Machine Learning with Graphs - Final Project Proposal: KAN-GPS

Yuval Hanuka
yuvalhanuka@mail.tau.ac.il

Yair Pickholz Berliner
Pickholz.yair@gmail.com

September 2024

1 Problem Statement

The primary goal of this project is to enhance the expressive power and interpretability of Graph GPS networks by integrating KAN layers into their architecture, resulting in what we term as **KAN-GPS**. Specifically, we aim to:

- Design two KAN-GPS architectures that incorporate KAN layers into the GPS network framework.
- Evaluate the performance of these architectures on benchmark datasets and compare them with standard GPS networks.
- Analyze the interpretability and continual learning capabilities of KAN-GPS by visualizing learned univariate functions and understanding their impact on the model's predictions.

2 Proposed Methodology

We propose to develop two architectures of KAN-GPS by integrating KAN layers into the GPS network: **Hybrid-KAN-GPS** and **KAN-GPS**. Below, we provide the layer equations for the standard GPS, Hybrid-KAN-GPS, and KAN-GPS architectures.

2.1 GPS Layer

The GPS layer [6] is defined as:

$$
\begin{aligned}
X^{(l+1)},\ E^{(l+1)} &= \mathrm{GPS}^{(l)}(X^{(l)}, E^{(l)}, A) \\
&\text{computed as:} \\
X_M^{(l+1)},\ E^{(l+1)} &= \mathrm{MPNN}^{(l)}(X^{(l)}, E^{(l)}, A) \\
X_T^{(l+1)} &= \mathrm{GlobalAttn}^{(l)}(X^{(l)}) \\
X_M^{(l+1)} &= \mathrm{BatchNorm}(\mathrm{Dropout}(X_M^{(l+1)}) + X^{(l)}) \\
X_T^{(l+1)} &= \mathrm{BatchNorm}(\mathrm{Dropout}(X_T^{(l+1)}) + X^{(l)}) \\
X^{(l+1)} &= \mathrm{MLP}^{(l)}(X_M^{(l+1)} + X_T^{(l+1)})
\end{aligned}
$$

2.2 Hybrid-KAN-GPS Layer

In the **Hybrid-KAN-GPS** architecture, we use the standard GPS layer equation except we replace the final MLP with a KAN layer:

$$
\begin{aligned}
X^{(l+1)},\ E^{(l+1)} &= \text{Hybrid-KAN-GPS}^{(l)}(X^{(l)}, E^{(l)}, A) \\
\text{computed as:} & \\
X_M^{(l+1)},\ E^{(l+1)} &= \text{MPNN}^{(l)}(X^{(l)}, E^{(l)}, A) \\
X_T^{(l+1)} &= \text{GlobalAttn}^{(l)}(X^{(l)}) \\
X_M^{(l+1)} &= \text{BatchNorm}(\text{Dropout}(X_M^{(l+1)}) + X^{(l)}) \\
X_T^{(l+1)} &= \text{BatchNorm}(\text{Dropout}(X_T^{(l+1)}) + X^{(l)}) \\
X^{(l+1)} &= \text{KAN}^{(l)}(X_M^{(l+1)} + X_T^{(l+1)})
\end{aligned}
$$

Here, $\text{KAN}^{(l)}$ denotes the KAN layer at layer l, replacing the traditional MLP.

2.3 KAN-GPS Layer

In the **KAN-GPS** architecture, we replace both the MPNN and Global Attention mechanisms with their KAN-based counterparts:

$$
\begin{aligned}
X^{(l+1)},\ E^{(l+1)} &= \text{KAN-GPS}^{(l)}(X^{(l)}, E^{(l)}, A) \\
\text{computed as:} & \\
X_M^{(l+1)},\ E^{(l+1)} &= \text{KAN-MPNN}^{(l)}(X^{(l)}, E^{(l)}, A) \\
X_T^{(l+1)} &= \text{KAN-GlobalAttn}^{(l)}(X^{(l)}) \\
X_M^{(l+1)} &= \text{BatchNorm}(\text{Dropout}(X_M^{(l+1)}) + X^{(l)}) \\
X_T^{(l+1)} &= \text{BatchNorm}(\text{Dropout}(X_T^{(l+1)}) + X^{(l)}) \\
X^{(l+1)} &= \text{KAN}^{(l)}(X_M^{(l+1)} + X_T^{(l+1)})
\end{aligned}
$$

In this architecture:

- $\text{KAN-MPNN}^{(l)}$ is the KAN-based Message Passing Neural Network at layer l.
- $\text{KAN-GlobalAttn}^{(l)}$ is the KAN-based Global Attention mechanism at layer l.

By constructing these two architectures, we aim to investigate the impact of integrating KAN layers at different components of the GPS network. Specifically, we will assess how the substitution of MLPs with KANs in various parts of the model affects performance, interpretability, and continual learning capabilities.

3 Continual Learning

Kolmogorov-Arnold Networks (KANs) have shown potential in facilitating continual learning, which allows models to adapt to new tasks without forgetting previously learned information [1]. This is particularly valuable when models encounter sequential data or need to be updated incrementally.

In our project, we will focus on **Knowledge Retention**, which refers to the model's ability to maintain performance on previously learned tasks while learning new ones. We aim to evaluate the continual learning capabilities of the KAN-GPS architectures by testing their knowledge retention. Specifically, we will assess how well the KAN-GPS models preserve their understanding of earlier tasks when incrementally trained on new graph datasets or tasks, compared to standard GPS networks.

By examining knowledge retention, we hope to demonstrate the practical advantages of KAN-GPS models in dynamic environments where data evolves over time and continual learning is essential.

4 Evaluation and Datasets

We will examine the proposed models on benchmark datasets from the **Open Graph Benchmark (OGB)** [7] across different tasks. These tasks include node classification, link prediction, and graph classification. Using OGB datasets ensures that our evaluations are standardized and comparable to existing models.

For testing continual learning, we will simulate a sequential learning scenario where the models are trained on a sequence of tasks or datasets. We will monitor the performance on both new and previously learned tasks to evaluate knowledge retention and adaptability.

5 Interpretability

To offer interpretability, we will:

- **Visualization of Learned Functions**: Visualize the learned univariate spline functions in KAN layers to understand feature transformations in the data context.
- **Model Analysis**: Attempt to understand the relation between the MPNN and Global Attention layers in the model context, exploring how KAN layers influence the interaction between local and global information processing.

In this project, we aim to enhance the expressive power and interpretability of Graph GPS networks by integrating KAN layers, resulting in the KAN-GPS architectures. We will measure the success of our project through:

- **Improved Performance**: Achieving better results on several benchmark datasets compared to standard GPS and other baseline models. If the KAN-GPS models outperform existing methods on tasks such as node classification, link prediction, and graph classification, it will demonstrate the effectiveness of integrating KAN layers.
- **Practical Utility and Benefits**: Even if the models do not achieve the best performance, demonstrating that KAN-GPS models have good enough accuracy while offering advantages in continual learning and interpretability would be a significant contribution.

By evaluating both quantitative performance metrics and qualitative aspects like continual learning and interpretability, we aim to provide a comprehensive assessment of the KAN-GPS architectures. Success in these areas would confirm the potential of integrating KAN layers into GPS networks and contribute to the advancement of graph neural network research.

6 Literature Review

6.1 Kolmogorov-Arnold Neural Networks by Ziming Liu et al.

Ziming Liu et al. [1] introduced Kolmogorov-Arnold Networks (KANs), inspired by the Kolmogorov-Arnold representation theorem. They proposed replacing linear weights in neural networks with learnable univariate spline functions, enhancing the expressive power of the models. The authors demonstrated advantages in interpretability and continual learning capabilities of KANs. They applied KANs to various tasks, including function approximation and solving partial differential equations.

6.2 GKAN: Graph Kolmogorov-Arnold Networks by Mehrdad Kiamari et al.

Mehrdad Kiamari et al. [2] proposed GKAN, extending KANs to graph-structured data. They developed two architectures, GKAN Architecture 1 and GKAN Architecture 2, integrating KAN layers into graph neural networks. The authors evaluated GKAN on node classification tasks and showed promising results. They suggested that GKANs could serve as a foundation for integrating KANs into various graph deep learning schemes.

6.3 KAGNNs: Kolmogorov-Arnold Networks Meet Graph Learning by Roman Bresson et al.

Roman Bresson et al. [3] introduced KAGNNs, integrating KANs into Graph Neural Networks like GIN and GCN. They replaced MLP components with KAN layers to enhance the expressive power of the models. The authors evaluated KAGNNs on node classification, graph classification, and regression tasks. They found that KAGNNs can be valid alternatives to traditional MLP-based models, offering advantages in certain scenarios.

6.4 Graph GPS: A General, Powerful, Scalable Graph Transformer Architecture by Anton Rampášek et al.

Anton Rampášek et al. [6] proposed Graph GPS, a general framework for building graph neural networks. They combined local message passing with global attention mechanisms to capture both local and global graph structures effectively. The authors demonstrated that GPS networks can handle complex graph data and achieved state-of-the-art results on various benchmark datasets. Their work provides a foundation for integrating different components, like KANs, into graph neural network architectures.

References

[1] Ziming Liu, Yixuan Wang, Sachin Vaidya, et al. *Kolmogorov-Arnold Neural Networks.* arXiv preprint arXiv:2404.19756, 2024.

[2] Mehrdad Kiamari, Mohammad Kiamari, Bhaskar Krishnamachari. *GKAN: Graph Kolmogorov-Arnold Networks.* arXiv preprint arXiv:2406.06470, 2024.

[3] Roman Bresson, Giannis Nikolentzos, George Panagopoulos, et al. *KAGNNs: Kolmogorov-Arnold Networks Meet Graph Learning.* arXiv preprint arXiv:2406.18380, 2024.

[4] Gianluca De Carlo, Andrea Mastropietro, Aris Anagnostopoulos. *Kolmogorov–Arnold Graph Neural Networks.* arXiv preprint arXiv:2406.18354, 2024.

[5] Fan Zhang, Xin Zhang. *GraphKAN: Enhancing Feature Extraction with Graph Kolmogorov Arnold Networks.* arXiv preprint arXiv:2406.13597, 2024.

[6] Anton Rampášek, Davide Bacciu, Guy Wolf, et al. *Graph GPS: A General, Powerful, Scalable Graph Transformer Architecture.* arXiv preprint arXiv:2205.12454, 2022.

[7] Weihua Hu, Matthias Fey, Marinka Zitnik, et al. *Open Graph Benchmark: Datasets for Machine Learning on Graphs.* arXiv preprint arXiv:2005.00687, 2020.