Spoken Utterances Guiding Chef's Assistant Robots SUGAR

Maria Di Maro, Antonio Origlia, Francesco Cutugno
Università degli Studi di Napoli 'Federico II'

Outline to the task

- SUGAR at Evalita 2018
  - Introduction to the task
  - Corpus Collection
  - Corpus Annotation
- Participants
- Evaluation
  - Metrics
  - Results
- Conclusion

SUGAR at EVALITA 2018

Introduction to the task

- Spoken Language Understanding:
  - Grammars (McGlashan et al., 1992)
  - Frame semantics (Wang, 2010)
  - Bag of words (Yao et al., 2013)
  - Semantic-syntactic trees (Miller et al., 1966)
  - Intent classification (Tur and Deng, 2011)

SUGAR at EVALITA 2018

Introduction to the task

- Intent Classification-based task to train systems to understand spoken commands
    - Authentic spoken data collected in a simulated natural context
    - Manual Annotations for training purposes
    - Automatic extraction of semantic predicates to enable a (robotic) system to perform an action (cooking context)
- Development of a suitable baseline

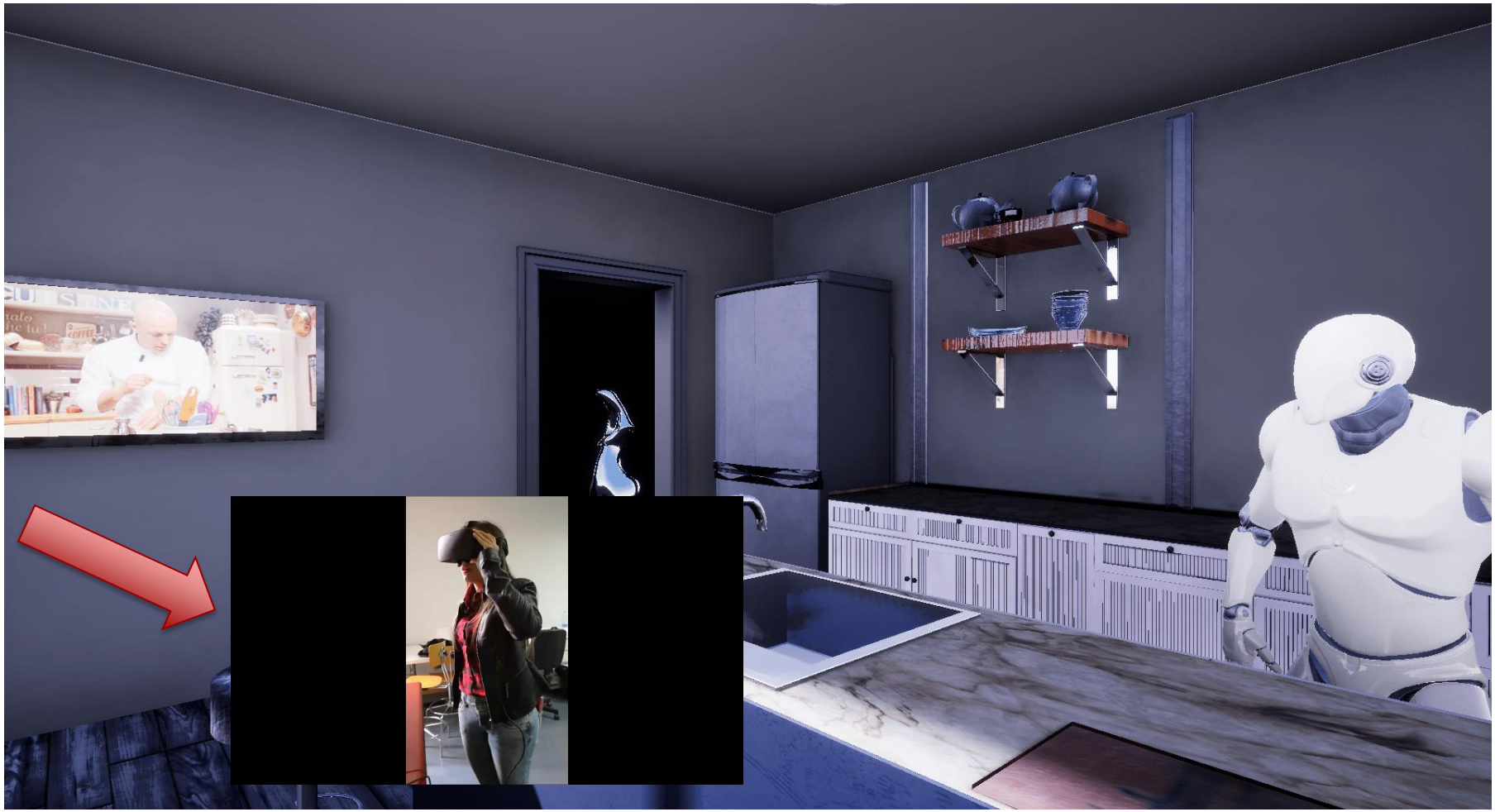
SUGAR at EVALITA 2018

Corpus Collection

- 3D virtual kitchen with Bastian “the interactive chef”
- Silent video frames showing actions
- Recording of user utterances in accomplishing recipes

SUGAR at EVALITA 2018

Corpus Collection



SUGAR at EVALITA 2018

Corpus Annotation

- Generic predicates + open domain-dependent set of parameters

  put(pot, fire)
  put(egg, bowl)

- Training set: 1721 actions
- Test set: 572 actions

- Transcriptions were not provided
- Data could be extended with external ontologies

SUGAR at EVALITA 2018

Corpus Annotation: Action Templates

Predicate	Arguments
prendere	quantità, [ingredienti]/recipiente
aprire	quantità, [ingredienti], recipiente
mettere	quantità, utensile/[ingredienti], elettrodomestico, modalità
sbucciare	quantità, [ingredienti], utensile
schiacciare	[ingredienti, utensile
passare	[ingredienti], utensile
grattare	[ingredienti], utensile
girare	[ingredienti], utensile
togliere	utensile/prodotto, elettrodomestico
aggiungere	quantità, [ingredienti], utensile/recipiente/ elettrodomestico/[ingredienti], modalità
mescolare	[ingredienti], utensile, modalità
impastare	[ingredienti]
separare	parte/[ingredienti],ingrediente/utensile
coprire	recipiente/[ingredienti], strumento
scoprire	recipiente/[ingredienti]
controllare	temperatura, ingrediente
cuocere	quantità, [ingredienti], utensile, modalità

SUGAR at EVALITA 2018

Training set - Example

```
1; prendere(uovo)
2; aprire(uovo, ciotola)
3; aggiungere(30 g, latte, ciotola)
4; aggiungere(pizzico, sale, *ciotola*)
5; mettere(pentola, fuoco)
6; mescolare(uova)
7; aggiungere(filo, olio, pentola)
8; aggiungere(uova, pentola)
9; girare(frittata)
10; togliere(padella, fuoco)
```

SUGAR at EVALITA 2018

Corpus Annotation

Challenging Scenarios:

- Implicit arguments
- Co-reference
- Arguments expressed within the verb (i.e. instrumental verbs)
- Generalization of actions (i.e. *sciogliere(lievito, acqua) => mescolare([lievito, acqua]))*

Participants

FBK-HLT-NLP	Anonymous System
• System 1 Encoder-Decoder approach with memory of previous sentences • System 2 Sequence to sequence modelling with synthetic data generation	Deep Neural Network System: • Word embeddings lexicon trained on a corpus of recipes (4.5 million words) as features • Two Bi-LSTM layers for the encoder (the first for the token sequences, the second to embed arguments in a vector) • Two Bi-LSTM layers for the decoder (the first in charge of decoding the sequence of arguments and the second of decoding the sequence of tokens) • Multi-task neural network to classify the actions, detect the implicitness and predict the arguments.

Evaluation

Metrics

- The proposed system correctly detects the requested action and all its parameters
- The proposed system asks for repetition
- The proposed system correctly detects the requested action but it assigns wrong parameters
- The proposed system misses the action

Evaluation

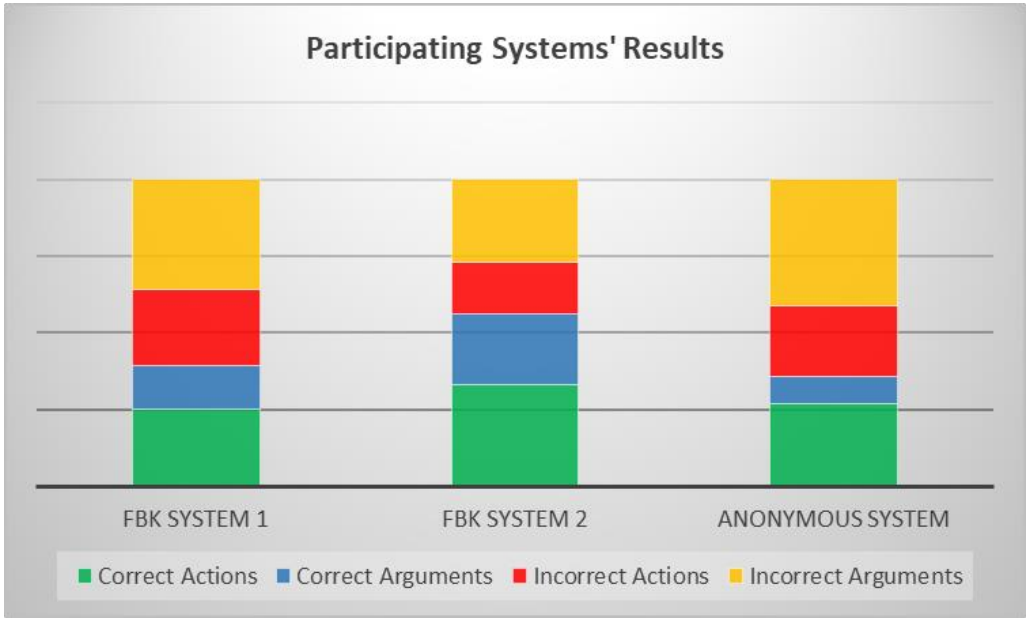
Metrics: Output

- Action id (listing number of predicate + number of action)
- Boolean Value indicating if the predicate has been recognized
- Number of expected arguments
- Distance between expected arguments and system outputted arguments (Levensthein distance)
- Number of arguments for which the system asked for repetition

Reference File: prendere(500 g, panna)
Output File: prendere(500 g, latte)
Evaluation Output: 1_1 (1, 2, **1**, 0)

Evaluation

Results



	Correct Actions	Correct Arguments	Incorrect Actions	Incorrect Arguments
FBK System 1	50,16	28,31	49,83	71.68
FBK System 2	**66,36**	**46,22**	33,64	53,78
Anonymous System	55,89	17,46	46,11	82,54

Conclusion

- Further analysis should be carried out to efficiently solve semantic recognition tasks
- Errors analysis is needed
- Enlargement of the corpus
- Rule-based module?
- Multilingualism and Multimodality


Thank you!

Spoken Utterances Guiding Chef's Assistant Robots SUGAR

Maria Di Maro, Antonio Origlia, Francesco Cutugno
Università degli Studi di Napoli 'Federico II'