ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

FouriER: Link Prediction by Mixing Tokens with Fourier-enhanced MetaFormer

Thanh Vu[1,2], Huy Ngo[1,2], Bac Le[1,2] and Thanh Le[1,2] *

1- Faculty of Information Technology, University of Science, Ho Chi Minh City, Vietnam

2- Vietnam National University, Ho Chi Minh City, Vietnam

Abstract. Knowledge graph link prediction has been researched for many years. With the steady development of data, the demand for missing link prediction in knowledge bases is growing. In this study, we propose FouriER, a model using Fourier transforms integrated into MetaFormer architecture to learn features from embeddings better but more computationally cost-effective than the self-attention mechanism in Transformer models. Furthermore, we transform embeddings to a 2D form and stack them that benefit the model in learning interactions between entities and relations more efficiently. As a result, we found that our model outperformed baseline models on two benchmark datasets in our experiments.

1 Introduction

A knowledge graph is a form of knowledge base that consists of entities and relations that represent the real world as triples, with each triple consisting of a subject, relation, and object. They have several uses that benefit daily living, like as search engines and recommendation systems. However, the current knowledge graphs all have the same problem: missing relationships. To address this issue, adding missing connections to knowledge graphs makes them more meaningful; covering more relationships allows applications to be performed more efficiently. The task of filling missing links like that is called link prediction. Transformer models have recently been used in numerous domains, including natural language processing and computer vision, and they perform well in many tough issues. As a result, this is a promising direction for link prediction in knowledge graphs.

In this paper, we introduce FouriER, a model based on MetaFormer [1] architecture (a general architecture derived from Transformers by omitting the token mixer) that uses the Fourier Transform as a token mixer to extract essential features from knowledge graph embeddings, allowing the model to predict better.

Our main contributions are as follows:

- To the best of our knowledge, our model is the first model to apply a computer vision model combined with Fourier Transform in link prediction task to help the model learn relations more effectively.
- The results show that FouriER outperforms the baseline models, indicating that our approach is more efficient than the baselines.

*This research is supported by the research funding from the Faculty of Information Technology, University of Science, Ho Chi Minh City, Vietnam.

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

2 Related work

Many approaches have been proposed for the link prediction task in the knowledge graph. Translation-based approach, such as TransE, [2] models the relation as a translation operator from head entity to tail entity. TransR [3] models entities and relations in distinct spaces. Rule-based approach, such as RARL [4], discovers rules in knowledge and uses them to infer missing relationships. Semantic information-based approach uses semantic information about entities and relations to define scoring functions for triples. RESCAL [5] uses a tensor product as a scoring function, with entities being vectors and relations being matrixes that capture the interaction between these entities. DistMult [6] simplifies RESCAL by using diagonal matrices to represent relations. Neural network-based approach uses many layers to learn the embedding of triples better. R-GCN [7] uses the message-passing framework of the graph convolution network to aggregate the embedding of neighboring entities and then exploit them with a decoder such as DistMult. Inheriting ideas from R-GCN, RA-GCN [8] improves the propagation formula for updating entity or node information and extracts additional entity and relationship information. ConvE [9] uses convolutional and fully connected layers to capture the interaction between entities and relations. ConvKB [10] represents a triple as a 3-column matrix where each column vector represents a triple element. By using a convolution layer, feature vectors representing triples are extracted and used to calculate a score to predict whether the triple is valid or not. KBGAN [11] leverages adversarial learning to provide high-quality negative samples. DMACM [12] uses attention mechanisms to explore the directional relational characteristics and implicit fine-grained features in the triple. In this paper, we leverage MetaFormer [1] architecture and Fourier transforms to capture information from embedding more effectively.

3 Proposed method

3.1 Problem and notations

A knowledge graph $\mathcal{G}$ is formalized as $\mathcal{G} = \{(s, r, o)|(s, o) \in \mathcal{E}, r \in \mathcal{R}\}$, which is a set of triples. Where $\mathcal{E}$ represents the set of entities, $\mathcal{R}$ represents the set of relations, and s, r, o represent the subject, relation, and object, respectively. Entities and relations are represented by embedding, they are denoted as $\mathbf{e}_s$, $\mathbf{e}_r$, $\mathbf{e}_o$ for the embedding of subject, relation and object respectively. The number of entities and relations is represented by n_e and n_r, while the dimension of the embedding vector representing the entity and relationship is symbolized by d_e and d_r. The purpose of the link prediction task is to predict the missing entity or relation using the triples already present in the knowledge graph. Our model takes as input a query q where $q = (s, r)$ and makes predictions about the missing object o for each sample (1-X scoring fashion, in which, the query is scored against a candidate set of entities instead of all entities).

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

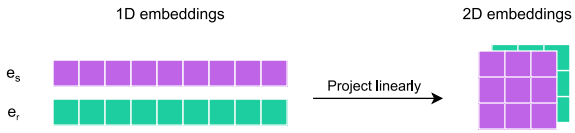

Fig. 1: Visualization of transforming 1D embeddings into image-like 2D embeddings

3.2 Model architecture

In this section, we go over the proposed model in depth. Our model leverages the strengths of the PoolFormer [1] - a MetaFormer-based architecture with a simple Pooling layer acting as a token mixer layer, which has been used in computer vision problems. We combined it with a Fast Fourier Transform (FFT) layer inspired by FNet [13] as a token mixer layer, to extract significant features from knowledge graph embeddings. The overall architecture of our proposed model is visualized in Fig. 2 while the specific information inside the FouriER block is illustrated in Fig. 3.

Our model is divided into four major stages, with the output of the previous stage being downsampled to retain the most crucial features for missing entity prediction. Each stage has a varied number of FouriER blocks, stages from 1 to 4 have a block number of 4, 4, 12 and 4 respectively.

Query embeddings: We augmented the input data by linearly transforming the embedding vectors from 1D to 2D matrix with the goal of helping our model learn better entity-to-relation relationships, both of them should have the same size after being transformed. Then we stack the two embeddings from the entity and the relation (which can be viewed as transforming into two channels in the image). Making embedding two-dimensional helps them adapt to the model's architecture as well as can help the Channel MLP learn interactions through stacking entities and relations. Fig. 1 shows how embeddings are fused and stacked together to be a relational query.

The process of transforming and stacking embedding can be formalized as:

$$q = \Phi_s[\phi(\mathbf{e}_s), \phi(\mathbf{e}_r)] \tag{1}$$

where $\phi(\mathbf{e}_s) = \mathbf{e}_s\mathbf{W}_s + b_s$ and $\phi(\mathbf{e}_r) = \mathbf{e}_r\mathbf{W}_r + b_r$ and Φ_s denotes the stacking operator. The parameters of the weight matrices and bias vectors are learnt in the training phase so that embeddings are going to be fit into the model better.

Discrete Fourier Transform (DFT): We use FFT in our implementation to reduce the computation time of calculating DFT. We use FFT to speed up the model because FFT has no parameters to be optimized instead of using multi head attention. Given a sequence $\{x_n | n \in [0, N-1]\}$, the DFT is defined by:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N} nk}, \ 0 \leq k \leq N-1 \tag{2}$$

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

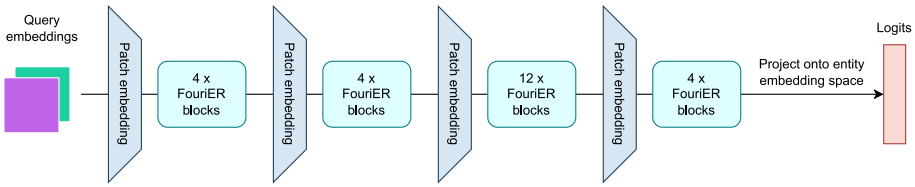

Fig. 2: Overall architecture of FouriER

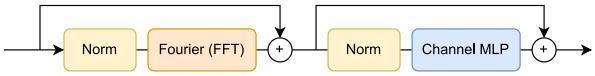

Fig. 3: A FouriER block

FouriER block: The main component consists of an FFT layer to extract features from embeddings and an MLP layer which is used to capture relationships between channels (entity-relation interactions). There are 2 skip connections to preserve information that might be lost after going through too many layers because of the multi-stage architecture of the proposed model.

Scoring function: After passing a forward pass through all of the FouriER blocks, the result is mapped into the embedding space $\mathbb{R}^{d_e}$ by a linear transformation. The scoring function can be determined as:

$$\psi(s, r, o) = f(FouriER(q))\mathbf{e}_o + b_o \tag{3}$$

where q is the query embeddings, f is a nonlinear function, for which we are currently using ReLU, $\mathbf{e}_o$ and b_o are the embeddings and translational vector of candidate objects for projecting results into embedding space, respectively.

Loss function: We choose the standard Binary Cross Entropy (BCE) loss to train our model. The logits from the scoring function is applied a sigmoid function and then fed into the loss function together with labels. We also apply label smoothing to make the model more robust and generalized.

4 Experiments

We have evaluated our model against 2 benchmark datasets: WN18RR [9] and FB15k-237 [14]. Statistics of benchmark datasets are shown in the Table 1. Our evaluation metrics are Mean Ranking (MR), Mean Reciprocal Ranking (MRR), Hits@{1,3,10}. We also apply a uniform negative sampling procedure to produce more samples, where the set of corrupted triples is: $\mathcal{G}' = \{(s', r, o)|(s' \in \mathcal{E} \setminus s) \wedge (r \in \mathcal{R})\} \cup \{(s, r, o')|(o' \in \mathcal{E} \setminus o) \wedge (r \in \mathcal{R})\}$. Our model is optimized on the MRR metric during training with a learning rate of 0.0001, embedding size of 400 and label smoothing with probability of 0.1.

Results of TransE are adopted from [10]; DistMult and R-GCN from [9]; others are taken from their original papers. Our model is better than the traditional model using convolution, ConvE, or modern approaches such as RA-GCN

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

Dataset	n_e	n_r	Train	Val	Test
WN18RR	40943	11	86835	3034	3134
FB15k-237	14541	237	272115	17535	20466

Table 1: Statistics of benchmark datasets

Model	WN18RR					FB15k-237				
	MRR	MR	H@1	H@3	H@10	MRR	MR	H@1	H@3	H@10
TransE	.226	2300	–	–	.501	.294	347	–	–	.465
DistMult	.430	5110	.390	.440	.490	.241	254	.155	.263	.419
R-GCN	–	–	–	–	–	.248	–	.153	.258	.417
ConvE	.430	4187	.400	.440	.520	.325	244	.237	.356	.501
KBGAN	.214	–	–	–	.472	.278	–	–	–	.458
DMACM	.230	**552**	–	–	**.540**	.270	244	–	–	.440
RARL	.360	–	.351	–	.409	.320	–	.251	–	.491
RA-GCN	–	–	–	–	–	.249	–	–	–	.417
FouriER (ours)	**.437**	5911	**.413**	**.447**	.478	**.331**	**219**	**.242**	**.363**	**.510**

Table 2: Our proposed model outperforms baseline models.

Model	WN18RR					FB15k-237				
	MRR	MR	H@1	H@3	H@10	MRR	MR	H@1	H@3	H@10
PoolFormer	.432	5995	.407	.441	.476	.317	276	.232	.350	.487
FouriER	**.437**	**5911**	**.413**	**.447**	**.478**	**.331**	**219**	**.242**	**.363**	**.510**

Table 3: The full version of our model produces better results.

or DMACM, proving the efficiency of the proposed architecture. Table 2 compares benchmark datasets from various baselines to our model.

Ablation study: We also compare the original PoolFormer model to FouriER in the link prediction task. The results show that utilizing FouriER with the addition of FFT improves feature extraction, so that the results on every dataset are enhanced when compared to the method of employing a simple pooling layer as a token mixer. Table 3 compares the whole model to the ablated model.

5 Conclusions and future work

In this paper, we propose a novel model based on MetaFormer architecture and an FFT layer that can help extract more information from embeddings more efficiently. Our proposed model outperforms the baseline models on two benchmark datasets, yielding highly positive results. We also conducted an ablation study to figure out the beneficial effects of the FFT layer to the model. However, inversed relations have not yielded particularly promising results so that we intend to improve inversed relation prediction in the future.

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

References

[1] Weihao Yu, Mi Luo, Pan Zhou, Chenyang Si, Yichen Zhou, Xinchao Wang, Jiashi Feng, and Shuicheng Yan. Metaformer is actually what you need for vision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10819–10829, June 2022.

[2] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013.

[3] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. Learning entity and relation embeddings for knowledge graph completion. *Proceedings of the AAAI Conference on Artificial Intelligence*, 29(1), Feb. 2015.

[4] Giuseppe Pirrò. Relatedness and tbox-driven rule learning in large knowledge bases. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(03):2975–2982, Apr. 2020.

[5] Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. A three-way model for collective learning on multi-relational data. In *Proceedings of the 28th International Conference on International Conference on Machine Learning*, ICML'11. Omnipress, 2011.

[6] Bishan Yang, Scott Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. Embedding entities and relations for learning and inference in knowledge bases. In *Proceedings of the International Conference on Learning Representations (ICLR) 2015*, May 2015.

[7] Michael Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne vanÂ den Berg, Ivan Titov, and Max Welling. Modeling relational data with graph convolutional networks. In *The Semantic Web*.

[8] Anqi Tian, Chunhong Zhang, Miao Rang, Xueying Yang, and Zhiqiang Zhan. Ra-gcn: Relational aggregation graph convolutional network for knowledge graph completion. In *Proceedings of the 2020 12th International Conference on Machine Learning and Computing*, ICMLC 2020, pages 580–586. Association for Computing Machinery, 2020.

[9] Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. Convolutional 2d knowledge graph embeddings. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1), Apr. 2018.

[10] Dai Quoc Nguyen, Tu Dinh Nguyen, Dat Quoc Nguyen, and Dinh Phung. A novel embedding model for knowledge base completion based on convolutional neural network. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 327–333. Association for Computational Linguistics, June 2018.

[11] Liwei Cai and William Yang Wang. KBGAN: Adversarial learning for knowledge graph embeddings. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1470–1480. Association for Computational Linguistics, June 2018.

[12] Jin Huang, TingHua Zhang, Jia Zhu, Weihao Yu, Yong Tang, and Yang He. A deep embedding model for knowledge graph completion based on attention mechanism. *Neural Computing and Applications*, 33, 08 2021.

[13] James Lee-Thorp, Joshua Ainslie, Ilya Eckstein, and Santiago Ontanon. FNet: Mixing tokens with Fourier transforms. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4296–4313. Association for Computational Linguistics, July 2022.

[14] Kristina Toutanova and Danqi Chen. Observed versus latent features for knowledge base and text inference. In *3rd Workshop on Continuous Vector Space Models and Their Compositionality*. ACL - Association for Computational Linguistics, July 2015.