Multimodal Graph Contrastive Learning for Cross-Domain Recommendation Performance

Assignee Research

June 2, 2026

Abstract

This report synthesises findings from 14 peer-reviewed papers addressing the following research question: How do multimodal extensions of graph contrastive learning models (e.g., combining visual and textual features) perform on cross-domain recommendation tasks compared to pure graph-based approaches. Multimedia collections are more than ever growing in size and diversity. Effective multimedia retrieval systems are thus critical to access these datasets from the end-user perspective and in a scalable way. 10 claims were extracted from source literature; 2 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Unsupervised Visual and Textual Information Fusion in Multimedia Retrieval - A Graph-based Point of View. Research question: How do multimodal extensions of graph contrastive learning models (e.g., combining visual and textual features) perform on cross-domain recommendation tasks compared to pure graph-based approaches like LGKAT, evaluated using metrics like NDCG and MRR?.

2 Methodology

Systematic literature search across multiple databases yielded 14 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.5/10.

3 Results

14 papers retrieved. 10 claims extracted; 2 independently verified. Quality review score: 4.5/10.

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
The cross-media similarity mechanism described in the paper has proven to give top-ranked retrieval results on several I	×	0.10
The experiments aim to study the different settings one could apply using the generalization proposed in Eq. 25 and Eq.	×	0.03
The goal is to establish guidelines on the combination of visual and textual information in CBMIR using graph-based meth	✓	0.23
The experiments examine the impact of several parameters on the cross-media and random walk method, including initializa	×	0.11
The experiments investigate the late combination of the text query-based semantically filtered multimedia scores with th	✓	0.16
The experiments address the benefits of a multimedia query compared to a text-only query and the benefits of linearly co	×	0.03
The experiments study the conditions under which it is beneficial to proceed to a late fusion of similarity matrices bef	×	0.03
The cross-media similarity mechanism involves finding the most similar items in the collection with regard to the textua	×	0.07
The cross-media similarities are defined as cmtv(q, .) = K(st(q, .), k) $\cdot$ Sv.	×	0.07
The operator K(., k) takes as input a vector and gives a zero value to elements whose score is strictly lower than the k	×	0.02

References

- http://arxiv.org/abs/2206.07869v1
- http://arxiv.org/abs/1401.6891v1
- http://arxiv.org/abs/2510.22799v1