Building Robots that work with people

Presented By: Jean Oh | Project Scientist, RI
hyaejino@andrew.cmu.edu

Scribed By: Abhishek Bhatia, Pranav Maheshwari, Rushat Gupta Chadha, Tushar Chugh
Summary

Robots are increasingly becoming key players in human–robot teams. To become effective teammates, robots must possess profound understanding of an environment, be able to reason about the desired commands and goals within a specific context, and be able to communicate with human teammates in a clear and natural way. To address these challenges, Jean and other people at NREC have developed an intelligence architecture that combines cognitive components to carry out high-level cognitive tasks, semantic perception to label regions in the world, and a natural language component to reason about the command and its relationship to the objects in the world. Jean in her guest lecture, discussed about the recent developments and her research in particular in the related domain by demonstrating some results from outdoor experiments with mobile robots and with robots having capabilities of manipulation. [1]

Introduction

A multidisciplinary approach to robotics has the potential to create competent human–robot teams. Different people might look for different attributes in robots to collaborate with. Some people might prefer that robots should have a predictable behavior, some will want more intelligence, other might ask for clear communication skills and for other the look and feel of robots might matter.

In terms of technology Autonomous Robots is driven by various other disciplines. There has been a decent progress already made in the field of path/motion planning, obstacle avoidance, controls, terrain/object recognition, 3D sensing and spatial reasoning. But there are still a lot of challenges to be solved in the domain of perception, semantic and communication (between robot–robots, robots–other devices and robot–human). The domains are difficult but necessary in order to make robots work with people.
Intelligent Robot Teammates

In the present state, the robots understand the commands from the users which are mathematical or expressed in terms of coordinates. For example, we can direct the robots to reach at a particular location with some coordinated \((x,y,z)\) by avoiding obstacles coming along the way or to pick up an object kept at \((x_1, y_1, z_2)\) and move to \((x_2, y_2, z_2)\). The ability to learn new skills is missing.

Therefore, there is a motivation to introduce robots with cognitive abilities. A robot for example should be able to understand and execute a command like ‘Go to the back of the building that is left of the car’.

These attributes to cognitive abilities include:

- Natural Communication
- Semantic Perception
- Spatial Reasoning
- Semantic understand of tasks and environment

RCTA Intelligent Architecture

The architecture that has been developed by the team at NREC to create robots with cognitive abilities is discussed in this section. At the higher level, the architecture is divided into three level which refers to the world model and is interfaced to the robot platform.

The higher most level deals with contest related templates and is called as mission level. Here the uncertainty is quite high as the robot can only interpret things without any knowledge from perception and the user. The middle layers is the attention level which deals with geometric reasoning. Here, uncertainty is medium as through perception we can figure out shapes and with planning things and own location can be localized with small uncertainty. The lower most layer is interfaces through the control and hence has quite a low uncertainty. This layer is termed as the interaction layer.

Figure 2: RCTA High Level Architecture
The three layers are explained further in figure 3. Here notice the multi-modal interface that is the layer of interaction of robots with the human team mates. This interface communicated with Natural Language Units to get the commands from the user which can be converted to corresponding actions. The mission planner which is driven by the multi-modal interface then drives the action to navigate, search or observe.

**Figure 3: RCTA Detailed Architecture**

**Example**

With the architecture explained above a mobile robot was given the following command

“Stay to the left of the building; navigate to a traffic barrel that is behind the building.”

The environment to be navigated is shown in figure 4.
Figure 4: Test area for mobile robot

Figure 5 shows how the robot interprets the command to understand the goal, classify the world and navigates to the destination. Here:

1. Semantic Classification
2. Labeled 3-D points cloud
3. Walls detected and merged into a wall cluster
4. A building is predicted
5. A traffic barrel is hypnotized behind the predict
6. Path planning satisfying spatial constraint “left of”
7. A real traffic barrel is detected, path re-planned [2]
The most important thing to think here is that, the language has been used as a sensor which makes robot more cognitive and robot human interaction more natural.

**Perception**

Challenges with perception lie in poor performance which can be quantified in terms of recall and precision. **RECALL** is the ratio of the number of relevant records retrieved to the total number of relevant records in the database. It is usually expressed as a percentage. **PRECISION** is the ratio of the number of relevant records retrieved to the total number of irrelevant and relevant records retrieved. It is usually expressed as a percentage.

![Figure 6: Relation between Precision and Recall](image)

Present day perception systems suffer from low recall and precision. For instance, a poor object detection algorithm either won’t or incorrectly detect the objects present in the environment.

**Semantic Understanding**

In a given environment, developing a semantic understanding of the scene is essential. As can be seen in the scene below, executing a command such as “Navigate left of a car to a traffic barrel near the gas pump”, isn’t really possible due to the various uncertainties.

![Figure 7: World model developed via perception](image)
The semantic model can be vastly improved by fusing multiple sensors and building a probabilistic model for label prediction of the objects present in the environment. These redundancies reduces the failure rate of the perception and the results, compared to the above figure, are much cleaner, as can be seen below.

Figure 8: Cleaner world model developed by fusion of multiple sensor data

Language too plays an important role when used as a sensor. Having a model for Natural Language Processing helps in providing context and removing uncertainties by defining specific objects, actions, etc. This is generally done by creating custom data structures which stores metadata regarding all the symbols in the robot’s environment.

```
{
  "tsbVersion": 6.0,
  "action": "navigate",
  "object": "robot",
  "mode": "quickly",
  "goal":
  {
    "relation": "front",
    "name": "TrafficBarrel",
    "symbols": [],
    "selectedSymbol": 0,
    "spatialConstraints":
    {
      "opType": "AND",
      "terms": [
        "relation": "left",
        "name": "Building",
        "symbols": [],
        "selectedSymbol": 0
      ]
    }
  },
  "symbols": [
    {
      "id": 17,
      "p": 0.25
    },
    {
      "id": 3,
      "p": 0.75
    }
  ],
  "selectedSymbol": 3,
}
```

Figure 9: Example Data structure to define symbols in the environment
**Imitation learning**

Imitation learning is a way to train the robot by giving examples and expecting it to imitate the same behavior. In case of navigation, (shown in the figure below) there are 2 ways of getting from pt A to pt B viz., quickly and covertly. Once it gets several training examples for covert, the robot will learn to generate a cost map which mimics the behavior.

![Figure 10: Quick and covert paths as defined by user](image)

Inside the robot, an optimization algorithm is used (see above). The machine generates a path and compares it with the training examples and it tries to minimize the difference between the 2 paths until it reaches a really similar path.

These approaches can be used for language grounding too. Once the relations between the language commands to the navigation are learnt, we can evaluate all the candidates and then use Computer vision to know what to do. The math involved to do so is –

![Figure 11: Math behind Imitation learning and Cost map generation](image)
Where the weights are learnt through imitation learning

For example –

![Figure 12: Semantic perception via object grounding](image)

If the user says, “grab the cup which is on the right side”

**Grab the cup** would make the algorithm choose cup 1 as it has a higher probability but **on the right** means the algorithm will be forced to choose the cup 2 even though it has a lower probability and this makes sure that we get the desired results even with poor Computer Vision results.

The making of the probabilistic map also involves decoding the language and predicting objects based on directions finally making sense of the world from what it sees to what its told to do. Semantic perception helps in hypothesizing an environment beyond physical sensor ranges. For example, the statement “**A traffic barrel that is behind the building**” makes the robot make a probabilistic map beyond what it sees (which is just a wall).

![Figure 13: Predicting objects based on directions](image)
The steps involved would be to –

1. Building suggests there is a structure behind the wall (rectangular) – the probability of the object is then generated.

2. The word behind narrows it down to the portion behind the building

3. The word traffic barrel makes sure that the robot knows what the shape and size is and it randomly assigns a location for the barrel till it reaches there, sees it and updates the location.

Jean Oh and her team performed all the experiments on a Clearpath Husky with the following add ons –

![Figure 14: Robot Platform](image)

**Human Teammate**

It is ironical in a way that the robots are designed to help humans, but with increasing complexities in the design, implementation and also the environment where the robot has to function, there is a need for humans to help robots in the decision making. The experiment below demonstrates how extra information from the human, helped the robot to converge faster towards the goal position. The intention for the robot was to arrive at a goal position that was towards the right of the shed, but human instructions also included “Get off the grass”, which helped the robot decide what the ‘right’ of the shed was in an easier way and finally arrive to its goal position faster.
How human intervention can help robots reach the target position faster:

- If the information provided is unambiguous.
- If the robot and human interact during the overall course of operation, robot can provide its thought process and human can comment and suggest alternatives.
- The robot can also learn a lot from observing the human trajectories and actions. There are certain things which are not detected by robot’s sensors, like doors or threats, but can be detected by observing the human trajectory around those.
- Besides, a lot of data mining techniques could also be used to increase interactions between robots and humans (multi-human multi-robot teams) and help robots with complex manipulation tasks.
References

1. Toward Mobile Robots Reasoning Like Humans: Jean et.al | RCTA–AAAI 2015
2. RCTA Semantic Navigation: Jean et.al
3. Robot videos are available on YouTube:

https://www.youtube.com/channel/UCj1AYMQiEOHS3W0GH5zykQ