

Wasserstein GAN for Knowledge Graph Completion with Continuous Learning

Erdem Önal 

Abstract—Knowledge Graph Completion predicts missing links in incomplete knowledge bases. Generative Adversarial Networks can create informative negative samples, but many methods struggle with mode collapse and vanishing gradients when working with discrete graph data. Many models also remain fixed and do not adapt as data changes. This work presents a Wasserstein GAN model for Knowledge Graph Completion. It uses the Wasserstein distance to make training more stable on the DBLP Computer Science Bibliography dataset and produces valid RDF triples. Incremental retraining is performed once per day via a scheduled CI workflow. Experimental results on DBLP show improved novelty and diversity compared to uniform negative sampling. This allows stable generative modeling and ongoing adaptation.

Index Terms—Knowledge Graph Completion, Wasserstein GAN, Continuous Learning, RDF.

I. INTRODUCTION

Knowledge Graphs such as the DBLP Computer Science Bibliography store entities and relations as RDF triples. They are often incomplete. Knowledge Graph Completion tries to infer missing facts from what is already known.

Uniform negative sampling is inefficient, as noted by Cai et al. [1], because random entities are easy to tell apart from true facts. Adversarial learning can produce harder negative samples, but with discrete data it can be unstable and lead to vanishing gradients [2].

Knowledge changes over time. Static embedding models fail to accommodate temporal updates in evolving knowledge graphs [3].

This study presents a system that:

- 1) Uses a Wasserstein GAN to learn the distribution of valid triples and avoid problems of conventional GANs.
- 2) Runs a continuous learning pipeline with GitHub Actions to update the model using new DBLP data.
- 3) Evaluates generated triples using novelty and diversity metrics.

II. RELATED WORK

A. Adversarial Learning for Knowledge Graph Completion

KBGAN [1] improved Knowledge Graph Completion by generating plausible negative samples. This made the discriminator more effective. However, it uses policy gradients to handle discrete sampling, which can be unstable and slow.

B. Wasserstein GANs in Graphs

Dai et al. [2] applied the Wasserstein distance to GANs for discrete data. This helped reduce vanishing gradients. The method in this paper uses this idea in a system that is updated regularly.

C. Continual Learning

Most knowledge graph embedding models assume data does not change. Daruna et al. [3] showed the need for continual learning to handle new concepts. Here, a CI/CD pipeline retrains the model every day.

III. METHODOLOGY

A. Data Processing

The DBLP Computer Science Bibliography is the source dataset. The XML dump is streamed. High-quality publications are kept. A schema of *Publication*, *Author*, *Venue*, and *Year* is extracted to build RDF triples.

B. WGAN Architecture

Following [2], the model has:

- **Generator (G):** Takes random noise z and a relation r to produce a tail embedding t' . It minimizes the Wasserstein distance to real data.
- **Discriminator (D):** Scores triples (h, r, t) for plausibility. It outputs a scalar instead of a probability.

C. Continuous Learning Pipeline

Implemented with GitHub Actions:

- 1) A daily job starts the process.
- 2) The system loads the latest checkpoint and updates weights with new DBLP data.
- 3) Results, including generated RDF triples, are published to a web dashboard.

IV. EXPERIMENTS AND METRICS

A. Metrics

The evaluation uses:

- **Novelty:** The share of generated triples not in the training data.
- **Diversity:** The range of generated relations, showing if mode collapse is avoided.

B. Results

Experiments with DBLP show stable Wasserstein GAN loss and correct prediction of valid author and venue pairings not in the training data.

V. CONCLUSION

This approach combines adversarial sampling, Wasserstein distance stability, and continual learning to predict future scientific collaborations.

REFERENCES

- [1] L. Cai and W. Y. Wang, "KBGAN: Adversarial learning for knowledge graph embeddings," in *Proc. NAACL-HLT*, 2018, pp. 1470–1480.
- [2] Y. Dai, S. Wang, X. Chen, C. Xu, and W. Guo, "Generative adversarial networks based on Wasserstein distance for knowledge graph embeddings," *Knowledge-Based Systems*, vol. 190, p. 105165, 2020.
- [3] A. Daruna, M. Gupta, M. Sridharan, and S. Chernova, "Continual learning of knowledge graph embeddings," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1128–1135, 2021.