Adaptive Negative Pair Weighting in Contrastive Loss for Multimodal Foundation Models

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 13 peer-reviewed papers addressing the following research question: How does adaptive negative pair weighting in contrastive loss functions impact the convergence speed and final accuracy of multimodal foundation models on out-of-distribution benchmarks. 11 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Batch-Softmax Contrastive Loss for Pairwise Sentence Scoring Tasks. Research question: How does adaptive negative pair weighting in contrastive loss functions impact the convergence speed and final accuracy of multimodal foundation models on out-of-distribution benchmarks?.

2 Methodology

Systematic literature search across multiple databases yielded 13 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.5/10.

3 Results

13 papers retrieved. 11 claims extracted; 1 independently verified. Quality review score: 4.5/10.

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.

5 Extracted Claims

Claim	Verified	Confidence
Pointwise approaches for training models for pairwise sentence scoring tasks, such as mean squared error (MSE), do not t	✓	0.17
For two pairs with correct target scores (0.4, 0.5), the loss function would equally penalize answers like (0.3, 0.6) an	×	0.03
The first pair (0.3, 0.6) is better as it keeps the correct ranking, while the second one (0.5, 0.4) does not.	×	0.04
Pairwise approaches, such as triplet loss, directly learn an ordering.	×	0.06
Constructing pairs or triplets in the training set is hard due to the difficulty in finding non-trivial negative example	×	0.04
The BSC loss treats all other possible pairs of examples in the batch as negatives, requiring only positive pairs for tr	×	0.11
The BSC loss function is defined as LBSC(X) = L0(X) + L1(X), where L0(X) and L1(X) are components involving softmax oper	×	0.04
The softmax operation in the BSC loss is applied by rows.	×	0.04
The BSC loss function includes a temperature parameter τ.	×	0.05
The component L0(X) of the BSC loss can be rewritten as L0(X) = -1/(mτ) * Σ (qT_i a_i) + 1/m * Σ log(Σ exp(qT_i a_j / τ)	×	0.01
The BSC loss function is mathematically similar to the one used in contrastive learning.	×	0.06

References

- http://arxiv.org/abs/2506.09781v2
- http://arxiv.org/abs/2110.15725v1
- http://arxiv.org/abs/2111.06934v1