MT-Clinical BERT: Scaling Clinical Information Extraction with Multitask Learning

arXiv.org

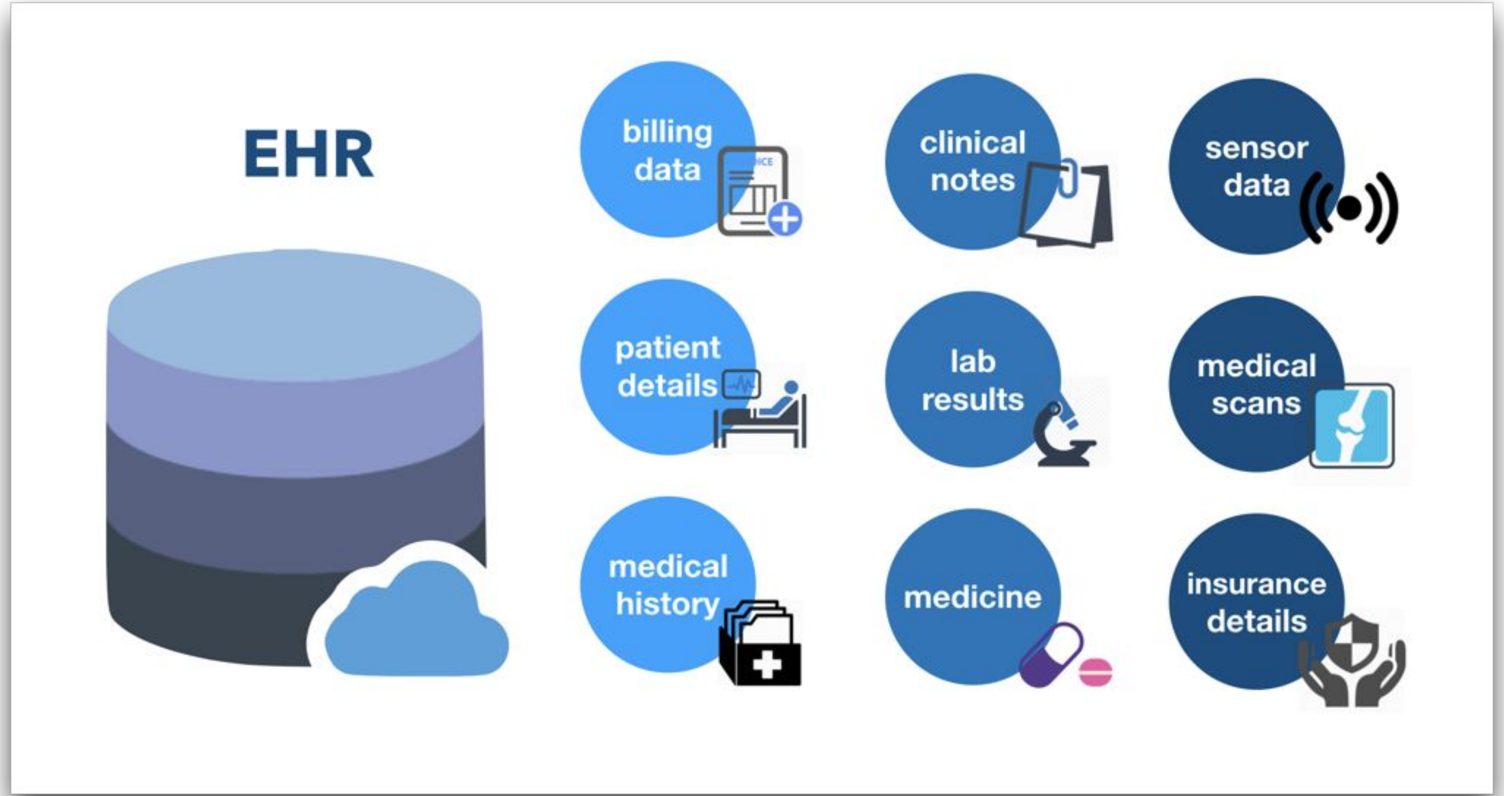
GitHub

VCU

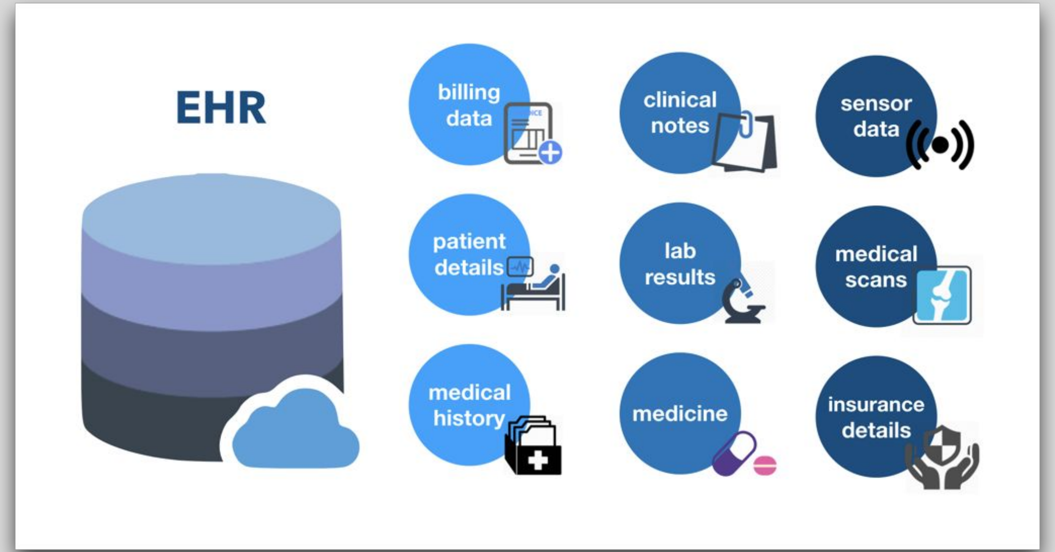
Andriy Mulyar and Bridget T. McInnes
@andriy_mulyar @btmcinnes

NLP@VCU
Virginia Commonwealth University
In Submission, AMIA 2020.

We train a **single** neural network to simultaneously perform **multiple** clinical note information extraction tasks at near state-of-the-art performance.



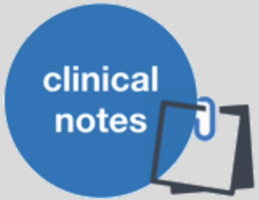
Source: https://goku.me/blog/deep-learning-with-ehr-systems



Can we model / predict:

- Patient Readmission
- Patient Mortality
- Reasons for discharge

EHR



Consultation Notes

Discharge Summaries

Historical Note

Procedure Notes

Progress Note

Admission Date: [**2115-2-22**] Discharge Date: [**2115-3-19**]

Date of Birth: [**2078-8-9**] Sex: M

Service: MEDICINE

Allergies:
Vicodin

Attending:[**First Name3 (LF) 4891**]
Chief Complaint:
Post-cardiac arrest, asthma exacerbation

Major Surgical or Invasive Procedure:
Intubation
Removal of chest tubes placed at an outside hospital
R CVL placement

History of Present Illness:
Mr. [**Known lastname 3234**] is a 36 year old gentleman with a PMH signifciant with dilated cardiomyopathy s/p AICD, asthma, and HTN admitted to an OSH with dyspnea now admitted to the MICU after PEA arrest x2. The patient initially presented to LGH ED with hypoxemic respiratory distress. While at the OSH, he received CTX, azithromycin, SC epinephrine, and solumedrol. While at the OSH, he became confused and subsequently had an episode of PEA arrest and was intubated. He received epinephrine, atropine, magnesium, and bicarb. In addition, he had bilateral needle thoracostomies with report of air return on the left, and he subsequently had bilateral chest tubes placed. After approximately 15-20 minutes of rescucitation, he had ROSC. He received vecuronium and was started on an epi gtt for asthma and a cooling protocol, and was then transferred to [**Hospital1 18**] for further evaluation. Of note, the patient was admitted to LGH in [**1-4**] for dyspnea, and was subsequently diagnosed with a CAP and asthma treated with CTX and azithromycin. Per his family, he has also had multiple admissions this winter for asthma exacerbations.

Can we identify and extract:

- Problems, Treatments and Tests.
- Drugs and drug-related information.
- Protected Health Information.

Admission Date: [**2115-2-22**] Discharge Date: [**2115-3-19**]

Date of Birth: [**2078-8-9**] Sex: M

Service: MEDICINE

Allergies:
Vicodin

Attending:[**First Name3 (LF) 4891**]
Chief Complaint:
Post-cardiac arrest, asthma exacerbation

Major Surgical or Invasive Procedure:
Intubation
Removal of chest tubes placed at an outside hospital
R CVL placement

History of Present Illness:
Mr. [**Known lastname 3234**] is a 36 year old gentleman with a PMH signifciant with dilated cardiomyopathy s/p AICD, asthma, and HTN admitted to an OSH with dyspnea now admitted to the MICU after PEA arrest x2. The patient initially presented to LGH ED with hypoxemic respiratory distress. While at the OSH, he received CTX, azithromycin, SC epinephrine, and solumedrol. While at the OSH, he became confused and subsequently had an episode of PEA arrest and was intubated. He received epinephrine, atropine, magnesium, and bicarb. In addition, he had bilateral needle thoracostomies with report of air return on the left, and he subsequently had bilateral chest tubes placed. After approximately 15-20 minutes of rescucitation, he had ROSC. He received vecuronium and was started on an epi gtt for asthma and a cooling protocol, and was then transferred to [**Hospital1 18**] for further evaluation. Of note, the patient was admitted to LGH in [**1-4**] for dyspnea, and was subsequently diagnosed with a CAP and asthma treated with CTX and azithromycin. Per his family, he has also had multiple admissions this winter for asthma exacerbations.

Can we model clinical language understanding:

- Logical entailment.
- Semantic similarity.
- Question answering.

Our work introduces and validates a method that can train a **single** neural network to perform **all** of these tasks.

Leverage numerous existing datasets.

Encompassing information extraction and language understanding tasks.

Spanning multiple languages.

Stratifying multiple clinical note sources.

Task	Dataset	Metric	Description	# Train Inst.	# Test Inst.
STS	n2c2-2019[10]	Pearson Rho	Sentence Pair Semantic Similarity	1,641	410
Entailment	MedNLI[11]	Accuracy	Sentence Pair Entailment	12,627	1,422
	MedRQE[12]	Accuracy	Sentence Pair Entailment	8,588	302
NER	n2c2-2018[3]	Micro-F1	Drug and Adverse Drug Event	36,384	23,462
	i2b2-2014[4]	Micro-F1	PHI de-identification	17,310	11,462
	i2b2-2012[2]	Micro-F1	Events	16,468	13,594
	i2b2-2010[1]	Micro-F1	Problems, Treatments and Tests	27,837	45,009
	quaero-2014[13]	Micro-F1	UMLS Semantic Groups (French)	2,695	2,260

History of Present Illness:

Mr. [**Known lastname 3234**] is a 36 year old gentleman with a PMH signifciant with dilated PROBLEM cardiomyopathy s/p AICD, asthma, and HTN admitted to an OSH with dyspnea now admitted to the MICU after PEA PROBLEM arrest x2. The patient initially presented to LGH ED with hypoxemic respiratory PROBLEM distress. While at the OSH, he azithromycin TREATMENT , SC epinephrine TREATMENT , and solumedrol TREATMENT . While at the OSH, he became confused and subsequently had an episode of PEA PROBLEM arrest and was intubated. He received epinephrine TREATMENT , atropine TEST , magnesium, and bicarb TEST . with report of air return on the left, and he subsequently had bilateral chest tubes placed. After approximately 15-20 minutes of rescucitation, he had ROSC. He received vecuronium TREATMENT and was started on an epi gtt TREATMENT for asthma PROBLEM and a cooling protocol TREATMENT , and was then transferred to [**Hospital1 18**] for further evaluation. Of note, the patient was admitted to LGH in [**1-4**] for dyspnea PROBLEM , and was subsequently diagnosed with a PROBLEM CAP and asthma treated with CTX and azithromycin TREATMENT . Per his family, he has also had multiple admissions this winter for asthma PROBLEM exacerbations.

Named Entity Recognition

Imposes structure on the note by extracting relevant entities from text.

Standardized as a token classification objective.

Input:
Sequence of tokens.

Output:
Sequence of discrete labels.

NER	n2c2-2018[3]	Micro-F1	Drug and Adverse Drug Event	36,384	23,462
	i2b2-2014[4]	Micro-F1	PHI de-identification	17,310	11,462
	i2b2-2012[2]	Micro-F1	Events	16,468	13,594
	i2b2-2010[1]	Micro-F1	Problems, Treatments and Tests	27,837	45,009
	quaero-2014[13]	Micro-F1	UMLS Semantic Groups (French)	2,695	2,260

Sentence 1

In the ED, initial VS revealed T 98.9, HR 73, BP 121/90, RR 15, O2 sat 98% on RA.

Sentence 2

The patient is hemodynamically stable.

Does Sentence 2 **logically entail** Sentence 1?

Language Entailment

Assesses logical relationships between spans of clinical text.

Standardized as a span classification objective.

Input:

Two sequences of tokens.

Output:

[entails, contradicts, neutral]

Entailment	MedNLI[11]	Accuracy	Sentence Pair Entailment	12,627	1,422
	MedRQE[12]	Accuracy	Sentence Pair Entailment	8,588	302

Sentence 1
In the ED, initial VS revealed T 98.9, HR 73, BP 121/90, RR 15, O2 sat 98% on RA.

Sentence 2
The patient is hemodynamically stable.

How **semantically similar** is Sentence one to Sentence two.

Semantic Similarity

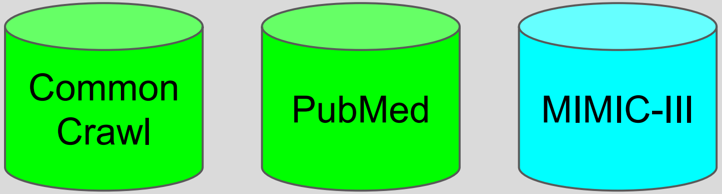
Measures degree of semantic similarity between text spans.

Standardized as a span regression objective.

Input:
Two sequences of tokens.
Output:
[1,5]

STS	n2c2-2019[10]	Pearson Rho	Sentence Pair Semantic Similarity	1,641	410

To achieve a unified model for all tasks, we multitask finetune a clinically pre-trained stack of Transformers.



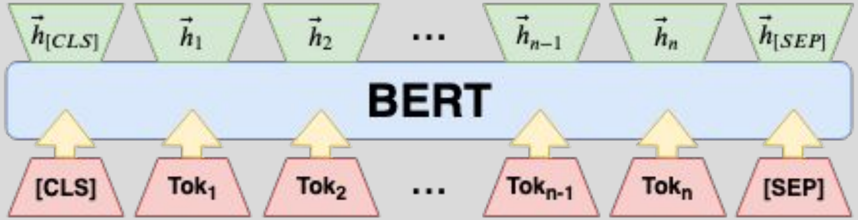
+

Self-supervision

over an

Expressive enough NN

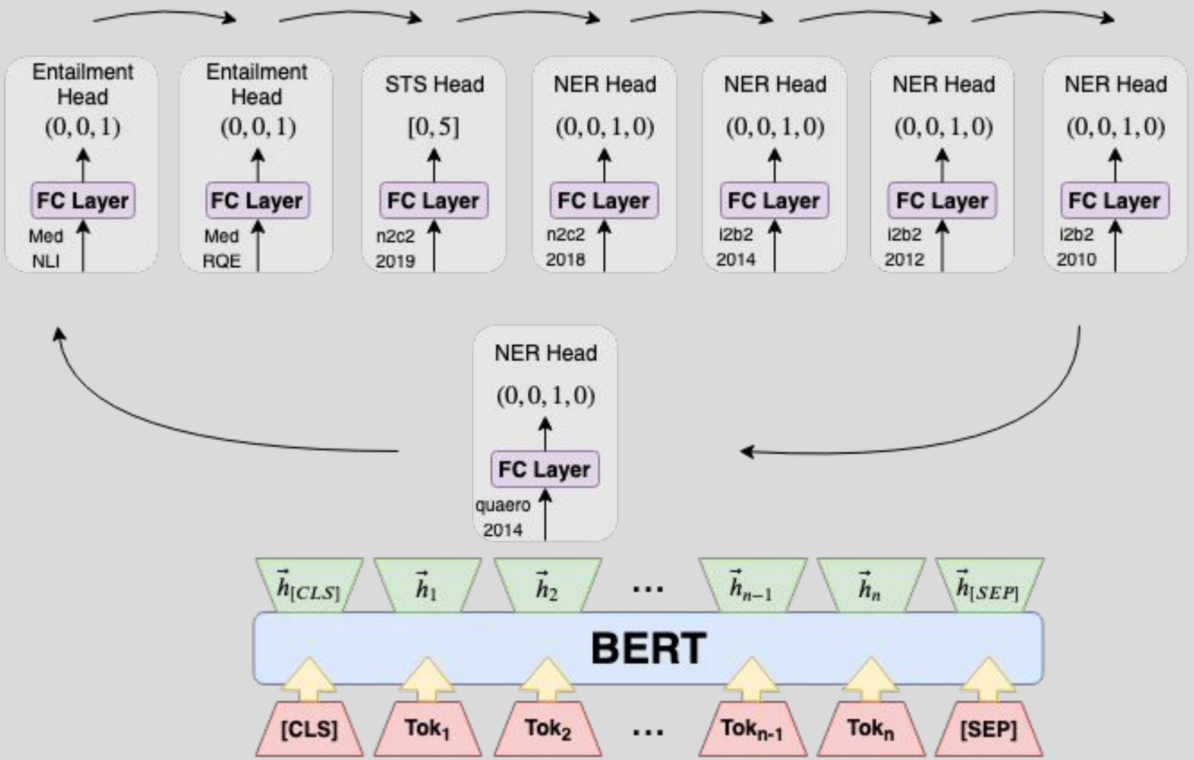
=



BERT

Pre-trained stacks of Transformers such as BERT have enough expressivity to easily adapt to a single language task.

Can we adapt to multiple tasks at once?



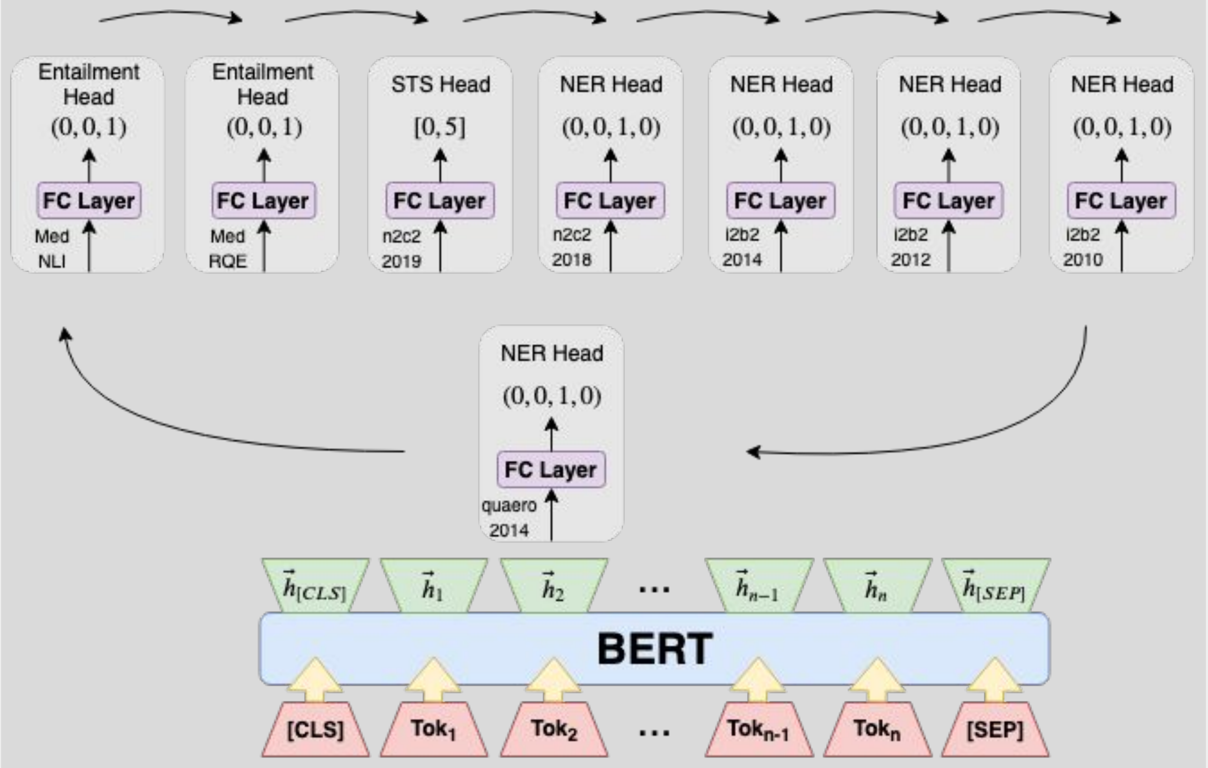
Multitask Learning

Regularize a model by forcing it to adapt against multiple loss signals.

- May help performance
- May learn more robust features.

but

- Often be difficult to train.
- Often difficult to scale.

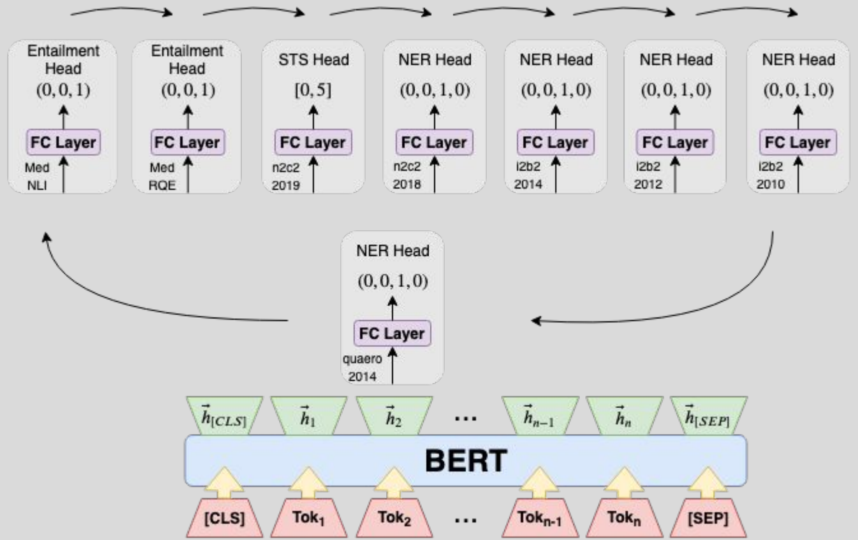


Multitask Learning

But we made it work!

How:

- Clinical tasks differ in objectives or text distribution, but still confine with-in clinical notes.
- Multiple loss objectives prevent traditional multitasking techniques (loss averaging), but task sampling works in practice!



Multitask Learning

Round Robin Task Sampling

Idea:

Interleave **task data sampling** and batched stochastic gradient descent parameter updates.

Algorithm 1 MT-Clinical BERT Training Schedule

Require: θ_E: pre-trained Transformer encoder.
Require: $\theta_H = \{\theta_{h_1}, \ldots, \theta_{h_n}\}$: n task-specific heads.

```
Randomly initialize θ_{h_i} ∀i ∈ {1, ..., n}
while all batches from largest task dataset are not sampled do
    Sample a batch D_i for each θ_{h_i} ∈ θ_H
    for each (θ_{h_i}, D_i) do                          ▷ One round robin iteration
        Let θ = θ_E ∘ θ_{h_i}                            ▷ Outputs of encoder into head θ_h
        θ' = θ − α∇_θ L(θ, D_i)
        Update θ with θ'
    end for
end while
```

This training schedule **successfully** learns a universal clinical text feature encoder.

MT-Clinical BERT

A **single** pre-trained clinical model that performs competitively for **multiple tasks.**

- Previously we had task specific models, now we only need one!
- Sharing representations amongst objectives shows potential performance gains.
- We can cross languages, task types and note sources.

	MT-Clinical BERT	Optimized Clinical BERT	Clinical Bert[8]
n2c2-2019	86.7 (−0.5)	87.2	-
MedNLI	80.5 (−2.3)	82.8	82.7
MedRQE	76.5 (−3.6)	80.1	-
n2c2-2018	87.4 (−0.7)	88.1	-
i2b2-2014	91.9 (−3.6)	95.5	92.7
i2b2-2012	**84.1** (+0.2)	83.9	78.9
i2b2-2010	89.5 (−0.3)	89.8	87.8
quaero-2014	49.1 (−6.4)	55.5	-

MT-Clinical BERT

A **Single** pre-trained clinical model that performs competitively for **multiple tasks.**

- We can achieve massive gains at inference time.
    - A single forward pass of BERT involves hundreds of millions of CPU operations.
      Now the BERT encoder only needs **one** pass through for **all** tasks!

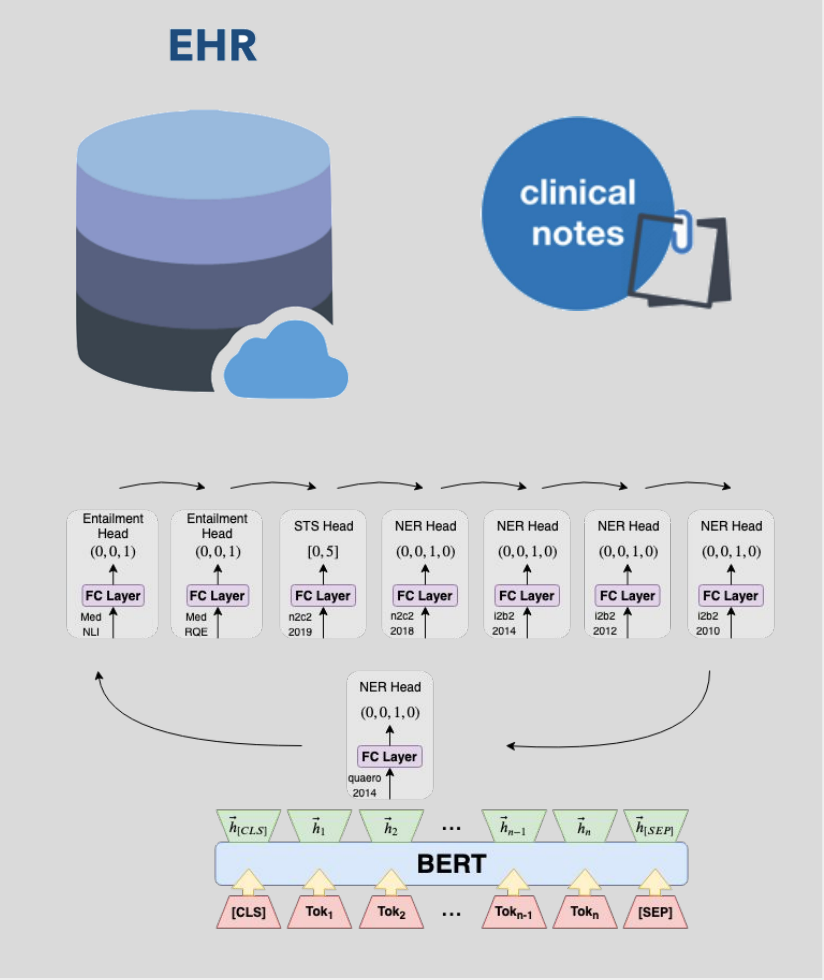
	MT-Clinical BERT	Optimized Clinical BERT	Clinical Bert[8]
n2c2-2019	86.7 (−0.5)	87.2	-
MedNLI	80.5 (−2.3)	82.8	82.7
MedRQE	76.5 (−3.6)	80.1	-
n2c2-2018	87.4 (−0.7)	88.1	-
i2b2-2014	91.9 (−3.6)	95.5	92.7
i2b2-2012	**84.1** (+0.2)	83.9	78.9
i2b2-2010	89.5 (−0.3)	89.8	87.8
quaero-2014	49.1 (−6.4)	55.5	-

MT-Clinical BERT

A **Single** pre-trained clinical model that performs competitively for **multiple tasks.**

- Task orthogonality presents some difficulties.
  - Span level objectives are hard to learn when you predominantly have token level tasks.
  - Shifting data distributions such as crossing lingual boundaries.

	MT-Clinical BERT	Optimized Clinical BERT	Clinical Bert[8]
n2c2-2019	86.7 (−0.5)	87.2	-
MedNLI	80.5 (−2.3)	82.8	82.7
MedRQE	76.5 (−3.6)	80.1	-
n2c2-2018	87.4 (−0.7)	88.1	-
i2b2-2014	91.9 (−3.6)	95.5	92.7
i2b2-2012	**84.1** (+0.2)	83.9	78.9
i2b2-2010	89.5 (−0.3)	89.8	87.8
quaero-2014	49.1 (−6.4)	55.5	-



Whats next?

Can we:

- Leverage multitask learning to use less annotated data?
- Pack more tasks into the encoder?
- Exploit self-supervised pre-training signal **during** multitasking?

We train a **single** neural network to simultaneously perform **multiple** clinical note information extraction tasks at near state-of-the-art performance.

arXiv.org

GitHub

VCU
Andriy Mulyar and Bridget T. McInnes
@andriy_mulyar @btmcinnes
NLP@VCU
Virginia Commonwealth University
In Submission, AMIA 2020.