Universal Output Layers Accelerate Convergence in Multilingual ASR Systems

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 10 peer-reviewed papers addressing the following research question: What is the impact of UML's universal output layer on model convergence speed and training stability in multilingual ASR compared to conventional WPM-based models when evaluated on LibriSpeech and. 9 claims were extracted from source literature; 2 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 5.3/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: UML: A Universal Monolingual Output Layer for Multilingual ASR. Research question: What is the impact of UML's universal output layer on model convergence speed and training stability in multilingual ASR compared to conventional WPM-based models when evaluated on LibriSpeech and Common Voice benchmarks?.

2 Methodology

Systematic literature search across multiple databases yielded 10 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 5.3/10.

3 Results

10 papers retrieved. 9 claims extracted; 2 independently verified. Quality review score: 5.3/10.

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
UML is a monolingual output layer shared by all languages.	✓	0.29
UML enables the use of more WPMs with fewer parameters compared to alternative methods.	×	0.07
UML uses only one H $\times$ max(V1, . . . , VL)-dimensional output layer to model the sum of WPMs across all languages.	×	0.11
The method using a conventional output layer for all multilingual WPMs requires a H $\times$ (sum of Vl for all languages)-dime	✓	0.15
The methods in [21, 22] use L separate monolingual output layers whose dimensions are H$\times$V1, . . . , H$\times$VL.	×	0.09
UML requires H $\times$ max(V1, . . . , VL) parameters to model the sum of WPMs across all languages.	×	0.05
In UML, each WPM is determined jointly by the LID and the output node index.	×	0.13
LIDs need to be taken into account in inference in UML.	×	0.02
At test-time, let x, y, and z be the input, output, and LID prediction sequences of an utterance, y* is the decoding res	×	0.03

References

- http://arxiv.org/abs/2011.03118v1
- http://arxiv.org/abs/2105.14779v2
- http://arxiv.org/abs/2302.11186v1