Recent Healthcare Applications and Methods

CNeRG Reading Group
Soumyadeep, Ankan
2nd December 2021

Talk Outline

- Different Medical Data-types - clinical, genetic, graph, text
- Methods in different Medical Applications
  - Self-supervision (autoencoder, transformer), graphs
- Learning Methods - Zero shot and Supervised.
- Medical QnA Understanding
- Summary of contributions
- Future Research Directions

Rise of Health Track - confs, tracks

- Started in 2020 - ACM Transactions on Computing for Healthcare
- New conference - ACM Conference on Health, Inference, and Learning
- Prevalent track in top-tier conferences
  - Health on the Web in WebConf
  - BioKDD - oldest workshop, going on for 18 years
- Well-known workshops and Labs
  - ACL BioNLP
  - Cambridge Lab - Inspiration Seminar (https://www.vanderschaar-lab.com/)

CIKM 2021 Accepted Full Paper stats

- **COVID and Bioinformatics** constitute around 3% of papers
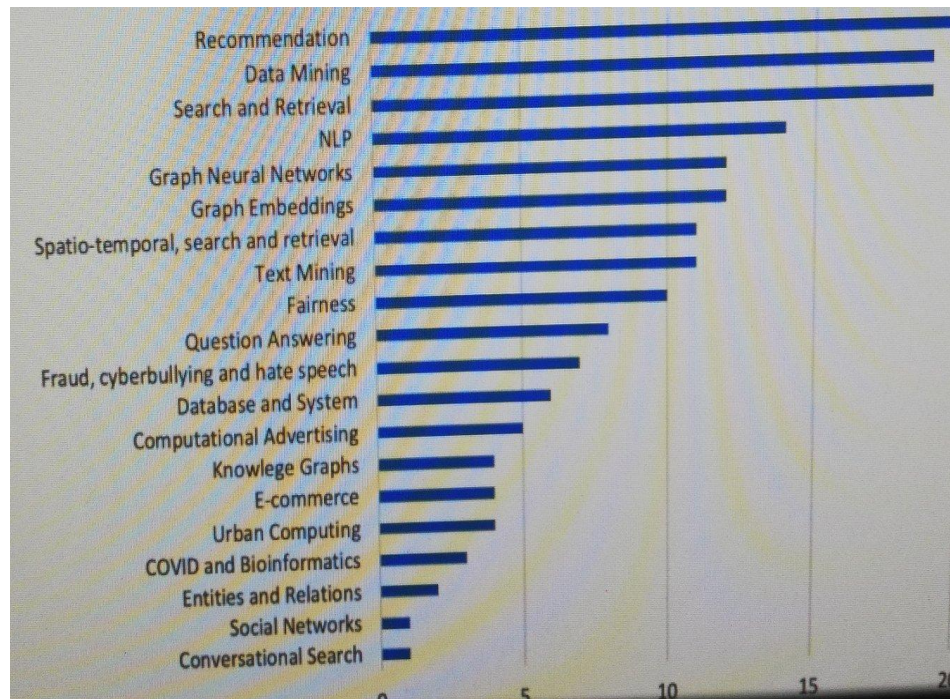
Retrospective studies not going to clinical practice

- Multi-stakeholder involvement of patients, medical professionals, care-givers, insurance companies, government
    - Few published papers are actually being translated into clinical practice
   - Prospective validation primarily done by randomized clinical trials

Clinical Data (2 papers)

- Sparse, high dimensional, treated as tabular data
  - Features do not have spatial or semantic structure like in image or language
  - Usually low number of patients, but more than genetic data
  - 10% of features have <30% missing data

|	datasrc	lrrkid	event_id	alc_NOTREGN	alc_avgltrg	alclf100	alclfreg	alcagstr	alcyrstr	alcstrdk	alcagstp	alcyrstp	alcstpdk	alcnot	a
1	1	10001	V01	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
2	2	10002	V01	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
3	3	10003	V01	5	2	1	1	NA	1975	NA	NA	1980	NA	0	
4	4	10004	V01	5	2	1	0	NA	NA	NA	NA	NA	NA	NA	
5	5	10005	V01	5	2	0	NA	NA	NA	NA	NA	NA	NA	NA	
6	6	10006	V01	5	2	0	NA	NA	NA	NA	NA	NA	NA	NA	
7	7	10007	V01	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
8	8	10008	V01	1	2	1	NA	NA	NA	-9	40	NA	NA	1	
9	9	10009	V01	1	2	NA	0	NA	NA	NA	NA	NA	NA	NA	
10	10	10010	V01	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
11	1	20001	V01	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
12	2	20002	V01	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
13	3	20003	V01	5	2	1	0	NA	NA	NA	NA	NA	NA	NA	
14	4	20004	V01	5	2	0	NA	NA	NA	NA	NA	NA	NA	NA	
15	5	20005	V01	5	2	0	NA	NA	NA	NA	NA	NA	NA	NA	
16	6	20006	V01	5	2	1	1	25	NA	NA	NA	NA	-5	0	
17	7	20007	V01	5	2	1	1	18	1998	NA	NA	NA	NA	0	
18	9	20009	V01	1	2	NA	0	NA	NA	NA	NA	NA	NA	NA	
19	10	20010	V01	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
20	1	30001	V01	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
21	2	30002	V01	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	

Genetic Data (2 papers)

- Genetic data - very high-dimensional (millions), very small vocabulary (5: A, T, C, G, N), small number of patients
  - DNA is a sequence of nucleotides (A, T, C, G)
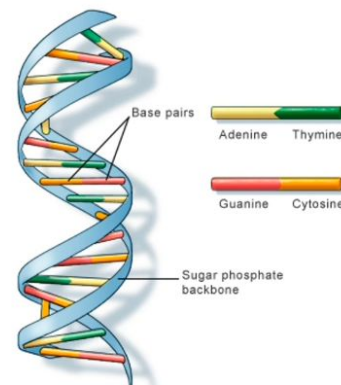
U.S. National Library of Medicine

Graphs (2 papers)

- NLP - biomedical abstracts (most common is PubMed), clinical notes, radiology reports)
- Protein-protein interaction network, gene-disease-gene network

Source: https://het.io/hnep/
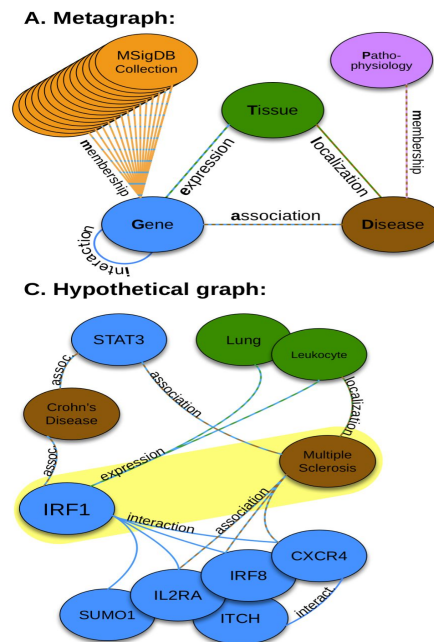
VIME: autoencoder model for tabular data (NIPS 2020)

- For tabular data, standard augmentations of image or NLP does not work
- Introduces novel pretext task of masked vector estimation for self-supervised setting
  - Previously, pretext task involved recovering original from corrupted form and/or mask vector

VIME: autoencoder model for tabular data

Self-supervised version
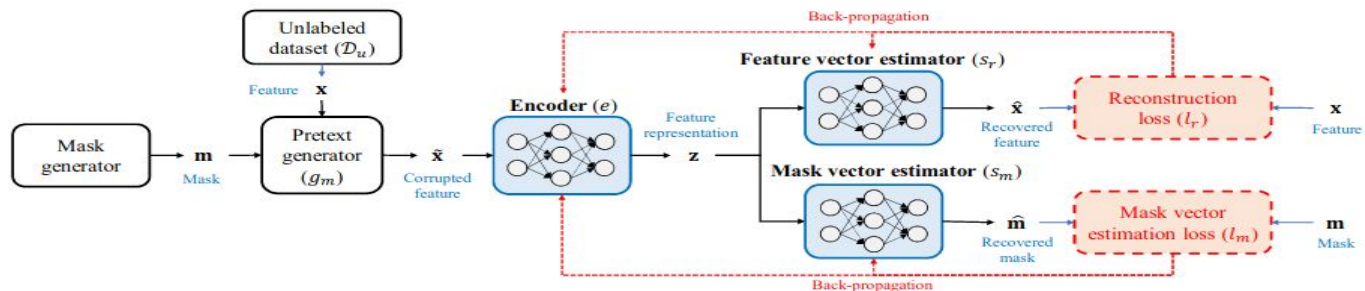
Figure 1: Block diagram of the proposed self-supervised learning framework on tabular data. (1) Mask generator generates binary mask vector ($\mathbf{m}$) which is combined with an input sample ($\mathbf{x}$) to create a masked and corrupted sample ($\tilde{\mathbf{x}}$), (2) Encoder (e) transforms $\tilde{\mathbf{x}}$ into a latent representation ($\mathbf{z}$), (3) Mask vector estimator (s_m) is trained by minimizing the cross-entropy loss with $\mathbf{m}$, feature vector estimator (s_r) is trained by minimizing the reconstruction loss with $\mathbf{x}$, (4) Encoder (e) is trained by minimizing the weighted sum of both losses.

VIME: Extending the Success of Self- and Semi-supervised Learning to Tabular Domain, NeurIPS 2020

Semi-supervised version
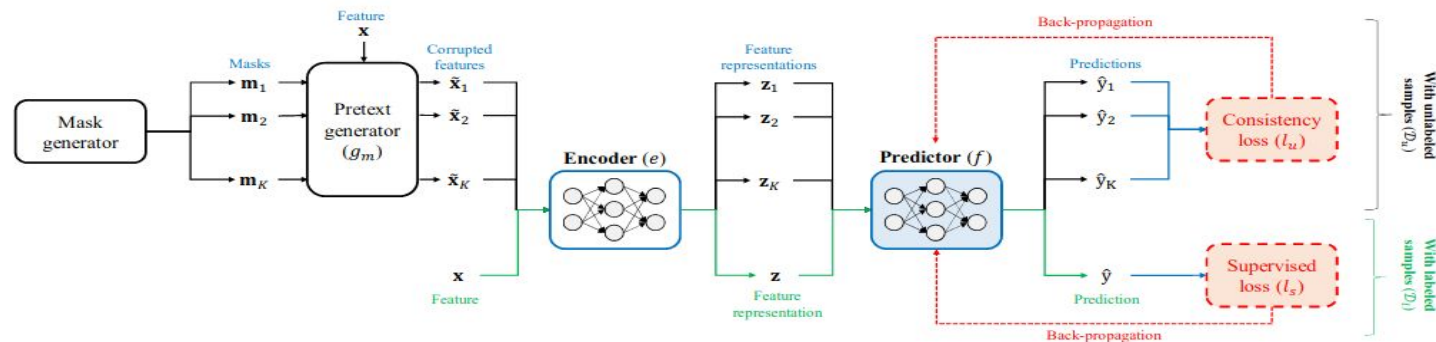
Figure 2: Block diagram of the proposed semi-supervised learning framework on tabular data. For an unlabeled sample $\mathbf{x}$ in $\mathcal{D}_u$, (1) Mask generator generates K-number of mask vectors and combine each of them with $\mathbf{x}$ to generate the corrupted samples $\tilde{\mathbf{x}}_k$, $k = 1, ..., K$ via pretext generator (g_m), (2) Encoder (e) transforms these corrupted samples into latent representations $\mathbf{z}_k$, $k = 1, ..., K$ as K different augmented samples, (3) Predictive model is trained by minimizing the supervised loss on ($\mathbf{x}, y$) in $\mathcal{D}_l$ and the consistency loss on the augmented samples ($\mathbf{z}_k$, $k = 1, ..., K$) jointly. The block diagram of the proposed self- and semi-supervised learning frameworks on exemplary tabular data can be found in the Supplementary Materials (Figure 2).

VIME: Extending the Success of Self- and Semi-supervised Learning to Tabular Domain, NeurIPS 2020

Feature Imputation and Label Prediction

- Feature Imputation: missing feature values are estimated using observed values
- Label prediction: downstream labels are learnt directly from incomplete data
- Previous strategies include GANs or autoencoders
  - Do not completely utilize feature values from other observations
  - Bias due to initializing missing values with default values

Handling Missing Data with Graph Representation Learning

- GRAPE, a general framework for feature imputation and label prediction in the presence of missing data
- Formulate as a bipartite graph representation
  - observations and features are two types of nodes
  - Observed feature values are attributed edges between the nodes
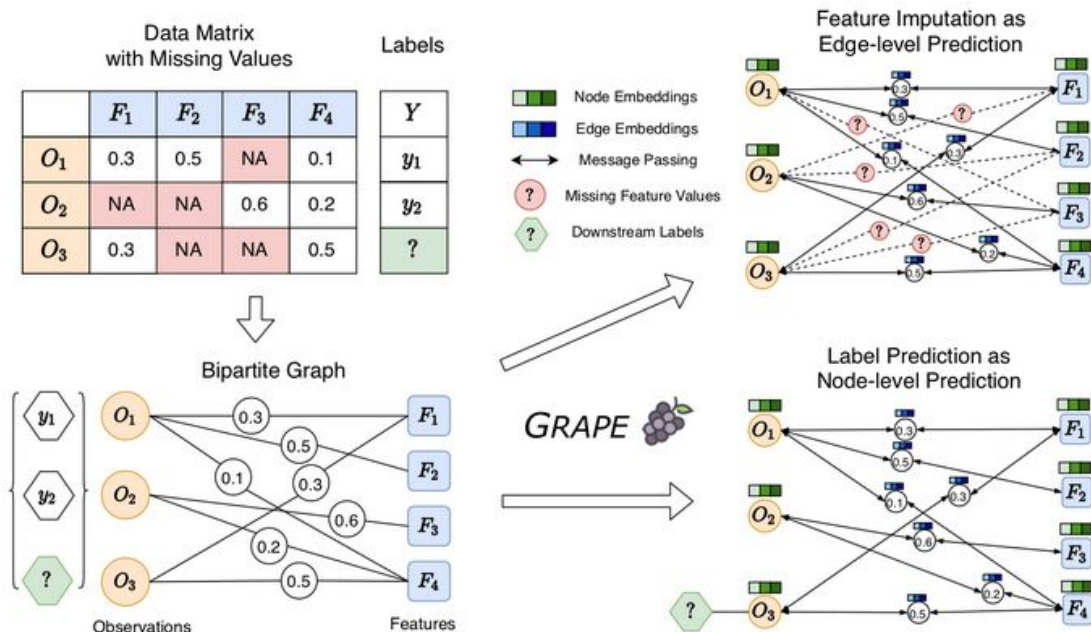
Figure 1: Overview of the proposed GRAPE framework

Handling Missing Data with Graph Representation Learning. Jiaxuan You*, Xiaobai Ma*, Daisy Yi Ding*, Mykel Kochenderfer, Jure Leskovec. Neural Information Processing Systems (NeurIPS), 2020.

GRAPE background (GraphSAGE, NeurIPS 2017)

First layer
$$h_v^0 = x_v$$
Feature vector

Trainable weights

During message passing, concatenate the learnt edge embedding

Hidden layers
$$h_v^k = \sigma(w_k \cdot f_{agg}(h_u^{k-1}, u \in N(v)), B_k h_v^{k-1})$$

Update edge embedding by new trainable matrix

Aggregate function

Concatenation

Final layer
$$z_v = h_v^K$$

Source: Explanation of GraphSAGE paper ([Link](#))
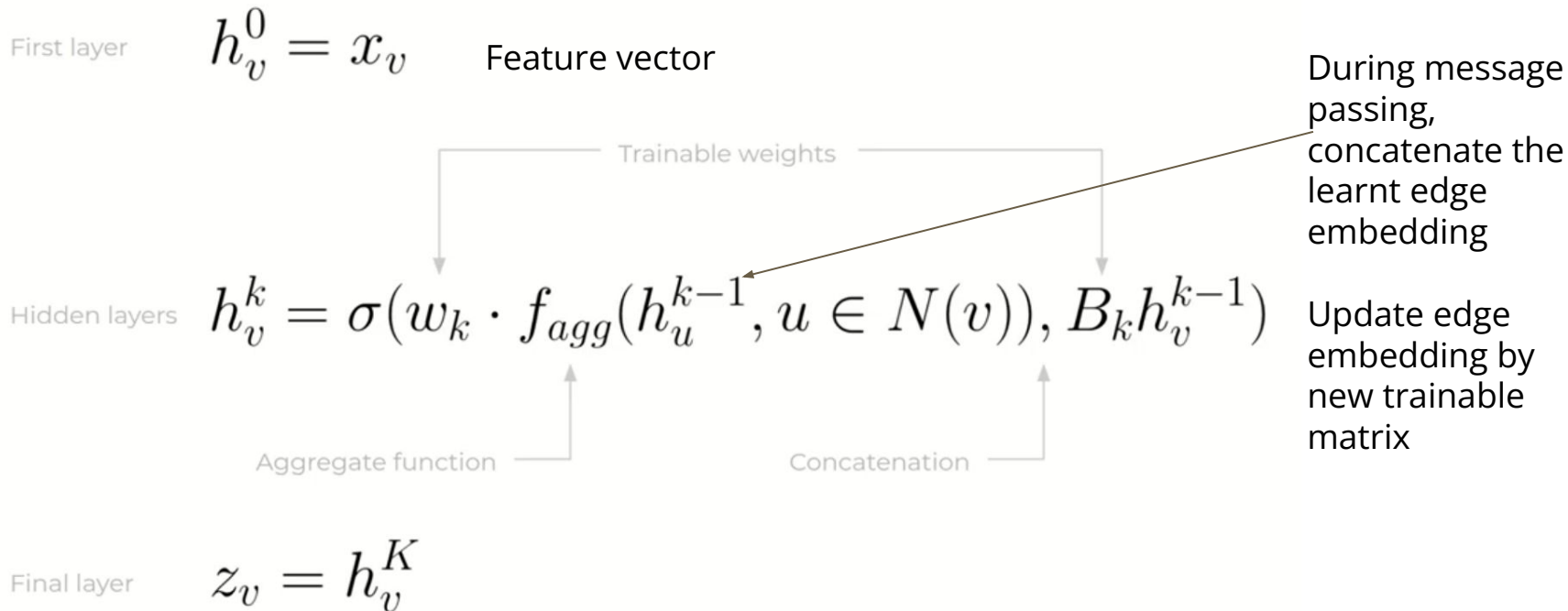
GRAPE Algorithm

GRAPE GNN architecture. Given that our bipartite graph $\mathcal{G}$ has important information on its edges, we modify GraphSAGE architecture by introducing *edge embeddings*. At each GNN layer l, the message passing function takes the concatenation of the embedding of the source node $\mathbf{h}_v^{(l-1)}$ and the edge embedding $\mathbf{e}_{uv}^{(l-1)}$ as the input:

$$\mathbf{n}_v^{(l)} = \text{AGG}_l\Big(\sigma(\mathbf{P}^{(l)} \cdot \text{CONCAT}(\mathbf{h}_v^{(l-1)}, \mathbf{e}_{uv}^{(l-1)}) \mid \forall u \in \mathcal{N}(v, \mathcal{E}_{drop}))\Big) \tag{1}$$

where AGG_l is the aggregation function, σ is the non-linearity, $\mathbf{P}^{(l)}$ is the trainable weight, $\mathcal{N}$ is the node neighborhood function. Node embedding $\mathbf{h}_v^{(l)}$ is then updated using:

$$\mathbf{h}_v^{(l)} = \sigma(\mathbf{Q}^{(l)} \cdot \text{CONCAT}(\mathbf{h}_v^{(l-1)}, \mathbf{n}_v^{(l)})) \tag{2}$$

where $\mathbf{Q}^{(l)}$ is the trainable weight, we additionally update the edge embedding $\mathbf{e}_{uv}^{(l)}$ by:

$$\mathbf{e}_{uv}^{(l)} = \sigma(\mathbf{W}^{(l)} \cdot \text{CONCAT}(\mathbf{e}_{uv}^{(l-1)}, \mathbf{h}_u^{(l)}, \mathbf{h}_v^{(l)})) \tag{3}$$

where $\mathbf{W}^{(l)}$ is the trainable weight. To make edge level predictions at the L-th layer:

$$\hat{\mathbf{D}}_{uv} = \mathbf{O}_{edge}(\text{CONCAT}(\mathbf{h}_u^{(L)}, \mathbf{h}_v^{(L)})) \tag{4}$$

The node-level prediction is made using the imputed dataset $\hat{\mathbf{D}}$:

$$\hat{\mathbf{Y}}_u = \mathbf{O}_{node}(\hat{\mathbf{D}}_{u\cdot}) \tag{5}$$

where $\mathbf{O}_{edge}$ and $\mathbf{O}_{node}$ are feedforward neural networks.

Key takeaways

- Under this graph representation, we formulate the feature imputation as an edge-level prediction task, and the label prediction as a node-level prediction task
- GRAPE yields 20% lower mean absolute error (MAE) for feature imputation, and 10% lower MAE for label prediction, compared with the best baselines
- Observe consistent gains under different missing data ratios (10 to 70%)
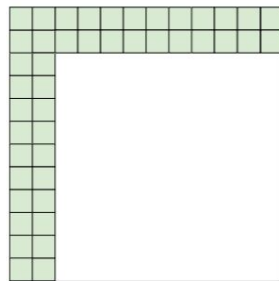
BigBird (NIPS 2020)

- BigBird Transformer - A scalable model for really long sequences
- Modified attention
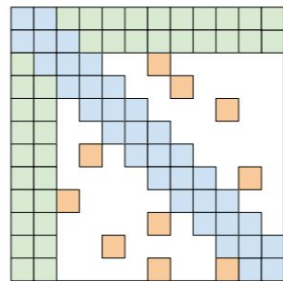 - global + window + random attention
  - 512 → 4096 tokens
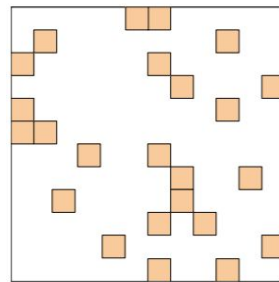
(a) Random attention
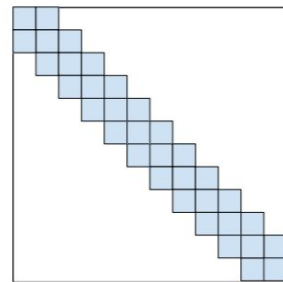
(b) Window attention

(c) Global Attention

(d) BIGBIRD

17

Gene Sequence Data

- Raw sequences are randomly split into docs
- Tokenization into words of ~8BP per token
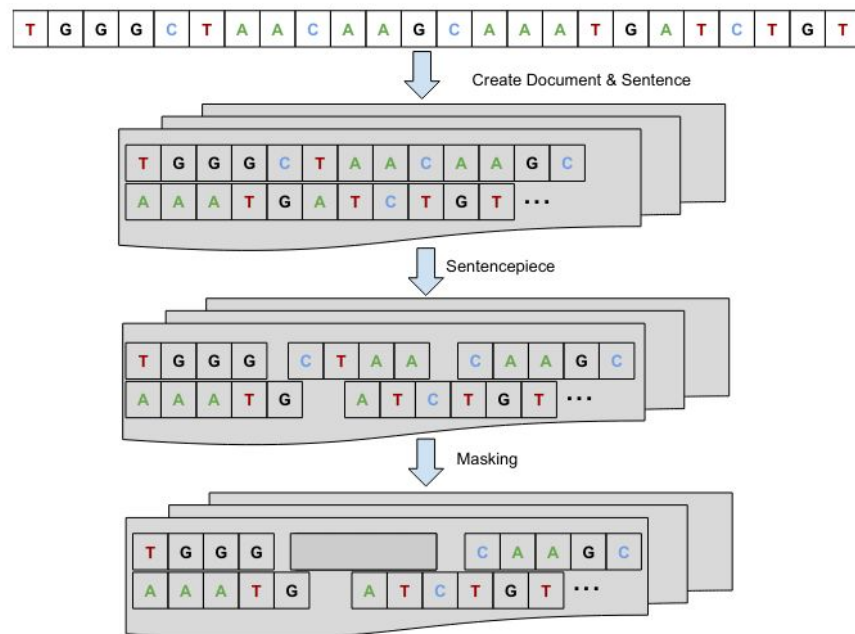- Masking 10% of tokens

Figure 7: Visual description of how the masked language modeling data was generated from raw DNA dataset. The raw DNA sequences of GRCh37, where split at random positions to create documents with 50-100 sentences where each sentence was 500-1000 base pairs (bps). Thus each document had a continuous strand of 25000-100,000 bps of DNA. This process was repeated 10 times to create 10 sets of document for each chromosome of GRCH37. The resulting set of documents was then passed through Sentencepiece that created tokens of average 8bp. For pretraining we used masked language model and masked 10% of the tokens and trained on predicting the masked tokens.

Gene expression prediction using transformers

- Input: one-hot vectors
  - A= [1, 0, 0, 0]
  - 196,608 base-pairs (bp)
- Convolution layer with pooling reduces it to 1536
- Transformer blocks capture long range interactions
- Cropping layer trims 320 bp on each side
  - Regulatory element occur usually on any one side
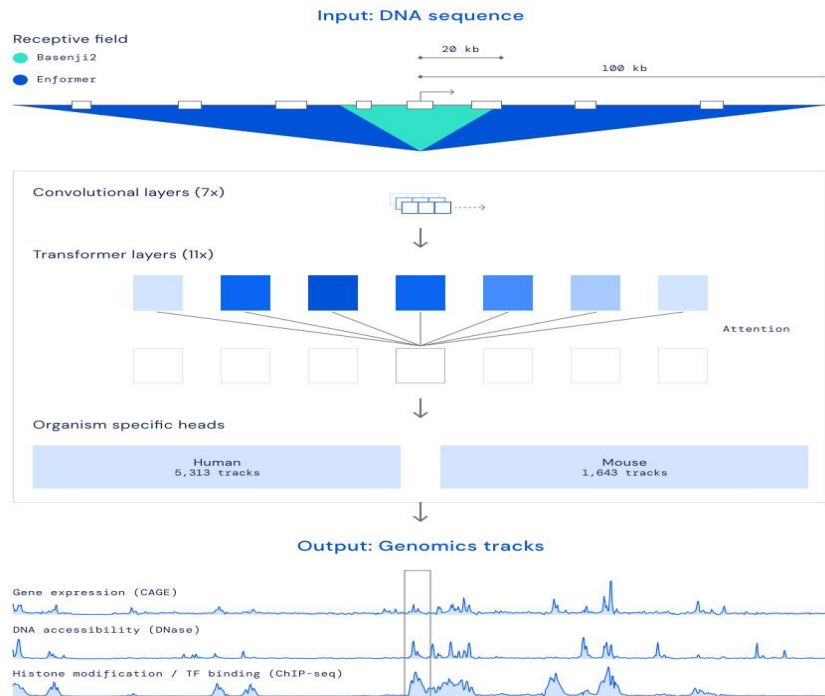
Effective gene expression prediction from sequence by integrating long-range interactions. Nature Methods 2021

Multi-domain Pretraining Analysis

- Out of domain tasks like clinical NER, relation extraction, require additional in-domain pretraining
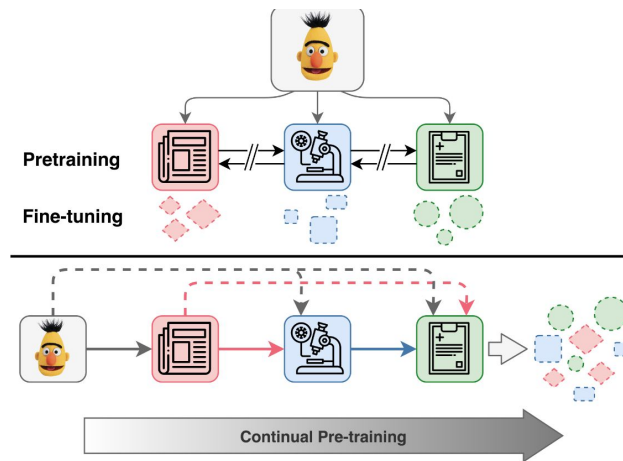- Continual pre-training suffers from catastrophic forgetting (CF)

Figure 1: Traditional approaches (top) train independent domain specific language models (newswire, bio-medical, and clinical) which share no cross domain knowledge. They are further fine-tuned on their respective in-domain tasks. Our approach (bottom) shows how several domains are introduced in sequence, with knowledge retention using mitigation techniques across all domains. Here the final model has the capability to properly fine-tune on any domain specific task.

An Empirical Investigation Towards Efficient Multi-Domain Language Model Pretraining (EMNLP 2020)

Multi-domain Pretraining Analysis

- Explore CF Mitigation strategies
  - Elastic Weight Consolidation
  - Learning rate control: Learning rate is decayed as we move towards input layer
  - Experience replay: reuse data across multiple domains

Model	CoLA	SST-2	MRPC	STS-B	QQP	MNLI	QNLI	RTE	WNLI
BERT$_{BASE}$	57.82	92.09	86.74	88.13	87.49	84.01	90.79	64.98	53.52
BioBERT	37.78	89.68	88.44	87.40	86.96	83.19	89.79	60.29	28.17
Delta	20.04	2.41	-1.69	0.73	0.53	0.82	1.01	4.69	25.35

Table 1: Performance drop of BioBERT after further pre-training on Pubmed articles. The last row shows a positive value indicating the degree to which performance has dropped, and a negative value when it has increased.

An Empirical Investigation Towards Efficient Multi-Domain Language Model Pretraining (EMNLP 2020)
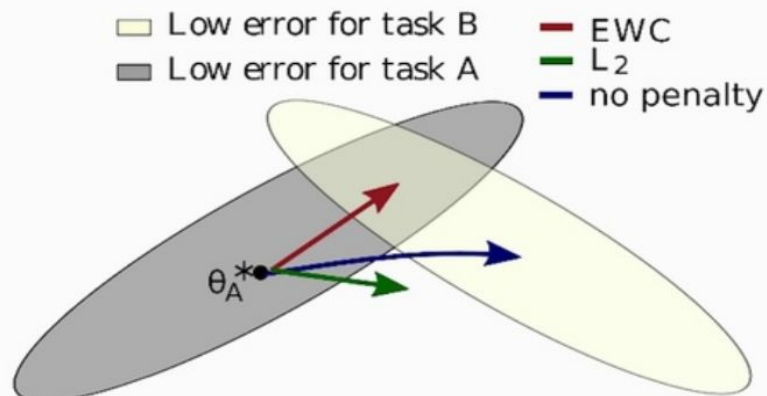
Elastic Weight Consolidation

EWC acts as a regularizer, where different parameter changes weighed differently

- EWC (Kirkpatrick et al. 2017) makes use of a Bayesian factorization of model representation.

$$F_{i,i} = \frac{1}{N} \sum_{j=1}^{N} \left(\frac{\partial \mathcal{L}_A^{(j)}}{\partial \theta_i} \right)^2$$

- This isolates the posterior of a learned task (A) while maintaining the objective of a current task (B).

$$\mathcal{L}(\theta) = \mathcal{L}_B(\theta) + \sum_i \lambda F_{i,i} (\theta_i - \theta_{A,i}^*)^2$$

Low error for task B
Low error for task A

EWC
L₂
no penalty

θ_A^*

Frozen parameters from task A

Figure from Kirkpatrick et al. (2017)

Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., ... & Hassabis, D. (2017). Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13), 3521-3526.

Data Selection Strategies

- Large redundancy exists in training data used for LM pretraining
- Data Selection step: Identify salient data points
  - Model gradient
  - Latent representation
- Uniform sampling works best

Learning Methods

- Zero Shot Based
- Supervised

Zero Shot Based Approach

❏   <u>Medical entity retrieval</u> is an integral component for understanding medical conditions and treatment.

❏   <u>Medical knowledge graphs</u> (KG) contains rich semantics including large numbers of synonyms as well as its curated graphical structures.

❏   Current approaches tend to work well on specific medical domains but generalize poorly to unseen sub-specialties.

❏   Zero-shot entity retrieval model is useful without requiring any human annotation.

Zero-shot Medical Entity Retrieval without Annotation: Learning From Rich Knowledge Graph Semantics. ACL 2021.

Zero Shot Based Approach

Entity retrieval
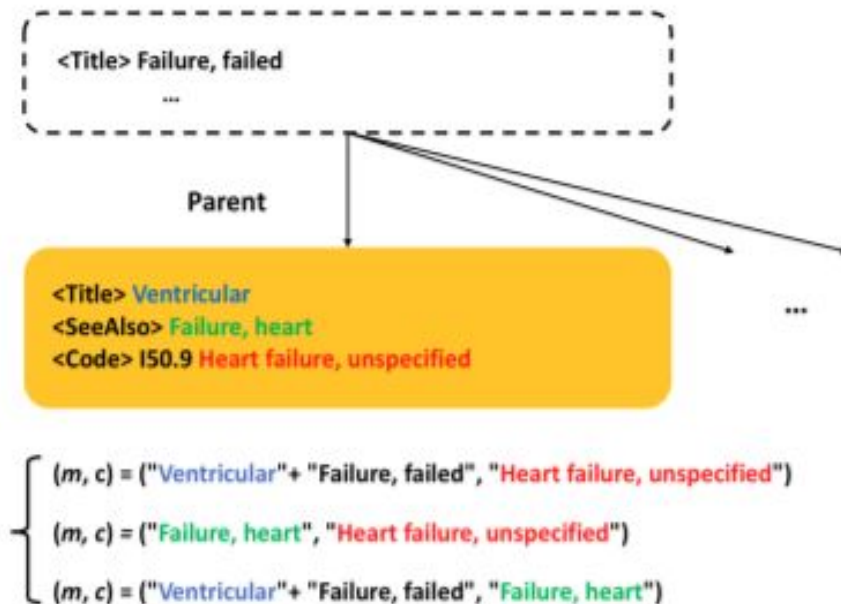
Given a mention m and a concept c $\in$ KG = {c1, c2, ..., cn}, the goal is to learn a similarity measurement S(m, c) , so that the most relevant concept is assigned the highest score. A concept is also referred to as a node in a KG.

Zero-shot entity retrieval

1) zero-shot on <u>mentions only</u>, which assumes unseen mentions but allows seen concepts at test time.

2) zero-shot on <u>mentions and concepts</u>, which assumes both to be unseen at test time.

Zero-shot Medical Entity Retrieval without Annotation: Learning From Rich Knowledge Graph Semantics. ACL 2021.

Zero Shot Based Approach

- Learning tasks by constructing mention-concept pairs (m, c).

- Design learning tasks by finding very similar or closely related textual descriptions and use them to construct (m, c) pairs.
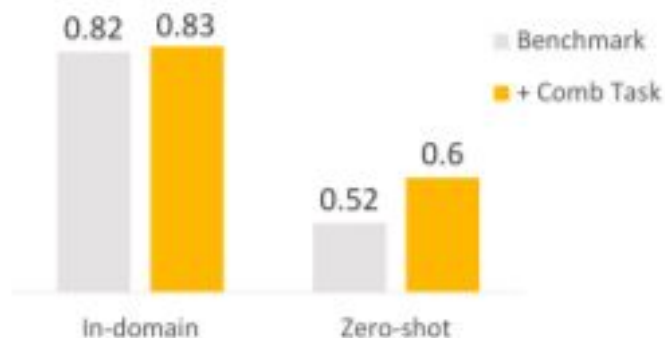

Zero-shot Medical Entity Retrieval without Annotation: Learning From Rich Knowledge Graph Semantics. ACL 2021.

Zero Shot Based Approach: KGs

1. **ICD-10:** The 10th version of the International Statistical Classification of Diseases, Clinical Modification (ICD-10). It contains over 69K concepts, organized in a tree structure of parent-child relationships.
2. **SNOMED:** Systematized Nomenclature of Medicine – Clinical Terms (SNOMED) is a standardized clinical terminology used for the electronic exchange of clinical health information with over 360K active concepts.
3. **UMLs:** The Unified Medical Language System (UMLS) is a compendium of a large number of curated biomedical vocabularies with over 1MM concepts.

Zero-shot Medical Entity Retrieval without Annotation: Learning From Rich Knowledge Graph Semantics. ACL 2021.

Zero Shot Based Approach: Datasets

1. **MedMention (Mohan and Li, 2019)** is a publicly available corpus of 4,392 PubMed1 abstracts with biomedical entities annotated with UMLS concepts.
2. **COMETA (Basaldella et al., 2020)** is one of the largest public corpora of social media data with SNOMED annotations. It provides four train, dev, test splits: Stratified-General (SG), StratifiedSpecific (SS), Zeroshot-General (ZG), ZeroshotSpecific (ZS).
3. **3DNotes(Zhu et al., 2020)** It has two sets of annotations: one with ICD-10 (ICD split), another with SNOMED (SN split). The annotation follows the i2b2 challenge (Uzuner et al., 2011) guidelines.

Zero-shot Medical Entity Retrieval without Annotation: Learning From Rich Knowledge Graph Semantics. ACL 2021.

Zero Shot Based Approach: Performance

Zero-shot Medical Entity Retrieval without Annotation: Learning From Rich Knowledge Graph Semantics. ACL 2021.

Supervised Approach for MIE

- Electronic Medical Records (EMRs) have become key components of modern medical care systems.
- Many doctors suffer from writing them, which is time-consuming and tedious.
- Automatically converting doctor-patient dialogues into EMRs can effectively remove the heavy burdens of doctors, making them more deliberate to communicate with their patients.
- Easy approach - Supervised data, i.e., dialogue-EMR pairs are needed. Data is hard to acquire in medical domain due to the privacy policy.

MIE: A Medical Information Extractor towards Medical Dialogues. ACL 2020

Supervised Approach for MIE

- Propose a Medical Information Extractor (MIE) towards medical dialogues.
- MIE is able to extract mentioned symptoms, surgeries, tests, other information and their corresponding status.
- MIE uses a deep matching architecture, taking dialogue turn-interaction into account.

MIE: A Medical Information Extractor towards Medical Dialogues. ACL 2020

Supervised Approach for MIE

Dialogue Window	Annotated Labels
Patient: Doctor, could you please tell me is it premature beat?	Symptom: Premature beat (doctor-pos)
Doctor: Yes, considering your Electrocardiogram. Do you feel palpitation or short of breath?	Test: Electrocardiogram (patient-pos)
	Symptom: Cardiopalmus (patient-neg)
Patient: No. Can I do radiofrequency ablation?	Symptom: Dyspnea (patient-neg)
Doctor: It is worth considering. Any discomfort in chest?	Surgery: Radiofrequency ablation (doctor-pos)
Patient: I always have bouts of pain.	Symptom: Chest pain (patient-pos)

MIE: A Medical Information Extractor towards Medical Dialogues. ACL 2020

Supervised Approach for MIE : Datasets

- Doctor-patient dialogues from a Chinese medical consultation website, Chunyu-Doctor. The dialogues are already in text format.
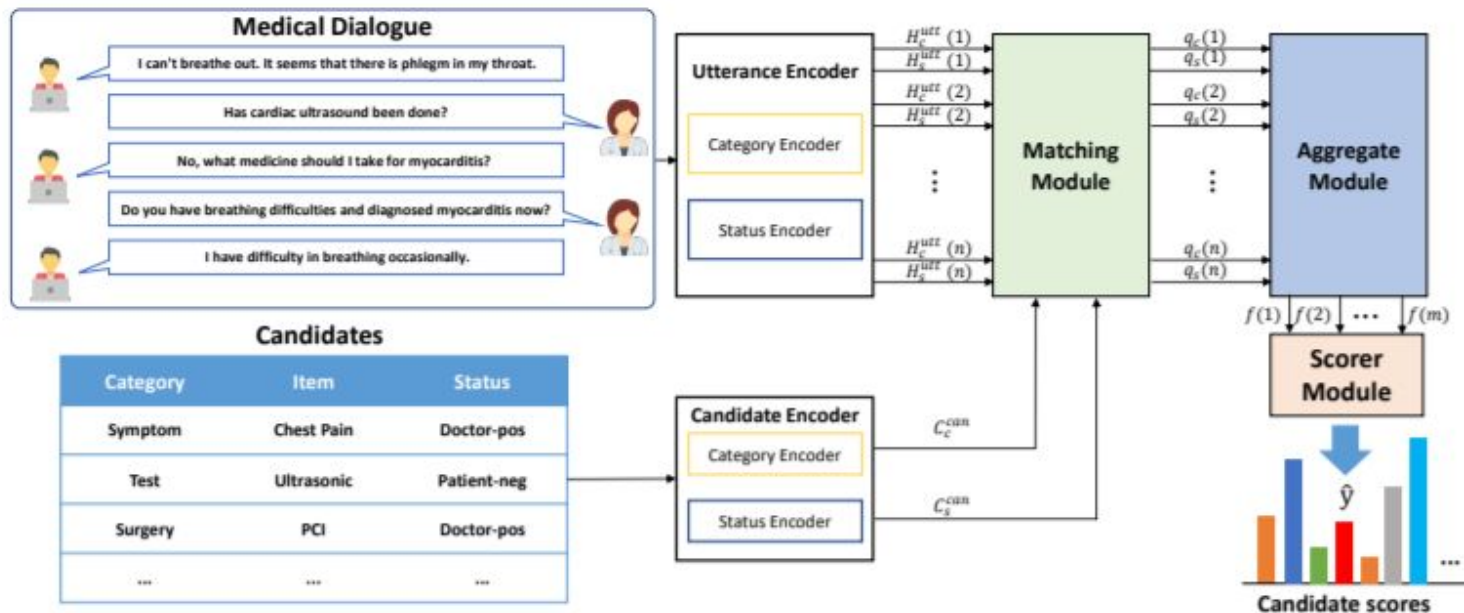- In all, authors annotate 1,120 dialogues.

	Dialogue	Window	Symptom	Surgery	Test	Other info
Train	800	12931	21420	839	8879	1363
Dev	160	2587	4254	119	1680	259
Test	160	2694	4878	264	1869	327
Total	1120	18212	30552	1222	12428	1949

	Patient-pos	Patient-neg	Doctor-pos	Doctor-neg	Unknown
Symptom	15119	1782	1655	910	11086
Surgery	169	48	698	10	297
Test	5589	303	4443	44	2049
Other info	550	1399	-	-	1505

MIE: A Medical Information Extractor towards Medical Dialogues. ACL 2020

Supervised Approach for MIE : Datasets

Category	Item	Status
Symptom	Backache Perspiration Hiccups Nausea Cyanosis Fever Fatigue Abdominal discomfort ...	patient-positive (appear) patient-negative (absent) doctor-positive (diagnosed) doctor-negative (exclude) unknown
Surgery	Interventional treatment Radiofrequency ablation Heart bypass surgery Stent implantation	patient-positive (done) patient-negative (not done) doctor-positive(suggest) doctor-negative (deprecated) unknown
Test	B-mode ultrasonography CT examination CT angiography CDFI Blood pressure measure-ment Ultrasonography MRI Thyroid function test Treadmill test ...	patient-positive(done) patient-negative (not done) doctor-positive(suggest) doctor-negative (deprecated) unknown
Other info	Sleep Diet Mental condition Defecation Smoking Drinking	patient-positive (normal) patient-negative (abnormal) unknown

MIE: A Medical Information Extractor towards Medical Dialogues. ACL 2020

Supervised Approach for MIE : Model

MIE: A Medical Information Extractor towards Medical Dialogues. ACL 2020

Supervised Approach for MIE : Performance

Model	Window-level									Dialogue-level								
	Category			Item			Full			Category			Item			Full		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Plain-Classifier	67.21	63.78	64.92	60.89	49.20	53.81	53.13	49.46	50.69	93.57	89.49	90.96	83.42	73.76	77.29	61.34	52.65	56.08
MIE-Classifier-single	80.51	76.39	77.53	76.58	64.63	68.30	68.20	61.60	62.87	97.14	91.82	93.23	91.77	75.36	80.96	71.87	56.67	61.78
MIE-Classifier-multi	**80.72**	**77.76**	**78.33**	76.84	**68.07**	**70.35**	67.87	64.71	64.57	96.61	**92.86**	**93.45**	90.68	82.41	84.65	68.86	62.50	63.99
MIE-single	78.62	73.55	74.92	76.67	65.51	68.88	69.40	64.47	65.18	96.93	90.16	92.01	94.27	79.81	84.72	75.37	63.17	67.27
MIE-multi	80.42	76.23	77.77	**77.21**	66.04	69.75	**70.24**	**64.96**	**66.40**	**98.86**	91.52	92.69	**95.31**	**82.53**	**86.83**	76.83	64.07	69.28

MIE: A Medical Information Extractor towards Medical Dialogues. ACL 2020

Supervised Approach for MIE : Prediction

Patient: I have atrial fibrillation, heart failure, anemia and loss my appetite.
Doctor: Hello! How long did them last? Did you examine blood routine?
Patient: Yes.
Doctor: Is there coronary heart disease?
Patient: No.

(a)

Patient: I have atrial fibrillation, heart failure, anemia and loss my appetite.
Doctor: Hello! How long did them last? Did you examine blood routine?
Patient: Yes.
Doctor: Is there coronary heart disease?
Patient: No.

(b)

Patient: I have atrial fibrillation, heart failure, anemia and loss my appetite.
Doctor: Hello! How long did them last? Did you examine blood routine?
Patient: Yes.
Doctor: Is there coronary heart disease?
Patient: No.

(c)

MIE: A Medical Information Extractor towards Medical Dialogues. ACL 2020

Supervised Approach for MIE : Prediction

MIE: A Medical Information Extractor towards Medical Dialogues. ACL 2020

Clinical Embeddings

BERT

Bio-BERT

 1) Clinical BERT, initialized from BERT-Base (Alsentzer et al. 2019)

2) Clinical BioBERT, initialized from Bio-BERT (Alsentzer et al. 2019)

1) MedNLI natural language inference task (Romanov and Shivade, 2018)
2) Four i2b2 named entity recognition (NER) tasks, all in IOB format (Ramshaw and Marcus, 1995) -
   a) i2b2 2006 (Uzuner et al., 2007)
   b) i2b2 2010 (Uzuner et al., 2011)
   c) i2b2 2012 (Sun et al., 2013a,b)
   d) i2b2 2014 (Stubbs et al., 2015)

Training : Entire embedding model procedure took roughly 17 - 18 days of computational runtime using a single GeForce GTX TITAN X 12 GB GPU (and significant CPU power and memory for pre-processing tasks)

Publicly Available Clinical BERT Embeddings. MIT-Harvard-Microsoft Research Collaboration. Clinical NLP Workshop @ ACL 2019

Clinical Embeddings

Model	MedNLI	i2b2 2006	i2b2 2010	i2b2 2012	i2b2 2014
BERT	77.6%	93.9	83.5	75.9	92.8
BioBERT	80.8%	**94.8**	86.5	78.9	**93.0**
Clinical BERT	80.8%	91.5	86.4	78.5	92.6
Discharge Summary BERT	80.6%	91.9	86.4	78.4	92.8
Bio+Clinical BERT	**82.7%**	94.7	87.2	**78.9**	92.5
Bio+Discharge Summary BERT	**82.7%**	94.8	**87.8**	78.9	92.7

Accuracy (MedNLI) and Exact F1 score (i2b2) across various clinical NLP tasks.

Model		Disease			Operations			Generic	
	Glucose	Seizure	Pneumonia	Transfer	Admitted	Discharge	Beach	Newspaper	Table
BioBERT	insulin exhaustion dioxide	episode appetite attack	vaccine infection plague	drainage division transplant	admission sinking hospital	admission wave sight	coast rock reef	news official industry	tables row dinner
Clinical	potassium sodium sugar	headache stroke agitation	consolidation tuberculosis infection	transferred admitted arrival	admission transferred admit	disposition transfer transferred	shore ocean land	publication organization publicity	scenario compilation technology

Nearest neighbors for 3 sentinel words for each of 3 categories

Publicly Available Clinical BERT Embeddings. MIT-Harvard-Microsoft Research Collaboration. Clinical NLP Workshop @ ACL 2019

Medical QnA Understanding

Medical QnA Understanding

❏ Users of medical question answering systems often submit long and detailed questions.

❏ There is a growing number of approaches to medical question understanding:

  ❏ Query relaxation (Ben Abacha and Zweigenbaum, 2015; Lei et al., 2020)
  ❏ Question entailment (Ben Abacha and Demner-Fushman, 2016, 2019b; Agrawal et al., 2019)
  ❏ Question summarization (Ben Abacha and Demner-Fushman, 2019a)
  ❏ Question similarity (Ben Abacha and Demner-Fushman, 2017; Yan and Li, 2018; McCreery et al., 2019).

A Gradually Soft Multi-Task and Data-Augmented Approach to Medical Question Understanding. ACL 2021

Medical QnA Understanding

- ❏   2 Tasks -
  - ❏   Multi-task learning of medical question summarization
  - ❏   Medical RQE.
- ❏ The input to both tasks is a pair of medical questions.
- ❏ The first question is called a Consumer Health Question (CHQ), and the second question is called a Frequently Asked Question (FAQ).
- ❏ The CHQ is written by a patient and is usually longer and more informal, whereas the FAQ is usually a single sentence question written by a medical expert.
- ❏ The purpose of both tasks is to match a CHQ to an FAQ, and ultimately to an expert-written answer that matches the FAQ.

Medical QnA Understanding

❏ Establish an equivalence between the tasks of question summarization and Recognizing Question Entailment (RQE) using their definitions in the medical domain.

**Source User-written Question or *Consumer Health Question (CHQ)*:**
SUBJECT: Morgellon Disease. MESSAGE: It appears as if I have had this horrible disease for many, many years and it is getting worst. I am trying to find a physician or specialist in the South Carolina area who can treat me for this medical/mental disease. It seems as if this disease has "NO" complete treatment and it is more least a disability!

Reference Summarized Question or *Frequently Asked Question (FAQ)*:
What are the treatments for Morgellon Disease, and how can I find physician(s) in South Carolina who specialize in it?

A Gradually Soft Multi-Task and Data-Augmented Approach to Medical Question Understanding. ACL 2021

Medical QnA Understanding

❏ use equivalence observation to propose a scheme for data augmentation.
❏ Simultaneous multi-task learning model architecture and learning objective.
❏ Finally, soft parameter-sharing scheme.

A Gradually Soft Multi-Task and Data-Augmented Approach to Medical Question Understanding. ACL 2021
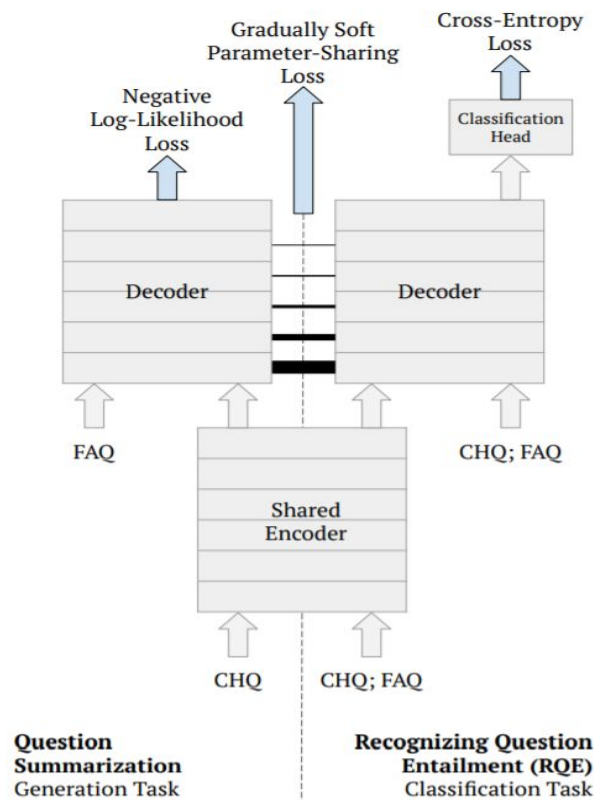
Medical QnA Understanding : Error Model

Given a CHQ embedding x, the corresponding FAQ embedding y, and the entailment label $l_{entail} \in \{0, 1\}$, we optimize the following multi-task learning loss function where BCE is binary cross entropy, and λ is a hyperparameter between 0 and 1.

$$\mathcal{L}_{\mathrm{MTL}}(\theta) = -\lambda * \log p(\mathbf{y}|\mathbf{x}; \theta)$$
$$+ (1 - \lambda) * \mathrm{BCE}\left([\mathbf{x}; \mathbf{y}], l_{entail}; \theta\right)$$

Medical QnA Understanding : Datasets

DATASET	TRAIN	DEV	TEST
MeQSum	400	100	500
HealthCareMagic	181,122	22,641	22,642
iCliniq	24,851	3,105	3,106
MEDIQA RQE	8,588	302	230

A Gradually Soft Multi-Task and Data-Augmented Approach to Medical Question Understanding. ACL 2021

Medical QnA Understanding : Performance
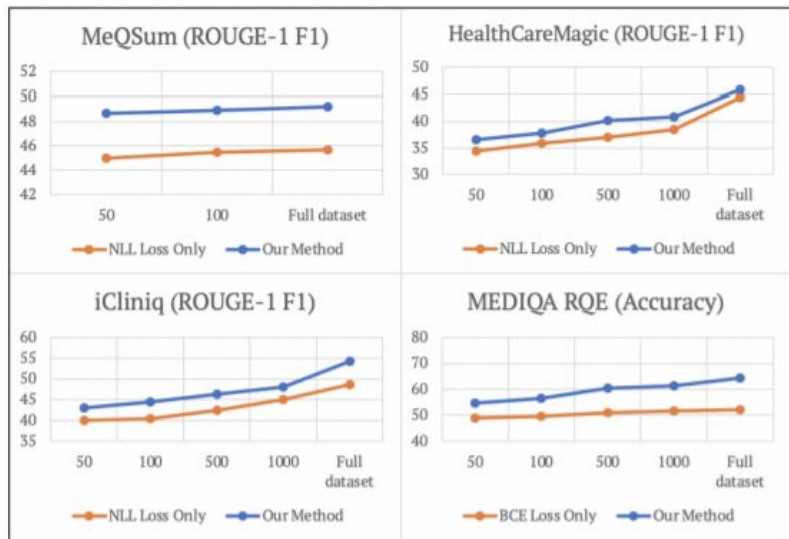
Figure 4: Test set 4-run average performance of our method compared to single-task BART in low-resource settings. Full dataset results are shown for comparison.

A Gradually Soft Multi-Task and Data-Augmented Approach to Medical Question Understanding. ACL 2021

Medical QnA Understanding : Output

Source User-written Question or *Consumer Health Question (CHQ)*:
SUBJECT: Morgellon Disease. MESSAGE: It appears as if I have had this horrible disease for many, many years and it is getting worst. I am trying to find a physician or specialist in the South Carolina area who can treat me for this medical/mental disease. It seems as if this disease has "NO" complete treatment and it is more least a disability!

Reference Summarized Question or *Frequently Asked Question (FAQ)*:
What are the treatments for Morgellon Disease, and how can I find physician(s) in South Carolina who specialize in it?

BART Trained on Summarization Loss Only (Baseline):
Where can I find physician(s) who specialize in morgellon disease?

Our Gradually Soft Multi-Task and Data-Augmented Model:
Where can I find a physician or specialist in South Carolina who can treat Morgellon Disease?

A Gradually Soft Multi-Task and Data-Augmented Approach to Medical Question Understanding. ACL 2021

Publicly Available Datasets

Publicly Available Datasets

Build large-scale medical dialog datasets – MedDialog.

    1) a Chinese dataset with 3.4 million conversations between patients and doctors, 11.3 million utterances, 660.2 million tokens, covering 172 specialties of diseases

    2) an English dataset with 0.26 million conversations, 0.51 million utterances, 44.53 million tokens, covering 96 specialties of diseases.

Each consultation consists of two parts:

    (1) description of patient's medical conditions;

    (2) conversation between patient and doctor.

The data is crawled from iclinic.com3 and healthcaremagic.com4 , which are two online platforms of healthcare services, including symptom self-checker, video consultation, online chat with doctors etc.

MedDialog: Large-scale Medical Dialogue Datasets. Zeng et al. , EMNLP 2020.

Publicly Available Datasets

Conversation history: 所谓的湿疹, 孩子从一周岁后起, 到换季的时候, 就周身发痒, 用手一抓破, 就一直是红色的, 水水的, 要过几长时间才会结痂, 留有黑色印迹. (Eczema. Since one year old, whenever changing seasons, the baby feels itchy all over the body. The pimples break when scratched. They are always red and swelling. It takes a long time to scab, leaving a black mark.)

Groundtruth response: 建议到江苏省皮肤病研究所就诊. (I suggest to visit the Dermatology Research Institute of Jiangsu Province.)

Transformer: 需要看到皮肤的形态才能给予具体建议. (I need to see the skin to give detailed suggestions.)

GPT: 如果是过敏性皮肤, 建议外用酮康唑乳膏. (If the skin is allergic, can take some Ketoconazole cream.)

BERT-GPT: 你去买一种依巴斯汀一天晚上吃一片试试看. (Can use Ebastine, one tablet every day at bedtime.)

MedDialog: Large-scale Medical Dialogue Datasets. Zeng et al. , EMNLP 2020.

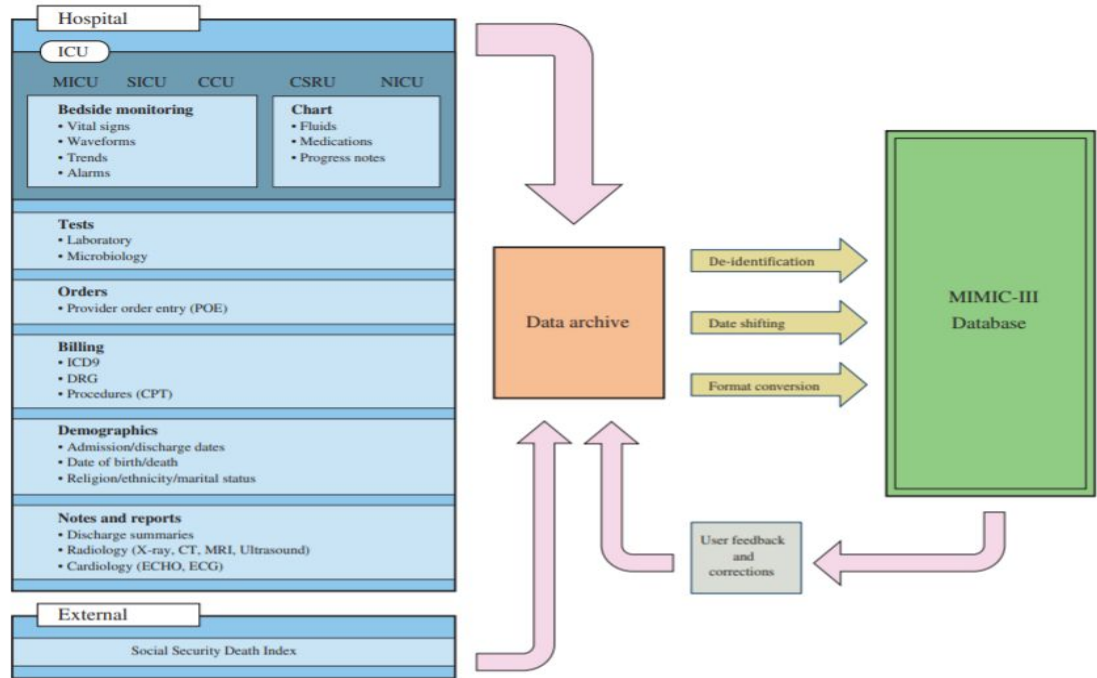# Publicly Available Datasets

Conversation history: (My voice is broken. In the past few years, it happens oftentimes.)
Groundtruth response: (Surgery choices: outpatient, surface anesthetic, remove small polyps using electrolaryngoscope; inpatient, general anesthesia, remove large polyps using laryngoscope and microscope.)
Transformer: (I suggest to take a laryngoscope test at a local Children's hospital. Then take a laryngoscope test at a better hospital, to confirm and select the best way for treatment.)
GPT: (Hello. Broken voice can be caused by many things. Is your problem at the throat or pharynx? If there is no abnormality at the throat, you can take a laryngoscope test. If abnormal, you should take a laryngoscope test.
BERT-GPT: (Hello. I suggest to take a laryngoscope test at a local hospital.)

MedDialog: Large-scale Medical Dialogue Datasets. Zeng et al. , EMNLP 2020.

Publicly Available Datasets

1. MIMIC-III, a freely accessible critical care database.

   Johnson et al. 2015

   Scientific data Journal (Nature)

Summary of Contributions

- Strategies to utilize or transfer knowledge from generic domain, helps improve performance in clinical data applications
- Transformers with wider receptive field or seeing longer sequence at once, helps identify long range relationships
- Proposed learning frameworks can mine rich semantics from a medical KG.

Future Research Directions

- Use of labeled data of a single hospital department to improve the MDG performance on other departments without labeled data by transfer learning or zero-shot learning
- Develop Universal Health Care Knowledge Graph
- Extend the MIE work in multi-lingual scenario
- Data augmentations used in SSL approaches may lead to biologically invalid samples → interpretable-by-design or human-in-the-loop
- Incorporating domain knowledge in self-supervised setting
  - Additional pretraining objective
  - Use domain tools like KGs, interaction networks as part of training objective

Thank you for listening