Deep Learning and Sentiment Analysis for Human-Robot Interaction

Mattia Atzeni and Diego Reforgiato Recupero

Università degli Studi di Cagliari
Department of Mathematics and Computer Science
Via Ospedale 72, 09124, Cagliari, Italy
m.atzeni38@studenti.unica.it
diego.reforgiato@unica.it

Abstract. In this paper we present an ongoing work showing to what extent semantic technologies, deep learning and natural language processing can be applied within the field of Human-Robot Interaction. The project has been developed for Zora, a completely programmable and autonomous humanoid robot, and it aims at allowing Zora to interact with humans using natural language. The robot is capable of talking to the user and understanding sentiments by leveraging our external services, such as a Sentiment Analysis engine and a Generative Conversational Agent, which is responsible for generating Zora's answers to open-dialog natural language utterances.

Keywords: Deep Learning, Sentiment Analysis, Human-Robot Interaction, Word Embeddings, Natural Language Processing, LSTM

1 Introduction

Recent advances in automation and robotics are currently allowing humans to effectively cooperate with robots in several environments where the ultimate effectiveness of the interaction only depends on the actual success of accomplished tasks. This kind of scenarios is widespread in industrial settings and often requires a low degree of autonomy, thereby turning into a low-level and non-intuitive human-robot interaction, which cannot be fulfilled by inexperienced or untrained operators.

On the other hand, nowadays robots are starting to play important roles also in human environments, where a higher degree of autonomy, as well as a richer and more complex level of human-robot interaction, is required [4]. This gives rise to some main challenges and research opportunities, as robots should be able to cooperate with humans using natural language. Current research is mainly focusing towards supervised and unsupervised approaches for the semantic parsing of natural language commands into formal meaning representations [1]. In this context, human-robot interaction also involves natural language processing, knowledge representation and commonsense reasoning. Recently, Semantic Web technologies have also been applied to efficiently perform fine-grained sentiment analysis, resolving holders and topics on Linked Data [12].

Hence, in this paper, we present a use case and an ongoing work which shows how effectively sentiment analysis and deep learning can be used for human-robot interaction. The work has been accomplished using Zora[1], an interactive and programmable humanoid robot built on top of NAO by Zorabot[2]. The robot is capable of interacting with humans using natural language, by leveraging our external services, which allow *(i)* to generate the robot reply to a natural language question and *(ii)* to perform sentiment analysis on the given textual input.

2 System Architecture

The Zora robot is completely programmable and the Choregraphe suite[3] allows to *(i)* combine and create different behaviors using a visual programming approach, *(ii)* create animations by means of an intuitive and dedicated user interface and *(iii)* test behaviors and animations on simulated virtual robots, or directly on the real one. The system has been designed so that all the heavy computations run on a server, while the robot is only responsible for the interaction with the final user. Figure 1 shows the high-level architecture of the system.

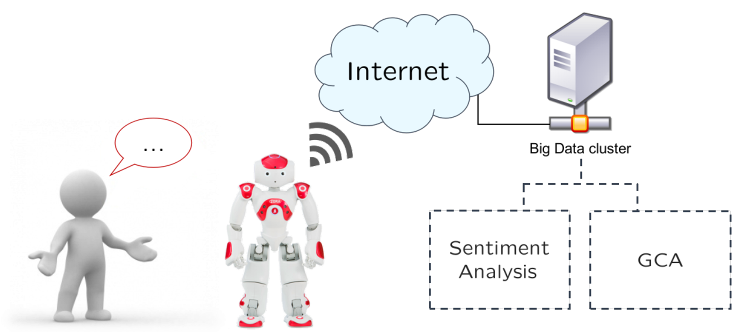

Fig. 1. High-level architecture of the system.

The robot is equipped with four microphones: two at the front of the head, and two at the back. Hence, the robot can easily record the human voice, which is contextually analyzed and turned into text by a speech recognition module powered by Nuance[4]. However, to improve performances, we are currently relying on cloud computing also for speech recognition. This allows us to pre-process

[1] http://zorarobotics.be/index.php/en/who-am-i
[2] http://zorarobotics.be/index.php/en/
[3] http://doc.aldebaran.com/1-14/software/choregraphe/index.html
[4] https://www.nuance.com

the sound recorded by the robot, in order to remove fan noise, which is quite disturbing for converting the human voice into written text. The resulting audio file is then sent to *IBM Watson Speech to Text*[5], to perform speech recognition.

Zora's reply to the natural language input can either be *(i)* an action triggered by a simple command from the user, such as "sit down" or "move your hand", or *(ii)* a natural language answer to the question asked by the user.

Hence, the result given by the speech recognition module is analyzed by means of *QiChat*[6], a language specifically designed to handle dialogs between the robot and humans. This allows checking whether the user input matches some simple command aimed at triggering an action corresponding to predefined behaviors which can be directly accomplished by the robot. In this case, Zora starts a new behavior to satisfy the user request.

If the user input does not match any predefined action, then it is sent to our RESTful web services, where the text is analyzed by a Sentiment Analysis engine and by a Generative Conversational Agent (GCA). This allows the robot to answer any question asked by the user and to understand sentiments hidden in the text. The following sections dig into more details about the Sentiment Analysis module and the GCA.

3 Sentiment Analysis

The sentiment analysis system has been implemented following our previous work in [2] and [5], but it has been significantly redesigned and improved to achieve better performances. The system has been implemented in Java and makes use of Stanford CoreNLP [8] for standard preprocessing tasks, such as sentence detection, tokenization, POS tagging and stop-words removal. On the other hand, feature extraction and classification have been implemented using Deeplearning4j[7]. Classification relies on recurrent neural networks (RNNs) and, more precisely, LSTM (Long Short-Term Memory), while feature extraction is achieved using neural word embeddings, obtained by applying the global log-bilinear regression model GloVe [10], trained on Wikipedia 2014 and Gigaword 5, for a total of 6 billion tokens and resulting in 300-dimensional word vectors.

The system has been trained for both polarity detection and subjectivity detection. Polarity detection has been achieved by training the model on the union of two datasets: *(i)* the *Large Movie Review Dataset*[8], introduced in [7], a balanced dataset consisting of 50000 highly polarized movie reviews extracted from IMDb[9], and *(ii)* a dataset containing one million reviews extracted from Amazon. This dataset is balanced as well, and it has been introduced in [11], within the Semantic Sentiment Analysis Challenge at ESWC2017.

[5] `https://www.ibm.com/watson/services/speech-to-text/`
[6] `http://doc.aldebaran.com/2-4/naoqi/interaction/dialog/aldialog_syntax_toc.html`
[7] `https://deeplearning4j.org/`
[8] `http://ai.stanford.edu/~amaas/data/sentiment/`
[9] `http://www.imdb.com/`

The system has also been trained for subjectivity detection on a dataset which contains 5000 objective sentences and 5000 subjective sentences [9]. The performances of the system have been measured using 10-fold cross-validation, achieving an accuracy of 0.88 for polarity detection and 0.90 for subjectivity detection. Such results show a considerable improvement with respect to our previous work in [2], which was already ranked first at ESWC 2017 Challenge on Semantic Sentiment Analysis [11].

4 Generative Conversational Agent

The Generative Conversational Agent (GCA) allows the robot to perform open-domain dialog generation, that is creating meaningful and coherent responses given the dialog history. The implementation of the GCA is based on previous work in [6] and relies on sequence to sequence (seq2seq) modeling, thereby being language independent and capable of implicitly learning both semantics and syntax. Figure 2 shows the GCA model.

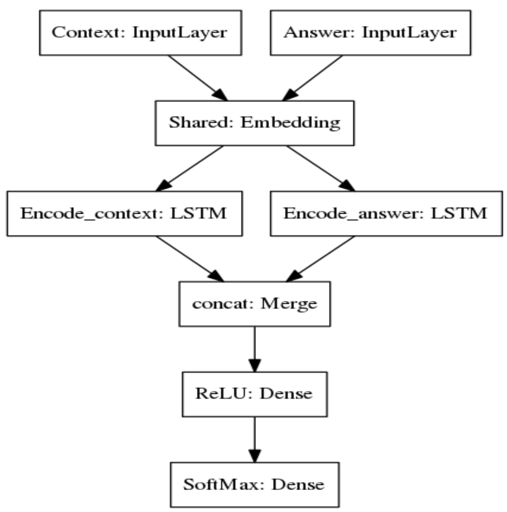

Fig. 2. The Generative Conversational Agent model.

The GCA architecture assumes the same distributions for input and output words, and, subsequently, an embedding layer, based on GloVe word embeddings, is shared between the encoding and decoding processes. The model makes use of two LSTM networks to process respectively the dialog context and the incomplete answer generated up to the current token. The resulting vectors are concatenated and provided to dense layers that predict the current token of the answer. The GCA has been trained on the *Cornell Movie Dialogs Corpus* [3], which includes more than 300000 natural language utterances.

5 Conclusion and Demo Showcase

In this paper, we have introduced a project aimed at making the Zora robot capable of *(i)* talking with humans using natural language, *(ii)* executing specific actions triggered by natural language commands and *(iii)* understanding sentiments by leveraging an external RESTful web service. The robot, along with how it is possible to effectively and easily use the Choregraphe suite and visual programming techniques to create new behaviors, will be showcased during the demo. A video showing an example of a basic interaction with Zora is available at: `https://youtu.be/hoJYHLecZMs`.

References

1. Atzeni, M., Atzori, M.: Towards Semantic Approaches for General-Purpose End-User Development. In: 2018 Second IEEE International Conference on Robotic Computing. pp. 369–376 (2018)
2. Atzeni, M., Dridi, A., Reforgiato Recupero, D.: Fine-Grained Sentiment Analysis on Financial Microblogs and News Headlines. In: Dragoni, M., Solanki, M., Blomqvist, E. (eds.) Semantic Web Challenges - 4th SemWebEval Challenge at ESWC 2017, Revised Selected Papers. vol. 769, pp. 124–128. Springer (2017)
3. Danescu-Niculescu-Mizil, C., Lee, L.: Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs. In: Workshop on Cognitive Modeling and Computational Linguistics, ACL (2011)
4. Dautenhahn, K.: Roles and functions of robots in human society: Implications from research in autism therapy. Robotica 21(4), 443–452 (Aug 2003)
5. Dridi, A., Atzeni, M., Reforgiato Recupero, D.: FineNews: fine-grained semantic sentiment analysis on financial microblogs and news. International Journal of Machine Learning and Cybernetics (2018), https://doi.org/10.1007/s13042-018-0805-x
6. Ludwig, O.: End-to-end adversarial learning for generative conversational agents. CoRR abs/1711.10122 (2017), http://arxiv.org/abs/1711.10122
7. Maas, A.L., Daly, R.E., Pham, P.T., Huang, D., Ng, A.Y., Potts, C.: Learning word vectors for sentiment analysis. In: Proc. of the 49th Annual Meeting of the ACL: Human Language Technologies. pp. 142–150 (June 2011)
8. Manning, C.D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S.J., McClosky, D.: The Stanford CoreNLP natural language processing toolkit. In: Association for Computational Linguistics (ACL) System Demonstrations. pp. 55–60 (2014)
9. Pang, B., Lee, L.: A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In: Proc. of the ACL (2004)
10. Pennington, J., Socher, R., Manning, C.D.: Glove: Global vectors for word representation. In: Empirical Methods in Natural Language Processing (EMNLP). pp. 1532–1543 (2014), http://www.aclweb.org/anthology/D14-1162
11. Reforgiato Recupero, D., Cambria, E., Di Rosa, E.: Semantic Sentiment Analysis Challenge at ESWC2017. In: Dragoni, M., Solanki, M., Blomqvist, E. (eds.) Semantic Web Challenges. pp. 109–123. Springer International Publishing (2017)
12. Reforgiato Recupero, D., Consoli, S., Gangemi, A., Nuzzolese, A.G., Spampinato, D.: A semantic web based core engine to efficiently perform sentiment analysis. In: Presutti, V., Blomqvist, E., Troncy, R., Sack, H., Papadakis, I., Tordai, A. (eds.) The Semantic Web: ESWC 2014 Satellite Events. pp. 245–248. Springer (2014)