

What is the impact of valence-arousal-dominance (VAD) representation format versus categorical emotion annotations

Assignee Research

June 10, 2026

Abstract

Current large speech language models (Speech-LLMs) often exhibit limitations in empathetic reasoning, primarily due to the absence of training datasets that integrate both contextual content and paralinguistic cues. In this work, we propose two approaches to incorporate contextual paralinguistic information into model training: (1) an explicit method that provides paralinguistic metadata (e.g., emotion annotations) directly to the LLM, and (2) an implicit method that automatically generates novel training question-answer (QA) pairs using both categorical and dimensional emotion annotations also

1 Introduction

This paper examines: Incorporating Contextual Paralinguistic Understanding in Large Speech-Language Models. Research question: What is the impact of valence-arousal-dominance (VAD) representation format versus categorical emotion annotations on the performance of OpenPangu-7B-MLA in empathetic speech understanding across different genres?.

2 Methodology

Systematic literature search across multiple databases yielded 14 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.0/10.

3 Results

14 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 3.0/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

References

- <http://arxiv.org/abs/2508.07273v1>
- <http://arxiv.org/abs/2510.22758v2>
- <http://arxiv.org/abs/2601.22873v1>