

Article

Not peer-reviewed version

---

# Enhancing Emotion Detection with Sentiment Analysis Insights

---

Ethan Parker , [Wyne Nasir](#) , Olivia Benson \*

Posted Date: 1 April 2024

doi: 10.20944/preprints202404.0059.v1

Keywords: sentiment analysis; transfer learning; natural language processing



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

# Enhancing Emotion Detection with Sentiment Analysis Insights

Ethan Parker, Wyne Nasir and Olivia Benson \*

Tufts University

\* Correspondence: olivia.benson@tufts.edu

**Abstract:** The exploration and understanding of sentiments and emotions through textual analysis are foundational to advancements in natural language processing (NLP). While sentiment analysis, which involves discerning the polarity of text, has been extensively explored, the nuanced detection of complex emotions within textual content is still an emerging area of study. In this paper, we introduce EmoLeverage, a model leveraging Transformer architecture and enhanced with a novel Fusion Adapter module. This model aims to deepen emotion detection capabilities by drawing insights from fundamental sentiment analysis tasks, focusing exclusively on textual data. EmoLeverage sets a new standard in the field by demonstrating that deep, nuanced understanding of emotions is achievable through text alone, challenging and surpassing existing methodologies in emotion identification on key datasets. This is a significant step forward, considering the complexity of human emotions and the subtleties involved in their expression through language. By concentrating on textual input, EmoLeverage offers a promising avenue for applications where audio or visual data is not available, highlighting the versatility and potential of text-based emotional analysis. Our approach illustrates the untapped potential of leveraging sophisticated NLP techniques to enhance the detection and comprehension of a wide array of emotions from text. The success of EmoLeverage in outperforming conventional models in emotion recognition tasks underscores the model's effectiveness in parsing and understanding the intricate expressions of human emotions through linguistic cues alone. This achievement paves the way for more accurate, efficient, and nuanced emotion detection tools, potentially transforming various sectors by providing deeper insights into human emotional states through text.

**Keywords:** sentiment analysis; transfer learning; natural language processing

## 1. Introduction

The exploration of sentiment analysis has long captivated scholars within the realm of natural language understanding. This field seeks to discern the sentiment polarity embedded within signals, spanning across audio, visual, and textual modalities. A closely allied field, emotion recognition, ventures beyond to classify sentiments into more detailed emotions like anger, happiness, or sorrow. Various strategies have been employed to tackle the challenges of text-based sentiment analysis and emotion detection [1–3]. Rule-based methods, reliant on grammatical and logical constructs alongside lexicons for emotion or polarity assignment, have been superseded by machine learning techniques [4]. These advanced techniques learn word-emotion relationships, with recent efforts pivoting towards the fine-tuning of Transformer models, which excel due to their sophisticated multi-head attention mechanisms. "EmoLeverage" represents a significant leap forward, incorporating transfer learning, multi-task learning, and class imbalance mitigation strategies to refine textual emotion recognition methods.

Transfer learning leverages the knowledge acquired from one task to enhance performance on another [5–8]. This strategy has become instrumental in achieving cutting-edge results across various NLP tasks, with Transformer-based [14,15] pre-trained models like BERT [18], RoBERTa [20], XLNet [21] leading the charge.

The multi-task learning technique trains a singular model across multiple tasks, offering a more efficient alternative to task-specific model fine-tuning [24]. Notable methodologies include the Multi-Task Deep Neural Network (MT-DNN), which employs a shared transformer encoder with task-specific

outputs, and innovations in training procedures that utilize knowledge distillation to boost MT-DNN performance. These methods cultivate a common representation across tasks, facilitating a more integrated learning approach [27].

Addressing class imbalance—a scenario where certain classes are underrepresented in the training data—is crucial for the success of many AI tasks. Strategies to counteract this imbalance include re-sampling of minority class data and adjusting loss functions to account for class frequency. In the computer vision domain [31,32], the focal loss modification of cross-entropy loss has been proposed to specifically tackle this issue.

Our research is primarily concerned with textual data. Understanding the sentiment or emotion underpinning sentences or paragraphs unlocks potential for various applications, including the creation of empathetic conversational agents and mental health assessment tools.

Although the task of sentiment analysis, which involves categorizing text into basic sentiments such as positive, negative, or neutral, has been extensively studied, the challenge intensifies when it comes to identifying specific emotions. This challenge is compounded by the prevalence of data imbalance across emotional categories within many datasets. "EmoLeverage" emerges as a solution, incorporating knowledge from simpler tasks to overcome the challenges of data imbalance. It employs a Transformer-based architecture integrated with a Fusion Adapter module, specifically designed to leverage sentiment analysis insights.

Our approach not only contends with state-of-the-art multimodal models on the CMU-MOSEI dataset [36] but does so by exclusively utilizing textual information. The primary contributions of this study are as follows. The introduction of EmoLeverage, a novel approach that utilizes sentiment analysis insights to enhance emotion recognition tasks. The fusion of a Transformer-based architecture with Adapter layers, optimizing the model to benefit from pre-trained language models and auxiliary tasks, thereby streamlining the training process and addressing class imbalance effectively.

## 2. Methodology

To propel the field of emotion detection forward and overcome the limitations of previous approaches, we have conceptualized and crafted a pioneering framework named "EmoLeverage." This innovative model intricately integrates the concepts of transfer learning and multi-task learning with a specialized strategy aimed at addressing the prevalent challenge of class imbalance. Central to our methodology is the employment of a sophisticated language model, specifically BERT [18], selected for its unparalleled capacity to grasp the complexities of language, an essential attribute for the accurate detection of emotions. The symbiotic relationship between sentiment analysis and emotion detection forms the cornerstone of our approach, guiding us to forge a model that harmoniously blends insights from these interrelated domains. By doing so, EmoLeverage not only enhances the precision of emotion detection but also utilizes the wealth of data available from sentiment analysis tasks.

Furthermore, a critical aspect of our design is the deliberate focus on mitigating the issue of class imbalance, a common hurdle in emotion detection datasets, ensuring that our model remains robust across diverse emotional expressions.

Delving into the architecture and methodology of EmoLeverage, we introduce a sophisticated amalgamation of the BERT Transformer encoder [18,42] with a tailored assembly of Adapter layers, each fine-tuned for specific tasks. This intricate architecture builds upon the foundational structure of the BERT encoder, configured in its base form, which coordinates a series of twelve encoder layers augmented with token, sentence, and positional embeddings. This elaborate encoding process culminates in a dedicated classification head, consisting of two sequential feedforward layers, tasked with interpreting the final hidden state associated with the [CLS] token.

The strategic inclusion of Adapter layers within each encoder layer is a testament to our commitment to task-specific fine-tuning. These layers, designed in alignment with the architecture proposed by Pfeiffer *et al.* [27], Wang *et al.* [44], execute a dimensionality reduction through a feedforward layer, followed by a non-linear transformation, and a subsequent expansion back to the encoder's original

hidden dimension. The choice of a dimensionality reduction factor of 16, as advocated by Pfeiffer *et al.* [27], strikes an optimal balance between the addition of task-specific parameters and the maintenance of exceptional model performance.

In addressing multiple tasks, EmoLeverage deploys an array of Adapter layers operating in concert, each imbued with insights specific to its designated task. The integration of these varied knowledge streams is adeptly managed through the AdapterFusion technique [27], initiating a composite learning phase where the pre-trained encoder and individual Adapter layers are fixed, thereby concentrating training efforts on the classification and fusion layers.

To confront and rectify the class imbalance inherent in emotion detection datasets, we have devised a customized version of the Binary Cross-Entropy (BCE) Loss, aptly suited for the nuances of multi-label classification. This refined loss function is expressed as:

$$L = \sum_{n=1}^N \sum_{c=1}^C -w_c [y_{n,c} \log(\sigma(x_{n,c})) + (1 - y_{n,c}) \log(1 - \sigma(x_{n,c}))] \quad (1)$$

In this formula,  $N$  symbolizes the batch size,  $C$  represents the total number of classes,  $x_{n,c}$  indicates the output of the model for class  $c$  of sample  $n$ , and  $w_c$  signifies the class-specific positive weighting factor, determined as:

$$w_c = \frac{\text{number of negative samples for class } c}{\text{number of positive samples for class } c}$$

This weighting mechanism, derived from the empirical data of the training set, is ingeniously calibrated to enhance recall in instances dominated by negative samples for class  $c$ , and bolster precision in scenarios where positive samples prevail.

Although explorations into adapting the focal loss for multi-label classification were conducted, these investigations concluded that it did not offer a significant advantage over the traditional BCE loss in terms of augmenting our model's efficacy. Through this detailed exposition, EmoLeverage stands unveiled as a beacon of innovation in the NLP landscape, charting a new course for emotion detection with its ground-breaking architecture and meticulously crafted methodologies. Our ambition with EmoLeverage is not only to surpass existing benchmarks but to redefine them, establishing a new standard of excellence in the detection and analysis of emotional expressions in text.

### 3. Experiments

#### 3.1. Setups

To rigorously evaluate the effectiveness of the "EmoLeverage" model, we conducted extensive experiments using three datasets and executed several ablation studies to dissect the contribution of individual components within our method. We engaged with the following datasets in our experimentation:

**CMU-MOSEI [36]:** Renowned for its multimodal composition, the CMU-MOSEI dataset encompasses visual, acoustic, and textual data from approximately 23,500 sentences extracted from diverse video sources. While it's structured for multimodal model training, our investigation focuses solely on textual data. The dataset annotations include sentiment on a continuum of [-3,3] and Ekman emotions (joy, sadness, anger, surprise, disgust, and fear) on a [0, 3] scale [50]. For sentiment analysis, sentiments are bifurcated into negative (labels below 0) and non-negative (labels 0 or above). Emotions are categorized as non-existent (label is 0) or present (label above 0), with the possibility of concurrent emotions for a single sentence. Model performance on CMU-MOSEI is evaluated using binary accuracy (A) and F1 scores (F1) for each distinct emotion, alongside an aggregate non-weighted mean accuracy and an overall weighted F1 score to offer a comprehensive performance measure.

**SST-2 [51] & IMDB [52]:** The SST-2 dataset offers over 60,000 sentences from movie reviews, while the IMDB dataset comprises 50,000 reviews, both annotated for sentiment analysis with a binary (positive or negative) classification. These datasets were sourced through the HuggingFace Datasets

library<sup>1</sup>, facilitating direct integration and uniform evaluation metrics, specifically binary accuracy (A), mirroring the CMU-MOSEI dataset's assessment criteria.

Our experimental framework employs BERT<sub>base</sub> (cased) [18] as the foundational pretrained model, characterized by its 12 encoder layers and a hidden dimension of 768. Incorporating Adapter and AdapterFusion layers into each encoder layer, we further refine the model's capacity to handle the intricacies of emotion and sentiment analysis. The architecture of the classification head is devised with dual linear layers, the first mirroring the transformer's hidden size (768) and the second tailored to the label count (6), interlinked with *tanh* activation functions. The input to the classification mechanism is derived from the BERT model's final hidden state corresponding to the initial [CLS] token.

All models undergo training with the AdamW optimizer [53], adhering to a linear rate scheduler, a learning rate set at 1e-5, and a weight decay parameter of 1e-2. The training process spans 10 epochs, incorporating an early stopping mechanism activated after a 3-epoch period without validation metric improvements. Leveraging the Adapter-Transformers library [54], we seamlessly integrate Adapter and AdapterFusion layers, enhancing the model's functionality. The results, averaged over three distinct runs, demonstrate the model's robustness and consistency across evaluations.

In our analysis, we trained two variant models to explore the impact of knowledge fusion: the first, Fusion<sub>3</sub>, combines insights from within the CMU-MOSEI dataset tasks (binary sentiment analysis, 7-class sentiment, and emotion classification), while the second, Fusion<sub>5</sub>, extends this integration by incorporating additional insights from the SST-2 and IMDB sentiment analysis tasks. This dual-fusion approach aims to unravel the synergistic potential of cross-task knowledge, aspiring to elevate the "EmoLeverage" model's predictive accuracy and adaptability in discerning nuanced emotional expressions and sentiment orientations across diverse textual contexts.

### 3.2. Results and Discussions

In our comprehensive assessment of "EmoLeverage" for emotion detection on the CMU-MOSEI dataset, we contrasted its performance against both a fine-tuned BERT model and a model enhanced with task-specific adapters employing an identical classification head structure. Furthermore, we juxtaposed these results with the latest state-of-the-art model for this dataset [55], which leverages a Transformer architecture utilizing textual and auditory modalities for emotion detection.

**Table 1.** Comparative performance on CMU-MOSEI for emotion detection tasks. Adapter: Utilizes BERT with task-specific adapters. Fusion<sub>3</sub>: Enhances BERT with a triple adapter fusion for CMU-MOSEI tasks, encompassing binary sentiment, seven-class sentiment, and emotion classification. Fusion<sub>5</sub>: Extends BERT with a quintuple adapter fusion, incorporating both CMU-MOSEI tasks and sentiment analysis from SST-2 & IMDB datasets.

Model	Joy A/F1	Sadness A/F1	Anger A/F1	Surprise A/F1	Disgust A/F1	Fear A/F1	Overall A/F1
State-of-Art <sup>1</sup>	66.0/71.7	73.9/17.8	81.9/17.3	89.2/3.5	86.5/45.3	90.6/0.0	81.5/40.5
BERT	66.3/69.0	69.4/42.8	74.2/44.3	85.8/21.9	83.1/53.1	83.8/18.7	77.1/51.8
Adapter	67.3/69.4	66.3/46.1	70.4/48.5	73.4/26.5	77.3/52.3	70.9/22.7	70.9/53.7
Fusion <sub>3</sub>	67.5/70.5	66.5/44.4	72.5/47.3	81.4/25.9	79.0/52.9	81.1/21.1	74.7/53.6
Fusion <sub>5</sub>	67.5/70.7	69.1/44.6	73.1/47.5	81.3/26.6	79.9/53.0	82.2/20.3	75.5/53.7

<sup>1</sup> Accuracy and F1 scores are adapted from [55], with F1 scores recalculated for clarity, as original scores were presented in a weighted format.

Notably, models employing our specialized loss function surpassed the F1-scores of the prevailing state-of-the-art, particularly highlighting the Fusion models' adeptness in balancing accuracy and

<sup>1</sup> <https://huggingface.co/datasets>

F1-scores across the spectrum of emotions. Given the dataset's skewed emotion distribution—where all except *joy* are markedly imbalanced—accuracy alone falls short as a holistic performance metric. The F1-score, hence, emerges as a more reliable gauge of model capability in distinguishing between emotions. This discrepancy underscores the Fusion models' superior performance, attributing to their integration of multi-task knowledge, which evidently enhances model efficacy.

Expanding upon the experimental analysis, we observed that while the Adapter model showcases commendable F1 scores, its accuracy metrics do not align with the Fusion models. This discrepancy underscores the added value of amalgamating insights across various tasks, as embodied by the Fusion models. Notably, the Fusion<sub>5</sub> model, with its encompassing approach that integrates additional sentiment analysis tasks beyond the CMU

-MOSEI dataset, slightly edges out the Fusion<sub>3</sub> model in performance. This comparative analysis further highlights the Fusion<sub>5</sub> model's nuanced understanding and processing of emotional content, attributed to its broader knowledge base.

**Table 2.** Distribution of positive samples across emotion classes within CMU-MOSEI.

	Joy	Sadness	Anger	Surprise	Disgust	Fear
Positive Samples (%)	52	25	21	10	17	8

**Table 3.** Comparative analysis of model parameters, highlighting the efficiency and scalability of "EmoLeverage".

Model	Total Parameters	Trainable Parameters
BERT (fine-tuned)	108.3M	108.3M
Adapter	109.8M	1.5M
Fusion <sub>3</sub>	132.8M	21.8M
Fusion <sub>5</sub>	134.6M	21.8M

This detailed exposition demonstrates "EmoLeverage"'s prowess in leveraging complex emotional datasets with significantly fewer trainable parameters, showcasing its efficiency and scalability. By harnessing insights from a wide array of tasks and datasets, "EmoLeverage" not only achieves but in some aspects surpasses the benchmark set by existing models, promising a new horizon in emotion detection research.

In this detailed examination of our novel approach, "EmoLeverage," we delve into its performance metrics across various emotional categories within the CMU-MOSEI dataset. We compare its efficacy against traditional BERT models fine-tuned for emotion detection and models that incorporate task-specific adapters. Moreover, we juxtapose these findings with the latest advancements in the field, represented by a state-of-the-art Transformer-based model that harnesses both textual and auditory data for emotion detection.

A critical aspect of model performance, particularly in addressing emotions that are underrepresented in the dataset, is the choice of loss function. We explored the efficacy of several loss functions, including the conventional Binary Cross-Entropy (BCE), Focal Loss (FL), and a novel loss function proposed in this study (PL). The comparative results are showcased in Table 4.

**Table 4.** Comparative analysis on emotion detection in CMU-MOSEI, evaluating different loss functions. BCE: Binary Cross-Entropy loss, FL: Focal Loss, PL: The newly proposed loss function tailored for this study.

Loss Function	Emotional Categories						
	Joy A/F1	Sadness A/F1	Anger A/F1	Surprise A/F1	Disgust A/F1	Fear A/F1	Overall A/F1
BCE	<b>67.9/71.5</b>	75.8/22.2	78.5/25.4	90.5/1.3	85.6/48.5	91.7/0.5	81.7/42.8
FL	67.7/70.9	75.8/24.9	78.4/23.1	90.5/0.6	85.6/46.1	91.7/0.0	81.6/42.2
PL	67.5/70.7	69.1/ <b>44.6</b>	73.1/ <b>47.5</b>	81.3/ <b>26.6</b>	79.9/ <b>53.0</b>	82.2/ <b>20.3</b>	75.5/ <b>53.7</b>

The analyses reveal that while the classic Binary Cross-Entropy loss and Focal loss exhibit similar performance metrics, the adoption of our proposed loss function (PL) notably enhances the F1-scores across the board, except for *joy*, albeit at a slight compromise in accuracy for some emotions. This indicates the proposed loss function's adeptness in balancing the detection of prevalent and scarce emotional expressions within the dataset.

### 3.3. Adapter Performance on Sentiment Analysis Tasks

Furthermore, we explored the utility of single Adapters in capturing task-specific knowledge which, when combined through fusion in "EmoLeverage," enriches the model's understanding and recognition of emotional nuances. The performance of these single Adapters on additional sentiment analysis tasks, specifically within the CMU-MOSEI dataset as well as on the SST-2 and IMDB datasets, was rigorously analyzed.

**Table 5.** Performance metrics on sentiment analysis within the CMU-MOSEI dataset, showcasing the efficacy of BERT when fine-tuned and when enhanced with task-specific adapters.

Model	Binary Sentiment Analysis Accuracy	Seven-Class Sentiment Analysis Accuracy
State-of-Art <sup>1</sup>	84.2	45.5
BERT	<b>84.3</b>	<b>46.8</b>
Adapter	83.9	46.5

<sup>1</sup> State-of-the-art accuracy scores as reported in [55].

These task-specific Adapters, when compared to a fully fine-tuned BERT model, demonstrated comparable, and in some instances superior, performance. This suggests that Adapters efficiently encapsulate pertinent task-specific insights with a significantly reduced computational footprint. The fusion of these insights, facilitated by "EmoLeverage," leverages related sentiment analysis tasks to enhance emotion detection capabilities.

**Table 6.** Sentiment analysis performance on the SST-2 and IMDB tasks comparing fine-tuned RoBERTa, XLNet, BERT models against BERT enhanced with task-specific Adapters.

Model	SST-2 Accuracy	IMDB Accuracy
RoBERTa	<b>94.8</b>	94.5
XLNet	93.4	<b>95.1</b>
BERT	93.5	94.0
Adapter	92.6	93.7

In summary, the Adapter models exhibit a remarkable capacity to mirror the performance of fully fine-tuned models across sentiment analysis tasks, further substantiating the premise that Adapter-based approaches, particularly when integrated through "EmoLeverage," present a formidable strategy

for emotion detection and sentiment analysis, promising enhanced model efficiency and interpretability.

#### 4. Conclusions and Future Work

EmoLeverage stands out as a significant advancement in the field of emotion detection, setting new records in F1-scores and showcasing the potential of leveraging purely textual data to understand complex emotional states. The exceptional performance of "EmoLeverage" on the CMU-MOSEI dataset underscores the effectiveness of our model in navigating the nuances of emotion recognition, particularly in scenarios where only textual information is available.

One of the pivotal achievements of this work is demonstrating the capability of text-based emotion detection systems to achieve, and in some aspects surpass, the accuracy of multi-modal systems, solely through the strategic analysis of text. This marks a substantial leap forward in the pursuit of more accessible, efficient, and focused emotion recognition technologies that can operate independently of visual or auditory data inputs.

However, the journey doesn't end here. The scarcity of large-scale, high-quality datasets specifically tailored for emotion detection in textual content remains a significant barrier to the advancement of this research domain. It is imperative that future studies not only continue to refine models like "EmoLeverage" for even greater accuracy across a broader spectrum of emotions but also contribute to the development and dissemination of more comprehensive and diverse datasets. These datasets will be crucial for testing and validating the efficacy of purely text-based emotion detection models as they evolve.

As we look forward to the evolution of emotion detection technologies, the contributions of "EmoLeverage" provide a robust foundation for future research. The model's adaptability and the insights it offers into the complex landscape of human emotions through text set the stage for further innovations in this fascinating intersection of natural language processing and emotional intelligence. Our next steps involve not only the continued enhancement of "EmoLeverage" but also a concerted effort to collaborate with the broader research community in curating expansive datasets that mirror the rich tapestry of human emotions. This collaborative endeavor will undoubtedly propel the field of emotion detection into new realms of possibility and precision.

#### References

1. Udochukwu, O.; He, Y. A Rule-Based Approach to Implicit Emotion Detection in Text. *NLDB*, 2015.
2. Tan, L.I.; Phang, W.S.; Chin, K.O.; Patricia, A. Rule-Based Sentiment Analysis for Financial News. 2015 IEEE International Conference on Systems, Man, and Cybernetics, 2015, pp. 1601–1606. doi:10.1109/SMC.2015.283.
3. Seal, D.; Roy, U.; Basak, R., Sentence-Level Emotion Detection from Text Based on Semantic Rules; 2019; pp. 423–430. doi:10.1007/978-981-13-7166-0\_42.
4. Wu, S.; Fei, H.; Qu, L.; Ji, W.; Chua, T.S. NExT-GPT: Any-to-Any Multimodal LLM. *CoRR* **2023**, *abs/2309.05519*.
5. Abdul-Mageed, M.; Ungar, L. EmoNet: Fine-Grained Emotion Detection with Gated Recurrent Neural Networks. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers); Association for Computational Linguistics: Vancouver, Canada, 2017; pp. 718–728. doi:10.18653/v1/P17-1067.
6. Fei, H.; Ren, Y.; Ji, D. Retrofitting Structure-aware Transformer Language Model for End Tasks. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, 2020, pp. 2151–2161.
7. Tang, D.; Qin, B.; Feng, X.; Liu, T. Target-Dependent Sentiment Classification with Long Short Term Memory. *CoRR* **2015**, *abs/1512.01100*, [1512.01100].
8. Ma, L.; Zhang, L.; Ye, W.; Hu, W. PKUSE at SemEval-2019 Task 3: Emotion Detection with Emotion-Oriented Neural Attention Network. Proceedings of the 13th International Workshop on Semantic Evaluation; Association for Computational Linguistics: Minneapolis, Minnesota, USA, 2019; pp. 287–291. doi:10.18653/v1/S19-2049.

9. Wu, S.; Fei, H.; Li, F.; Zhang, M.; Liu, Y.; Teng, C.; Ji, D. Mastering the Explicit Opinion-Role Interaction: Syntax-Aided Neural Transition System for Unified Opinion Role Labeling. *Proceedings of the Thirty-Sixth AAAI Conference on Artificial Intelligence*, 2022, pp. 11513–11521.
10. Shi, W.; Li, F.; Li, J.; Fei, H.; Ji, D. Effective Token Graph Modeling using a Novel Labeling Strategy for Structured Sentiment Analysis. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2022, pp. 4232–4241.
11. Fei, H.; Zhang, Y.; Ren, Y.; Ji, D. Latent Emotion Memory for Multi-Label Emotion Classification. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, pp. 7692–7699.
12. Wang, F.; Li, F.; Fei, H.; Li, J.; Wu, S.; Su, F.; Shi, W.; Ji, D.; Cai, B. Entity-centered Cross-document Relation Extraction. *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, 2022, pp. 9871–9881.
13. Zhuang, L.; Fei, H.; Hu, P. Knowledge-enhanced event relation extraction via event ontology prompt. *Inf. Fusion* **2023**, *100*, 101919.
14. Park, S.; Kim, J.; Jeon, J.; Park, H.; Oh, A. Toward Dimensional Emotion Detection from Categorical Emotion Annotations. *CoRR* **2019**, *abs/1911.02499*, [1911.02499].
15. Acheampong, F.A.; Nunoo-Mensah, H.; Chen, W. Transformer models for text-based emotion detection: a review of BERT-based approaches. *Artificial Intelligence Review* **2021**. doi:10.1007/s10462-021-09958-2.
16. Li, J.; Xu, K.; Li, F.; Fei, H.; Ren, Y.; Ji, D. MRN: A Locally and Globally Mention-Based Reasoning Network for Document-Level Relation Extraction. *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, 2021, pp. 1359–1370.
17. Fei, H.; Wu, S.; Ren, Y.; Zhang, M. Matching Structure for Dual Learning. *Proceedings of the International Conference on Machine Learning, ICML, 2022*, pp. 6373–6391.
18. Devlin, J.; Chang, M.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *CoRR* **2018**, *abs/1810.04805*, [1810.04805].
19. Cao, H.; Li, J.; Su, F.; Li, F.; Fei, H.; Wu, S.; Li, B.; Zhao, L.; Ji, D. OneEE: A One-Stage Framework for Fast Overlapping and Nested Event Extraction. *Proceedings of the 29th International Conference on Computational Linguistics*, 2022, pp. 1953–1964.
20. Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *CoRR* **2019**, *abs/1907.11692*, [1907.11692].
21. Yang, Z.; Dai, Z.; Yang, Y.; Carbonell, J.G.; Salakhutdinov, R.; Le, Q.V. XLNet: Generalized Autoregressive Pretraining for Language Understanding. *CoRR* **2019**, *abs/1906.08237*, [1906.08237].
22. Fei, H.; Li, F.; Li, B.; Ji, D. Encoder-Decoder Based Unified Semantic Role Labeling with Label-Aware Syntax. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, pp. 12794–12802.
23. Li, B.; Fei, H.; Li, F.; Wu, Y.; Zhang, J.; Wu, S.; Li, J.; Liu, Y.; Liao, L.; Chua, T.S.; Ji, D. DiaASQ: A Benchmark of Conversational Aspect-based Sentiment Quadruple Analysis. *Findings of the Association for Computational Linguistics: ACL 2023*, 2023, pp. 13449–13467.
24. Liu, X.; He, P.; Chen, W.; Gao, J. Multi-Task Deep Neural Networks for Natural Language Understanding. *CoRR* **2019**, *abs/1901.11504*, [1901.11504].
25. Fei, H.; Ren, Y.; Ji, D. Boundaries and edges rethinking: An end-to-end neural model for overlapping entity relation extraction. *Information Processing & Management* **2020**, *57*, 102311.
26. Li, J.; Fei, H.; Liu, J.; Wu, S.; Zhang, M.; Teng, C.; Ji, D.; Li, F. Unified Named Entity Recognition as Word-Word Relation Classification. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022, pp. 10965–10973.
27. Pfeiffer, J.; Kamath, A.; Rücklé, A.; Cho, K.; Gurevych, I. AdapterFusion: Non-Destructive Task Composition for Transfer Learning. *CoRR* **2020**, *abs/2005.00247*, [2005.00247].
28. Fei, H.; Wu, S.; Li, J.; Li, B.; Li, F.; Qin, L.; Zhang, M.; Zhang, M.; Chua, T.S. LasUIE: Unifying Information Extraction with Latent Adaptive Structure-aware Generative Language Model. *Proceedings of the Advances in Neural Information Processing Systems, NeurIPS 2022*, 2022, pp. 15460–15475.
29. Fei, H.; Ren, Y.; Zhang, Y.; Ji, D.; Liang, X. Enriching contextualized language model from knowledge graph for biomedical information extraction. *Briefings in Bioinformatics* **2021**, *22*.
30. Wu, S.; Fei, H.; Ji, W.; Chua, T.S. Cross2StrA: Unpaired Cross-lingual Image Captioning with Cross-lingual Cross-modal Structure-pivoted Alignment. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2023, pp. 2593–2608.

31. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection, 2018, [[arXiv:cs.CV/1708.02002](https://arxiv.org/abs/cs.CV/1708.02002)].
32. Fei, H.; Zhang, M.; Ji, D. Cross-Lingual Semantic Role Labeling with High-Quality Translated Training Corpus. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 7014–7026.
33. Wu, S.; Fei, H.; Ren, Y.; Ji, D.; Li, J. Learn from Syntax: Improving Pair-wise Aspect and Opinion Terms Extraction with Rich Syntactic Knowledge. Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, 2021, pp. 3957–3963.
34. Li, B.; Fei, H.; Liao, L.; Zhao, Y.; Teng, C.; Chua, T.; Ji, D.; Li, F. Revisiting Disentanglement and Fusion on Modality and Context in Conversational Multimodal Emotion Recognition. Proceedings of the 31st ACM International Conference on Multimedia, MM, 2023, pp. 5923–5934.
35. Fei, H.; Liu, Q.; Zhang, M.; Zhang, M.; Chua, T.S. Scene Graph as Pivoting: Inference-time Image-free Unsupervised Multimodal Machine Translation with Visual Scene Hallucination. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2023, pp. 5980–5994.
36. Bagher Zadeh, A.; Liang, P.P.; Poria, S.; Cambria, E.; Morency, L.P. Multimodal Language Analysis in the Wild: CMU-MOSEI Dataset and Interpretable Dynamic Fusion Graph. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers); Association for Computational Linguistics: Melbourne, Australia, 2018; pp. 2236–2246. doi:10.18653/v1/P18-1208.
37. Wu, S.; Fei, H.; Cao, Y.; Bing, L.; Chua, T.S. Information Screening whilst Exploiting! Multimodal Relation Extraction with Feature Denoising and Multimodal Topic Modeling. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2023, pp. 14734–14751.
38. Fei, H.; Wu, S.; Ren, Y.; Li, F.; Ji, D. Better Combine Them Together! Integrating Syntactic Constituency and Dependency Representations for Semantic Role Labeling. Findings of the Association for Computational Linguistics: ACL/IJCNLP 2021, 2021, pp. 549–559.
39. Wu, S.; Fei, H.; Zhang, H.; Chua, T.S. Imagine That! Abstract-to-Intricate Text-to-Image Synthesis with Scene Graph Hallucination Diffusion. *Advances in Neural Information Processing Systems* **2024**, *36*.
40. Fei, H.; Wu, S.; Ji, W.; Zhang, H.; Chua, T.S. Empowering dynamics-aware text-to-video diffusion with large language models. *arXiv preprint arXiv:2308.13812* **2023**.
41. Qu, L.; Wu, S.; Fei, H.; Nie, L.; Chua, T.S. Layoutllm-t2i: Eliciting layout guidance from llm for text-to-image generation. Proceedings of the 31st ACM International Conference on Multimedia, 2023, pp. 643–654.
42. Ando, R.K.; Zhang, T. A Framework for Learning Predictive Structures from Multiple Tasks and Unlabeled Data. *Journal of Machine Learning Research* **2005**, *6*, 1817–1853.
43. Fei, H.; Li, F.; Li, C.; Wu, S.; Li, J.; Ji, D. Inheriting the Wisdom of Predecessors: A Multiplex Cascade Framework for Unified Aspect-based Sentiment Analysis. Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI, 2022, pp. 4096–4103.
44. Wang, G.; Ying, R.; Huang, J.; Leskovec, J. Improving Graph Attention Networks with Large Margin-based Constraints. NeurIPS-Workshop, 2019.
45. Fei, H.; Chua, T.; Li, C.; Ji, D.; Zhang, M.; Ren, Y. On the Robustness of Aspect-based Sentiment Analysis: Rethinking Model, Data, and Training. *ACM Transactions on Information Systems* **2023**, *41*, 50:1–50:32.
46. Zhao, Y.; Fei, H.; Cao, Y.; Li, B.; Zhang, M.; Wei, J.; Zhang, M.; Chua, T. Constructing Holistic Spatio-Temporal Scene Graph for Video Semantic Role Labeling. Proceedings of the 31st ACM International Conference on Multimedia, MM, 2023, pp. 5281–5291.
47. Fei, H.; Ren, Y.; Zhang, Y.; Ji, D. Nonautoregressive Encoder-Decoder Neural Framework for End-to-End Aspect-Based Sentiment Triplet Extraction. *IEEE Transactions on Neural Networks and Learning Systems* **2023**, *34*, 5544–5556.
48. Zhao, Y.; Fei, H.; Ji, W.; Wei, J.; Zhang, M.; Zhang, M.; Chua, T.S. Generating Visual Spatial Description via Holistic 3D Scene Understanding. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2023, pp. 7960–7977.
49. Fei, H.; Li, B.; Liu, Q.; Bing, L.; Li, F.; Chua, T.S. Reasoning Implicit Sentiment with Chain-of-Thought Prompting. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 2023, pp. 1171–1182.
50. Ekman, P. An argument for basic emotions. *Cognition and Emotion* **1992**, *6*, 169–200, [<https://doi.org/10.1080/02699939208411068>] doi:10.1080/02699939208411068.

51. Socher, R.; Perelygin, A.; Wu, J.; Chuang, J.; Manning, C.D.; Ng, A.; Potts, C. Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank. *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing; Association for Computational Linguistics: Seattle, Washington, USA, 2013*; pp. 1631–1642.
52. Maas, A.L.; Daly, R.E.; Pham, P.T.; Huang, D.; Ng, A.Y.; Potts, C. Learning Word Vectors for Sentiment Analysis. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies; Association for Computational Linguistics: Portland, Oregon, USA, 2011*; pp. 142–150.
53. Loshchilov, I.; Hutter, F. Fixing Weight Decay Regularization in Adam. *CoRR* **2017**, *abs/1711.05101*, [[1711.05101](#)].
54. Pfeiffer, J.; Rücklé, A.; Poth, C.; Kamath, A.; Vulić, I.; Ruder, S.; Cho, K.; Gurevych, I. AdapterHub: A Framework for Adapting Transformers. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, 2020*, pp. 46–54.
55. Delbrouck, J.; Tits, N.; Brousmiche, M.; Dupont, S. A Transformer-based joint-encoding for Emotion Recognition and Sentiment Analysis. *CoRR* **2020**, *abs/2006.15955*, [[2006.15955](#)].

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.