# YNU-HPCC at SemEval-2026 Task 2: Contrastive Calibration and Temporal Modeling for Continuous Valence-Arousal Prediction

**Xin Lan, Jin Wang and Xuejie Zhang**
School of Information Science and Engineering
Yunnan University
Kunming, China
xinlan@stu.ynu.edu.cn, {wangjin,xjzhang}@ynu.edu.cn

## Abstract

This paper addresses continuous affect modeling in SemEval-2026 Task 2 through two task-specific architectures tailored to static state estimation and dynamic change prediction. To mitigate semantic ambiguity and annotation subjectivity in Subtask 1, a hard-prompt-based regression model is developed and enhanced with unsupervised contrastive learning (SimCSE) and supervised contrastive calibration (SCL) grounded in an external affect lexicon. This design improves the structural consistency and scale stability of textual representations in the Valence–Arousal (V/A) space. For Subtask 2a, which involves irregular time intervals and historical dependencies, a Time-Aware LSTM architecture is introduced to integrate current affective states with temporally enriched historical trajectories. Experimental results show that the YNU-HPCC system ranks 2nd in both subtasks. In Subtask 1, the Valence and Arousal scores are 0.677 and 0.528, respectively; in Subtask 2a, they are 0.692 and 0.647.

## 1 Introduction

Valence and Arousal (V/A) form an established continuous representation framework for modeling human emotional states in sentiment analysis. SemEval-2026 Task 2 focuses on modeling affective states in the V/A space and their temporal evolution using longitudinal, first-person, self-reported texts (Soni et al., 2026). This task comprises two related yet distinct subtasks: Subtask 1 performs V/A state regression, while Subtask 2a predicts affective changes between adjacent time steps.

As continuous affective dimensions, V/A represent emotions through numerical regression rather than discrete classification, inherently characterized by strong subjectivity and annotation noise (Han et al., 2021; Ghosal et al., 2020; Tits et al., 2019). Consequently, SemEval-2026 Task 2 presents multiple challenges. In Subtask 1, the mapping between textual semantics and continuous affective values is highly ambiguous. Compounding this difficulty, Subtask 2a introduces a temporal dimension. Models must not only interpret individual texts but also capture affective evolution under irregular time intervals, modeling the nonlinear influence of historical states on future affective changes.

Recent affect modeling research has shifted from traditional supervised learning toward representation enhancement based on large-scale pretrained language models (PLMs). To improve representation stability, contrastive learning methods such as SimCSE (Gao et al., 2021) have been introduced to alleviate anisotropies in PLM embedding spaces. Meanwhile, supervised contrastive learning (SCL) (Khosla et al., 2020), combined with external affect lexicons, has become an effective technique for aligning representation spaces with affective dimensions. In temporal modeling, sequence architectures based on Transformers (Tang et al., 2014) or LSTM variants equipped with time-aware mechanisms (Baytas et al., 2017; Poria et al., 2019) have demonstrated effectiveness in capturing dynamic temporal patterns.

Building upon prior work, two specialized architectures are developed: one for static affect regression and the other for time-aware affect change prediction. The former improves the stability of the mapping between textual representations and the V/A space through representation optimization, while the latter models affective evolution trends by integrating historical affect states with temporal information. For Subtask 1, a regression framework is constructed grounded in semantic smoothing and external knowledge calibration. Building upon hard prompting, which explicitly activates affective priors, contrastive learning is employed to enhance representation robustness and alignment within the V/A space. For Subtask 2a, a time-aware sequential modeling architecture is designed that leverages historical affect trajectories enriched with



*Proceedings of the The 20th International Workshop on Semantic Evaluation (2026)*, pages 295–301
July 3-4, 2026 ©2026 Association for Computational Linguistics

temporal signals to capture irregular evolutionary patterns. Experimental results indicate that the proposed strategy effectively tackles uncertainty and temporal modeling challenges in continuous affect prediction, confirming the effectiveness of the differentiated framework and yielding competitive performance on the leaderboard.

The rest of this paper is organized as follows. Section 2 reviews related work. Sections 3 and 4 introduce the modeling frameworks for Subtask 1 and Subtask 2a, respectively. Section 5 describes the experimental setup and presents empirical results. Section 6 concludes the paper.

## 2 Related Work

In static affect representation learning, PLMs' embedding spaces often exhibit anisotropy, limiting their direct applicability to regression tasks. To address this issue, researchers have incorporated affect lexicons to constrain representations within continuous affective spaces. For instance, Poria et al. (2019) explored label mapping and prediction strategies for dimensional affect modeling. To enhance discriminative power, contrastive learning has been introduced to mitigate subjective noise; Pinitas et al. (2022) integrated continuous annotations into representation learning through supervised contrastive objectives. Additionally, Zhong et al. (2019) employed affect-enriched graph networks to strengthen word-level representations, while Tian et al. (2020) leveraged large-scale weakly supervised data for contrastive pretraining. More recently, Bulla and Mongiovì (2024) demonstrated that carefully designed hard prompt templates can activate affective priors in pretrained models, thereby improving continuous emotion regression performance. These studies validate the effectiveness of semantic calibration in continuous spaces.

In dynamic affect modeling, prior work emphasizes the temporal evolution of emotions, typically incorporating historical states via RNN-based or sequential architectures to capture temporal dependencies. To address irregular time intervals in real-world scenarios, Li et al. (2020) explicitly introduced time interval information to modulate the influence of historical states. Sawhney et al. (2020) further proposed time-gated mechanisms to model affective decay and nonlinear evolution patterns. Despite these advances, integrating time-aware mechanisms with PLMs remains underexplored. However, prior work rarely integrates representation calibration and temporal modeling under a task-differentiated framework for continuous affect prediction.

## 3 Affect State Estimation (Subtask 1)

Given the distinct challenges of static V/A mapping and dynamic state transitions, this system adopts independent modeling strategies for the two subtasks. The overall architecture of the proposed system is illustrated in Figure 1.

### 3.1 Input Reconstruction.

During downstream fine-tuning, a template function $T(\cdot)$ is employed to restructure the input text:

$$x_{\text{prompt}} = T(x) \tag{1}$$

The template explicitly incorporates the tokens Valence and Arousal into the input sequence, guiding the model to attend to semantic cues associated with continuous affective dimensions during encoding. This template is applied exclusively during task-specific fine-tuning and inference and is not involved in representation pretraining.

### 3.2 Multi-Stage Training Strategy.

An affect-aware encoder $E_\theta(\cdot)$ is constructed and trained under a three-stage optimization scheme to progressively enhance its sensitivity to continuous Valence–Arousal dimensions.

**Unsupervised Semantic Smoothing (SimCSE).** Unsupervised contrastive learning (SimCSE) is first applied to improve the isotropy and stability of sentence representations. Positive pairs are generated via dropout-induced stochastic perturbations of the same input, while in-batch samples serve as negatives. Optimization with the InfoNCE objective encourages semantically consistent representations.

**Supervised Contrastive Calibration with Affect Lexicon (SCL).** To align representations with continuous affective dimensions, supervised contrastive calibration is performed using a lexicon annotated with V/A values. During this stage, Transformer parameters remain frozen while the embedding matrix is updated. Lexicon entries are organized via K-nearest neighbors in the V/A space, and a supervised contrastive loss reduces distances among affectively similar words.

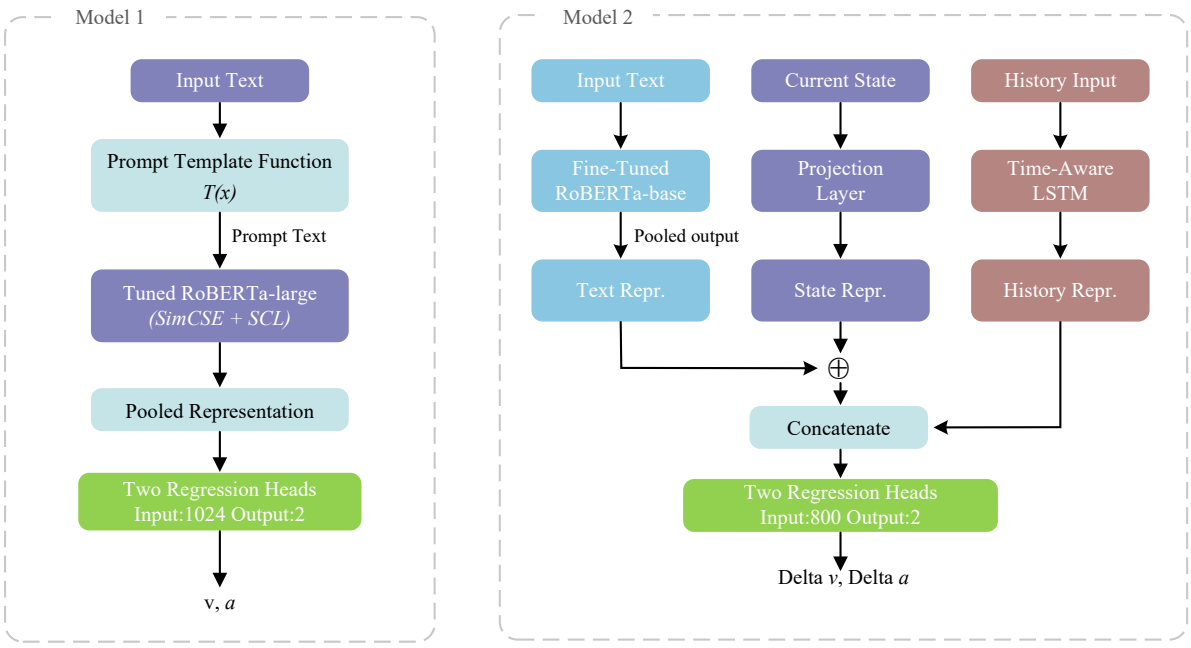


Figure 1: The overall architecture of the proposed system.

**Supervised Regression Fine-Tuning on Official Data.** Initializing with SCL-calibrated encoder weights, the system incorporates the prompt template and uses a linear regression head to predict V/A values directly. The model is optimized by minimizing the Mean Squared Error (MSE), which drives the encoder to adapt to the target task distribution while preserving semantic expressiveness and enabling precise mapping from textual features to continuous affective coordinates.

### 3.3 Regression Head Design

The encoder's pooler output serves as the input to the regression head. After passing through a dropout layer, the high-dimensional semantic representation is projected via a single fully connected layer to a two-dimensional output corresponding to V/A, producing the final prediction.

## 4 Affect Change Prediction (Subtask 2a)

### 4.1 Input Reconstruction

At time step $t$, the model processes three inputs: the current text, the current affective state, and the historical affect trajectory. A parallel encoding structure is employed, followed by feature-level fusion.

**Text Representation.** The input text is encoded using RoBERTa-Base. The encoder's pooled output serves as the sentence-level semantic representation.

**Current Affective State.** First, the continuous V/A values are concatenated into a two-dimensional vector and projected through a transformation block (Linear $\rightarrow$ GELU $\rightarrow$ Dropout) to obtain a 768-dimensional affective embedding aligned with the textual feature space.

**Historical Information.** The historical sequence, consisting of the past h affective states (V/A) and their corresponding time intervals $\Delta t$, is fed into a Time-Aware LSTM module. During temporal iteration, a time gate is generated by applying a linear transformation followed by a Sigmoid activation to the time interval $\Delta t$. This time-gate modulates the decay of the previous cell state, enabling adaptive attenuation based on elapsed time. The LSTM unit then updates its hidden state by combining the decayed cell state with the affective input at the current step through standard gating mechanisms. The hidden state at the final time step is extracted as the historical representation, resulting in a 32-dimensional feature vector.

### 4.2 Fusion Layer

First, the textual semantic representation and the projected current affective embedding are combined element-wise. The resulting feature is then concatenated with the LSTM-produced historical feature vector, yielding an 800-dimensional fused representation.

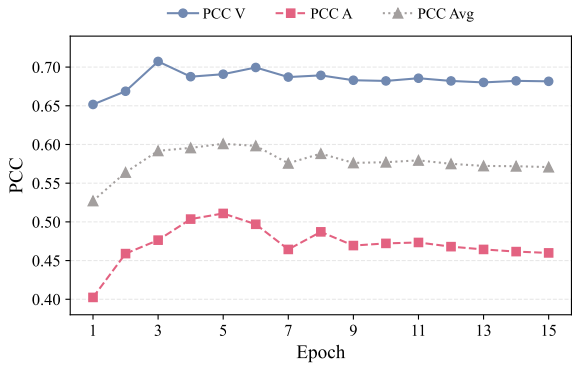


Figure 2: Performance trends on the validation set across different training epochs for Subtask 1.

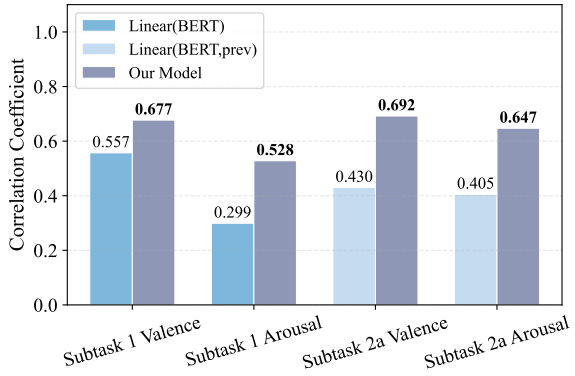


Figure 3: Comparison between the model and official baselines on Subtask 1 and Subtask 2a.

### 4.3 Regression Head Design

The fused representation is fed into two independent MLP regression heads to decouple Valence and Arousal change prediction. Each head consists of a dropout layer, a linear layer that compresses the feature dimension to 64, a GELU activation, and a final output layer that predicts the change values $\widehat{\Delta v}_t$ and $\widehat{\Delta a}_t$. Formally,

$$\begin{aligned} z_t &= \mathrm{Pool}(E_\phi(x_t)) + \mathrm{Proj}(y_t), \\ h_t &= \mathrm{Concat}(z_t, \mathrm{T{-}LSTM}(H_t)), \\ \widehat{\Delta v}_t &= \mathrm{Head}_{\Delta v}(h_t), \\ \widehat{\Delta a}_t &= \mathrm{Head}_{\Delta a}(h_t). \end{aligned} \tag{2}$$

Here, $\mathrm{Pool}(\cdot)$ denotes the pooled text embedding, $\mathrm{Proj}(\cdot)$ is the state projection layer, and $\mathrm{T{-}LSTM}(\cdot)$ models temporal dynamics.

## 5 Experiment

### 5.1 Dataset Description

This study addresses the two subtasks of SemEval-2026 Task 2 using different data sources and partitioning strategies tailored to their respective modeling objectives. For Subtask 1, the official training data serves as the primary supervised dataset, supplemented with external resources for auxiliary training. Specifically, the training split of GoEmotions (Demszky et al., 2020) is used for unsupervised representation learning, while `Ratings_Warriner_et_al.csv` (Warriner et al., 2013), which provides word-level V/A annotations, is employed for affective representation calibration. The system adopts a seen/unseen user split strategy. The validation set consists of all samples from a subset of unseen users and a portion of samples from seen users, accounting for 20% of the total data, while the remaining 80% is used for training. For Subtask 2a, only official data are used, without any external resources. The dataset is partitioned by user, and within each user's subset, an 80/20 split is applied for training and validation.

### 5.2 Implementation Details

For Subtask 1, RoBERTa-Large is utilized as the backbone text encoder, while RoBERTa-Base is adopted for Subtask 2a. For Subtask 1, the system first performs unsupervised SimCSE pretraining on the GoEmotions dataset to obtain robust sentence representations. The system then uses word-level Valence–Arousal annotations from `Ratings_Warriner_et_al.csv` for supervised contrastive calibration. The input text is reformulated using manually designed hard prompt templates, followed by full-parameter regression fine-tuning on the official labeled data. The model outputs continuous V/A values.

For Subtask 2a, only the official dataset is used. The model takes the current text representation, the current affective state, and the user's historical affect sequence as inputs. A Time-Aware LSTM module processes the historical sequence, and its output is fused with other features to predict affective change values.

All textual inputs are tokenized using the RoBERTa tokenizer. Both subtasks adopt MSE as the loss function and are optimized using the AdamW optimizer.

For Subtask 1, the learning rate is set to $1 \times 10^{-5}$ with a warmup ratio of 0.1. For Subtask 2a, a layered learning rate strategy is adopted: the pretrained backbone is assigned a lower learning rate of $2 \times 10^{-5}$, while task-specific modules—including the Time-Aware LSTM and regression heads—use a higher learning rate of $1 \times 10^{-3}$

| Model | Prompting | Calibration | PCC (Avg) | CCC (Avg) |
|---|---|---|---|---|
| RoBERTa-Large | No | No | 0.5767 | 0.5507 |
| w/ Prompting | Yes | No | 0.5838 | 0.5616 |
| Ours | Yes | Yes | **0.6009** | **0.5731** |

Table 1: Ablation study on Task 1. Comparison of the baseline with Prompting and Calibration variants.

| Model | State Context | Temporal Modeling | PCC (Avg) | CCC (Avg) |
|---|---|---|---|---|
| RoBERTa-Base | No | No | 0.3026 | 0.2654 |
| w/ State Context | Yes | No | 0.5690 | 0.5122 |
| Ours (Time-Aware) | Yes | Yes | **0.6161** | **0.5751** |

Table 2: Ablation study on Task 2. Evaluating the effectiveness of State Context and Temporal Modeling.

to accelerate downstream convergence.

### 5.3 Model Selection Strategy

For Subtask 1, due to the limited size of the official dataset and the absence of a public test set, a validation-guided retraining strategy is adopted. The optimal training epoch is determined based on validation performance (Figure 2). Subsequently, under the same hyperparameter configuration and selected epoch, the model is retrained on the entire official training set to produce the final submission. This procedure enables objective model selection while maximizing the utilization of limited labeled data.

For Subtask 2a, the system directly uses the checkpoint with the best validation performance for final prediction.

### 5.4 Ablation Study

To verify the contribution of each key module, ablation experiments are conducted on both Subtask 1 and Subtask 2a, as shown in Tables 1 and 2.

For Subtask 1, incorporating prompting yields consistent performance improvements, particularly in CCC, suggesting that explicit affective semantic guidance contributes to continuous emotion modeling. Building on this, introducing Affect Calibration further improves PCC and CCC. This indicates that aligning the representation space with an affect lexicon helps mitigate scale bias and subjective noise in continuous regression.

For Subtask 2a, the text-only baseline performs poorly. Introducing the current affective state as contextual input substantially improves prediction accuracy. Further incorporating time-aware historical modeling yields the best results across PCC and CCC, highlighting the importance of jointly modeling affective states and temporal dynamics for affect change prediction.

Overall, the results demonstrate that each module provides distinct and complementary contributions across the two subtasks.

### 5.5 Results and Analysis

The proposed system is evaluated against the official baselines (Figure 3).

On Subtask 1, this model achieves 0.677/0.528 (Valence/Arousal), outperforming Linear (BERT) (0.557/0.299) with lower MAE. The substantial improvement in Arousal indicates that contrastive calibration and prompting enhance structural alignment in the V/A space, particularly for dimensions more susceptible to annotation variance.

On Subtask 2a, correlations improve from 0.430/0.405 to 0.692/0.647 (+0.262/+0.242). The larger improvement in Valence indicates that temporal modeling effectively captures gradual affective transitions, while Arousal remains comparatively more volatile, reflecting the intrinsically higher volatility of Arousal in longitudinal affect prediction.

## 6 Conclusion

This work addresses continuous affect modeling in SemEval-2026 Task 2 through two task-specific models: a static regression model and a time-aware change prediction model. For Subtask 1, representation optimization improves alignment in the Valence–Arousal space and consistently outperforms the official baseline. For Subtask 2a, modeling historical states and temporal information enhances Valence prediction, while Arousal remains less stable, suggesting greater temporal variability. Future work will focus on more robust fusion strategies to improve stability and generalization.

## Acknowledgement

This work was supported by the National Natural Science Foundation of China (NSFC) under Grant Nos.61966038 and 62266051. The authors would like to thank the anonymous reviewers for their constructive comments.

## References

Inci M. Baytas, Cao Xiao, Xi Zhang, Fei Wang, Anil K. Jain, and Jiayu Zhou. 2017. Patient subtyping via time-aware lstm networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '17, page 65–74, New York, NY, USA. Association for Computing Machinery.

Luana Bulla and Misael Mongiovì. 2024. Adequate prompting improves performance of regression models of emotional content. In *Proceedings of the 2024 International Conference on Information Technology for Social Good*, GoodIT '24, page 135–142, New York, NY, USA. Association for Computing Machinery.

Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. 2020. GoEmotions: A dataset of fine-grained emotions. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4040–4054, Online. Association for Computational Linguistics.

Tianyu Gao, Xingcheng Yao, and Danqi Chen. 2021. SimCSE: Simple contrastive learning of sentence embeddings. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 6894–6910, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Deepanway Ghosal, Navonil Majumder, Alexander Gelbukh, Rada Mihalcea, and Soujanya Poria. 2020. COSMIC: COmmonSense knowledge for eMotion identification in conversations. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2470–2481, Online. Association for Computational Linguistics.

Wei Han, Hui Chen, Alexander Gelbukh, Amir Zadeh, Louis-philippe Morency, and Soujanya Poria. 2021. Bi-bimodal modality fusion for correlation-controlled multimodal sentiment analysis. In *Proceedings of the 2021 International Conference on Multimodal Interaction*, ICMI '21, page 6–15, New York, NY, USA. Association for Computing Machinery.

Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. 2020. Supervised contrastive learning. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, Red Hook, NY, USA. Curran Associates Inc.

Jiacheng Li, Yujie Wang, and Julian McAuley. 2020. Time interval aware self-attention for sequential recommendation. In *Proceedings of the 13th International Conference on Web Search and Data Mining*, WSDM '20, page 322–330, New York, NY, USA. Association for Computing Machinery.

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, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Kosmas Pinitas, Konstantinos Makantasis, Antonios Liapis, and Georgios N. Yannakakis. 2022. Supervised contrastive learning for affect modelling. In *Proceedings of the 2022 International Conference on Multimodal Interaction*, ICMI '22, page 531–539, New York, NY, USA. Association for Computing Machinery.

Soujanya Poria, Devamanyu Hazarika, Navonil Majumder, Gautam Naik, Erik Cambria, and Rada Mihalcea. 2019. MELD: A multimodal multi-party dataset for emotion recognition in conversations. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 527–536, Florence, Italy. Association for Computational Linguistics.

Ramit Sawhney, Harshit Joshi, Saumya Gandhi, and Rajiv Ratn Shah. 2020. A time-aware transformer based model for suicide ideation detection on social media. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7685–7697, Online. Association for Computational Linguistics.

Nikita Soni, H. Andrew Schwartz, Ryan L. Boyd, Phi Long Bui, Syeda Mahwish, August Håkan Nilsson, Adithya V Ganesan, Lyle Ungar, Niranjan Balasubramanian, and Saif M. Mohammad. 2026. SemEval-2026 task 2: Predicting variation in emotional valence and arousal over time from ecological essays. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*. Association for Computational Linguistics.

Duyu Tang, Furu Wei, Nan Yang, Ming Zhou, Ting Liu, and Bing Qin. 2014. Learning sentiment-specific word embedding for Twitter sentiment classification. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1555–1565, Baltimore, Maryland. Association for Computational Linguistics.

Hao Tian, Can Gao, Xinyan Xiao, Hao Liu, Bolei He, Hua Wu, Haifeng Wang, and Feng Wu. 2020. SKEP: Sentiment knowledge enhanced pre-training for sentiment analysis. In *Proceedings of the 58th Annual*

*Meeting of the Association for Computational Linguistics*, pages 4067–4076, Online. Association for Computational Linguistics.

Noé Tits, Fengna Wang, Kevin El Haddad, Vincent Pagel, and Thierry Dutoit. 2019. Visualization and Interpretation of Latent Spaces for Controlling Expressive Speech Synthesis Through Audio Analysis. In *Interspeech 2019*, pages 4475–4479.

Amy Warriner, Victor Kuperman, and Marc Brysbaert. 2013. Norms of valence, arousal, and dominance for 13,915 english lemmas. *Behavior research methods*, 45.

Peixiang Zhong, Di Wang, and Chunyan Miao. 2019. Knowledge-enriched transformer for emotion detection in textual conversations. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 165–176, Hong Kong, China. Association for Computational Linguistics.