

Next Utterance Ranking Based On Context Response Similarity

Basma El Amel Boussaha · Nicolas Hernandez · Christine Jacquin · Emmanuel Morin

Laboratoire des Sciences du Numérique de Nantes (LS2N)

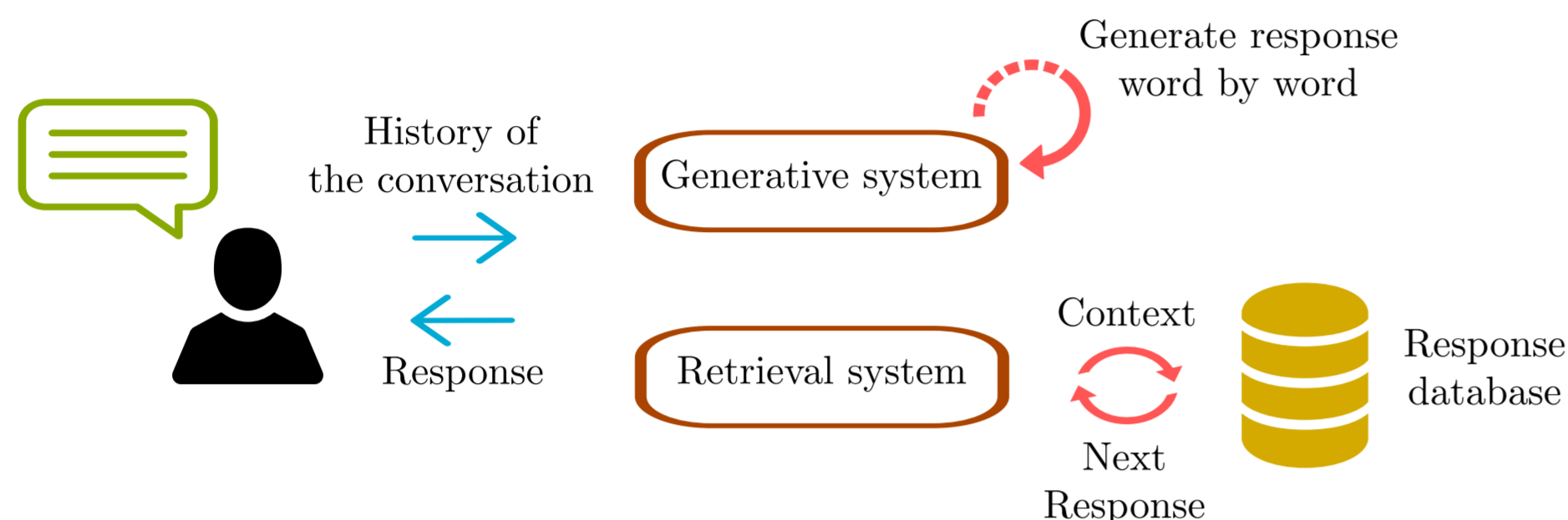
Université de Nantes, 44322 Nantes Cedex 3, France

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



Context

- Increasing need to develop machines able to converse with humans and help them performing their daily tasks.
- Two categories of dialogue systems: generative and retrieval systems.
- *Task specific dialogue systems*: restaurant recommendation, hotel booking, etc. vs *chit-chat systems*: Cortana, Alexa, Replika, etc.
- Available data and computing power helped building data-driven dialogue systems.



Problem

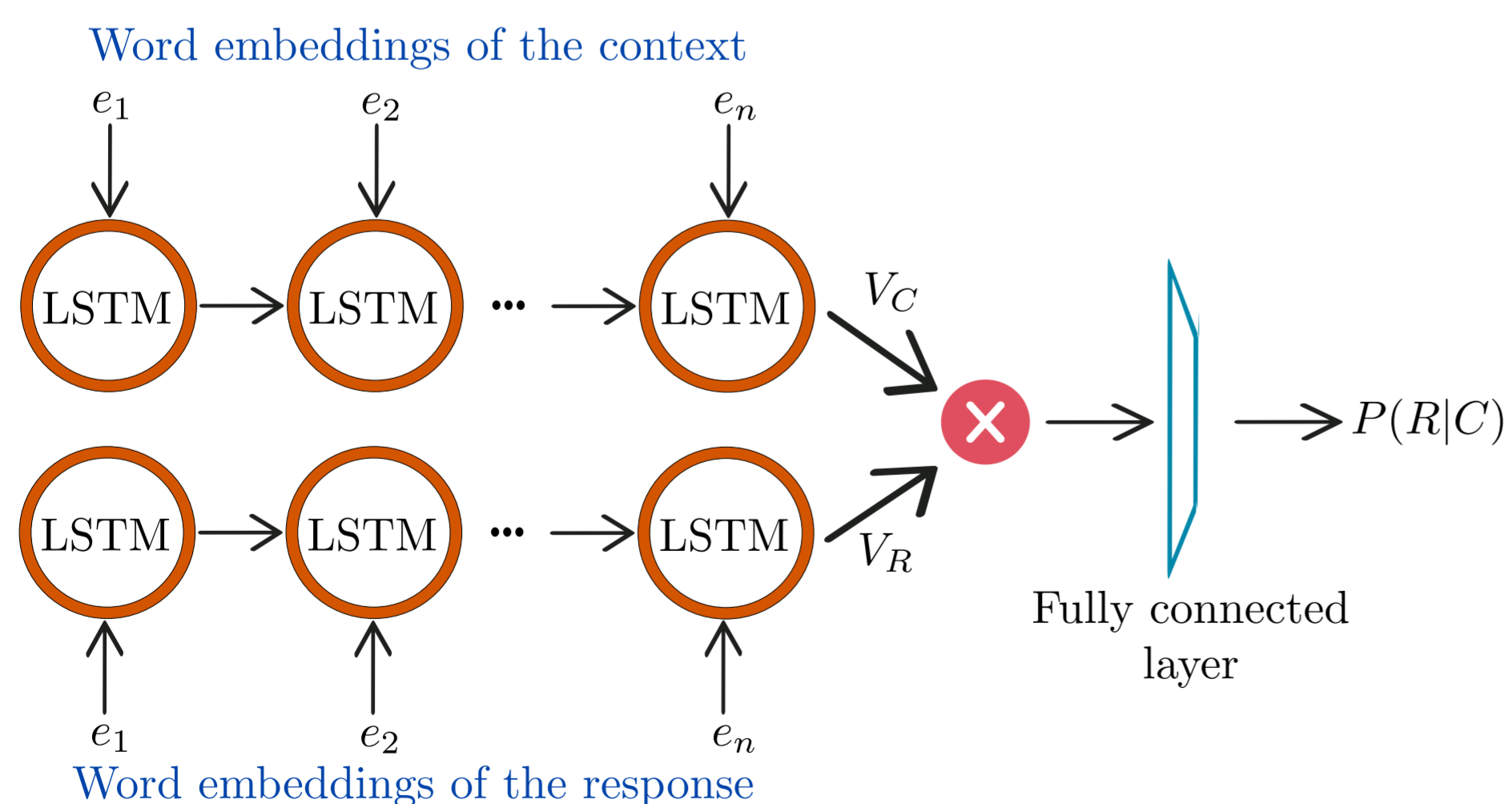
We address the problem of retrieving the next utterance of a given conversation's context and a set of candidate responses. The problem consists of:

- Understanding the context of the conversation.
- Identifying important information (words, phrases and sentences).
- Matching these information between the context and the candidate response.
- Modeling the relationship between the context and the candidate response.

Context
u1 Hi, I can not longer access the graphical login screen on ubuntu 12.04
u2 what exactly happen?
u3 I can't remember the error message, would it have auto-logged to a file or should I reboot quick?
u4 you mean it won't automatically start and what happen then?
u5 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
r1 for me lightdm often won't start automatically either. It show me console tty1 instead and I have to start lightdm manually ✓
r2 what about sources.list ? ✗

Model

Based on the dual encoder of [1], we propose a new response ranking system trained in an end-to-end fashion with supervised learning.



1. Represent the input words using word embeddings (initialized with Glove).
2. Word embeddings are fed in chronological order into a shared LSTM neural network.
3. The hidden layer of the network is updated every time an embedding is fed.
4. A cross product is computed between the context and the response vectors and transformed into a probability using a fully connected layer and sigmoid.
5. This classification probability is used in order to rank the candidate responses.

Experiments and results

We evaluated our system on a large public dataset: the Ubuntu Dialogue Corpus.

Dataset

# utterances (total)	7,100,000
# turns (total)	5,139,574
# words (total)	100,000,000
Min. # turns per dialogue	3
Avg. # turns per dialogue	4.94
Avg. # word per turn	10.34
# train samples	1,000,000
# test samples	18,920
# validation samples	19,560

Evaluation results

- We followed [1, 2, 3] and evaluated our approach using Recall@k metrics.
- Our utterance ranking system outperforms the state of the art systems on all metrics with a good margin.

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 [5]	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 %

Error Analysis

In some cases our model was not able to retrieve the best utterance among the candidate responses. The reasons of this bad ranking may include:

- The nature of the corpus: the candidate responses were randomly sampled from the hole corpus, without any human evaluation.
- Our model fails to rank short and general responses.

Context	Candidate responses
u1 http://www.howtogeek.com how to add screensavers to ubuntu 12.04 see also http://askubuntu.com questions how can i change or install	0.99 it's only annoying when the cursor drag really slowly
u2 ok it won't become an issue on system upgrade	0.87 apt-get install hwinfo
u3 then you probably just need to log out back in to restart indicator messages	0.85 ok what is that ok just figure it out you just help me out haha
	0.27 thank you

Conclusion

- We proposed in this work a utterance ranking system based on dual encoder.
- Experimental results show that our approach brings significant improvements compared to the state of the art systems.
- Our new approach based on semantic and syntactic similarities between the context and response allows to better distinguish between good and bad responses.

Bibliography

- [1] R. Lowe, N. Pow, I. Serban, and J. Pineau, "The ubuntu dialogue corpus: A large dataset for research in unstructured multi-turn dialogue systems," in *SIGDIAL*, 2015.
- [2] Y. Wu, W. Wu, C. Xing, M. Zhou, and Z. Li, "Sequential matching network: A new architecture for multi-turn response selection in retrieval-based chatbots," in *ACL*, 2017.
- [3] Z. Xu, B. Liu, B. Wang, C. Sun, and X. Wang, "Incorporating loose-structured knowledge into conversation modeling via recall-gate lstm," in *IJCNN*, 2017.
- [4] R. T. Lowe, N. Pow, I. V. Serban, L. Charlin, C.-W. Liu, and J. Pineau, "Training end-to-end dialogue systems with the ubuntu dialogue corpus," *Dialogue & Discourse*, vol. 8, no. 1, pp. 31–65, 2017.
- [5] R. Kadlec, M. Schmid, and J. Kleindienst, "Improved deep learning baselines for ubuntu corpus dialogs," in *Workshop on Machine Learning for Spoken Language Understanding and Interaction at NIPS*, 2015.