Retriever Ensemble Size and Factual Accuracy Degradation Under Adversarial Attacks

Assignee Research

June 8, 2026

Abstract

This report synthesises findings from 16 peer-reviewed papers addressing the following research question: What is the correlation between retriever ensemble size and factual accuracy degradation under gradient-based adversarial attacks across News and Science domains in AmbiEval. 13 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 2.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Adversarial Diffusion Attacks on Graph-based Traffic Prediction Models. Research question: What is the correlation between retriever ensemble size and factual accuracy degradation under gradient-based adversarial attacks across News and Science domains in AmbiEval?.

2 Methodology

Systematic literature search across multiple databases yielded 16 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 2.5/10.

3 Results

16 papers retrieved. 13 claims extracted; 0 independently verified. Quality review score: 2.5/10.

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
Szegedy et al. (2013) discovered that adversarial samples are low-probability but densely distributed in deep neural net	×	0.11
Goodfellow et al. showed that generating adversarial samples is sufficient when DNNs demonstrate linear behaviors in hig	×	0.04
An experiment involving 99 smartphones moving slowly on a handcart caused Google Maps to identify an empty street as a c	×	0.01
Mobile phone-based mapping services like Google Maps and AutoNavi estimate traffic states based on GPS trajectories sent	×	0.05
On the LA dataset, the ST-GCN model has a baseline metric value of 5.46.	×	0.05
On the LA dataset, the A3T-GCN model has a baseline metric value of 12.74.	×	0.05
On the HK dataset, the ST-GCN model has a baseline metric value of 23.34.	×	0.05
Under attack conditions on the LA dataset, the ST-GCN model exhibits a performance degradation of 8.32%.	×	0.08
Under attack conditions on the LA dataset, the A3T-GCN model exhibits a performance degradation of 22.77%.	×	0.07
Under attack conditions on the HK dataset, the A3T-GCN model exhibits a performance degradation of 70.21%.	×	0.07
Traffic prediction tasks include traffic state prediction, demand prediction, and trajectory prediction.	×	0.14
Traffic state prediction includes the prediction of traffic flow, speed, and travel time.	×	0.09
Traffic data is represented in non-Euclidean space due to its association with the topological structure of road network	×	0.04

References

- `http://arxiv.org/abs/2307.02055v1`

- http://arxiv.org/abs/2006.16545v1
- http://arxiv.org/abs/2104.09369v1