공학석사 학위논문

Incorporation of Dialogue Properties in Decoupling-Fusing Network for Incomplete Utterance Rewriting

대화 속성을 고려한 분리-융합 네트워크 기반의 불완전 발화 다시쓰기 모델 구축

2023년 8월

서울대학교 대학원

협동과정 바이오엔지니어링 전공

민 정 원

Incorporation of Dialogue Properties in Decoupling–Fusing Network for Incomplete Utterance Rewriting

Academic adviser Jinwook Choi

Submitting a Master's thesis of Engineering
August 2023

Interdisciplinary Program in Bioengineering
The Graduate School
Seoul National University

Jeongwon Min

Confirming the Master's thesis written by
Jeongwon Min
August 2023

Chair ______________________________
Hyung Jin Yoon, M.D. / Ph.D.

Vice Chair ______________________________
Jinwook Choi, M.D. / Ph.D.

Examiner ______________________________
Jae Sung Lee, Ph.D.

Abstract

Incorporation of Dialogue Properties in Decoupling-Fusing Network for Incomplete Utterance Rewriting

Jeongwon Min
Interdisciplinary Program in Bioengineering
The Graduate School
Seoul National University

Dialogue-based Incomplete Utterance Rewriting (IUR) represents the task of transforming context-dependent utterances within a dialogue, which require contextual information to be fully comprehended, into self-contained expressions. With the proliferation of AI-based chatbots in recent years, it has become increasingly crucial for dialogue systems to effectively process and understand conversations involving human agents. To this end, IUR has emerged as a promising approach to enhance the overall performance of dialogue systems.

Advancements in Transformer-based Pretrained Large Language Models, such as BERT and GPT, have significantly contributed to the progress in this field. Leveraging their ability to capture contextual dependencies and generate coherent text, these models have paved the way for enhanced performance in IUR. The integration of these state-of-the-art language models into IUR has yielded substantial advancements.

In this study, we present a novel approach to enhance the performance of the Dialogue-based IUR task by leveraging the intrinsic properties of dialogues. By introducing utterance and speaker information into the model and effectively capturing the context and dynamics of the conversation, our proposed method enables the generation of self-contained utterances. Our method contributes to enhancing the overall understanding of the conversation flow and facilitates the generation of coherent and contextually appropriate rewritten utterances. Another key advantage of our method is its ability to achieve state-of-the-art performance without requiring additional labor-intensive annotation of coreference links between the antecedents and referents. This alleviates the burden of manual annotation, making our approach more scalable and efficient compared with previous methods that heavily rely on such annotations.

Through extensive evaluation, we demonstrate the effectiveness of our approach in achieving improved performance, surpassing existing approaches in the field. Overall, our proposed method improves the performance of IUR without the need for labor-intensive coreference link annotations by using the intrinsic properties of dialogues.

Keywords: rewriting, dialogue, utterance, coreference, ellipsis, Decoupling-Fusing, GPT-2

Student Number: 2021-20816

Contents

Abstract **i**

Contents **iii**

List of Figures **v**

List of Tables **vi**

1 Introduction **1**

1.1 Background 1

1.2 Task Definition 2

2 Related Work **6**

2.1 Incomplete Utterance Rewriting 6

2.2 Baseline Model 7

3 Materials and Method **11**

3.1 Datasets 11

3.1.1 CREAD 12

3.1.2 CANARD 12

3.1.3 Task 13

3.1.4 MedDial 13

3.2 Proposed Model 14

3.2.1 Contextualized Representation 16

3.2.2 Decoupling-Fusing Layer 17

3.2.3 Rewriting Classification 23

3.2.4 Speaker Identification 24

3.2.5 Utterance Rewriting 25

3.2.6 Learning Objectives 25

3.3 Experimental Settings 27

3.4 Evaluation Metrics 28

4 Results and Analysis **30**

4.1 Results 30

4.2 Ablation Study 35

4.3 Analysis 36

5 Conclusion **40**

Appendix **48**

A Masks 48

B Utterance Rewriting Annotation Guideline 49

국문초록 **53**

List of Figures

1.1 Example of an incomplete utterance (coreference) in a dialogue. 3

1.2 Example of an incomplete utterance (ellipsis) in a dialogue. 4

2.1 The overall model archietecture of the baseline model `CREAD` [12]. 8

2.2 The datasets constructed by `CREAD` [12]. 9

3.1 The proposed model for dialogue-based incomplete utterance rewriting. 15

3.2 The inputs for the model. 16

3.3 The simplified version of introduced masks for constructing utterance and speaker-aware representations via MHSA [22]. 20

3.4 The Decoupling part using different masks ranging from M_1 to M_4. . . 20

3.5 The Fusing or decoupled vector representations, $\hat{E_U}$ and $\hat{E_S}$. 22

4.1 The comparison of the utterance rewriting results from the baseline model and our proposed model for "coreference" in `CREAD` dataset. . . . 37

4.2 The comparison of the utterance rewriting results from the baseline model and our proposed model for "ellipsis" in `CREAD` dataset. 38

4.3 The utterance rewriting results from our proposed model in `MedDial` dataset. 39

A.1 An example dialogue. 48

A.2 Introduced mask M_1 for the example dialogue 48

List of Tables

3.1 Statistics of the datasets utilized in the evaluation of our proposed method. . . . 12

4.1 The utterance rewriting results on CREAD dataset. . . . 30

4.2 The utterance rewriting results on CANARD dataset. . . . 32

4.3 The utterance rewriting results on Task dataset. . . . 33

4.4 The utterance rewriting results on MedDial dataset 34

4.5 The ablation study on CREAD dataset. . . . 35

1 Introduction

1.1 Background

The task of Dialogue-based Incomplete Utterance Rewriting (IUR) focuses on addressing the inherent challenges posed by incomplete utterances in dialogues, which often involve the phenomena of coreference and ellipsis. As utterances devoid of complete meaning might require the earlier parts of the conversation and the context to be properly interpreted, incomplete utterances in a dialogue pose a challenge for recent dialogue systems and chatbots.

The task of Dialogue-based IUR involves taking such incomplete utterances and rewriting them into complete and stand-alone forms by capturing the original intended meaning of the speaker and considering the context of the conversation. Incomplete utterances in dialogues often fall into two main types: coreference and ellipsis.

Coreference refers to a phenomenon in language in which a word refers back to another word, phrase, or entity earlier in the conversation [1]. The word it refers back to is called the antecedent. This is often seen with pronouns like “he”, “she”, “it”, “this”, “that”, etc. Ellipsis is a phenomenon in which a word or phrase is left out because it is implied and understood in the context [2]. Dialogue-based IUR involves correctly identifying and filling in such ambiguous references or gaps. Properly understanding and handling these incomplete utterances are critical for developing more sophisticated and effective dialogue systems because this could contribute to the comprehension capabilities of AI-based dialogue systems.

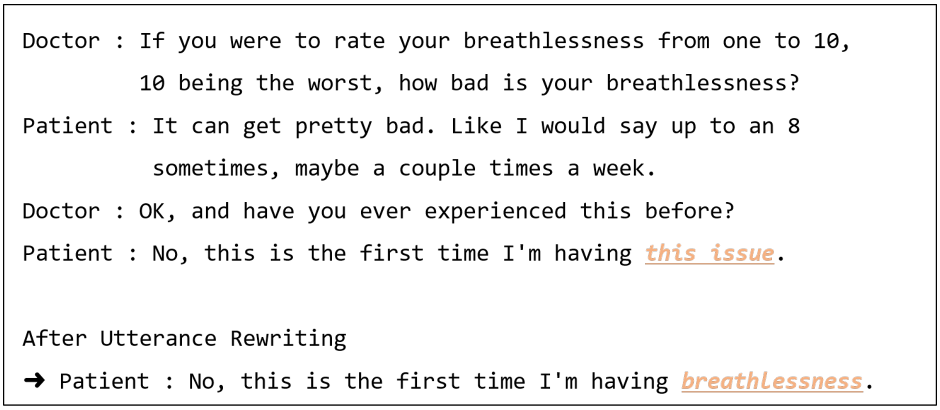
While Dialogue-based IUR is considered a stand-alone task, it is also viewed as one of the components of end-to-end dialogue systems because of its importance. This is adopted as an intermediate step for various purposes, such as dialogue summarization [3] and conversational question answering [4]. Although dialogue-based IUR is still challenging, it is regarded as an essential task in NLP in improving the quality and accuracy of automated dialogue systems and significantly enhancing human-machine interactions.

1.2 Task Definition

The objective of Dialogue-based IUR is to transform last utterance of the dialogue into a self-contained utterance. The primary focus is on cases in which the last utterance contains the phenomenon of either coreference or ellipsis and cannot be fully understood without considering the dialogue history.

In such situations, the model aims to generate a rewritten version of the last utterance that is self-comprehensible and does not rely on the dialogue context for understanding. The model generates a self-contained version of the incomplete utterance by leveraging the information and context provided by the preceding dialogues.

The goal of IUR is to produce a reformulated version of the last utterance that maintains its semantic fidelity while ensuring that it can be understood in isolation without requiring any reference to the preceding dialogue context. By achieving this objective, the IUR task aims to enhance the interpretability, coherence, and effectiveness of dialogue systems, facilitating more seamless and meaningful interactions in a dialogue.



Doctor : If you were to rate your breathlessness from one to 10,
10 being the worst, how bad is your breathlessness?

Patient : It can get pretty bad. Like I would say up to an 8
sometimes, maybe a couple times a week.

Doctor : OK, and have you ever experienced this before?

Patient : No, this is the first time I'm having ***this issue***.

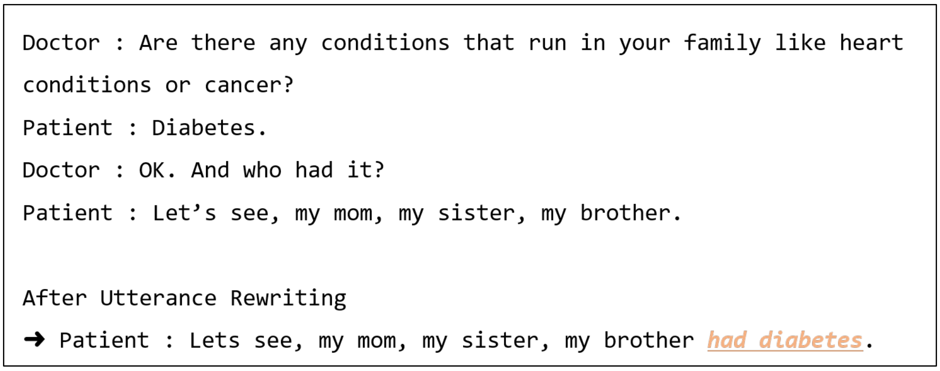
After Utterance Rewriting

➜ Patient : No, this is the first time I'm having ***breathlessness***.

Figure 1.1 Example of an incomplete utterance (coreference) in a dialogue.

The above dialogue in Figure 1.1 is an example dialogue from one of the IUR datasets that we use in the experiment. The patient's last statement "No, this is the first time I'm having this issue." is an incomplete utterance. If the last utterance were taken out of context, the meaning of "this issue" could be ambiguous, indicating that it cannot be understood without the dialogue context. This is an instance of coreference, a linguistic phenomenon where a pronoun or noun phrase refers back to a previously mentioned entity.

The task of Dialogue-based IUR, in this case, is to rewrite the incomplete utterance so that it becomes a complete sentence that could stand alone outside of the dialogue context while maintaining its original meaning. This involves identifying that "this issue" is referring to "breathlessness" and rewriting the sentence accordingly: "No, this is the first time I'm having breathlessness." This rewritten sentence conveys the same idea as the original, but now it is a complete, self-contained statement. It does not require any additional context to be understood even after it is taken out of the dialogue.



Doctor : Are there any conditions that run in your family like heart conditions or cancer?
Patient : Diabetes.
Doctor : OK. And who had it?
Patient : Let’s see, my mom, my sister, my brother.

After Utterance Rewriting
➜ Patient : Lets see, my mom, my sister, my brother ***had diabetes***.

Figure 1.2 Example of an incomplete utterance (ellipsis) in a dialogue.

The dialogue presented in Figure 1.2 is another illustrative example extracted from the same datasets referenced earlier. It is another type of an incomplete utterance, called ellipsis, which is a prevalent phenomenon found in dialogue interactions. Ellipsis is a linguistic phenomenon in which a word or phrase that is crucial to the structure or meaning of a sentence is intentionally left out. This happens because human interlocutors implicitly understand the context of the conversation without explicitly mentioning the parts.

In the above dialogue, the phrase “had diabetes” is omitted in the patient’s last response. It is understood in the context considering the doctor’s previous question, but the sentence “Let’s see, my mom, my sister, my brother.” is not a complete form on its own. The task of Dialogue-based IUR in this context involves inferring the missing information from the surrounding dialogue and inserting it back into the sentence to create a stand-alone statement. After being rewritten, the patient’s response becomes “Let’s see, my mom, my sister, my brother had diabetes.” This makes it a complete sentence that maintains its original meaning but can now be understood outside of the original dialogue context.

Understanding and accurately handling the elliptical construction are crucial for a system to effectively interpret and engage in human-like dialogue. The objective of our study is to develop a model that effectively leverages the inherent properties of dialogues in the task of utterance rewriting. Specifically, the focus of our study is on the reformulation of the last utterance within a dialogue, rather than rewriting the entire incomplete utterances throughout the dialogue, which aligns with the scope of previous approaches in this field.

2 Related Work

2.1 Incomplete Utterance Rewriting

The field of Dialogue-based IUR has gained substantial attention within the research community in recent years. Numerous researchers have acknowledged the importance of addressing the inherent challenges posed by incomplete utterances in dialogues, and have dedicated their efforts to devising effective models and methodologies.

Prior studies in the field have explored various approaches, most of which include the utilization of sequence-to-sequence models and copy mechanisms. These approaches treat the IUR task as a machine translation task, aiming to transform incomplete utterances into self-contained ones.

One notable work by Malmi et al. introduces a sequence labeling approach for sentence rewriting, which conceptualizes text generation as a text editing task [5]. This proposed method offers the advantage of fast inference time while achieving the performance comparable to vanilla sequence-to-sequence models.

Quan et al. employ an end-to-end sequence-to-sequence model for IUR [6]. Their model consists of two encoders: one reads the user utterance, while the other processes the dialogue context. The decoder module then generates the complete utterance, facilitating the transformation from incomplete to self-contained utterances. Within the decoder, either a copy mechanism [7] or a modified gated copy mechanism [8] is incorporated.

Another notable approach by Pan et al. adopts a pick-and-combine model for the IUR task and constructs a large-scale annotated dataset to support their research [9]. Their proposed model leverages a sequence-to-sequence architecture with attention and a pointer generative network.

There have been recent works that have taken a fresh approach to the IUR task, deviating from the conventional perspective of treating it as a text translation task.

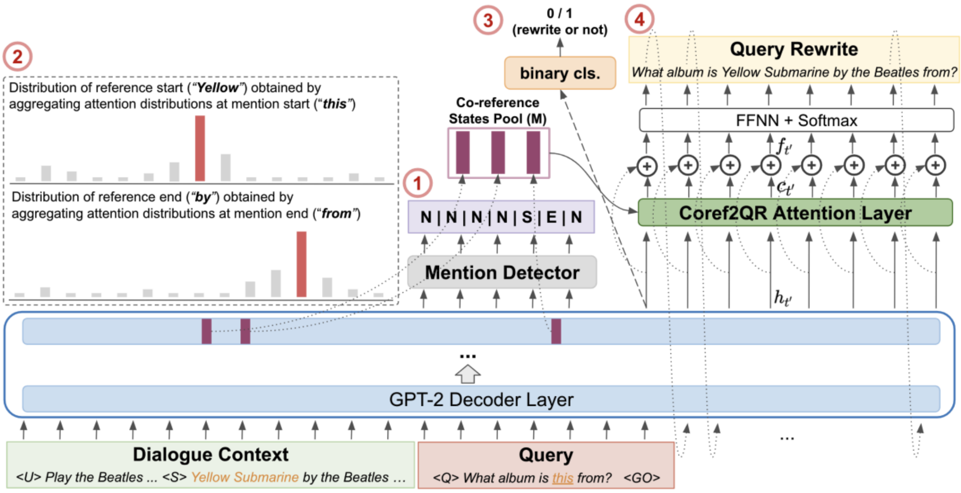
Qian et al. introduce a unique perspective by formulating the task as a semantic segmentation problem, drawing inspiration from the field of computer vision [10]. Instead of generating the entire utterance from scratch, their approach incorporates edit operations to shape the problem as the prediction of a word-level edit matrix, encompassing operations such as substitution, insertion, and none. This innovative formulation allows for fine-grained control over the generation process and facilitates more accurate and contextually appropriate utterance rewriting.

Additionally, Huang et al. present a novel semi-autoregressive generator that combines the efficiency and flexibility of autoregressive text generation and sequence labeling for text editing [11].

2.2 Baseline Model

Our baseline model is the framework proposed in [12]. The overall architecture is illustrated in Figure 2.1.

In [12], a novel joint learning framework is introduced to address the task of Dialogue-based IUR. To generate accurate rewritten utterances, they incorporate a unique attention layer called “Coref2QR”, which leverages information obtained during coreference resolution to enhance the rewrite generation process. Specifically, in this attention layer, the antecedents and referents are allowed to attend to each other. This enables the model to incorporate coreference information into the input representation and utilize it during the utterance rewriting step.



2
Distribution of reference start ("Yellow") obtained by aggregating attention distributions at mention start ("this")
Distribution of reference end ("by") obtained by aggregating attention distributions at mention end ("from")
3
0 / 1
(rewrite or not)
binary cls.
Co-reference States Pool (M)
1
N | N | N | N | S | E | N
Mention Detector
4
Query Rewrite
What album is Yellow Submarine by the Beatles from?
FFNN + Softmax
$f_{t'}$
$c_{t'}$
Coref2QR Attention Layer
$h_{t'}$
...
GPT-2 Decoder Layer
...
Dialogue Context
<U> Play the Beatles ... <S> Yellow Submarine by the Beatles ...
Query
<Q> What album is this from? <GO>

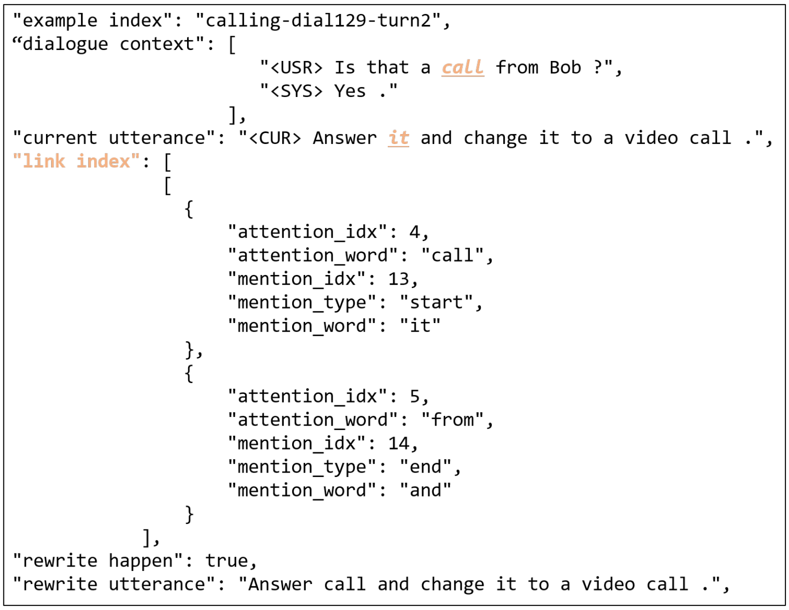
Figure 2.1 The overall model archietecture of the baseline model CREAD [12].

Furthermore, a novel joint learning framework is adopted for the utterance rewriting task. The main tasks include identifying coreference links between the last utterance of the dialogue and the dialogue context, binary classification to determine whether rewriting is necessary or not, and generating self-contained utterances. These tasks collectively contribute to improving the model's ability to accurately rewrite incomplete utterances in dialogues.

By leveraging the proposed attention mechanism and adopting a joint learning framework, [12] aims to enhance the performance of its baseline model in the task of Dialogue-based IUR.

Another notable contribution of their work is the construction of a specialized dataset for IUR. This dataset is designed to facilitate the implementation of the

“Coref2QR” attention mechanism by including the annotation of coreference links between the antecedents and referents in the dataset as in Figure 2.2. In particular, the dataset includes index labels indicating the possible antecedents and referents, if they exist, for each data sample. These links provide explicit information about the coreference relationships within the dialogues.



```
"example index": "calling-dial129-turn2",
“dialogue context": [
                        "<USR> Is that a call from Bob ?",
                        "<SYS> Yes ."
                     ],
"current utterance": "<CUR> Answer it and change it to a video call .",
"link index": [
               [
                 {
                     "attention_idx": 4,
                     "attention_word": "call",
                     "mention_idx": 13,
                     "mention_type": "start",
                     "mention_word": "it"
                 },
                 {
                     "attention_idx": 5,
                     "attention_word": "from",
                     "mention_idx": 14,
                     "mention_type": "end",
                     "mention_word": "and"
                 }
             ],
"rewrite happen": true,
"rewrite utterance": "Answer call and change it to a video call .",
```

Figure 2.2 The datasets constructed by CREAD [12].

In this study, we introduce a novel approach by transforming the “Coref2QR” attention layer into the Decoupling-Fusing layer, which will be elaborated on in subsequent chapters. Although the utilization of “Coref2QR” may have contributed to the performance, it requires manual labeling of coreference information in the dataset, a process that is both expensive and labor-intensive. Apart from the CREAD dataset released by the authors, the other existing datasets for IUR do not contain coreference links. This highlights the scarcity of available resources with coreference information for training and evaluating using “Coref2QR” module.

Additionally, the baseline model's capability to leverage the intrinsic properties of a dialogue remains limited, as it processes the input sequence as if it were dealing with free texts. Considering the primary objective of the task, we have incorporated the distinctive characteristics of dialogues into the design of our proposed model. By doing so, we aim to capture and leverage the utterance and speaker information of dialogues, ultimately improving the performance of utterance rewriting.

By transforming the "Coref2QR" attention layer into the integration of dialogue-specific design elements, we strive to overcome the limitations of the baseline approach and harness the inherent characteristics of dialogues for more effective utterance rewriting. This modification significantly contributes to the enhancement of utterance rewriting performance.

3 Materials and Method

In this chapter, we aim to provide a comprehensive overview of both the training datasets utilized in our study, the proposed methodology, and the evaluation metrics.

3.1 Datasets

The main objective of our research is to conduct a comprehensive comparison between the baseline model [12] and our proposed method in the task of utterance rewriting. To achieve this, we mainly focus on evaluating the performance of both models using the same dataset.

In order to identify the potential of our model to a broader application, we conduct experiments on three additional multi-turn dialogue-based utterance rewriting datasets, which comprise two publicly available datasets and one specifically-constructed dataset. Details on each dataset is shown in Table 3.1. For the public datasets including the main datasets from [12], we have employed the identical data split used in the respective publications in order to ensure consistency and comparability with the original papers, thereby aiming to maintain a standardized approach to our experimentation process and facilitate the accurate comparisons and evaluations.

Table 3.1 Statistics of the datasets utilized in the evaluation of our proposed method.

	CREAD [12]	**CANARD** [13]	**Task** [6]	**MedDial**
Train	16.0K	32K	2.2K	319
Dev	1.9K	3.3K	0.5K	39
Test	1.9K	3.3K	-	42

3.1.1 CREAD

The first and the main dataset employed in our study is referred to as `CREAD` [12]. This particular dataset was tailored specifically for the purpose of IUR, by building upon an existing coreference resolution dataset called `MuDoCo` [14]. The dataset comprises multi-turn dialogues with an average of 2.6 utterances per dialogue, which spans across six distinct domains (`Calling`, `Messaging`, `Music`, `News`, `Reminders` and `Weather`).

One noteworthy aspect that sets this dataset apart from other publicly available datasets is its unique annotation of coreference links, which was illustrated earlier in the Figure 2.2 in Chapter 1. The coreference links between the antecedent and the referent are annotated in cases where such cases exist. This distinctive annotation plays a crucial role in the proposed model, allowing for the effective utilization of these coreference links. By introducing the corerefence information into the model, the authors aim to improve the performance of models.

3.1.2 CANARD

Another publicly-available dataset utilized in our study is called `CANARD` [13]. This dataset was developed by building upon an existing dataset called `QuAC` [15], which is widely used for the task of Conversational Question Answering [16]. The construction of CANARD involved creating self-contained questions based on the conversation topic and the dialogue context from the original QuAC dataset.

One notable distinction of this dataset as compared to other datasets, lies in the nature of the target utterances for rewriting. Unlike other datasets where the target utterances are typically in the form of plain statements, the target utterances in this dataset are in the form of questions.

3.1.3 Task

The final publicly available dataset used in our study for the additional evaluation is Task [6]. This dataset was manually annotated, taking CamRest676 dataset [17] as its foundation, which specifically focuses on the Restaurant domain. One of the notable characteristics of the dataset is that each utterance was annotated with both coreference and ellipsis versions, whenever possible. This comprehensive annotation process resulted in the creation of incomplete utterances which encompass either ellipsis or co-reference versions. Importantly, no new versions were generated outside of the original base dataset.

3.1.4 MedDial

To extend the application of IUR to the Medical domain, we undertook the construction of a dataset encompassing conversations between medical professionals and patients. It is worth noting that questions and remarks from non-experts or laypeople in the medical field may lack proper structure and explicitness. These utterances often contain a significant number of coreferences and ellipsis that require resolution and comprehension, especially when discussing health-related issues [18]. To this end, these characteristics in medical-related dialogues necessitate effective utterance rewriting for tailored adjustments in domain-specific assistive dialogue systems.

In order to collect relevant dialogues from a hospital setting, we performed additional annotation on the dataset presented in [19]. From this dataset, we randomly selected the samples, and these selected dialogues underwent a re-editing process to form a total of four utterances per dialogue. This curated dataset provides a valuable resource for training and evaluating models in the `Medical` domain, enabling the exploration of effective strategies for handling incomplete utterances in healthcare conversations. For more comprehensive information regarding the data construction and annotation samples, we refer readers to Appendix B.

3.2 Proposed Model

The task of Dialogue-based IUR can be formulated as follows. Given a multi-turn dialogue $\mathcal{D} = \{u_1, u_2, ..., u_n\}$ where n represents the number of utterances, and each utterance $u_i = \{w_{i,1}, w_{i,2}, ..., w_{i,m}\}$ consists of m tokens with $w_{i,j}$ denoting the j-th token in the sequence u_i, the objective of IUR is to generate a rewritten version u_n^* for the last utterance of the dialogue u_n. The rewritten utterance u_n^* should possess the same underlying meaning or semantic content as the original utterance u_n. Additionally, it should be self-contained and independently comprehensible, without any reliance on the dialogue history $\{u_1, u_2, ..., u_{n-1}\}$.

Figure 3.1 provides a visual representation of the proposed model, illustrating the architecture and components employed in our approach. Our proposed model for utterance rewriting is built upon the GPT-2 architecture [20]. By utilizing the GPT-2 architecture as the basis for our model, we benefit from its ability to capture complex language patterns and generate coherent text. To generate the rewritten utterance

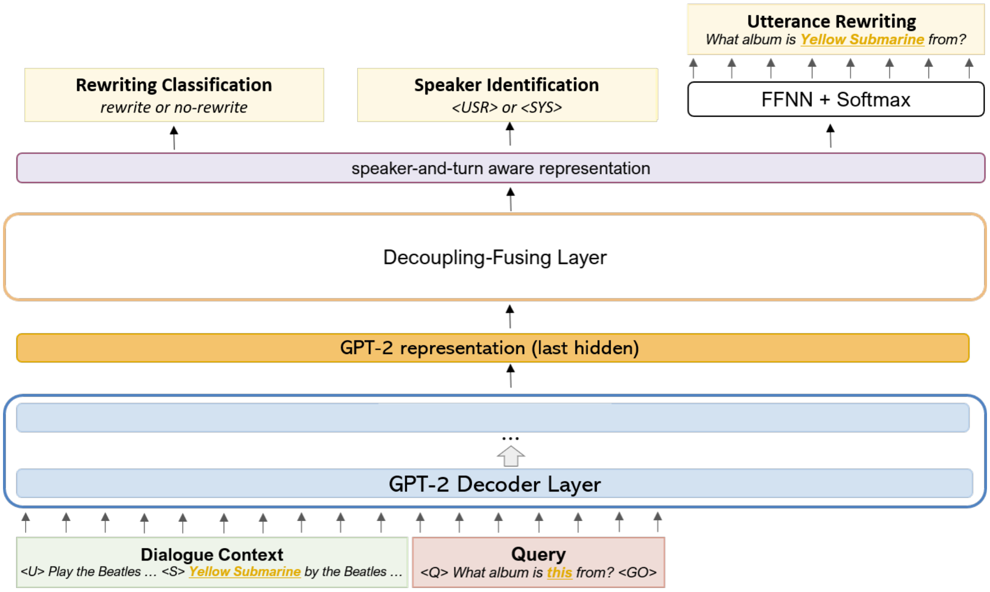

Utterance Rewriting
What album is Yellow Submarine from?
Rewriting Classification
rewrite or no-rewrite
Speaker Identification
<USR> or <SYS>
FFNN + Softmax
speaker-and-turn aware representation
Decoupling-Fusing Layer
GPT-2 representation (last hidden)
...
GPT-2 Decoder Layer
Dialogue Context
<U> Play the Beatles ... <S> Yellow Submarine by the Beatles ...
Query
<Q> What album is this from? <GO>

Figure 3.1 The proposed model for dialogue-based incomplete utterance rewriting.

u_n^*, our approach leverages the inherent properties observed in human dialogues and incorporates them into the generation model using Decoupling-Fusing network. By doing so, we aim to capture the essence of dialogue dynamics and incorporate it into the rewriting process. By leveraging the strengths of the GPT-2 architecture and incorporating dialogue-specific adaptations, our model aims to achieve improved performance in generating self-contained and contextually appropriate rewritten utterances.

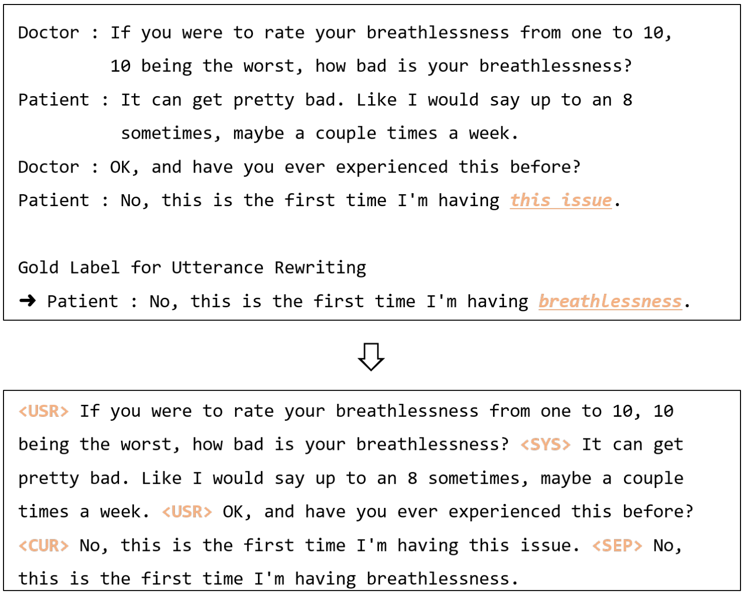
The proposed model comprises three main steps: (1) determining if the last utterance requires rewriting or not, (2) predicting which the speaker each token corresponds to, and (3) generating the rewritten utterance if necessary.

By implementing these three tasks, our proposed model adopts a joint learning framework that aims to effectively handle incomplete utterances within dialogues and produce high-quality rewritten utterances. The subsequent sections delve into further details and specifics for each part of the model, providing a comprehensive understanding of the proposed approach.

3.2.1 Contextualized Representation

To construct the input sequence, we concatenate all the context utterances to form the input representation for a specific utterance u_i. Furthermore, in order to distinguish the boundaries between individual utterances within the concatenated sequence, we introduce additional special tokens to indicate the respective speakers.

Specifically, we add the tokens <USR> and <SYS> to represent the first and the second speaker in the input sequence. Also, for the last utterance, we append <CUR> as its corresponding special token. By incorporating these special tokens to mark the boundaries and denote the speakers within the input sequence, our model gains a better understanding of the dialogue structure and can effectively leverage the contextual information provided by each speaker in generating the rewritten utterance.



```
Doctor : If you were to rate your breathlessness from one to 10,
         10 being the worst, how bad is your breathlessness?
Patient : It can get pretty bad. Like I would say up to an 8
          sometimes, maybe a couple times a week.
Doctor : OK, and have you ever experienced this before?
Patient : No, this is the first time I'm having this issue.

Gold Label for Utterance Rewriting
➜ Patient : No, this is the first time I'm having breathlessness.
```

```
<USR> If you were to rate your breathlessness from one to 10, 10
being the worst, how bad is your breathlessness? <SYS> It can get
pretty bad. Like I would say up to an 8 sometimes, maybe a couple
times a week. <USR> OK, and have you ever experienced this before?
<CUR> No, this is the first time I'm having this issue. <SEP> No,
this is the first time I'm having breathlessness.
```

Figure 3.2 The inputs for the model.

For example, if the example dialogue in Figure 1.1 is fed into the model during the training phase, all the utterances in the dialogue would be concatenated, with the sepcial tokens in between as illustrated in Figure 3.2. Thus, the first speaker "Doctor" is changed into the token <USR>, and "Patient" into <SYS> while training the model.

We utilize the output from the last layer of the GPT-2 model to obtain contextualized representations for the input sequence. This allows us to capture the contextual information and embeddings for each token in the input sequence effectively. Here, we obtain the contextualized representations for each token denoted as $h_m \in \mathbb{R}^d$, where m represents the position of the input token and d represents the embedding size. These representations encapsulate the contextual understanding and semantic information of each token within the dialogue context, facilitating subsequent steps in our proposed model.

3.2.2 Decoupling-Fusing Layer

In this section, we introduce Decoupling-Fusing network, which is an additional Multi-Head Self-Attention (MHSA) with some masks. The method of "Decoupling-Fusing" is a methodology proposed in dialogue-based NLP tasks [21] [22]. The mechanism involves separating or "decoupling" different aspects of the dialogue, which in our study is the utterance and speaker information, constructing the representation for each aspect separately, and then "fusing" them back together to form a composite representation that takes both aspects into account. This approach can potentially allow the model to incorporate the interactions between the speaker and the utterance more effectively, thereby capturing the semantics of the dialogue context.

This MHSA mechanism positioned at the topmost layer of the of the last layer of GPT-2 and enhances the model's ability to capture and incorporate utterance and speaker information of a dialogue into the contextualized representation for each token.

Mathematically, the MHSA can be formulated as follows:

$$Attention(Q, K, V, M) = softmax(\frac{QK^T}{\sqrt{d_k}} + M)V \tag{3.1}$$

In this equation, Q, K, and V represent the query, key, and value matrices, respectively. These matrices are constructed using learnable parameters $W^Q \in \mathbb{R}^{d \times d_{\text{query}}}$, $W^K \in \mathbb{R}^{d \times d_{\text{key}}}$, and $W^V \in \mathbb{R}^{d \times d_{\text{value}}}$, where d_{query}, d_{key}, and d_{value} indicate the dimensions of Q, K, and V, respectively.

MHSA applies a softmax operation to the scaled dot-product of the query and key matrices, divided by the square root of the dimension d_k. The operation of scaled dot-product of the query and key matrices outputs attention weights, which indicate the relevance or importance of each element in the input sequence. Here, the mask M is applied to this attention weights to ensure which elements are attended to and which are not. Finally, the attention weights are used to weight the corresponding values in the value matrix V. This results in the attended representations that capture the specific information for each token.

In this study, we introduce four masks $\{M_k\}_{k=1}^{4} \in \mathbb{R}$ to incorporate utterance and speaker information in a dialogue following the previous work [22], which is defined as follows:

$$
\begin{aligned}
M_1[i,j] &= \begin{cases} 0, & \text{if } \mathbb{T}_i = \mathbb{T}_j \\ -\infty, & \text{otherwise} \end{cases} \\
M_2[i,j] &= \begin{cases} 0, & \text{if } \mathbb{T}_i \neq \mathbb{T}_j \\ -\infty, & \text{otherwise} \end{cases} \\
M_3[i,j] &= \begin{cases} 0, & \text{if } \mathbb{S}_i = \mathbb{S}_j \\ -\infty, & \text{otherwise} \end{cases} \\
M_4[i,j] &= \begin{cases} 0, & \text{if } \mathbb{S}_i \neq \mathbb{S}_j \\ -\infty, & \text{otherwise} \end{cases}
\end{aligned} \tag{3.2}
$$

In Equation 3.2, the variables i and j represent different token positions within the entire dialogue. $\mathbb{T}_i$ denotes the index of the utterance in which the i^{th} token is located, while $\mathbb{S}_i$ signifies the speaker associated with the i^{th} token. The masks M_1, M_2, M_3, and M_4 are designed to guide the attention mechanism to attend to the specific part of a dialogue.

The illustration of each mask is depicted in Figure 3.3. M_1 and M_2 are used to construct utterance information. Specifically, M_1 facilitates attending to the tokens within the current utterance, allowing the model to capture the intra-utterance dependencies. An example dialogue and its corresponding M_1 is depicted in Figure A.1 and Figure A.2 respectively. On the other hand, M_2 enables attending to tokens from other utterances in the dialogue, enabling the model to capture inter-utterance dependencies and context.

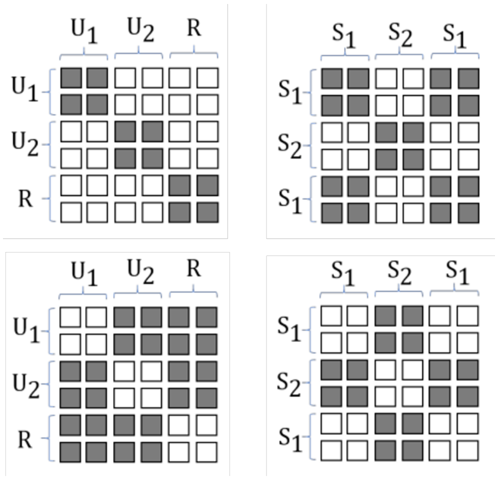


Figure 3.3 The simplified version of introduced masks for constructing utterance and speaker-aware representations via MHSA [22].

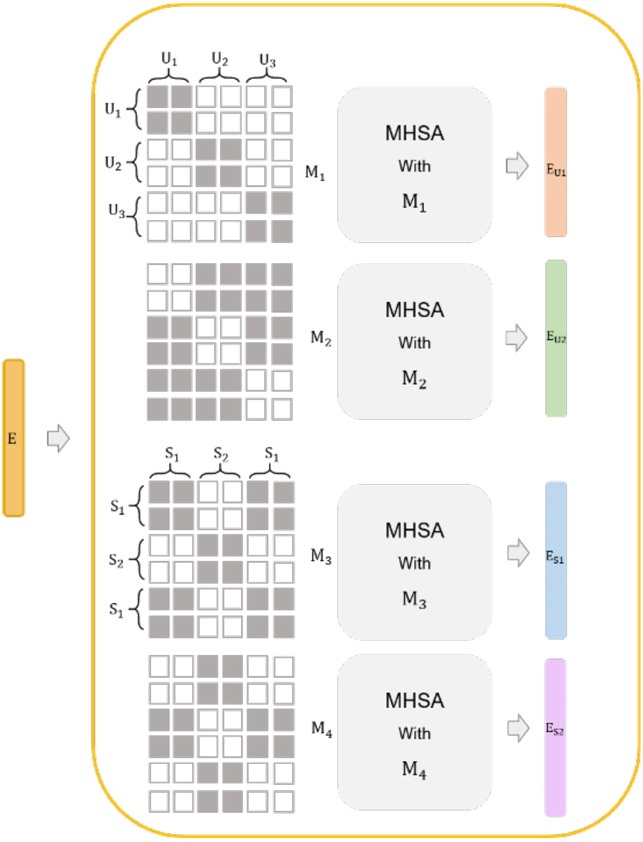


Figure 3.4 The Decoupling part using different masks ranging from M_1 to M_4.

M_3 and M_4 are employed to capture speaker information in the dialogue. M_3 focuses on attending to the tokens from the same speaker, facilitating the model's understanding of the speaker's individual style, preferences, and patterns within the dialogue. Conversely, M_4 guides the attention mechanism to attend to tokens from different speakers, enabling the model to capture the dynamics and interactions between speakers in the dialogue.

By utilizing these masks, the attention mechanism is guided to attend to specific aspects of the dialogue, and outputs the same utterance-aware representation E_{U1}, different utterance-aware representation E_{U2}, same speaker-aware representation E_{S1}, and different speaker-aware representation E_{S2}.

$$MHSA(E, M_i), i \in \{1, 2, 3, 4\} \tag{3.3}$$

This tailored attention mechanism enhances the model's ability to capture relevant information and dependencies within the dialogue, resulting in a more comprehensive and accurate representation of both utterance and speaker information. The term "Decoupling" in the "Decoupling-Fusing" mechanism specifically refers to this above step in which specific information-contained representations are constructed separately.

In order to "fuse" the utterance-aware and speaker-aware representation, we use a fully connected layer to incorporate the resulting four representations. Specifically, the formulation of the incorporation is denoted as follows:

$$\begin{aligned} E_1 &= ReLU(FC([E - \tilde{E}, E \odot \tilde{E}])) \\ E_2 &= ReLU(FC([E - \bar{E}, E \odot \bar{E}])) \\ \hat{E} &= Sigmoid(FC([E_1; E_2])) \end{aligned} \tag{3.4}$$

Let E denote the original representations output from GPT-2. For obtaining utterance-aware representation, $\tilde{E} = E_{U1}$ and $\bar{E} = E_{U2}$, and for speaker representation, $\tilde{E} = E_{S1}$ and $\bar{E} = E_{S2}$.

The resulting vectors are utterance-aware representation $\hat{E_U}$ and speaker-aware representation $\hat{E_S}$, respectively. To incorporate these two representations and construct the final representation for implementing the following three tasks in our model, we compute an element-wise summation of the two vectors as following Figure 3.5:

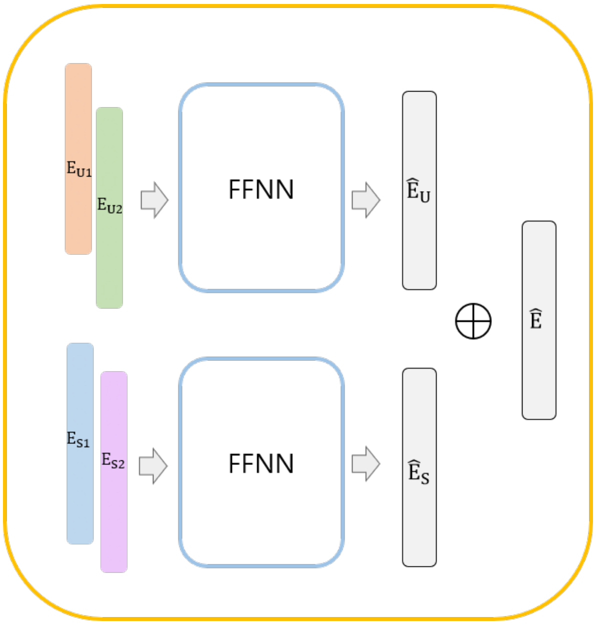


Figure 3.5 The Fusing or decoupled vector representations, $\hat{E_U}$ and $\hat{E_S}$.

The final representation $\hat{E}$ obtained from Decoupling-Fusing network serves as a crucial input for the subsequent tasks in our proposed model. We hypothesize that this representation encompasses the utterance and speaker information captured by the Decoupling-Fusing layer. To encode contextualized linguistic information, we introduce the original GPT-2 representation along with $\hat{E}$, using weights of 0.7 for $\hat{E}$ and 0.3 for the GPT-2 vector representation.

3.2.3 Rewriting Classification

The first task of our model is to predict whether the last utterance in the given input sequence requires rewriting, following the approach presented in [12]. This task serves as a binary classification, where the model determines whether the last utterance needs to be rewritten or not.

As depicted in Figure 3.1, our model includes a classifier component. This classifier consists of a two-layer feed-forward neural network followed by a softmax layer. It takes the input representation after the Decoupling-Fusing step and predicts a vector with two entries, representing the rewrite and no-rewrite classes.

The rewrite class encompasses instances where coreference or ellipsis is present, indicating that rewriting is necessary. On the other hand, the no-rewrite class includes instances where the last utterance is already self-contained and does not require any modifications.

Crucially, the model only generates the rewritten utterance when the binary prediction is true, signifying that the classifier has determined that rewrite is necessary. In cases where the classifier predicts the no-rewrite class, the input utterance is directly copied as the output without any modifications.

The efficacy of this binary classifier is demonstrated in [12], where it achieves a high accuracy. This classifier serves as a filter that minimizes the risk of incorrectly rewriting already self-contained utterances. Moreover, it allows the subsequent generation process to focus solely on rewriting incomplete utterances during training, enhancing the overall performance and reliability of the utterance rewriting task.

3.2.4 Speaker Identification

The model's second step involves predicting the speaker for each token in the dialogue. We humans easily recognize which token is spoken by which speaker even from a random written dialogue. It is evident that humans possess a natural ability to discern the speaker associated with each token effortlessly.

However, as the input sequence of this model is a form of concatenated utterances in the whole dialogue, it becomes imperative for the model to accurately determine the correspondence between tokens and their respective speakers. Given the concatenated nature of the input sequence, explicitly identifying the speaker associated with each token becomes a challenging task. Hence, the model must rely on its learned understanding of linguistic patterns, contextual cues, and other relevant features to successfully attribute tokens to their respective speakers.

For instance, consider the scenario where the model encounters the token "rate" within the "Doctor" 's first utterance in Figure 1.1. In this case, the model is expected to accurately predict the speaker associated with that token as "Doctor". Similarly, when encountering the token "pretty" within the "Patient"'s first utterance, the model should correctly identify the speaker of that token as "Patient".

The correct prediction of speakers for each token enables the model to generate rewritten utterances that align with the intended speaker and accurately represent the dialogue context. This speaker prediction capability enhances the model's ability to produce coherent and contextually appropriate rewritten utterances that reflect the dynamics and speaker roles within the dialogue. By assigning the appropriate speaker labels to the tokens, the model gains an understanding of the dialogue flow and consider the speaker's perspective when generating the rewritten utterance. Thus, the model's ability to discern the speaker-token relationship within the concatenated utterances is crucial.

3.2.5 Utterance Rewriting

In the final step of the proposed approach, the model proceeds with the generation process based on its binary decision of whether to rewrite the last utterance or not. Here, the rewritten utterance is generated utilizing the constructed utterance-and-speaker aware vector $\hat{E}$.

By utilizing the vector representation from Decoupling-Fusing layer, the model incorporates the comprehensive understanding of the dialogue context, including the individual utterances and their respective speakers. This enriched vector serves as the foundation for generating the rewritten utterance in a manner that aligns with the contextual and speaker-aware nature of the task.

With this vector representation, the model ensures that the generated output maintains the necessary information and characteristics of the dialogue, leading to more effective and faithful utterance rewriting.

3.2.6 Learning Objectives

During the training phase, an input sequence is constructed by concatenating the dialogue context, the last utterance, and the rewritten last utterance, employing the teacher-forcing methodology. To train the model effectively, we adopt a joint-learning methodology of the three distinct tasks illustrated above: Rewriting Classification, Speaker Identification, and Utterance Rewriting.

The objective of the Rewriting Classification task involves binary classification, where the model predicts the class label (rewrite or no-rewrite). The loss function used for this task is the two-class cross-entropy, which measures the dissimilarity between the predicted class p^{CLS} and the ground truth label $\mathbf{y}^{CLS}$.

$$L^{CLS} = -log((\mathbf{y}^{CLS})^T p^{CLS}) \tag{3.5}$$

For the Speaker Identification task, the loss function employed is also the two-class cross-entropy. This choice of loss function is suitable because all the utilized multi-turn dialogue-based utterance rewriting datasets consist of dialogues involving only two speakers.

The predicted class for token k's Speaker Identification is denoted as p_k^{IDEN} and the ground truth label is represented as $\mathbf{y}_k^{IDEN}$. The two-class cross-entropy loss function effectively measures the dissimilarity between the predicted speaker identification and the true speaker labels. By learning to minimize this loss, the model becomes proficient in assigning the correct speaker to each token, contributing to the generation of contextually appropriate and speaker-aware rewritten utterances.

$$L^{IDEN} = \sum_{k=1}^{l} -log((\mathbf{y}_k^{IDEN})^T p_k^{IDEN}) \tag{3.6}$$

In the Utterance Rewriting task, the approach is similar to the standard language modeling task. The loss function employed is the cross-entropy, which measures the dissimilarity between the predicted sequence p^{GEN} and its corresponding ground-truth sequence $\mathbf{y}^{GEN}$.

During training, the model generates a sequence of tokens p^{GEN}, and the objective is to minimize the cross-entropy loss between this generated sequence and the actual ground-truth sequence $\mathbf{y}^{GEN}$. The cross-entropy loss is computed at each time step t in the sequence, evaluating the dissimilarity between the predicted tokens and the true tokens.

By optimizing this loss function, the model learns to generate rewritten utterances that closely match the ground-truth utterances. This objective fosters the development of accurate and contextually appropriate rewritten utterances, enabling the model to effectively reformulate incomplete utterances in a dialogue.

The use of cross-entropy as the loss function ensures that the model's generation process aligns with the desired ground-truth sequences, promoting the acquisition of coherent and accurate utterance rewriting capabilities.

$$L^{GEN} = \sum_{t=1}^{T} -log((\mathbf{y}_t^{GEM})^T p_t^{GEN}) \tag{3.7}$$

The final loss is the sum of all these losses:

$$L = L^{CLS} + L^{IDEN} + L^{GEN} \tag{3.8}$$

3.3 Experimental Settings

In our model, the GPT-2 decoder layers and the classification layers are initialized with pre-trained weights from the GPT-2 small model, leveraging the knowledge and representations learned during pre-training. To fine-tune the model, we employ the Adam optimizer [23] with an initial learning rate of $5e - 05$. The training process is conducted on a single GPU (GeForce RTX 3090).

The batch size varies for each dataset: 15 for `CREAD`, 24 for `Task`, 5 for `CANARD`, and 24 for `MedDial`. These batch sizes are carefully chosen to balance computational efficiency and model performance. We use early stopping during training, where the model stops further training if it does not show performance improvement for over 5 epochs. We use the performance of utterance rewriting on the F1 score of utterance rewriting for the validation set as the criterion.

3.4 Evaluation Metrics

To ensure a fair comparison with previous studies, we adhere to the same evaluation metrics for each dataset. These evaluation metrics vary depending on the specific datasets.

For evaluating the quality of the generated utterances, we report the standard BLEU [24] scores, which measures the similarity between the generated sentences and the target sentences. The BLEU score provides a quantitative measure of how closely the generated utterances align with the desired target utterances.

In addition to the BLEU score, we also adopt an F1 score, as introduced in [13], to assess the quality of the rewritten parts within the generated sentences. Especially for `CREAD` dataset, the F1 score specifically focuses on comparing the machine-generated words with the ground truth words considering only the ellipsis or co-reference parts of the user utterances. This allows us to gauge the accuracy and fidelity of the rewritten sections. Following the previous work [12], we also present the performance of the coreference resolution in the generated rewritten queries, as RM ratio. The RM ratio indicates the percentage of successfully generated referents in the ground-truth coreference links. A higher RM ratio signifies a higher quality of coreference resolution, as it reflects the model's ability to accurately identify and generate referents in the rewritten queries.

Additionally, we incorporate ROUGE measures as part of the evaluation process. ROUGE measures assess the overlapping n-grams between the generated rewritten utterances and the corresponding reference utterances. Specifically, we employ the standard ROUGE metric, which considers the n-gram overlap between the generated and reference utterances. This metric provides a quantitative measure of the degree of similarity in terms of shared n-grams. Furthermore, we utilize $\mathrm{ROUGE_L}$ measure, which focuses on identifying the longest matching sequence between the generated and reference utterances. $\mathrm{ROUGE_L}$ metric emphasizes capturing the most substantial and comprehensive similarities in terms of sequence alignment.

By incorporating these evaluation metrics, we aim to provide a comprehensive assessment of the performance and quality of the generated rewritten queries, taking into account both overall sentence similarity and the specific aspects related to coreference and ellipsis.

4 Reuslts and Analysis

In this chapter, we present the outcomes of the experiments conducted as part of this study, focusing on the evaluation of the four distinct datasets mentioned in Section 3.1. Again, as the purpose of this study focuses on comparing the baseline model and our proposed model, the analysis encompasses the rewriting result of the dataset `CREAD`. Yet, the results includes the model's performance on every dataset, examining various metrics and indicators to assess the effectiveness of the proposed approach. Thus, in this section, we aim to provide an in-depth understanding of the model's capabilities and its ability to address the challenges of dialogue-based utterance rewriting.

4.1 Results

As discussed previously, the evaluation metrics employed for each dataset exhibit significant variability. Thus, we proceed to follow the specific evaluation methodology employed for each individual dataset in accordance with the existing approaches.

Table 4.1 The utterance rewriting results on `CREAD` dataset.

	Prec.	**Rec.**	**F1**	**BLEU**	**RM**
CREAD (Tseng et al., 2021 [12])	61.0	59.5	60.2	90.2	**82.0**
Proposed Model	**65.1**	**62.6**	**63.8**	**90.6**	81.5

For the first part of the Results section, we present a comprehensive analysis of the results obtained from the CREAD dataset in Table 4.5. The table includes metrics such as Precision, Recall, F1 score, BLUE, and RM.

As explained above, the scores are calculated confining the sections that the rewriting is required. The best scores for each metric are highlighted in bold, and the results of the baseline model are sourced from its respective research paper [12].

Upon analyzing the results, we observe that our proposed method surpasses the baseline model in all evaluated metrics except for RM. Notably, the Precision metric demonstrates a remarkable improvement of 4.1%, and the F1 score exhibits a notable increase of 3.6%. These findings highlight the effectiveness of our proposed model in the task of utterance rewriting.

It is important to see why RM showed lower performance from our method. The baseline model adopts a novel attention module that leverages explicit coreference links in the last utterance of a dialogue and the dialogue context. However, our model achieves comparable results without using this explicit information.

Overall, our proposed model showcases exceptional performance in the task of utterance rewriting in terms of the result of automatic evaluation, outperforming the baseline approach and highlighting the significance of leveraging dialogue-specific properties for enhanced performance without the need for explicit coreference annotations.

Secondly, we present a rewriting results obtained from the CANARD dataset illustrated in Table 4.2. Regarding the BLEU scores, our proposed model exhibits superior performance compared to all baseline models. The improvement in BLEU scores indicates a substantial enhancement in capturing the semantic aspects of utterance rewriting. This improvement signifies the model's ability to generate rewritten

Table 4.2 The utterance rewriting results on `CANARD` dataset.

	$\mathbf{B_1}$	**$\mathbf{B_2}$**	**$\mathbf{B_4}$**	**$\mathbf{R_1}$**	**$\mathbf{R_2}$**	**$\mathbf{R_L}$**
Pro-Sub (Elgohary et al., 2019 [13])	60.4	55.3	47.4	73.1	63.7	73.9
Ptr-Gen (See et al., 2017 [8])	67.2	60.3	50.2	78.9	62.9	74.9
RUN (Liu et al., 2020 [10])	70.5	61.2	49.1	79.1	61.2	74.7
RaST (Hao et al., 2021 [25])	53.5	47.6	38.1	62.7	50.5	61.9
MST (Jin et al., 2022 [26])	66.6	59.9	48.7	79.5	64.1	79.0
HCT (Jin et al., 2022 [26])	68.7	62.3	52.1	**80.0**	**66.5**	**79.4**
SARG (Huang et al., 2021 [12])	60.4	55.3	47.4	73.1	63.7	73.9
Proposed Model	**74.8**	**66.4**	**54.5**	77.7	64.2	75.0

utterances that closely align with the desired target utterances, demonstrating its effectiveness in capturing the intended meaning.

However, it is worth noting that our model did not achieve the state-of-the-art performance in terms of ROUGE scores. While the ROUGE scores may not reach the highest level, the overall performance suggests that our model successfully captures important aspects of the rewriting process and produces coherent and contextually appropriate rewritten utterances.

These results highlight the model's strength in improving the semantic quality of the rewritten utterances, as evidenced by the substantial increase in BLEU scores. Although further improvements may be necessary to achieve the state-of-the-art performance in terms of ROUGE scores, the overall performance of our proposed model demonstrates its effectiveness and competitiveness in the field of dialogue-based utterance rewriting.

Thirdly, we present a rewriting results obtained from the `Task` dataset illustrated in Table 4.3.

Table 4.3 The utterance rewriting results on `Task` dataset.

	EM	$\mathbf{B_4}$	$\mathbf{F_1}$
Ellipsis Recovery [6])	50.4	74.1	44.1
GECOR 1 (Quan et al., 2019 [12])	68.5	83.9	66.1
GECOR 2 (Quan et al., 2019 [12])	66.2	83.0	66.2
RUN (Liu et al., 2020 [10])	**69.2**	85.6	70.6
Proposed Model	64.6	**88.0**	**71.8**

Analyzing the results presented in Table 4.3, it is evident that our proposed model did not achieve outstanding performance in the `Task` dataset in terms of EM. Notably, the EM (Exact Match) score is particularly low compared to the highest result achieved by the RUN model [10]. This discrepancy can be attributed to the fact that our proposed model generates the last utterance of the dialogue from scratch, while the RUN model utilizes a copy mechanism that allows it to essentially copy the last utterance. As a result, the RUN model tends to make fewer mistakes compared to our model.

However, despite not excelling in terms of the EM score, our proposed model outperforms other models in terms of BLEU and F1 scores. This indicates a significant overlap between the gold rewritten utterance and the predictions made by our model. The high F1 score highlights the model's ability to capture and reproduce key elements from the target utterances, demonstrating its proficiency in generating rewritten utterances that closely align with the desired outputs.

Although our model may lag behind in terms of exact match accuracy, the impressive performance in terms of F1 scores emphasizes its capability to produce rewritten utterances that exhibit substantial agreement with the gold standard references. This suggests that while the model may not generate the exact same utterance as the reference, it is adept at capturing the essential information and context required for effective utterance rewriting.

Finally, we present the results of the rewrite obtained from the MedDial dataset, which we specifically constructed for evaluating our proposed model in the Medical domain. As MedDial is a very small dataset for the finetuning a deep learning model, we conduct an experiment using the weights obtained from training the baseline datasets CREAD, and further train and evaluate with MedDial.

Table 4.4 The utterance rewriting results on MedDial dataset

	Prec.	**Rec.**	**F1**	**BLEU**	**ROUGE$_1$**	**ROUGE$_2$**	**ROUGE$_L$**
Proposed Model	82.95	67.16	74.23	74.62	87.05	81.51	86.70

The results obtained from the MedDial dataset, which was exclusively constructed by our team, provide a unique perspective on the performance of our proposed model in the context of medical dialogues. As the dataset was specifically designed to evaluate our model's effectiveness in the Medical domain, the results obtained are exclusive to our proposed model.

The evaluation of the MedDial dataset demonstrates promising results, as indicated by the significantly higher Precision, Recall, and F1 scores compared to the CREAD dataset. These results affirm the efficacy of our model in the task of utterance rewriting within the medical domain. Notably, it is important to consider that the composition of the data samples requiring rewriting differs between the MedDial and CREAD datasets. In the MedDial dataset, over 50% of the utterances contain ellipsis, whereas the CREAD dataset has a higher proportion of data samples belonging to the no-rewrite class, accounting for over 70% out of total data samples.

Despite the difference in data composition, our model showcases superior performance in utterance rewriting for dialogues between medical professionals and patients. This suggests that our model effectively handles the challenges posed by medical dialogues, particularly in capturing and reformulating contextually incomplete utterances. The higher scores achieved in the medical domain highlight the model's proficiency in addressing the unique linguistic characteristics and complexities inherent in medical conversations.

In summary, the results obtained from the `MedDial` dataset demonstrate its effectiveness and superiority in the task of utterance rewriting within the context of medical dialogues. These results validate the model's ability to handle the specific challenges posed by dialogues between medical professionals and patients, contributing to the advancement of dialogue systems in the medical domain.

4.2 Ablation Study

In this section, we examine the impact of individual components within our integrated model on the performance of utterance rewriting. In comparison to the baseline model, we enhance the system by incorporating a Decoupling-Fusing layer and integrating the task of Speaker Identification. We aim to identify how these parts contribute to the performance improvement.

Table 4.5 The ablation study on `CREAD` dataset.

	Prec.	**Rec.**	**F1**	**BLEU**	**RM**
Proposed Model	**65.1**	**62.6**	**63.8**	**90.6**	81.5
- Decoupling-Fusing	64.7	58.7	61.5	90.1	81.8
- Speaker Identification	58.9	59.3	59.1	90.0	**82.4**

In terms of RM, we did not observe notable improvements resulting from these components. The rationale behind this lies in the fact that, unlike existing approaches that employ a copy mechanism, our proposed model generates self-contained utterances entirely from scratch. As a result, there are instances where our model produces considerably more sensible utterances than the provided gold rewritten answer. In these instances, the model does not receive high scores in terms of RM evaluation despite the correct semantic meaning conveyed by the generated utterance.

However, in general, both components in our proposed model have made significant contributions to the performance of utterance rewriting. Notably, when excluding the Speaker Identification task, the model exhibits a significant decline in performance. This underscores the essential role of speaker information in effectively rewriting incomplete dialogue utterances. The result from the ablation study suggests that incorporating distinctive dialogue properties, which differentiate it from plain text, leads to the overall enhancements in utterance rewriting.

4.3 Analysis

Our proposed method for IUR integrates utterance and speaker information in a dialogue. This can significantly enhance the performance and efficacy in most datasets for IUR. The improvement can be attributed to the following reasons: increased contextual understanding, improved sequential processing and enhanced coherence.

First, conversations inherently have a structure where every utterance builds upon previous ones, creating a rich context. Both the speaker's identity and their utterance in the conversation contribute to this context. Considering these factors in a model can enhance the understanding of the context, leading to more accurate identification and resolution of incomplete utterances.

For instance, the speaker information can provide clues when there are multiple people in a single dialogue.

Speaker A : Call my brother and sister.
Speaker B : Which brother? Harry or Mike?
Speaker A : The second.
Speaker B : Okay calling Mike. Which sister? Tiffany or Ronda?
Speaker A : Actually could you add both of *them* to the call?

Gold Rewritten Utterance : Actually could you add *both Tiffany and Ronda* to the call?
Rewritten Utterance from the baseline model :
Actually could you add *Ronda Ronda* to the call?
Rewritten Utterance from the proposed model :
Actually could you add *both of Tiffany and Ronda* to the call?

Figure 4.1 The comparison of the utterance rewriting results from the baseline model and our proposed model for "coreference" in CREAD dataset.

In Figure 4.1, there are more than two people in a dialogue even except for the speakers directly participating in the conversation. "them" in the last utterance of the dialogue here refers to "Tiffany and Ronda", but the baseline model generates the wrong tokens "Ronda Ronda". On the other hand, our proposed model correctly generates who "them" in the dialogue refer to, owing to utilizing the speaker-aware representations from Decoupling-Fusing network.

Second, the utterance in a dialogue signifies the sequence of the conversation. The model can understand the conversation better when it considers this sequence, making the interpretation and rewriting of incomplete utterances more contextually appropriate. For instance, an utterance might not explicitly represent what it actually means as in Figure 4.2 because the referent is already mentioned several utterances earlier.

Speaker A : Can you play Frosty the snowman?

Speaker B : You have 3 different recordings of that song. Which one would you prefer?

Speaker A : Who are the recordings by?

Speaker B : There is the original by Burl Ives, one by Frank Sinatra, and one by the Kinks.

Speaker A : Please play the original.

Gold Rewritten Utterance : Please play the original ***song***.

Rewritten Utterance from the baseline model :

Please play the original ***Frosty the snowman***.

Rewritten Utterance from the proposed model :

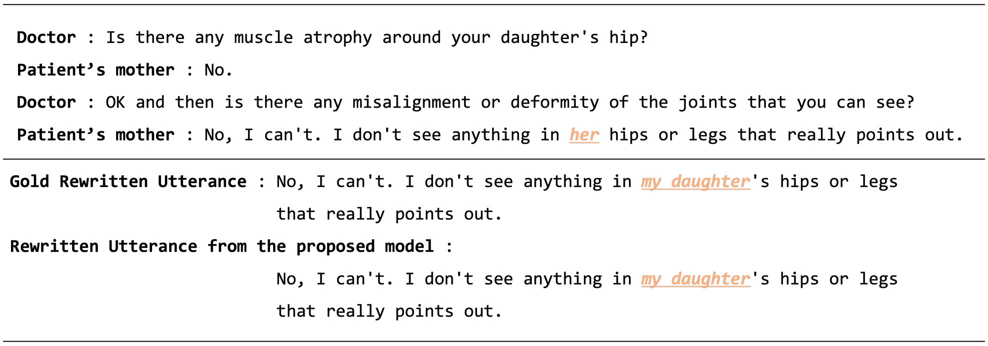
Please play the original ***recording of Frosty the snowman.***

Figure 4.2 The comparison of the utterance rewriting results from the baseline model and our proposed model for "ellipsis" in CREAD dataset.

In this example dialogue, the phenomenon of ellipsis has occurred in the last utterance. The baseline model recovered the omitted word "Frosty the snowman" correctly. Our model, on the other hand, generates more specifically, additionally using the word "recording of" present in the previous utterances. In the cases as the above example, utterance information helps the sequence and flow of the overall dialogue, guiding the model in its resolution of the ellipsis by effectively capturing and using the word in the context.

Lastly, utterance and speaker information contribute to the cohesion and coherence of the dialogue. Cohesion relates to how sentences connect with each other, while coherence is about how each utterance contributes to the overall meaning of the conversation. Understanding who is speaking and when allows the model to maintain the narrative's coherence, leading to more natural and accurate rewritten utterances.

By incorporating utterance and speaker information into the model, we succeed in aligning the model more closely with the way human conversation works. As illustrated in in Figure 4.3, the model successfully rewrites the last utterance in a



Doctor : Is there any muscle atrophy around your daughter's hip?

Patient's mother : No.

Doctor : OK and then is there any misalignment or deformity of the joints that you can see?

Patient's mother : No, I can't. I don't see anything in ***her*** hips or legs that really points out.

Gold Rewritten Utterance : No, I can't. I don't see anything in ***my daughter***'s hips or legs that really points out.

Rewritten Utterance from the proposed model :
No, I can't. I don't see anything in ***my daughter***'s hips or legs that really points out.

Figure 4.3 The utterance rewriting results from our proposed model in `MedDial` dataset.

dialogue by substituting the word "her" with "my daughter". Remarkably, despite the presence of the term "your daughter" in the dialogue contexxt, our proposed model effectively captures the cohesion and coherence of the conversation and refrains from directly copying the word "your" from the context.

In summary, our proposed model exhibits a better understanding and processing of the complexities within human dialogue, resulting in the improvement in the performance of the IUR task.

5 Conclusion

In this study, we introduce a novel method to enhance the performance of the Dialogue-based Incomplete Utterance Rewriting task. By capitalizing on the advancements of Transformer-based Pretrained Language Models, we use GPT-2 as the backbone architecture for our model.

The key concept behind our model is to leverage the intrinsic properties unique to dialogues, which differ from those found in free-text contexts. Specifically, we consider utterance and speaker information as crucial components in a dialogue and utilize a Decoupling-Fusing layer to capture the aforementioned properties in the representation. This approach allows us to create an utterance and speaker-aware vector representation, which captures the nuances and dynamics of the conversation. Furthermore, we facilitate joint learning by using three specific tasks within our model. By jointly training on these tasks, we could enhance the overall performance of the IUR.

To evaluate the effectiveness of our proposed model, we conducted extensive testing on three publicly available datasets and a specially constructed dataset. Through rigorous evaluation on multiple datasets, including publicly available and specially constructed ones, we demonstrate the effectiveness and superiority of our approach. The integration of utterance and speaker information, coupled with the power of GPT-2, enables our model to achieve remarkable results in rewriting dialogues and producing self-contained utterances.

The lack of curated datasets in the medical domain poses challenges in training robust and contextually aware AI dialogue models for healthcare-related tasks. Without access to comprehensive and domain-specific datasets, the performance and reliability of AI-based systems in the medical domain may be compromised. Therefore, the development of the appropriate datasets is crucial to unlock the full potential of AI-based dialogue systems in healthcare. Collaborative efforts and a commitment to develop in-domain datasets should be made to foster the advancement and adoption of transformative technologies, which would ultimately leads to enhanced healthcare outcomes and improved patient care.

Bibliography

[1] Andrew Radford. In: *English Syntax: An Introduction*. Cambridge University Press, 2004.

[2] Jason Merchant. "Ellipsis: A survey of analytical approaches". In: *The Oxford Handbook of Ellipsis*. Oxford University Press, Dec. 2018. ISBN: 9780198712398. DOI: 10.1093/oxfordhb/9780198712398.013.2. eprint: https://academic.oup.com/book/0/chapter/353990361/chapter-ag-pdf/45907301/book_41718_section_353990361.ag.pdf. URL: https://doi.org/10.1093/oxfordhb/9780198712398.013.2.

[3] Yue Fang et al. "From spoken dialogue to formal summary: An utterance rewriting for dialogue summarization". In: *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Seattle, United States: Association for Computational Linguistics, July 2022, pp. 3859–3869. DOI: 10.18653/v1/2022.naacl-main.283. URL: https://aclanthology.org/2022.naacl-main.283.

[4] Gangwoo Kim et al. "Learn to Resolve Conversational Dependency: A Consistency Training Framework for Conversational Question Answering". In: *Association for Computational Linguistics (ACL)*. 2021.

[5] Eric Malmi et al. "Encode, Tag, Realize: High-Precision Text Editing". In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language*

Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong, China: Association for Computational Linguistics, Nov. 2019, pp. 5054–5065. DOI: 10.18653/v1/D19-1510. URL: https://aclanthology.org/D19-1510.

[6] Jun Quan et al. "GECOR: An End-to-End Generative Ellipsis and Co-reference Resolution Model for Task-Oriented Dialogue". In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong, China: Association for Computational Linguistics, Nov. 2019, pp. 4547–4557. DOI: 10.18653/v1/D19-1462. URL: https://aclanthology.org/D19-1462.

[7] Jiatao Gu et al. "Incorporating Copying Mechanism in Sequence-to-Sequence Learning". In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Berlin, Germany: Association for Computational Linguistics, Aug. 2016, pp. 1631–1640. DOI: 10.18653/v1/P16-1154. URL: https://aclanthology.org/P16-1154.

[8] Abigail See, Peter Liu, and Christopher Manning. "Get To The Point: Summarization with Pointer-Generator Networks". In: *Association for Computational Linguistics*. 2017. URL: https://arxiv.org/abs/1704.04368.

[9] Zhufeng Pan et al. "Improving Open-Domain Dialogue Systems via Multi-Turn Incomplete Utterance Restoration". In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong, China: Association for Computational Linguistics, Nov. 2019, pp. 1824–

1833. DOI: 10.18653/v1/D19-1191. URL: https://aclanthology.org/D19-1191.

[10] Qian Liu et al. “Incomplete Utterance Rewriting as Semantic Segmentation”. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*. 2020.

[11] Mengzuo Huang et al. “SARG: A Novel Semi Autoregressive Generator for Multi-turn Incomplete Utterance Restoration”. In: *Proceedings of the AAAI Conference on Artificial Intelligence* 35.14 (May 2021), pp. 13055–13063. DOI: 10.1609/aaai.v35i14.17543. URL: https://ojs.aaai.org/index.php/AAAI/article/view/17543.

[12] Bo-Hsiang Tseng et al. “CREAD: Combined Resolution of Ellipses and Anaphora in Dialogues”. In: *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Online: Association for Computational Linguistics, June 2021, pp. 3390–3406. DOI: 10.18653/v1/2021.naacl-main.265. URL: https://aclanthology.org/2021.naacl-main.265.

[13] Ahmed Elgohary, Denis Peskov, and Jordan Boyd-Graber. “Can You Unpack That? Learning to Rewrite Questions-in-Context”. In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong, China: Association for Computational Linguistics, Nov. 2019, pp. 5918–5924. DOI: 10.18653/v1/D19-1605. URL: https://aclanthology.org/D19-1605.

[14] Scott Martin, Shivani Poddar, and Kartikeya Upasani. "MuDoCo: Corpus for Multidomain Coreference Resolution and Referring Expression Generation". English. In: *Proceedings of the Twelfth Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association, May 2020, pp. 104–111. ISBN: 979-10-95546-34-4. URL: `https://aclanthology.org/2020.lrec-1.13`.

[15] Eunsol Choi et al. "QuAC: Question Answering in Context". In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Brussels, Belgium: Association for Computational Linguistics, Oct. 2018, pp. 2174–2184. DOI: `10.18653/v1/D18-1241`. URL: `https://aclanthology.org/D18-1241`.

[16] Siva Reddy, Danqi Chen, and Christopher D. Manning. "CoQA: A Conversational Question Answering Challenge". In: *Transactions of the Association for Computational Linguistics* 7 (2019), pp. 249–266. DOI: `10.1162/tacl_a_00266`. URL: `https://aclanthology.org/Q19-1016`.

[17] Tsung-Hsien Wen et al. "Conditional Generation and Snapshot Learning in Neural Dialogue Systems". In: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Austin, Texas: Association for Computational Linguistics, Nov. 2016, pp. 2153–2162. DOI: `10.18653/v1/D16-1233`. URL: `https://aclanthology.org/D16-1233`.

[18] Halil Kilicoglu, Marcelo Fiszman, and Dina Demner-Fushman. "Interpreting Consumer Health Questions: The Role of Anaphora and Ellipsis". In: *Proceedings of the 2013 Workshop on Biomedical Natural Language Processing (BioNLP 2013)*. Sofia, Bulgaria: Association for Computational Linguistics, 2013, pp. 54–62.

[19] Faiha Fareez et al. "A dataset of simulated patient-physician medical interviews with a focus on respiratory cases". In: *Scientific Data* 9.313 (June 2022). ISSN: 2052-4463. DOI: 10.1038/s41597-022-01423-1. URL: https://doi.org/10.1038/s41597-022-01423-1.

[20] Alec Radford et al. "Language Models are Unsupervised Multitask Learners". In: (2019).

[21] Yiyang Li, Hai Zhao, and Zhuosheng Zhang. "Back to the Future: Bidirectional Information Decoupling Network for Multi-turn Dialogue Modeling". In: *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*. Abu Dhabi, United Arab Emirates: Association for Computational Linguistics, Dec. 2022, pp. 2761–2774. URL: https://aclanthology.org/2022.emnlp-main.177.

[22] Longxiang Liu et al. "Filling the Gap of Utterance-aware and Speaker-aware Representation for Multi-turn Dialogue". In: *CoRR* abs/2009.06504 (2020). arXiv: 2009.06504. URL: https://arxiv.org/abs/2009.06504.

[23] Diederik P. Kingma and Jimmy Ba. "Adam: A Method for Stochastic Optimization". In: *CoRR* abs/1412.6980 (2014).

[24] Kishore Papineni et al. "Bleu: a Method for Automatic Evaluation of Machine Translation". In: *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*. Philadelphia, Pennsylvania, USA: Association for Computational Linguistics, July 2002, pp. 311–318. DOI: 10.3115/1073083.1073135. URL: https://aclanthology.org/P02-1040.

[25] Jie Hao et al. "RAST: Domain-Robust Dialogue Rewriting as Sequence Tagging". In: *Proceedings of the 2021 Conference on Empirical Methods in Natural*

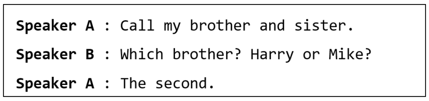
Language Processing. Online and Punta Cana, Dominican Republic: Association for Computational Linguistics, Nov. 2021, pp. 4913–4924. DOI: 10.18653/v1/2021.emnlp-main.402. URL: https://aclanthology.org/2021.emnlp-main.402.

[26] Lisa Jin et al. "Hierarchical Context Tagging for Utterance Rewriting". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 36.10 (June 2022), pp. 10849–10857. DOI: 10.1609/aaai.v36i10.21331. URL: https://ojs.aaai.org/index.php/AAAI/article/view/21331.

Appendix

A Masks

In this section, we present an example dialogue and its corresponding M_1 mask. This mask guides the model to attend to the tokens in the same utterance, thereby capturing the information of identical turn information.



Speaker A : Call my brother and sister.
Speaker B : Which brother? Harry or Mike?
Speaker A : The second.

Figure A.1 An example dialogue.

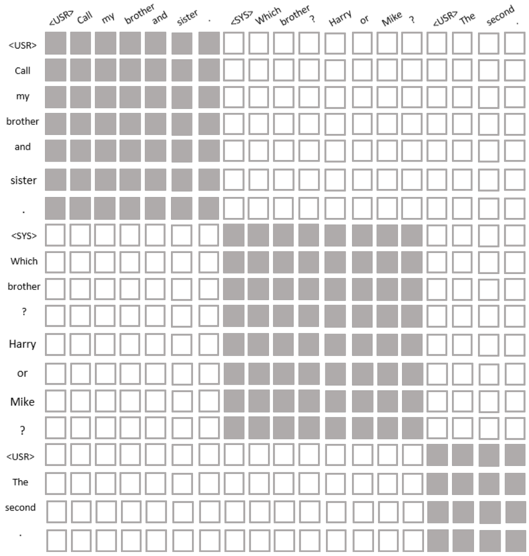
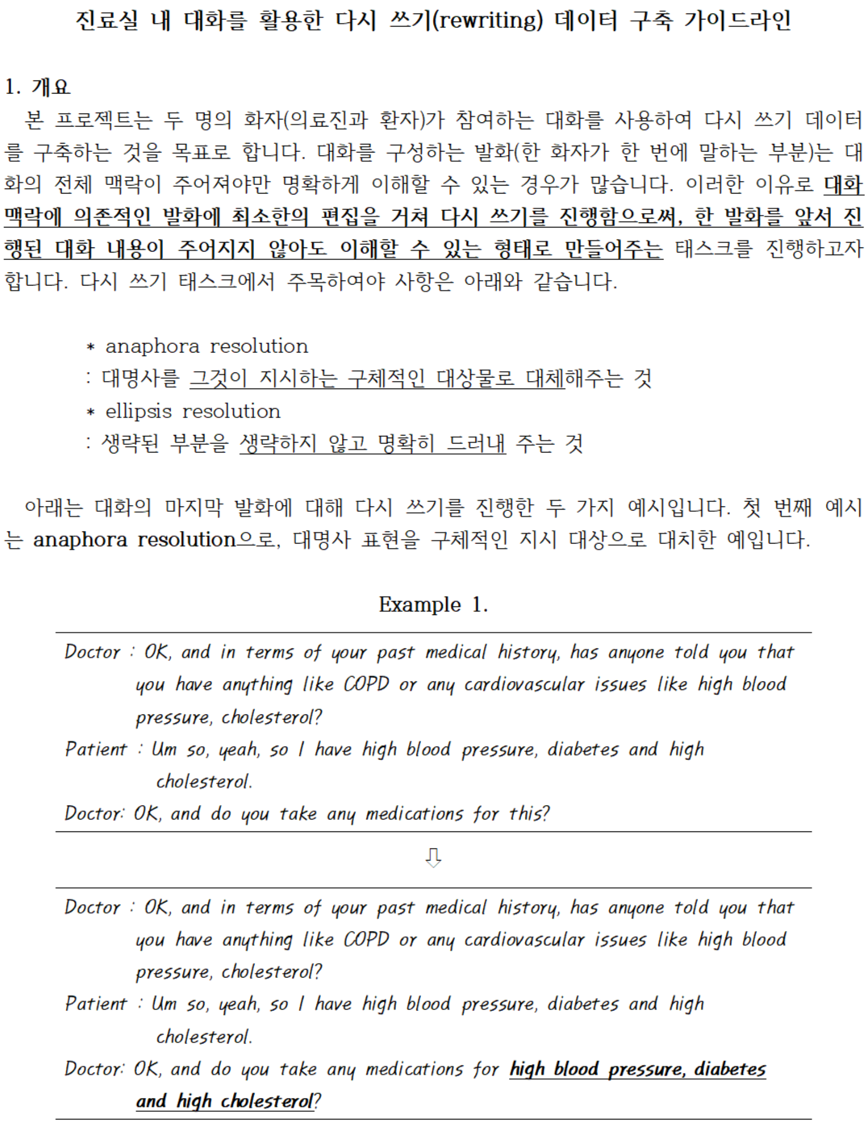


Figure A.2 Introduced mask M_1 for the example dialogue .

B Utterance Rewriting Annotation Guideline

In this section, we present an annotation guideline used for the **MedDial** rewriting dataset construction. The guideline was created based on [12].



진료실 내 대화를 활용한 다시 쓰기(rewriting) 데이터 구축 가이드라인

1. 개요

본 프로젝트는 두 명의 화자(의료진과 환자)가 참여하는 대화를 사용하여 다시 쓰기 데이터를 구축하는 것을 목표로 합니다. 대화를 구성하는 발화(한 화자가 한 번에 말하는 부분)는 대화의 전체 맥락이 주어져야만 명확하게 이해할 수 있는 경우가 많습니다. 이러한 이유로 **대화 맥락에 의존적인 발화에 최소한의 편집을 거쳐 다시 쓰기를 진행함으로써, 한 발화를 앞서 진행된 대화 내용이 주어지지 않아도 이해할 수 있는 형태로 만들어주는** 태스크를 진행하고자 합니다. 다시 쓰기 태스크에서 주목하여야 사항은 아래와 같습니다.

* anaphora resolution
: 대명사를 그것이 지시하는 구체적인 대상물로 대체해주는 것
* ellipsis resolution
: 생략된 부분을 생략하지 않고 명확히 드러내 주는 것

아래는 대화의 마지막 발화에 대해 다시 쓰기를 진행한 두 가지 예시입니다. 첫 번째 예시는 anaphora resolution으로, 대명사 표현을 구체적인 지시 대상으로 대치한 예입니다.

Example 1.

Doctor : OK, and in terms of your past medical history, has anyone told you that you have anything like COPD or any cardiovascular issues like high blood pressure, cholesterol?
Patient : Um so, yeah, so I have high blood pressure, diabetes and high cholesterol.
Doctor: OK, and do you take any medications for this?

⇩

Doctor : OK, and in terms of your past medical history, has anyone told you that you have anything like COPD or any cardiovascular issues like high blood pressure, cholesterol?
Patient : Um so, yeah, so I have high blood pressure, diabetes and high cholesterol.
Doctor: OK, and do you take any medications for ***high blood pressure, diabetes and high cholesterol****?*

가장 마지막 발화인 “*OK, and do you take any medications for this?*”만 떼어서 보았을 때 ‘이것(*this*)’이 의미하는 바가 어떤 것인지 명확히 알 수 없습니다. ‘이것(*this*)’이 지칭하는 대상은 이전의 대화 맥락에서 제시된 ‘*high blood pressure, diabetes, and high cholesterol*’임을 알 수 있습니다. 다시 쓰기를 통해 대명사인 ‘이것(*this*)’을 구체적인 지시 대상으로 대치시켜주면, 비로소 마지막 발화가 완전하게 이해될 수 있습니다.

다음 예시는 ellipsis resolution으로, 발화에서 생략된 부분을 대화 맥락에서 찾아 복원시켜준 예입니다.

Example 2.

Doctor : What brings you in here today?
Patient : Yeah, I've just been feeling breathless and it's getting worse.
So I wanted to check it out.
Doctor : OK, and when did you first start feeling breathless?
Patient : About two months ago.

⇩

Doctor : What brings you in here today?
Patient : Yeah, I've just been feeling breathless and it's getting worse.
So I wanted to check it out.
Doctor : OK, and when did you first start feeling breathless?
Patient : ***I' ve been feeling breathless*** *about two months ago.*

가장 마지막 발화인 “*About two months ago.*”만 떼어서 보았을 때 어떤 것이 ‘약 두 달 전’이었는지 알 수 없습니다. 해당 발화에서 생략된 표현이자 이전의 대화 맥락에서 제시된 ‘*I' ve been feeling breathless*(숨이 가쁜 증상이 시작된 지가)’를 복원시키면, 비로소 마지막 발화가 완전하게 이해될 수 있습니다.

다시 쓰기 태스크의 목표는 위와 같이 발화에서 **맥락이 요구되는 부분을 다시 써 줌**으로써 대화 내에서 **불명확한 부분을 줄이고 모든 발화를 완전한 형태로 복원**시켜주는 것을 목표로 합니다. 이때, 다시 쓰기가 요구되는 부분이 아닌 다른 부분은 원래 발화 그대로를 사용함으로써 **최소한의 편집**을 가하는 것을 목표로 합니다.

2. 다시 쓰기 원칙

(1) 대화의 마지막 발화에 대해 다시 쓰기를 진행함.
(2) 다시 쓰기 유형을 명시함.
(3) 대상이 되는 발화를 paraphrase하거나 요약해서는 안 됨.
(4) 최대한 문맥에서 등장했던 표현을 사용하여 다시 쓰기를 진행함.
(5) 동일 발화에서 파악할 수 있는 부분에 대해서는 다시 쓰기를 진행하지 않음.

Example 3.

Doctor : How may I help you?

Patient : Hi, yes it's nice to meet you. I've been having this cough that's been going on for the last few days and I have had difficulty breathing.

Doctor : Oh, I'm so sorry to hear that. When did this start?

⇩

Doctor : How may I help you?

Patient : Hi, yes it's nice to meet you. I've been having this cough that's been going on for the last few days and I have had difficulty breathing.

Doctor : Oh, I'm so sorry to hear that ***you’ ve been having cough that’ s been going on for the last few days and have had difficulty breathing.***
When did this start?

(1) 대화의 마지막 발화에 대해 다시 쓰기를 진행함.

마지막 의료진의 발화인 ‘*Doctor : Oh, I'm so sorry to hear that. When did this start?*‘에 대해 다시 쓰기를 진행합니다.

(2) 다시 쓰기 유형을 명시함.

마지막 발화에서 대명사 표현(anaphora)이 사용되었는지 생략 표현(ellipsis)이 사용되었는지, 혹은 다시 쓰기가 필요하지 않는 발화(no-rewrite)인지 명시합니다. 위 발화에서는 *that* 이후의 표현이 생략되었으므로, 다시 쓰기는 ellipsis resolution에 해당합니다.

(3) 대상이 되는 발화를 paraphrase하거나 요약해서는 안 됨.

다시 쓰기를 진행할 때 기본적으로 마지막 발화의 구조를 그대로 사용합니다. 즉, 위의 예시에서 ‘*Doctor : Oh, I'm so sorry to hear that. When did this start?*‘는 생략 표현을 복원하는 편집만 진행하고 이외의 부분은 그대로 둡니다.

(4) 최대한 문맥에서 등장했던 표현을 사용하여 다시 쓰기를 진행함.

대상이 되는 발화를 다시 쓰기 할 때 대화에서 등장하였던 표현을 활용합니다. 위 대화에서 의료진이 유감으로 생각하는 것은 '환자가 오래간 기침을 해왔고 숨 쉬는 것이 어렵다는 점' 이므로, *that* 이후에 '*you' ve been having cough that' s been going on for the last few days and have had difficulty breathing.*'만 추가합니다. 문맥 표현을 그대로 사용한다면 '*I*'로 써야 하지만, 다시 쓰기를 진행하는 발화의 화자를 고려했을 때 인칭을 바꾸어 '*you*'로 써줍니다.

(5) 동일 발화에서 파악할 수 있는 부분에 대해서는 다시 쓰기를 진행하지 않음.

대상이 되는 발화를 다시 쓰면 '*Oh, I'm so sorry to hear that you' ve been having cough that' s been going on for the last few days and have had difficulty breathing. When did this start?*'입니다. 대명사 표현인 this는 <u>동일 발화 내에서 지칭 대상을 찾을 수 있기 때문에 다시 쓰기를 진행하지 않습니다</u>.

국 문 초 록

대화 기반 불완전 발화 다시쓰기(Dialogue-based Incomplete Utterance Rewriting, IUR)는 대화 내에서 문맥 정보가 있어야 이해할 수 있는 발화(context-dependent utterance)를 문맥과 독립적으로 떼어 놓아도 이해 가능한 완전한 발화(self-contained utterance)로 다시 작성해주는 태스크이다. 최근 AI 기반 챗봇이 널리 사용됨에 따라 시스템이 대화를 효과적으로 처리하고 이해하는 것의 중요성이 대두되고 있다. 이 점에서 IUR은 대화 시스템의 전반적인 성능을 향상시키기 위한 접근법 가운데 하나로서 주목 받았다.

최근의 BERT와 GPT와 같은 Transformer 기반 사전 훈련 대형 언어 모델은 이 분야의 발전에 크게 기여하였다. 이러한 모델들은 언어 자체에 대한 이해를 기반으로 IUR의 성능 향상을 이끌어냈다.

본 연구에서는 대화 기반 IUR 태스크의 성능을 향상시키기 위해 대화의 본질적 특성을 활용한 새로운 접근 방식을 제안한다. 제안된 방법론은 Transformer 기반 사전 훈련 대형 언어 모델인 GPT-2에 화자 및 발화 정보를 도입함으로써, 대화의 문맥을 효과적으로 포착하여 완전한 발화를 생성할 수 있게 한다. 모델은 일반적인 글(free-texts)과 구분되는 대화의 특성을 고려할 수 있게 되어, 일관성 있고 문맥에 적합한 발화를 생성할 수 있었다. 기존 연구에서와는 달리 참조하는 단어와 실제 대상의 상호 참조 정보에 대한 추가적인 어노테이션 없이도 좋은 성능을 달성할 수 있었다는 것 또한 제안된 모델이 기여한 부분이다. 이는 훈련용 데이터 구축에 있어서 라벨링 작업의 부담을 줄여주어, 기존 접근법에 비해 더 확장 가능하고 효율적인 접근 방식을 제공할 수 있다는 것을 의미한다.

제안된 방법론을 평가하기 위해 우리는 공개된 IUR 데이터셋 세 가지와 직접 구축한 추가적인 데이터셋에 대한 포괄적인 평가를 실시하였다. 그 결과 대부분의 평가 지표에서 제안된 방법론이 기존 접근 방식보다 향상된 성능을 달성하였다. 제안된 방법론은 모델에 대화 데이터가 가지는 고유의 특성을 반영하여 성능 향상을 이끌어냈으며 상호 참조 라벨링에 의존한 기존의 접근법에 비해 훨씬 적용하기 용이하다고 판단되어, 결과적으로 대화 기반 발화 다시쓰기에 기여하였음을 확인할 수 있었다.

키워드: 다시쓰기, 대화, 발화, 상호참조, 생략, 분리-융합, GPT-2

학 번: 2021-20816