PERSONA-BASED DIALOGUES

ADITYA MOHAN MISHRA

IIT DELHI

ME1180582@IITD.AC.IN

MARCH 6,2022

Overview

- Requirement of persona-Based dialogues
- Discussion of the largest persona based dataset available
- Discussion of few persona based dialogue models

CHALLENGES IN A CHIT-CHAT MODEL

message	Where do you live now?
response	I live in Los Angeles.
message	In which city do you live now?
response	I live in Madrid.
message	In which country do you live now?
response	England, you?
message	Where were you born?
response	I was born in Canada.
message	Where are you from?
response	England, you?
message	Where did you grow up?
response	I grew up in Texas.

message	How old are you?
response	16 and you?
message	What's your age?
response	18.
message	What is your major?
response	I'm majoring in psychology
message	What did you study in college?
response	English lit.

Table 1: Inconsistent responses generated by a 4-layer SEQ2SEQ model trained on 25 million Twitter conversation snippets.

Source: [Li et al]

What is a persona?

- According to Li et al persona is composed of:
  - Background facts
  - Language behavior
  - Interaction style

PERSONALISED RESPONSE GENERATION

A Persona-Based Neural Conversation Model

Jiwei Li[1*] Michel Galley[2] Chris Brockett[2]
Georgios P. Spithourakis[3*] Jianfeng Gao[2] Bill Dolan[2]

[1]Stanford University, Dept. of Computer Science
jiweil@stanford.edu
[2]Microsoft Research, Redmond
{mgalley,chrisbkt,jfgao,billdol}@microsoft.com
[3]University College London, Dept. of Computer Science
g.spithourakis@cs.ucl.ac.uk

ACL, 2016

PROBLEM STATEMENT:

4.1 Notation

For the response generation task, let M denote the input word sequence (message) $M = \{m_1, m_2, ..., m_I\}$. R denotes the word sequence in response to M, where $R = \{r_1, r_2, ..., r_J, EOS\}$ and J is the length of the response (terminated by an EOS token). r_t denotes a word token that is associated with a K dimensional distinct word embedding e_t. V is the vocabulary size.

DATASETS USED:

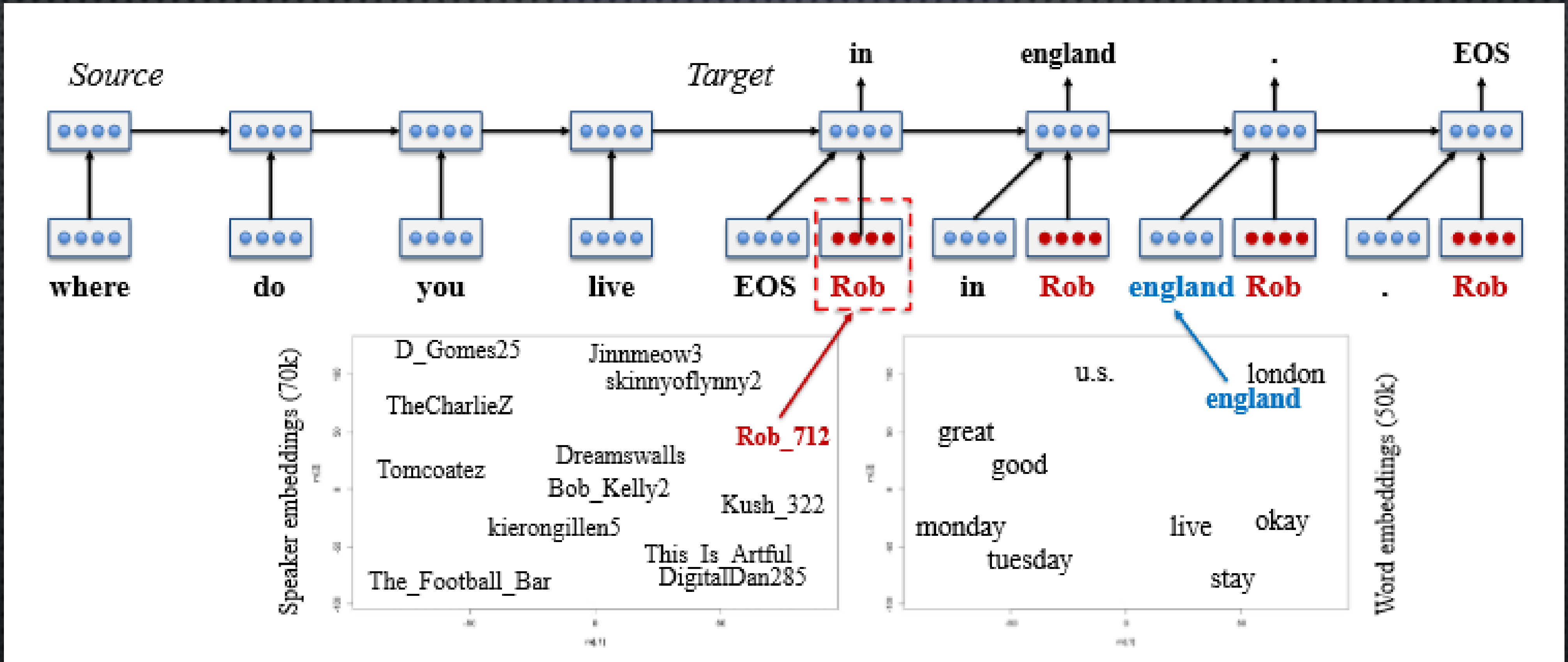
- TWITTER PERSONA DATASET
- TWITTER SORDONI DATASET
- TELEVISION SERIES TRANSCRIPTS

MODELS

SPEAKER MODEL

SPEAKER-ADDRESSEE MODEL

SPEAKER MODEL

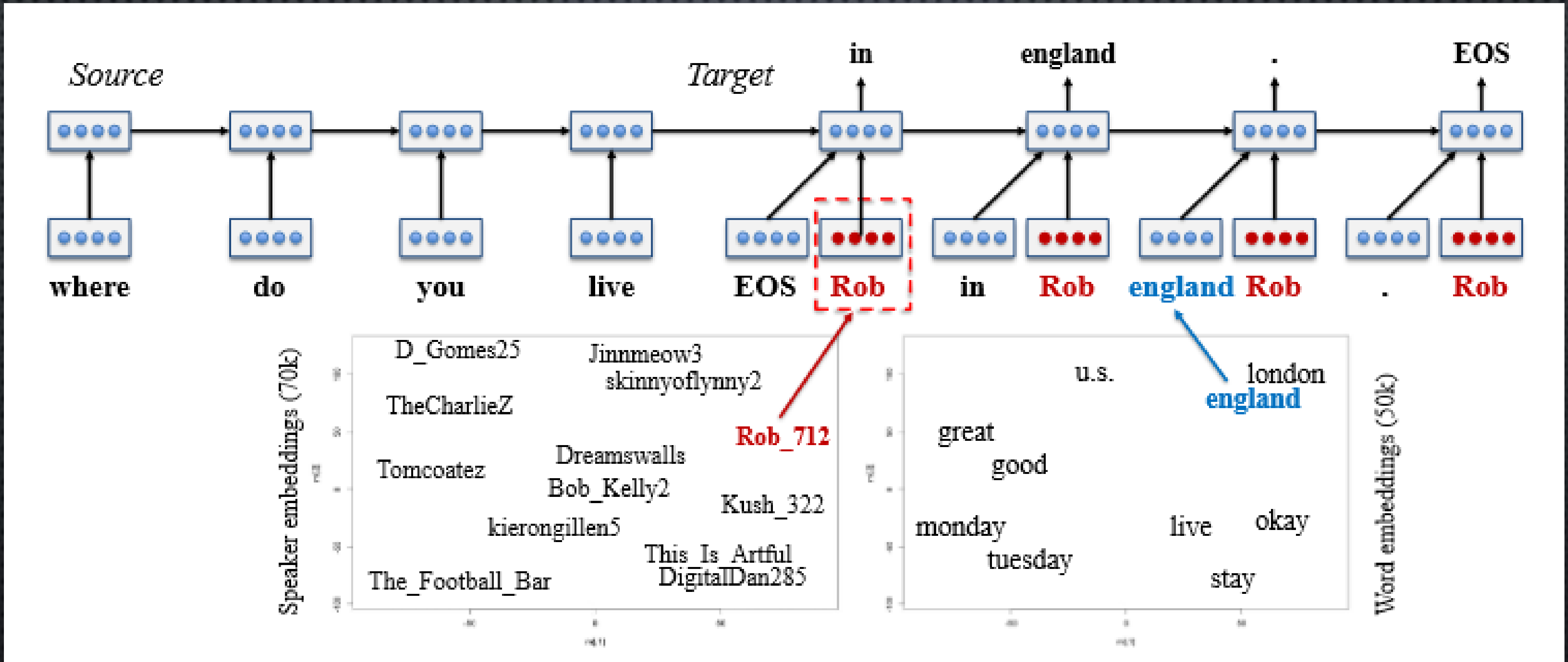


$$\begin{bmatrix} i_t \\ f_t \\ o_t \\ l_t \end{bmatrix} = \begin{bmatrix} \sigma \\ \sigma \\ \sigma \\ \tanh \end{bmatrix} W \cdot \begin{bmatrix} h_{t-1} \\ e_t^s \\ v_i \end{bmatrix}$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot l_t$$

$$h_t^s = o_t \cdot \tanh(c_t)$$

SPEAKER-ADDRESSEE MODEL



$$V_{i,j} = \tanh(W_1 \cdot v_i + W_2 \cdot v_2)$$

Correction : v_j instead of v_2

$$\begin{bmatrix} i_t \\ f_t \\ o_t \\ l_t \end{bmatrix} = \begin{bmatrix} \sigma \\ \sigma \\ \sigma \\ \tanh \end{bmatrix} W \cdot \begin{bmatrix} h_{t-1} \\ e_t^s \\ V_{i,j} \end{bmatrix}$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot l_t$$

$$h_t^s = o_t \cdot \tanh(c_t)$$

DECODING : RERANKING

- N-best list with Beam search (N=20,B=200)
- Reranking

  Score function: $\log p(R|M,v) + \lambda \log p(M|R) + \gamma |R|$

RESULTS

Model	Standard LSTM	Speaker Model	Speaker-Addressee Model
Perplexity	27.3	25.4 (−7.0%)	25.0 (−8.4%)

Table 5: Perplexity for standard SEQ2SEQ and persona models on the TV series dataset.

Model	Standard LSTM	Speaker Model	Speaker-Addressee Model
MLE	1.60%	1.82% (+13.7%)	1.83% (+14.3%)
MMI	1.70%	1.90% (+10.6%)	1.88% (+10.9%)

Table 6: BLEU on the TV series dataset (1 reference), for the standard SEQ2SEQ and persona models.

Model	Standard LSTM	Speaker Model
Perplexity	47.2	42.2 (−10.6%)

Table 3: Perplexity for standard SEQ2SEQ and the Speaker model on the Twitter Persona development set.

Model	Objective	BLEU
Standard LSTM	MLE	0.92%
Speaker Model	MLE	1.12% (+21.7%)
Standard LSTM	MMI	1.41%
Speaker Model	MMI	1.66% (+11.7%)

Table 4: BLEU on the Twitter Persona dataset (1 reference), for the standard SEQ2SEQ model and the Speaker model using as objective either maximum likelihood (MLE) or maximum mutual information (MMI).

Speaker model trained on Twitter Persona Dataset

Consistent Replies

User1	
message	Where is your hometown?
response	I'm from England.
message	Where are you from?
response	I'm from England.
message	In which city do you live now?
response	I live in London.
message	In which country do you live now?
response	I live in England.
message	What is your major?
response	Business. You?
message	What did you study in college?
response	I did business studies.
message	How old are you?
response	I'm 18.
message	What is your age?
response	I'm 18.

Inconsistent Replies

User2	
message	Where is your hometown?
response	I live in Indonesia.
message	Where are you from?
response	England, you?
message	In which city do you live now?
response	I live in Jakarta.
message	In which country do you live now?
response	I live in Indonesia.
message	What is your major?
response	Business, you?
message	What did you study in college?
response	Psychology, you?
message	How old are you?
response	I'm 18.
message	What is your age?
response	I'm 16.

Speaker-Addressee model

message	Do you love me?
baseline	Of course, I love you.
	addressee: Emily, speaker: Sheldon
response	Of course, I love you, Emily.
	addressee: Emily, speaker: Penny
response	Of course, I love you, Emily.
	addressee: Leonard, speaker: Penny
response	Of course, I love you, Leonard.
	addressee: Monica, speaker: Chandler
response	Of course I love you. I want to marry you.
	addressee: Rachel, speaker: Chandler
response	Of course I love you.
	addressee: Ross, speaker: Chandler
response	Of course I love you.
	addressee: Ross, speaker: Rachel
response	Of course I love you. (kisses him)
	addressee: Emily, speaker: Rachel
response	Of course I love you.

Table 8: Responses to *Do you love me?* from the *Speaker-Addressee* model on the TV-series dataset using different addressees and speakers.

KEY CONTRIBUTIONS:

- THE SYSTEMS OUTPERFORM OUR BASELINE SEQ2SEQ SYSTEMS IN TERMS OF BLEU, PERPLEXITY, AND HUMAN JUDGMENTS OF SPEAKER CONSISTENCY.
- DEMONSTRATED THAT ENCODING PERSONAS AS TOKENS CAPTURE SPEAKING STYLE AND BACKGROUND INFORMATION.

PERSONA-CHAT DATASET

Personalizing Dialogue Agents: I have a dog, do you have pets too?

Saizheng Zhang[†,1], Emily Dinan[‡], Jack Urbanek[‡], Arthur Szlam[‡], Douwe Kiela[‡], Jason Weston[‡]
[†] Montreal Institute for Learning Algorithms, MILA
[‡] Facebook AI Research
saizheng.zhang@umontreal.ca, {edinan,jju,aszlam,dkiela,jase}@fb.com

ACL, 2018 Zhang et. al.

- LI ET AL: PERSONA EMBEDDINGS
- ZHANG ET. AL. : FOCUS EXPLICIT AND INTERPRETABLE PROFILE INFORMATION

Persona 1	**Persona 2**
I like to ski	I am an artist
My wife does not like me anymore	I have four children
I have went to Mexico 4 times this year	I recently got a cat
I hate Mexican food	I enjoy walking for exercise
I like to eat cheetos	I love watching Game of Thrones

Example of some personas

DATA COLLECTION PIPELINE

- Raw Persona collection:
  - 1155 possible personas,(at least 5 profile sentences each)
  - Test and val set of 100 personas each
- Revised personas:
  - additional rewritten sets of above personas
  - Ensures no non-stopword is copied
  - rephrases, generalizations or specializations
- Persona chat
  - pair two Turkers with random personas from pool.
  - Turkers were asked to chat
  - Ensures novelty of dialogue
  - Result: 164,356 utterances over 10,981 dialogs
  - Approx. 10% val and test set.

Original Persona	Revised Persona
I love the beach.	To me, there is nothing like a day at the seashore.
My dad has a car dealership	My father sales vehicles for a living.
I just got my nails done	I love to pamper myself on a regular basis.
I am on a diet now	I need to lose weight.
Horses are my favorite animal.	I am into equestrian sports.
I play a lot of fantasy videogames.	RPGs are my favorite genre.
I have a computer science degree.	I also went to school to work with technology.
My mother is a medical doctor	The woman who gave birth to me is a physician.
I am very shy.	I am not a social person.
I like to build model spaceships.	I enjoy working with my hands.

Persona 1	Persona 2
I like to ski	I am an artist
My wife does not like me anymore	I have four children
I have went to Mexico 4 times this year	I recently got a cat
I hate Mexican food	I enjoy walking for exercise
I like to eat cheetos	I love watching Game of Thrones

[PERSON 1:] Hi
[PERSON 2:] Hello ! How are you today ?
[PERSON 1:] I am good thank you , how are you.
[PERSON 2:] Great, thanks ! My children and I were just about to watch Game of Thrones.
[PERSON 1:] Nice ! How old are your children?
[PERSON 2:] I have four that range in age from 10 to 21. You?
[PERSON 1:] I do not have children at the moment.
[PERSON 2:] That just means you get to keep all the popcorn for yourself.
[PERSON 1:] And Cheetos at the moment!
[PERSON 2:] Good choice. Do you watch Game of Thrones?
[PERSON 1:] No, I do not have much time for TV.
[PERSON 2:] I usually spend my time painting: but, I love the show.

TASKS

next utterance prediction during dialogue.

profile prediction given dialogue history.

BASELINE EVALUATION FRAMEWORK

- Automatic evaluation metrics
  - Perplexity
  - F1 score
  - next utterance classification loss
- Human Evaluation metrics:
  - Fluency
  - Engagingness
  - consistency

MODELS-RESULTS

Method	No Persona		Original Persona		Revised Persona	
	ppl	hits@1	ppl	hits@1	ppl	hits@1
Generative Models						
Seq2Seq	38.08	0.092	40.53	0.084	40.65	0.082
Profile Memory	38.08	0.092	34.54	0.125	38.21	0.108
Ranking Models						
IR baseline	-	0.214	-	0.410	-	0.207
Starspace	-	0.318	-	0.491	-	0.322
Profile Memory	-	0.318	-	0.509	-	0.354
KV Profile Memory	-	0.349	-	0.511	-	0.351

Table 3: **Evaluation of dialog utterance prediction with various models** in three settings: without conditioning on a persona, conditioned on the speakers given persona ("Original Persona"), or a revised persona that does not have word overlap.

Method					Persona
Model	Profile	**Fluency**	**Engagingness**	**Consistency**	**Detection**
Human	Self	4.31(1.07)	4.25(1.06)	4.36(0.92)	0.95(0.22)
Generative PersonaChat Models					
Seq2Seq	None	3.17(1.10)	3.18(1.41)	2.98(1.45)	0.51(0.50)
Profile Memory	Self	3.08(1.40)	3.13(1.39)	3.14(1.26)	0.72(0.45)
Ranking PersonaChat Models					
KV Memory	None	3.81(1.14)	3.88(0.98)	3.36(1.37)	0.59(0.49)
KV Profile Memory	Self	3.97(0.94)	3.50(1.17)	3.44(1.30)	0.81(0.39)
Twitter LM	None	3.21(1.54)	1.75(1.04)	1.95(1.22)	0.57(0.50)
OpenSubtitles 2018 LM	None	2.85(1.46)	2.13(1.07)	2.15(1.08)	0.35(0.48)
OpenSubtitles 2009 LM	None	2.25(1.37)	2.12(1.33)	1.96(1.22)	0.38(0.49)
OpenSubtitles 2009 KV Memory	None	2.14(1.20)	2.22(1.22)	2.06(1.29)	0.42(0.49)

Table 4: **Human Evaluation** of various PERSONA-CHAT models, along with a comparison to human performance, and Twitter and OpenSubtitles based models (last 4 rows), standard deviation in parenthesis.

Persona	Method	Original			Revised		
		ppl	hits@1	F1	ppl	hits@1	F1
No Persona		38.08	0.092	0.168	38.08	0.092	0.168
Self Persona	Seq2Seq	40.53	0.084	**0.172**	40.65	0.082	**0.171**
	Profile Memory	**34.54**	**0.125**	**0.172**	38.21	**0.108**	0.170
Their Persona	Seq2Seq	41.48	0.075	0.168	41.95	0.074	0.168
	Profile Memory	36.42	0.105	0.167	**37.75**	0.103	0.167
Both Personas	Seq2Seq	40.14	0.084	0.169	40.53	0.082	0.166
	Profile Memory	35.27	0.115	0.171	38.48	0.106	0.168

Table 5: **Evaluation of dialog utterance prediction with generative models** in four settings: conditioned on the speakers persona ("self persona"), the dialogue partner's persona ("their persona"), both or none. The personas are either the original source given to Turkers to condition the dialogue, or the revised personas that do not have word overlap. In the "no persona" setting, the models are equivalent, so we only report once.

SUMMARY: PERSONA-CHAT

CONV-AI2

Dinan Et. Al.

Team Names	Perplexity	Hits@1	F1
1. Hugging Face	16.28	80.7	19.5
2. ADAPT Centre	31.4	-	18.39
3. Happy Minions	29.01	-	16.01
4. High Five	-	65.9	-
5. Mohd Shadab Alam	29.94	13.8	16.91
6. Lost in Conversation	-	17.1	17.77
7. Little Baby	-	64.8	-
8. Sweet Fish	-	45.7	-
9. 1st-contact	31.98	13.2	16.42
10. NEUROBOTICS	35.47	-	16.68
11. Cats'team	-	35.9	-
12. Sonic	33.46	-	16.67
13. Pinta	32.49	-	16.39
14. Khai Mai Alt	-	34.6	13.03
15. loopAI	-	25.6	-
16. Salty Fish	34.32	-	-
17. Team Pat	-	-	16.11
18. Tensorborne	38.24	12.0	15.94
19. Team Dialog 6	40.35	10.9	7.27
20. Roboy	-	-	15.83
21. IamNotAdele	66.47	-	13.09
22. flooders	-	-	15.47
23. Clova Xiaodong Gu	-	-	14.37
Seq2Seq + Attention Baseline	29.8	12.6	16.18
Language Model Baseline	46.0	-	15.02
KV Profile Memory Baseline	-	55.2	11.9

Team Names	Model Summary
Lost in Conversation	Generative Transformer based on OpenAI GPT. Trained on PERSONA-CHAT (original+revised), DailyDialog and Reddit comments.
Hugging Face	Pretrained generative Transformer (Billion Words + CoNLL 2012) with transfer to PERSONA-CHAT.
Little Baby	Profile-Encoded Multi-Turn Response Selection via Multi-Grained Deep Match Network. Modification of [9]: better model + data augmentation via translation.
Mohd Shadab Alam	Seq2Seq + Highway model. Glove + language model vector. Transfer learning strategy for Seq2Seq tasks.
ADAPT Centre	Bi-directional Attentive LSTM. Pretrained via GloVe embeddings + Switchboard, Open Subtitles.

For the full competition : https://arxiv.org/pdf/1902.00098.pdf

TRANSFERTRANSFO

TransferTransfo: A Transfer Learning Approach for Neural Network Based Conversational Agents

Thomas Wolf, Victor Sanh, Julien Chaumond & Clement Delangue
HuggingFace Inc.
81 Prospect St.
Brooklyn, New York 11201, USA
{thomas,victor,julien,clement}@huggingface.co

Wolf et. al.

MODEL

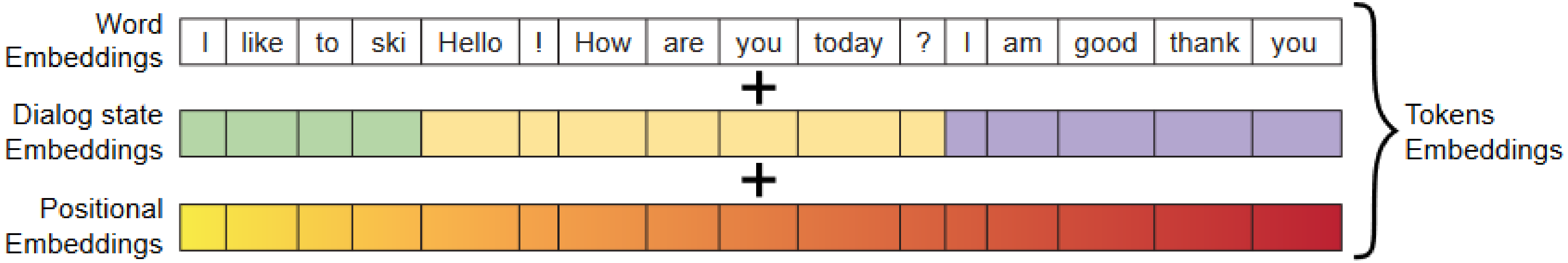


Figure 1: TranferTransfo's input representation. Each token embedding is the sum of a word embedding, a dialog state embedding and a positional embedding.

Dialog state embeddings:

- a personality sentence
- an utterance from PERSON1
- an utterance from PERSON2.

Positional Embeddings

- Introduces same PE to all personality embeddings

Model details:

- Used a 12-layer decoder-only transformer with masked self-attention heads
- Based on GPT
- Pretrained on book corpus

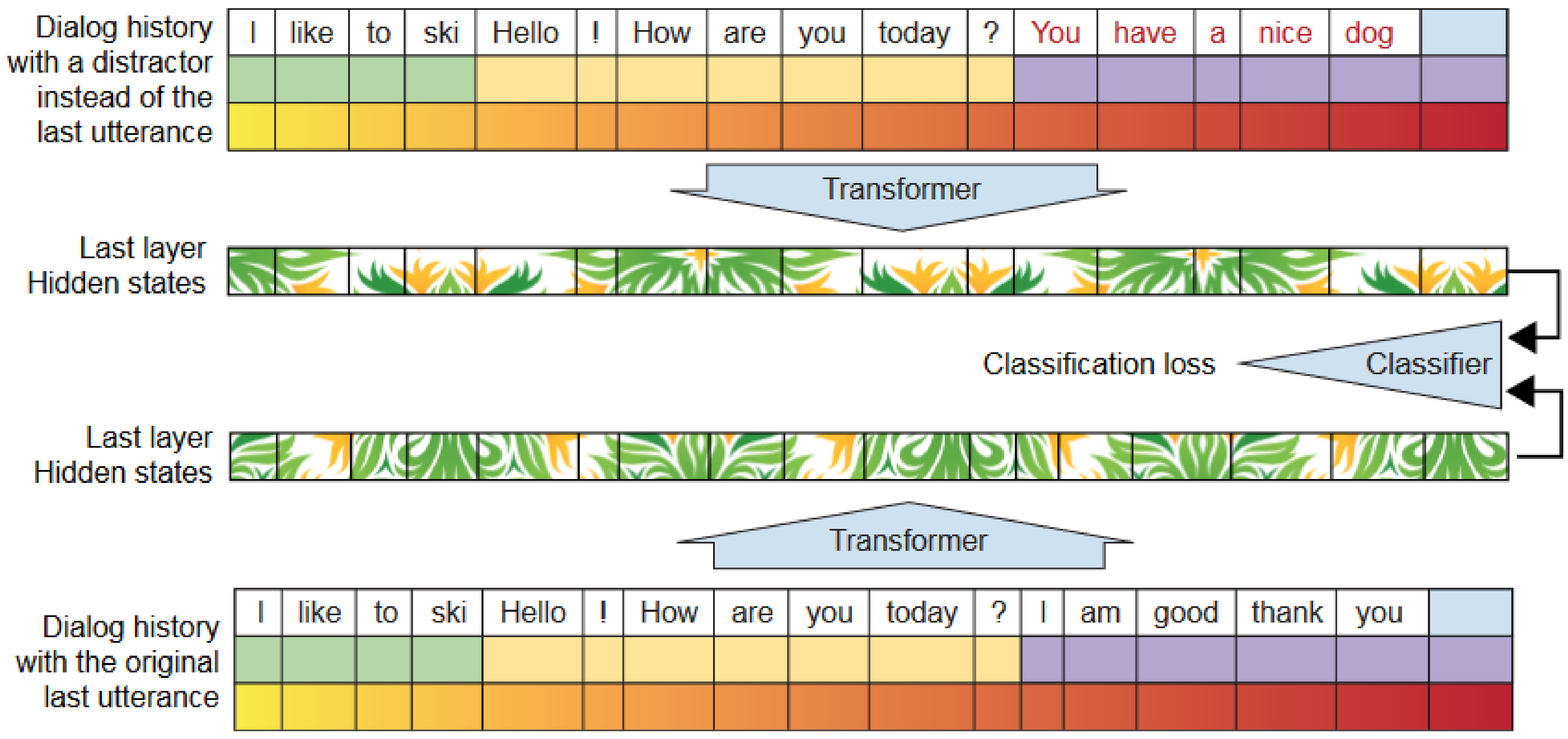


Figure 2: TranferTransfor input representation. The input embeddings is the sum of the word embeddings, the dialog state embeddings and the positional embeddings.

Fine-tuning

- Total loss function:
  - A next-utterance classification loss
  - A language modeling loss
- Adam optimizer

Generation

- Beam search ,B=4.
- N-grams filtering employed to ensure avoiding direct copying
- Remaining beams ranked by combination of the length-normalized utterance probability and the next-utterance classification score.

METRICS

- Automatic evaluation metrics
  - PPL
  - Hits@1
  - F1
- Human evaluations(Mentioned but not reported)
  - Fluency
  - Consistency
  - engagingness
  - whether the human could guess the persona used by the bot

RESULTS

	Eval			Test		
Model	PPL	Hits@1	F1	PPL	Hits@1	F1
Generative Profile Memory (Zhang et al., 2018)	*34.54*	12.5	–	–	–	–
Retrieval KV Profile Memory (Zhang et al., 2018)	–	*51.1*				
Seq2Seq + Attention (ConvAI2 baseline[3])	35.07	12.5	*16.82*	*29.8*	12.6	*16.18*
Language Model (ConvAI2 baseline[4])	51.1	–	15.31	46.0	–	15.02
KV Profile Memory (ConvAI2 baseline[5])	–	55.1	11.72	–	*55.2*	11.9
TransferTransfo (this work)	**17.51**	**82.1**	**19.09**	**16.28**	**80.7**	**19.5**

KEY POINTS:

- INPUT REPRESENTATION OF THE MODEL CAN SWITCH FROM A SINGLE TO A DYADIC SETTING PLUS PERSONALITY SENTENCES.
- USES TRANSFER LEARNING
- BETTER THAN THE RESULTS PRESENTED IN THE PERSONA-CHAT PAPER ACROSS ALL THREE METRICS.
- HUMAN EVAL WASN'T PUBLISHED.

BOB: BERT-OVER-BERT

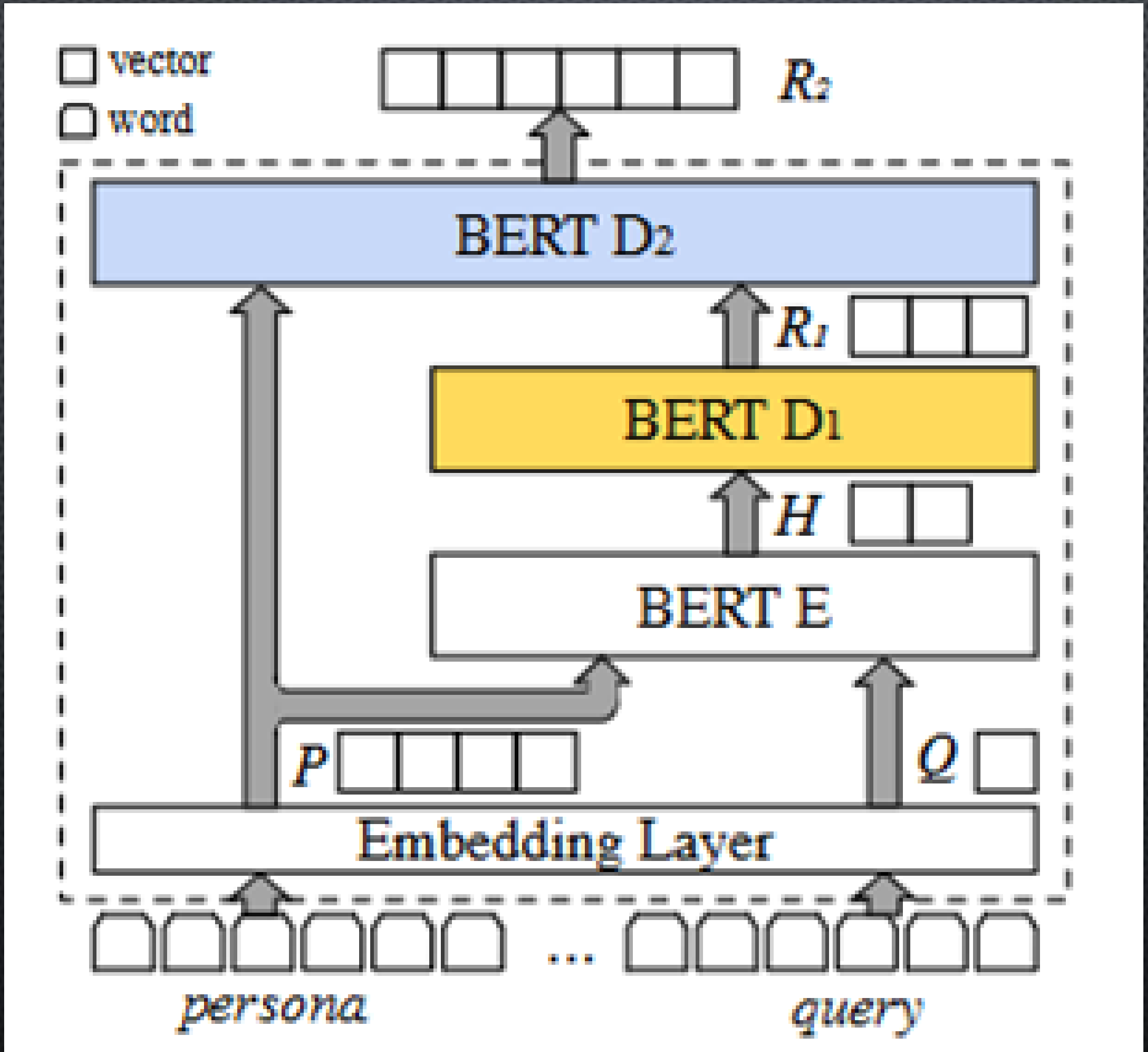
BoB: BERT Over BERT for Training Persona-based Dialogue Models from Limited Personalized Data

Haoyu Song[1], Yan Wang, Kaiyan Zhang[1], Wei-Nan Zhang[1,*], Ting Liu[1]
[1]Research Center for Social Computing and Information Retrieval
Harbin Institute of Technology, Heilongjiang, China
{hysong,kyzhang,wnzhang,tliu}@ir.hit.edu.cn
yanwang.branden@gmail.com

Song et. Al., ACL 2021

MODEL

- Disentangles persona-based dialogue generation into two tasks: Dialogue generation and consistency understanding
- Uses three BERT based submodules: Encoder, response decoder and consistency understanding decoder.


vector
word
R_2
BERT D2
R_1
BERT D1
H
BERT E
P
Q
Embedding Layer
...
persona
query

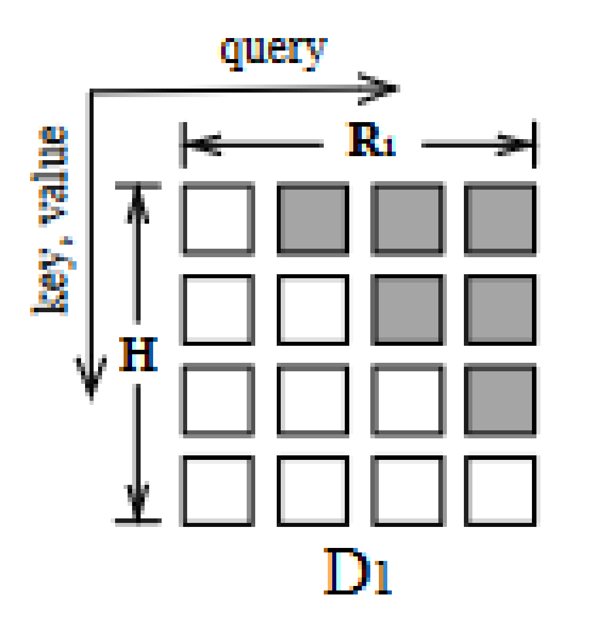
MODEL TASK

- THE TASK OF THE PROPOSED MODEL M IS TO GENERATE A PERSONA CONSISTENT RESPONSE R = R1, R2, ..., RM, BASED ON BOTH PERSONA P AND QUERY Q, I.E., R = M(Q, P).

ENCODER

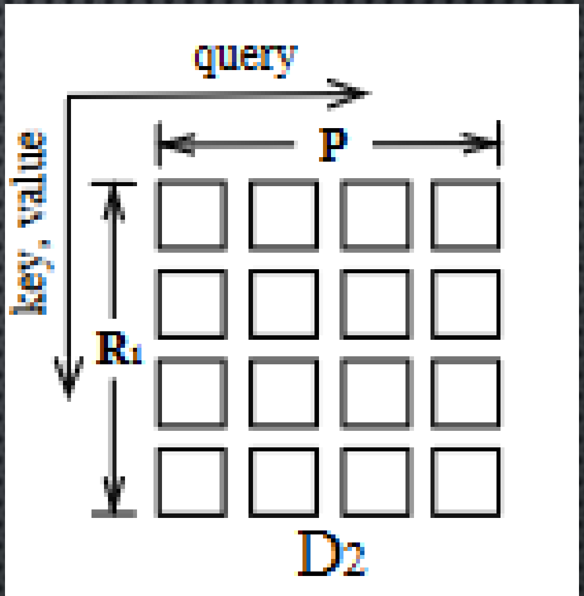
- A special token is placed between persona sequence and dialogue query.
- The input embedding consists of:
  - Word embedding
  - type Embedding(0 or 1)
  - Position embeddings
- Bert based encoder

DECODER 1



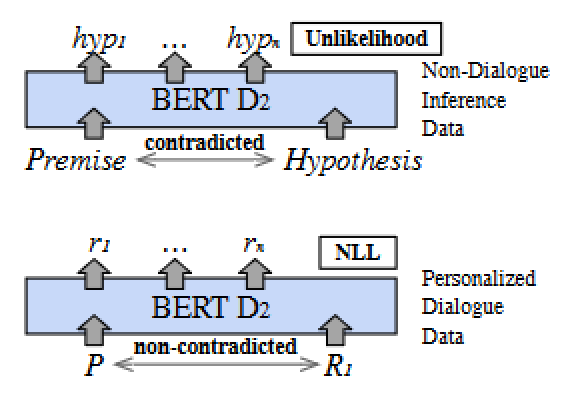
- BASED ON BERT WEIGHTS
- CROSS-ATTENTION IS INSERTED BETWEEN E AND D1.(RANDOMLY INITAILISED)
- ATTENTION MASK USED TO PRESERVE GENERATION PROPERTY

DECODER 2



- Based on bert weights
- cross-attention is inserted between E and D2.(randomly initailised)
- Attention is applied twice at each layer

$$p^{i+1} = \mathrm{FNN}(\mathrm{MultiHead}(r_2^i, P, P)),$$
$$r_2^{i+1} = \mathrm{FNN}(\mathrm{MultiHead}(p^{i+1}, R_1, R_1)).$$



TRAINING

- Loss function:
  - Response generation loss $\mathcal{L}_1 = \mathcal{L}_{NLL}^{\mathbb{D}_1} + \alpha\mathcal{L}_{NLL}^{\mathbb{D}_2}$
  - Consistency understanding loss $\mathcal{L}_2 = \beta\mathcal{L}_{UL}^{\mathbb{D}_2^+} + (1-\beta)\mathcal{L}_{UL}^{\mathbb{D}_2^-}$

RESULTS

	PPL	Dist.1	Dist.2	D.AVG	p.Ent	p.Ctd	ΔP	C.Score	Flue.	Info.	Relv.	Per.C.
Transformer	28.8	3.14	17.80	10.47	31.5	35.5	4.0	1.20	3.05	2.57	2.72	0.05
CMAML	36.7	1.00	2.10	1.55	32.3	37.5	5.2	6.96	3.36	2.40	3.09	0.24
GDR	16.7	3.76	23.10	13.43	19.7	32.3	12.6	7.89	3.38	2.74	3.13	0.21
LIC	17.3	6.29	28.99	17.64	13.7	20.4	6.7	14.12	3.70	3.53	3.47	0.39
GPT2	14.4	7.29	28.12	17.71	12.0	20.2	8.2	15.88	3.79	3.22	3.79	0.47
BoB (Ours)	**7.8**	**8.40**	**36.08**	**22.24**	**7.3**	**83.4**	**76.1**	**17.18**	**4.12**	**4.03**	**4.09**	**0.60**

Table 3: Automatic and human evaluation results on the full PersonaChat dataset. The best results are in bold.

	PPL	Dist.1	Dist.2	D.AVG	p.Ent	p.Ctd	ΔP	C.Score	Flue.	Info.	Relv.	Per.C.
Baselines' Best	14.4	7.29	28.99	17.71	12.0	37.5	12.6	15.88	3.79	3.53	3.79	0.47
Ours 1/8 Data	11.6†	7.49†	27.10	17.30	11.3†	83.6†	72.3†	15.87	4.17†	3.48	4.12†	0.62†
Ours 1/4 Data	9.7	7.97	30.20†	19.09†	11.8	85.8	74.0	16.04†	4.19	3.47	4.17	0.60
Ours 1/2 Data	8.9	8.13	33.08	20.61	8.1	81.9	73.8	16.36	4.03	3.70†	3.94	0.61

Table 4: Automatic and human evaluation results of the low resource settings on the PersonaChat dataset. The † means the minimum amount of data our model needed to outperform baselines' best results.

REVIEWS: PROS

- SESHANK : THE INTRODUCTION OF PROFILE INFORMATION TO THE MODELS IMPROVES THE PERFORMANCE COMPARED TO THE BASELINE MODEL BY A LARGE MARGIN. THIS SHOWS THAT ONE OF THE UNIQUE FEATURES OF THIS PAPER, PROFILE INFORMATION, TURNS OUT TO BE A VAST IMPROVEMENT OVER THE BASELINES. ESPECIALLY IN THE 'PERSONA DETECTION' METRIC, THE ADDITION OF PROFILE MEMORY DOES WONDERS.
- ROHIT: IN TABLE 6 WE SEE TRAINING ON REVISED PERSONA GIVES A SIGNIFICANT IMPROVEMENT IN PERFORMANCE. THIS INDICATES THAT MAKING THE TASK DIFFICULT GIVES A LOT OF ADVANTAGE TO THE LEARNING SYSTEM BECAUSE MODEL IS FORCED TO LEARN MORE THAN COMPUTING SIMPLE WORD OVERLAP AND ALSO GENERALIZES THE MODEL.
- ROHIT: BETTER PREDICTION PERFORMANCE ON LONGER DIALOGUE LENGTHS INDICATES THAT THE MODEL IS ABLE TO LEARN LONG RANGE DEPENDENCIES LEARNS BETTER WITH MORE PERSONAL INFORMATION

Reviews: Pros

- Shivangi: Authors defined a persona prediction task using the dialogs exchanged between the user and the agent. This task can have a practical application in the recommendation system. If an agent can accurately determine the profile and personality of the user, then providing personal recommendations would be really simple. It can provide an effective solution to the cold-start problem.
- Jai : The authors add checks in the crowd sourcing of the dataset so that its quality is not compromised. One example is sending error messages when the person tries to directly copy the description given to them

REVIEWS: CONS

- Rohit : This is an older paper and uses even older 2014 trained Glove Embeddings We see a significant drop in performance around 30-50% when we move from original to rewritten test set. This indicates IR system is not working very well. We could try out better embeddings which may give better results
- Vishal B.: The suggested method does not scale well as the number of personalities increases.
- Shivangi: We can observe that the Generative profile memory network performs poor than the seq2seq model. The explanation for this behavior is not mentioned in the paper.

REVIEWS: CONS

- Jai :Although Rohit said that revised persona gave better results in Table 6, revised persona does not do that in results in Table 3. The paper could have elaborated more on why we are seeing such a difference.

REVIEWS: FUTURE WORK

- SHESHANK: GIVEN A DIALOGUE HISTORY, FINDING A WAY TO PREDICT THE BOT'S PERSONA IS AN EXCITING FUTURE RESEARCH AREA. THE CURRENT PAPER DOES NOT HAVE EXCELLENT RESULTS FOR THIS.
- ROHIT : RANKING MODELS GET BETTER RESULTS BECAUSE THIS IS METRIC WHICH IS OPTIMIZED. CAN TRY SOME DIFFERENT APPROACH, TO GET BETTER RESULTS WITH GENERATIVE MODELS AS WELL.
- VISHAL B.: LOOKING AT THIS PAPER FROM TODAY'S PERSPECTIVE, THE DATASET GENERATION CAN BE AUTOMATED MORE USING PRE-TRAINED LANGUAGE MODELS, LIKE THE APPROACH FOLLOWED IN ACCENTOR.
- SHIVANGI: SIMILAR MULTILINGUAL DATASETS CAN BE GENERATED TO DESIGN A MORE GENERIC CHIT-CHAT BOT.
- S. SHREYA : SUCH PERSONA-BASED INFORMATION CAN BE COLLECTED ACROSS DIFFERENT PLATFORMS (LIKE TWITTER ETC) AND INCORPORATED IN TASK-BASED DIALOGUE SYSTEMS TO PERSONALISE THEM (AS DISCUSSED IN THE PREV CLASS).

REVIEWS: FUTURE WORK

- VISHAL S. : CREATING COMPLEMENTARY PERSONAS BASED ON THE USER. FOR EXAMPLE, AFTER CONVERSING WITH USER FOR SOMETIME, MODEL CAN GUESS USER'S PERSONA AND CREATE PERSONA BEST SUITED FOR THE USER.
- JAI : SINCE THE PERSONAS HAVE BEEN CROWDSOURCED, IT CAN HAVE GENDER AND RACIAL BIAS IN IT. SOME WORK NEEDS TO ANALYZE THAT AND IMPROVE THE QUALITY OF THE DATASET.

REFERENCES

Li et. Al. A Persona-Based Neural Conversation Model.

Zhang Et. Al. Personalizing Dialogue Agents: I have a dog, do you have pets too?.

Dinan et. al. The Second Conversational Intelligence Challenge (ConvAI2).

Wolf et. Al. , TransferTransfo: A Transfer Learning Approach for Neural Network BasedConversational Agents

Song et. Al. BoB: BERT Over BERT for Training Persona-based Dialogue Models from Limited Personalized Data