Exemplar Encoder Decoder for Neural Conversation Generation

By

Gaurav Pandey, Danish Contractor, Vineet Kumar and Sachindra Joshi

IBM Research AI

Generative Models for Conversations

- Context encoder: *(1)* RNN *(2)* hierarchical RNN
- Decoder: RNN
- Objective: log probability of GT response given context.
- Can generate novel responses for novel contexts!!

Retrieval Models for Conversations

- Retrieve a response from a nearest neighbor index constructed from the training data.

- Can be used for closed domain problems.

- Advantages:
  - Answers are grounded in the domain.
  - Easy to prune answers according to requirements.

- Disadvantage:
  - Can not generate novel responses.

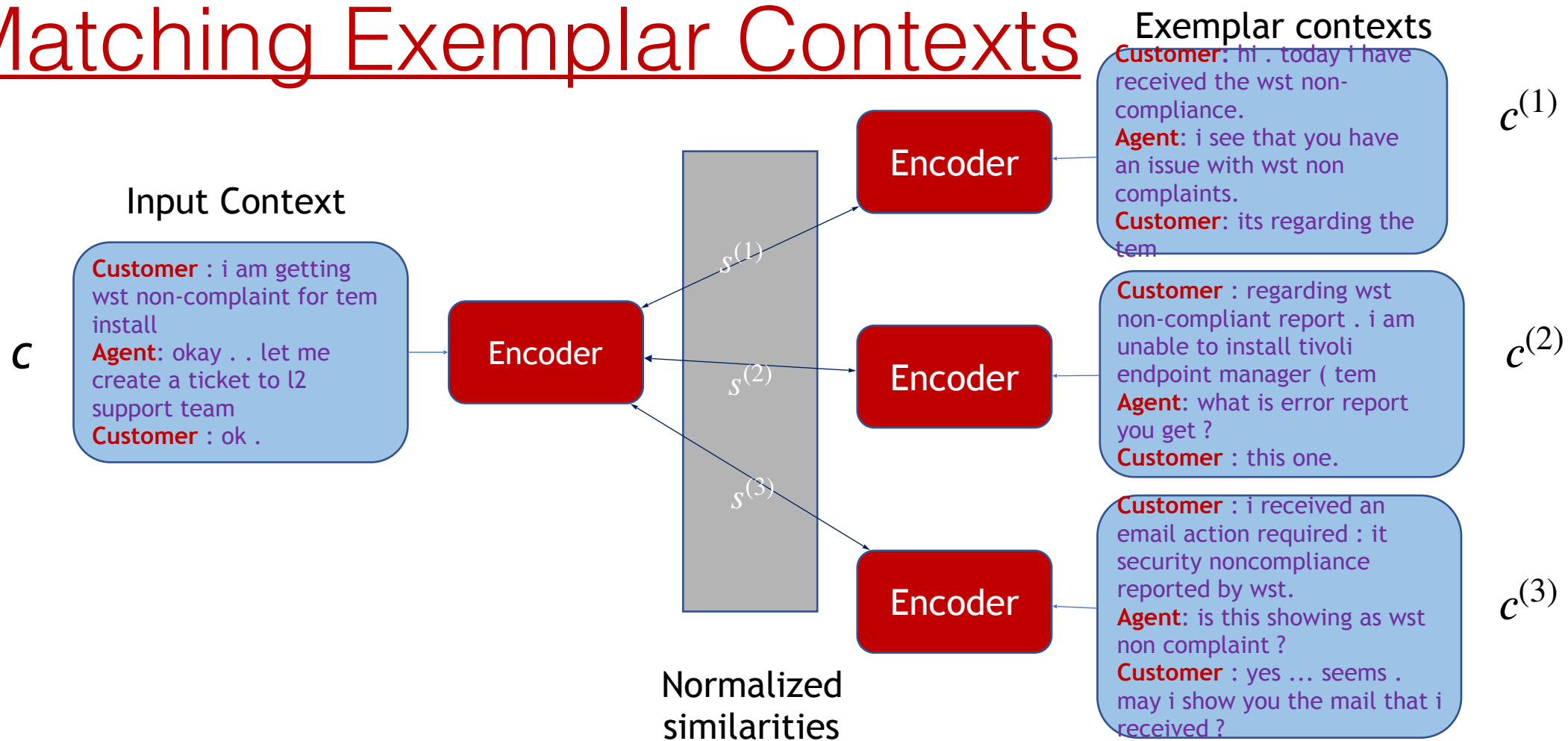Can we use generative models to fix this?

Exemplar Encoder Decoder

- Build an index from all context-response pairs offline.
- For each context c:
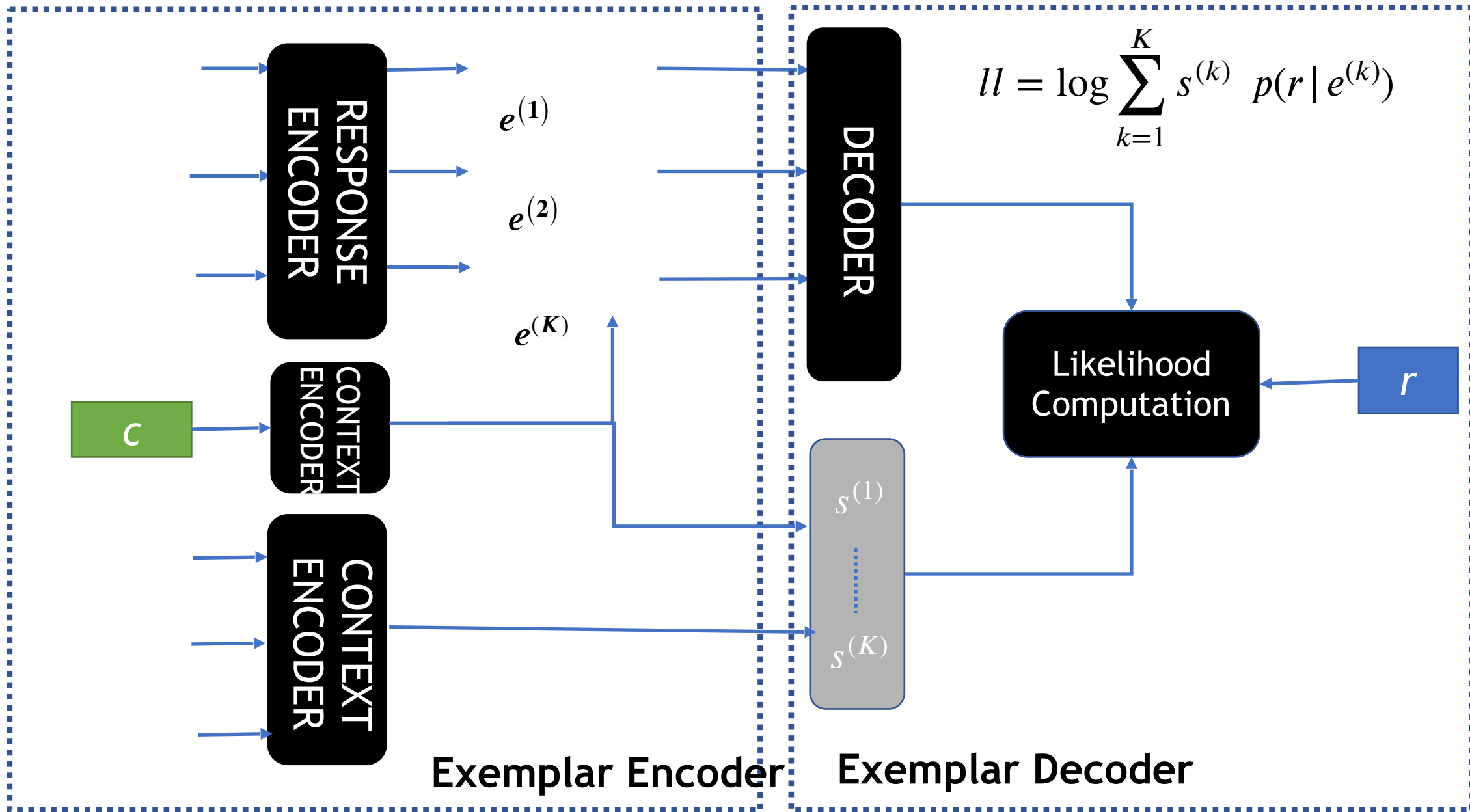 - Retrieve a set of *exemplar* contexts and corresponding responses.

| Input Context | Index | Exemplar conversations |

- Match the *exemplar* contexts with c and get the similarities.
- Use these similarities to weigh the *exemplar* responses.

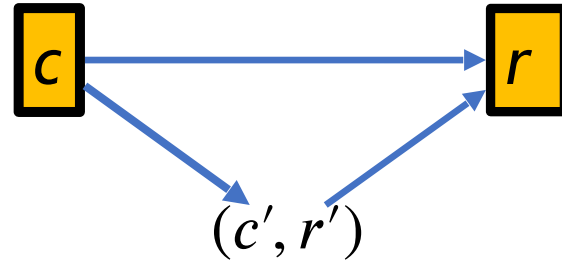# Matching Exemplar Contexts

Exemplar contexts

Input Context

c

Customer : i am getting wst non-complaint for tem install
Agent: okay . . let me create a ticket to l2 support team
Customer : ok .

Encoder

Normalized similarities

$s^{(1)}$

$s^{(2)}$

$s^{(3)}$

Encoder

Customer: hi . today i have received the wst non-compliance.
Agent: i see that you have an issue with wst non complaints.
Customer: its regarding the tem

$c^{(1)}$

Encoder

Customer : regarding wst non-compliant report . i am unable to install tivoli endpoint manager (tem
Agent: what is error report you get ?
Customer : this one.

$c^{(2)}$

Encoder

Customer : i received an email action required : it security noncompliance reported by wst.
Agent: is this showing as wst non complaint ?
Customer : yes ... seems . may i show you the mail that i received ?

$c^{(3)}$

The normalized similarities are used to weigh the exemplar responses.

Analyzing the Objective

Think of exemplar contexts and responses as latent variables

$$\log p(r\,|\,c) = \log \sum_{(c',r')} p(r\,|\,c, r')\ \ p(c'\,|\,c)$$

$$\leq \log \sum_{1 \leq k \leq K} p(r\,|\,c, r^k)\ \ p(c^k\,|\,c)$$

$$= \log \sum_{1 \leq k \leq K} p(r\,|\,e^{(k)})\ \ s^{(k)}$$

Evaluation

- Exemplar Encoder Decoder
  - Hierarchical Recurrent Encoder
  - TF-IDF for retrieving exemplar conversations
- Datasets used:
  - Ubuntu Dialogue Corpus
  - IBM Tech Support Dataset
- Comparison Metrics
  - Activity and Entity metrics
  - Embedding metrics

Activity and Entity metrics

These metrics compare the precision, recall and F1 score of specific nouns and ver present in the generated response as compared to the groundtruth response.

Ubuntu Dialogue Corpus

	Activity			Entity			Tense	Cmd
Model	P	R	F1	P	R	F1	Acc.	Acc.
LSTM*	1.7	1.03	1.18	1.18	0.81	0.87	14.57	94.79
VHRED*	**6.43**	4.31	4.63	3.28	2.41	2.53	20.2	92.02
HRED*	5.93	4.05	4.34	2.81	2.16	2.22	22.2	92.58
EED	6.42	**4.77**	**4.87**	**3.8**	**2.91**	**2.99**	**31.73**	**95.06**

For comparison, the retrieval only model has an activity F1 score of **4.23** and
entity F1 score of **2.72** respectively.

Embedding metrics

- These metrics compare the word embeddings of the generated response with the words of the groundtruth response.

- These metrics do not correlate with human judgements for Ubuntu Corpus[1].

	Ubuntu			Tech Support		
Model	Average	Greedy	Extrema	Average	Greedy	Extrema
HRED	0.5131	0.4103	**0.3263**	0.6372	0.5100	0.4355
EED	**0.5528**	**0.4172**	0.3134	**0.6420**	**0.5125**	**0.4356**

[1]How NOT To Evaluate Your Dialogue System: An Empirical Study of Unsupervised Evaluation Metrics for Dialogue Response Generation

Generated and retrieved responses

	Input Context (C) and exemplar response (R)	Generated/Ground Truth Responses
1	**C:** i want to check ntfs file systems. is fsck ok with ntfs. → no , you 'd need to use chkdsk from windows. → **R:** i already symlinked ntfsfix to fsck. **ntfs** and am scanning the unmounted devices	**HRED:** i know , but i dont know how to do that **VHRED:** how can i do that ? **EED:** i am trying to use **ntfs** 3g i have a windows partition. **GT:** i dont have windows installed
2	**C:** anyone help me ? → seems that its detecting grub , have you tried checking continue without installing grub ? **R:** try the live **cd**... you ca n't boot into linux unless **grub install**	**HRED:** yes i did , but i have no idea how to fix it **VHRED:** yes i did. i dont know how to do it **EED:** nope not yet. **grub install cd** is not working **GT:** no i did not.
3	**C:** any ideas on my mplayer problem ? → reinstall. apt-get purge mplayer then reinstall **R:** but half my system depends on this **package** ...	**HRED:** I did that , but it didn't work **VHRED:** I dont know how to do that **EED:** does n t work. **package** does n t work **GT:** that does n't seem as a solution

Discussion

- A generative model that utilizes similar conversations for response generation.

- Can generate novel responses while ensuring that the responses are grounded in the domain.

- Incorporating retrieved conversations during generation improves performance as evident from several metrics.

- The proposed idea is general and can be used for image captioning and neural machine translation.