

Master thesis

Jannik Gut
28. April 2021

Content

- Autoencoder
 - Papers
 - Application
- EHR dataset
 - Papers
 - Dataset
 - Application
- Discussion
 - Which topic to propose?

Dynamic joint variational autoencoder

- All graphs are already generated and can be used to train
- One part is the graph autoencoder from Welling
 - Kullback-Leibler etc.
- The other part is assuming that the changes are smooth
 - Punish Gaussian random walk differences from the current graph compared to the previous graph
- Good results on authorship graphs

Multimodal Network Embedding via Attention based Multi-view Variational Autoencoder

- Two kinds of multimodal
 - Nodes are images with title
 - Nodes have different kinds of edges (here views)
- Nodes get one hidden layer computed with an RNN on the title using attention from the image (content view)
- Network view generate using DeepWalk (I think)
- Late vs. early fusion
 - Keeping separate → can't use useful relations like “same location”
 - Connected → concatenated early and autoencoded together → can lead to uneven weights
 - **Mixed** → encode separate first, then encode together, sample from this and decode separately

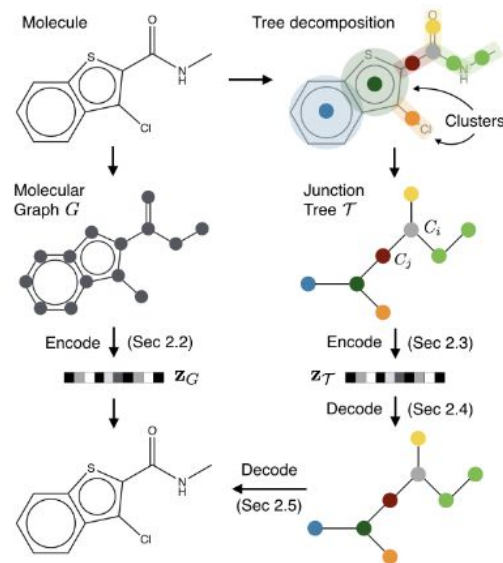
Heterogeneous graph transformer

- Weights for all the things
- Uses attention based on meta relation triplets, this enables for all types to keep their space
 - Starting node type
 - Relation type
 - Ending node type
- Attention, 8 heads
 - Start is key
 - End is query
 - Attention matrix is relation type dependent
- Special sampling for dense subgraphs
- Performs well in HUGE author graph

Junction Tree Variational Autoencoder for Molecular Graph

Generation

- There are 780 structures pre selected
 - Single edges and rings
- Autoencoder on two levels:
- Tree structure
 - Tree of structures
 - First, check if new structure, second determine which one
- Fine-grained connectivity
 - Different parts where to dock onto ring
-
- Hidden graph state computed using bond type dependent GNN weights
- Teacher forcing in training
- Two levels guarantee validity of molecule



Representation Learning for Networks in Biology and Medicine

Task	Goal	Data	Methods	Problems
generating protein structure	generate valid proteins	amino acid sequences	generate edges based on distance generate 3D fingerprints	have to adhere validity
characterising protein interactions	fill blanks of PPI network	protein-protein interactions (PPI)	embed both proteins and then predict link construct knowledge graph with NLP on PubMed	very noisy
predicting protein phenotypes	identify protein functions	gene ontology, structural data of proteins, PPI	hierarchical graph on GO terms Transformers	
cell type aware protein representation learning	predict if a gene is expressed by a sequence (node label prediction)	PPI	protein embeddings	different cells have different embeddings
integrating gene expression data	identify disease specific markers to classify diseases	gene expression data often as matrix, PPI	condgen (GCN,VAE,GAN)	gene expression data can be noisy and variational PPI does not have all the information needed
leveraging non-coding elements	add non-coding information to methods	interaction networks between coding and non-coding RNA snippets (lncRNA, miRNA and diseases)	node2vec or DeepWalk for node embeddings	non-coding snippets are long experiments are non-complete
fusing genome-wide data	improve disease classification models by fusing coding and non-coding elements	coding elements and non-coding elements in as two knowledge graphs	fusing the two knowledge graphs	
disease classification using subgraphs	form a subgraph consisting of the symptoms of a patient to predict his disease	human phenotype ontology (HPO)	subGNN	
generating drug structure	generate valid drug structures	nodes as molecules, edges as bonds	use edge features, JT-VAE	valid drug
characterising drug interactions	regression on interaction edges	nodes as compounds, edges indicate similarity chemical structures and side effects as nodes	TDA, node2vec	labour and cost intensive experiments to test
quantifying drug efficacy	regression on drug efficacy	gene expression data, gene ontology, drug similarity medical subject headings (MeSH),PPI	TDA, DeepWalk	combinational explosion
cell-line specific prediction of drug pairs	edge regression, where edge means interaction	PPI, protein-drug interactions, drug-drug interactions	transfer learning	drug effects on the body are not uniform
characterising diseases through medical imaging	construct graphs from images to do classification	cell graphs from MRI	cgexplainer	interpretability is crucial
personalising medical knowledge networks	edge regression	medical knowledge graph, EHR	predict on hierarchical knowledge graph via GAT,word2vec or LINE RNN for a sequence of drugs/diseases mixed pooling multi-view self-attention autoencoder	
integrating electronic health records (EHR) for personalised medicine	edge regression on patients to new disease or treatments	bioentities (e.g. treatment), patient as meta-node	add a patient as a meta-node	

Multimodal generation idea

- There are some relations which are easy to annotate, but others are hard/expensive
- Use an autoencoder to generate tentative edges, which then can be tested
- Best dataset
 - Different graphs, not one knowledge graphs
 - Graphs not huge
 - Amount of relation types not huge
- Drug side effects discovery?

Latent Patient Network Learning for Automatic Diagnosis

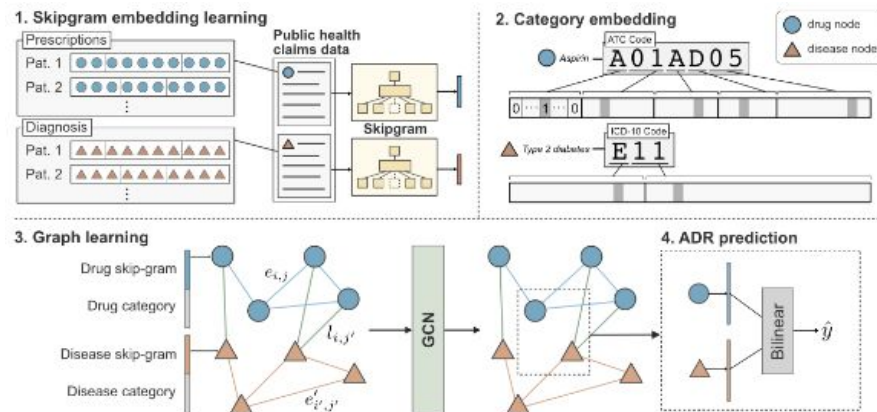
- Patients are nodes
 - Infer best connections/adjacency matrix
 - Nodes are embedded using Graph AE (MLP)
 - Edge weight proportional to distance between embeddings
-
- Patients influence each other?
 - Each patient should be for their own
 - Hard to explain to patients, I guess

GAMENet: Graph Augmented MEmory Networks for Recommending Medication Combination

- Recommend one-hot encoded medication
- Patients embedded using RNN as patients often have many visits
- MANN (Memory augmented neural nets)
 - Attention
 - Memory/knowledge from DDI graph
- One content loss and a DDI loss
- DDI Thresholded training
 - Usually only content loss, chance of only DDI loss proportional to overshoot over threshold

Drug-disease Graph: Predicting Adverse Drug Reaction Signals via Graph Neural Network with Clinical Data

- Node embedding based on skip gram on data and ICD-10
 - (Potential for pre-training?)
- Edges constructed
 - Based on distance between homogeneous
 - Based on dataset regularity for heterogeneous
- Link prediction after GNN rounds



Integrating biomedical research and electronic health records to create knowledge-based biologically meaningful machine-readable embeddings

- SPOKE dataset underlying
- Pagerank (in 2019)
- Weird disambiguation based on BMI

SubGNN

- Adds another layer of hierarchy with components
- Information flows through 3 separate channels
 - Position (in the graph, in contrast to other components)
 - Neighbourhood features (internal and external neighbourhood)
 - Structure (internal connectivity)
- Use anchor patches to generate messages to components
- Aggregate on component
- Aggregate on subgraph
-
- Widely outperforms other methods also on medical datasets on classification or more theoretical measures like density or cut ration.
- Patients are a set of conditions
 - Mapping possible, but how exactly, conditions usually are not connected
- Attention instead of standard message passing performs worse
 - Attention more interpretable, maybe something possible?

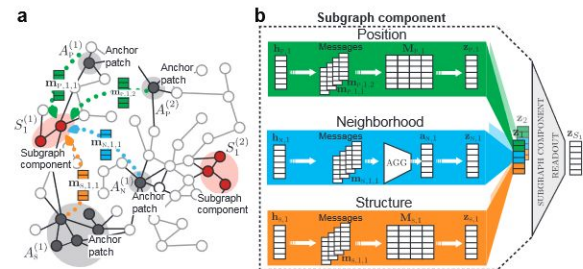


Figure 2: SUBGNN architecture. a. Property-specific messages $MSG_x^{A \rightarrow S}$ are propagated from anchor patches A_x to components of subgraph S . Here, $m_{x,c,i}$ denotes property- x message from i -th anchor patch $A_x^{(i)}$ to c -th component $S^{(c)}$. b. SUBGNN specifies three channels, which are each designed to capture a distinct subgraph property. Channel outputs z_x are concatenated to produce a final subgraph representation z_S .

SubGNN

- Adds another layer of hierarchy with components
- Information flows through 3 separate channels
 - Position (in the graph, in contrast to other components)
 - Neighbourhood features (internal and external neighbourhood)
 - Structure (internal connectivity)
- Use anchor patches to generate messages to components
- Aggregate on component

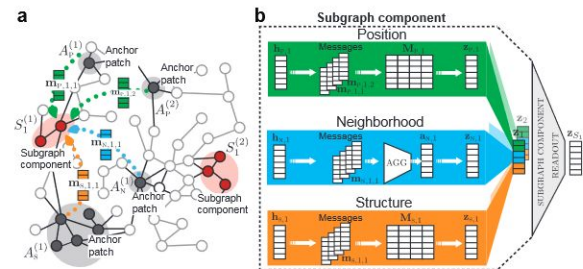


Figure 2: SUBGNN architecture. a. Property-specific messages $MSG_x^{A \rightarrow S}$ are propagated from anchor patches A_x to components of subgraph S . Here, $m_{x,c,i}$ denotes property- x message from i -th anchor patch $A_x^{(i)}$ to c -th component $S^{(c)}$. b. SUBGNN specifies three channels, which are each designed to capture a distinct subgraph property. Channel outputs z_x are concatenated to produce a final subgraph representation z_S .

Model extensions. Our implementation of SUBGNN initializes component embeddings with pre-trained node embeddings, which could alternatively be learned end-to-end. While we specify similarity functions for each channel, we note that γ_x can be replaced with any function measuring the relevance of anchor patches to subgraph components, including learnable similarity functions. Finally, while this paper focuses on subgraph classification, the representation learning modules we present can be trivially extended to any unsupervised, semi-supervised, or other prediction tasks, simply by modifying the loss functions. SUBGNN can also be integrated into other frameworks as a sub-module for end-to-end training.

Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis

- Phenotype discovery using clustering
- Outcome prediction
- Representation learning
- De-identification
- Interpretability
 - Attention
 - Constraints
 - Clustering
 - Mimic
- Pre-training/transfer learning were useful before

BAG dataset

Data:

- Patient attributes (age, sex, etc.)
- Medical conditions
- Treatment
- Way of exit
- (recurring visits?, I think not)

A set of conditions and treatments, which may be represented as a simple bi-partite graph of conditions and treatments, with some attributes

Questions:

- Predict way of exit (classification)
- Find missing entries (may be many)

Find missing elements

- Operate on Knowledge graph with conditions and treatments (SPOKE?)
- Skip-gram (out-of-bag prediction)
 - Use SubGNN to represent patient as a query
 - Pre-compute other nodes as answers
 - Rank best missing nodes to find not tested conditions or treatments
- If this is skip-gram, is transformer sensible? Use ICD-hierarchy as types for attention.
- May predict one or many nodes with different decoding strategies
 - Set of nodes
 - Top-k, other
 - Another subgraph representation
 - Dynamic subgraph
- Impact of pre-training of classification
- Impact of curriculum training with starting with easier graphs
- Baselines
 - [Set based neural nets](#)
 - Clustering around metanode to get one patient node
- Transfer learning from US dataset?
- Interpretable?

Discussion about which topic

- When I started I wanted to challenge me as far as I probably ever will
 - Also theoretically
- Autoencoder HUGE field
 - My initial ideas were all already found years ago
 - What is state of the art?
 - No flash of genius
 - Even if, already found
- Could the found autoencoder be used in the field?
- Can EHR be (theoretically) challenging?