Graph Neural Network Approaches to EHR Data

David Pitt

HMC Math 189AC

dpitt@g.hmc.edu

December 7, 2021

Overview

Motivation

Survey information-theoretic approaches to structuring and learning from the same EHR dataset

- Learning from a graph structure
- Learning the structure itself
- Generating a better structure
- More complex organization/representation

General terminology

- Consider a graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$
- Nodes $\mathcal{V}$ represented by feature matrix X, class vector Y
- Adjacency matrix A
- Embeddings of feature vectors in latent space $= H$
- Standard nonlinear activation function σ
- Trainable parameters W (and sometimes b)

HORDE

Construct a multi-modal EHR graph

- Graph contains multiple classes of node: patient, event, and concept
- Patients are connected to concepts and events that relate to their hospital visits
- Graph varies in time (discrete visit sequence)
- Edges $\mathcal{E} = \{\mathcal{E}_t\}_{t=0}^{n}$
- Time-invariant embeddings computed through MPNN process:

$$h_i^l = \sigma\left(\sum_{v_j \in \mathcal{V}_{inv}} \frac{h_j^{l-1} W_{inv}^l}{|N(v_i)| \cdot |N(v_j)|} \right)$$

- Time-variant node embeddings updated using MPNN, then passed through LSTM

LSTM-GNN

The majority of diseases and procedures occur relatively infrequently

- Connect similar cases in a graph of patients
- A becomes $\mathcal{M}$, a relatedness matrix

$$\mathcal{M}_{ij} = a \sum_{\mu=1}^{m} (\mathcal{D}_{i\mu}\mathcal{D}_{j\mu}(d_{\mu}^{-1} + c)) - \sum_{\mu=1}^{m} (\mathcal{D}_{i\mu} + \mathcal{D}_{j\mu})$$

- $\mathcal{D}$ is a row matrix of diagnoses
- MPNN neighborhood is chosen using k most related neighbors
- Hyperparameters a, c, k

LSTM-GNN

- X is divided into X_{var} and X_{inv}, time-variant and time-invariant
- Time-variant features are fed into LSTM, invariant to a GCN
- Embeddings concatenated and classified by a FCN

Attention models

Adjacency matrix not known a priori

- Learn edges using the transformer architecture
- There are several self-attention mechanisms

GCT

First application of the transformer to a graph

- A is naively initialized with conditional probabilities of co-occurrence
- Some connections are forbidden: operations at each step are multiplied by mask M

$$\hat{A}^{(j)} = \text{softmax}\left(\frac{C^{(j-1)}W_Q^{(j)}(C^{(j-1)}W_K^{(j)})^T}{\sqrt{d}} \cdot M\right)$$

- C represents output of last encoding

VGNN

Key advance: variational

- Features passed through embedding: $X \to H$
- Model learns distribution $p(z_i \mid h_i)$
- Sample from $q(\hat{h}_i \mid z_i)$
- K-head attention predicts $\widehat{A}$
- Attention coefficients $e_{ij} = \sigma(a^T[Wh_i \mid\mid Wh_j])/\sqrt{dim(h_i)}$
- $A_{ij} = \frac{exp(e_{ij})}{\sum_{p \in N(i)} exp(e_{ip})}$
- $H^{(l+1)} = \text{FFN}\ [A^{(l)}(H^{(l)}W^{(l)} + b^{(l)})]$

GRAM

Key advance: new hierarchical structure

- Concept nodes are placed in a DAG, where depth encodes specificity
- MPNN creates embeddings for concepts
- Visits are represented by a row matrix of concept vectors
- Sequence of visit embeddings passed through RNN, final output is a softmax disease prediction

GraphSMOTE

Key advance: node interpolation

- GraphSAGE layer creates embeddings
- Add more nodes of a highly underrepresented class

$$h^1_{v'} = (1 - \delta) \cdot h^1_v + \delta \cdot h^1_{nn},$$

- Predict edges for new nodes:

$$E(u, v) = \text{softmax}(\sigma(h^1_v \cdot S \cdot h^1_u)),$$

- Final embeddings are passed through one MPNN and one linear layer

RioGNN

Key advance: iterative node deletion

- Penalize dissimilar embeddings within a class using similarity
$$\mathcal{S}(u, v) = 1 - \mathcal{D}(u, v),$$
- We define a Markov Decision Process with an action space of node and edge deletions
- Reward function is the average similarity of a neighborhood
- After pruning, the model uses one MPNN layer to update embeddings
- Repeated over several layers
- Similarity is fed into loss

Walk-GNN

Key advance: a new kind of recommender system

- Construct a knowledge graph of patient and concept nodes
- Agent walks along graph
- Define an MDP, state space $s_t = \{p_e, e_t, h_t\}$
- Action space $a_t = (r_{t+1}, e_{t+1})$ (walking to a new node).
- Reward function prioritizes walking to disease nodes that match diagnosis
- Agent learns optimal walk policy π_t

EHR Datasets

Two main datasets used in medical ML

- MIMIC-III contains information over three years at Beth Israel Hospital
- eICU contains information from same time period from 300+ ICUs across the US
- Deidentified, tagged patient records, including encounters, admission data, prescriptions, lab results and formal diagnoses
- Newer versions of the dataset include doctors' notes in natural language

Results

Model	AUPRC
HORDE [1]	0.433
LSTM-GNN [2]	0.386
GCT [3]	0.581
VGNN [4]	0.603
GRAM [5]	0.575
RioGNN [7]	0.572*

**Provided their own MIMIC matrix, not standard*

Takeaways

- Graph RL has huge potential in health record analysis
- New approaches continue to arise
- ML outperforms humans at constructing knowledge graphs
- No need to constrain to traditional MPNNs

References

[1] Harmonized representation learning on dynamic EHR graphs
https://www.sciencedirect.com/science/article/pii/S153204642030054X?via
[2] Predicting patient outcomes with representation learning
https://paperswithcode.com/paper/predicting-patient-outcomes-with-graph
[3] Learning the Graphical Structure of Electronic Health Records with Graph Convolutional Transformer
https://paperswithcode.com/paper/graph-convolutional-transformer-learning-the
[4] Variationally Regularized Graph-based representation learning for electronic health records
https://paperswithcode.com/paper/graph-neural-network-on-electronic-health
[5] GRAM: Graph-based attention model for healthcare representation learning
https://paperswithcode.com/paper/gram-graph-based-attention-model-for
[6] GraphSMOTE: Imbalanced Node Classification on Graphs with Graph Neural Networks https://github.com/TianxiangZhao/GraphSmote
[7] Reinforced Neighborhood Selection Guided Multi-Relational Graph Neural Networks https://arxiv.org/pdf/2104.07886.pdf
[8] Sun, Zhoujian et al. "Interpretable Disease Prediction based on Reinforcement Path Reasoning over Knowledge Graphs." Oct 2020.
https://arxiv.org/pdf/2010.08300.pdf

The End