西湖大學
WESTLAKE UNIVERSITY

DialogSum Challenge: Summarizing Real-Life Scenario Dialogues

Yulong Chen	Zhejiang University, Westlake University
Naihao Deng	University of Michigan, Ann Arbor
Yang Liu	University of Edinburgh
Yue Zhang	Westlake University, Westlake Institute for Advanced Study

INLG 2021

The *DialogSum Challenge*

Task Overview

The task asks a model to ***generate a summary*** given a piece of dialogue text.

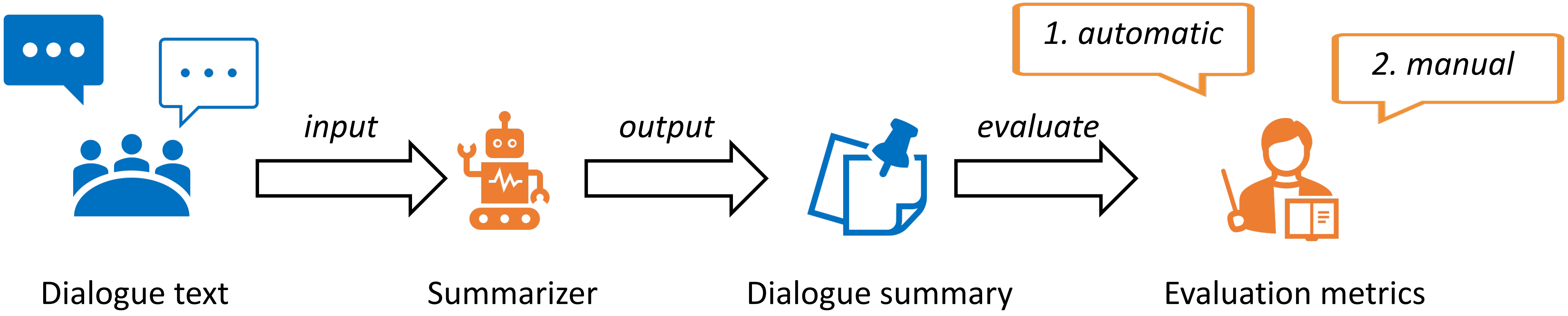


Figure 1: Task overview of *DialogSum Challenge*. Both automatic evaluation and manual evaluation from multiple aspects have been conducted to measure model performance.

Motivation

Why Dialogue Summarization?

Existing research on text summarization focuses on *monologic* texts.

(CNN) — Growing awareness of the climate crisis means conscientious travelers are increasingly looking for alternatives to trips by airplane.

For many, that means traveling by high-speed train, but on Wednesday UK-based company Hybrid Air Vehicles (HAV) released the latest details of its airship, which boasts a far smaller carbon footprint than a conventional passenger plane.

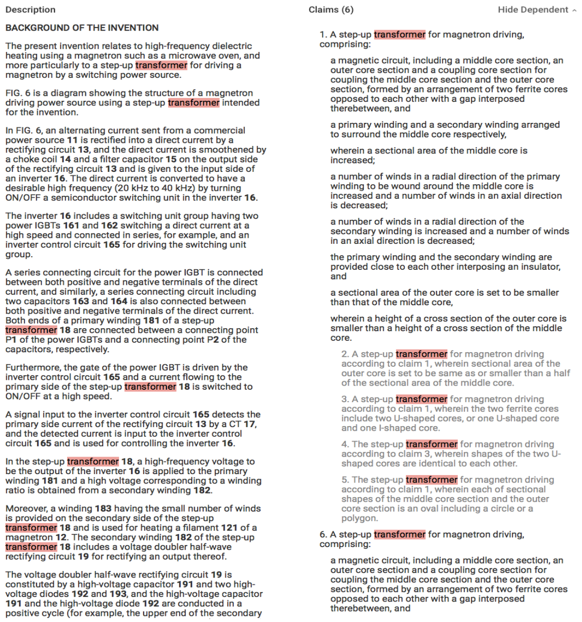
The Airlander 10 aircraft will seat up to 100 passengers and operate with 90% fewer emissions than conventional aircraft, the company said in a press release.

The airship requires less fuel than a conventional aircraft due to a combination of "buoyant lift from helium, aerodynamic lift, and vectored thrust," according to HAV.

A hybrid electric and jet fuel model will be available by 2025, and a fully electric version by 2030, according to the press release.

The Airlander 10 aircraft will operate with 90% fewer emissions than conventional aircraft, say its developers.

(a) A CNN article



Description

BACKGROUND OF THE INVENTION

The present invention relates to high-frequency dielectric heating using a magnetron such as a microwave oven, and more particularly to a step-up transformer for driving a magnetron by a switching power source.

FIG. 6 is a diagram showing the structure of a magnetron driving power source using a step-up transformer intended for the invention.

In FIG. 6, an alternating current sent from a commercial power source **11** is rectified into a direct current by a rectifying circuit **13**, and the direct current is smoothened by a choke coil **14** and a filter capacitor **15** on the output side of the rectifying circuit **13** and is given to the input side of an inverter **16**. The direct current is converted to have a desirable high frequency (20 kHz to 40 kHz) by turning ON/OFF a semiconductor switching unit in the inverter **16**.

The inverter **16** includes a switching unit group having two power IGBTs **161** and **162** switching a direct current at a high speed and connected in series, for example, and an inverter control circuit **165** for driving the switching unit group.

A series connecting circuit for the power IGBT is connected between both positive and negative terminals of the direct current, and similarly, a series connecting circuit including two capacitors **163** and **164** is also connected between both positive and negative terminals of the direct current. Both ends of a primary winding **181** of a step-up transformer **18** are connected between a connecting point P**1** of the power IGBTs and a connecting point P**2** of the capacitors, respectively.

Furthermore, the gate of the power IGBT is driven by the inverter control circuit **165** and a current flowing to the primary side of the step-up transformer **18** is switched to ON/OFF at a high speed.

A signal input to the inverter control circuit **165** detects the primary side current of the rectifying circuit **13** by a CT **17**, and the detected current is input to the inverter control circuit **165** and is used for controlling the inverter **16**.

In the step-up transformer **18**, a high-frequency voltage to be the output of the inverter **16** is applied to the primary winding **181** and a high voltage corresponding to a winding ratio is obtained from a secondary winding **182**.

Moreover, a winding **183** having the small number of winds is provided on the secondary side of the step-up transformer **18** and is used for heating a filament **121** of a magnetron **12**. The secondary winding **182** of the step-up transformer **18** includes a voltage doubler half-wave rectifying circuit **19** for rectifying an output thereof.

The voltage doubler half-wave rectifying circuit **19** is constituted by a high-voltage capacitor **191** and two high-voltage diodes **192** and **193**, and the high-voltage capacitor **191** and the high-voltage diode **192** are conducted in a positive cycle (for example, the upper end of the secondary

Claims (6) Hide Dependent

1. A step-up transformer for magnetron driving, comprising:

a magnetic circuit, including a middle core section, an outer core section and a coupling core section for coupling the middle core section and the outer core section, formed by an arrangement of two ferrite cores opposed to each other with a gap interposed therebetween, and

a primary winding and a secondary winding arranged to surround the middle core respectively,

wherein a sectional area of the middle core is increased;

a number of winds in a radial direction of the primary winding to be wound around the middle core is increased and a number of winds in an axial direction is decreased;

a number of winds in a radial direction of the secondary winding is increased and a number of winds in an axial direction is decreased;

the primary winding and the secondary winding are provided close to each other interposing an insulator, and

a sectional area of the outer core is set to be smaller than that of the middle core,

wherein a height of a cross section of the outer core is smaller than a height of a cross section of the middle core.

2. A step-up transformer for magnetron driving according to claim 1, wherein sectional area of the outer core is set to be same as or smaller than a half of the sectional area of the middle core.

3. A step-up transformer for magnetron driving according to claim 1, wherein the two ferrite cores include two U-shaped cores, or one U-shaped core and one I-shaped core.

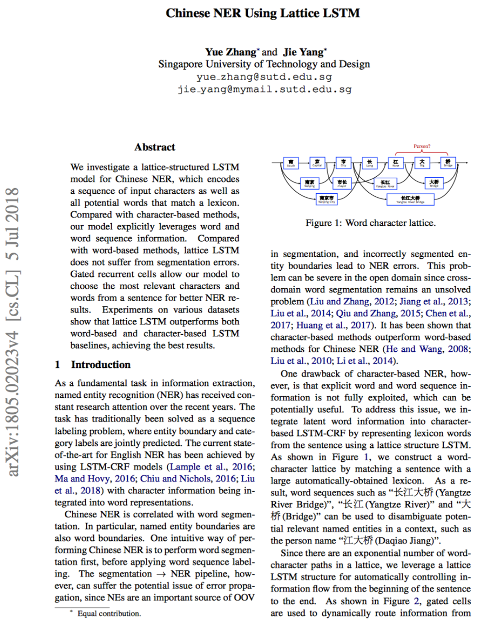
4. The step-up transformer for magnetron driving according to claim 3, wherein shapes of the two U-shaped cores are identical to each other.

5. The step-up transformer for magnetron driving according to claim 1, wherein each of sectional shapes of the middle core section and the outer core section is an oval including a circle or a polygon.

6. A step-up transformer for magnetron driving, comprising:

a magnetic circuit, including a middle core section, an outer core section and a coupling core section for coupling the middle core section and the outer core section, formed by an arrangement of two ferrite cores opposed to each other with a gap interposed therebetween, and

(b) A patent paper



arXiv:1805.02023v4 [cs.CL] 5 Jul 2018

Chinese NER Using Lattice LSTM

Yue Zhang* and **Jie Yang***
Singapore University of Technology and Design
yue_zhang@sutd.edu.sg
jie_yang@mymail.sutd.edu.sg

Abstract

We investigate a lattice-structured LSTM model for Chinese NER, which encodes a sequence of input characters as well as all potential words that match a lexicon. Compared with character-based methods, our model explicitly leverages word and word sequence information. Compared with word-based methods, lattice LSTM does not suffer from segmentation errors. Gated recurrent cells allow our model to choose the most relevant characters and words from a sentence for better NER results. Experiments on various datasets show that lattice LSTM outperforms both word-based and character-based LSTM baselines, achieving the best results.

1 Introduction

As a fundamental task in information extraction, named entity recognition (NER) has received constant research attention over the recent years. The task has traditionally been solved as a sequence labeling problem, where entity boundary and category labels are jointly predicted. The current state-of-the-art for English NER has been achieved by using LSTM-CRF models (Lample et al., 2016; Ma and Hovy, 2016; Chiu and Nichols, 2016; Liu et al., 2018) with character information being integrated into word representations.

Chinese NER is correlated with word segmentation. In particular, named entity boundaries are also word boundaries. One intuitive way of performing Chinese NER is to perform word segmentation first, before applying word sequence labeling. The segmentation → NER pipeline, however, can suffer the potential issue of error propagation, since NEs are an important source of OOV

* Equal contribution.

Figure 1: Word character lattice.

in segmentation, and incorrectly segmented entity boundaries lead to NER errors. This problem can be severe in the open domain since cross-domain word segmentation remains an unsolved problem (Liu and Zhang, 2012; Jiang et al., 2013; Liu et al., 2014; Qiu and Zhang, 2015; Chen et al., 2017; Huang et al., 2017). It has been shown that character-based methods outperform word-based methods for Chinese NER (He and Wang, 2008; Liu et al., 2010; Li et al., 2014).

One drawback of character-based NER, however, is that explicit word and word sequence information is not fully exploited, which can be potentially useful. To address this issue, we integrate latent word information into character-based LSTM-CRF by representing lexicon words from the sentence using a lattice structure LSTM. As shown in Figure 1, we construct a word-character lattice by matching a sentence with a large automatically-obtained lexicon. As a result, word sequences such as "长江大桥 (Yangtze River Bridge)", "长江 (Yangtze River)" and "大桥 (Bridge)" can be used to disambiguate potential relevant named entities in a context, such as the person name "江大桥 (Daqiao Jiang)".

Since there are an exponential number of word-character paths in a lattice, we leverage a lattice LSTM structure for automatically controlling information flow from the beginning of the sentence to the end. As shown in Figure 2, gated cells are used to dynamically route information from

(c) An academic paper

Figure 2: Examples of monologic text summarization data.

Motivation

Why Dialogue Summarization?

Dialogue, as an important communicative channel, has received significantly less attention.

(a) An academic discussion

(b) A job interview

(c) A business negotiation

Figure 3: Examples of dialogues under different scenarios.

Motivation

Why Dialogue Summarization?

Summarization dialogues can be beneficial to both business and personal uses.

- Find common needs for businesses
- Record complains from customers
- Keep personal medical records
- Manage personal schedule
- …

However, we find that dialogue summarization is very ***different*** from monologue summarization and contain ***unique challenges***.

Challenges in *DialogSum*

Challenges from Source Dialogue Text

Dialogues contain ***special linguistic phenomena***, making dialogues difficult to be encoded using ordinary representation technologies.

- **Dialogue discourse structures**
- **Dialogue terms**
- **Coreference and ellipsis**

For example, two utterances can be closely related even where there is ***a large distance*** between them.

Such phenomenon is common in spoken dialogues, in particular, ***negotiations*** and ***procedures***.

DIALOGUE - A:
#Person1#: Hello. 9-1- 1. Can I help you?
#Person2#: **I need the police.**
#Person1#: **What happened?**
#Person2#: My neighbor hit my brother on the head. He's bleeding.
#Person1#: **Give me your address.**
#Person2#: 176 Wooden Street East.
#Person1#: All right. **The police and an ambulance are on the way.** In the meantime, find a clean cloth and press it firmly over the wound. This will slow the bleeding.
#Person2#: I will, but please hurry.
#Person1#: Help will be there in just a few minutes.

Figure 4: An example for dialogue discourse structure.

Challenges in *DialogSum*

Challenges from Source Dialogue Text

DIALOGUE - B:
#Person_1#: Good morning. What can I do for you?
#Person_2#: I'm in Room 309. I'm checking out today. Can I have my bill now?
#Person_1#: Certainly. Please wait a moment. Here you are.
#Person_2#: Thanks. Wait…What's this? The 30 dollar for?
#Person_1#: Excuse me… The charge for your laundry service on Nov. 20th.
#Person_2#: But I didn't take any laundry service during my stay here. **I think you have added someone else's.**
#Person_1#: Ummm…Sorry, would you mind waiting a moment? We check it with the department concerned.
#Person_2#: No. As long as we get this straightened out.
#Person_1#: I'm very sorry. There has been a mistake. We'll corrected the bill. Please take a look.
#Person_2#: Okay, here you are.
#Person_1#: Goodbye.

SUMMARY – B1: #Person_2# is checking out and asks #Person1# for the bill. #Person1# gives #Person_2# a **wrong** bill at first then corrects it.
SUMMARY – B2: #Person_1# helps #Person_2# correct a **mischarged** bill on laundry service and helps #Person_2# check out.
SUMMARY – B3: #Person_2# finds #Person_2# being **mischarged**. #Person_1# corrects the bill and #Person_2# pays for it.

UniLMv2: #Person_2# is checking out. #Person_1# finds #Person_2# has added someone else's laundry service . #Person_1# apologizes and will correct the bill.

Transformer: #Person_2# checks out with #Person_2#'s assistance and thinks they'll be very sorry for the laundry service.

Ellipsis:
"I think you have added someone else's" ⇒
"I think you have added someone else's *laundry service on my bill*"

Coreference:
"*my* bill" ⇒ "*#Person_2#'s* bill"

Figure 5: Case study with paired human-annotated/system-generated summaries.

Challenges in *DialogSum*

Pragmatics – Summarizing beyond Sentence's Semantics

DIALOGUE - C:	SUMMARIES
#Person_1#: Hello, so how are we feeling today? **#Person_2#:** Things are going well for me, doctor. **#Person_1#:** Am I correct in thinking that you are here for your annual physical? **#Person_2#: Yes, I am applying for new health insurance, and I need a physical examination to qualify.** **#Person_1#: Your basic physical exam will include lungs, heart, blood levels, and eyes, ears, and nose.** **#Person_2#:** I've been having a little trouble breathing. Would you look into that, please? **#Person_1#:** We can do an allergy test, and later I can send you for an asthma test. **#Person_2#:** I would appreciate it. When you give me a blood test, what are you looking for? **#Person_1#: I am going to check your cholesterol, blood sugar, and white blood cell count.** **#Person_2#:** I am expecting the tests to go well. I have been taking good care of myself.	**SUMMARY – C1:** #Person_2# **wants** to do an annual physical examination to apply for new health insurance and says #Person_2#'s breathing is not good. #Person_1# explains the items and will do tests on #Person_2#'s breathing. **SUMMARY – C2:** #Person_1# **explains** the checking items in #Person_2#'s annual physical examination and will do test to look into #Person_2's breathing. **SUMMARY – C3:** #Person_2# is going through an annual physical examination to apply for new health insurance, and #Person_2# asks #Person_1# to look into the breathing. **UNILMV2:** #Person_2# comes to #Person_1#'s annual physical to apply for new health insurance. #Person_1# will do an allergy test, an asthma test, and a blood test. **Transformer:** #Person_2# goes to #Person_1# for an annual physical examination. #Person_1# will send #Person_1# for an asthma test and what #Person_2# eats.

Figure 6: Case study with paired human-annotated/system-generated summaries.

The *DialogSum Challenge*

Data

We have prepared the DIALOGSUM dataset for this challenge.

DIALOGSUM contains 13,460 dialogues and their corresponding summaries, carefully annotated by experts. It contains dialogue under rich scenarios, including diverse task oriented dialogues.

To ensure the fairness, we have built a hidden test set consisting of 100 dialogues and their paired summaries.

Submissions

IITP-CUNI:

- Pre-trained BART large model
- Auxiliary task of extractive summarization from AMI dataset

UoT:

- Pre-trained BART large model, further tuned on CNN/Daily News Corpus
- Penalize longer summaries in decoding
- Post-process summaries to resolve generation errors

TCS_WITM:

- Pre-trained PEGASUS large model, further tuned on CNN/Daily News Corpus
- Concatenate topics with the dialogue text to feed to the model

The *DialogSum Challenge*

Evaluation

Our evaluation contains both automatic evaluation metric and human evaluation.

- **Automatic Evaluation**
  ROUGE and BERTScore

- **Human Evaluation**
  - Scale of -1 to 1
    - Coreference Information
    - Intent Identification
    - Discourse Relation
    - Objective Description
  - Scale of 0 to 5
    - Standard Summarization Metrics: Fluency, Consistency, Relevance and Coherence
    - Overall Score

Submissions

	Public Test Set				Hidden Test Set			
Model	R1	R2	RL	BERTScore	R1	R2	RL	BERTScore
Human	**53.35**	**26.72**	**50.84**	92.63	-	-	-	-
GoodBai	**47.61**	**21.66**	45.48	**92.72**	49.66	**26.03**	**48.55**	91.69
UoT	47.29	21.65	**45.92**	92.26	**49.75**	25.15	46.50	**91.76**
IITP-CUNI	47.26	21.18	45.17	92.70	45.89	21.88	43.16	91.13

Table 1: Scores by automatic metrics for each submission and human results. We embolden the top scores among models, as well as the human score if it is the highest among all the scores.

	Public Test Set				Hidden Test Set			
Model	R1	R2	RL	BERTScore	R1	R2	RL	BERTScore
TCS_WITM	47.02	21.20	44.90	90.13	50.32	25.59	47.40	91.81

Table 2: Scores by automatic metrics for the submission from TCS_WITM. The model submitted by TCS_WITM predicts 3 summaries based on the 3 topics in the public test set. We take the highest score among the 3 summaries to calculate the scores.

Submissions

	Public Test Set						Hidden Test Set					
Model	CoRef	Dis	Obj	Intent	Summ	Over	CoRef	Dis	Obj	Intent	Summ	Over
Perfect Score	1.00	1.00	1.00	1.00	5.00	5.00	1.00	1.00	1.00	1.00	5.00	5.00
GoodBai	0.96	0.86	**1.00**	0.72	4.12	3.96	**0.90**	**0.90**	**1.00**	0.70	**4.20**	**4.15**
UoT	**0.98**	**0.92**	**1.00**	**0.80**	**4.18**	**4.08**	0.75	0.75	**1.00**	**0.80**	4.00	3.70
IITP-CUNI	0.88	0.66	0.96	0.76	3.94	3.64	0.75	0.85	**1.00**	0.70	3.80	3.70

Table 4: Prediction results by one of the annotators of the *DialogSum* dataset. “CoRef”, “Dis”, “Obj”, “Intent”, “Summ”, “Over” indicates coreference information, discourse relation, objective description, intent identification, standard summarization metrics and overall scores, respectively. We embolden the best scores for each column.

	Public Test Set						Hidden Test Set					
Model	CoRef	Dis	Obj	Intent	Summ	Over	CoRef	Dis	Obj	Intent	Summ	Over
TCS_WITM	0.88	0.90	1.00	0.82	4.20	4.10	0.90	0.80	0.84	0.70	3.95	3.80

Table 5: Prediction results by one of the annotators of the *DialogSum* dataset for TCS_WITM.

Take-Aways

Size of the corpus affects the performance:

- UoT (tuned with CNN/Daily News Corpus) outperforms IITP-CUNI (tuned with AMI dataset)
- This might be that CNN/Daily News Corpus (312,000 articles) is larger than AMI dataset (137 meetings)

Incorporating topics can help the performance:

- TCS_WITM which concatenates topics with the dialogue text achieves 50.32 ROUGE-1 score on the hidden test case

Rooms for improvement:

- Even the best-performed model still underperforms humans by a margin larger than 5.0 for all the ROUGE scores

Deviation of BERTScore from Human scores:

- Although Goodbai achieves the best BERTScore on the public test set, it is TCS_WITM with a lower BERTScore achieves the best Overview score.

Error Analysis

The model makes a mistake in terms of identifying the intent of #Person2#.
—>
This suggests reasoning over the discourse to figure out Tom is #Person2# might be a challenging task for the model.

Dialogue:
#Person1#: What time is it, Tom?
#Person2#: Just a minute. It's ten to nine by my watch.
#Person1#: Is it? I had no idea it was so late. I must be off now.
#Person2#: What's the hurry?
#Person1#: I must catch the nine-thirty train.
#Person2#: You've plenty of time yet. The railway station is very close. It won't take more than twenty minutes to get there.

Gold:
Summary: #Person1# is catching a train. Tom asks #Person1# not to hurry.

IITP-CUNI:
Prediction: #Person1# and Tom are in a hurry to catch the nine-thirty train.

Figure 3: An error example where the model fails to distinguish the intent between #Person1# and Tom (#Person2#). [2]

Error Analysis

The models fail to capture the context of the dialogue in the generated summaries.
—>
The model might only capture salient information instead of information wanted in the gold summary.

Dialogue:
#Person1#: Excuse me, could you tell me how to get to the school clinic? I've lost my way.
#Person2#: Yes. Go straight ahead till you come to the traffic lights, turn left there and it's the first turning on the right.
#Person1#: Straight ahead to the traffic lights, left and then right.
#Person2#: That's it. It'll take you about five minutes.
#Person1#: Thank you very much.

Gold:
Summary: #Person1# is lost on the way to the school clinic. #Person2# shows #Person1# the correct direction.

IIPT-CUNI:
Prediction: #Person2# shows #Person1# the way to the school clinic.
UoT:
Prediction: #Person2# shows #Person1# the way to the school clinic.
TCS_WITM:
Prediction: #Person2# tells #Person1# how to get to the school clinic.
GoodBai:
Prediction: #Person2# tells #Person1# how to get to the school clinic.

Figure 2: An example where all the predicted summaries miss the the context #Person1# is lost, while all of the gold summaries contain this context. However, all the predicted summaries successfully capture the event of #Person2# shows #Person1# the direction. [2]

QA

Thank you for listening.