Parsing Natural Language Sentences into Robot Actions

Gianluca Bardaro[1], Danilo Dessí[2], Enrico Motta[1], Francesco Osborne[1], and Diego Reforgiato Recupero[2]

[1] Knowledge Media Institute, The Open University, Milton Keynes (UK)
{gianluca.bardaro, enrico.motta, francesco.osborne}@open.ac.uk.it
[2] Department of Mathematics and Computer Science, University of Cagliari (Italy)
{danilo_dessi, diego.reforgiato}@unica.it

Abstract. In this paper we propose a Natural Language Processing engine that allows the NAO humanoid robot to execute natural language actions spoken by the user. To this aim we created an ontology that describes body parts, actions, and incompatibilities between actions. The system can work in two modes: *stateless* and *stateful*. In *stateless* mode, the robot returns to its default position after each action. In *stateful* mode, it performs the actions sequentially and may refuse a command if incompatible with the robot current state. Our system handles compound and multiple expressions that the robot understands and performs.

Keywords: Language Understanding · Humanoid Robot · Robot Action Ontology · Human-Robot Dialogue · Ontology Design.

1 Introduction

[3]Recent innovations in the field of Robotics and Artificial Intelligence (AI) have led to the development of more and more robotic oriented applications and to the belief that *there is a 50% chance of AI and robotics outperforming humans in all tasks in 45 years and of automating all human jobs in 120 years* [2]. Quickly, social robots are taking hold and we are starting to see them in different countries where they are employed in several domains. Their common goal is to interact more effectively and efficiently with humans. Human-robot interaction involves a wide set of cutting edge technologies focused within the domain of Natural Language Processing (NLP) [3], Semantic Web [4], Knowledge Representation, AI and Commonsense Reasoning; successfully works have already appeared in literature such as [1], where authors effectively proved how sentiment analysis and deep learning approaches can be used for human-robot interaction.

This poster paper focuses on the understanding of natural language expressions in the English language related to action commands that are given by humans to Zora[4], an interactive and programmable humanoid robot built on top of NAO[5] that we have employed. In particular, we created an NLP engine

[4] http://zorarobotics.be/index.php/en/who-am-i
[5] http://zorarobotics.be/index.php/en/

that analyzes the input expressions in natural language and map them to a set of actions and body parts described in a domain ontology. These actions are related to movements of its legs, arms, hands, head, eyes, the whole body (e.g., walking forward, backward, aside, sitting, crouching) and of its eyes (e.g., rolling, twinkling). The robot was programmed (i) to perform the identified actions, (ii) to ask the user to specify missing elements if the action command is not complete (e.g., if the user asks to raise its arm, the robot will ask the user which one - left or right - it needs to raise), (iii) to tell the user whether the expression he/she said is not an action command that the robot can perform, and, finally, (iv) to perform the next recognized action only if it is compatible with its current position (state). A video showing the proposed system is publicly available[6].

2 System Architecture

The Zora robot is completely programmable and the Choregraphe suite[7] allows *(i)* combining and create different behaviours using a visual programming approach, *(ii)* creating animations by means of an intuitive and dedicated user interface, and *(iii)* testing behaviors and animations on simulated virtual robots, or directly on the real one. It is equipped with four microphones and can easily record the human voice, which is contextually analyzed and turned into text by a speech recognition module powered by Nuance[8]. Figure 1 shows the architecture of the system. The user interacts with the robot, which sends his/her input to the cloud through the Internet for the highly expensive computations. In the cloud we have the NLP engine running, which takes a text as input and uses a pipeline of NLP tools and semantic technologies to return a list of commands for Zora to perform. In presence of a partial command (when the user does not indicate some elements in his/her command to identify a unique action), the NLP engine is able to detect and process further those statements. As an example, in the text *Hello Zora, now raise your arm*, the adjective left or right has not been specified. In such a case, the robot is informed by the NLP engine and interacts with the user by asking whether he/she meant the right or left arm of the robot.

3 Understanding Natural Language Expressions

After the speech to text, the user's sentence is used as input for the NLP Engine we have developed. The engine uses Stanford CoreNLP[9] and returns an RDF that contains the base forms of words, their part of speech (POS), dependency and parse trees.

The first thing performed by the NLP engine is to look for those tokens that have been labeled as verbs by the POS tagger. In addition, it exploits the extracted syntactic dependencies relationships for finding tokens that are linked to the identified verbs (e.g., phrasal verbs) to better capture the semantics of the command. For example in the sentence *Zora, now stand up*, the robot action

[6] https://www.youtube.com/watch?v=CC9NzlbF0gQ
[7] http://doc.aldebaran.com/1-14/software/choregraphe/index.html
[8] https://www.nuance.com
[9] https://stanfordnlp.github.io/CoreNLP/

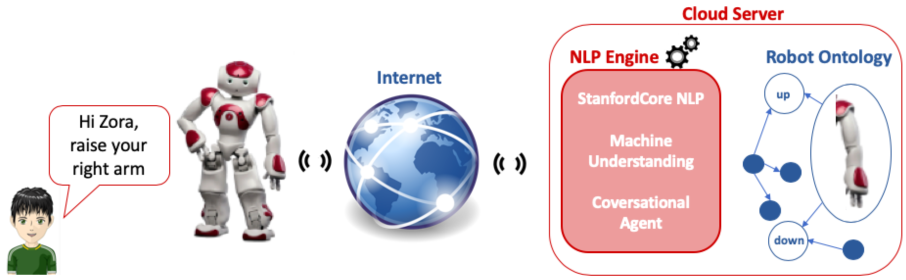

Fig. 1. High-level architecture of the system.

is generated by using the extracted phrasal verb *stand up*. Then, the detected expression is compared with a list containing verbs and synonyms. If a match is detected, then the NLP engine looks for syntactic dependencies to find possible direct object dependency. The list of robot body parts, their related synonyms, and adjectives (e.g., up, down, left, right) are searched using the object as a key to find a match. When the user indicates plural body parts (e.g., arms, legs), as in *Zora, move your hands down*, possible adjectives are ignored, as there is no need to specify either left or right. On the other hand, when the identified body part is a hand, arm or leg, the associated adjective (left or right), must be identified to detect one action only to be executed by the robot. For such a purpose, the NLP engine adopts the adjectival dependency of the extracted verb. Therefore, with all this extracted information, the engine can query an OWL ontology that includes the robot body parts and all the related actions.

The ontology[10] consists of two parts. The first part includes two main classes, `RobotAction` and `RobotBody`. `RobotAction` includes three subclass: `BaseAction`, describing all the basic movements that Zora performs with its parts, `Compound-Action`, which define complex action including multiple basic actions, and `Simple-Action` defining general movements Zora performs with its entire body such as walking, rotating, lying back, and so on. Hands, arms, legs and all physical parts of Zora that can be involved in basic actions are defined within the `RobotBody` class. The ontology includes 48 individuals of type `RobotBody` and `RobotAction`. The understanding of natural language expressions has been included in the second part of the ontology. The list of alternative keywords (individuals) for all possible actions that Zora can perform and all its body parts are respectively defined within the `ActionWord` and `BodyPartWord` classes. For example, the individual `sit` of the class `ActionWord` is related to the sit action for the robot and includes different `synonym` relations to the list of expressions that convey the same meaning (i.e. *sit down*, *seat*, and so on).

If the part of the body has not been correctly specified, the robot can ask the user to specify what is missing.

[10] http://www.w3id.org/zoraActions

For example, after the command *Zora, please raise your leg*, the robot will reply back asking the user to specify which leg (if left or right) he/she meant. Zora will perform the action only once the user specifies the missing information. There are two modes our system can work: *stateless* and *stateful*. In *stateless* mode, an identified action is performed by the robot which returns then to its default position. In *stateful* mode, the robot performs the actions sequentially without returning to its default posture and may refuse a command if it is incompatible with its current state.

Let us suppose that the robot is standing on its left leg and the user asks the robot to raise its right leg. Then, the robot will reject the user's request and will warn the user about the incompatibility issue.

The current state of Zora is checked by the NLP engine. The compatibility of the next action to be performed is then checked as well as soon as the current state of the robot is identified. In presence of incompatibility issues, the robot informs the user with a vocal message. All combinations of actions and postures have been tested to generate the list of incompatibilities.

Two subsequent actions were defined as incompatible if they resulted in the robot falling down or being physically unable to perform the second action.

The source code of the NLP engine, the Choregraphe program we developed and the entire list of incompatibilities are publicly available[11].

4 Conclusion

In this poster paper, we propose and share with the community a new system based on NLP and semantic technologies that allows the Zora humanoid robot to *(i)* recognize natural language expressions indicating action commands, *(ii)* ask the user to specify a term if the given command lacks some specific term, *(iii)* execute sequential and compound actions triggered by natural language commands, and *(iv)* understand if the performance of a given action is compatible with the current position of the robot when it is in *stateful* mode.

References

1. Atzeni, M., Recupero, D.R.: Deep learning and sentiment analysis for human-robot interaction. In: ESWC 2018 Satellite Events, Heraklion, Crete, Greece, Revised Selected Papers. pp. 14–18 (2018), https://doi.org/10.1007/978-3-319-98192-5_3
2. Grace, K., Salvatier, J., Dafoe, A., Zhang, B., Evans, O.: When will AI exceed human performance? evidence from AI experts. CoRR **abs/1705.08807** (2017), http://arxiv.org/abs/1705.08807
3. Hameed, I.A.: Using natural language processing (nlp) for designing socially intelligent robots. In: 2016 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob). pp. 268–269 (Sep 2016). https://doi.org/10.1109/DEVLRN.2016.7846830
4. Kobayashi, S., Tamagawa, S., Morita, T., Yamaguchi, T.: Intelligent humanoid robot with japanese wikipedia ontology and robot action ontology. In: 2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI). pp. 417–424 (March 2011). https://doi.org/10.1145/1957656.1957811

[11] Data and code: https://github.com/hri-unica/Humanoid-Robot-Obeys-Human-Action-Commands-through-a-Robot-Action-Ontology