CYUT at SemEval-2026 Task 3: Multi-Task Dimensional Aspect Sentiment Regression with Polar Multi-Zone Labeling in VA Space

Shih-Hung Wu*, **Xian-Yan Chen**, **Yi-Min Jian**
Department of Computer Science and Information Engineering, Chaoyang University of Technology, Wufeng, Taichung, Taiwan
shwu@cyut.edu.tw, s11427605@gm.cyut.edu.tw, s11427601@gm.cyut.edu.tw

Abstract

This paper describes CYUT's system for SemEval-2026 Task 3 Track B, a multilingual aspect-based dimensional sentiment regression task. We formulate the task as continuous Valence–Arousal (VA) prediction and adopt a multi-task learning (MTL) framework with auxiliary tasks automatically derived from gold VA annotations, including polarity, intensity, and quadrant classification. However, these coarse-grained labels may still suffer from regional imbalance in the VA space, leaving some regions with insufficient auxiliary supervision. To address this issue, we extend the system with Polar Multi-Zone Labeling (PMZL) and use its seven-zone variant, PMZL-7. PMZL-7 partitions the VA plane into one core neutral region and six non-central zones based on the directional distribution of non-central samples. This design reduces the risk of auxiliary-label imbalance while supplementing directional information that conventional auxiliary tasks cannot directly capture. We evaluate XLM-R and two generative pretrained models. Results show that PMZL-7 is strongly model-dependent: it provides more stable improvements for Qwen2 and Ministral, while its effect on XLM-R is less consistent. On the official test set, our system achieves the best performance on the Nigerian-Pidgin subset among all participating systems.

1 Introduction

SemEval-2026 Task 3 Track B requires systems to predict continuous Valence–Arousal (VA) scores for a given input text and target aspect (Yu et al., 2026; Becker et al., 2026). Compared with classification settings based on discrete emotion or stance labels, VA representations provide a more fine-grained description of emotional direction, response intensity, and boundary-ambiguous samples, making them suitable for multilingual and cross-domain sentiment analysis (Russell, 1980; Buechel and Hahn, 2017; Mohammad et al., 2017; Küçük and Can, 2020).

In continuous affect prediction, a common strategy is to derive coarse-grained auxiliary labels from gold VA annotations and train them jointly with the main task. Polarity, intensity, and quadrant classification can provide complementary structural signals and stabilize representation learning in the VA space (Caruana, 1997; Buechel et al., 2018; Park et al., 2021; Xie et al., 2021). However, these labels are usually constructed using fixed thresholds or fixed quadrant partitions. When the data distribution is imbalanced across emotional directions, some auxiliary classes may contain too few samples, leading to unstable auxiliary supervision. Thus, if auxiliary label design does not consider the actual VA-space distribution, it may amplify the imbalance of the original data.

Based on this motivation, we introduce an additional partition-based auxiliary labeling method on top of polarity, intensity, and quadrant objectives. We view the VA space as a polar coordinate plane centered at the neutral point and divide it into multiple regions. We call this method Polar Multi-Zone Labeling (PMZL), and use its seven-zone version, PMZL-7. PMZL-7 assigns low-magnitude samples near the center to a core neutral region, then divides the remaining samples into six non-central zones according to their directional angle distribution, labeled Zone 1 to Zone 6. This design does not replace existing auxiliary tasks; instead, it provides a finer-grained and data-distribution-aware directional label to alleviate auxiliary class imbalance caused by fixed spatial partitioning.

In our shared task system, PMZL-7 is incorporated into the multi-task learning framework as an additional auxiliary task alongside polarity, intensity, and quadrant classification. Through ablation experiments, we compare different backbone models with and without PMZL-7 to examine whether polar-coordinate-based seven-zone label-

*Corresponding author

Proceedings of the The 20th International Workshop on Semantic Evaluation (2026), pages 153–159

ing can provide additional supervision for VA regression while reducing auxiliary label imbalance, and whether its effect varies across model architectures.

The main contributions of this paper are as follows:

- We present the CYUT system for SemEval-2026 Task 3 Track B, combining VA regression with polarity, intensity, quadrant classification, and PMZL-7 in a multi-task learning framework.
- We incorporate PMZL-7 as a polar-coordinate-based partition labeling method to alleviate class imbalance caused by fixed auxiliary label partitioning.
- We analyze PMZL-7 through ablation experiments and show that its benefit is model-dependent: it is more stable for generative models, but does not show a consistent advantage for XLM-R.

We release the system code for reproducibility: https://github.com/YiMin0130/CYUT-SemEval2026-Task3.git.

2 Related Work

2.1 Dimensional Sentiment Prediction

Dimensional affect representation describes emotional or stance states using continuous coordinates rather than a small set of discrete categories (Russell, 1980; Schlosberg, 1954). In natural language processing, EmoBank and related shared tasks have made VA/VAD regression a common evaluation setting (Buechel and Hahn, 2017; Yu et al., 2017; Lee et al., 2022). Compared with conventional sentiment classification or stance detection, dimensional representations better preserve emotional ambiguity, intensity variation, and continuous transitions (Mohammad et al., 2017; Küçük and Can, 2020). SemEval-2026 Task 3 further brings this type of continuous prediction into an aspect-based setting, making continuous dimensional prediction the core evaluation objective (Yu et al., 2026; Becker et al., 2026).

2.2 Multi-task Learning in Continuous Affect Prediction

Multi-task learning (MTL) improves model generalization by sharing representations across related tasks (Caruana, 1997). In continuous affect prediction, this strategy is particularly natural because polarity, intensity, and other auxiliary labels are structurally related to VA coordinates. Previous studies have shown that appropriate multi-task designs can improve model stability and prediction performance in multidimensional emotion prediction and low-resource settings (Buechel et al., 2018; Buechel and Hahn, 2018; Zhu et al., 2019; Mukherjee et al., 2021). For shared task system papers, these auxiliary tasks are usually better understood as comparable components in system design rather than as independent methodological claims.

2.3 Partition-based Auxiliary Labeling in Shared Tasks

In the context of this study, our main concern is whether another labeling method based on VA-space partitioning can be added on top of existing auxiliary tasks. Previous studies on the VA space suggest that, in addition to numerical values themselves, the direction, distance, and regional relationships of samples in the space may also be informative (Schlosberg, 1954; Cowen and Keltner, 2017; Park et al., 2021; Xie et al., 2021). These studies provide a reasonable starting point for treating partition-based labels as a complementary source of supervision in continuous affect prediction.

In this paper, we position PMZL-7 as a multi-zone auxiliary labeling method defined in a polar coordinate system for shared task system design. Specifically, we adopt the seven-zone version, PMZL-7, and examine whether it can serve as a complementary supervision signal beyond conventional polarity, intensity, and quadrant classification. Therefore, PMZL-7 is essentially a partition-based auxiliary labeling scheme.

3 Methods

3.1 Task Definition

Given an input text $T = \{w_1, \ldots, w_n\}$ and a target aspect x, the model is required to predict the corresponding VA coordinates:

$$y = (v, a) \in \mathbb{R}^2, \tag{1}$$

where v and a denote valence and arousal, respectively, and both values fall within the range $[1, 9]$. In this paper, we use VA regression as the main task and automatically derive multiple auxiliary labels from the gold VA annotations during training to form a multi-task learning framework.

3.2 Evaluation Metric

Following the official evaluation protocol of SemEval-2026 Task 3 Track B, we use Root Mean Square Error (RMSE) as the main evaluation metric. Since the task requires the simultaneous prediction of valence and arousal, the metric is defined as follows:

$$RMSE_{VA} = \sqrt{\sum_{i=1}^{N} \frac{\left(V_p^{(i)} - V_g^{(i)}\right)^2 + \left(A_p^{(i)} - A_g^{(i)}\right)^2}{N}} \tag{2}$$

where N is the number of samples. $V_p^{(i)}$ and $A_p^{(i)}$ denote the predicted valence and arousal of the i-th sample, respectively, while $V_g^{(i)}$ and $A_g^{(i)}$ denote the corresponding gold values. A lower RMSE indicates better model performance.

In addition, according to the official requirements, the final VA outputs must be constrained to the range $[1, 9]$ and rounded to two decimal places.

3.3 Per-aspect Sample Construction

Each text may contain multiple target aspects, and each aspect corresponds to an independent VA annotation. To avoid interference among signals from different aspects and to follow the official aspect-level evaluation protocol, we convert the original samples into per-aspect instances. For an original sample $(T, \{x_i\}_{i=1}^m)$, we construct:

$$(T, x_i) \rightarrow (v_i, a_i), \quad i = 1, \ldots, m. \tag{3}$$

Thus, each training instance contains only a single text–aspect pair and its corresponding VA annotation.

3.4 Instruction-based Formatting and Output Schema

For generative backbone models, we reformulate the regression task into an instruction-style format to improve output consistency and parsing stability. The input template is:

$$Input = \text{“}\texttt{Text: } T\texttt{, Aspect: } x\text{”} \tag{4}$$

During inference, the model outputs only the numerical result using the following fixed format:

$$Output = \text{“}\texttt{V\#A}\text{”} \tag{5}$$

where `V` and `A` denote the predicted valence and arousal, respectively, and are formatted to two decimal places, such as “`3.50#6.20`”. If the output cannot be parsed directly, we apply regex-based recovery and clip the numerical values to the valid range $[1, 9]$.

3.5 Auxiliary Tasks

In addition to the main VA regression task, we automatically derive four auxiliary tasks from the gold VA annotations in the training data. All auxiliary labels are generated automatically from the data and are used only during training.

(1) Polarity and intensity classification. To provide coarse-grained but stable affective signals, we define two three-class classification tasks:

- **Polarity (valence).** Samples with $V \geq 5.5$ are labeled as positive; samples with $V \leq 4.5$ are labeled as negative; the remaining samples are labeled as neutral.
- **Intensity (arousal).** Samples with $A \geq 6.0$ are labeled as excited; samples with $A \leq 4.0$ are labeled as calm; the remaining samples are labeled as medium.

(2) Quadrant classification. To capture the interaction between valence and arousal, we divide the VA plane into four quadrants using the center point $(5, 5)$. This task provides a coarse-grained joint structure for the two-dimensional affective space.

(3) Polar Multi-Zone Labeling (PMZL). In addition to the conventional auxiliary tasks, we further introduce Polar Multi-Zone Labeling (PMZL) as an additional auxiliary objective. In this paper, we use the seven-zone version, denoted as PMZL-7. This design is not intended to replace polarity, intensity, or quadrant classification. Instead, it serves as a partition-based auxiliary labeling method based on the polar coordinate system to supplement angular information in the VA space.

For each sample (v, a), we first perform a polar coordinate transformation:

$$r = \sqrt{(v-5)^2 + (a-5)^2} \tag{6}$$

$$\theta = atan2(a-5,\ v-5) \tag{7}$$

where r denotes the distance from the sample to the center, and θ denotes its directional angle in the VA plane.

The construction of PMZL-7 labels consists of two steps:

Step 1: Core neutral region. Samples close to the center have very small radii, making their directional angles highly sensitive to small perturbations. Directly assigning these samples to directional zones may lead to unstable labels. Therefore, we first assign the closest 20% of training samples to the core neutral region.

Step 2: Six non-central zones. For the remaining non-central samples, we do not use fixed-angle partitioning. Instead, we divide them into six approximately equal-mass segments according to the angle distribution in the training data. This results in one core neutral region and six non-central zones, namely PMZL-7. The six outer zones are labeled sequentially as Zone 1 to Zone 6.

Figures 1 to 4 show the partitioning schemes and class distributions of different auxiliary labels on the Chinese training set. It is worth noting that the seven classes shown in Figure 4 have relatively similar distributions. In contrast, the class distributions produced by conventional partitioning methods are more imbalanced. This observation indicates that the labels derived by PMZL-7 are relatively more balanced. Whether this helps improve model performance is examined in the subsequent experiments.

3.6 Backbone Architectures and Parameter-Efficient Fine-tuning

We compare encoder-based and decoder-based models for the multilingual VA regression task.

Encoder-based model (XLM-R). We use XLM-RoBERTa-base (XLM-R) as the discriminative baseline model (Conneau et al., 2020). XLM-R learns robust cross-lingual representations through large-scale multilingual masked language pretraining. In this paper, we feed its final hidden representation into a linear regression head to predict the two-dimensional VA coordinates.

Decoder-based LLMs. We use Qwen2-7B-Instruct from the Qwen model family and Ministral-7B from Mistral AI as generative backbones(Yang et al., 2024; Jiang et al., 2023). To reduce the training cost under the multi-task setting, we adopt Low-Rank Adaptation (LoRA) for parameter-efficient fine-tuning (Hu et al., 2021). Its formulation is as follows:

$$h = W_0x + \Delta Wx = W_0x + BAx \qquad (8)$$

where W_0 denotes the frozen pretrained weights, A and B are trainable low-rank matrices, and the rank $r \ll d$. This design significantly reduces the number of trainable parameters while preserving pretrained knowledge.

3.7 Multi-task Training Objective

We first construct polarity, intensity, quadrant, and PMZL-7 labels based on the training data of each language, and then merge all aspect-level instances for multilingual joint training. For PMZL-7, the core-region threshold and zone boundaries are estimated independently for each language to preserve language-specific distributional characteristics.

For generative large language models (LLMs), all tasks are formulated as constrained generation problems and jointly optimized within the shared model. For XLM-R, we use a shared encoder with task-specific output heads. In all settings in this paper, the model jointly learns five objectives: the main VA regression task and four auxiliary tasks, including polarity, intensity, quadrant classification, and PMZL-7.

For generative models, we use equal weights for

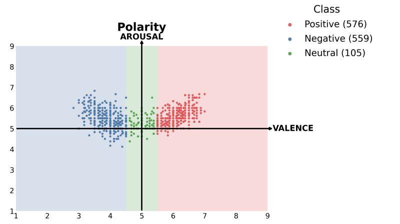

Figure 1: Partitioning scheme and class distribution of polarity classification.

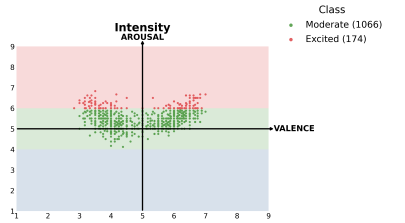

Figure 2: Partitioning scheme and class distribution of intensity classification.

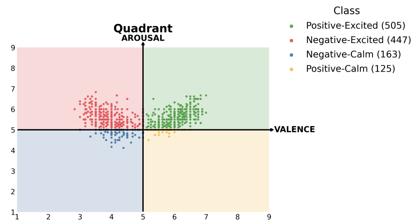

Figure 3: Partitioning scheme and class distribution of quadrant classification.

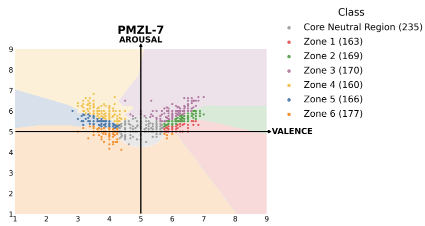

Figure 4: PMZL-7: seven zones defined in the polar coordinate system of the VA space.

Language	Domain	Train	Dev	Test	Total
English (ENG)	Environmental Protection	922	200	1020	2142
Chinese (ZHO)	Environmental Protection	700	100	600	1400
German (DEU)	Politics	683	34	263	980
Nigerian Pidgin (PCM)	Politics	1049	119	331	1499
Swahili (SWA)	Politics	1375	123	266	1764

Table 1: Dataset statistics for SemEval-2026 Task 3 Track B.

the main task and auxiliary tasks. For XLM-R, the main regression loss is given the highest priority, while the auxiliary losses are assigned relatively lower weights.

For fine-tuning generative models, we use QLoRA/LoRA with rank $r = 16$, $\alpha = 32$, and dropout $= 0.05$. The target modules include `q/k/v/o_proj` and `gate/up/down_proj`. Other training parameters are as follows: the learning rate is 1×10^{-4}, the batch size is 1, gradient accumulation is 4, the maximum input length is 512, the number of training epochs is 1, and we use bf16 precision with gradient checkpointing.

4 Data and Preprocessing

We use the official dataset of SemEval-2026 Task 3 Track B (Yu et al., 2026; Becker et al., 2026). Each text may contain one or more target aspects, and each aspect is associated with a pair of continuous VA annotations. The values range from 1 to 9, where 5 corresponds to neutral valence and medium arousal. Table 1 summarizes the languages, domains, and split sizes covered in this task.

During preprocessing, we first convert multi-aspect samples into multiple per-aspect instances. For generative models, all inputs are reformatted using the instruction-style template described above, and the outputs are restricted to the `V#A` format. If the model output does not fully follow the required format, we apply regex-based parsing and range clipping to ensure that the submitted results comply with the official requirements.

5 Experimental Results

We use the official RMSE metric to compare different backbone models and multi-task settings. As shown in Table 2, the impact of multi-task learning differs across model families. Compared with the No-MTL setting, the best multi-task variant of XLM-R reduces the average RMSE from 1.7339 to 1.5528. Ministral improves from 1.5673 to 1.3950, while the complete 5-task setting of Qwen2 also achieves its best average result of 1.5184. These results show that, under the multilingual shared task setting, multi-task learning is generally helpful for VA regression, but different backbones absorb auxiliary supervision in different ways.

This paper specifically focuses on whether PMZL-7 provides additional benefits beyond the existing auxiliary tasks. The ablation results show that, for Qwen2 and Ministral, removing PMZL-7 leads to higher average RMSE than the complete 5-task model. This indicates that PMZL-7 provides additional complementary signals for these two models. For XLM-R, however, the w/o PMZL-7 variant slightly outperforms the complete model, suggesting that the benefit of PMZL-7 is strongly model-dependent rather than consistently effective across all architectures.

Regarding the importance of individual auxiliary tasks, intensity is the most stable key signal. When intensity is removed, all three backbones show clear performance degradation, indicating that coarse-grained arousal supervision is important for continuous VA prediction. In contrast, the effect of quadrant supervision is more model-dependent. For Qwen2, removing quadrant causes performance degradation, whereas for Ministral, the best result appears in the w/o Quadrant setting. This suggests that different backbones rely on coarse-grained regional structures and angular partition information to different degrees.

On the official test set (Table 3), we submit a unified 5-task system. Qwen2-7B achieves the best overall average RMSE of 1.3999, followed by Ministral-7B with 1.4817 and XLM-R with 1.6612. For individual languages, Qwen2 performs best on ENG, DEU, and PCM, Ministral performs best on ZHO, and XLM-R is relatively stable on SWA. Notably, among our systems, Qwen2 achieves the best result on PCM, with an RMSE of 1.1024.

Overall, the experimental results can be summarized in three points. First, PMZL-7 is more stable for generative models. Second, PMZL-7 does not show a stable advantage for XLM-R. Third, when

Backbone	Setting	VA	Polarity	Intensity	Quadrant	PMZL-7	ENG	DEU	ZHO	PCM	SWA	Avg↓
Qwen2	No-MTL	✓	–	–	–	–	1.9231	1.4435	0.5826	1.5095	2.2110	1.5339
Qwen2	w/o Polarity	✓	–	✓	✓	✓	1.8640	1.5119	0.5514	1.5177	2.3078	1.5506
Qwen2	w/o Intensity	✓	✓	–	✓	✓	1.8936	1.4965	0.5533	1.5929	2.4846	1.6042
Qwen2	w/o Quadrant	✓	✓	✓	–	✓	1.9090	1.4986	0.5608	1.5521	2.3671	1.5775
Qwen2	w/o PMZL-7	✓	✓	✓	✓	–	**1.7961**	**1.4349**	0.5424	1.5728	2.3846	1.5462
Qwen2	MTL (5-task)	✓	✓	✓	✓	✓	1.8912	1.5384	**0.5236**	**1.4481**	**2.1908**	**1.5184**
XLM-R	No-MTL	✓	–	–	–	–	2.1337	1.5711	0.7227	1.7855	2.4567	1.7339
XLM-R	w/o Polarity	✓	–	✓	✓	✓	**2.0761**	1.7607	0.7119	**1.1236**	2.0918	**1.5528**
XLM-R	w/o Intensity	✓	✓	–	✓	✓	2.1721	1.8826	0.9069	1.3583	2.1049	1.6849
XLM-R	w/o Quadrant	✓	✓	✓	–	✓	2.0945	1.8178	0.6814	1.6729	2.1796	1.6892
XLM-R	w/o PMZL-7	✓	✓	✓	✓	–	2.2338	**1.5083**	**0.6442**	1.4426	**1.9784**	1.5615
XLM-R	MTL (5-task)	✓	✓	✓	✓	✓	2.1217	1.7988	0.7166	1.2819	1.9907	1.5820
Ministral	No-MTL	✓	–	–	–	–	1.9414	1.4506	0.7752	1.3391	2.3303	1.5673
Ministral	w/o Polarity	✓	–	✓	✓	✓	1.8331	**1.3765**	0.6332	1.1642	2.0881	1.4190
Ministral	w/o Intensity	✓	✓	–	✓	✓	1.8187	1.5518	0.6270	1.2497	2.2231	1.4941
Ministral	w/o Quadrant	✓	✓	✓	–	✓	**1.7568**	1.4230	0.6498	1.0811	**2.0642**	**1.3950**
Ministral	w/o PMZL-7	✓	✓	✓	✓	–	1.8011	1.6342	**0.6263**	0.9737	2.1507	1.4372
Ministral	MTL (5-task)	✓	✓	✓	✓	✓	1.7601	1.5471	0.6767	**0.9645**	2.1284	1.4154

Table 2: Ablation results on the development sets of five languages.

Backbone	ENG	DEU	ZHO	PCM	SWA	Avg↓
Official Baselines:						
mBERT	2.6985	2.3294	1.2756	3.2152	2.7835	2.4604
Mistral-3 14B	1.6430	1.5910	0.7400	1.7390	2.2990	1.6024
Our Systems (5-task):						
Qwen2-7B	**1.6331**	**1.4827**	0.6771	**1.1024**	2.1042	**1.3999**
Ministral-7B	1.7793	1.5153	**0.6631**	1.1851	2.2659	1.4817
XLM-R	2.1330	1.6592	0.7056	1.7634	**2.0446**	1.6612

Table 3: RMSE results on the official test sets of five languages. Our multi-task models outperform the official encoder-based baseline (mBERT) and decoder-based baseline (Mistral-3 14B) in terms of average performance.

combined with the results in Table 3, our method performs better on PCM. Based on these observations, PMZL-7 is better understood as an additional auxiliary task worth comparing and analyzing, rather than as a fixed design that is universally effective for all models.

6 Conclusion and Future Work

This paper describes the CYUT system for SemEval-2026 Task 3 Track B. We formulate the task as VA regression and combine polarity, intensity, quadrant classification, and PMZL-7 in a multi-task learning framework. This paper adds a seven-zone design based on the polar coordinate system on top of the original auxiliary supervision. The purpose is to supplement directional information that conventional auxiliary tasks cannot easily express and to reduce the auxiliary label imbalance that may result from fixed partitioning of the VA space.

The experimental results show that PMZL-7 is more stable for generative models, but it does not show a consistent advantage for XLM-R. At the same time, our method achieves the best result on Nigerian Pidgin. These results indicate that, in addition to conventional polarity, intensity, and quadrant classification, auxiliary labels based on polar coordinate partitioning are also worth comparing and analyzing in shared task system design. It should be noted that PMZL-7 does not consistently improve all models. Therefore, its role is better understood as a design that can alleviate auxiliary label distribution imbalance and supplement directional supervision signals, rather than as a universally effective performance enhancement method.

Future work can be extended in three directions. First, we will compare different partition granularities, such as 5, 6, 8, or 9 zones, to analyze the relationship among the number of zones, class balance, and multilingual data distributions. Second, we will explore the possibility of extending hard partitioning to soft zone assignment, in order to examine whether smoother regional labels can further improve the learning of boundary samples. Third, we will further analyze the distributional differences among languages in the VA space and introduce more fine-grained calibration strategies to reduce multilingual prediction errors.

Acknowledgments

This work was supported by the National Science and Technology Council under Grant NSTC114-2221-E-324-006.

References

Jonas Becker, Liang-Chih Yu, Shamsuddeen Hassan Muhammad, Jan Philip Wahle, Terry Ruas, Idris Abdulmumin, Lung-Hao Lee, Nelson Odhiambo, Lilian Wanzare, Wen-Ni Liu, Tzu-Mi Lin, Zhe-Yu Xu, Ying-Lung Lin, Jin Wang, Maryam Ibrahim Mukhtar, Bela Gipp, and Saif M. Mohammad. 2026. Dimstance: Multilingual datasets for dimensional stance analysis. *Preprint*, arXiv:2601.21483.

Sven Buechel, Anneke Buffone, Barry Slaff, Lyle Ungar, and João Sedoc. 2018. Modeling empathy and distress in reaction to news stories. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4758–4765, Brussels, Belgium. Association for Computational Linguistics.

Sven Buechel and Udo Hahn. 2017. Emobank: Studying the impact of annotation perspective and representation format on dimensional emotion analysis. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics.

Sven Buechel and Udo Hahn. 2018. Word emotion induction for multiple languages as a deep multi-task learning problem. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1907–1918. Association for Computational Linguistics.

Rich Caruana. 1997. Multitask learning. *Machine Learning*, 28(1):41–75.

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. Association for Computational Linguistics.

Alan S. Cowen and Dacher Keltner. 2017. Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proceedings of the National Academy of Sciences*, 114(38):E7900–E7909.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*.

Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, György Lengyel, Guillaume Lample, Luc Saulnier, and 1 others. 2023. Mistral 7b. *arXiv preprint arXiv:2310.06825*.

Dilek Küçük and Fazli Can. 2020. Stance detection: A survey. *ACM Computing Surveys*, 53(1).

Lung-Hao Lee, Jheng-Hong Li, and Liang-Chih Yu. 2022. Chinese EmoBank: Building valence-arousal resources for dimensional sentiment analysis. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 21(4):1–19.

Saif M. Mohammad, Parinaz Sobhani, and Svetlana Kiritchenko. 2017. Stance and sentiment in tweets. *ACM Transactions on Internet Technology*, 17(3):1–23.

Rajdeep Mukherjee, Aakash Naik, Saptarshi Poddar, Somak Dasgupta, and Niloy Ganguly. 2021. Understanding the role of affect dimensions in detecting emotions from tweets: A multi-task approach. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2305–2309.

Sungjoon Park, Jiseon Kim, Seonghyeon Ye, Jaeyeol Jeon, Hee Young Park, and Alice Oh. 2021. Dimensional emotion detection from categorical emotion. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 4367–4380. Association for Computational Linguistics.

James A. Russell. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178.

Harold Schlosberg. 1954. Three dimensions of emotion. *Psychological Review*, 61(2):81–88.

Hongfei Xie, Eduard Hovy, and Diyi Yang. 2021. A multi-dimensional relation model for dimension score prediction. *Information Sciences*, 579:832–844.

An Yang and 1 others. 2024. Qwen2 technical report. *arXiv preprint arXiv:2407.10671*.

Liang-Chih Yu, Jonas Becker, Shamsuddeen Hassan Muhammad, Idris Abdulmumin, Lung-Hao Lee, Ying-Lung Lin, Jin Wang, Jan Philip Wahle, Terry Ruas, Alexander Panchenko, Ilseyar Alimova, Kai-Wei Chang, Lilian Wanzare, Nelson Odhiambo, Bela Gipp, and Saif M. Mohammad. 2026. SemEval-2026 task 3: Dimensional aspect-based sentiment analysis (DimABSA). In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*. Association for Computational Linguistics.

Liang-Chih Yu, Lung-Hao Lee, Jin Wang, and Kam-Fai Wong. 2017. IJCNLP-2017 task 2: Dimensional sentiment analysis for chinese phrases. In *Proceedings of the IJCNLP 2017 Shared Tasks*, pages 17–24. Asian Federation of Natural Language Processing.

Siyi Zhu, Yu Zhang, and Xiaobing Li. 2019. Adversarial attention modeling for multi-dimensional emotion regression. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 631–640. Association for Computational Linguistics.