Build Your Own Chatbot

LYU2203
Chau Tak Ho (1155143501)
Wang Weixiao (1155141608)
Supervised by Prof. Michael R. Lyu

Agenda

01 Introduction

02 Background Knowledge

03 Methodology

04 Experiment & Analysis

05 Conclusion

06 Demo

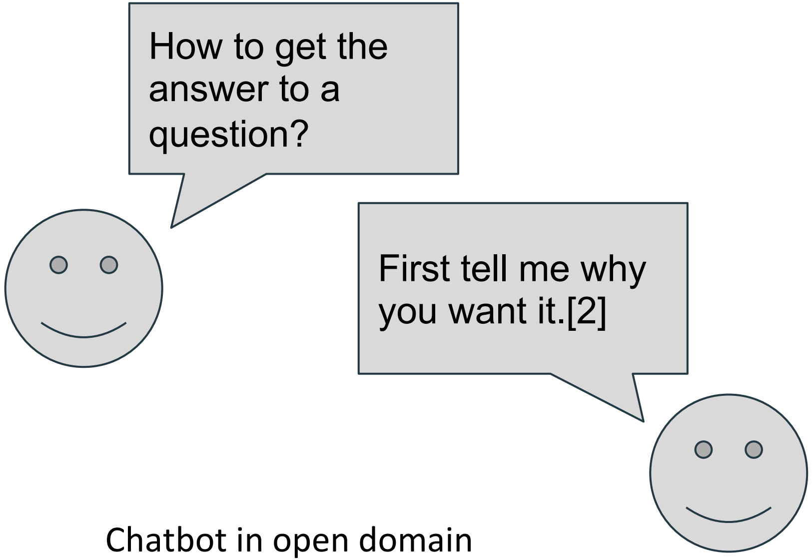
01 Introduction

Introduction

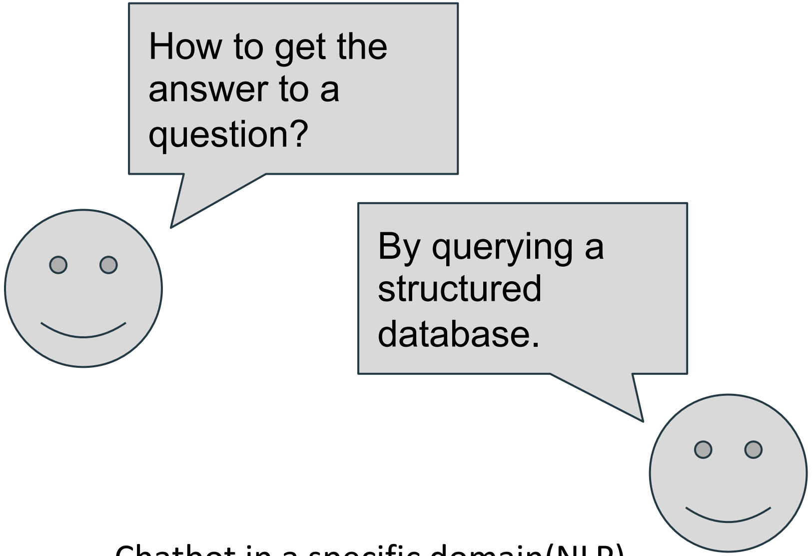
- What is a chatbot?
  - Chatbot is a conversational computer program that aim at simulating natural human conversation[1]
  - Example: Siri, Alexa

Objective

- We want to build a chatbot that can talk with you about a specific domain with some “background knowledge“ in that domain, e.g. terminologies, fundamental facts.



Chatbot in open domain



Chatbot in a specific domain(NLP)

Objective

- Question-Answering
  - One of the main task of a chatbot
  - Process a question → output an answer
- Closed-domain
  - Focus on a specific domain or field
  - Advantages :
    - Accurate
    - Efficient

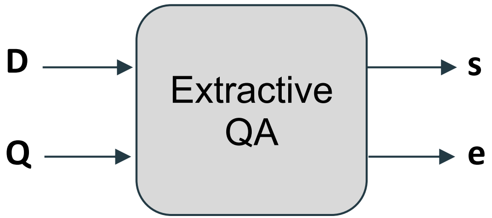
02 Background Knowledge

Extractive QA

A method of directly finding answers in the given document.

Document : D = {d1, d2, d3, ..., dn}

Question : Q = {q1, q2, q3, ..., qm}



Answer : A = {ds, ..., de}

The Amazon rainforest (Portuguese: Floresta Amazônica or Amazônia; Spanish: Selva Amazónica, Amazonía or usually Amazonia; French: Forêt amazonienne; Dutch: Amazoneregenwoud), also known in English as Amazonia or the Amazon Jungle, is a moist broadleaf forest that covers most of the Amazon basin of South America. This basin encompasses 7,000,000 square kilometres (2,700,000 sq mi), of which 5,500,000 square kilometres (2,100,000 sq mi) are covered by the rainforest.[3]

Which name is also used to describe the Amazon rainforest in English?

Amazonia or the Amazon Jungle s=26, e=30

Where is Amazon rainforest?

South America s=44, e=45

What is Amazon rainforest?

a moist broadleaf forest s=32, e=35

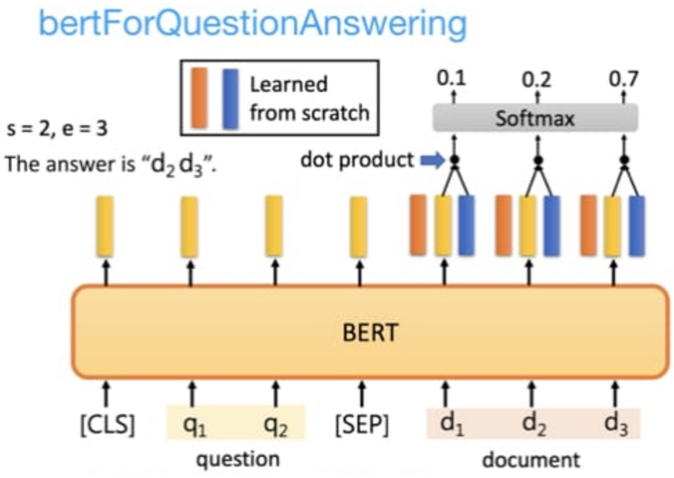
Extractive QA Model

- roberta-base-squad2[3]
- Start vector S, End vector E, output embedding Ti
- Probability = dot product between Ti and S followed by a softmax :

$$P_i = \frac{e^{S \cdot T_i}}{\sum_j e^{S \cdot T_j}}$$

- Score of a candidate span from position i to position j :

$$S \cdot T_i + E \cdot T_j$$



BERT fine-tuning for QA[4, 5]

Question Decomposition

- Extractive QA cannot handle some complicated questions
  - Extractive QA get the information based on the question
  - Unable to answer questions that involves knowledge spread over the document
- Question Decomposition
  - Split a question into different sub-questions
  - To get information step by step

Question Decomposition Model

- DecompRC [6]
- Split question into different sub-question of 3 types:
  - Bridging
  - Intersection
  - Comparison

Type	**Bridging (47%)** requires finding the first-hop evidence in order to find another, second-hop evidence.
Q	Which team does the **player** named 2015 Diamond Head Classic's MVP play for?
Q1	Which player named 2015 Diamond Head Classic's MVP?
Q2	Which team does ANS play for?
Type	**Intersection (23%)** requires finding an entity that satisfies two independent conditions.
Q	Stories USA starred ✓ which actor and comedian ✓ from 'The Office'?
Q1	Stories USA starred which actor and comedian?
Q2	Which actor and comedian from 'The Office'?
Type	**Comparison (22%)** requires comparing the property of two different entities.
Q	Who was born earlier, Emma Bull or Virginia Woolf?
Q1	Emma Bull was born when?
Q2	Virginia Woolf was born when?
Q3	Which_is_smaller (Emma Bull, ANS) (Virgina Woolf, ANS)

Question Decomposition Model

- PointerC
- S = [s1, . . . , sn] denote a sequence of n words in the input sequence
- Encodes S using BERT

$$U = \text{BERT}(S) \in \mathbb{R}^{n \times h}$$

- Pointer score matrix

$$Y = \text{softmax}(UW) \in \mathbb{R}^{n \times c}$$

- Extracts c indices that yield the highest joint p

$$\text{ind}_1, \ldots, \text{ind}_c = \underset{i_1 \le \cdots \le i_c}{\text{argmax}} \prod_{j=1}^{c} \mathbb{P}(i_j = \text{ind}_j)$$

Algorithm 1 Sub-questions generation using Pointer$_c$.[2]

procedure GENERATESUBQ(Q : question, Pointer$_c$)

/ Find q_1^b and q_2^b for Bridging */*

$\text{ind}_1, \text{ind}_2, \text{ind}_3 \leftarrow \text{Pointer}_3(Q)$

$q_1^b \leftarrow Q_{\text{ind}_1:\text{ind}_3}$

$q_2^b \leftarrow Q_{:\text{ind}_1} : \text{ANS} : Q_{\text{ind}_3:}$

article in $Q_{\text{ind}_2-5:\text{ind}_2} \leftarrow$ 'which'

/ Find q_1^i and q_2^i for Intersecion */*

$\text{ind}_1, \text{ind}_2 \leftarrow \text{Pointer}_2(Q)$

$s_1, s_2, s_3 \leftarrow Q_{:\text{ind}_1}, Q_{\text{ind}_1:\text{ind}_2}, Q_{\text{ind}_2:}$

if s_2 starts with wh-word **then**

$q_1^i \leftarrow s_1 : s_2,\ q_2^i \leftarrow s_2 : s_3$

else

$q_1^i \leftarrow s_1 : s_2,\ q_2^i \leftarrow s_1 : s_3$

/ Find q_1^c, q_2^c and q_3^c for Comparison */*

$\text{ind}_1, \text{ind}_2, \text{ind}_3, \text{ind}_4 \leftarrow \text{Pointer}_4(Q)$

$\text{ent}_1, \text{ent}_2 \leftarrow Q_{\text{ind}_1:\text{ind}_2}, Q_{\text{ind}3:\text{ind}_4}$

$op \leftarrow \text{find\_op}(Q, \text{ent}_1, \text{ent}_2)$

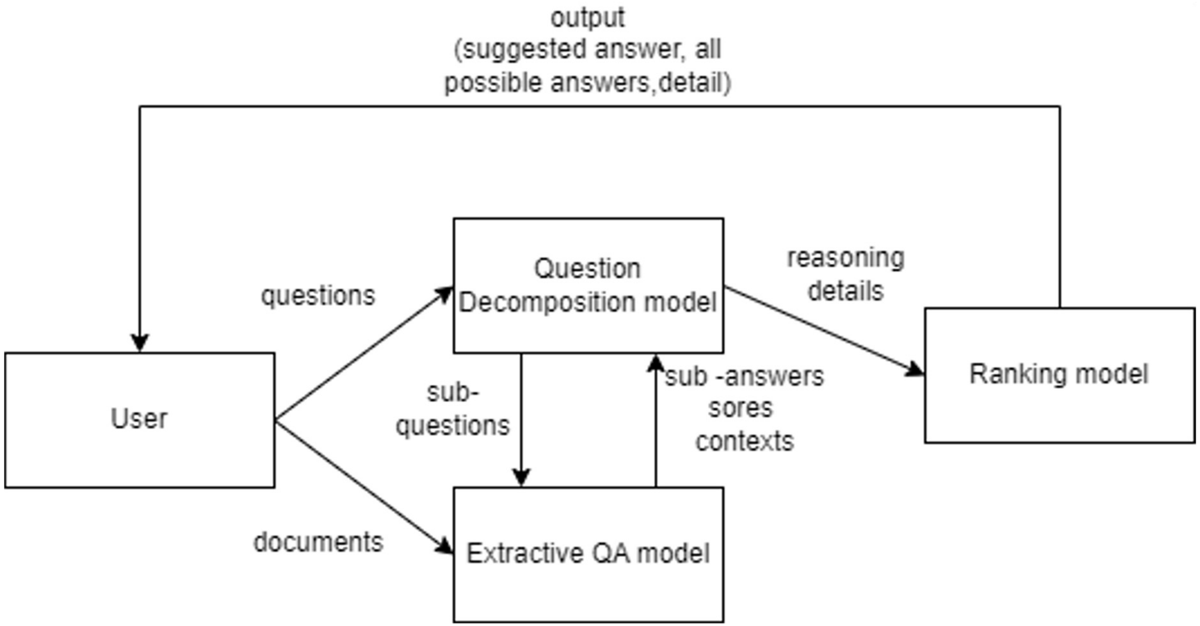
$q_1^c, q_2^c \leftarrow \text{form\_subq}(Q, \text{ent}_1, \text{ent}_2, op)$

$q_3^c \leftarrow op\ (\text{ent}_1, \text{ANS})\ (\text{ent}_2, \text{ANS})$

03 Methodology

Architecture

A question decomposition model that decomposing the compositional questions.
A extractive QA model that finding the answers of the questions in the documents.
A Ranking model that rank all candidate answers, and decide the most possible one.



Architecture:

Question Decomposition model

The question decomposition model is modified from the decomposition functions of DecompRC[30]. It read the question from the user, if it is a compositional question, the model decomposes it into sub-questions.

Example:
Original question: Which kind of tools does the method that replaced statistical methods in NLP research belongs to?
Sub-question1: What replaced statistical methods in NLP research?
Sub-question2: Which kind of tools does [ANSWER] belongs to?

In this example, once the QA model get the answer to sub-question1, it will replace the [ANSWER] token in sub-question2.
And the answer to sub-question2 will be the final answer to the original question. As their may be more than 1 answers of sub-question1, they form different sub-question2.

Architecture:

Extractive QA model

The QA model is fine-tuned from roberta-base-squad2[3]). It is an extractive QA model and it is trained on the basic knowledge of a specific field in advance.

It will generate the answers of a question according to given documents, and also give them corresponding score. Then, the answers, together with their score and the context where the model found them, will be send back.

In practice, we set the model to keep at most 5 most possible answers of a question.

Architecture:

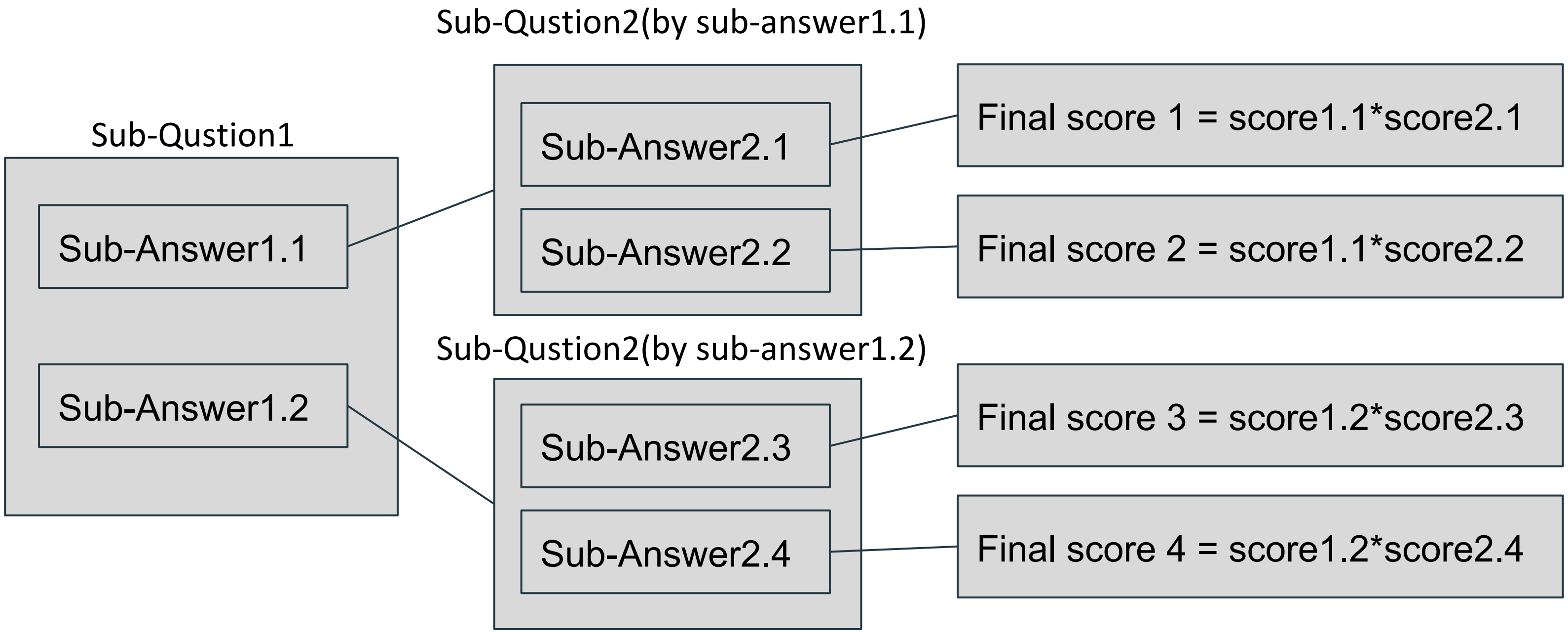
Answer Ranking model

As the extractive QA model could find more than one possible answers for each question, we build a answer ranking model to rank all final answers, and decide the most possible one.
In practice, we define the final score as the product of the scores of the sub-answers.

Architecture:

Answer Ranking model

Example:



04 Experiment

Experiment：Test on manually-made datasets

We made several datasets contain a document and a set of questions made from the original texts of the documents, the documents are from different fields.

We made questions by modifying the original sentences, example:

Statistical methods in NLP research have been largely replaced by neural networks -

Many different classes of what algorithms have been applied to natural-language-processing tasks?

Field	Document	Simple Question	Compositional Question
NLP	9426 words	20	28
Bank Service	4452 words	93	5

Experiment：Test on manually-made datasets

We test whether it can:
1. Decompose the sentence correctly.
2. Find the context that contains the answer.
3. Output the correct format of the answer, especially for those containing terms or special phrases.

Dataset	Decomposition	Context	Answer
NLP	21 of 28	37 of 41	34 of 41
Bank Service	5 of 5	89 of 98	86 of 98

Experiment: Limitations

Question Decomposition model: Not all the questions can be decomposed intuitively.

Context1: Machine learning algorithms build a model based on sample data, known as training data, in order to make predictions or decisions without being explicitly programmed to do so.[33] **Context2:** Decision trees is a earliest-used machine learning algorithm that produced systems of hard if-then rules similar to existing hand-written rules.[32]
Question: Decision trees build a model based on what?
subQ1: decision trees build which model? **subQ2:** [ANSWER]based on what?
Correct subQ1: What kind of algorithm does decision trees belongs to? **Correct subQ2:** [ANSWER] build a model based on what?
Final answer: sample data

Experiment: Limitations

Question Decomposition model: Some compositional questions do not need decomposition, and even could not be answered correctly by solve these sub-questions.

Context: Most higher-level NLP applications involve aspects that emulate intelligent behaviour and apparent comprehension of natural language.[32]	
Question: higher-level NLP applications involve aspects that emulate what?	
subQ1: higher-level NLP applications involve what aspect? **subQ2:** [ANSWER]emulate what?	

Experiment: Limitations

Extractive QA model: The meaning of a sentences may be change by a single word.

Question: How can I apply for a Credit Card in Hong Kong before leaving my home country?
Context found: After you leave your home country and move to Hong Kong, you can apply for a Credit Card by visiting one of our branches...   **Correct context:** You can apply for a Credit Card even before you leave your home country. Fill out an application and provide all the necessary documents and we will review your application...

Experiment: Limitations

Extractive QA model: The model could not output the answer in correct format in some cases.

Case1: The reasoning logic is implicit.
Context: Many words have more than one meaning; we have to select the meaning which makes the most sense in context.[32]
Question: What kind of words we need to select the meaning of them?
Answer: Many words **Correct answer:** words that have more than one meaning
Case2: Answers directly extracted from context are not complete or miss some restrictions
Context: Machine translation Automatically translate text from one human language to another.[32]
Question: Machine translation translate text into what?
Answer: another **Correct answer:** another human language

05 Conclusion

Conclusion

01 Different NLP theories are studied and different QA algorithms and model are investigated

02 Build a QA system based on extractive QA, and use a question decomposition model to improve its performance

03 System are evaluated and limitations are found

Future work

01 Solving the problems of the current system

02 Improving the system to a chatbot system

03 Exploring some practical use of our system

06

Demo

Q&A

Thank You!

Reference

[1] G. Caldarini, S. Jaf, and K. McGarry, “A literature survey of recent advances in chatbots,” Information, vol. 13, no. 1, p. 41, 2022.

[2] I. Pandorabots, “Chat with me!,” *Kuki*. [Online]. Available: https://chat.kuki.ai/chat.

[3] B. Chan, T. Möller, M. Pietsch, and T. Soni, “Deepset/Roberta-base-squad2 · hugging face,” *deepset/roberta-base-squad2 · Hugging Face*. [Online]. Available: https://huggingface.co/deepset/roberta-base-squad2.

[4] Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv:1810.04805* (2018).

[5] H.-yi Lee, “Elmo, Bert, GPT,” *Youtube*, 07-Jun-2019. [Online]. Available: https://www.youtube.com/watch?v=UYPa347-DdE.

[6]S. Min, V. Zhong, L. Zettlemoyer, and H. Hajishirzi, “Multi-hop reading comprehension through question decomposition and rescoring,” in ACL, 2019. [Online]. Available: https://github.com/shmsw25/DecompRC