GraphStar Sentiment Analysis

Graph data models are a set of subjects/ entities (nodes) and their inter-relationships (edges). In a simplest classification setting, the model attempts to predict unlabeled nodes by the surrounding labeled nodes. The graph domain is also categorized into spectral and non-spectral approaches. Spectral approaches often work with methods to represent the graph. Non-spectral approaches often conduct local-convolutions on the graph in the neighborhood nodes. In order to capture the global state, a local conventional model has to increase its depth, but this may encounter an over-smoothing issue. On the other hand, a non-local representation to capture long-range dependencies uses fully-connected-nodes attention and aggregates all features as a weighted sum which results in system overhead and high computational complexity.

GraphStar is able to map the global state effectively without incurring system overhead and heavy computation costs.

GraphStar trains the model to: (1) perform inductive tasks on previously unseen graph data; (2) aggregate both local and long-range information, making the model globally aware, extracting high-level abstraction typically not represented in individual node features; (3) the relays serve as a hierarchical representation of the graphs and can be used directly for graph classification tasks.

GraphStar can be used in three general graph tasks like:

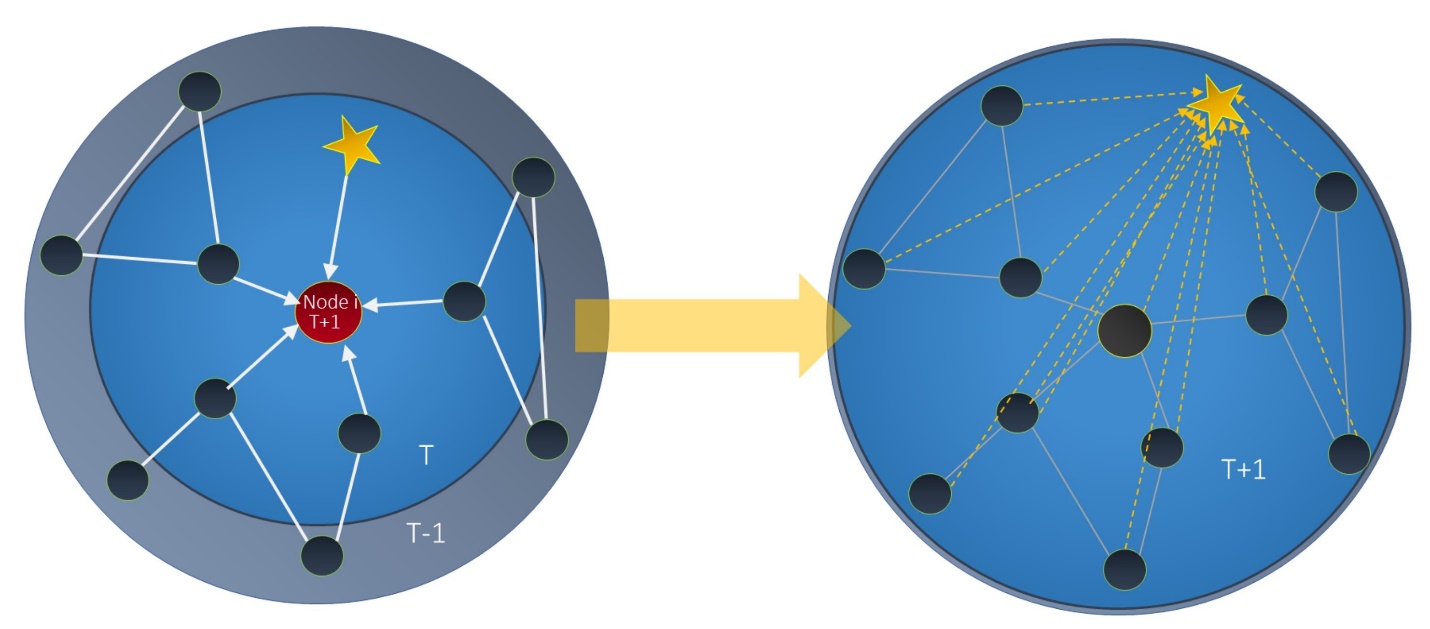
* Node classification: predict a property of a node
* Graph classification: categorize different graphs
* Link prediction: predict whether there are missing links between two nodes.

**GraphStar Model Architecture:**

Step 1: Initial Representation of the Star

Step 2: Real Node Update

Step 3: Star Update

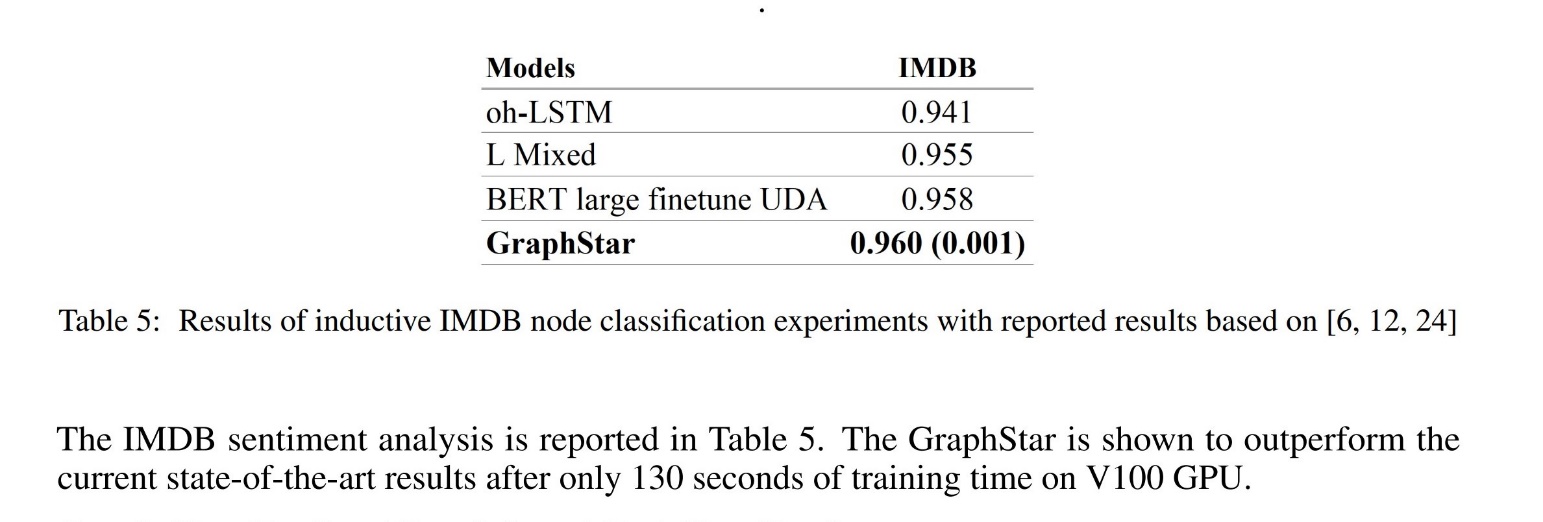


**IMDB Sentiment Analysis:**

GraphStar propose a new method to tackle sentiment analysis based on node classification and use IMDB-binary dataset as an example. This dataset was originally 4 not a graph task; it is usually treated as a natural language processing task (NLP). It is a binary sentiment analysis consisting of 50,000 reviews labeled as either positive or negative.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **# Nodes** | **# Features** | **# Classes** | **# Training Nodes** | **# Validation Nodes** | **# Test Nodes** |
| 50000 | 1024 | 2 | 23003 (12 graph) | 1997 (1 graph) | 25000 (13 graphs) |

GraphStar turns the pure NLP task into a graph data task based on document embedding. First, for IMBD, the model uses a pre-trained large Bert model to document encoding, and treats every film review as a node in a graph. Then it links the nodes (film reviews) which belong to the same topic and create a graph dataset. This approach is highly generalizable to most topic-specific classification tasks.



**Citations:**

Haonan, Lu, et al. "Graph star net for generalized multi-task learning." *arXiv preprint arXiv:1906.12330* (2019).

GitHub Code: https://github.com/graph-star-team/graph\_star