CS598 Project: Deep Learning for Healthcare

Team 61: Deepa N Veeravalli & Sudhir Koundinya Nagesh

Introduction

We have chosen "Pre-training of Graph Augmented Transformers for Medical Recommendation" paper, authored by Junjun Shang, Tengfei Ma, Cao Xiao and Jimeng Sun, for our final project. The authors propose G-BERT graph learning model for medication recommendation.

The advances in Deep Learning technologies, availability of resources and data have provided great opportunities in the field of medical recommendation. The authors argue that existing medication recommendation models often learn the representations from the longitudinal EHR data from a small number of patients with multiple visits, ignoring patients with single visit (*selection bias*) and ignore the complex hierarchical relationships among diagnosis, medications, symptoms, and diseases (lack of hierarchical knowledge), which can lead to suboptimal recommendations. To address these issues and enhance the prediction and interpretability, they propose G-BERT graph learning model. The proposed model leverages medical knowledge encoded in large-scale medical corpora and single visit patient information to pre-train the GAT model to learn the underlying structure of medical concepts and relationships. The purpose of pre-training is mainly to leverage patient record with single visit and provide model trainings with good initializations.

The pre-trained GAT model is then fine-tuned on patient medical records with multiple visits, to personalize the medication recommendation for each patient. The model can capture not only the direct relationships between medications and symptoms but also the indirect relationships between medications through shared diseases or symptoms.

Method

The authors evaluate their proposed approach on MIMIC-III dataset and show that it outperforms state-of-the-art medication recommendation methods. The statistics of the data is provided below.

Stats	Single-Visit	Multi-Visit
# of patients	30,745	6,350
avg # of visits	1.00	2.36
avg # of dx	39	10.51
avg # of rx	52	8.80
# of unique dx	1,997	1,958
# of unique rx	323	145

Table 1: Data Statistics (dx for diagnosis and rx for medication)

The overall framework of G-BERT consists of three main parts: Ontology Embedding, BERT and fine-tuned classifier, which is described below.
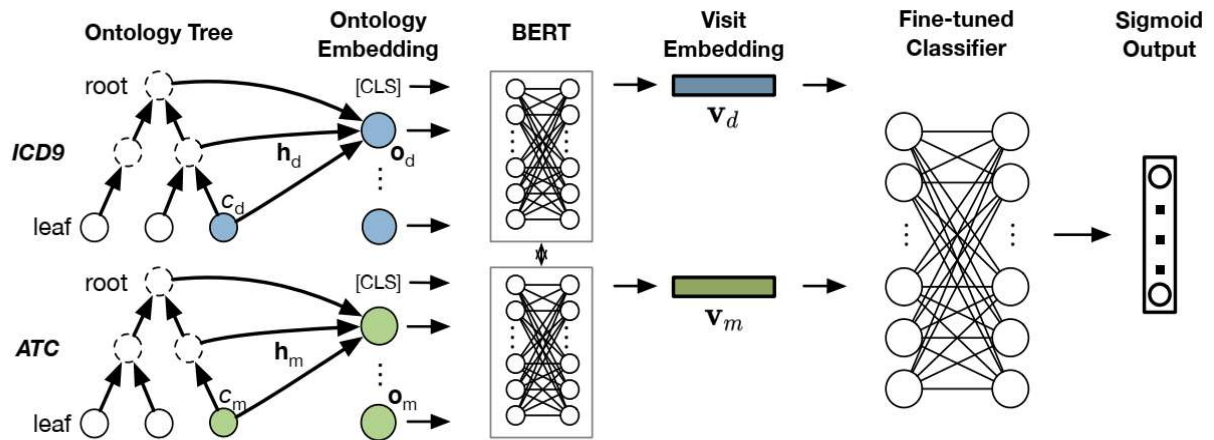
Fig 1: G-BERT Framework

The specific enhancements discussed are:

- Pre-training to leverage more data - adapt the framework of BERT and pre-train the model on each visit of the EHR data to leverage the single visit data that were not fit for training in earlier medication recommendation models.
- Medical ontology embedding with graph neural networks
- Ontology embedding for medical code laid in leaf nodes by cooperating ancestors' information based on graph attention networks.
- Next, input set of diagnosis and medication ontology embedding separately to shared weight BERT which is pre- trained using single visit data.
- Finally, concatenate the mean of all previous visit embeddings and the last visit embedding as input and fine-tune the prediction layers for medication recommendation tasks.

The proposed approach is interesting and innovative because it addresses the key limitations of existing medication recommendation models, such as selection bias, the lack of consideration of complex hierarchical relationships among diagnosis, medications, symptoms, and diseases and increases interpretability. By using a pre-trained GAT model, the proposed approach can provide more personalized and effective medication recommendations for patients, leading to improved patient outcomes.

Result

The authors evaluated their proposed approach on MIMIC-III dataset. They compared the performance of G-BERT model with other models like Logistic Regression, LEAP, RETAIN, GRAM and GAMENet models. Below table provides the performance comparison of G-BERT with other models.

Methods	Jaccard	PR-AUC	F1	# of parameters
LR	0.4075	0.6716	0.5658	-
GRAM	0.4176	0.6638	0.5788	3,763,668
LEAP	0.3921	0.5855	0.5508	1,488,148
RETAIN	0.4456	0.6838	0.6064	2,054,869
GAMENet$^-$	0.4401	0.6672	0.5996	5,518,646
GAMENet	0.4555	0.6854	0.6126	5,518,646
G$-$BERT$_{G-,P-}$	0.4186	0.6649	0.5796	2,634,145
G$-$BERT$_{G-}$	0.4299	0.6771	0.5903	2,634,145
G$-$BERT$_{P-}$	0.4236	0.6704	0.5844	3,034,045
G$-$BERT	**0.4565**	**0.6960**	**0.6152**	3,034,045

Table 2: Performance Comparison

As part of the reproduction study, we would like to try to achieve Jaccard Similarity score >= 0.35, PR-AUC >= 0.50 and F1 score >= 0.50 using the same MIMIC-III dataset.

As part of ablation, we plan to use the medical code embedding without ontology as input, with/without pre-training and compare the performance with the G-BERT. This will help us understand the effectiveness of medical ontology in medication prediction tasks. To study the impact of the transformer architecture, we might compare the performance of the proposed method with other transformer-based models or non-transformer-based models.

Conclusion

In summary, we would like to reproduce the results from G-BERT model using MIMIC-III synthetic data made available in the git repository (https://github.com/jshang123/G-Bert). We plan to develop our own code but on a need basis we may plan to reuse some of the existing code. We will use our personal Windows/M1/M2 Macs and will transition to google cloud service if we hit performance issues. That being said, the paper provides some information on the hardware used for their experiments. They used an NVIDIA V100 GPU with 16GB memory for pre-training the transformer model on a large corpus of medical records. For fine-tuning the model on the medication recommendation task, they used a NVIDIA TITAN XP GPU with 12GB memory. Based on this information, it is likely that a GPU with at least 12GB memory would be needed for fine-tuning the model on the medication recommendation task. As an extension we will try to apply the same on MIMIC IV dataset.

References

Asif Qamar, *Alumni UIUC, SupportVectors AI Labs.*