Empowering Conversational Agents with Situated Natural Language Communication Skills by Exploiting Deep Reinforcement Learning Techniques

Alessandro Suglia *

Interaction Lab, Heriot-Watt University
Edinburgh Centre for Robotics
Edinburgh, Scotland, UK
as247@hw.ac.uk

Abstract. Since the early days of Artificial Intelligence, designing an agent able to interact with humans was considered a remarkable goal, aiming to replicate human cognitive capabilities in computer. In this work, we outline the main topics of a prospective PhD project whose aim is to exploit Deep Reinforcement Learning techniques to learn to interact with users in Natural Language conversations.

1 Background and Motivation

Natural Language (NL) is by far the easiest and most powerful communication device we possess, so it is reasonable to require an intelligent machine to be able to communicate through language [1]. Since the early days of Artificial Intelligence, researchers have tried to design systems able to communicate with humans by using different techniques ranging from simple *Rule-based* systems [2] to advanced data-driven *Spoken Dialogue systems* [3]. However, these systems rely on complex hand-engineered features (i.e., dialogue acts and slots) that are used to provide meaningful information about the utterance pronounced by the user and that are exploited by the system to easily retrieve relevant information for the user from a database. Despite the apparent effectiveness of this approach for specific domains, it is clear that this kind of supervision will not allow these systems to scale due to the high variability of Natural Language utterances and the high cost and effort of data annotation for each new application.

A dialogue is an activity between an hearer and a speaker which is characterized by the following characteristics: it is *temporal* because the answer of the system in a specific moment in time, has an impact on the state of the dialogue and it is possible to appreciate its quality only at the end of the dialogue. In

* A. Suglia is a 1st year PhD student in the CDT of "Robotics and Autonomous Systems" and is funded by an *EPSRC* studentship from the *Edinburgh Centre for Robotics*.

addition, a dialogue is incredibly *dynamic* because users usually talk about different topics in the same conversation without any problem. The above-mentioned characteristics have been described in [4]. A relevant characteristic that should be considered is that conversations are highly *contextualized*. In fact, all the agents involved in a conversation tend to align to a given topic or argument which depend on the specific situation in which the conversation is taking place. Moreover, users have specific preferences so it is fundamental to take them into account when a system should provide an effective service.

In recent years, the Deep Learning (DL) research field has designed models composed by different computational layers able to learn increasingly abstract representations that are specifically optimized for the task that the system should solve by using gradient-based learning algorithms [5]. In particular, inspired by the recent success of the *Sequence-to-Sequence* [6] model in Machine Translation, researchers have tried to learn neural network models able to "translate" the user utterance into the correct system response by leveraging large-scale corpora composed of many dialogue turns [7]. However, it has been demonstrated that this approach leads to models that generate the most frequent responses in the dataset. Thus, it is not able to capture the real semantics and intent of the user during the dialogue [8]. In addition, these models have demonstrated good performance only on large datasets obtained from movie transcripts [9] or *Ubuntu Chat dataset* [10] that completely lack incremental phenomena [11] that are incredibly common in real conversations such as restarts and corrections.

Reinforcement Learning (RL) techniques have been adopted in order to allow a system to acquire long-term planning capabilities that will help it to satisfy user goals. The system learns to interact with the user by optimizing a numerical value called *reward*, which represents an estimate of how good a given action is in a particular state [12]. In the long-term, it learns to generate appropriate trajectories of actions by maximizing a cumulative reward obtained in a given dialogue. Researchers have adopted slots to describe state and actions so as to cast the dialogue completion as an RL problem [4]. However, intelligent conversational agents should be able to interact with humans by *learning to compose meaningful sentences*. Just like a baby in its early stages that learns to speak by putting together simple words and by receiving supervision by his/her relatives about how good or bad they are.

The present work will discuss a prospective research agenda that I will conduct with Prof. Oliver Lemon [1] and with his research group. The proposal will try to outline possible ways to answer some of the following research questions:

R 1: Is it possible to train a conversational agent to interact within real world contexts consisting of embodied agents and situated objects?
R 2: Is it possible to train a conversational agent to generate accurate *contextualized* responses for the user?
R 3: Is the system able to adapt seemly to different domains by exploiting what it has previously learned?

[1] Director of the Interaction Lab at Heriot-Watt University, Edinburgh (https://sites.google.com/site/olemon/)

2 Project Agenda

Learning to interact with humans and the environment it is a desiderata for real world artificial agents having to support effectively humans during their daily tasks. We are convinced that an RL agent equipped with Deep Learning models may represent a sensible starting point for developing an effective embodied conversational agent. It is worth to note that, despite the recent success of Deep Reinforcement Learning algorithms (DRL) [13], they are still limited: they are not able to generalize the learned knowledge in a given task (e.g., learning to play Go) to different tasks (e.g., learning to play Risk) and they require an incredible number of examples (experiences) before they are able to grasp the relevant features that are required to succeed in the task. This is an incredible limitation for effective dialogue systems that are supposed to learn in an *online fashion* the user preferences and exploit them to provide personalized responses.

The potential of DL models to learn disentangled representations of the world combined with RL algorithms will help us to design an agent able to communicate with the user by *learning to compose sentences*. We decide to frame this task as an RL problem. Suppose that the agent receives in a given timestep t a state s_t and, exploiting its behaviour policy π, it samples *elementary symbols of language* [2] w_t until it decides to stop by generating a *dedicated termination symbol* w_{eos}. At the end of the sentence, it receives a reward which represents the correctness of the generated sentence according to the state s_t. Different reward functions have been proposed for different *Natural Language Understanding* tasks (see [14,15]). In these works, the agent receives a reward only after T timesteps (i.e., when the agent completes the sentence). Generally, this time lag can prevent the agent to understand the actual effects of its actions during the sentence generation process. In order to cope with this problem, we propose to leverage different linguistic resources with which is possible to evaluate the "grammaticality" and "naturalness" of the sentence, as proposed by [16]. In addition, these measures can be exploited to support the agent in the sentence generation task, providing it with additional "supervision". Indeed, especially in the early stages of training, this can be really beneficial to provide the agent with rewards associated to specific groups of *symbols of language* so that it can understand whether the generated symbols association is sensible. This incremental training strategy is in line with the so called *Curriculum Learning* training method [17]. We will also consider task-specific reward functions because we intend to evaluate the agent's capabilities to generate sentences for the task at hand[3].

We aim to investigate how to integrate different sensory modalities into conversational agents. Recently proposed DRL conversational agents have been equipped only with text-based interfaces which are quite restrictive [18]. On the other hand, we intend to equip the agent with several vision and audio modules to acquire additional multi-modal signals that it should leverage to complete its

[2] We talk about symbols because they can represent words or characters.

[3] For instance, we can think about measures that evaluate the number of correct movie recommendations for a given user.

tasks more effectively. For instance, it may use them to enrich its supervision signals with specific nuances coming from the user responses considering facial expressions or the tone of voice. DL models for vision and audio have achieved remarkable results in real world scenarios [19,20]. Therefore, it is reasonable to integrate them in a complete architecture for an embodied dialogue system. It is worth noting that we do not plan to integrate already pre-trained vision models as they are (as proposed by [21]) but we are convinced that an effective conversational agent should be able to optimize jointly all its modules, in an end-to-end fashion. Indeed, in this way the system will be able to learn latent representation that try to align visual representations to word representations, in an attempt to solve the *Language Grounding problem* [22].

The prospective PhD project may be divided in four macro phases: 1) study of the literature related to Conversational agents and DRL; 2) development of an advanced 3D environment as a testbed for embodied conversational agents; 3) development of a DRL model able to solve different tasks in the designed environment; 4) deployment of the system in real world robots[4] or in real world personal assistants like *Alexa* [5].

3 Conclusions

In this work we describe a prospective PhD project about DRL for real world Conversational Agents. We underline the relevance of the RL framework and its integration with DL techniques for the development of effective dialogue systems. We propose to tackle the problem by first designing a 3D environment which will be used as a testbed for conversational agents and then we will design a DRL based agents able to solve different tasks in the environment. Furthermore, we envision to deploy the trained model in real world robots.

References

1. Mikolov, T., Joulin, A., Baroni, M.: A roadmap towards machine intelligence. arXiv preprint arXiv:1511.08130 (2015)
2. Weizenbaum, J.: Eliza: a computer program for the study of natural language communication between man and machine. Communications of the ACM **9**(1) (1966) 36–45
3. Young, S.J.: Probabilistic methods in spoken–dialogue systems. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences **358**(1769) (2000) 1389–1402
4. Rieser, V., Lemon, O.: Reinforcement learning for adaptive dialogue systems: a data-driven methodology for dialogue management and natural language generation. Springer Science & Business Media (2011)
5. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553) (2015) 436–444

[4] See the equipment of the ROBOTARIUM at the Edinburgh Centre for Robotics: http://www.edinburgh-robotics.org/robotarium

[5] The Interaction Lab is a finalist in the Alexa Prize Challenge: https://goo.gl/FzoqMe

6. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. In: Advances in neural information processing systems. (2014) 3104–3112
7. Vinyals, O., Le, Q.: A neural conversational model. arXiv preprint arXiv:1506.05869 (2015)
8. Shao, Y., Gouws, S., Britz, D., Goldie, A., Strope, B., Kurzweil, R.: Generating high-quality and informative conversation responses with sequence-to-sequence models. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017. (2017) 2200–2209
9. Serban, I.V., Sordoni, A., Bengio, Y., Courville, A., Pineau, J.: Building end-to-end dialogue systems using generative hierarchical neural network models. In: Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI-16). (2016)
10. Uthus, D.C., Aha, D.W.: The ubuntu chat corpus for multiparticipant chat analysis. In: AAAI Spring Symposium: Analyzing Microtext. Volume 13. (2013) 01
11. Aist, G., Allen, J., Campana, E., Gallo, C.G., Stoness, S., Swift, M., Tanenhaus, M.K.: Incremental dialogue system faster than and preferred to its nonincremental counterpart
12. Sutton, R.S., Barto, A.G.: Reinforcement learning: An introduction. Volume 1. MIT press Cambridge (1998)
13. Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., et al.: Mastering the game of go with deep neural networks and tree search. Nature **529**(7587) (2016) 484–489
14. Ranzato, M., Chopra, S., Auli, M., Zaremba, W.: Sequence level training with recurrent neural networks. arXiv preprint arXiv:1511.06732 (2015)
15. Rennie, S.J., Marcheret, E., Mroueh, Y., Ross, J., Goel, V.: Self-critical sequence training for image captioning. arXiv preprint arXiv:1612.00563 (2016)
16. Novikova, J., Dušek, O., Curry, A.C., Rieser, V.: Why we need new evaluation metrics for nlg. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. (2017) 2231–2242
17. Bengio, Y., Louradour, J., Collobert, R., Weston, J.: Curriculum learning. In: Proceedings of the 26th annual international conference on machine learning, ACM (2009) 41–48
18. Serban, I.V., Sordoni, A., Lowe, R., Charlin, L., Pineau, J., Courville, A.C., Bengio, Y.: A hierarchical latent variable encoder-decoder model for generating dialogues. (2017)
19. Bojarski, M., Del Testa, D., Dworakowski, D., Firner, B., Flepp, B., Goyal, P., Jackel, L.D., Monfort, M., Muller, U., Zhang, J., et al.: End to end learning for self-driving cars. arXiv preprint arXiv:1604.07316 (2016)
20. Oord, A.v.d., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., Kavukcuoglu, K.: Wavenet: A generative model for raw audio. arXiv preprint arXiv:1609.03499 (2016)
21. Part, J.L., Lemon, O.: Teaching robots through situated interactive dialogue and visual demonstrations. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017. (2017) 5201–5202
22. Clark, H.H., Brennan, S.E.: Grounding in communication. Perspectives on socially shared cognition **13**(1991) (1991) 127–149