Kolmogorov-Arnold Neural Network's optimization and architecture analysis

https://doi.org/10.31713/MCIT.2024.049

Ihor Serhiienko
igorserj2000@gmail.com

Abstract – **This study aims to conduct an in-depth analysis of the Kolmogorov-Arnold neural network architecture and its functioning principles, comparing it with the Multi-Layer Perceptron and identifying possible optimization paths. The Kolmogorov-Arnold Network, introduced in May 2024, retains a fully connected structure but introduces trainable activation functions on the edges instead of fixed activation functions at the nodes. This allows for increased modeling accuracy and efficiency, with splines helping KAN networks learn and adapt in a controlled manner.**

***Keywords* – artificial intelligence, Kolmogorov-Arnold network, neural network, multi-layer perceptron**

I. Introduction (Heading 1)

This year, in the field of deep learning, the Kolmogorov-Arnold Network (KAN) introduced by Jiming Liu, Xuan Wang, Sachin Vaidya, and others has gained significant attention. It can be used as an alternative to the currently popular Multi-Layer Perceptron (MLP) in some cases. Inspired by the Kolmogorov-Arnold representation theorem, the authors developed KAN, which shows extraordinary potential for the future of machine learning. The new network has many advantages compared to MLP, including higher accuracy and the ability to learn from small datasets, with the drawback of slower training speed.

Problem Statement: Conduct an in-depth analysis of the Kolmogorov-Arnold neural network architecture and its functioning principles, including a comparison with the Multi-Layer Perceptron and identifying possible optimization paths.

Objective: Study the architecture of the Kolmogorov-Arnold neural network, its applications, and identify effective approaches to optimize training algorithms.

II. Article analysis

In the article [5], the author provides brief information about the network itself and the KAN architecture. The Kolmogorov-Arnold Network is a new neural network architecture introduced in May 2024.

Comparison of KAN with MLP [4]: KAN retains a fully connected structure but introduces trainable activation functions on the edges instead of fixed activation functions at the nodes. In KAN, activation functions can be optimized during training, which increases modeling accuracy and efficiency. Splines help KAN networks learn and adapt in a more controlled manner, creating smooth curves that can change locally without altering the entire shape.

The Kolmogorov-Arnold theorem explains the use of more than one hidden layer in neural networks, as noted in study [1]. The theorem shows that any continuous function can be represented by a specific network with two hidden layers. Proposed modifications to the theorem allow the smoothness properties of the function to be transferred to the outer function, which can be well approximated by ReLU networks. Instead of two hidden layers, it is more natural to interpret the theorem as a deep neural network, where most layers are needed to approximate the inner function. New versions of the theorem are easier to prove and more practical for application in deep neural networks.

Kolmogorov-Arnold Networks (KANs) represent a significant advancement in neural network design by utilizing the Kolmogorov-Arnold Representation (KAR) theorem in conjunction with B-splines, resulting in a dynamic and powerful model. The KAR theorem provides a method to break down complex functions into simpler components. KANs implement this principle at every edge within the network, transforming each connection between neurons into a learnable B-spline activation function.

Advantages from Articles [4] and [5]:

- Improved accuracy compared to MLP.

- Better model interpretability.

- Combines the strengths of splines and MLP.

- KAN can achieve better results with fewer parameters. For example, a 2-layer KAN with a width of 10 can be 100 times more accurate and efficient than a 4-layer MLP with a width of 100 [4].

Disadvantages from Articles [4] and [5]:

- KAN cannot use GPUs, making their training 10 times slower compared to MLP.

- Scaling KAN is challenging due to computational complexity and memory requirements.

In Article [3], the authors use the KAN model for time series forecasting:

- Using splines to parameterize activation functions, allowing them to learn dynamically.

- Advantage over traditional Multi-Layer Perceptron (MLP) in satellite traffic prediction tasks.

- Ensuring more accurate results with fewer parameters.

Problems arising from noise in the dataset are described in article [2]. A small amount of noise in the data significantly degrades the performance of KAN. The author proposes a data pre-filtering technique using diffusion maps and increasing the volume of training data. Using both methods (pre-filtering and data augmentation) reduces the negative impact of noise, but optimizing the filtering parameters is challenging.

Potential applications and future research directions are described in article [5]:

- Improving the efficiency of large language models.
- Enhancing the interpretability of AI systems.
- Few-shot learning.
- Improved knowledge representation and reasoning.

One possible optimization direction is adapting KAN to work on GPUs, which would significantly increase training speed. Additionally, parallel computing and specialized hardware solutions can be applied to reduce training time without losing accuracy.

III. Conclusions

KAN is a promising architecture that can significantly improve machine learning in tasks where interpretability and accuracy are important. However, to fully utilize its potential, issues related to slow training speed and computational resource optimization need to be addressed. Recommended approaches include implementing GPU support for KAN, using parallel computing, and introducing specialized hardware solutions to more effectively use the architecture in practical machine learning tasks.

References

[1] The Kolmogorov–Arnold representation theorem revisited. URL: https://www.sciencedirect.com/science/article/pii/S0893608021000289

[2] Hu Y. et al. KAN: Kolmogorov-Arnold Networks. arXiv. 2023. URL: https://arxiv.org/abs/2407.14882.

[3] Huang Y., Yu W., Li T. KAN: A Formal Study of Kolmogorov-Arnold Networks. arXiv. 2023. URL: https://arxiv.org/abs/2405.08790.

[4] Dotessec. Kolmogorov-Arnold Network (KAN): A Game Changer in the Deep Learning World? Medium. URL: https://medium.com/@dotessec/kolmogorov-arnold-network-kan-a-game-changer-in-the-deep-learning-world-f04771989f67.

[5] Ahmed Z. What is the new Neural Network Architecture KAN (Kolmogorov Arnold Networks) Explained. Medium. URL: https://medium.com/@zahmed333/what-is-the-new-neural-network-architecture-kan-kolmogorov-arnold-networks-explained-d2787b013ade.