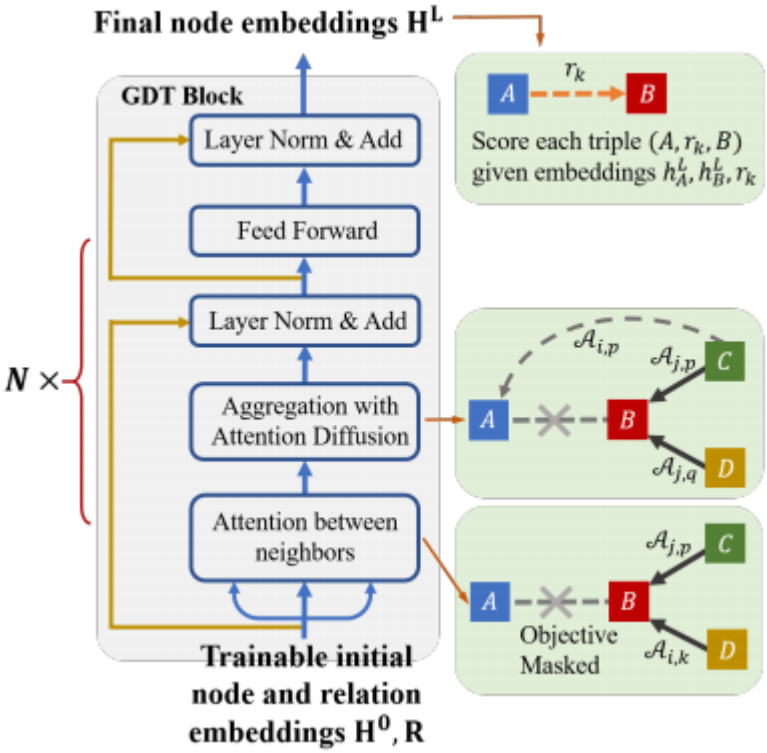
Docket #: S20-271

State-of-the-Art Graph Diffusion Transformer for Natural Language Processing

Researchers at Stanford have developed a potentially best-in-class method for performing knowledge graph completion tasks. Their innovation, called Graph Diffusion Transformer (GDT), advances the state of the art for completion of knowledge graphs, namely node classification and link prediction, and can be applied widely, e.g., medical knowledge graphs. Transformer architecture introduced the notion of self-attention (allowing the model to direct its focus and pay attention to different parts of the input) which leads to high performance in many natural language processing tasks. However, extending the notion to complex relational structures, such as graphs, remains a challenge. **The Stanford innovation provides a scalable self-attention mechanism for graph data**. It diffuses the attention scores from neighboring nodes to non-neighboring nodes, thus benefiting from the expressiveness of full self-attention. Experimental results on standard semi-supervised node classification as well as the knowledge graph completion show that GDT achieves state-of-the-art results.



GDT architecture. Each GDT block consists of attention computation, attention diffusion, layer normalization, feed forward layers, and 2 residual connections for each block. GDT blocks can be stacked to constitute a deep model. As illustrated on the right, context-dependent attention is achieved via the attention diffusion process. Here A, B, C, D ? V are nodes in the graph. (image credit: the inventors)

Stage of Development

Experimental results on standard semi-supervised node classification as well as the knowledge graph completion show that GDT achieves state-of-the-art results: GDT achieves up to 5:7% relative error reduction over previous state-of-the-art on Cora, Citeseer, and Pubmed. GDT also obtains the best performance on a large-scale Open Graph Benchmark dataset. On knowledge graph completion, GDT advances state-of-the art on WN18RR and FB15k-237 across four different performance metrics.

Applications

- Widely applicable to knowledge graph completion tasks
- Analytics of graphical data, e.g., online retail, social networks, search engines

Advantages

- Outperforms state-of-the-art methods on the standard tasks of node classification and knowledge graph completion
- Enables context-dependent attention between any pair of nodes in the graph
- Enhances large-scale structural information and learns more informative attention distribution

Innovators

- Guangtao Wang
- Zhitao Ying
- Jurij Leskovec
- Jing Huang

Licensing Contact

David Mallin

Licensing Manager, Physical Sciences

Email