Position Embeddings: Connections to NTK and Fourier Features

Anonymous authors
Paper under double-blind review

Abstract

This paper explores the concept of position embeddings in deep learning models, with a particular focus on their connections to Neural Tangent Kernel (NTK) and Fourier features. We discuss how position embeddings enable sequence models to capture spatial or temporal relationships, and how they relate to the broader concepts of NTK and Fourier feature mappings. Our analysis sheds light on the theoretical underpinnings of these techniques and their practical implications for model performance and generalization.

1 Introduction

Position embeddings have become a crucial component in many deep learning architectures, particularly in natural language processing and computer vision tasks. They allow models to incorporate spatial or temporal information about the input sequence, enabling the network to distinguish between elements based on their position. This paper explores the theoretical foundations of position embeddings and their connections to two important concepts in machine learning: the Neural Tangent Kernel (NTK) and Fourier features.

2 Position Embeddings

Position embeddings are vector representations that encode the position of elements in a sequence. They are typically added to the input embeddings before being processed by the main network architecture. There are several types of position embeddings:

1. Learned position embeddings
2. Sinusoidal position embeddings
3. Relative position embeddings

The choice of position embedding can significantly impact model performance and generalization capabilities.

3 Neural Tangent Kernel (NTK)

The Neural Tangent Kernel is a theoretical framework that describes the behavior of wide neural networks during training (**?**). It provides a connection between neural networks and kernel methods, offering insights into the dynamics of neural network training and their generalization properties.

3.1 Connection to Position Embeddings

The NTK perspective can help us understand the role of position embeddings in the following ways:

1. Feature mapping: Position embeddings can be viewed as a form of feature mapping, similar to the implicit feature space induced by the NTK.

2. Infinite-width limit: As the width of a neural network approaches infinity, its behavior can be described by the NTK. Position embeddings contribute to this limiting behavior.

3. Generalization: The NTK provides insights into how neural networks generalize, which can be extended to understand the generalization properties of models with position embeddings.

4 Fourier Features

Fourier features are a technique for mapping input data to a higher-dimensional space using sinusoidal functions. They have been shown to improve the performance of neural networks, particularly in tasks involving continuous inputs (**?**).

4.1 Connection to Position Embeddings

The relationship between Fourier features and position embeddings is particularly interesting:

1. Sinusoidal embeddings: The sinusoidal position embeddings used in the Transformer architecture (**?**) can be seen as a form of Fourier feature mapping.

2. Spectral bias: Both Fourier features and certain types of position embeddings help neural networks overcome the spectral bias, allowing them to learn high-frequency functions more easily.

3. Periodic patterns: Fourier features and position embeddings both enable models to capture periodic patterns in data, which is crucial for many sequence modeling tasks.

5 Theoretical Analysis

Let's consider a simple position embedding function $\phi(x)$ for a 1D input x:

$$\phi(x) = [\sin(\omega_1 x), \cos(\omega_1 x), \sin(\omega_2 x), \cos(\omega_2 x), ..., \sin(\omega_d x), \cos(\omega_d x)] \tag{1}$$

where $\omega_i = 1/10000^{2i/d}$ for $i = 0, 1, ..., d/2 - 1$.

This embedding function has interesting connections to both NTK and Fourier features:

1. NTK connection: The kernel induced by this embedding approximates the NTK of a wide neural network with periodic activation functions.

2. Fourier connection: This embedding is essentially a Fourier feature mapping with a specific choice of frequencies.

6 Practical Implications

Understanding the connections between position embeddings, NTK, and Fourier features has several practical implications:

1. Architecture design: Insights from NTK and Fourier analysis can guide the design of more effective position embedding schemes.

2. Initialization strategies: The NTK perspective suggests initialization strategies that can improve training dynamics for models with position embeddings.

3. Transfer learning: The connection to Fourier features suggests that models with well-designed position embeddings may transfer more effectively to tasks with different input distributions.

7 Conclusion

Position embeddings play a crucial role in modern deep learning architectures. By exploring their connections to the Neural Tangent Kernel and Fourier features, we gain valuable insights into their theoretical properties and practical applications. This understanding can guide the development of more effective and generalizable models for sequence modeling tasks.

References