**GC-LSTM: Graph Convolution Embedded LSTM for Dynamic Link Prediction**

**What is dynamic link prediction?**

-Predict future network structure based on historic network information. It is an efficient feature learning tool for complex networks. Example applications as social network predicting, economy predicting, biology predicting and industry predicting etc.

**What is the importance of Dynamic link prediction?**

-Major task is learning how to learn the different evolution modes of each node.

-For example, two nodes have the same network structure at first, but their different evolution models reflect different social strategies. These differences are reflecting their social needs, change their social relationship in the future.

-Most dynamic link prediction methods take advantage of historic network information and compact them into one network to predict network of next moment.

**LSTM**

-Specific form of RNN that can process time-dependent data of long-term dependence.

-Used in image field, video field, language model, speech recognition and machine translation.

-Adaptively capture the dependencies among multi-dimensional interactions based on the learned representations for each time slot in dynamic networks.

**GCN (Graph Convolutional Network)**

-Learn the structural characteristics of network data, thereby implementing various tasks, such as network representation learning, node classification.

**GC-LSTM**

-New deep model which propose a novel end-to-end dynamic link prediction.

-Capable of handling links that are going to appear or disappear.

-Effectively handle high-dimensional, time-dependent and sparse structural sequence data.

**Overall framework**

* Encoder model is Graph Convolution Network (GCN) embedded LSTM, using GCN to learn network structure of the cell state c and the hidden state h of each moment snapshot, while using LSTM to learn the temporal information of the state of each link.
* Decoder is a fully connected layer network to convert the extracted features mapping back to the original space.
* GC-LSTM will output the predicted network and implement link prediction in a unified fashion.

**GC-LSTM Model**

* Utilize the LSTM to solve long-term dependency problems and effectively learn temporal feature of the dynamic graph. The link states of each node in the dynamic network may use LSTM to implement timing prediction, when predict links at the next moment.
* Relies on two state values, the hidden state h which is used to extract the input information at the last time, and the cell state c which is used to save the long-term information.
* First step is to decide what information will be thrown away from the previous cell state.
* Next step is to update the cell state.

**Loss Function and Model Training**

-The purpose of training the entire GC-LSTM model is to improve the accuracy of the dynamic link prediction.

-Adopt the Adam optimizer to optimize our model.

**Experimental datasets**

-Carry out the experiments on four real-world datasets, with each one representing a dynamic network. Most networks are human relation networks, with nodes representing humans and links showing their connections.

* CONTACT and HYPERTEXT09: Human contact dynamic networks of face-to-face proximity.
* ENRON and RADOSLAW [48]: Email networks

**Experimental baseline methods**

* CN: one of the similarity-based prediction methods, a link is more likely to exist between two nodes with more common neighbors.
* node2vec: a network embedding method by mapping the nodes of a network from a high dimensional space to a lower dimensional vector space.
* LINE: Integrates local and global information to learn node representations.
* TNE: Model network evolution as a Markov process and then uses the matrix factorization to learn the embedding vectors for all nodes.

**Evaluation Metrics**

* AUC
* GMAUC
* Error Rate

**Experiment Result**

-GC-LSTM model outperforms all baseline methods in all dynamic networks for both the short-term and long-term prediction capabilities.

-Most of the baseline methods may predict more invalid links than the number of links that exist, resulting in a relatively large Error Rate

-GC-LSTM model has better performance in dynamic link prediction accuracy

-GC-LSTM model not only uses LSTM to learn the timing characteristics of the sequence network but also uses GCN to learn the network characteristics of each snapshot.

-In the four data sets, ENRON is the most unsatisfactory in each evaluation index. This may be because the evolutionary model of ENRON has changed greatly during the evolution process, so the evaluation results of our pre- 9 trained GC-LSTM model on the test set are relatively poor

-GC-LSTM method, although Error Rate+ is slightly larger than Error Rate-, our method’s Error Rate+ is still the smallest of all comparison experiments.

-GC-LSTM model has the lowest Error Rate on all four data sets, both for short-term and long-term prediction.