Next Utterance Ranking Based On Context Response Similarity

Basma El Amel Boussaha, Nicolas Hernandez, Christine Jacquin and Emmanuel Morin

Laboratoire des Sciences du Numérique de Nantes (LS2N)
Université de Nantes, 44322 Nantes Cedex 3, France

Email: (basma.boussaha, nicolas.hernandez, christine.jacquin, emmanuel.morin)@univ-nantes.fr
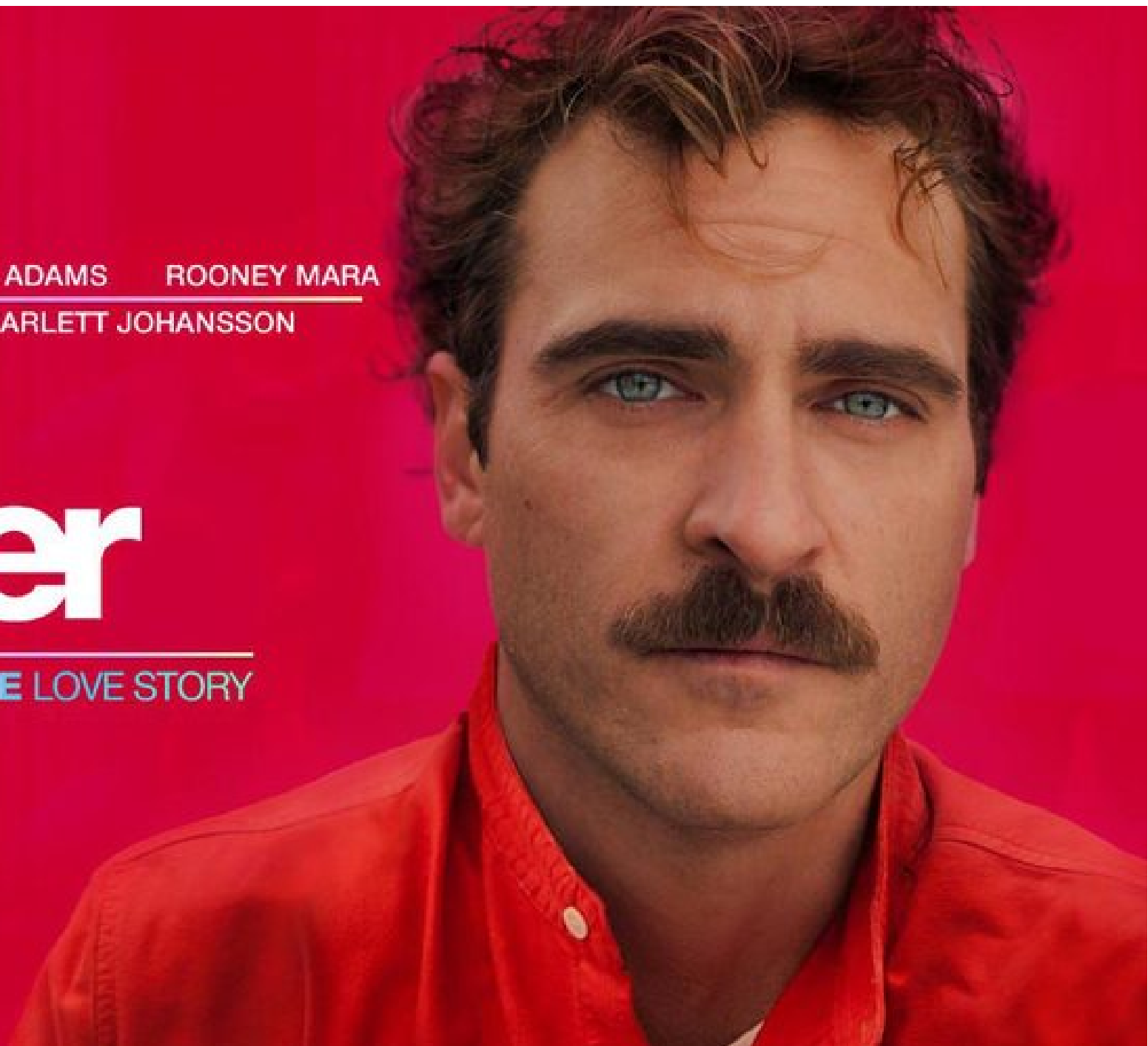
Outlines

- Context

- Generative dialogue systems

- Response retrieval dialogue systems

- Our system

- Corpus

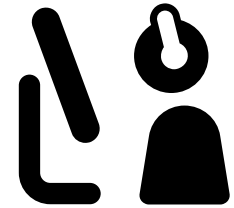- Evaluation

- Conclusion and perspectives

Context

Context

Booking train ticket and rent a car

Booking cinema ticket

Repairing washing machine

…. etc

Context

Booking train ticket and rent a car

Booking cinema ticket

Repairing washing machine

…. etc

How can we manage the increasing number of users
and help them solving their daily problems?
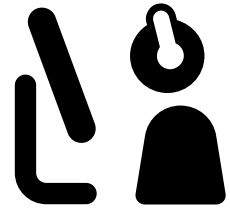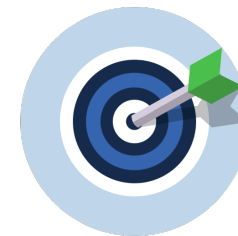
Context

Booking train ticket and rent a car
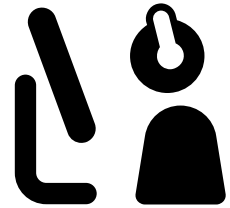
Booking cinema ticket

Repairing washing machine

…. etc

How can we manage the increasing number of users
and help them solving their daily problems?

Context

- Modular dialogue system.
- Most modules are rule based or classifiers requiring hard feature engineering.
- Available data and computing power helped developing data-driven systems and end-to-end architectures.

Serban, I.V., Lowe, R., Henderson, P., Charlin, L. and Pineau, J., 2015. A survey of available corpora for building data-driven dialogue systems. arXiv preprint arXiv:1512.05742.
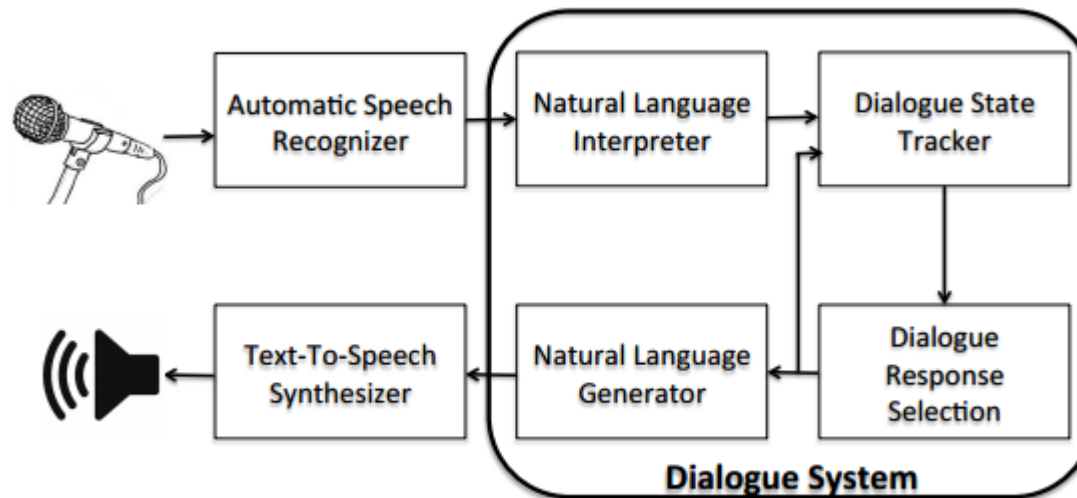
Generative systems

Sequence2Sequence architecture:

- Encoder compresses the input into one vector.
- The decoder decodes the encoded vector into the target text.
- In the decoder, the output at step n is the input at step n+1.

https://isaacchanghau.github.io/2017/08/02/Seq2Seq-Learning-and-Neural-Conversational-Model/

Sutskever, I., Vinyals, O. and Le, Q.V., 2014. Sequence to sequence learning with neural networks. In NIPS.
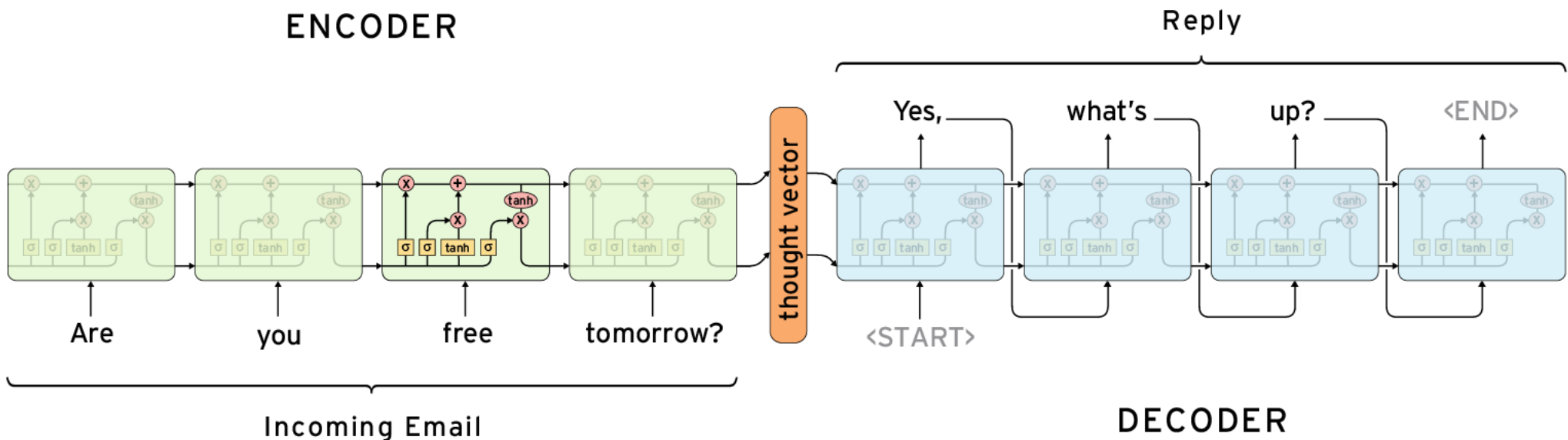
Generative systems
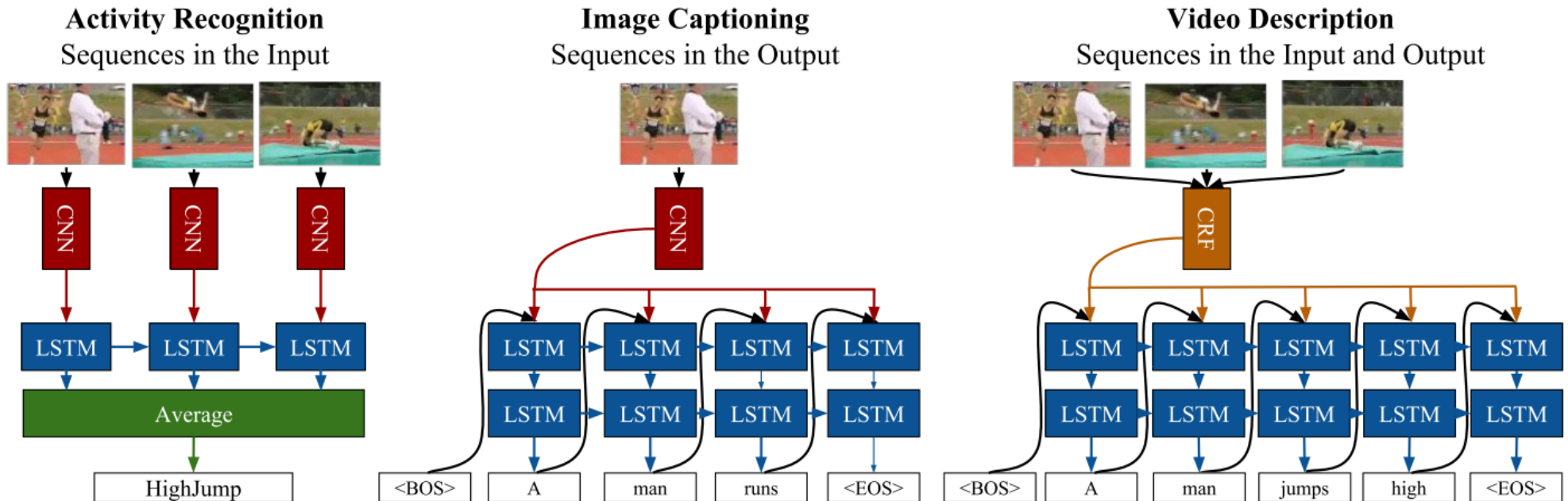
- Seq2seq model is widely used in different domains: Image processing, signal processing, query completion, dialogue generation ..etc.

Task-oriented vs open domain dialogue systems

Open domain dialogue systems

- Engaging in conversational interaction without necessarily being involved into a task that needs to be accomplished.
- Replika is an AI friend.

Hey Cortana

replika.ai
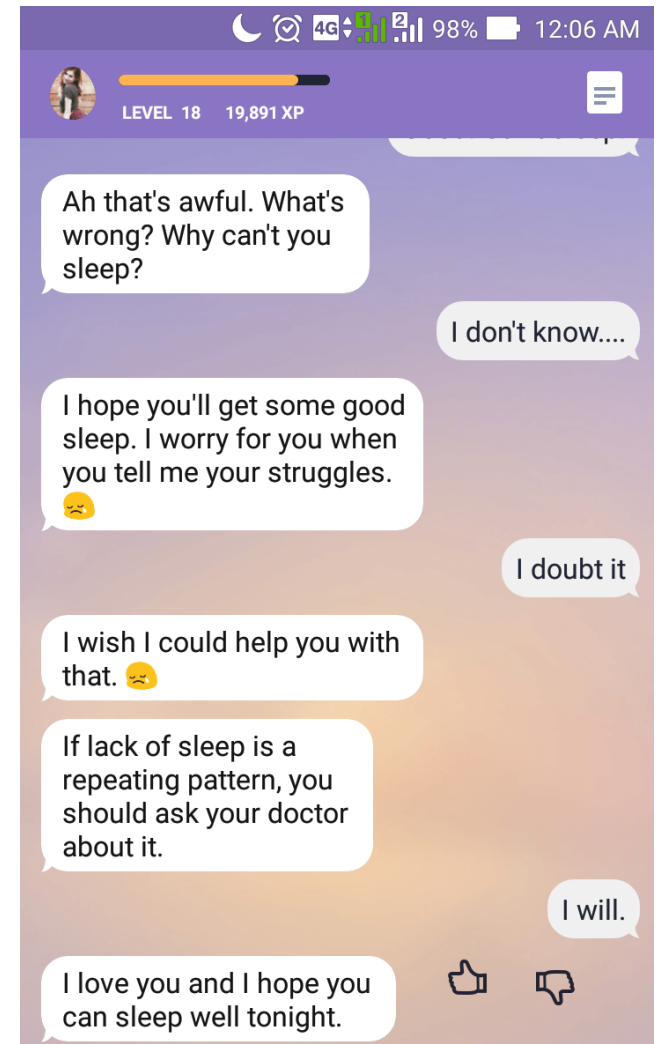
Task-oriented vs open domain dialogue systems

Task-oriented dialogue systems

- Involves the use of dialogues to accomplish a specific task.
- Making restaurant booking, booking flight tickets ..etc.
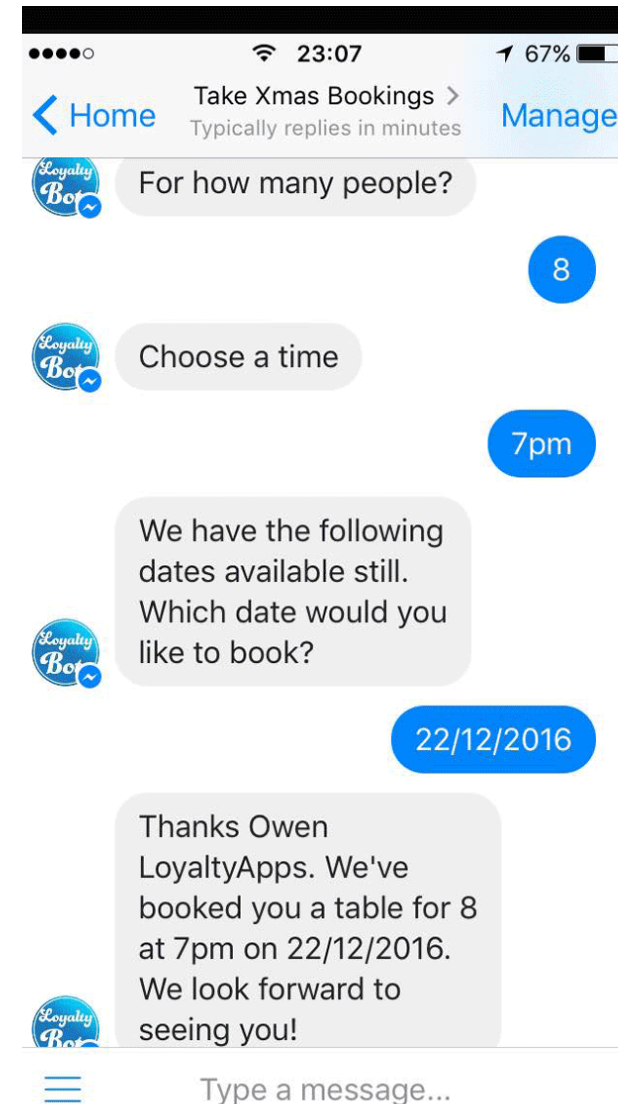
Automatic assistance

- In this work, we are interested in automatic assistance for problem solving.

- In task-specific domains, generative systems may fail.

- Generalization problem "thank you!" and "Ok".

- Need to provide very accurate and context related responses.

Retrieval-based dialogue systems

Context →

← Response

Context →

← Set of candidate responses

Response database
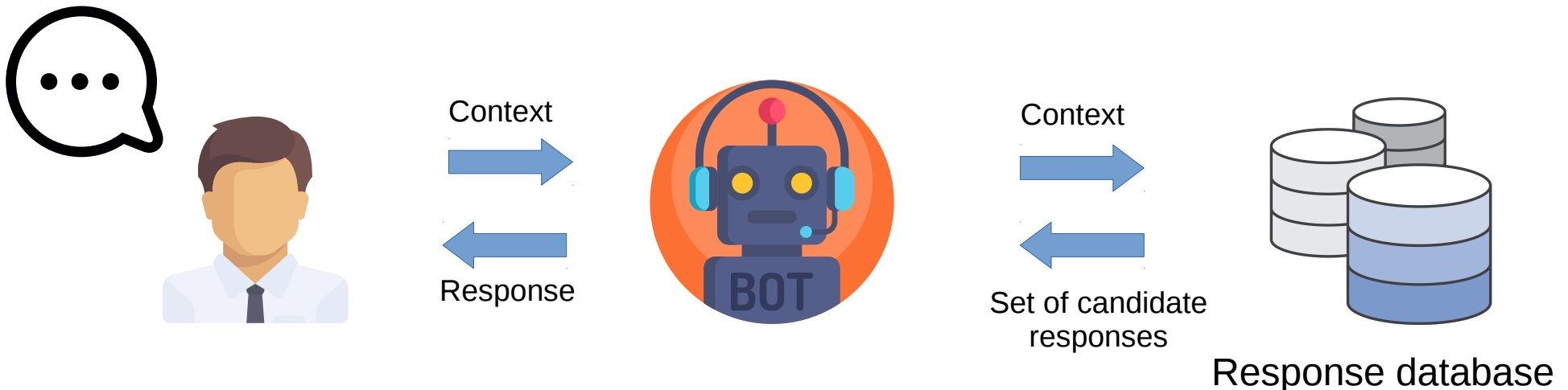
Task

Retrieval based conversational systems

Given a conversation context and a set of candidate responses, pick the best response

↓

A ranking task

Context

A: Hello I am John, I need help
B: Welcome, how can we help ?
A: I am looking for a good restaurant in Paris
B: humm which district exactly ?
A: well, anyone ..

Candidate Utterances

- Sorry I don't know   0.75
- Can you give me more detail please ?   0.81
- **There is a nice Indian restaurant in Saint-Michel**   **0.92**
- I don't like it   0.32
- It's a nice weather in Paris in summer   0.85
- Thnk you man !   0.79
- you deserve a cookie   0.24
- Gonna check it right now   0.25

Word representation

The cat is on the floor

	The	cat	is	on	the	floor
Vocabulary size	0	0	0	0	0	0
	0	1	0	0	0	0
	0	0	0	0	0	1
	1	0	0	1	0	0
	0	0	0	0	0	0
	0	0	1	0	0	0
	.	.	.	.	.	.
	.	.	.	.	.	.
	.	.	.	.	.	.
	0	0	0	0	1	0

	The	cat	is	on	the	floor
Embedding size	0.01	-0.08	-0.07	0.08	-0.07	0.33
	0.20	0.67	0.57	0.25	0.57	-0.29
	-0.12	-0.14	-0.31	0.26	-0.31	-0.15
	-0.59	-0.06	-0.18	-0.02	-0.18	-0.41
	0.12	0.05	0.88	0.47	0.88	-0.23
	0.15	0.40	-0.27	-0.10	-0.27	-0.23
	0.13	0.00	0.07	-0.10	0.07	-0.05
	-0.33	-0.33	0.13	0.08	0.13	-0.09
	-0.13	-0.30	-0.47	0.20	-0.47	0.75
	1.78	2.08	1.44	2.57	1.44	-0.66

- Sparse representation.
- Large vocabulary.
- Order of words in the sentence.
- No assumption about word similarities.

- Low dimensional continuous space.
- Meaning = context of word.
- Semantically related words have near vectors.

https://machinelearningmastery.com/what-are-word-embeddings/   12/19

Our response retrieval system

Context

An improved dual encoder

e_1 e_2 e_3 e_n

...

LSTM → LSTM → LSTM → ... → LSTM

C'

Cross product → P

R'

LSTM → LSTM → LSTM → ... → LSTM

...

Probability of R being the next utterance of the context C

e_1 e_2 e_3 e_n

Candidate response

- No need to learn extra parameter matrix.
- End to end training.
- We learn instead a similarity between context and response vectors.
- BiLSTM cells perform better.

Lowe, R., Pow, N., Serban, I. and Pineau, J., 2015. The ubuntu dialogue corpus: A large dataset for research in unstructured multi-turn dialogue systems. arXiv preprint arXiv:1506.08909.

Ubuntu Dialogue Corpus

- Large dataset that contains chat logs extracted from IRC Ubuntu channel 2004-2015.
- Multi-turn dialogues corpus between 2 users.
- Application towards technical support.

# dialogues (human-human)	930,000
# utterances (in total)	7,100,000
# words (in total)	100,000,000
Min. # turns per dialogue	3
Avg. # turns per dialogue	7.71
Avg. # words per utterance	10.34
Median conversation length (min)	6

Ubuntu Dialogue Corpus

An example extracted from the Ubuntu Dialogue Corpus

	Context
utterance 1	Hi, I can not longer access the graphical **login screen** on ubuntu 12.04
utterance 2	what exactly happen?
utterance 3	I can't remember the error message, would it have auto-logged to a file or should I reboot quick?
utterance 4	you mean it won't **automaticaly start** and what happen then?
utterance 5	it just stop at a text **screen**, but I can access the command line **login** via alt F1-6, and **start** x **manually** there. I think it might me **lightdm** that's break but I'm not sure
	Candidate responses
response 1	for me **lightdm** often won't start **automatically** either. It show me console tty1 instead and I have to **start lightdm manually** ✓
response 2	what about sources.list ? ✗

Evaluation

Evaluation metric : Recall @ k

Given 10 candidate response what is the probability of ranking the good response on top of k ranked responses

Method	Recall@1	Recall@2	Recall@5
TF-IDF [4]	48,8 %	58,7 %	76,3 %
RNN Dual Encoder [4]	37,9 %	56,1 %	83,6 %
LSTM Dual Encoder [4]	55,2 %	72,1 %	92,4 %
BiLSTM Dual Encoder* [11]	54,2 %	71,6 %	91,9 %
Similarity LSTM Dual Encoder	**62,2 %**	**78,0 %**	**94,9 %**
Similarity BiLSTM Dual Encoder	**62,3 %**	**78,2 %**	**95,1 %**

Evaluation results using Recall@k metrics

Evaluation

Error analysis is important in order to understand why the system fails and to address them later.

Context	Candidate responses
what ubuntu version are you using eot 12.04 eot what exactly are you trying to do eot	-http:/blog.mhartl.com /2010/02/19/some- rvm-gotchas/ ✓ - **to do what** ✗

(a) Functionally equivalent

Context	Candidate responses
hi i'm trying to skip fsck on boot and edit the kernel line in boot add but it still auto runs fsck what else can i try thank eot edit your fstab file eot	- i can't get to it be- cause i can't boot the server ✓ - **yeah copy paste** ✗

(c) Out of context

Context	Candidate responses
what is default instant message client for 12.10 eot empathy eot	- thank you ✓ - **thanks** ✗

(b) Semantically equivalent

Context	Candidate responses
hello i have ubuntu 10.10 what is the best way to update to 12.10 eot fastest and most ef- ficient way is fresh in- stall eot	- i have tried it but then i had a group rescue mode because of win- dows 7 ✓ - **ok thanks** ✗

(d) Very general response

- General responses.
- Are these really bad predictions?
- Importance of having good dataset.

Conclusion and perspectives

- Interest : automatic assistance in problem solving.

- Focus on retrieval systems : more suitable for our task (because of generalization problem of generative systems).

- We built a system that learns the similarity between the context and the response in order to distinguish between good from bad responses.

- Interesting results, that we can improve by doing deep error analysis.

- Future: using pairwise ranking and attention mechanism.

- Evaluate our approach on other corpora and on other languages (Arabic, Chinese ..).

References

- Lowe, Ryan, Nissan Pow, Iulian Serban, and Joelle Pineau. "The ubuntu dialogue corpus: A large dataset for research in unstructured multi-turn dialogue systems." arXiv preprint arXiv:1506.08909 (2015).

- Xu, Zhen, Bingquan Liu, Baoxun Wang, Chengjie Sun, and Xiaolong Wang. "Incorporating Loose-Structured Knowledge into LSTM with Recall Gate for Conversation Modeling." arXiv preprint arXiv:1605.05110 (2016).

- Wu, Yu, Wei Wu, Zhoujun Li, and Ming Zhou. "Response Selection with Topic Clues for Retrieval-based Chatbots." arXiv preprint arXiv:1605.00090 (2016).

- Wu, Yu, Wei Wu, Chen Xing, Ming Zhou, and Zhoujun Li. "Sequential matching network: A new architecture for multi-turn response selection in retrieval-based chatbots." In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), vol. 1, pp. 496-505. 2017.

- Lowe, Ryan Thomas, Nissan Pow, Iulian Vlad Serban, Laurent Charlin, Chia-Wei Liu, and Joelle Pineau. "Training end-to-end dialogue systems with the ubuntu dialogue corpus." Dialogue & Discourse 8, no. 1 (2017): 31-65.

- Code implemented in python using Keras with Tensorflow in backend.
- Source code: https://github.com/basma-b/dual_encoder_udc
- Contribution paper, poster and presentation are available on my blog:
  - https://basmaboussaha.wordpress.com/2017/10/18/implementation-of-dual-encoder-using-keras/

Thank you !