A System for Interactive Learning in Dialogue with a Tutor

Slides by Samir

George (*one on the right side in the image*)

- Ability to learn, communicate and act autonomously (*cognitive system*)
- Learns from visual conceptual models of colours and two basic shapes (*compact and elongated*)
- Learning
  - *Tutor-driven learning*
- Communicating
  - *Situated tutor-assisted learning*
- Autonomously
  - *Non-situated tutor-assisted learning*

Introduction

- *Objective* - Design a cognitive system that acquire visual concepts through interacting with human
- *Property of Cognitive system* - Interactive continuous learning from information obtain from vision and language
- *Previous work*
    - System architecture and integration
    - Learning
    - Social interaction
- *Focus of this work*
    - Forming beliefs about the state of the world
        - Integration of visual perception
        - Processing of linguistics information
    - Used beliefs for learning process and updating the current representations

System Competencies and Representations

- Vision
- Visual learning and recognition
- Beliefs
- Situated dialogue
- Behaviour generation
- The integrated system

Vision

- System uses a generic bottom-up 3D attention mechanism
- Objects are place on a table or any other supporting surface to make the problem of generic segmentation of the unknown objects tractable
- System detect supporting planes using a variant of particle swarm optimisation
- Spaces of interest (*SOIs*)
- Segmentation mask to learn colour and shape

Visual learning and recognition

- Visual concepts are represented by generative models
- Form of probability density function (*pdf*) over the feature space and are constructed in an online fashion from new observations
- Continuous learning
  - Extracting the visual data in the form of multi-dimensional features
  - *OdKDE* to estimate the pdf in the multi-dimensional feature space

Beliefs

- Unit of information describing an entity
- Constrains spatio-temporally and epistemically
- Spatio-temporally

  Stating where and when the described entity is assumed and exist

- Epistemically

  Stating for which agent(s) the information contained in the belief holds

- Epistemic status of beliefs

  1. *Private beliefs*
  2. *Attributed beliefs*
  3. *Shared beliefs*

Situated Dialogue

- Intention recognition problem
- Main components in charge of robot's language competence

1. *Language understanding* - the process of recognising the intention behind the human utterance

2. *Dialogue management* - given a context update, DM selects which action to perform, robot's intention to act

3. *Language production* - process of realising the robot's intention given the situated context

Behaviour generation

- **AI Planning**

    1. *Goal generation and management*

        - Generating goals from the results of sensing and internal processing, and selection which of this to pass to the planning
        - This is done based on the potential information gain and associated cost

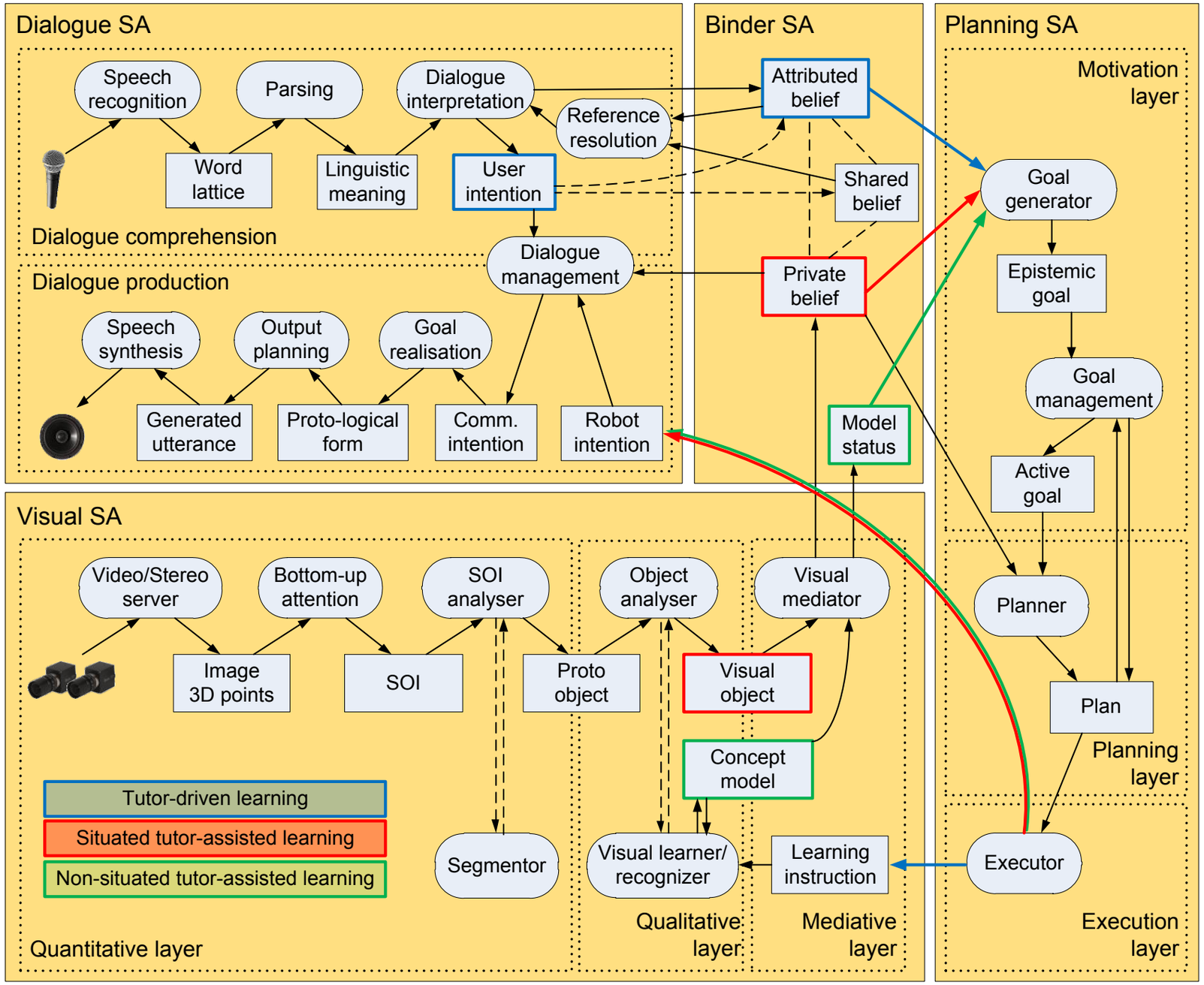
    2. *Planning*

        - Classical planner - finding sequences of actions turning the initial state into a state satisfying the goal formula given an initial state, a set of actions and a goal formula
        - Continue planner

    3. *Execution*

        - A set of mediator components that trigger other components when a plan requires it

The integrated system



Learning Mechanism

- Tutor driven learning(TD)
- Situated tutor-assisted learning(TA)
- Non-situated tutor-assisted learning(NS)
- Sample Dialogue

H: Hi, Robot.
R: Hi, there.
H places an unknown object in front of the robot:
Do you know what this is?
R: I don't know.

H: This is an elongated object.
R: Let me see.
R verifies the assertion and modifies its model: OK. TD

R: What colour is this object?
H: It is blue.
R: OK. TA

H: Is this object red?
R: No, it is blue.
H places a green ball on the table:
Could you tell me what shape it is?
R: It is compact.
H: Is the compact object green?
R: Probably yes.

H places another object on the table.
R: Is this object red?
H: No. It is blue.
R: OK. TA

H removes the objects from the table.
R: Would you show me something yellow?
H places a banana in front of the robot:
This is yellow.
R: I see. Is it also elongated?
H: Yes, it is.
OK. NS

H places several objects on the table: What colour is the compact object?
R: It is orange.
H: Thanks.
R: Good bye.

Experiment results

- **System evaluation**
  - Dataset
    - 1120 images and 129 objects
  - Training and testing
    - 500 pairs of images for training and remaining for testing
  - Performance of the model was evaluated in terms of recognition rate
  - Average recognition rate on each step was taken from 3 runs

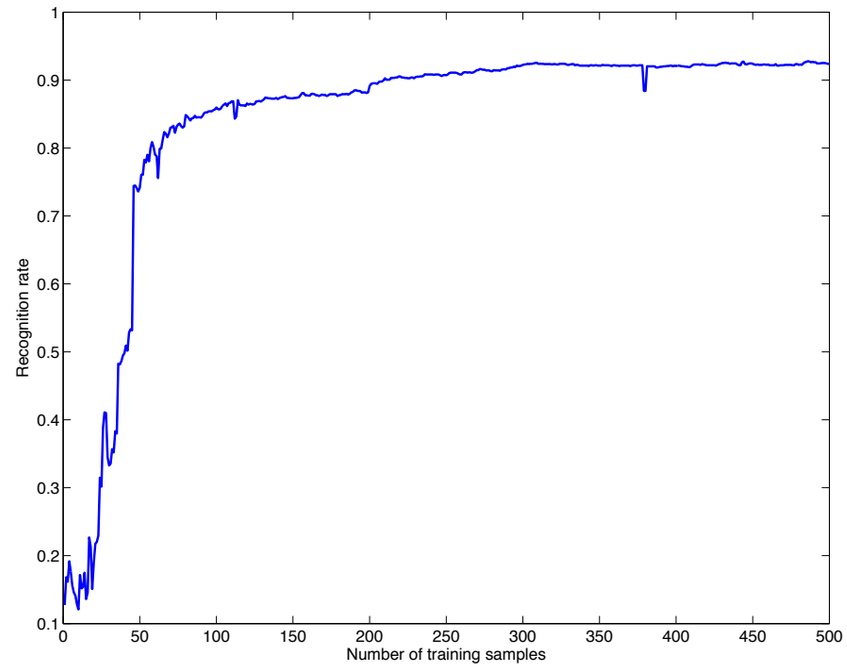


Fig. 7. System evaluation - recognition rate.

Experimental results

- **Evaluation of learning mechanism**
  - Performance on all three learning mechanism
    - Tutor driven (TD)
      - Variant 1 - (TD_{rnd}) training images were randomly chosen
      - Variant 2 - (TD_{seq}) Model were first initialise with 5 images from every class and then the object was presented in a sequence by presenting all objects of the first class, then the second and so on
    - Situated tutor assisted (TA)
    - Non-situated tutor assisted (NS)

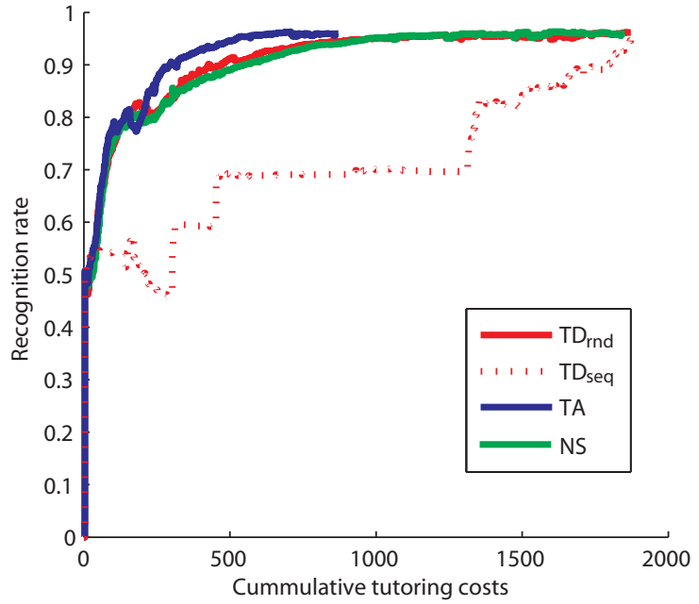


Fig. 8. Evaluation of different learning strategies.

Conclusion

- Continuous learning of visual concepts in dialogue with a tutor
- How beliefs about the world are created by processing visual and linguistics information
- How this beliefs are used for planning the system behaviour to extend the knowledge
- Robotic implementation is based on a distributed asynchronous architecture