Manifold-Based Semantic Scoring vs. Cosine Similarity in Low-Resource Cross-Lingual QA

Assignee Research

June 3, 2026

Abstract

This report synthesises findings from 4 peer-reviewed papers addressing the following research question: What is the comparative robustness of manifold-based semantic scoring versus cosine similarity in cross-lingual open QA benchmarks when evaluated on low-resource languages. 15 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: NMIXX: Domain-Adapted Neural Embeddings for Cross-Lingual eXploration of Finance. Research question: What is the comparative robustness of manifold-based semantic scoring versus cosine similarity in cross-lingual open QA benchmarks when evaluated on low-resource languages?.

2 Methodology

Systematic literature search across multiple databases yielded 4 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.8/10.

3 Results

4 papers retrieved. 15 claims extracted; 0 independently verified. Quality review score: 3.8/10.

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
The training corpus consists of 18.8k high-confidence triplets , ,.	×	0.11
Hard negatives () are generated according to our four-axis financial semantic-shift taxonomy (temporal drift, perspecti	×	0.06
Positives () combine two complementary strategies: (1) In-domain paraphrase and (2) Exact translation.	×	0.03
We adopt a temperature-scaled triplet negative-log-likelihood loss.	×	0.00
We measure Spearman's in four STS suites: FinSTS, KorFinSTS, STS, and KorSTS.	×	0.08
We fine-tune seven license-compatible embedding models that can be run on our four-A100 setup.	×	0.08
Domain Adaptation Trade-offs: Domain-specific fine-tuning can erode performance on general-domain tasks.	×	0.08
Benchmarks such as FLUE [23] established core FinNLP tasks (sentiment, entity recognition, QA).	×	0.04
Retrieval-oriented suites like FinDER [4] and TWICE [12] demonstrated significant performance gaps between off-the-shelf	×	0.07
These studies highlight challenges such as market jargon, temporal semantic drift, and regulatory formality, which gener	×	0.06
Domain-specific jargon: e.g., 'bullish reversal' or 'EBITDA margin,' which are semantically dense and not well covered b	×	0.07
Temporal semantic drift: where terms like 'credit risk' or 'volatility' shift meaning depending on market cycles [25].	×	0.04
Regulatory formality and legalistic phrasing: Especially in SEC filings or supervisory disclosures [12].	×	0.02
FinGPT [33] offers an open source pipeline for mixed-source financial pretraining.	×	0.02
BloombergGPT [32] trains a massive 50B-parameter model on proprietary and public financial texts.	×	0.02

References

- https://arxiv.org/abs/2507.09601
- https://www.semanticscholar.org/paper/32b30a98f023a5a416cc7b7d3e0b1948d41462ce
- https://arxiv.org/abs/2506.08372