

Notes on: Graph Attention Networks

September 5, 2024

1 Abstract

Graph Attention Networks (GATs) is a novel network architecture that operate on graph-structured data, leveraging masked self-attentional layers to address the shortcomings of prior methods based on graph convolutions or their approximations. By stacking layers in which nodes are able to attend over their neighborhoods' features, we enable (implicitly) specifying different weights to different nodes in a neighborhood. This model is readily applicable to inductive and transductive problems.

2 Introduction

Graph Neural Networks (GNNs) were introduced in 2009, 2005 by Gori and Scarselli as a generalisation of recursive neural networks that can directly deal with a more general class of graphs. A convolution operation is defined in the Fourier domain by computing the eigendecomposition of the graph Laplacian, resulting in potentially intense computations and non-spatially localized filters.

However, spectral methods the learned filters depend on the Laplacian eigenbasis, which depends on the graph structure. Thus, a model trained on a specific structure can not directly be applied to a graph with a different structure.

Non-spectral approaches which define convolutions directly on the graph, operations on groups of spatially close neighborhoods. One of the challenges of these approaches is to define an operator which works with different sized neighborhoods and maintains the weight sharing property of CNNs.

Inspired by the challenges and advances in recent works, this paper introduces an attention-based architecture to perform node classification of graph-structured data. The idea is to compute the hidden representations of each node in the graph, by attending over its neighbors, following a self-attention mechanism.

The attention architecture has several interesting properties:

- the operation is efficient, since it is parallelizable across node-neighbor pairs

- it can be applied to graph nodes having different degrees by specifying arbitrary weights to the neighbors.
- the model is directly applicable to inductive learning problems including tasks where the model has to generalise to complete unseen graphs.