Learning language with robots

Simon Dobnik

31 European Summer School in Logic, Language and Information
12 August 2019, Riga Latvia
Several conversational interfaces

The longest river in the European Union, the Danube River is the second-longest river in Europe after Russia's Volga. It begins in the Black Forest region of Germany and runs through 10 countries (Germany, Austria, Slovakia, Hungary, Croatia, Serbia, Romania, Bulgaria, Moldova and Ukraine) on its way to the Black Sea.

About the Danube River - Viking River Cruises
www.vikingrivercruises.com/cruise-destinations/europe/rivers/danube/about.html

People also ask

Where is the Danube River located?
Where does the Danube begin and end?
Which way does the Danube river flow?
Where is the Danube River born?
Several conversational interfaces

“What is the distance between Oxford and London”

London, England is about 95 kilometres from Oxford, England by car, or about 84 kilometres as the crow flies.
Why robotics?

- A robot that can make sense of the world and interact with humans is very useful: assistants to people with disabilities, robots on rescue missions, just for fun, etc.
- Spatial cognition and action represent the core of human cognition and behaviour.
- Having access to robot’ sensors and actuators can give us a theoretical insight into language, spatial perception and action.
Properties of the robot’s world

The nature of robot’s world and language:

- Partially observable (sensory data is noisy and incomplete)
- Dynamic (changes over time)
- Continuous (real valued sensory data)
- Sequential (current decisions affect future actions)
- Contains other interacting agents
- Stochastic (outcomes of actions are non-deterministic)
Learning from environment

SLAM, (Newman and Durrant-Whyte, 2001)
http://www.youtube.com/watch?v=6afrMnEmXFI
Agents with different bodies (sensors and actuators) perceive an interact with the world differently.

Consequently, they also structure the world differently: the representations they learn will be different (“embodied mind”) (Maurice Merleau-Ponty and George Lakoff).

Is human-robot communication possible at all?
Situatedness

- Human and robot are situated in the same environment which imposes identical constraints on both kinds of representations.
- They can also interact with each other: see each other, jointly attend to each other and refer to the same situations (socialness).
- Perhaps the fact that they may internally operate with different representations is not that important.
Theory of mind

(Baron-Cohen et al., 1985)

The beliefs, desires and intentions of the other
Theory of mind and common ground

Human-Human Communication

Common Ground
- Mutual knowledge and shared beliefs, shared experience

Theory of Mind
- Other’s beliefs, desires, intention, knowledge, and perspectives

Shared Intentionality
- Cooperative motive and process, joint communication goal

Slide from (Chai, 2019)
Learning from linguistic interaction

Referring as a collaborative process (Clark and Wilkes-Gibbs, 1986)

- Speakers and addressees work together in the making of a definite reference
- Speaker proposes/invites a noun phrase
- Participants iteratively repair, expand, replace the referring expression until they reach mutual agreement
- Minimise joint effort
Situated conversation #1

(Dobnik, Howes, and Kelleher, 2015; Dobnik, Howes, Demaret, and Kelleher, 2016)
P2: 123: ok, so i see a red mug directly behind the red one on your left
P2: 124: probably next to the white with “funny top” that i cant see
P1: 125: it is just behind that and to my left/your right
P1: 126: behind from my perspective
P2: 127: and the red i can’t see is it to the left of the yellow?
P1: 128: yes, as you se it its left
P2: 129: ok, i mark it, and you mark the other red
P1: 130: yup
P1: 131: and the blue ones are one on the second row from you, to the right from you
P1: 132: one slightly to my left
P1: 133: and one in front of katie in the first row
P2: 134: yes, that’s the same
P1: 135: and the yellow are on between us to your far right
P1: 136: and one quite close to the corner on your left and katie’s right?
P2: 137: yes the same
From SCARE corpus (Stoia et al., 2008), 2.txt, l.38

DG: SIL AND uh WHAT WE GOTTA DO IS MOVE THE PICTURE TO THE OTHER WALL SIL [pause]
DF: SIL WHAT’S OTHER
DG: I
DF: OPPOSITE
DG: D-
DF: [pause]
DG: I DON’T KNOW the DEFINITION OF OTHER [pause]
DF: SIL
DG: SIL UM
DF: OPPOSITE WALL
DG: SIL
DF: [pause]
DG: I WOULD [pause] SIL HOW MANY WALLS ARE THERE SIL [pause]
DF: SIL WELL IT’S A ROOM SO THERE ARE FOUR WALLS [pause]
DG: SIL WELL SIL [pause] SIL PUT IT ON THE OPPOSITE WALL SIL [pause]
DG: SIL I CAN SAY THAT SIL [pause] SIL I CAN SAY THAT NUMBER SIL [pause] SIL NO SIL [pause] SIL OH THAT’S WHERE I HAVE TO MOVE IT TO SIL [pause] SIL THAT’S WHERE I HAVE TO MOVE IT TO SIL [NOISE LAURA NO YOU CAN DESCRIBE THAT THIS BUTTON CONTROLS IT] SIL WELL THERE IS A BUTTON THAT CONTROLS IT BUT DF: SIL
DG: OH
DF: CONTROLS WHAT
DG: SIL
DF: SIL [pause]
DG: NOW I UNDERSTAND
Expressing meaning with our body I

- Gestures and emotions
- Conversational resources
  - Non-verbal cues and information
  - ... but not any kind of movement and prosody.
- Help with coordination of conversation:
  - understanding and misunderstanding
  - turn-taking
  - topic progression
  - empathy
  - sarcasm
  - attitude
  - mood
  - ...
Expressing meaning with our body II

▶ Social referencing: film of Leonardo robot (Thomaz et al., 2005; Breazeal et al., 2006)

▶ Eye-gaze and multi-party dialogue: Furhat (Skantze, 2016)
Building robotic systems

A layered approach (Kruijff et al., 2007; Zender et al., 2008)
Building robotic systems

A layered approach (Kruijff et al., 2007; Zender et al., 2008)

- Integration of (independent) processes
- Information exchange and flow
- Temporal processing
- Information fusion
- Increased abstraction of representations
Building robotic systems, II
Data collection and offline learning (Dobnik, 2009)

**Linguistic observations**
- You’re turning right slowly.
- The table is behind the chair.

**Motion observations**
- iAgv/iPlatform: speed, heading

**Laser observations**
- SLAM
- Objects named manually

**Object/Motion Instance Creator**

**Offline learning**
- Parser
- Grammar

**Weka Classifier Builder**

**Normalise values**
Building robotic systems, III

pDescriber

Motion observations
iAgv/iPlatform: speed, heading

Laser observations

Normalise

SLAM

Objects
named manually

Normalise

iCommentary
on-line object localisation

A new scene

Weka classifiers

verb?
direction?
heading?
manner?
relation?

pDescriber

Am I moving?
Describe movement

Have I stopped?
Localise objects

iSpeaker

The table is to the left of the chair.
You’re moving backward right.
Building robotic systems, IV

pDialogue

Linguistic input

Where is the chair? iVoice

Go forward slowly.

Parser

Grammar

Chat agent

Motion

Object

Q&A

A new scene

pDialogue

Weka classifiers

speed?

heading?

lo_x?

lo_y?

Lexical entries

iCommentary

on-line object localisation

SLAM

Objects named manually

Laser observations

I’m very well, thank you. The chair is behind the table. I’m moving forward. <moving>
Robot operating system (ROS)

- **ROS**: Robot Operating System
- A middleware that assists writing robotic applications on the top of OS
- Processes and information flow between them
- Portable: several robots supported, https://robots.ros.org
- Easy to program
- http://wiki.ros.org/ROS/Introduction
ROS basics

- Each task separate process (a ROS node)
- ROS nodes communicate over network directly with each other
- They either publish or subscribe to information (ROS topics/services)
  /camera/rgb/image_color of type sensor_msgs/Image
- ROS master (roscore) coordinates the communication between the nodes

(Quigley et al., 2009)
Tutorial #4: writing your ROS node

by Mehdi Ghanimifard

In this tutorial we will give you some quick guidance how to start programming with ROS.

Code on Github
KILLE: Kinect Is Learning Language

▶ A framework for situated agents that learn grounded language incrementally and online from a human tutor
▶ Language technology:
  ▶ A platform for building models of situated language learning
  ▶ Apply existing LT tools in an interactive learning framework
▶ Linguistics and psychology:
  ▶ A platform for testing theories, formalisms and applications on grounding language in interaction
▶ A proof-of-concept evaluation of tutor assisted learning

(de Graaf, 2016; Dobnik and de Graaf, 2017)
KILLE: Kinect Is Learning LanguageE

Hardware

Software

RGB frames  depth frames
Freenect
Kinect driver
Kille Core
perceptual classification
clustering and SVM
Roscore
ROSDial
bridge to OpenDial
OpenDial
ASR (speech recognition)
TTS (text to speech)
process natural language
system output
speech input
variables
acknowledgement or recognition result

frames

topic

Roscore


Recognising objects

“This is a gnome.”
Dialogue games as learning strategies

Direct instruction: objects

**U:** This is a cup.

**S:** [Object is focused on and observed.]

**S:** OK, I learned a cup.
Dialogue games as learning strategies

Direct instruction: objects

U: This is a cup.
S: [Object is focused on and observed.]
S: OK, I learned a cup.

Direct instruction: spatial relations

U: The book is to the right of the mug.
S: [Objects are focused on and observed.]
S: OK. This relation is called “to the right of”.
S: [Objects are focused on and observed.]
U: Where are the objects?
S: The book is to the right of the mug.
Dialogue games as learning strategies

Reinforcement from feedback

U: What is this?
S: [Object is focused on and observed.]
S: This is a cup.
U: That is correct.
S: [Observation is added to the models.]
Dialogue games as learning strategies

Reinforcement from feedback

U: What is this?
S: [Object is focused on and observed.]
S: This is a cup.
U: That is correct.
S: [Observation is added to the models.]

Self-initiative: spatial relations

S: [Objects are focused on and observed.]
S: One of the objects is book, the other one is box. Please tell me where things are.
U: The book is to the left of the box.
S: OK, this relation is called “to the left of”.
Demo

https://www.youtube.com/watch?v=AsEgaka6tH0
Summary

- Challenges for working with robots in a dynamic world
- A brief overview of language in interaction
- Architectures of robotic systems and information fusion
- ROS (Robot Operating System) as a software platform for building interactive systems
- KILLe as a framework for learning grounded language incrementally and online from a human tutor


References IV


