

Research Summary

Shiwei Zhang

February 15, 2023

Overview

1. Introduction
2. Research at RMIT University
3. Research at Tencent Jarvis Lab
4. Summary

Introduction

A bit about myself

1. **Who am I:** a person never stop exploring, learning, imagining, and creating.
2. **Where am I from:** China(2021-), Australia(2008-2021), China(1985-2008).
3. **Where am I going:** Travel anywhere in the world.

My career

1. Applied medical NLP researcher (Tencent Jarvis Lab).
2. Software developer / data analyst (OracleCMS, HealthConnex).
3. Store manager (Juice bar, Bottle shop).

My hobbies

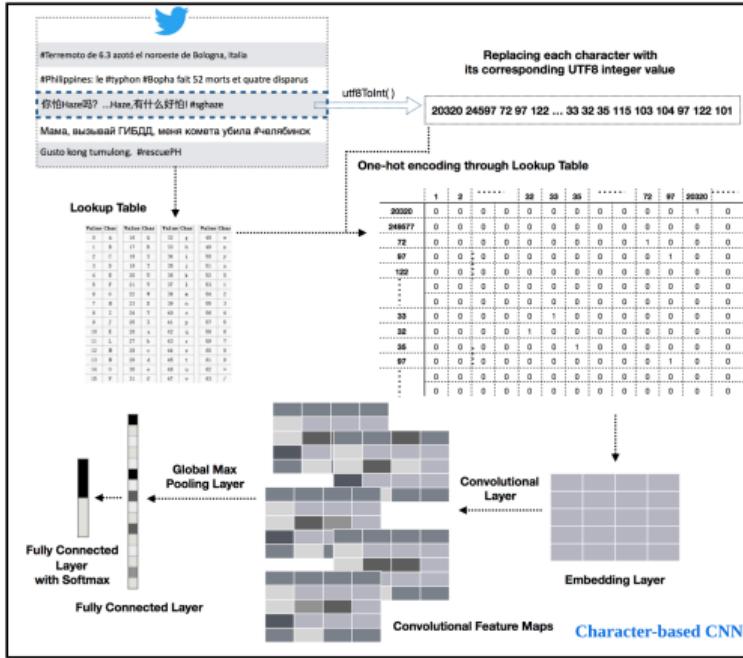
Traveling, trekking, hiking, mixed martial arts.

*Multilingual
Sentiment
Analysis*

Irony Detection

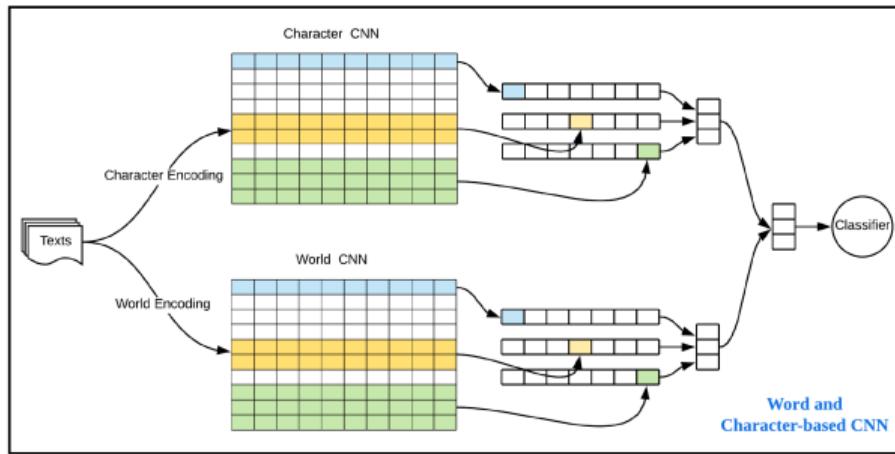
Product
Question
Answering

Multilingual Sentiment Analysis



Problem: Twitter supports over 40 languages and about 50% of tweets are non-English tweets.

Solution: Encoding the sequence of characters in a tweet into a sequence of numerical UTF-8 codes (corresponding UTF-8 integer values).

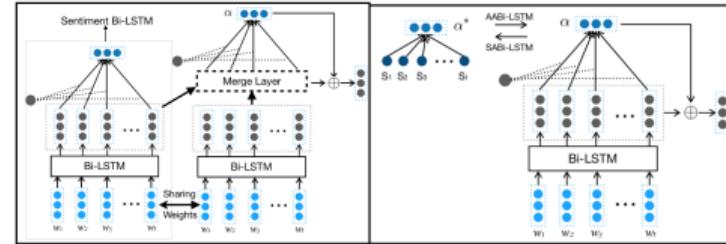


Irony Detection

Context incongruity is a commonly seen form of irony on Twitter, especially the contrast between the **positive** statement and the **negative** context.

Attention-based models (Bi-LSTM) often fail to capture words and phrases crucial for building the ironic intent.

We proposed our models to take advantage of these widely and readily available **sentiment resources** to improve the ability of attention-based model on detecting context incongruity.



picked an excellent day to get my hair done 😊
highly inflamed stomach is just what i like

i just love being ignored 😊 | 😒
picked an excellent day to get my hair done 😊
laptop speakers are too quiet for music & simply too loud for porn 😒 !
i like to think of myself as a broken down justin bieber - my philosophy professor

Product Question Answering

Are these headphones suitable for use while running?
asked on 2 July 2019

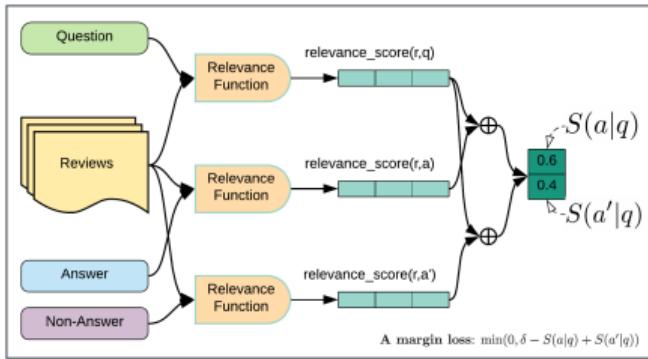
Yep used it while running on a treadmill and it didn't come off.
Chrishan [21 July 2019]

[Leave a comment](#) | Do you find this helpful? No | Report abuse
19 days

★★★★★ Genuinely surprised how good their True capsule wireless earbuds i
By Dan on 8 July 2019

Have been using soundpeats earbuds over the past few days and have come to the conclusion Sound has strong Bass, clean sounds and is loud to block out external sounds. Been using it at the gym, on runs and hikes and have not slipped out.

Reviews are relevant and helpful, but there is a lack of ground truth.

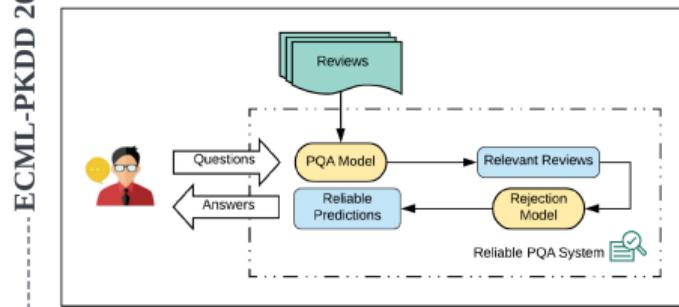


ICDM 2019

Q1: What is the chain for on the side?		
Top 3 Ranked Reviews	Prob	Conf
- This was driving me crazy but i see that another reviewer explained that grill has wire clip on chain to be used as extended match holder for igniting the gas if the spark mechanism fails to work or is worn out as sometimes happens with any gas grill.	0.99	0.82
- PS Could not figure out the purpose of that little chain with the clip attached to the outside of the grill even after reading entire manual.	0.95	0.54
- It is to replace an old portable that I have been using for about 10 years.'	0.91	0.40

Q2: Does this Dell Inspiron 14R i14RMT-7475s come with dell's warranty?		
Top 3 Ranked Reviews	Prob	Conf
- I don't really recommend the PC for people who wants install heavy games programs.	0.74	0.48
- The computer is nice, fast, light, ok.	0.12	0.01
- I bought the computer for my daughter.	0.05	0.00

1. Only keeping relevant and reliable reviews.
2. For unanswerable questions, return nil.



**Symptom
Detection**

**Disease
Prediction Using
EHR**

Medical-T5

Symptom Detection

录入学前信息-疾病信息

基本信息 疾病信息 既往病史 提交信息

Is this your first medical consultation or follow-up visit?

first consultation 初诊 重新回答

您本次就诊的原因是? Why you need medical consultation?

恶心 头痛 由尿 胸痛 胸闷 心悸 消瘦 下肢水肿 停止排便 嗜睡 全身麻木 气促 吞血 呼吸困难 发烧 便秘 出血不止 咳嗽

Others (If the above list does not contain symptoms that patients have, they can input their query below.)

感觉脖子上有疙瘩 (Feeling a pimple on the neck)

Determined symptoms

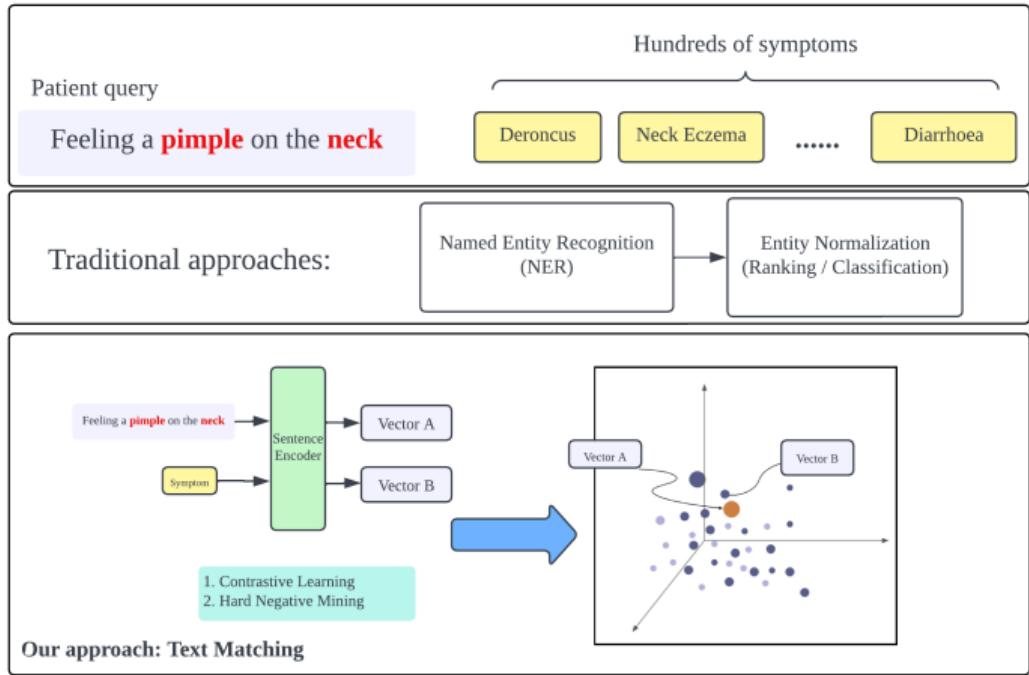
多选 Multiple Choice

颈部肿块 Deroncus 颈部皮疹 Neck Eczema 颈部麻木 Neck Numbness 颈部淋巴结肿大 Cervical Lymphadenopathy 皮疹 Erythra 都没有 None of them

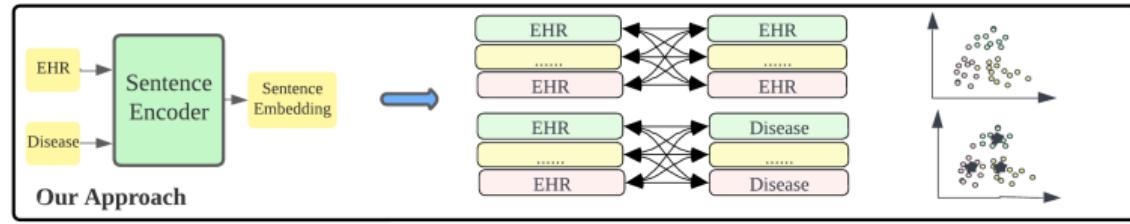
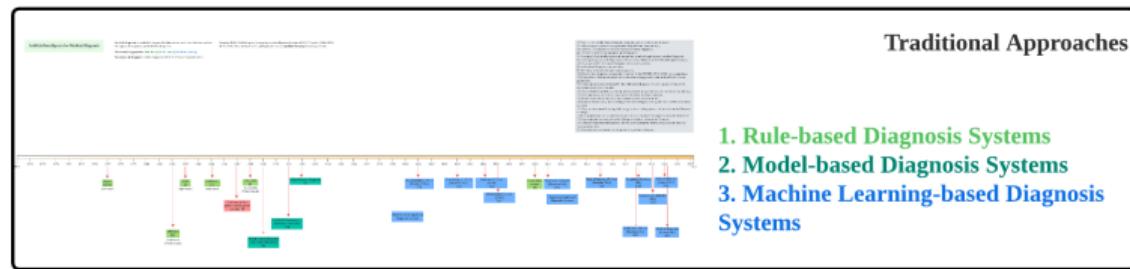
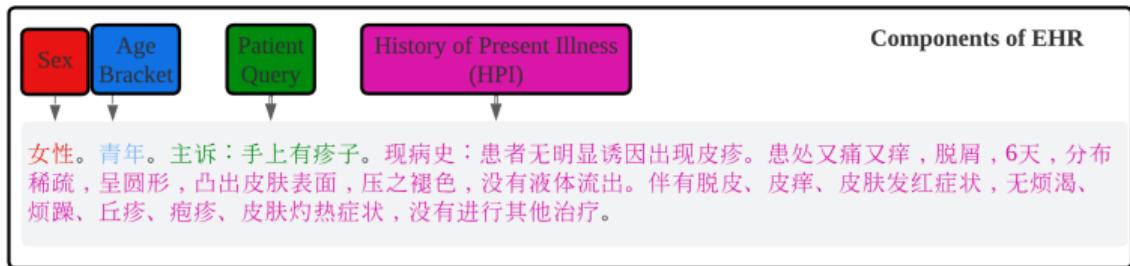
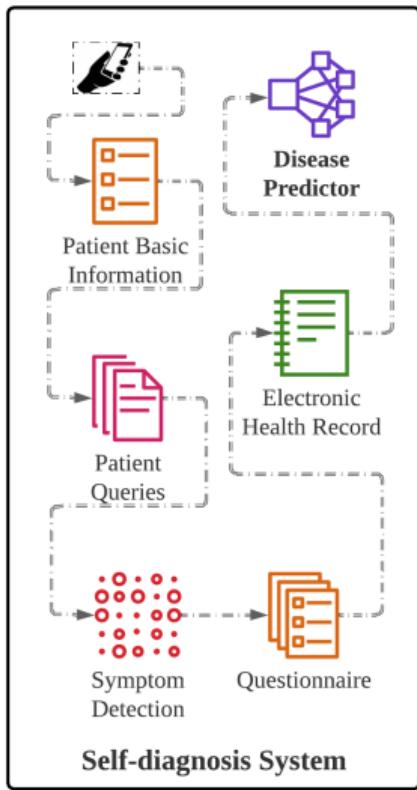
确定 Submit

1. Patient inputs query.

2. Patient will have a chance to confirm detected symptoms.

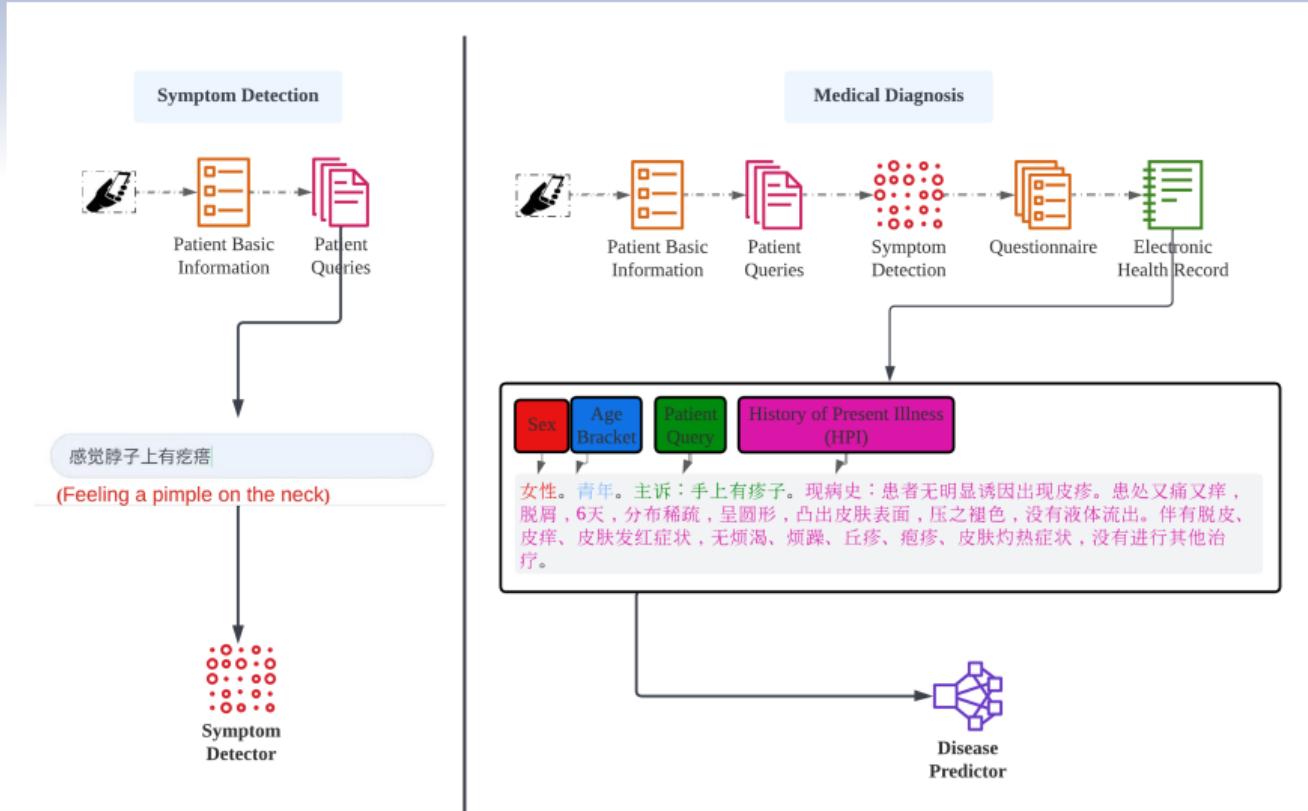


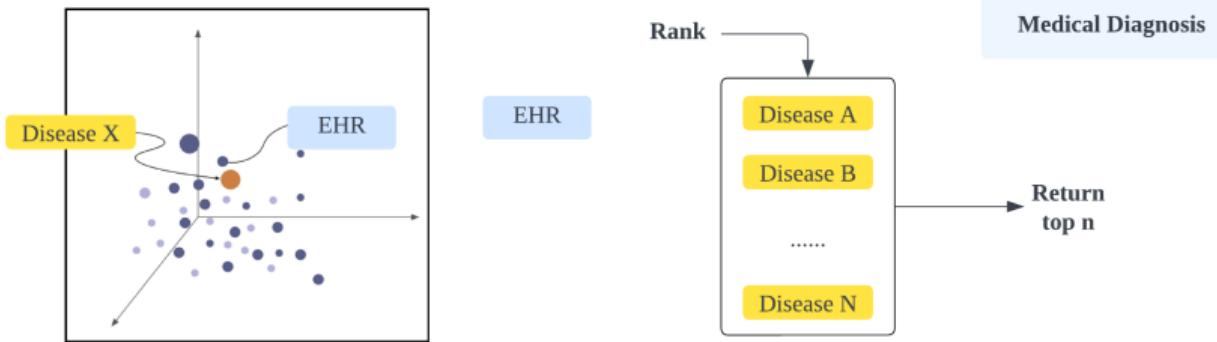
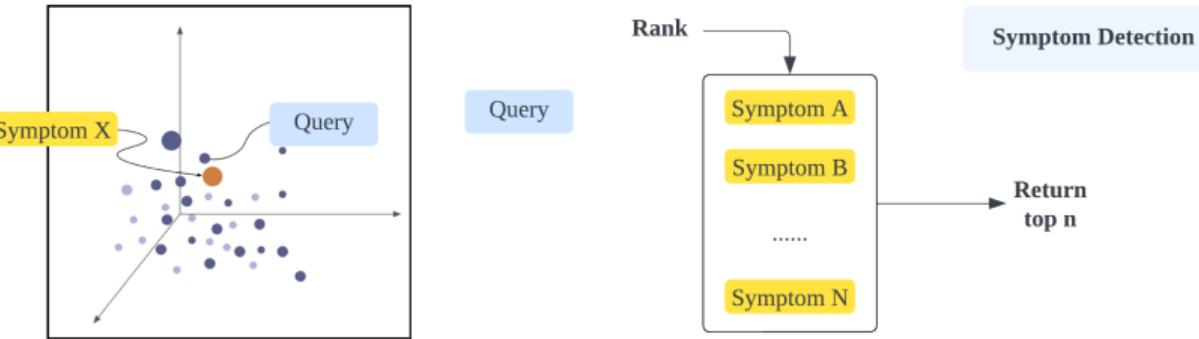
Disease Prediction using EHR



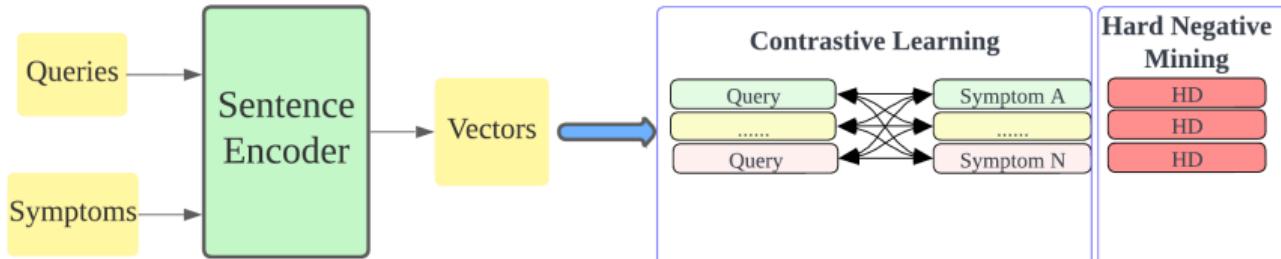
Summary of My Recent Research

1. **Tasks:** Symptom Detection and Medical Diagnosis using EHR.
2. **Methods:** Sentence Representation; Medical Multi-task Learning.

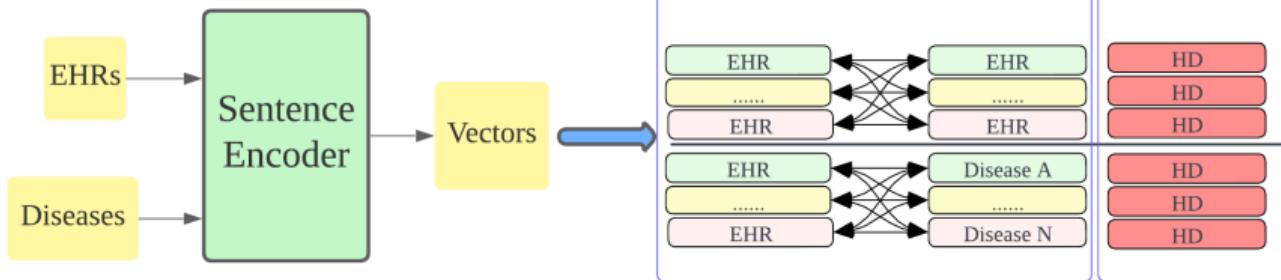


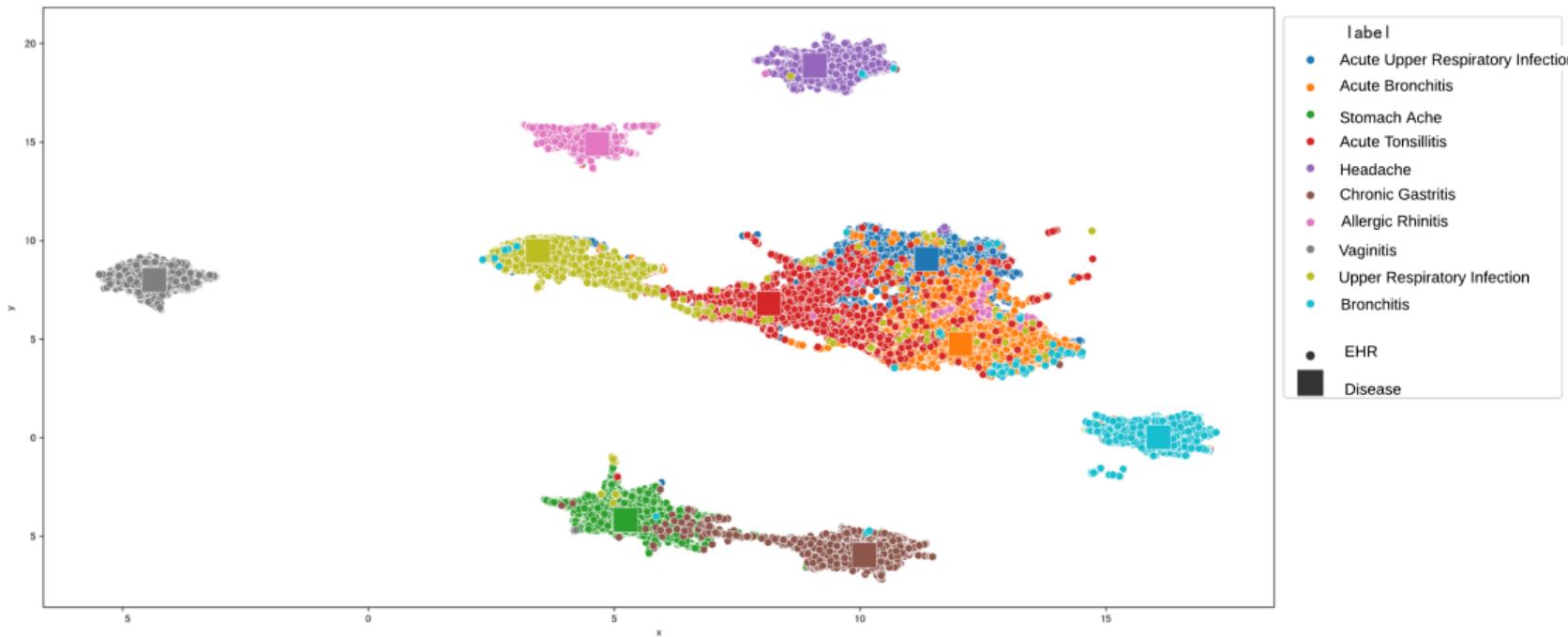


Symptom Detection

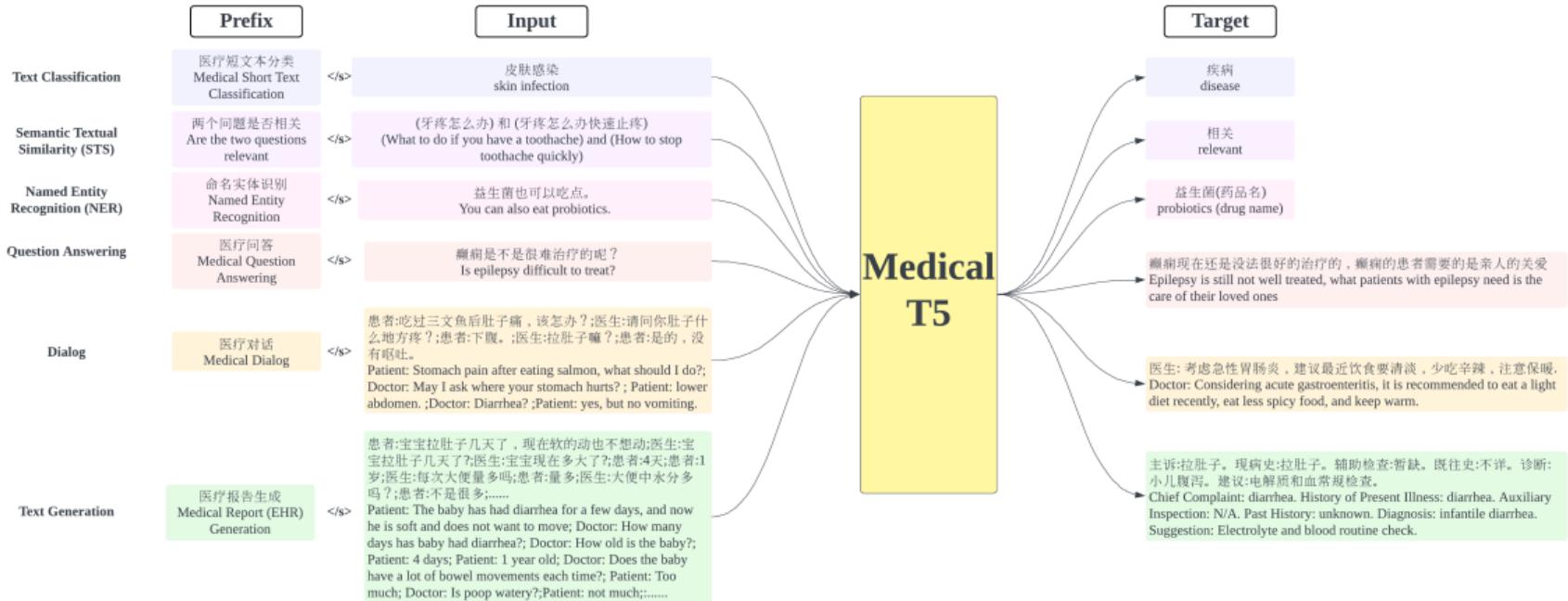


Medical Diagnosis



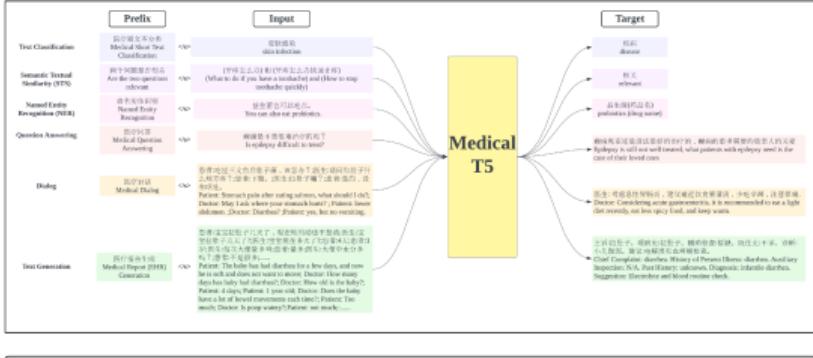


Medical-T5



Demo: <https://huggingface.co/spaces/zswvivi/ChineseMedicalT5>

Pros and Cons (Medical-T5)



Performances on CBLUE 1.0 (Chinese Biomedical Language Understanding Evaluation)									
	STS	QQR	QTR	QIC	CTC	CDD	CMeEE	CMeIE	Avg
BERT-base (110M)	87.82%	82.18%	66.18%	81.58%	82.49%	75.48%	80.74%	71.41%	77.15%
Medical-T5-base (220M)	87.26%	83.11%	65.67%	79.44%	73.71%	52.31% (red)	66.36% (red)	38.86% (red)	65.64%
Roberta-large (330M)	88.17%	83.68%	68.45%	81.63%	84.12%	76.64%	80.87%	71.94%	78.19%

Legend for columns:

- Semantic Textual Similarity (STS, QQR, QTR)
- Classification (QIC, CTC, CDD)
- NER-related Tasks (CMeEE, CMeIE)

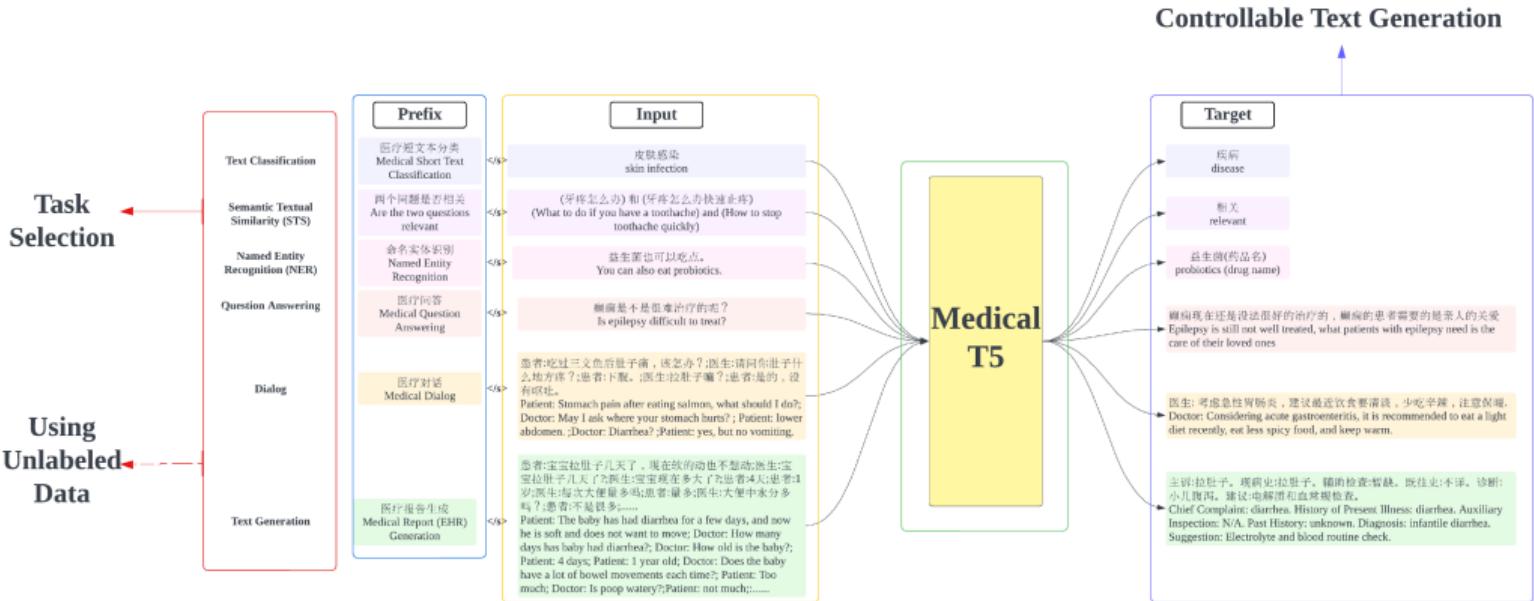
Advantages

1. Resource-efficient.
2. Knowledge transfer.

Disadvantages

1. Challenging to control text generation.
2. Time-consuming.

Research Directions (Medical-T5)



	RMIT University			Tencent Jarvis Lab			Next 1-2 years		2-5 years (Future)	
Task	Multilingual Sentiment Analysis	Irony Detection	Product Question Answering	Symptom Detection	Disease Prediction Using EHR	Medical Multi-task	Medical Multi-task	Medical Question Answering	Chatbot or Conversational QA	Sequence-to-sequence models
Model	Machine Learning Models (SVM, DT)	CNN	LSTM	Machine Learning Models (xgboost)	BERT	Sentence Transformer	Sequence-to-sequence models		Multi-task Learning	Controllable Text Generation
Achievements	Attention-based models	BERT	Mixture of Experts	T5	Contrastive Learning	Hard Negative Mining	Multi-task Learning	Controllable Text Generation		
Thoughts	Conformal Prediction	Weakly Supervised Learning	Multi-task Learning	Multi-task Learning	Controllable Text Generation	Academic Publications (KDD, COLING)				
	Best Paper Award (at IP&M 2091)	Academic Publications (ICDM, ECML-PKDD, IP&M)	Eventually, raw data will completely replace manually engineered features as inputs.	Most machine learning techniques are contrastive learning (implicit or explicit)	Through contrastive learning, models learn more fine-grained features.	Hard negatives are examples with fine-grained features.				
	Text-to-text is the future.			Text-to-text is better than text-to-logits on transfer learning.	Controllable text generation is the key component of text-to-text models.					