Batch-Size Scaling in Batch-Softmax Contrastive Loss and Its Impact on Representation Quality

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 8 peer-reviewed papers addressing the following research question: How does the scaling of batch size in batch-softmax contrastive loss affect the convergence speed and final representation quality, measured by SPEARMAN correlation or retrieval accuracy, on sentence. 11 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.7/10. 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 the scaling of batch size in batch-softmax contrastive loss affect the convergence speed and final representation quality, measured by SPEARMAN correlation or retrieval accuracy, on sentence similarity benchmarks like STS-B or SciDocS?.

2 Methodology

Systematic literature search across multiple databases yielded 8 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.7/10.

3 Results

8 papers retrieved. 11 claims extracted; 0 independently verified. Quality review score: 3.7/10.

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 like mean squared error (MSE) do not take the relative order of scores into account.	×	0.01
For target scores (0.4, 0.5), MSE penalizes the prediction (0.3, 0.6) equally to the prediction (0.5, 0.4).	×	0.00
The prediction (0.3, 0.6) maintains the correct ranking relative to targets (0.4, 0.5), whereas (0.5, 0.4) does not.	×	0.02
Pairwise approaches such as triplet loss allow models to directly learn an ordering.	×	0.07
Constructing pairs or triplets for training is difficult due to the challenge of finding non-trivial negative examples.	×	0.05
The Batch-Softmax Contrastive (BSC) loss treats all other possible pairs of examples in a batch as negatives.	×	0.12
The BSC loss requires only positive pairs for training.	×	0.12
The BSC loss function is defined as the sum of two components, L0(X) and L1(X).	×	0.01
In the BSC loss function, the softmax operation is applied by rows.	×	0.04
The BSC loss function includes a temperature parameter denoted as τ.	×	0.03
The component L0(X) of the BSC loss can be mathematically rewritten as the negative average of query-answer dot products	×	0.03

References

- http://arxiv.org/abs/2506.09781v2

- http://arxiv.org/abs/2302.02125v1
- http://arxiv.org/abs/2110.15725v1