A Multi-layer LSTM-based Approach for Robot Command Interaction Modeling

Martino Mensio[1], Emanuele Bastianelli[1], Ilaria Tiddi[2], Giuseppe Rizzo[3]

([1]Knowledge Media Institute, The Open Univerisity, UK. [2]Department of Computer Science, VU University Amsterdam, NL. [3]Istituto Superiore Mario Boella, Italy)

Human-Robot interaction

- indoor environment
- natural language commands stating:
  - action → **semantic frame**
  - argument → **frame element**

Challenges:

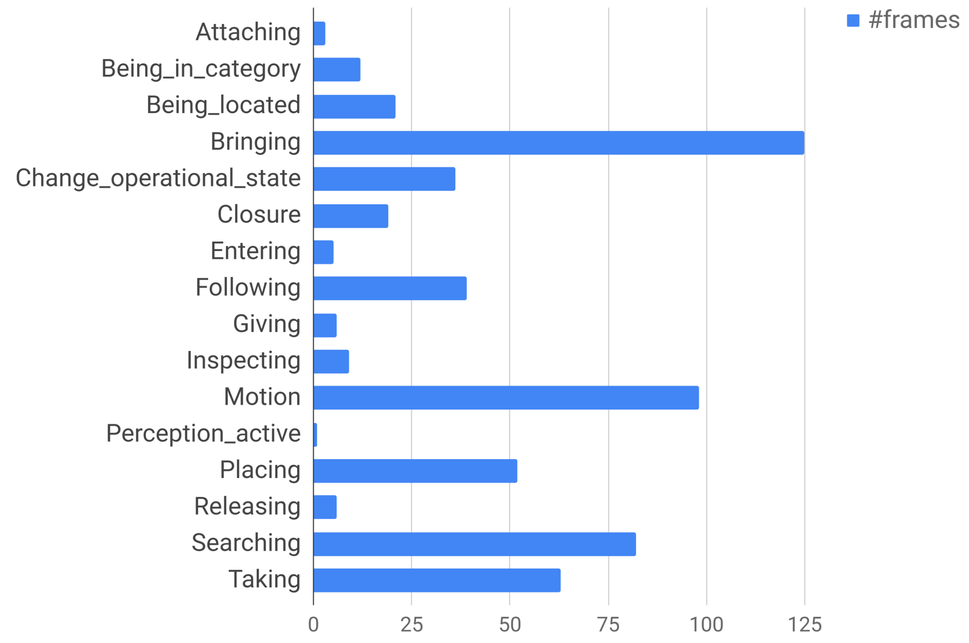
- **understand** the sentence of the user
- **instantiate** the corresponding command

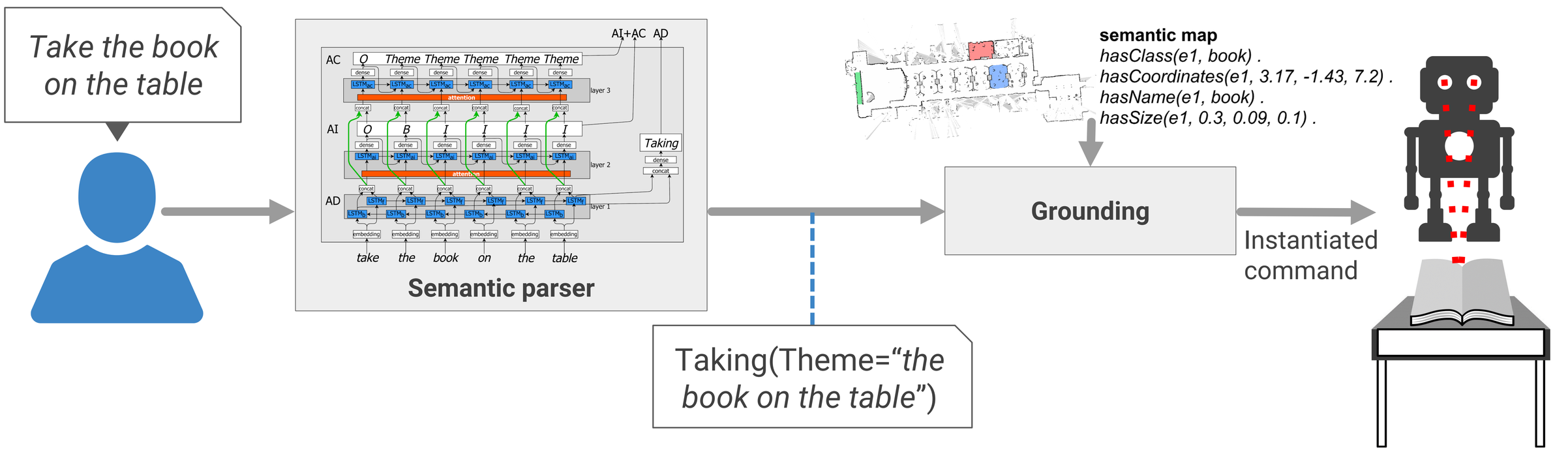
Preliminary studies

Dataset

HuRIC: FrameNet-annotated spoken commands

Distribution of samples over the frames





The LSTM neural network

Sequences of words processing using LSTM

- the word embeddings (GloVe) are taken for each word
- a bidirectional LSTM encoder provides contextualized representations of words
- other layers (LSTM decoders, dense, self-attention) specialize on the three different tasks

Grounding

Mapping words in a command to real-world entities:

1) A semantic map stores information about entities in the environment

2) Linking words to the entities' lexical information in the map:
   - semantic heads are extracted from each frame element
   - grounding through a Lexicalised Grounding Mechanism (LGM)

Semantic parsing results

Using 5-fold evaluation, each task is evaluated:

- AD: Action detection (identification of the Frame)
- AI: Argument Identification (detecting the spans of the arguments)
- AC: Argument Classification (assign a label to each argument)

F1	AD	AI	AC
BAS16	94.67%	90.74%	94.93%
3L-ATT	94.44%	94.73%	94.69%
3L-NO-ATT	95.37%	94.90%	91.90%
2L-ATT	96.29%	94.40%	92.30%
2L-NO-ATT	94.44%	94.50%	92.45%

Whole chain results

The results of the parser are fed into the grounding module and the performances are evaluated.

	Whole chain F1
BAS16	41.70%
3L-ATT	43.67%
3L-NO-ATT	41.92%
2L-ATT	44.54%
2L-NO-ATT	42.79%

Future works

- Study the values of attention towards explainability
- Design an interactive scenario for correction

References

B. Liu and I. Lane, "Attention-based recurrent neural network models for joint intent detection and slot filling," in INTERSPEECH. ISCA, 2016, pp. 685–689.

E. Bastianelli, G. Castellucci, D. Croce, L. Iocchi, R. Basili, and D. Nardi, "Huric: a human robot interaction corpus," in Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14). Reykjavik, Iceland: European Language Resources Association (ELRA), 2014.