VIETNAM NATIONAL UNIVERSITY HO CHI MINH CITY
HO CHI MINH CITY UNIVERSITY OF TECHNOLOGY
FACULTY OF COMPUTER SCIENCE AND ENGINEERING

REPORT

SPECIALIZED PROJECT

DEVELOPING AI CHATBOT APPLICATION: QUESTION REFORMULATION PROBLEM

MAJOR: COMPUTER SCIENCE

COUNCIL : **COMPUTER SCIENCE**
INSTRUCTORS : **QUẢN THÀNH THƠ, PH.D**
: **BÙI CÔNG TUẤN, M.Eng**
SECRETARY :

—o0o—

STUDENT : **ĐẶNG CÔNG KHANH - 2053105**

HO CHI MINH CITY, May 2024

VIETNAM NATIONAL UNIVERSITY HO CHI MINH CITY
HO CHI MINH CITY UNIVERSITY OF TECHNOLOGY
FACULTY OF COMPUTER SCIENCE AND ENGINEERING

REPORT

SPECIALIZED PROJECT

DEVELOPING AI CHATBOT APPLICATION: QUESTION REFORMULATION PROBLEM

MAJOR: COMPUTER SCIENCE

COUNCIL	: **COMPUTER SCIENCE**
INSTRUCTORS	: **QUẢN THÀNH THƠ, PH.D**
	: **BÙI CÔNG TUẤN, M.Eng**
SECRETARY	:

—o0o—

STUDENT	: **ĐẶNG CÔNG KHANH - 2053105**

HO CHI MINH CITY, May 2024

Instructor’s signature

____**Date**:______________

Assoc. Prof. Quan Thanh Tho, Ph.D (Project Instructor)

Faculty of Computer Science and Engineering

Declaration of Authenticity

The author - Đặng Công Khanh, declares that this Capstone Project was composed and implemented entirely by myself under the guidance and supervision of Assoc. Prof. Quản Thành Thơ and Mr. Bùi Công Tuấn at the Faculty of Computer Science and Engineering, Vietnam National University - Ho Chi Minh City University of Technology.

In the process of researching and implementing this Specialized Project, the author have referenced multiple previous studies from other authors. All of them have been fully and clearly stated in the References part.

This Specialized Project has not been published in any form under any circumstances.

The author,
Đặng Công Khanh

Acknowledgement

First and foremost, the author would like to express the most sincere gratitude to Associate Professor, Ph.D. Quản Thành Thơ and Mr. Bùi Công Tuấn. Thank you for allocating your valuable time to guide, give advice, orientate, and encourage me throughout the outline and implementation phase of this Specialized Project.

The author would also like to thank Mr. Phạm Quốc Nguyên and Mr. Lê Minh Khôi for the advice and questions on the aspects which the author did not think about carefully or did not know that was needed. Your questions and advice were a great help.

Next, the author is grateful that the lecturers at the Faculty of Computer Science and Engineering, Vietnam National University - Ho Chi Minh City University of Technology have delivered us sufficient fundamental knowledge, which serves as the springboard so that the author can compose this specialized project and contribute to Vietnamese Computer Science field.

In addition, the author would like to cheer himselves for the efforts when full-time study and project composition happen simultaneously. Although making the first steps was not easy, the author did not give up. By constructing a sensible plan and putting in the best efforts, the author were able to keep the pace and finish the project just in time.

Last but not least, the author would like to thank his family for always supporting him on the journey at university. My parents' diligence has become the largest source of motivation for him to work hard. Furthermore, seniors at the authors company also conscientiously guide and train the author in a professional working environment when taking part in practical application projects, which really shapes the individual thoughts and behaviors to be better.

The author,

Đặng Công Khanh

Abstraction

Artificial Intelligence (AI) applications, particularly chatbots, have become integral in modern society, driven by the need for efficient customer service solutions. In the wake of successful introductions like ChatGPT, many companies are actively investing in AI applications to align with customer preferences and market trends.

During this 4.0 revolution, to extend their popularity to more customers, many organizations have gone online by both making their accounts on social networks and making their organization's website. These accounts and websites usually have admins and mods to help or answer customers' questions anytime they are needed. For these admins and mods to be able to answer the questions, they have to be trained with knowledge about the company's production and spend time practicing answering questions. Yet not every employee performs the same, and they cannot answer questions all the time, so they have to take shifts. To reduce the cost of training, hiring, and performance, chatbot is a solution.

However, the adoption of chatbots presents challenges, particularly in linguistic diversity and accuracy. While numerous models cater to English, there's a noticeable scarcity of effective solutions for languages like Vietnamese. Moreover, inherent ambiguities in user queries often lead to incorrect responses, undermining user satisfaction. To address these challenges, one promising approach is the utilization of question reformulation methods. By refining user queries, these methods enhance the chatbot's comprehension, thereby improving information extraction and response accuracy.

The objective is to find a good reformulation method to help the model better understand the question, therefore increasing the confidence when extracting information to answer and ensuring more satisfactory answers to the user.

Contents

List of Tables

List of Figures

1

Introduction

In Chapter 1, the author will present an overview of the urgency of the topic, objectives, and scope of the project's research. The report outline will also be presented.

Contents

1.1 Problem statement

The advancement of artificial intelligence (AI) technologies, particularly in the domain of natural language processing (NLP), has revolutionized human-computer interactions. AI-powered chatbots have emerged as strong tools in various domains, facilitating efficient communication between users and systems. However, despite their widespread adoption, challenges persist in achieving powerful understanding and interpretation of user queries. Addressing this challenge requires innovative approaches in NLP research to enhance chatbots' ability to analyze and respond accurately to user input.

Because people use language in many different ways, it can be hard for chatbots to understand questions, and they often end up giving answers that aren't quite right. This error can happens for many reasons (e.g. the user did not express their intent clearly, ...). One reason that this happens is people sometimes don't finish their sentences, leaving out words or phrases because they think the chatbot will understand what they mean without having to repeat everything. They do this to make the conversation faster and avoid saying the same things over and over again. For example, someone might ask, "How do I get to that market?" without saying which market they mean. To understand what they're asking, the chatbot has to remember what was said before, like if they previously asked about a market called "Ha Dong". Then, the chatbot can figure out what they mean by "that" market.

But even if the chatbot knows what was said before, it's still hard to understand the current question completely. The chatbot has to connect the old information to the new question and figure out which details to add to make sense of it. Choosing the right details to add isn't easy, because the chatbot has to make the question clearer and understands the intention of the user. So, dealing with unknown intent sentences in questions is a tough problem for chatbots. But by finding better ways to understand and respond to these kinds of questions will make chatbots easier to use and more helpful in conversations with people.

1.2 Goals

This project is a part of a bigger plan to make a question-answering chatbot that helps university students to communicate with in Vietnamese language. The author aims to make it easier for students to get the information they need quickly and easily.

The main focus is on making the chatbot better at understanding questions from students. This can be done by getting it to look at the questions before answering them. This way, the chatbot can understand what the student is asking more accurately.

Basically, the goal is to make sure that when students ask questions, the chatbot understands them well and gives helpful answers. This can be achieved by changing the questions a bit to make them clearer if needed, but the important parts will remain unchanged. This will make it easier for students to use the chatbot and get the information they're looking for about university rules in Vietnamese, making the question less ambiguous while preserving the semantics of the question.

1.3 Scope

The project will involve the design and implementation of a reformulation model that contributes to the field of question reformulation as follows:

- Question Reformulation Model: The primary focus will be on developing a Vietnamese language question reformulation model. This model will employ techniques such as conference and ellipsis resolution to enhance the clarity and precision of user inquiries, thus improving the effectiveness of the chatbot system.
- Preprocessing Agent: the preprocessing agent tasked with optimizing user questions before they enter the chatbot. This agent will streamline the question processing workflow, making inquiries more understandable and easier to answer. By optimizing the input data, the overall functionality and user experience of the chatbot system is aimed to be enhanced.

1.4 Thesis structure

This Capstone Project comprises 6 key chapters. Chapter 1 provides a concise overview of the problem, detailing the approach and idea for solving it In Chapter 2, the thesis delves into foundational knowledge regarding the technologies integral to this solution. Chapter 3 focuses on presenting the authors' model design and analysis highlighting their strengths and weaknesses. As well as provides a detailed account of the implementation process for the model. This includes insights into building the coreference module. Chapter 4 discusses the methodology for system evaluation and presents the result for the final product of the author. Chapter 5 is about deployment and Chapter 6 addresses potential improvements and challenges encountered by the author during the course of the thesis.

2

Reformulation Domain Knowledge

In Chapter 2, the author will provide domain knowledge on Question Reformulation. Specifically, the author will show the related works and explain why the normal QA model needs further question reformulation. The author will go on to elaborate on different methods of reformulation: coreference resolution, ellipsis (anaphora) resolution and a little bit of abstraction about question decompose and recompose. Then, the author will explain how coreference resolution and ellipsis resolution reformulate questions, thus helping the model better understand the questions. Finally, the author points out where the reformulation model helps the QA chatbot, especially chatbot with Knowledge Graph (KG) answer questions more correctly and informatively.

Contents

2.1 Related works

Before any serious development, it is worth noting that there are many existing solutions that partially or fully solve the existing problems. Over the years, various approaches and techniques have been developed to address the challenges associated with question reformulation, each contributing to the advancement of this domain. Until the introduction of deep learning, which marked a significant milestone in the field. Techniques such as sequence-to-sequence (Seq2Seq) models and transformers revolutionized question reformulation. These models, particularly those based on architectures like BERT and GPT.

The co-reference resolution aims to link an precedent for each possible mention. The work of Kantor and Globerson, 2019[1] solved the problem in an end-to-end fashion by jointly detecting mentions and predicting co-references. More recently, Zheyu Ye, Jiangning Liu, Qian Yu and Jianxun Ju, 2021[2] propose an action based network (ActNet), which is composed of three modules: First, the encoder component to encode the input text. Second, the detecting component to detect the positions of co-reference and ellipsis in the current question as well as the corresponding actions. Third, the comprehension component to find the related co-referential or omitted information from context history.

2.2 Reformulation in Natural Language Preprocessing field

In the context of Natural Language Processing (NLP), reformulation refers to the process of modifying or rephrasing text or queries to improve their clarity, precision, or relevance for downstream tasks.

Recent research has seen the integration of reformulation techniques with other NLP tasks like paraphrase detection, semantic parsing, and context-aware question answering. These integrated approaches aim to develop more adaptable and powerful systems capable of effectively processing a wide range of user inputs. By combining multiple

NLP tasks, these systems can achieve higher accuracy and better user experience.

Within the broader scope of reformulation, specific challenges such as coreference resolution and ellipsis adding have collected considerable attention. Coreference resolution involves in identifying and linking pronouns and other referring expressions to their correct antecedents, which is very vital for maintaining coherence in reformulated sentences. Ellipsis adding, on the other hand, focuses on interpreting and filling in omitted information based on contextual clues.

2.3 Types of Reformulation

Reformulation techniques aim to enhance text or queries in various ways. Below are some methods of reformulation that the author had found.

2.3.1 Query Expansion

Adding additional terms or synonyms to a query to retrieve more relevant information.

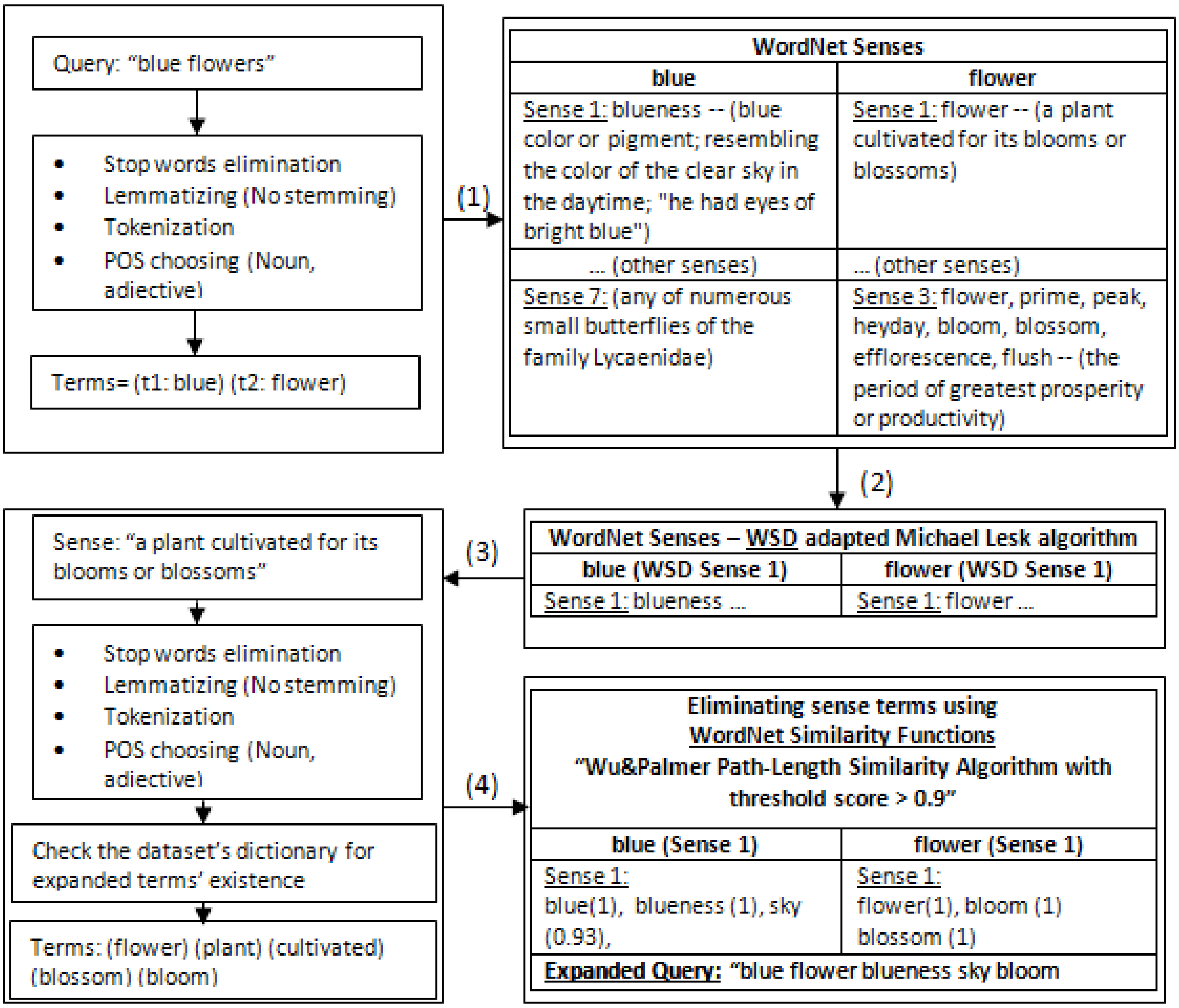
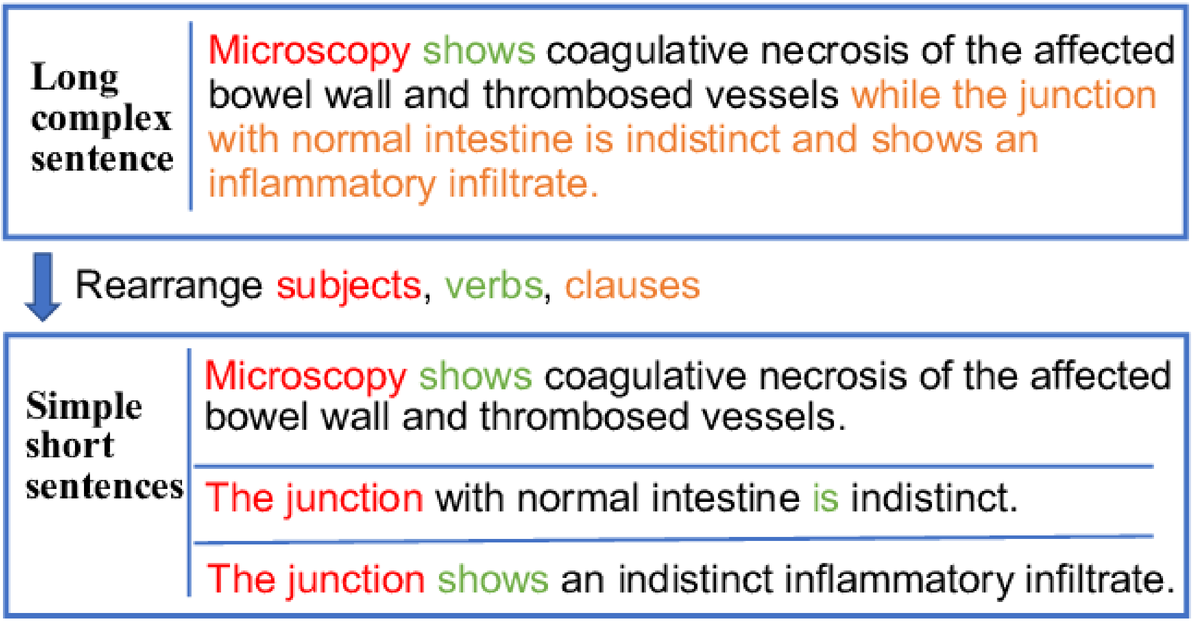

Query: "blue flowers"
Stop words elimination
Lemmatizing (No stemming)
Tokenization
POS choosing (Noun, adjective)
Terms= (t1: blue) (t2: flower)
(1)
WordNet Senses
blue
flower
Sense 1: blueness -- (blue color or pigment; resembling the color of the clear sky in the daytime; "he had eyes of bright blue")
Sense 1: flower -- (a plant cultivated for its blooms or blossoms)
... (other senses)
... (other senses)
Sense 7: (any of numerous small butterflies of the family Lycaenidae)
Sense 3: flower, prime, peak, heyday, bloom, blossom, efflorescence, flush -- (the period of greatest prosperity or productivity)
(2)
WordNet Senses – WSD adapted Michael Lesk algorithm
blue (WSD Sense 1)
flower (WSD Sense 1)
Sense 1: blueness ...
Sense 1: flower ...
(3)
Sense: "a plant cultivated for its blooms or blossoms"
Stop words elimination
Lemmatizing (No stemming)
Tokenization
POS choosing (Noun, adjective)
Check the dataset's dictionary for expanded terms' existence
Terms: (flower) (plant) (cultivated) (blossom) (bloom)
(4)
Eliminating sense terms using WordNet Similarity Functions
"Wu&Palmer Path-Length Similarity Algorithm with threshold score > 0.9"
blue (Sense 1)
flower (Sense 1)
Sense 1: blue(1), blueness (1), sky (0.93),
Sense 1: flower(1), bloom (1) blossom (1)
Expanded Query: "blue flower blueness sky bloom

Figure 2.1: Query Expansion Example. (Source: [12])

2.3.2 Sentence Simplification

Restructuring complex sentences into simpler, easier-to-understanding forms without altering the original meaning. This aids in improving readability and comprehension.



Long complex sentence

Microscopy shows coagulative necrosis of the affected bowel wall and thrombosed vessels while the junction with normal intestine is indistinct and shows an inflammatory infiltrate.

Rearrange subjects, verbs, clauses

Simple short sentences

Microscopy shows coagulative necrosis of the affected bowel wall and thrombosed vessels.

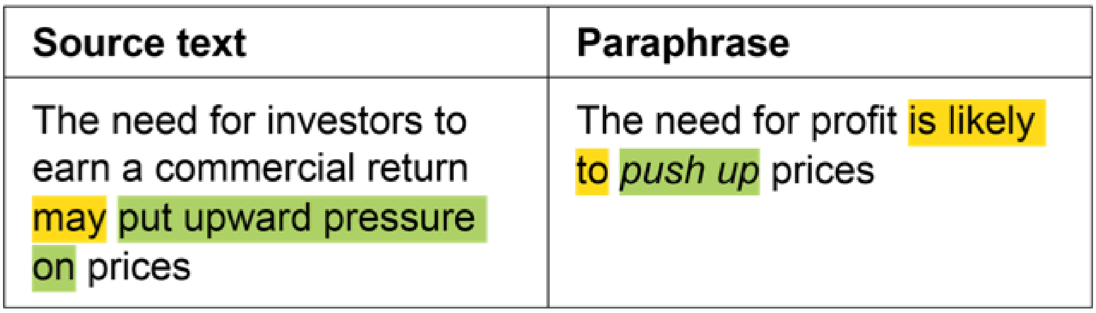
The junction with normal intestine is indistinct.

The junction shows an indistinct inflammatory infiltrate.

Figure 2.2: Sentence Simplification Example. (Source:[13])

2.3.3 Paraphrasing

Expressing the same meaning using different words or sentence structures. Paraphrasing can be a great help in text summarization, machine translation, and question answering.

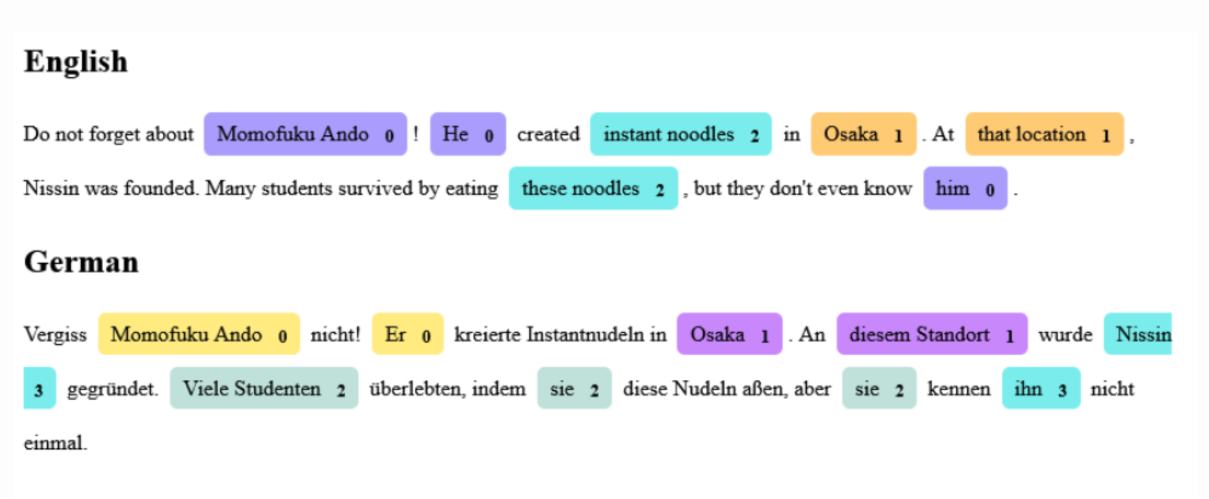


Source text	Paraphrase
The need for investors to earn a commercial return may put upward pressure on prices	The need for profit is likely to *push up* prices

Figure 2.3: Paraphrasing Example. (Source: [14])

2.3.4 Coreference Resolution

Resolving references to the same entity across a text. For instance, identifying that "he" refers to "John" in a subsequent sentence. Notice, this is the text coreference resolution, which is more general than the method coreference resolution that will be discussed later on.



English

Do not forget about Momofuku Ando 0 ! He 0 created instant noodles 2 in Osaka 1 . At that location 1 , Nissin was founded. Many students survived by eating these noodles 2 , but they don't even know him 0 .

German

Vergiss Momofuku Ando 0 nicht! Er 0 kreierte Instantnudeln in Osaka 1 . An diesem Standort 1 wurde Nissin 3 gegründet. Viele Studenten 2 überlebten, indem sie 2 diese Nudeln aßen, aber sie 2 kennen ihn 3 nicht einmal.

Figure 2.4: Coreference resolution example. (Source:[15])

2.3.5 Question Reformulation

Modify questions while preserving their intent. This is valuable in improving the accuracy of search engines or assisting chatbots and virtual assistants in understanding user queries better. The detail will be discussed in the next section.

2.4 What is Question Reformulation

The question reformulation or question de-contextualization problem is the task of rephrasing a question or removing the context from the question without changing its original meaning to help the computer better understand the user's intent.

2.5 Why we need Question Reformulation

Question Reformulation technique is often used in information retrieval tasks, and is usually seen in question-answering models to:

- Refine the question to improve search results: by changing the formation to make the question easier to understand or easier to extract the information, therefore retrieving more relevant and precise results.
- Make better communication between user and chatbot by overcoming the language ambiguity, complex sentence structure of the question, or handling synonyms, word variations, or different ways of user asking the same question.
- Assist in some other tasks by altering the wording or structure of the initial query.

2.6 Types of Reformulation

2.6.1 Coreference Resolution

Coreference resolution is the task of grouping mentions into entities. Thus, deciding whether to assign a given mention to a candidate entities[1]. Consider the questions:

"Hãy cho tôi biết phương thức xét tuyển 2 là gì?"
("What is the admission method of the second assessment?")
"Ngày xét tuyển của phương thức này là ngày mấy"
("What is the admission evaluation date of the assessment?")
“Điểm chuẩn phương thức trên của khoa cơ khí là bao nhiêu”
("What is the admission score of the above assessment for the mechanical engineering department?")

QA model of chatbot alone can not understand the word "the assessment" and the above assessment" of the second and third sentence since they refer to the previously mentioned "second assessment". The reason is that while they can greatly encode the questions and documents and find the most possible answers they can do it only with standalone questions. So if the user's question first refers to an entity where the answer span falls into the sentence in which has coreference to the mentioned entity, it will most likely reduce the confidence and accuracy of the model.

The task of the researched model is to solve this reference problem, making the question easier to understand for the machine, thus increasing the model's confidence in information retrieval and its accuracy.

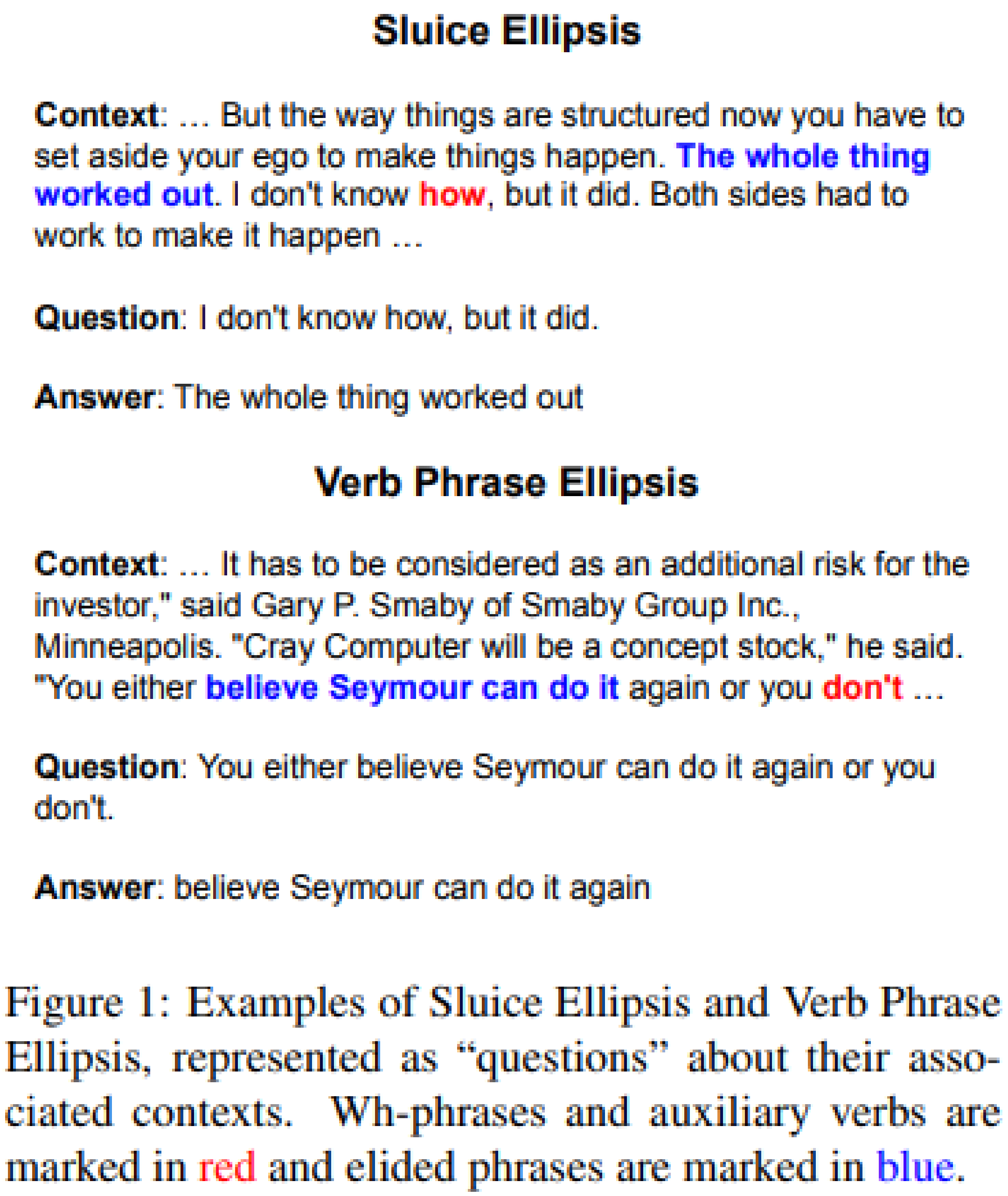
In this project, the author only focus on question resolution using pretrained coreference model and explore its usefulness in enhancing results in QA model in Vietnamese chatbot.

2.6.2 Ellipsis Resolution

Ellipsis is a linguistic phenomenon in which parts of a sentence are omitted, and have to be retrieved from discourse or real-world context.

Ellipsis is a form of anaphora (besides coreference) that often functions to reduce redundancy in language and improve discourse cohesion (Menzel 2017; Mitkov 1999). Languages provide various mechanisms to elide information, based on which different ellipses are defined in linguistics.

Ellipses are not very frequent in text[10] but for improving the accuracy of Natural Language Processing (NLP) systems that handle data with ellipses, they are important (Zhang et al. 2019; Dean, Cheung, and Precup 2016).[11]



Sluice Ellipsis

Context: … But the way things are structured now you have to set aside your ego to make things happen. **The whole thing worked out**. I don't know **how**, but it did. Both sides had to work to make it happen …

Question: I don't know how, but it did.

Answer: The whole thing worked out

Verb Phrase Ellipsis

Context: … It has to be considered as an additional risk for the investor," said Gary P. Smaby of Smaby Group Inc., Minneapolis. "Cray Computer will be a concept stock," he said. "You either **believe Seymour can do it** again or you **don't** …

Question: You either believe Seymour can do it again or you don't.

Answer: believe Seymour can do it again

Figure 1: Examples of Sluice Ellipsis and Verb Phrase Ellipsis, represented as "questions" about their associated contexts. Wh-phrases and auxiliary verbs are marked in red and elided phrases are marked in blue.

Figure 2.5: Ellipsis resolution example. (Source:[16])

3

Coreference Resolution Model

In Chapter 3, the author will describe the coreference resolution model that was utilized in this research. This section will include an exploration of key architecture that was adapted and its application in the developing of HCMUT chatbot.

Contents

3.1 The Goal of The Model and the position in QA system

As stated before, the goal in this project is to try to use a pretrained coreference resolution model to de-contextualize the questions from the user to improve the performance of the QA model (in this case the LLM chatbot model of HCMUT Admission Office). The first step for the model to achieve this goal is to be able to coreference the questions given by the user based on the chat history or what refered as the context.

The coreference agent sits before the chatbot agent doing the "preprocessing" job on the questions before feeding them to the chatbot. Assumed the chatbot is a black box and we do not have access to its internal component, we also assume that the information generated from the chatbot is 100% correct but additional information might be required by the user.

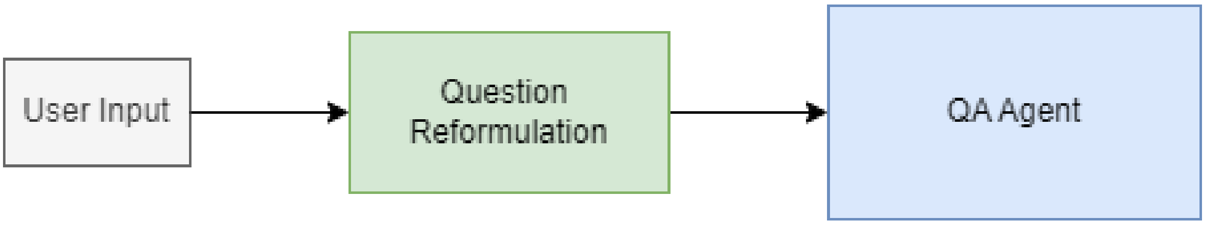


Figure 3.1: The model sits before the chatbot.

To the advice of Professor Quan Thanh Tho, the model consists of 2 layers. The first layer is the coreference resolution model. This model will take the query from the user, along with the chat history (the context), and try to replace the pronouns in the query (if any) with the correct entities extracted from the context. After that, the coreference agent will pass the reformulated question to the QA model to generate the answer.

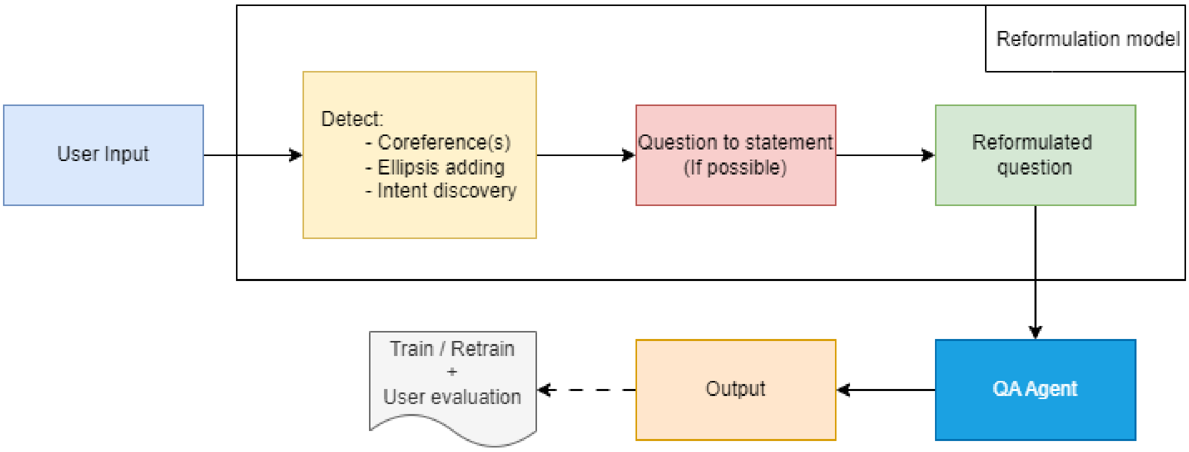


Figure 3.2: The flow of the data in coreference model.

3.2 Methodology

In this section, the author will introduce the model in detail. Formally speaking, given an utterance sequence $U = \{u_{-1}, ..., u_{-k}, ...u_{-n}\}$ where each utterance is a pair of coreferenced, English-translated questions Q_{-k} and answers A_{-k} at the k-th turn before the current question. The author denoted the current turn question and answer as Q_0, A_0 or Q, A for short.

Since the adopted model *en_core_web_sm* model works best for English and similarly structured languages, which Vietnamese does not, so a translation model have to be used to translate the questions and answers to English when storing to the history and to Vietnamese before passing the coreferenced query to the chatbot.

3.2.1 Translation Module

As a preprocessing process before coreferencing, the author adopted *facebook/mbart-large-50-many-to-many-mmt*[3] model to translate the Vietnamese queries and answers to English, the coreferenced query to Vietnamese, and formalized the processes into the formulas:

$$Q_E = F_{VE}(Q)$$

$$Q_{VCo} = F_{EV}(Q_{ECo})$$

$$A_E = F_{VE}(A)$$

where Q_E, Q_ECo Q_{VCo} are the English, English coreferenced and Vietnamese coreferenced queries respectively, A_E is the English-translated answer, F_{VE}, F_{EV} are translation models that translate its input from Vietnamese to English and from English to Vietnamese.

3.2.2 Coreference Module

In this module, the author adopt a SpaCy model, *en_core_web_sm* for both fast and accurate coreference. The English, concatenated string of query and history will be coreferenced by the model. In this model, the history will be the context to co-refer the translated question: $Q_ECo = ENCoref([U; Q_E]) - U$ where the $ENCoref$ is the coreference model, $[\cdot;\cdot]$ is the string concatenation operation.

3.3 Implementation

3.3.1 crosslingual-coreference library and SpaCy models

crosslingual-coreference library

crosslingual-coreference is a Python library that supports the coreferencing of documents between languages using SpaCy models with the assumption that cross-lingual embeddings should work for languages with similar sentence structures[7]. This library also supported SpaCy API to perform the coreference resolution operation

Spacy models

spaCy is a library for advanced Natural Language Processing in Python and Cython. It's built on the very latest research, and was designed from day one to be used in real products.[8]

spaCy comes with pretrained pipelines and currently supports tokenization and training for 70+ languages. It features state-of-the-art speed and neural network models for tagging, parsing, named entity recognition, text classification and more, multi-task learning with pretrained transformers like BERT, as well as a production-ready training system and easy model packaging, deployment and workflow management. spaCy is commercial open-source software, released under the MIT license.[8]

SpaCy pipelines, models, and APIs are easy to use. Along with its speed and accuracy, many researchers, developers, and practitioners in NLP-related fields use SpaCy. SpaCy also has a rich ecosystem, good documentation, and active community support, which makes it beginners-friendly.

en_core_web_sm Model

The *en_core_web_sm* is a SpaCy model with many supported pipelines suitable for NLP-related operations. This project adopted this model and combined it with the crosslingual-coreference library API to perform the resolution to the concatenated string of history and translated current query and the knowledge text input to the QA agent.

facebook/mbart-large-50-many-to-many-mmt

a Facebook's multilingual translation model, created using the original mBART model and extended to add extra 25 languages to support multilingual machine translation models of 50 languages[9] because of its popularity and fast translation. In this project, the author adopted this model to replace the QA agent in answering questions with given documents.

3.3.2 Implementation

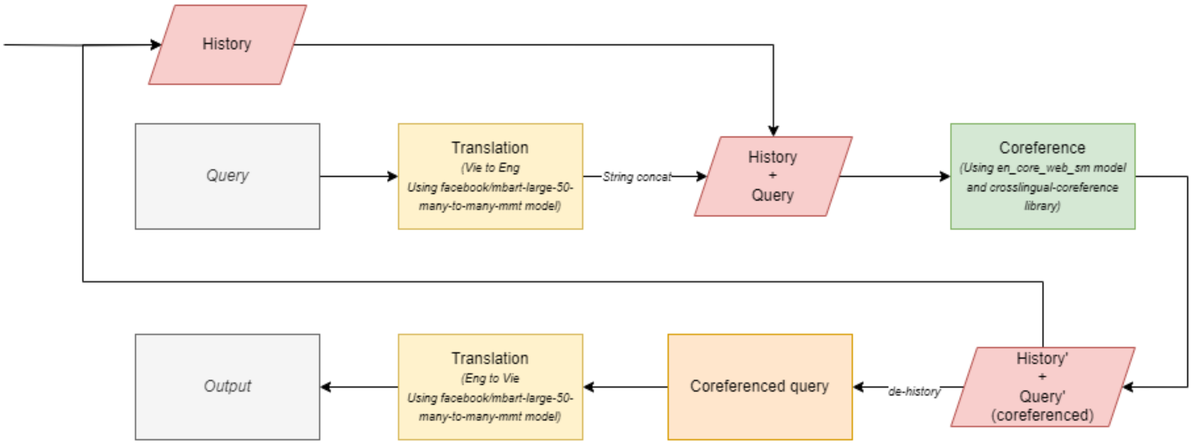

History
Query
Translation
(Vie to Eng
Using facebook/mbart-large-50-many-to-many-mmt model)
String concat
History
+
Query
Coreference
(Using en_core_web_sm model and crosslingual-coreference library)
Output
Translation
(Eng to Vie
Using facebook/mbart-large-50-many-to-many-mmt model)
Coreferenced query
de-history
History'
+
Query'
(coreferenced)

Figure 3.3: The model model pipeline with a pseudo QA agent.

The input of the model is a user's Vietnamese query. In the beginning, the input went through a translation model. In this research, *facebook/mbart-large-50-many-to-many-mmt*[9] is used to translate the query to English for a better coreference result as the *en_core_web_sm* model works best for English and similar structured languages.

After having translated the query, it is string-concatenated with the history to perform the coreference. In the beginning, the history is an empty string.

The coreferenced string will then be separated into two parts:

- The English coreferenced query, which is copied from the string will go through the translation model to translate into Vietnamese to pass to the chatbot model
- The coreferenced string (the concatenated history and query) will then be concatenated with the answer received from the chatbot to form a new history.

After the translation, the model will then return a Vietnamese answer, along with coreferenced query as an output.

3.4 Try to improve by creating Vietnamese coreference model

3.4.1 Design

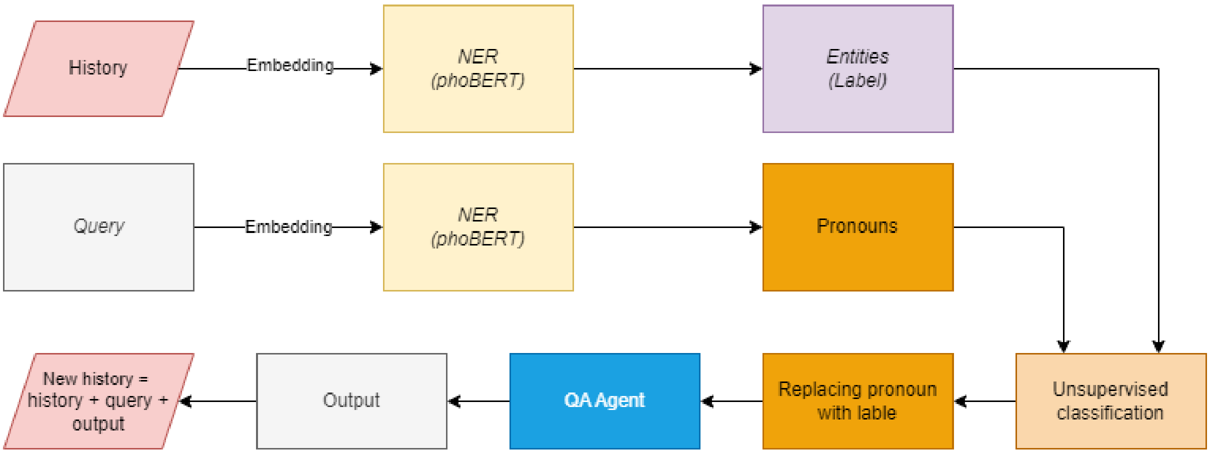


Figure 3.4: The Vietnamese model pipeline with a pseudo QA agent.

The author tried to improve the efficiency of the model by implementing a coreference resolution model that runs on Vietnamese corpus.

3.4.2 Idea

The idea was using the coreferenced history along with the inputted question, using the phoBERT model to perform Named Entity Recognition task to extract the all the appropriate entities from the history, using as labels. After that, performing Named Entity Recognition to extract the pronouns from the question to perform unsupervised classification with the labels to decide which entity to replace the pronoun. The following out the QA agent then can be execute without the need of using the translation model.

The author expected the improvement can decrease the execute time of the model significantly as the time-consuming translation model is now not in use.

3.4.3 Difficulties

Due to time limitation and inefficiency, the author were unable to put the improvement model into use. Also face some difficulties such as does not know the effective way to extract the label effectively due to the overlap property of coreference resolution problem, or still on research of a good way to perform the unsupervised classification task.

4

Result

In this chapter, the author will demonstrate the result of the adapted coreference resolution model in a QA model.

Contents

4.1 BERTSCORE

BERTSCORE is an automatic evaluation metric for text generation. Analogously to common metrics, BERTSCORE computes a similarity score for each token in the candidate sentence with each token in the reference sentence.
However, instead of exact matches, it computes token similarity using contextual embeddings. It evaluates using the outputs of 363 machine translation and image captioning systems. BERTSCORE correlates better with human judgments and provides stronger model selection performance than existing metrics. Finally, BERTSCORE is more robust to challenging examples when compared to existing metrics.[4]

In this project, the author decided to use BertScore as the main measure to evaluate the similarity of the answers output by the QA model to show that the co-reference model helps increase the confidence of the model and the information retrieved.

4.2 Result with Admission Dataset given by the Admission Office

The experimentation involved transforming and retesting the accuracy of the cross-lingual co-reference model using BertScore on specific data. The outcome demonstrated an improvement in sentence similarity by 0.6%, elevating it from 0.9300523887981068 to 0.9362728541547601. However, further validation is essential due to potential data

misalignment with the task requirements for co-reference. Co-reference needs historical conversational data where pronouns substitute nouns. However, the available data is a one-question-one-answer format, lacking the conversational context. Therefore, as a temporary measurement, the author self-handedly co-referenced and corefered the data as the context and concatenated string of history and question. After that, translating them into English. Thus, further experimenting with other data processing techniques or exploring alternative data-sets becomes crucial for certainty.

In conclusion, while the BertScore resulted in promising in elevating sentence similarity, the need for further validation still exists. The mismatch between data characteristics and co-reference requirements underscores the necessity for diverse data sources or refined data processing methods to ensure the consistency of the model's performance.

4.3 Result with the CoQA Dataset and roberta-base-squad2 model of deepset from Hugging Face Hub

The final aim was to assess the model's functionality using the CoQA dataset since a Vietnamese dataset was unavailable. Therefore, the author also skipped the question translation step, instead, only translate the human input answer. The model employed for this evaluation acted as a QA system using deepset/roberta-base-squad2. The result was small but still positive, enhancing the confidence of the model's answers by over 0.7% (from 0.360224369795273 to 0.36732871451948484) under the condition that the data underwent co-reference beforehand. This evaluation included of 10 dialogue segments and 141 questions for demonstration purpose and processing time estimation for later tests. The author hoped that with the help of Knowledge Graph (KG) the co-reference process for the data is no longer needed.

The confidence level in the model's answers witnessed an uplift, meeting the target increase of over 0.7%. However, the similarity between the original and extracted answers slightly decreased to 0.8179846718197777 from 0.8258722265561421 according to BertScore. Manual inspection revealed that the reduced similarity in the answers stemmed from the QA system retrieving more information and the pronouns got replaced by the entities, regardless of co-reference application. Both co-referenced and non-co-referenced instances exhibited similar discrepancies. However, at some point, the model did not work as well as expected. This shows that the project also need other methods to achieve better reformulation result.

For example, under the same condition of the QA model have not been train for the CoQA dataset:

- The question: "Did they want Cotton to change the color of her fur?" resulted in the model without coreference (Un-Coref) returning the answer "I only wanted to be more like you", while the model with coreferenced model (Coref) returned the answer "5 other sisters wouldn't want that", which math the result of "no".

- The question: "What guided RJ home?" resulted in the model Un-Coref returning the answer "his father's flashlight", the Coref model returned the answer "Reginald Eppes's father's flashlight". Both answers are correct to the result "The flashlight", but the answer of Coref model has more information than the Uncoref model.
- Sometimes, both models can be wrong in retrieving the answer, like in the question "What kind of dishes does she bring?", the Un-coref model returns the answer "Ipad", while the Coref model returned the answer "Lucy can hardly do dishes in return". Both are wrong from the result "hot soup and a container with rice, vegetables and either chicken, meat or shrimp, sometimes with a kind of pancake". However, the Coref model seems to have helped the QA model extract more accurate information.

Although there are also cases where the Coref model returned the wrong answer while the Un-coref model returned the right one, it is suspected that due to the incorrect coreference of the question or the searched document. This problem requires further research into different ways to reformulate the question and handle the document.

In general, the assessment showed an increase in the model's answer confidence, meeting the specified threshold. However, a slight decline in the similarity between original and extracted answers was observed, primarily due to the QA system's tendency to retrieve more information, regardless of co-reference application. The model's ability to handle pronoun-embedded questions indicates its utility in extracting additional details. Additionally, the failure of the model in some occasions suggested the author to explore methods to eliminate context co-reference dependency, potentially leveraging knowledge graphs, finding other ways to handling questions and data, opening a pathway for future developments.

Model advantages while testing with this dataset:

- Proficiency in handling questions embedded within passages containing pronouns.
- Enhanced extraction of additional information; nevertheless, a summarization model might still be necessary for users seeking more concise answers.
- Potential exploration to obviate the need for context co-reference, such as leveraging knowledge graphs (KG) to enrich the data pool.

5

Further improvement and conclusion

In the following chapter, the author will engage in a comprehensive exploration of avenues for further improvement within the context of the thesis. This discussion will encompass a detailed examination of potential enhancements, refining aspects of the research to bolster its overall depth and effectiveness

Contents

During the length of the thesis, the author briefly mentioned some improvements when discussing many of the functionalities of the system. The author would like to revisit some of those improvements, as well as discussing new potential improvements.

5.1 Lacking real data

Although the author received a dataset from the Admission Office containing questions that students asked in real-time chats with university technicians, it was challenging to fully utilize this dataset. The dataset was very noisy, and the data is preprocessed - which make significant information removed, rendering it incomplete. Furthermore, there was no metric available for the author to evaluate the coreferenced questions accurately.

5.2 Vietnamese Coreference Model

The author attempted to enhance the efficiency of the model by implementing a coreference resolution model that operates on a Vietnamese corpus. This improvement would significantly reduce the execution time of the model, as it would eliminate the need for the time-consuming translation model.

However, due to time constraints and inefficiencies, the author was unable to implement the improved model on time. Additionally, challenges were faced in effectively extracting labels due to the overlapping nature of the coreference resolution problem and ongoing research into finding an effective method for the unsupervised classification task.

So instead, the translation module is enhanced to make the model run slightly faster. Overall, it makes the model consistently runs query with less than 0.5 seconds per query. Which is quite acceptable in the authors' opinion.

5.3 Intent Discovery

Integrating intent discovery into a coreference model involves combining techniques from natural language understanding (NLU) and natural language processing (NLP) to enhance the model's ability to comprehend and predict user intents along with resolving references.

By combining intent discovery with coreference resolution, the model can provide more contextually accurate and relevant responses, enhancing the overall user experience. This integrated approach allows the system to better understand the user's underlying goals and reference points within their queries.

5.4 Conclusion

Question Reformulation is not a new problem in the NLP field. However, due to the popularity of LLMs, in-depth research on the problem is not many. Reformulation refers to the process of modifying or rephrasing text or queries to improve their clarity, precision, or relevance for downstream tasks. The question reformulation problem is the task of rephrasing a question or removing the context from the question without changing its original meaning to help the computer better understand the user's intent. The solution to question reformulation problem promises a way to make the question-answering model in the chatbot to be lighter, more robust and better at understanding the user's question.

The assessment of the model across various datasets has provided valuable insights. The crosslingualcoreference model showcased a commendable 0.6% increase in sentence similarity, demonstrating its potential for improvement. However, the dataset is not perfectly optimal for testing, meaning that it needed to be check more to be sure the results are reliable.

Similarly, the evaluation using the CoQA data-set reflected promising outcomes in enhancing answer confidence by over 0.7%. Nevertheless, a slight reduction in the similarity between original and extracted answers indicates the need for a deeper investigation into the model's information retrieval mechanisms.

Positively, the model exhibit strengths in specific areas, such as handling pronoun-embedded questions and co-reference resolution. Both assessments highlight the necessity for refining data processing techniques and exploring alternative data-sets to reinforce the models' reliability and effectiveness.

6

Task Evaluation and Task Description

Chapter 6 will be used to discuss the project evaluation, providing insights to how the author evaluated the outcome of the project. As well as describing all the tasks that the author has done in order to complete this project.

Contents

6.1 Project Evaluation

Throughout the project implementation, the author effectively employed an agile methodology, managed to finish the project through iterative sprints, each lasting approximately two weeks. At the conclusion of each sprint, the author conducted a comprehensive meeting to review the progress and strategize for the upcoming sprint.

The author communication with Assoc. Prof. Quan Thanh Tho and Mr. Bui Cong Tuan played a pivotal role. Their guidance and support had assisted in overcoming challenges, and they facilitated connections with stakeholders who supplied valuable real-world data and feedback. This collaboration enhanced the authors' understanding of the project, enable the author to align with stakeholder expectations.

Adopting an iterative and incremental approach, the author developed the project, focusing on delivering a minimum viable product with essential features and functionalities. Regular testing and evaluation were integral to maintaining the project's quality and performance. This experience provided valuable lessons, including the utilization of diverse technologies, effective teamwork, communication skills, and adept time and resource management. The author take pride in the project and aspire to contribute positively to the coreference field.

6.2 Task Description

6.2.1 Defining the problem and exploring potential solutions

Task	Description
Define the problem	Meetings with the instructors to comprehend the issue at hand about its significance, and the underlying reasons for the project necessity.
Researching	Involved an in-depth exploration of information related to the project's problem, drawing insights from diverse resources like academic papers and relevant data sources. To lay the groundwork for effective problem-understanding
Design workflow	Utilizing the identified problem, mapping out the comprehensive project pipeline, and outlining the project's current position within the larger project framework.
Explore potential solutions	Engaging in discussions with the instructors to receive guidance on project direction and technology choices for the overarching project.

Table 6.1: Table of task evaluation on defining the problem and exploring potential solutions

6.2.2 Developing the model

Task	Description
Selecting the library	Engaging in meetings with the instructors, reviewing academic papers to make an informed decision on selecting an optimal library.
Coding and Running Small Tests	Library experimental by coding and conducting small-scale tests to ensure the selected library's suitability for the project requirements.
Data preprocessing	Generating expected questions from the data-set's existing questions to provide further evaluation and insight into the model's performance.
Complete the code and run the data-set	Finalizing the coding process and executing it on the data-set for comprehensive analysis and evaluation of the model's functionality.
Evaluate result	Utilizing BertScore for assessing the model's efficacy, precision, and confidence.

Table 6.2: Table of task evaluation on Developing the model

References

[1] Ben Kantor, Amir Globerson (2019), Coreference Resolution with Entity Equalization. Retrieved from https://aclanthology.org/P19-1066.pdf

[2] Zheyu Ye, Jiangning Liu, Qian Yu, Jianxun Ju (2021), Action based Network for Conversation Question Reformulation. Retrieved from https://arxiv.org/abs/2111.14445

[3] Yuqing Tang, Chau Tran, Xian Li, Peng-Jen Chen, Naman Goyal, Vishrav Chaudhary, Jiatao Gu, Angela Fan (2020), Multilingual Translation with Extensible Multilingual Pretraining and Finetuning. Retrieved from https://arxiv.org/pdf/2008.00401.pdf

[4] Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi (2020), BERTSCORE: EVALUATING TEXT GENERATION WITH BERT. Retrieved from https://arxiv.org/pdf/1904.09675.pdf

[5] Zheyu Ye, Jiangning Liu, Qian Yu, Jianxun Ju (2021), Action based Network for Conversation Question Reformulation. Retrieved from https://arxiv.org/pdf/2111.14445.pdf

[6] Sheng-Chieh Lin, Jheng-Hong Yang, Rodrigo Nogueira, Ming-Feng Tsai, Chuan-Ju Wang and Jimmy Lin (2020), Conversational Question Reformulation via Sequence-to-Sequence. Retrieved from Architectures and Pretrained Language Models https://arxiv.org/pdf/2004.01909.pdf

[7] Introduction of crosslingual-coreference library. Retrieved from https://pypi.org/project/crosslingual-coreference/

[8] Introduction of SpaCy library. Retrieved from https://pypi.org/project/spacy/

[9] The adapted translation model. Retrieved from https://huggingface.co/facebook/mbart-large-50

[10] The reported frequency of noun ellipses is 1.99% (Khullar, Majmundar, and Shrivastava 2020); and that of VPE along with related phenomenon is 1% (Bos and Spenader 2011).[11]

[11] Payal Khullar(2021), Are Ellipses Important for Machine Translation?. Retrieved from https://direct.mit.edu/coli/article/47/4/927/106771/Are-Ellipses-Impo

[12] Query Expansion Example. Retrieved from https://www.researchgate.net/publication/228966554_DEU_at_ImageCLEF_2009_

[13] Sentence Simplification Example. Retrieved from https://www.researchgate.net/publication/353485830_Towards_Visual_Questio

[14] Paraphrasing Example. Retrieved from https://www.futurelearn.com/info/courses/en

[15] Coreference Resolution example. Retrieved from https://pypi.org/project/crosslingual-coreference/

[16] Ellipsis Resolution example. Retrieved from https://aclanthology.org/2021.eacl-main.68.pdf