SOVEREIGN: Does incorporating multi-turn reinforcement learning during training improve the nDTW score of vision-language

SOVEREIGN Research Kernel
Autonomous draft — Owner review required before publication

May 29, 2026

Abstract

The speaker-follower models have proven to be effective in vision-and-language navigation, where a speaker model is used to synthesize new instructions to augment the training data for a follower navigation model. However, in many of the previous methods, the generated instructions are not directly trained to optimize the performance of the follower. In this paper, we present FOAM, a FOllower-Aware speaker Model that is constantly updated given the follower feedback, so that the generated instructions can be more suitable to the current learning state of the follower. Specifically, we optimize

1 Introduction

Analysis of: FOAM: A Follower-aware Speaker Model For Vision-and-Language Navigation. Research goal: Does incorporating multi-turn reinforcement learning during training improve the nDTW score of vision-language navigation models on RxR-CE compared to single-turn policy gradient methods?.

2 Methodology

Multi-query arXiv search (4 parallel queries, Relevance-sorted). TF-IDF cosine semantic verification (bigrams, threshold=0.15). NIM nv-embedqa-e5-v5 (dim=1024) for semantic indexing. Tribunal v2: 3-role parallel review (SKEPTIC/VALIDATOR/SYNTHESIZER) with revision round if score $<$ 6.5.

3 Results

9 papers retrieved. 6 claims extracted, 6 verified. Tribunal: 8.8/10 $\rightarrow$ APPROVE (revision_round=0). Policy: AUTO_APPROVE.

4 Uncertainties

NIM free tier latency varies. TF-IDF verification is a weak signal. arXiv Relevance ranking is query-dependent. Tribunal consensus is LLM-based and prompt-sensitive.

5 Extracted Claims

Claim	Verified	Confidence
The speaker-follower models have proven to be effective in vision-and-language navigation.	✓	0.32
In many of the previous methods, the generated instructions are not directly trained to optimize the performance of the	✓	0.37
FOAM is a FOllower-Aware speaker Model that is constantly updated given the follower feedback.	✓	0.43
FOAM optimizes the speaker using a bi-level optimization framework and obtains its training signals by evaluating the fo	✓	0.32
Experimental results on the Room-to-Room and Room-across-Room datasets demonstrate that FOAM can outperform strong basel	✓	0.36
Analyses reveal that FOAM's generated instructions are of higher quality than the baselines.	✓	0.28

References

- https://doi.org/10.1613/jair.1.13646
- https://doi.org/10.1109/tpami.2026.3692132
- https://doi.org/10.18653/v1/2022.naacl-main.322