we introduce a **L**anguage-**a**gnostic **M**ulti **T**ask **L**earning (LaMTL) framework that integrates emotion prediction and emotion trigger word detection tasks. By fostering synergistic interactions between task-specific and task-agnostic representations, the LaMTL aims to mutually enhance emotional cues, ultimately improving the performance of both tasks. Additionally, we leverage large-scale language models to translate the training dataset into multiple languages, thereby fostering the formation of language-agnostic representations within the model, significantly enhancing the model's ability to transfer and perform well across multilingual data. Experimental results demonstrate the effectiveness of our framework across both tasks, with a particular highlight on its success in achieving second place in sub-task 2.

1 Introduction

Sentiment Analysis is an important task in Natural Language Processing (NLP), aiming to identify and assess the sentiment polarity in texts (Cambria, 2016). With the rapid development of social media and the Internet, sentiment analysis has become increasingly important in various fields such as customer service (Zvarevashe and Olugbara, 2018) and finance (Xing et al., 2020). Despite notable strides in sentiment analysis research (Jiang et al., 2023; Tu et al., 2023; Zhang et al., 2023; Hartmann et al., 2023; Zhong et al., 2023), challenges persist, particularly concerning foreign language texts where annotated data may be scarce.

Cross-lingual Sentiment Analysis (CLSA) (Liu, 2012) mitigates these challenges by utilizing resources from one or more source languages to assist in sentiment analysis for low-resource languages (Esuli et al., 2020). This approach addresses the lack of annotated corpora for many non-English languages, making it a crucial research area in NLP. The fundamental strategy entails the transfer and adaptation of knowledge across various linguistic contexts (Zhao et al., 2024). Building on the foundational principles of CLSA, our study further explores how these methodologies can be practically implemented to enhance model performance across diverse linguistic settings.

The main challenge in Shared Task 2 (Maladry et al., 2024) involves two key aspects: (1) Enhancing the model's capability to transfer knowledge to languages not present in the training data. (2) Developing strategies to effectively utilize complementarities between dual tasks given the limited availability of annotated data. To address the aforementioned challenges, we have developed a **L**anguage-**a**gnostic **M**ulti **T**ask **L**earning (LaMTL) framework that effectively navigates cross-lingual obstacles while simultaneously bridging shared emotional cues across dual tasks.

Specifically, we utilize ChatGPT[1] as a sophisticated tool for translation. By refining prompts, we translate the original tweets into Dutch, Russian, Spanish, and French, striving to maintain the original style as accurately as possible. By aligning the representations of identical tweets across different languages (Feng et al., 2022), we aim to develop a language-agnostic representation. Due to the complementary relationship between emotions and trigger words within tweets, we have designed a novel multi-task framework that includes both task-agnostic and task-specific encoders. The task-agnostic encoder captures task-invariant fea-

*Corresponding author.

[1]https://chat.openai.com/

Proceedings of the 14th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis, pages 476–482
August 15, 2024 ©2024 Association for Computational Linguistics

tures, facilitating the identification of underlying commonalities and related characteristics across tasks, while the task-specific encoder learns features unique to each task. To ensure comprehensive training, we incorporated multiple loss functions, including reconstruction loss and task prediction loss. This approach aims to improve the accuracy and robustness of multi-task learning by leveraging both shared and task-specific features, ultimately enhancing the framework's performance on various tasks.

Additionally, we conducted a rigorous evaluation of our approach utilizing the EXALT dataset[2], which substantiated its effectiveness. This comprehensive validation process led to our achievement of an esteemed second place in a highly competitive arena, as documented on the official leaderboard[3].

2 Related Evaluation Tasks

In recent years, CLSA has gained widespread attention. In 2013, NLP&CC organized a cross-language evaluation by releasing annotated English data from Amazon user reviews and unannotated Chinese reviews. This initiative facilitated the development of methods for cross-lingual sentiment analysis. SemEval-2017 Task 4 (Rosenthal et al., 2017) focused on multilingual sentiment analysis of Twitter posts. This task utilized product ratings from platforms such as Amazon, TripAdvisor, and Yelp, and included five subtasks for both Arabic and English. SemEval-2020 Task 9 (Patwa et al., 2020) concentrated on sentiment analysis of code-mixed tweets in Hinglish and Spanglish, providing annotated corpora and attracting 89 submissions. The top models achieved F1 scores of 75.0% for Hinglish and 80.6% for Spanglish. SemEval-2022 Task 10 (Barnes et al., 2022) introduced the first shared task on Structured Sentiment Analysis. Participants were required to predict sentiment graphs composed of a holder, target, expression, and polarity across seven datasets in five languages.

3 Methodology

In this section, we offer a comprehensive introduction to each component of the proposed LaMTL framework, illustrated in Fig. 1.

[2]https://huggingface.co/datasets/pranaydeeps/EXALT-v1
[3]https://lt3.ugent.be/exalt

3.1 Feature Extraction

We first employ ChatGPT to translate each English tweet $\mathbf{x}_i^e$ from the training dataset $\mathbf{D}$ into $\mathbf{x}_i^\psi$, where $\psi \in \{d, r, s, f\}$ corresponds to Dutch, Russian, Spanish, and French, respectively. Subsequently, we utilize a multilingual pretrained model as the foundational encoder to extract feature representations from the tweets across various languages.

Specifically, for a tweet $\mathbf{x}_i = \{s_1, s_2, \ldots, s_{\mathcal{N}_i}\}$, where $\mathcal{N}_i$ denotes the number of words of $\mathbf{x}_i$, the corresponding sequence of tokens resulting from the application of subword tokenization techniques such as WordPiece and Byte Pair Encoding (BPE) is denoted by $\{w_1, w_2, \ldots, w_{\widehat{\mathcal{N}_i}}\}$. $\widehat{\mathcal{N}_i}$ signifies the number of tokens corresponding to $\mathbf{x}_i$. The output of the last layer is denoted as $\mathbf{h}^e \in \mathbb{R}^{\widehat{\mathcal{N}_i} \times d_h}$ and $\mathbf{h}^\psi \in \mathbb{R}^{\widehat{\mathcal{N}_i^\psi} \times d_h}$. For each word s_j, its representation $\mathbf{h}_j \in \mathbb{R}^{d_h}$ is obtained by averaging the representations of its corresponding tokens.

3.2 Cross-Lingual Semantic Alignment

The Cross-Lingual Semantic Alignment (SA) Module aims to align semantic representations across language barriers. To achieve this, we employ the Mean Squared Error (MSE) as a reconstruction loss function. This function aims to minimize the semantic distance between translated non-English tweets and their English counterparts. It promotes the convergence of the feature vectors $\mathbf{h}^\psi$ of the translated tweets toward the feature vectors $\mathbf{h}^e$ of the original English tweets, ensuring consistent semantic representation across languages. The reconstruction loss L_{rec} can be represented as,

$$L_{rec} = \sum_{\psi} \mathrm{MSE}(\widehat{\mathbf{h}^e}, \widehat{\mathbf{h}^\psi}), \tag{1}$$

where $\widehat{\mathbf{h}^e}$ and $\widehat{\mathbf{h}^\psi}$ denote the [CLS] representation or the average of all tokens for $\mathbf{h}^e$ and $\mathbf{h}^\psi$.

3.3 Multi-Task Fusion

To effectively encapsulate the pertinent emotional cues in diverse tasks, we developed the Multi-Task Fusion (MTF) Module. This module integrates a task-invariant encoder alongside two task-specific encoders, each comprising a stacked structure of L transformer encoder layers. In MTF, each tweet $\mathbf{x}_i$ is processed by three distinct encoders. The first encoder, $\mathbf{E}_s$, is task-invariant and designed to learn a generalized representation across multiple tasks by positioning the learned features within a common subspace. The other two encoders, $\mathbf{E}_e$ and $\mathbf{E}_t$,

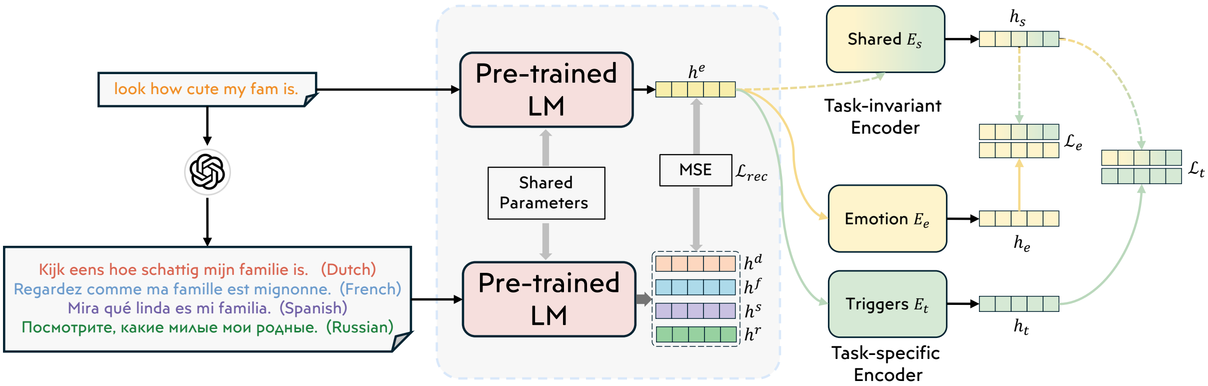

Figure 1: The proposed LaMTL framework. The graphical symbols directly correspond to the equations presented within the paper.

are task-specific. $\mathbf{E}_e$ is dedicated to emotion detection, focusing on the subtle emotional nuances embedded within the tweets. In contrast, $\mathbf{E}_t$ specializes in emotion trigger detection, identifying key triggers that indicate the presence or absence of specific emotional conditions. The process can be represented as:

$$\mathbf{h}_s = \mathbf{E}_s(\mathbf{h}^e, \Theta_s), \tag{2}$$

$$\mathbf{h}_\eta = \mathbf{E}_\eta(\mathbf{h}^e, \Theta_\eta), \eta \in \{e, t\}. \tag{3}$$

The encoder $\mathbf{E}_s$ shares parameters Θ_s across the two tasks, while $\mathbf{E}_\eta$ utilizes distinct parameters Θ_η for each task.

Finally, we concatenate the task-invariant representations $\mathbf{h}_s$ with the task-specific representation $\mathbf{h}_\eta$ and use a Multi-Layer Perceptron (MLP) with softmax for classification. Formally,

$$\mathcal{H}_\eta = \mathbf{h}_s \oplus \mathbf{h}_\eta, \tag{4}$$

$$\mathcal{P}_\eta = \mathrm{softmax}(\mathrm{MLP}_\eta(\mathcal{H}_\eta)), \tag{5}$$

$$\hat{\mathbf{y}}_\eta = \mathrm{argmax}(\mathcal{P}_\eta), \tag{6}$$

where $\oplus$ denotes the concatenation operation. Notably, we utilize $\hat{\mathbf{y}}_\eta$ as the prediction results for the task η.

3.4 Model Training

We utilize cross-entropy loss for the classification of the Cross-lingual Emotion Detection and Binary Trigger Word Detection tasks, denoted as $\mathcal{L}_e$ and $\mathcal{L}_t$, respectively. The computation process can be described as follows:

$$\mathcal{L}_\eta = -\frac{1}{N_\eta}\sum_{j=1}^{N_\eta}\sum_{k=1}^{C_\eta} \mathbf{y}_{\eta,[j][k]} \log(\mathcal{P}_{\eta,[j][k]}), \tag{7}$$

where N denotes the number of samples, C represents the number of classes for task η, $\mathcal{P}_{\eta,[j][k]}$ denotes the probability distribution for instance j over class k, and $\mathbf{y}_{\eta,[j][k]}$ is a binary indicator that equals 1 if class k is the correct classification for instance j in task η, and 0 otherwise. Specifically, C is set to 6 for task $\eta = e$ and to 2 for task $\eta = t$. For samples in the dataset containing two types of labels, we compute $\mathcal{L}_e$ and $\mathcal{L}_t$. For samples with only one type of label, we compute the loss specific to the corresponding task. To facilitate better-mixed learning, we apply a shuffling strategy to the dataset.

By combining the reconstruction loss and task-specific loss, our final loss function can be expressed as,

$$\mathcal{L} = \mathcal{L}_e + \mathcal{L}_t + \lambda_{rec}\mathcal{L}_{rec} + \lambda||\Theta||_2^2, \tag{8}$$

where L_e and L_t denote the classification loss for Cross-lingual Emotion Detection and Binary Trigger Word Detection, while λ represents the L2 regularization weight, and Θ signifies the set of all trainable parameters.

4 Experiments

4.1 Baselines

To demonstrate the efficacy of our approach, we fine-tuned multilingual pre-trained models using the official codebase[4], including Multilingual-BERT (Devlin et al., 2019), LaBSE (Feng et al., 2022), and Multilingual-E5-Large (Wang et al., 2024). Furthermore, due to the robust multilingual capabilities of LLM, we conducted experiments using various configurations. Specifically,

[4]https://github.com/pranaydeeps/WASSA24_EXALT

Methods	Emotion	Binary Triggers
EXALT-Baseline ♠	44.76	23.49
Multilingual-BERT♣	34.40	23.57
LaBSE♣	48.41	32.49
Multilingual-E5-large ♣	51.70	25.68
LLaMA2 + LoRAMoE ♢	49.03	57.05
LLaMA3 + LoRA ♢	<u>54.40</u>	<u>57.62</u>
GPT4 (Zero-shot)♢	52.57	-
LaMTL‡	**56.88**	**60.95**

Table 1: Comparison of F1 score (%) conducted for the EXALT datasets. The results are presented such that the highest performance is denoted in bold, and the second highest performance is underlined. ♠ indicates results obtained from Codalab, ♣ indicates our re-implemented using the official codebase, ♢ indicates the results of our implementation on the validation set, and ‡ indicates the results of our implementation on the test set.

Methods	Emotion	Binary Triggers
LaMTL‡	**56.88**	**60.95**
w/o SA‡	54.85	58.47
w/o MTF‡	54.90	59.84

Table 2: F1 score (%) for Ablation results.

we fine-tuned LLaMA-2 (Touvron et al., 2023) in conjunction with LoRAMoE (Dou et al., 2024) and LLaMA-3 (AI@Meta, 2024) with LoRA (Hu et al., 2022). Additionally, we performed zero-shot emotion detection experiments on GPT-4 (Achiam et al., 2023), and the designed prompt template can be found in Appendix A.

In ablation studies, 'w/o SA' denotes the removal of the SA module, and 'w/o MTF' indicates the removal of the MTF module.

4.2 Experimental Settings

In our experimental settings, we utilize a learning rate of 1e-4 with the AdamW optimizer to optimize the model parameters. We configured gradient accumulation to 4 and batch size to 8. In this study, we employ XLM-RoBERTa-Large (Conneau et al., 2020) as the backbone model. We configured the encoder in the MTF module as a single-layer transformer encoder. Additionally, the λ_{rec} parameter was strategically set to 0.05. The word embedding dimension d_h is 1024, and the maximum sequence length is 512. All experiments were conducted on a single RTX 4090 GPU, using BF16 precision to optimize both speed and computational efficiency.

Team	Binary Triggers
CTcloud	**61.58**
HITSZ-HLT	<u>60.95</u>
UWB	59.19
NLP_Newcomer	57.85
NYCU-NLP	56.36

Table 3: Top-5 F1 score (%) for Binary Trigger Word Detection. The results are presented such that the highest performance is denoted in bold, and the second highest performance is underlined.

4.3 Evaluation Metrics

We use the official metrics for evaluation. For the Cross-lingual Emotion Detection task, we use the Macro-averaged F1 score as the evaluation metric. For the Binary Trigger Word Detection task, we utilize the Token F1 score as the evaluation metric.

4.4 Experimental Results and Analysis

Comparative Results: Table 1 presents a comparative analysis of our LaMTL model against various cross-lingual baseline models and LLMs. Our LaMTL model consistently outperforms the baselines across both sub-tasks, demonstrating superior performance. In addition, LLMs also exhibit remarkable performance and will be a primary focus of our future research.

Ablation Studies: We conducted ablation experiments for our framework. According to the results shown in Table 2, the LaMTL model achieved F1 scores of 56.88% on the Emotion task and 60.95% on the Binary Triggers task. Removing the Cross-lingual Semantic Alignment (SA) module resulted in F1 score decreases of 2.03 and 2.48 percentage points for the Emotion and Binary Triggers tasks, respectively, indicating the importance of cross-linguistic feature semantic alignment, especially for the Binary Triggers Word Detection task. Similarly, removing the Multi-Task Fusion (MTF) module led to F1 score decreases of 1.98 and 1.11 percentage points for the Emotion and Binary Triggers tasks, respectively.

Leaderboard Results: Table 3 presents the performance of the top five teams in Binary Trigger Word Detection task. Our method achieves second place on the leaderboard.

5 Conclusions

In this paper, we propose a language-agnostic multi-task learning approach to address the challenge of interpretability in cross-lingual sentiment analysis.

Firstly, we designed a reconstruction loss to mitigate cross-lingual discrepancies. Secondly, we implemented a multi-task learning framework to share sentiment cues between two tasks, thereby enhancing performance in both tasks. Through these methods, our model effectively enhances cross-lingual capabilities and facilitates the sharing of emotional cues between multiple tasks, thereby achieving competitive performance.

6 Limitations

Although our LaMTL framework has demonstrated significant efficacy in cross-lingual sentiment detection and binary trigger word detection, there are several limitations that need to be addressed in future work. The use of large-scale language models for translation introduces potential biases and inaccuracies, especially for less common or informal texts in tweets. These translation inconsistencies can affect the quality of language-agnostic representations. While our framework has achieved commendable results, real-world applications might present additional challenges, such as subtle nuances specific to certain domains and the evolving use of language, which were not extensively explored in this study. Addressing these limitations in future research is crucial for enhancing the applicability and performance of cross-lingual sentiment analysis models.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (62176076), the Natural Science Foundation of Guangdong (2023A1515012922), the Shenzhen Foundational Research Funding (JCYJ20220818102415032), and the Guangdong Provincial Key Laboratory of Novel Security Intelligence Technologies (2022B1212010005).

References

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

AI@Meta. 2024. Llama 3 model card.

Jeremy Barnes, Laura Oberlaender, Enrica Troiano, Andrey Kutuzov, Jan Buchmann, Rodrigo Agerri, Lilja Øvrelid, and Erik Velldal. 2022. SemEval 2022 task 10: Structured sentiment analysis. In *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*, pages 1280–1295, Seattle, United States. Association for Computational Linguistics.

Erik Cambria. 2016. Affective computing and sentiment analysis. *IEEE Intelligent Systems*, 31(2):102–107.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Shihan Dou, Enyu Zhou, Yan Liu, Songyang Gao, Jun Zhao, Wei Shen, Yuhao Zhou, Zhiheng Xi, Xiao Wang, Xiaoran Fan, Shiliang Pu, Jiang Zhu, Rui Zheng, Tao Gui, Qi Zhang, and Xuanjing Huang. 2024. Loramoe: Alleviate world knowledge forgetting in large language models via moe-style plugin.

Andrea Esuli, Alejandro Moreo, and Fabrizio Sebastiani. 2020. Cross-lingual sentiment quantification. *IEEE Intelligent Systems*, 35(3):106–114.

Fangxiaoyu Feng, Yinfei Yang, Daniel Cer, Naveen Arivazhagan, and Wei Wang. 2022. Language-agnostic BERT sentence embedding. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 878–891, Dublin, Ireland. Association for Computational Linguistics.

Jochen Hartmann, Mark Heitmann, Christian Siebert, and Christina Schamp. 2023. More than a feeling: Accuracy and application of sentiment analysis. *International Journal of Research in Marketing*, 40(1):75–87.

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2022. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*.

Dazhi Jiang, Runguo Wei, Jintao Wen, Geng Tu, and Erik Cambria. 2023. Automl-emo: Automatic knowledge selection using congruent effect for emotion identification in conversations. *IEEE Transactions on Affective Computing*, 14(03):1845–1856.

Bing Liu. 2012. *Sentiment Analysis and Opinion Mining*. Synthesis Lectures on Human Language Technologies. Springer Cham.

Aaron Maladry, Pranaydeep Singh, and Els Lefever. 2024. Findings of the wassa 2024 exalt shared task on explainability for cross-lingual emotion in tweets. In *Proceedings of the 14th Workshop of on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis@ACL 2024*, Bangkok, Thailand.

Parth Patwa, Gustavo Aguilar, Sudipta Kar, Suraj Pandey, Srinivas PYKL, Björn Gambäck, Tanmoy Chakraborty, Thamar Solorio, and Amitava Das. 2020. SemEval-2020 task 9: Overview of sentiment analysis of code-mixed tweets. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 774–790, Barcelona (online). International Committee for Computational Linguistics.

Sara Rosenthal, Noura Farra, and Preslav Nakov. 2017. SemEval-2017 task 4: Sentiment analysis in Twitter. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 502–518, Vancouver, Canada. Association for Computational Linguistics.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. 2023. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*.

Geng Tu, Bin Liang, Ruibin Mao, Min Yang, and Ruifeng Xu. 2023. Context or knowledge is not always necessary: A contrastive learning framework for emotion recognition in conversations. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 14054–14067.

Liang Wang, Nan Yang, Xiaolong Huang, Linjun Yang, Rangan Majumder, and Furu Wei. 2024. Multilingual e5 text embeddings: A technical report. *arXiv preprint arXiv:2402.05672*.

Frank Xing, Lorenzo Malandri, Yue Zhang, and Erik Cambria. 2020. Financial sentiment analysis: An investigation into common mistakes and silver bullets. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 978–987, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Wenxuan Zhang, Xin Li, Yang Deng, Lidong Bing, and Wai Lam. 2023. A survey on aspect-based sentiment analysis: Tasks, methods, and challenges. *IEEE Transactions on Knowledge and Data Engineering*, 35(11):11019–11038.

Chuanjun Zhao, Meiling Wu, Xinyi Yang, Wenyue Zhang, Shaoxia Zhang, Suge Wang, and Deyu Li. 2024. A systematic review of cross-lingual sentiment analysis: Tasks, strategies, and prospects. *ACM Computing Surveys*, 56(7).

Qihuang Zhong, Liang Ding, Juhua Liu, Bo Du, Hua Jin, and Dacheng Tao. 2023. Knowledge graph augmented network towards multiview representation learning for aspect-based sentiment analysis. *IEEE Transactions on Knowledge and Data Engineering*, 35(10):10098–10111.

Kudakwashe Zvarevashe and Oludayo O. Olugbara. 2018. A framework for sentiment analysis with opinion mining of hotel reviews. In *2018 Conference on Information Communications Technology and Society (ICTAS)*, pages 1–4.

A Prompt Template for Emotion Detection

To employ GPT-4 for Cross-lingual Emotion Detection tasks, we have designed an effective prompt template, as illustrated in Figure 2.

```
Assuming the role of a tweet analyst, please analyze a tweet now.
Tweet: {}
Requirement:
Emotion: Identify the primary emotion from the following options:
["Anger", "Fear", "Joy", "Love", "Neutral", "Sadness"].
Explantation: provide an explanation in English for the identified emotion.
Please format your response in JSON as shown below:
{{
"Emotion": "<insert identified emotion here>",
"Explanation": "<provide explanation for the identified emotion here>"
}}
```

Figure 2: The designed prompt template for GPT-4.