Local context is not enough!

Towards Query Semantic and Knowledge Guided Multi-Span Medical Question Answering

Abhisek Tiwari[*], Aman Bhansali[$], Sriparna Saha[*], Pushpak Bhattacharyya[#] & Minakshi Dhar[**]
Indian Institute of Technology, [*]Patna, [$]Jodhpur, [#]Bombay, & [**]AIIMS Rishikesh

सत्यमेव जयते
Ministry of Education
Government of India

Introduction

- **Information Overload in Medicine** Medical knowledge is doubling approximately every 73 days, PLOS Medicine 2021
- **Timely and Accurate Information** Remote and underserved areas
- **Telemedicine and Telehealth** COVID-19, a survey conducted by the Pew Research Center, 2021, found that 80% of internet users have searched for health-related information online

Medical Question Answering

08/31/96 ascending aortic root replacement with homograft with omentopexy. The patient continued to be hemodynamically stable making good progress. Physical examination: BMI: 33.4 Obese, high risk. Pulse: 60. resp. rate: 18

Question: Has the patient ever had an abnormal BMI?

Answer: BMI: 33.4 Obese, high risk

Question: When did the patient last receive a homograft replacement ?

Answer: 08/31/96 ascending aortic root replacement with homograft with omentopexy.

Multi-Span Medical Question Answering

What are tips for managing my bipolar disorder?
Along with seeing your doctor and therapist and taking your medicines, simple daily habits can make a difference. Start with these strategies. *(22 words truncated)* Pay attention to your sleep. This is especially important for people with bipolar disorder... *(178 words truncated)* Eat well. There's no specific diet... *(29 words truncated)* Focus on the basics: Favor fruits, vegetables, lean protein, and whole grains. And cut down on fat, salt, and sugar. Tame stress. *(81 words truncated)* You can also listen to music or spend time with positive people who are good company. *(73 words truncated)* Limit caffeine. It can keep you up at night and possibly affect your mood. *(47 words truncated)* Avoid alcohol and drugs. They can affect how your medications work. *(118 words truncated)*

Research Questions

How can one drive a bike?

1. What does it mean and what do the listener expect *(Query Semantic)*
2. Learning from various time stamps *(Multi-Span)*
3. May include cycling experience *(Global Knowledge)*

RQ1: *Does the incorporation of query semantics affect the efficiency of extracting multi-span answers from documents?*

RQ2: *Is there a correlation between the question's intent and its type recognition?*

RQ3: *Does external medical knowledge provide the background and foundation to understand and extract answers from multiple paragraphs in a document?*

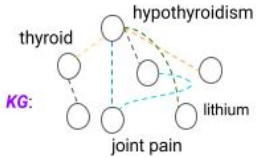
Context

Through the hormones it produces, the thyroid gland influences almost all of the metabolic processes in your body. Thyroid disorders can range from a small, harmless goiter (enlarged gland) that needs no treatment to life-threatening cancer. The most common thyroid problems involve abnormal production of thyroid hormones. Too much thyroid hormone results in a condition known as hyperthyroidism. Insufficient hormone production leads to hypothyroidism. Although the effects can be unpleasant or uncomfortable, most thyroid problems can be managed well if properly diagnosed and treated. All types of hyperthyroidism are due to an overproduction of thyroid hormones, but the condition can occur in several ways: Graves' disease: The production of too much thyroid hormone. Toxic adenomas: Nodules develop in the thyroid gland and begin to secrete thyroid hormones, upsetting the body's chemical balance; some goiters may contain several of these nodules. Subacute thyroiditis: Inflammation of the thyroid that causes the gland to "leak" excess hormones, resulting in temporary hyperthyroidism that generally lasts a few weeks but may persist for months. Pituitary gland malfunctions or cancerous growths in the thyroid gland: Although rare, hyperthyroidism can also develop from these causes. Hypothyroidism , by contrast, stems from an underproduction of thyroid hormones. Since your body's energy production requires certain amounts of thyroid hormones, a drop in hormone production leads to lower energy levels. Causes of hypothyroidism include: Hashimoto's thyroiditis : In this autoimmune disorder, the body attacks thyroid tissue. The tissue eventually dies and stops producing hormones. Removal of the thyroid gland: The thyroid may have been surgically removed or chemically destroyed.

Exposure to excessive amounts of iodide: Cold and sinus medicines, the heart medicine amiodarone, or certain contrast dyes given before some X-rays may expose you to too much iodine. You may be at greater risk for developing hypothyroidism if you have had thyroid problems in the past. Lithium: This drug has also been implicated as a cause of hypothyroidism. Untreated for long periods of time, hypothyroidism can bring on a myxedema coma, a rare but potentially fatal condition that requires immediate hormone treatment. Hypothyroidism poses a special danger to newborns and infants. ..... People who have received radiation treatment to the head and neck earlier in life, possibly as a remedy for acne, tend to have a higher-than-normal risk of developing thyroid cancer.

Query
What give rise to hypothyroidism?

Query Semantic
Intent: Cause, Query Type: What



Contributions

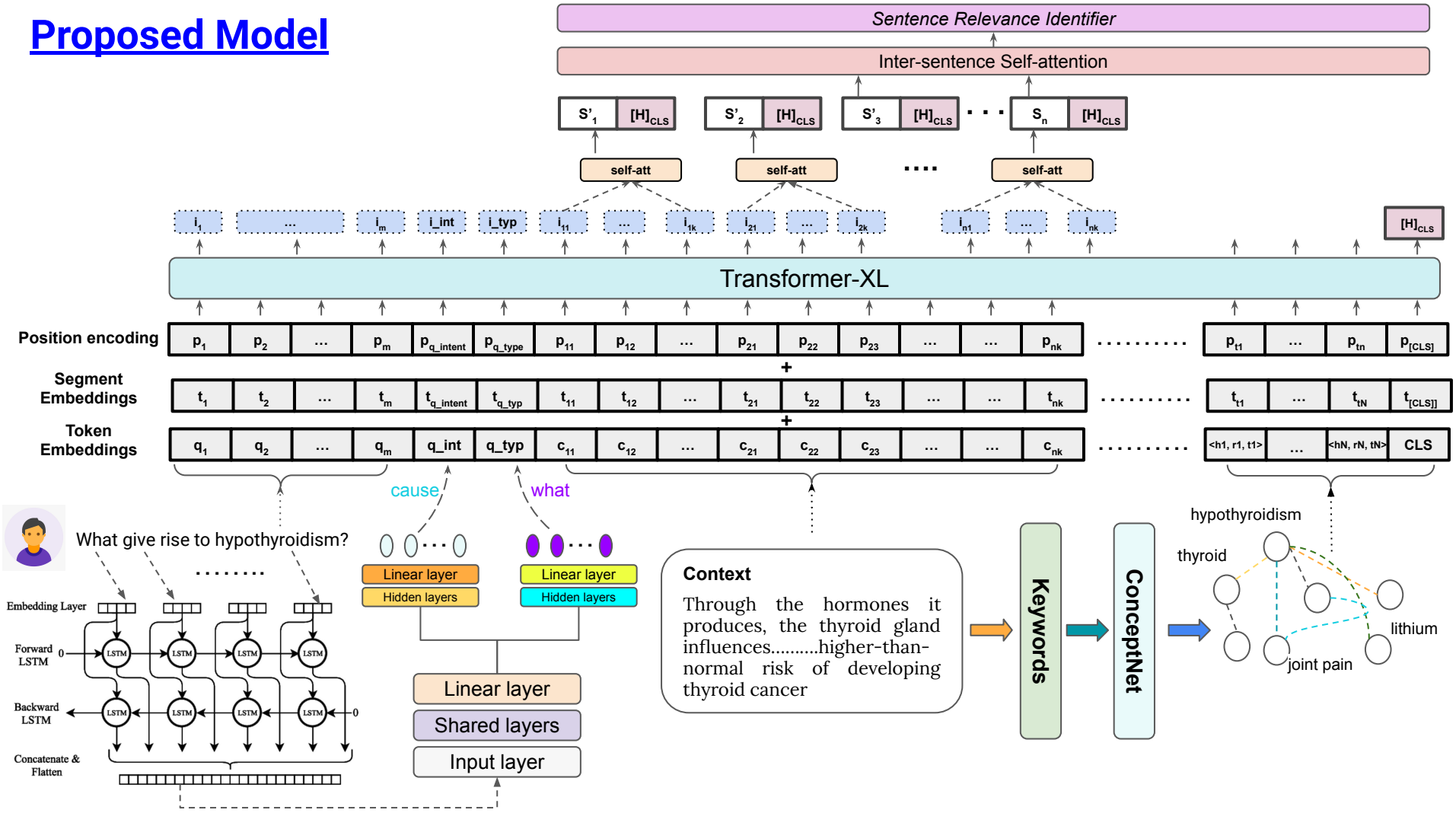
- **QueSemKnow Multi-Span Question Answering Framework:** We propose a query semantic and knowledge-guided multi-span question-answering framework that first extracts semantic information from a question, then extracts a relevant subset from the medical knowledge graph as per the underlying context (document), and finally identifies relevant sentences accordingly.

- **Multi-tasking Intent and Query type Identifier:** We propose a multi-task intent and query type identification model that exploits the interrelation between the intent and query type of a question.

- **QueSemSpan Corpora:** We curate large-scale semantic information annotated medical multi-span question answering corpus, which contains intent and question type for each context-question pair.

- **State-of-the-art Performance:** The proposed model surpasses the existing state-of-the-art models over multiple datasets across different evaluation metrics.

Proposed Model



Proposed Model

Knowledge Infusion

Algorithm 1 Context relevant Knowledge Distillation

Input Context ($C : s_1, s_2, s_3, ...s_n$) where s_i represents i^{th} sentence of the context (C) having n sentences
Output Context relevant Knowledge Graph (KG_C)
Initialization n_k: threshold for number of keywords from a single context, n_r: threshold for number of concepts for a keyword

KG_C = []
K[1, 2, ...n_k] = YAKE(C, n_k) $\Rightarrow K$: list of keywords
for entity in K **do**
    KGT_{entity} =[] $\Rightarrow KGT_{entity}$: KG triplet corresponding to keyword (entity)
    for j in range(0, n_r) **do**
        <r_j, h_j, t_j> = ConceptNet(entity, KGT_{entity})
        KGT_{entity} = KGT_{entity} + [r_j, h_j, t_j]
    end for
    KG_C = KG_C + KGT_{entity}
end for
return KG_C

Self attention layer

$$h_{ij} = w_s tanh(W_s \cdot C_{ij})$$
$$att_{ij} = Softmax_j(h_{ij})$$
$$S_i^{'} = \sum_{j=1}^{j=k} att_{ij} \cdot C_{ij}$$

Inter Self attention layer

$$si\_sa_{ij} = w_s \cdot tanh(W_s \cdot S_{ij}^{'})$$
$$\beta_{ij} = f_s(si\_sa_{ij})$$
$$S_i^{''} = \sum_{j=1}^{j=k} \beta_{ij} \cdot S_{ij}^{'}$$

$$f_s = ReLU[(\alpha - 1) \cdot a - \tau]^{\frac{1}{\alpha - 1}}$$

Sentence Relevance

$$y_i = softmax(W_o \cdot S_i^{''} + b_o)$$

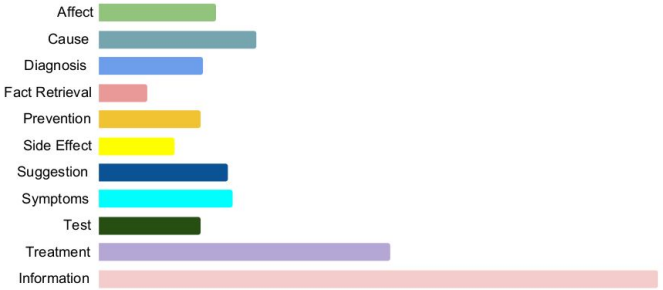
Loss Calculation

$$loss = -\sum_{j=1}^{j=N} \sum_{i=1}^{i=n} [y_i^{(j)} log(\hat{y}_i^{(j)}) + (1 - y_i^{(j)}) log(1 - \hat{y}_i^{(j)})]$$

Datasets

Dataset	#QA	Context	Intent	QA Type	Multi-Span	QuesSem
HealthQA [34]	8K	✗	✗	Ranking	✗	✗
MedQuaD [4]	47K	✗	✗	Ranking	✗	✗
Medication [5]	690	✗	✓	Ranking	✗	✗
MASH-QA [33]	35K	✓	✗	Extractive	✓	✓
QueSeMSpan (ours)	34.8K	✓	✓	Extractive	✓	✓

Entries	Value
# of question context pairs	34,808
# of context	5574
# of question annotated with intent and query type	5041
Avg. context length (in words)	696
Avg. answer length (in words)	67
# of intents	11
# of query types	12



Intent	%age	Query
information	59.30	Is there anything that I need to consider other than glycemic index when making dietary choices for diabetes?
treatment	14.07	What are some treatments for neutropenia? Will my doctor prescribe chemotherapy alone or with other treatments?
suggestion	6.57	What is the best way for people with HIV/AIDS to prevent the flu?
cause	6.29	What causes delusional disorder?
symptoms	5.89	How can one be certain of having gastritis?

Results & Discussion

Performance of different NLU models

Model	Accuracy(%)	F1-Score
FFNN	75.4	0.746
LSTM	81.4	0.782
BiLSTM	86.2	0.848

Model	Accuracy(%)	F1-Score
FFNN	94.8	0.943
LSTM	96.3	0.889
BiLSTM	98.1	0.968

Task	Accuracy (%)	F1-Score
Intent idn	**89.1 (2.9 ↑)**	**0.882 (0.034 ↑)**
Query type idn	**98.9 (0.8 ↑)**	**0.984 (0.016 ↑)**

Performance of existing and proposed multi-span question answering models

Model	F1-Score	EM1	CRP
BERT [8]	25.21	8.89	/
RoBERTA [18]	28.65	9.40	/
XLNet [32]	29.19	9.09	/
TANDA [9]	25.44	8.95	/
MultiCo [33]	50.81	17.80	93.22
QueSem (ours)	**55.29 (4.48 ↑)**	**21.15 (3.35 ↑)**	**94.15 (0.93 ↑)**
QueKnow (ours)	**53.10 (2.71 ↑)**	**19.55 (1.75 ↑)**	**93.72 (0.50 ↑)**
***QueSemKnow* (ours)**	**55.81 (5.00 ↑)**	**21.33 (3.53 ↑)**	**94.45 (1.23 ↑)**

Model	F1-Score	EM	CRP
QueSemKnow w/o Intent, Query and KG	50.81	17.80	93.22
QueSemKnow with only Intent	51.89	18.56	93.25
QueSemKnow with only Query type	52.62	19.21	93.66
QueSemKnow with only KG	53.10	19.55	93.72

Results & Discussion

Model	F1-Score	EM	CRP
BERT [8]	27.93	3.95	/
XLNet [32]	56.46	22.78	/
SpanBERT [13]	30.61	5.62	/
MultiCo [33]	59.38	26.40	95.65
QueSem (ours)	**60.03 (0.65 ↑)**	**26.72 (0.032 ↑)**	**95.68 (0.03 ↑)**
QueKnow (ours)	**63.76 (4.38 ↑)**	**27.71 (1.31 ↑)**	**96.13 (0.48 ↑)**
***QueSemKnow* (ours)**	**61.88 (2.50 ↑)**	**27.65 (1.25 ↑)**	**95.99 (0.34 ↑)**

Model	A	F	R	C	MSI	Avg.
MultiCo [33]	2.44	2.16	1.78	1.18	2.14	1.94
QueSem	2.58	2.36	1.96	1.24	2.32	2.09
QueKnow	2.74	2.34	1.65	1.39	2.48	2.12
QueSemKnow	**2.86**	**2.60**	**2.04**	**1.88**	**2.82**	**2.44**

Query

What medicines do doctors use to treat delusional disorder?

Context

Delusional disorder, previously called paranoid disorder, is a type of serious mental illness called a psychotic disorder. People who have it can't tell what's real from what is imagined. Delusions are the main symptom of delusional disorder….. **[52]** The primary medications used to attempt to treat delusional disorder are called antipsychotics. **[53]** Drugs used include: Conventional antipsychotics: Also called neuroleptics, these have been used to treat mental disorders since the mid-1950s. **[54]** They work by blocking dopamine receptors in the brain. **[55]** Dopamine is a neurotransmitter believed to be involved in the development of delusions. **[56]** Conventional antipsychotics include Chlorpromazine(Thorazine) Fluphenazine (Prolixin) Haloperidol (Haldol) Loxapine (Oxilapine) Perphenazine (Trilafon), Thioridazine (Mellaril), Thiothixene (Navane) ---. **[58]** Serotonin is another neurotransmitter believed to be involved in delusional disorder. **[59]** These drugs include: Aripiprazole (Abilify) Aripiprazole Lauroxil (Aristada) Asenapine (Saphris) Brexpiprazole (Rexulti) Cariprazine (Vraylar) Clozapine (Clozaril) Iloperidone (Fanapt) .--- disorder. **[60]** Tranquilizers might be used if the person has a very high level of anxiety or problems sleeping. **[61]** Antidepressants might be used to treat depression, which often happens in people with delusional disorder Psychotherapy can also be helpful, along with medications, as a way to help people better manage and cope with the stresses related to their delusional beliefs and its impact on their lives. **[62]** Psychotherapies that may be helpful in delusional disorder include: Individual psychotherapy can help the person recognize and correct the thinking that has become distorted. …… There's no known way to prevent delusional disorder. But early diagnosis and treatment can help lessen the disruption to the person's life, family, and friendships.

Gold [52, 61, 62]
MultCo [59]
QueSem [52,53,54,55,56,57,58,59,60,61]
QueKnow [52,53,54,55,56,57,58,59,60]
QueSemKnow [52,53,59, 60,61,62]

Conclusion

Summary

- In this work, we proposed a two-phased query semantic and knowledge guided medical question answering model that first extracts query semantic and relevant external medical knowledge in the first and identifies the relevance of sentences in the later phase
- We also developed a multi-task framework for identifying query intent and type, which exploits the interrelationship between tasks to recognize query semantics more effectively.
- Experimental results demonstrated the significant importance of infusing query semantic and external knowledge in multiple answer span question answering

Future Directions

- Knowledge Graph Traversal: Depth is good or width is good ? what is optimal ? 2 → 2 ?
- Towards an abstractive query response generation model for multi-span setting
- Is it possible to address multi-span question answering more efficiently by framing it as a sequence labeling challenge?

सत्यमेव जयते
Ministry of Education
Government of India
Prime Minister Research Fellowship (PMRF)

Microsoft Research

ecai 2023 KRAKÓW

Dziękuję !

Abhisek Tiwari
Prime Minister Research Fellow (PMRF)
Department of Computer Science and Engineering
Indian Institute of Technology, Patna, India
Website : https://abhisektiwari.github.io/
Email: abhisek_1921cs16@iitp.ac.in

Slide, Paper, Dataset & Code