Natural Language Processing for Robot Manipulator Control

Peter Khiem Duc Tinh Nguyen

Pengu20@student.sdu.dk

Bachelors contract

Supervisor: iñigo iturrate
co-supervisor: Martin Sundahl Laursen
co-supervisor: Jannik Skyttegaard Pedersen

University of Southern Denmark
Robot systems engineering

Date of Hand In: 31. January

Context

A rise in the use of robots in the industry is currently taking place. Such movement towards automatic production is often referred to as the "fourth industrial revolution". More robots in the industry means more robot control. The fourth industrial revolution therefore puts pressure on the limits of human-robot collaboration (HRC) [1]. The expansion of robots into human workspaces has paved the way for collaborative robots, which have the innate ability of HRC, as they are significantly safer to work with than their fully automated past counterparts[2]. But even then, the use of HRC is not what it could be. As the level of collaboration within the industry described by Wilhelm Bauer, resides mostly at the coexistence and synchronization level[3]. It's proposed that the lack of collaboration between the robot and the human stems from the complexity of work that a human must perform to control the robot. Assuming the statement to be true, then it's hypothesized that the use of natural language to control robots could enable, or at least ease, the transition into true collaboration between robots and humans in the industry.

Related work

In the field of natural language processing (NLP) for robot control, it is typical to conduct NLP on text as opposed to speech[4][5][6][7]. Translating voice to text, and processing text is therefore usually seen as two separate problems. Creating text from speech is usually overlooked, as it is such a popular field itself, that there is no benefit of making such a method specifically for robot control[8].
Processing text and parsing it into robot actions, is the main problem that will be addressed in this project.

A related paper is published, where the workflow for their natural language processing is by using a pre-built voice-to-text system based on pythons "SpeechRecognition" created by Google. The processing afterwards comprises an algorithm that detects any actionable commands within the text. With such a system and the use of robot vision, they were able to pick and place objects only using voice commands[5].

Another work that was done used a recurrent neural network to map text to predefined actions, such that the robot could learn what commands the sentences had[4].

Using neural networks has been a great success in the field of natural language processing[9][10][11]. Almost all mentioned papers, makes use of neural networks in some way[5][6][7]. Therefore, it will also be the main method used to solve the problem given in this bachelor project.

Project description

This project will be a study of verbal control for collaborative robots. The collaborative robot in question is considered being a Universal Robot[1]. The work flow for the project is shown in figure 1. A simple subset of verbal instructions will describe the robot movements. These instructions will be processed through a neural network, with every word and its corresponding grammatical category as its output. This type of categorization is called part of speech tagging (POS tagging). The neural network is considered to be made based as either a transformer architecture [10][11], or a recurrent neural network architecture[4].
The parser will analyse the neural network output (POS tagging), and determine the action by identifying important key information in the text. This key information may include specific keywords and their associated grammatical categories. After determining the action, the parser translates it into a programming language called URScript format, which can be understood by universal robots. This allows the robot to carry out the verbal action given.

[1]https://www.universal-robots.com/

The voice-to-text part is optional, as the best solution would be to use prebuild voice to text software.

The training data for the neural network will be provided using the Huggingface token classification datasets. [2]

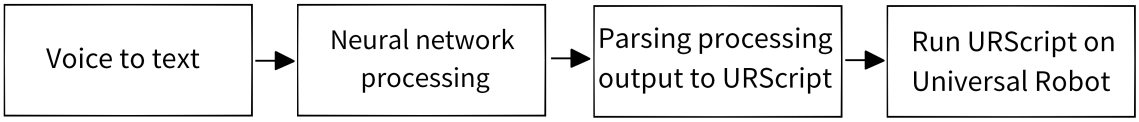

Fig. 1: Diagram of the work structure

Timeline

The suggested timeline for the project can be seen in figure 2, where the project is to start the 1. February and end the 31. of may. The first half of February will be spent researching related work. A month (from mid February to mid March) will be spent designing and training the neural network. another halv a month (from mid March to April) will be spent creating and testing URScripts. The month of April will be used to create the parser. The last month, may, will be spent getting the final results and finishing the report.

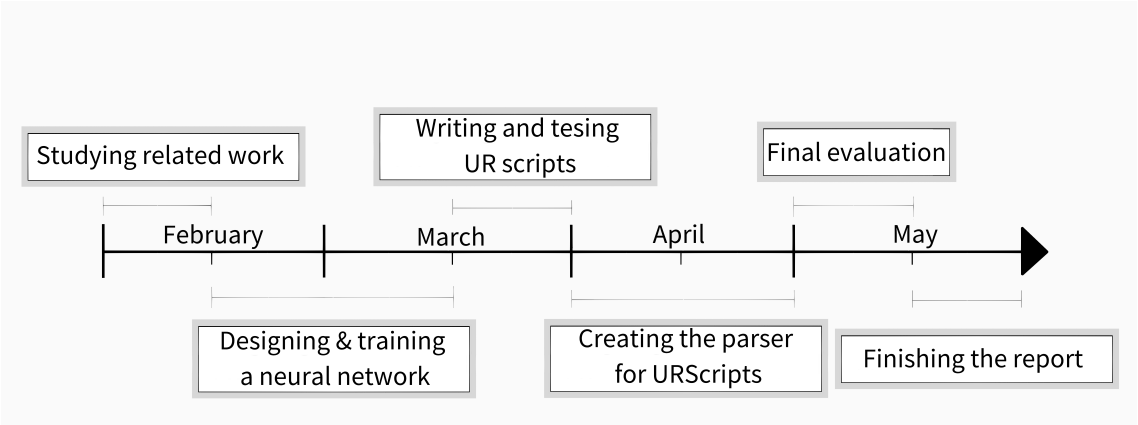

Fig. 2: Suggested project timeline

[2] https://huggingface.co/

Bibliography

[1] Olivares-Alarcos, Alberto. "On Inferring Intentions in Shared Tasks for Industrial Collaborative Robots". In: *Electronics* 8 (Nov. 2019).

[2] Contributor, TechTarget. *collaborative robot (cobot)*. URL: https://www.techtarget.com/whatis/definition/collaborative-robot-cobot.

[3] Wilhelm, Bauer et al. *Lightweight robots in manual assembly – best to start simply! Examining companies' initial experiences with lightweight robots*. Oct. 2016.

[4] Kahuttanaseth, Wittawin, Dressler, Alexander, and Netramai, Chayakorn. "Commanding mobile robot movement based on natural language processing with RNN encoderdecoder". In: *2018 5th International Conference on Business and Industrial Research (ICBIR)*. 2018, pp. 161–166. DOI: 10.1109/ICBIR.2018.8391185.

[5] Wu, Shih-Hung and Hong, Xie-Sheng. "Integrating Computer Vision and Natural Language Instruction for Collaborative Robot Human-Robot Interaction". In: *2020 International Automatic Control Conference (CACS)*. 2020, pp. 1–5. DOI: 10.1109/CACS50047.2020.9289768.

[6] Matuszek, Cynthia et al. "Learning to Parse Natural Language Commands to a Robot Control System". In: *Experimental Robotics: The 13th International Symposium on Experimental Robotics*. Ed. by Desai, Jaydev P. et al. Heidelberg: Springer International Publishing, 2013, pp. 403–415. ISBN: 978-3-319-00065-7. DOI: 10.1007/978-3-319-00065-7_28. URL: https://doi.org/10.1007/978-3-319-00065-7_28.

[7] Suárez Bonilla, Félix and Ruiz Ugalde, Federico. "Automatic Translation of Spanish Natural Language Commands to Control Robot Comands Based on LSTM Neural Network". In: *2019 Third IEEE International Conference on Robotic Computing (IRC)*. 2019, pp. 125–131.

[8] Wood, Laura. *Global Speech-to-text API Market to 2027*. URL: https://www.businesswire.com/news/home/20220504005949/en/Global-Speech-to-text-API-Market-to-2027---Size-Share-Industry-Trends-Analysis-Report---ResearchAndMarkets.com.

[9] Padmanabhan, Arvind. *Neural Networks for NLP*. URL: https://devopedia.org/neural-networks-for-nlp#cite-as.

[10] Vaswani, Ashish et al. "Attention is all you need". In: *Advances in neural information processing systems* 30 (2017).

[11] Devlin, Jacob et al. "Bert: Pre-training of deep bidirectional transformers for language understanding". In: *arXiv preprint arXiv:1810.04805* (2018).