VGNN
David Pitt
Introduction
Model
Data
Results

Variationally Regularized Graph Neural Network

David Pitt

HMC Math 189AC

dpitt@g.hmc.edu

November 28, 2021

Overview

1. Introduction
2. Model
   - Data
3. Results

Motivation

Only a small subset of problems have Euclidean data

- Represent other systems as a knowledge graph
- Support from some computational theories of mind – "semantic knowledge graph"
-

Motivation

Traditional GNNs learn to classify graphs $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, or elements of the sets $\mathcal{V}, \mathcal{E}$.

- Representation learning has been gaining traction
- Issue: lack of generality
- Models like GraphSAGE have a tendency to overfit
- Solution: introduce a generative element and regularization

Autoencoders

Key idea: **autoencoders filter out irrelevant information**

Split up into two components

- Encoder downsamples data
- Decoder generally upsamples to reproduce data
- Loss penalizes incorrect reconstructions

Graph autoencoders

We can apply the autoencoder method to a graph dataset

- Model takes in features X, adjacency matrix A
- Encoder downsamples to some latent space Z
- Decoder reconstructs $\widehat{A}$

Graph autoencoders

We can apply the autoencoder method to a graph dataset

- Model takes in features X, adjacency matrix A
- Encoder downsamples to some latent space Z
- Decoder reconstructs $\widehat{A}$

General overview

From the paper *Variationally Regularized Graph-based Representation Learning for Electronic Health Records* [Zhu and Razavian, 2021]

- Model takes in features X
- 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 $\widehat{A}$
- $\widehat{H}$ fed to FCN for classification

Encoding

K-head attention is used to construct A at every layer

- Iteratively downsample the graph
- Attention coefficients $e_{ij} = \sigma(a^T[Wh_i \mid\mid Wh_j])/\sqrt{dim(h_i)}$
- $A_{ij} = \frac{exp(e_{ij})}{\sum_{p \in N(i)} exp(e_{ip})}$
- $H^{(l+1)} = \text{FFN}\ [A^{(l)}(H^{(l)}W^{(l)} + b^{(l)})]$
- W and b form a linear layer

Latent Space and Decoding

- Output of encoder: latent variables Z
- Approximate $z_i \sim \mathcal{N}(\mu, \Sigma)$
- Sample z_i from prior distribution
- Reconstruct $\widehat{H}$ iteratively
- Learn function $\hat{h}_i \rightarrow \hat{y}_i$

EHR Datasets

Two main datasets used in medical ML

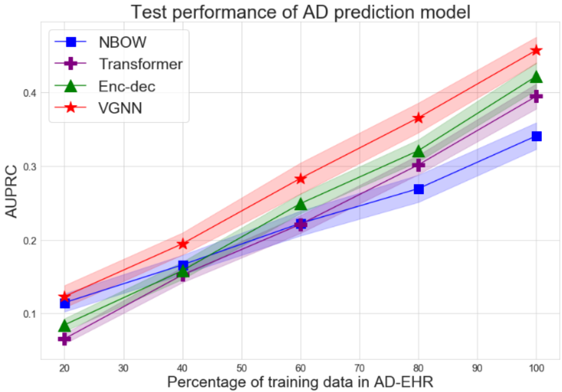
- MIMIC-III contains information over three years at Beth Israel Hospital
- eICU contains information from same time period from 300+ ICUs across the US
- Deidentified, tagged patient records, including encounters, admission data, prescriptions, lab results and formal diagnoses
- Newer versions of the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

Results

Table 2: Model evaluation on the test set using precision-recall curves (99% confidence interval)

Method	**AD-EHR** AUPRC	**AD-EHR** PPV@0.4Recall	**MIMIC-III Mortality** AUPRC	**eICU Readmission** AUPRC
Random Forest [4]	0.2316 ± 0.0043	0.0890 ± 0.0029	0.5976 ± 0.0056	0.3614 ± 0.0049
MLP[44]	0.3775 ± 0.0050	0.5623 ± 0.0182	0.6646 ± 0.0045	0.3639 ± 0.0045
RNN* [30]	0.2590 ± 0.0045	0.3038 ± 0.0041	—	—
CNN* [39]	0.3566 ± 0.0053	0.4267 ± 0.0056	—	—
NBOW [23]	0.3386 ± 0.0049	0.5265 ± 0.0138	0.6787 ± 0.0054	0.3730 ± 0.0049
Transformer [13]	0.3957 ± 0.0044	0.6844 ± 0.0165	0.6777 ± 0.0051	0.3792 ± 0.0042
GCT [13]	0.3409 ± 0.0040	0.5174 ± 0.0095	0.6810 ± 0.0046	0.3794 ± 0.0045
Enc-dec (Ours)	0.4216 ± 0.0047	0.6756 ± 0.0109	0.6962 ± 0.0051	0.3881 ± 0.0047
VGNN (Ours)	**0.4580 ± 0.0048**	**0.7489 ± 0.0075**	**0.7102 ± 0.0046**	**0.3986 ± 0.0050**

Results


Test performance of AD prediction model
NBOW
Transformer
Enc-dec
VGNN
AUPRC
0.1
0.2
0.3
0.4
20
30
40
50
60
70
80
90
100
Percentage of training data in AD-EHR

References

Zhu and Razavian (2021)
Variationally Regularized Graph-based Representation Learning for Electronic Health Records
Preprint

The End