Do Boat and Ocean Suggest Beach? Dialogue Summarization with External Knowledge

Tianqing Fang[1], Haojie Pan[1], Hongming Zhang[1], Yangqiu Song[1], Kun Xu[2], Dong Yu[2]

1: CSE, HKUST, Hong Kong, China

2: Tencent AI Lab, Bellevue WA, USA

Introduction

- Human can infer meaning from the situational context even though the meaning is not literally expressed.
- In human dialogues, we may think of *beach* when the scene is describing *ocean* and *abandoned boat*.
- Neural summarization models should take such inference into account.

A: Does the ***ocean*** appear calm or choppy?
B: Calm.
A: Can you see any other ***boats***?
B: One, also ***abandoned***.

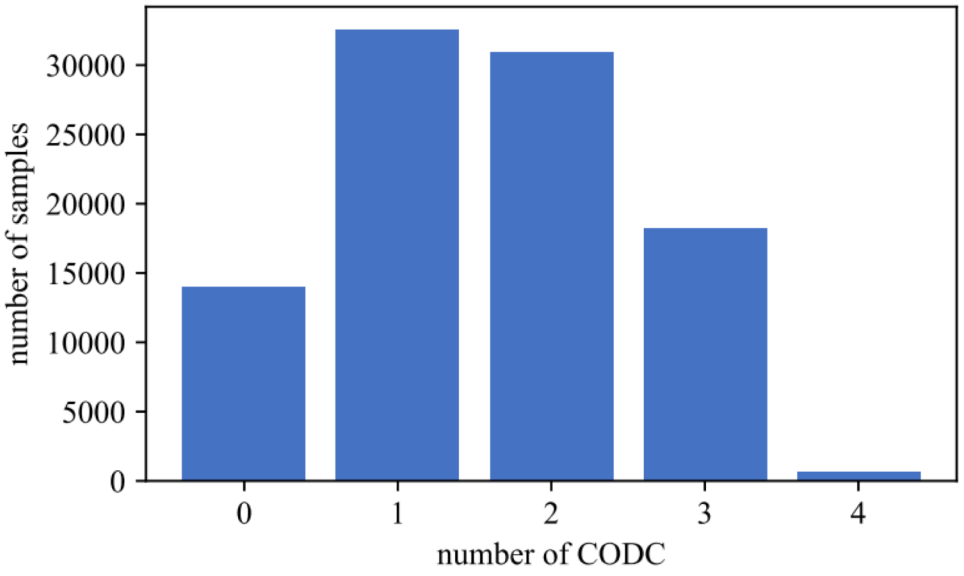
Ground : *A boat sitting on a **beach** next to an ocean*

A: Is ***boat*** on fire?
B: no , it 's exhaust I think.
A: what kind of water is it on , ***ocean*** , ***lake***?
B: ***ocean***.

Ground : *A small boat is sailing through the ocean*

Dataset

- Dialogues from VisDial dataset and image captions from MSCOCO. Aligned together.
- Define Concepts Out of Dialogue Context (CODC) as concepts that appears in the summarization while not appeared in the dialogue.

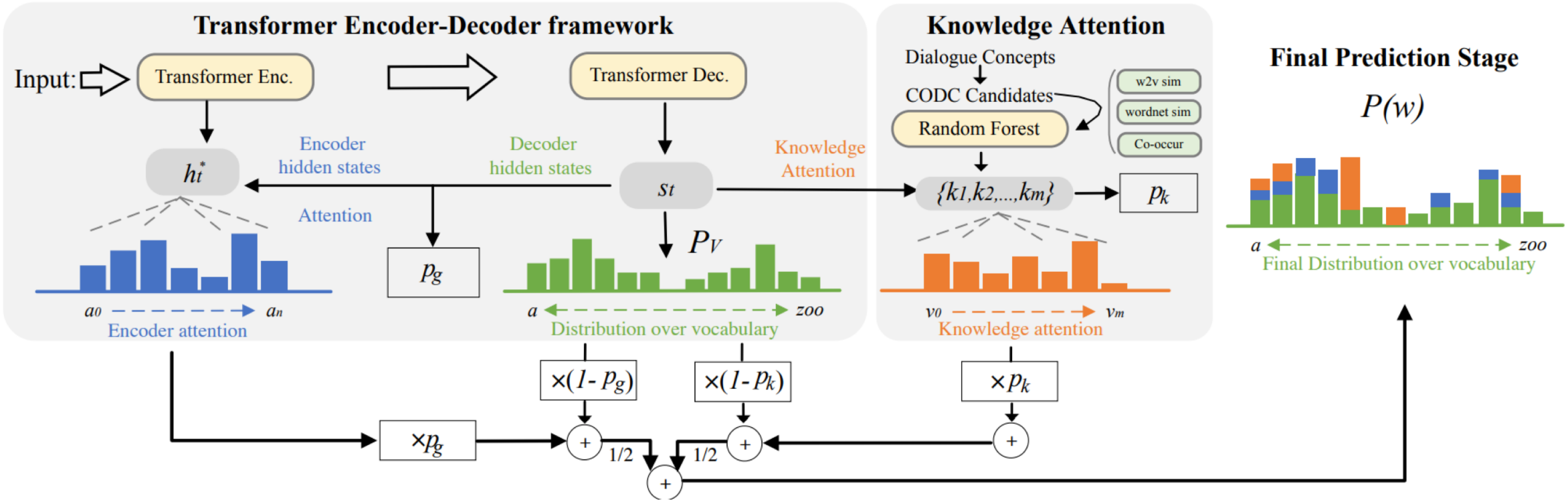


Description
Two small planes flying with each other in the *sky*

Dialogue Concepts	CODC
plane, *propeller*, *wing*, ...	*sky*

Dialogue
A: What color is the closest *plane*? B: Blue. A: What color is the other *plane*? B: Yellow. A: Are they flying side by side? B: Yes. A: Do you see any people? B: No. A: Do you see *propeller* on the closet *plane*? B: No. A: Do you see both *wings* on furthest *plane*? B: Yes. A: Do you see both *wings* on blue *plane*? B: Yes. A: Can you see the *ground*? B: *No*. A: Are there any other *planes*? B: No. A: Do you see any *clouds*? B: No.

Model


Transformer Encoder-Decoder framework
Input:
Transformer Enc.
Transformer Dec.
Encoder hidden states
Decoder hidden states
Knowledge Attention
Attention
h_t^*
s_t
p_g
P_V
a_0
a_n
Encoder attention
a
zoo
Distribution over vocabulary
Knowledge Attention
Dialogue Concepts
CODC Candidates
Random Forest
w2v sim
wordnet sim
Co-occur
$\{k_1,k_2,...,k_m\}$
p_k
v_0
v_m
Knowledge attention
Final Prediction Stage
$P(w)$
a
zoo
Final Distribution over vocabulary
$\times(1-p_g)$
$\times(1-p_k)$
$\times p_k$
$\times p_g$
+
1/2
1/2

Experiments

- The proposed Trans-KnowAttn model can outperforms it's baseline Trans-Copy in terms of both conventional automatic evaluation metrics, but also metrics regarding CODC inference.

Method	BLEU-4	METEOR	ROUGE-l	CIDEr	P_{CODC}	R_{CODC}	$F1_{CODC}$
BertSum	23.49	22.89	49.38	79.94	36.48	38.90	37.65
S2S-Attn	29.90	24.51	52.45	96.55	44.32	42.46	43.37
PGN	30.12	24.58	52.66	97.97	45.36	42.49	43.88
Pair-encoder	**31.26**	25.34	53.26	101.04	45.10	44.39	44.74
Trans-Copy	31.09	25.54	53.38	102.81	46.20	44.55	45.36
Trans-KnowAttn	31.22	**25.93**	**53.70**	**104.00***	**46.31**	**45.66***	**45.98***