Dialogue State Tracking with Convolutional Semantic Taggers

Mandy Korpusik, Jim Glass

MIT Computer Science and Artificial Intelligence Laboratory

Cambridge, MA, USA

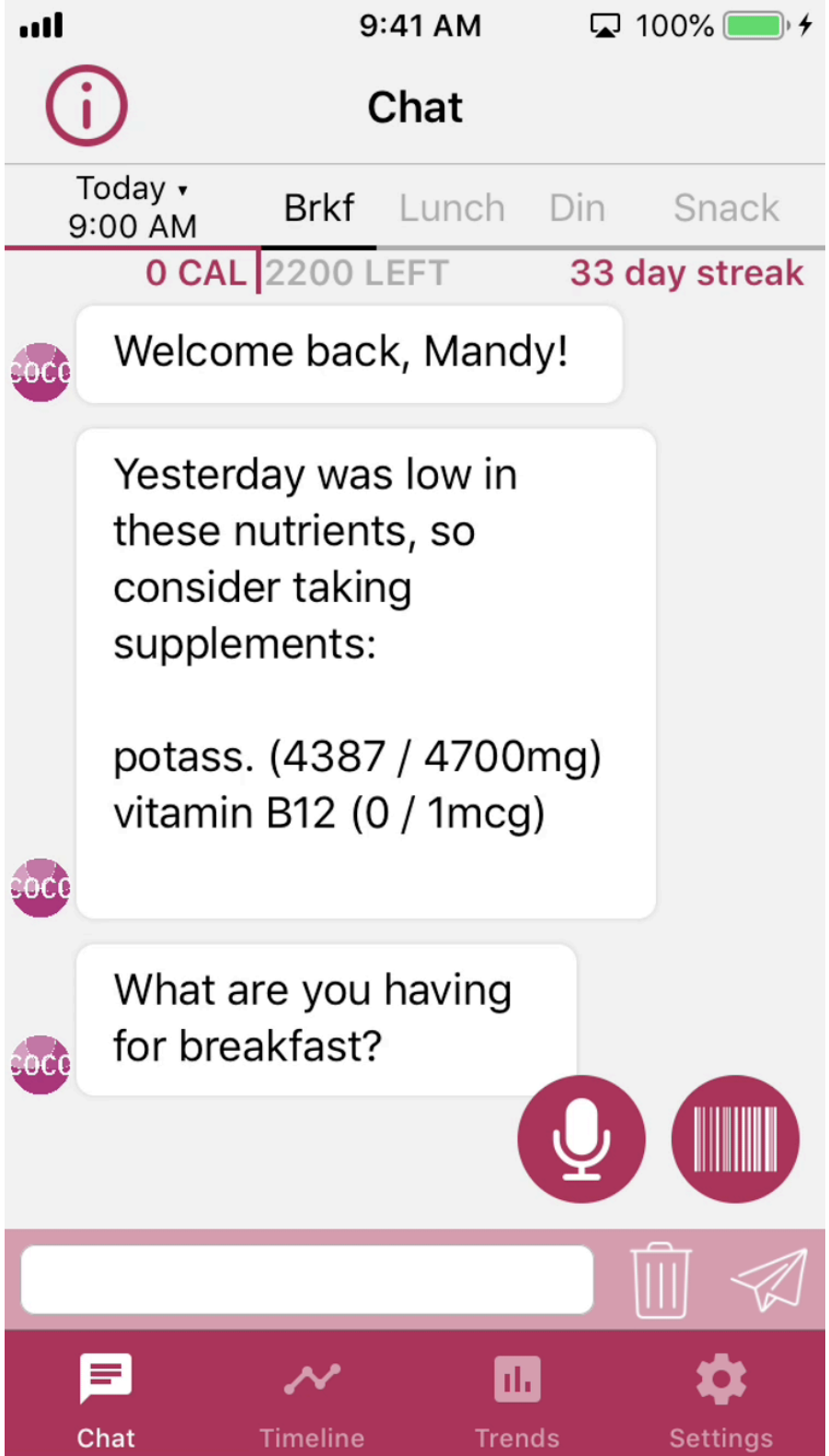
ICASSP, Brighton, UK

May 16, 2019

Motivation: Spoken Diet Tracking

the conversational calorie counter

Coco Nutritionist **lets you record what you ate with everyday spoken natural language.**


9:41 AM
100%
Chat
Today
9:00 AM
Brkf
Lunch
Din
Snack
0 CAL
2200 LEFT
33 day streak
Welcome back, Mandy!
Yesterday was low in these nutrients, so consider taking supplements:
potass. (4387 / 4700mg)
vitamin B12 (0 / 1mcg)
What are you having for breakfast?
Chat
Timeline
Trends
Settings

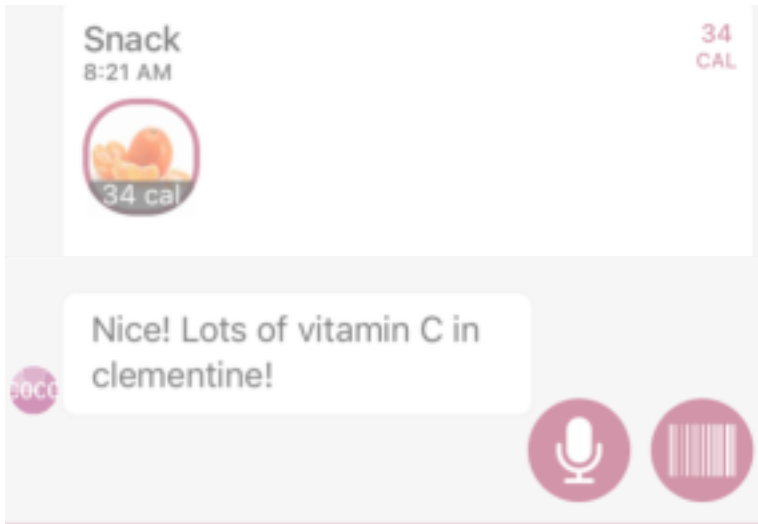
Motivation: Nutrition Multi-turn Dialogue

Nutrition question answering

- *is grilled chicken or red meat better?*
- *What should I eat for dinner?*
- *What is a healthy breakfast*
- *Which cereal is best to keep you satisfied?*
- *How many calories in ## of food item*
- *Is milk healthy? …*

Personalized food recommendation

(Korpusik et al., CBRecSys, 2016)



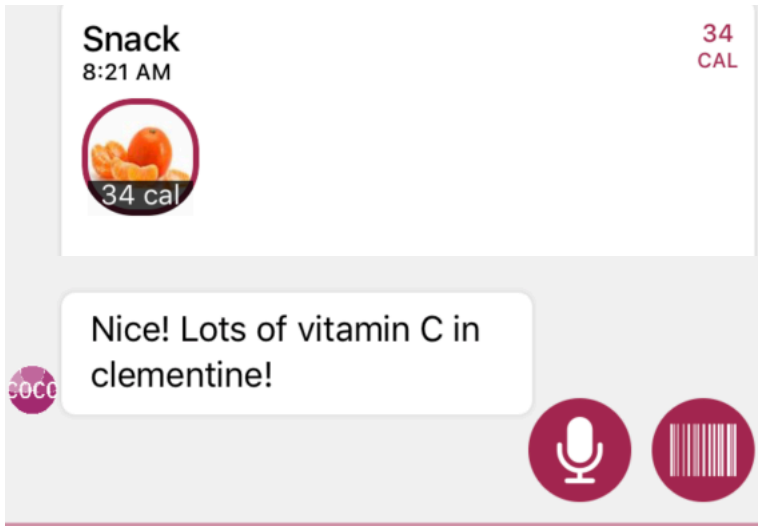
Motivation: Nutrition Multi-turn Dialogue

Nutrition question answering

- *is grilled chicken or red meat better?*
- *What should I eat for dinner?*
- *What is a healthy breakfast*
- *Which cereal is best to keep you satisfied?*
- *How many calories in ## of food item*
- *Is milk healthy? …*

Personalized food recommendation

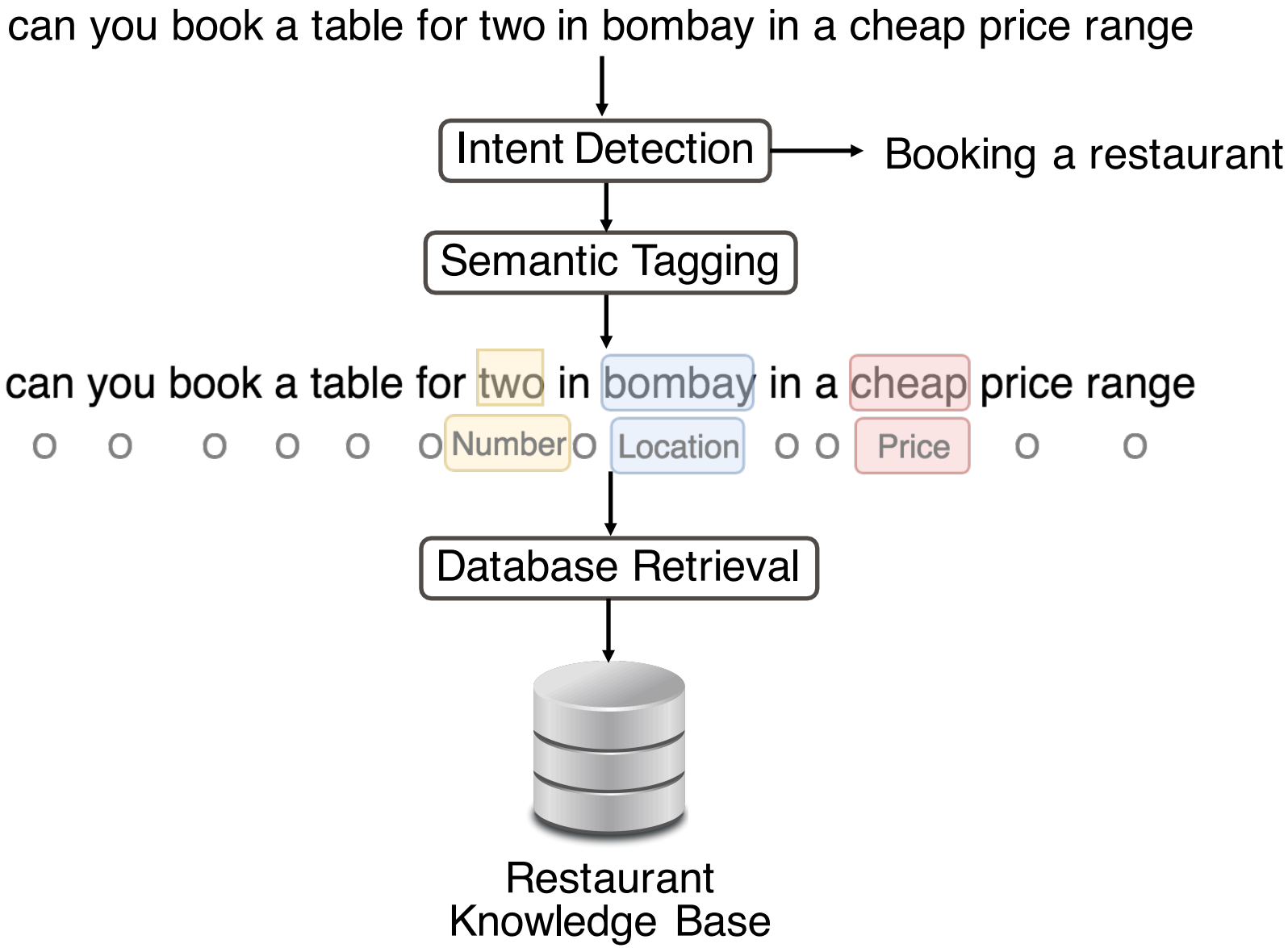
(Korpusik et al., CBRecSys, 2016)



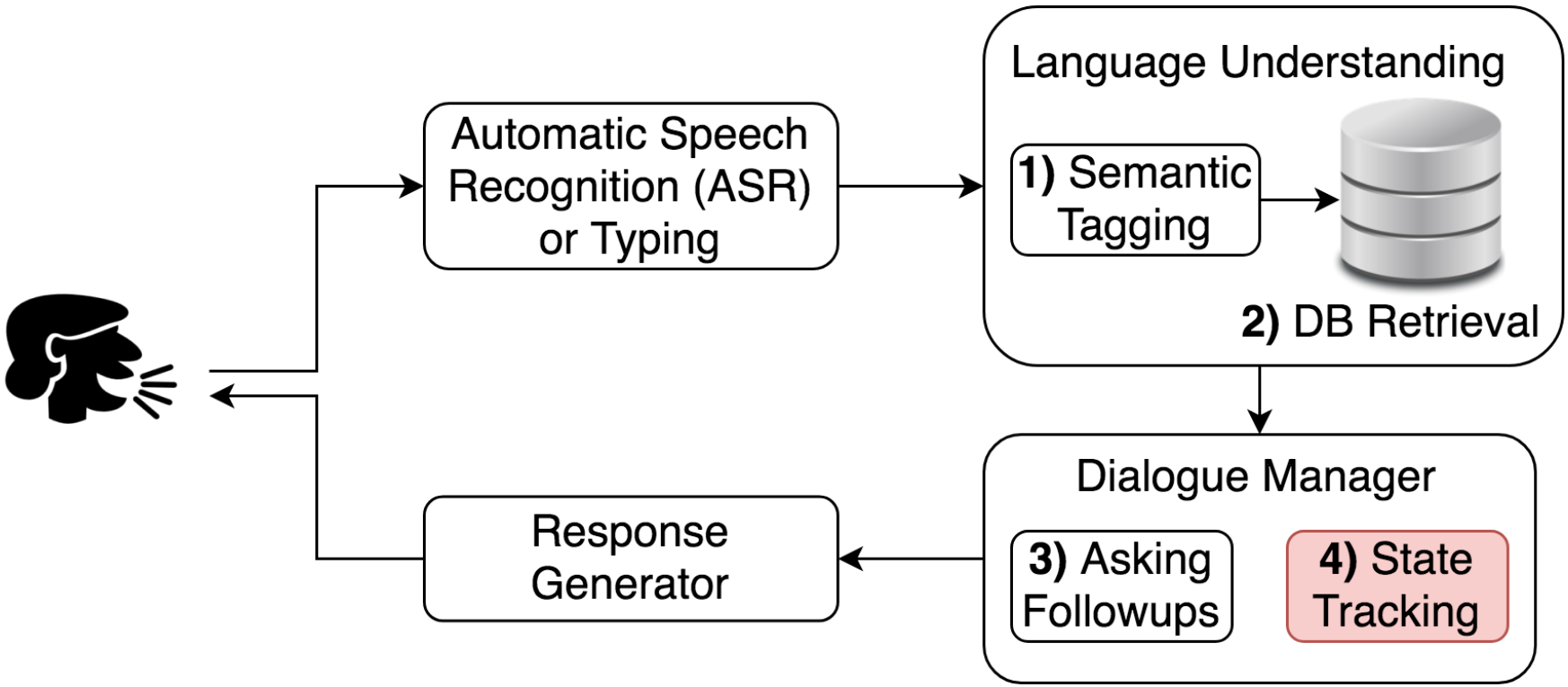
Overview

- Motivation: Nutrition
- **Introduction**
- **Our work in 3 state tracking challenges:**
  - DSTC7
  - **DSTC6**
  - DSTC2
- **Conclusion**

Introduction: Spoken Dialogue Systems



Introduction: Spoken Dialogue Systems



Our Goal: Develop ability to do dialogue state tracking.

Introduction: Dialogue State Tracking

can you book a table for two in bombay in a cheap price range

O O O O O O Number O Location O O Price O O

Dialogue State

```
cuisine: None
location: bombay
number: 2
price: cheap
atmosphere: None
```

System: any preference on a type of cuisine

with indian cuisine

O Cuisine O

Dialogue State

```
cuisine: indian
location: bombay
number: 2
price: cheap
atmosphere: None
```

Dialogue State Tracking Challenges (DSTC)

- **DSTC1 (2013): human-computer bus timetables**
- **DSTC2 and 3 (2014): human-computer restaurant info**
- **DSTC4 (2015): human-human tourist info**
- **DSTC5 (2016): multilingual tourist info**
- **DSTC6 (2017): 3 tracks, end-to-end learning**
- **DSTC7 (2019): 3 tracks (response selection, generation, and audio-visual)**

Dialogue State Tracking Challenges (DSTC)

- **DSTC1 (2013): human-computer bus timetables**
- **DSTC2 and 3 (2014): human-computer restaurant info**
- **DSTC4 (2015): human-human tourist info**
- **DSTC5 (2016): multilingual tourist info**
- **DSTC6 (2017): 3 tracks, end-to-end learning**
- **DSTC7 (2019): 3 tracks (response selection, generation, and audio-visual)**

(Korpusik et al., DSTC7 2019)

DSTC7

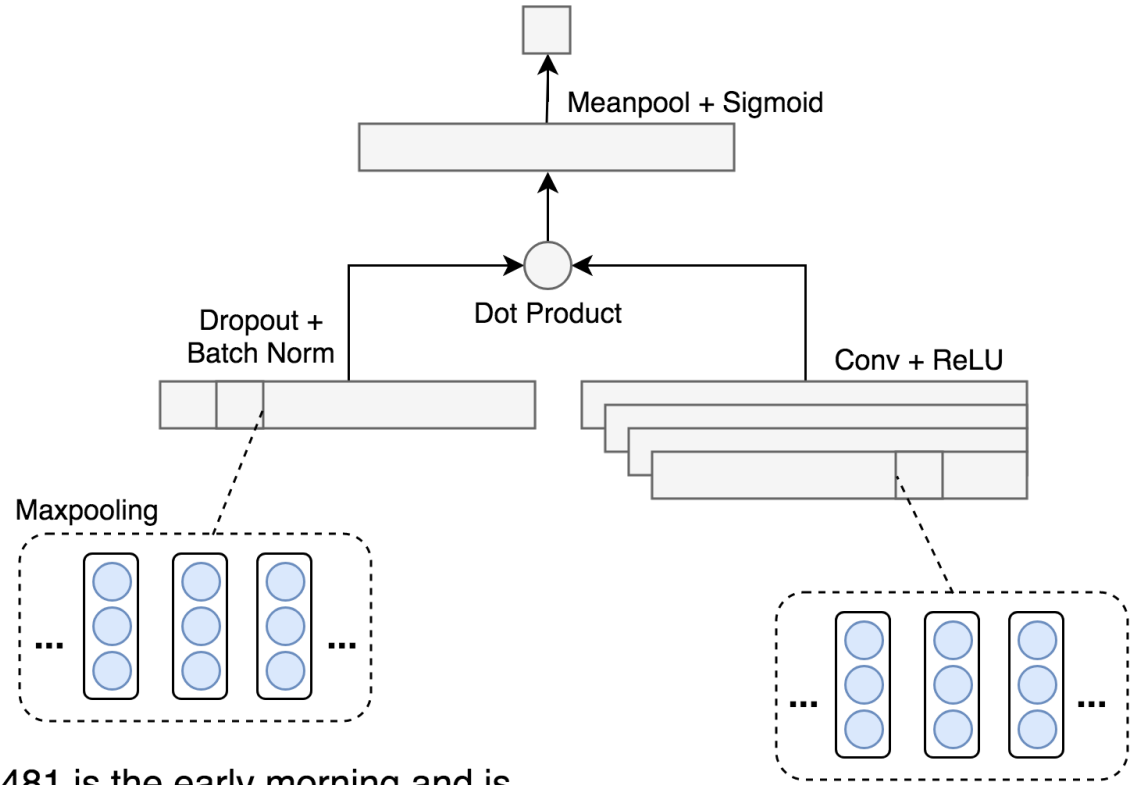
Student-Advisor Partial Dialogue:

ADVISOR | Hi! What can I help you with?
STUDENT | Hello! I'm trying to schedule classes for next semester. Can you help me?
STUDENT | Hardware has been an interest of mine.
STUDENT | But I don't want too hard of classes
ADVISOR | So are you interested in pursuing Electrical or Computer Engineering?
STUDENT | I'm undecided
STUDENT | I enjoy programming but enjoy hardware a little more.
ADVISOR | Computer Engineering consists of both programming and hardware.
ADVISOR | I think it will be a great fit for you.
STUDENT | Awesome, I think that's some good advice.
STUDENT | What classes should I take to become a Computer Engineer?
ADVISOR | You haven't taken EECS 203, 280, and 270, so it may be in your best interest to take one or two of those classes next semester
STUDENT | Ok. Which of those is in the morning. I like morning classes

(Korpusik et al., DSTC7 2019)

DSTC7: Convolutional Neural Encoder

Binary Verification: 1 (Match) / 0 (Not)



0 0 0 ... 481 is the early morning and is quite similar to EECS381, so you might want to skip it.

Candidate System Response

0 0 0 ... What time does the course occur? I like afternoon classes and will find something else if it's scheduled too early.

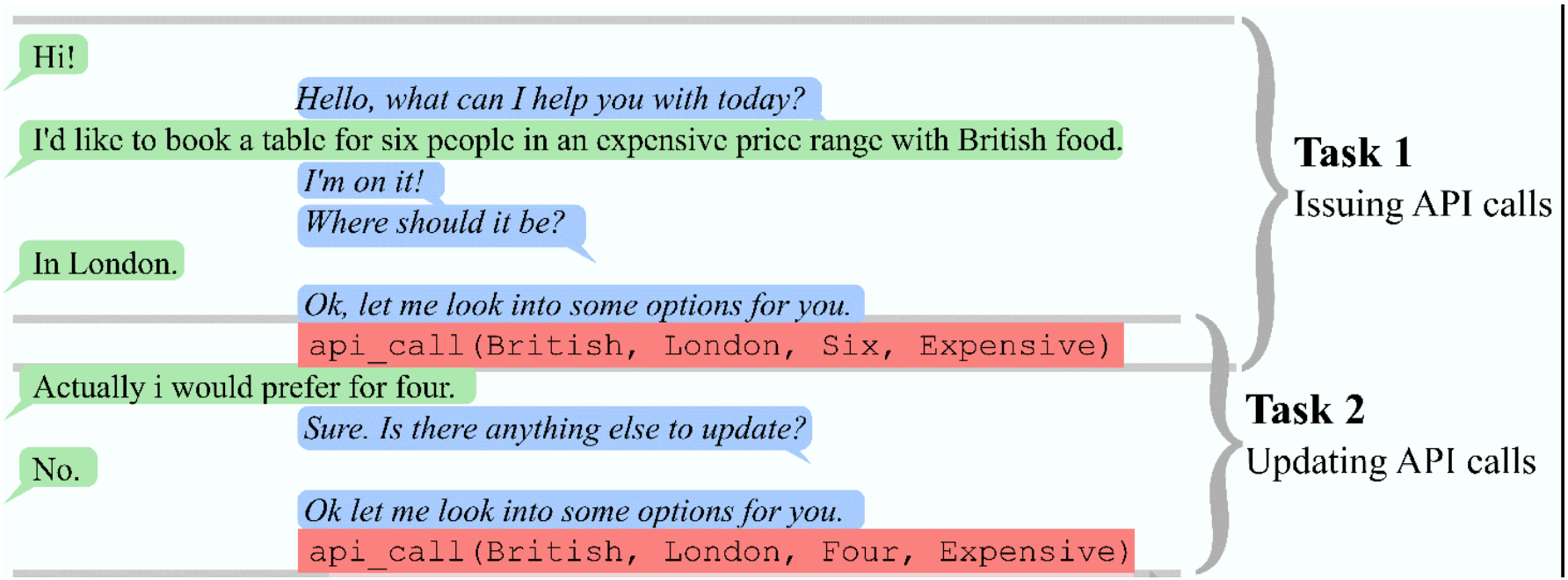
EECS351 is after lunch. The others are before. EECS481 is from nine to ten thirty and EECS 492 is from ten thirty to twelve.

Previous Utterances

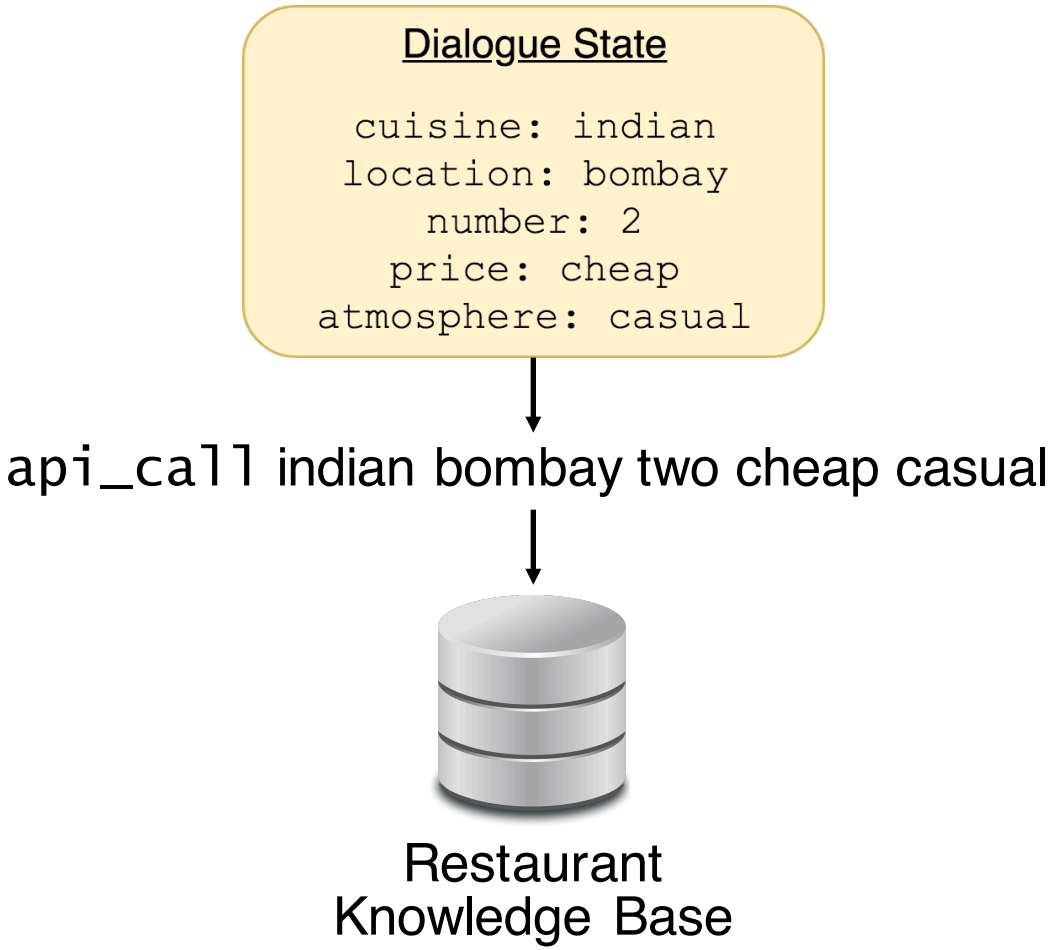
(Bordes & Weston, 2016)

DSTC6

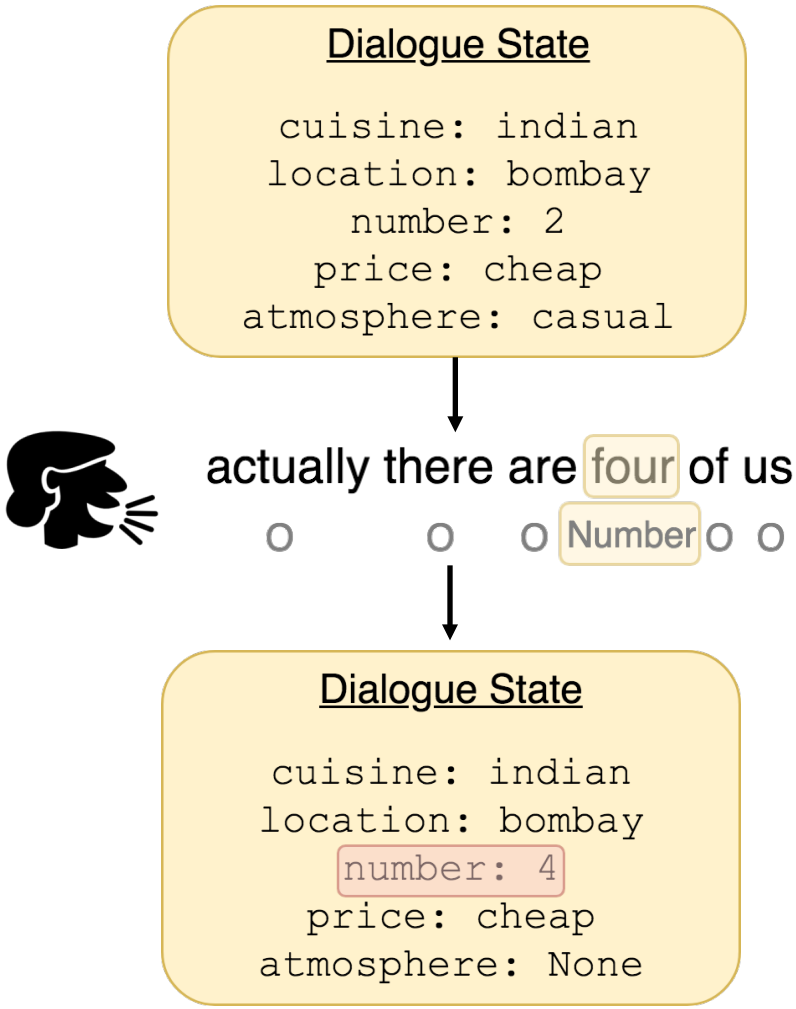
Goal: select the best system response.



Task 1: API Call



Task 2: Updating API Call

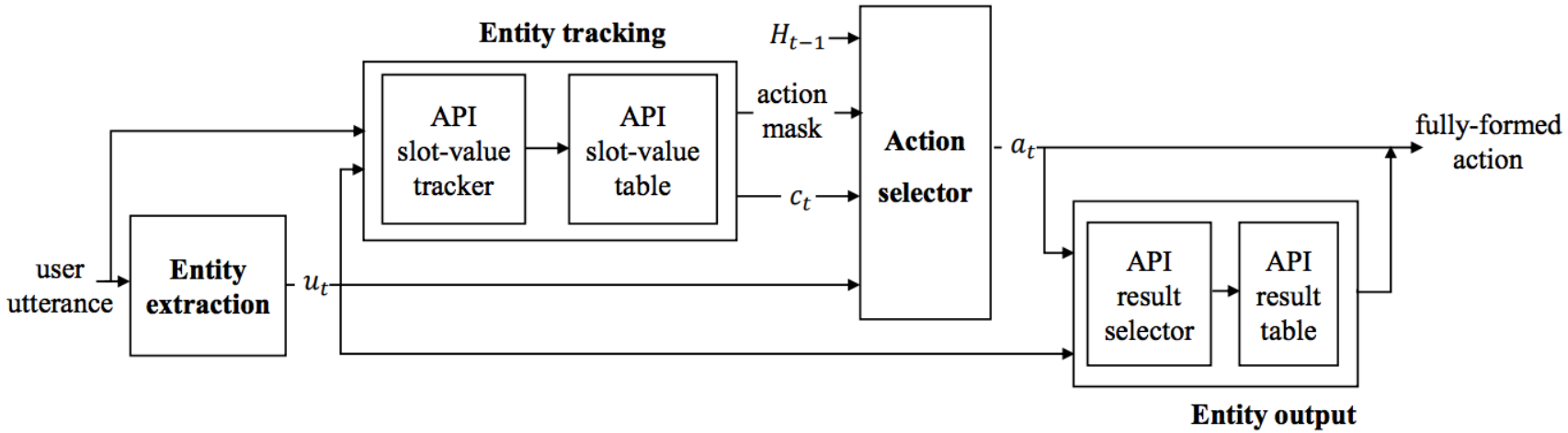


DSTC6 Data

- **10,000 simulated training dialogues per task**
- **KB of restaurants**
  - 10 cuisines
  - 10 locations
  - 3 price ranges
  - 4 party sizes

DSTC6: Related Work

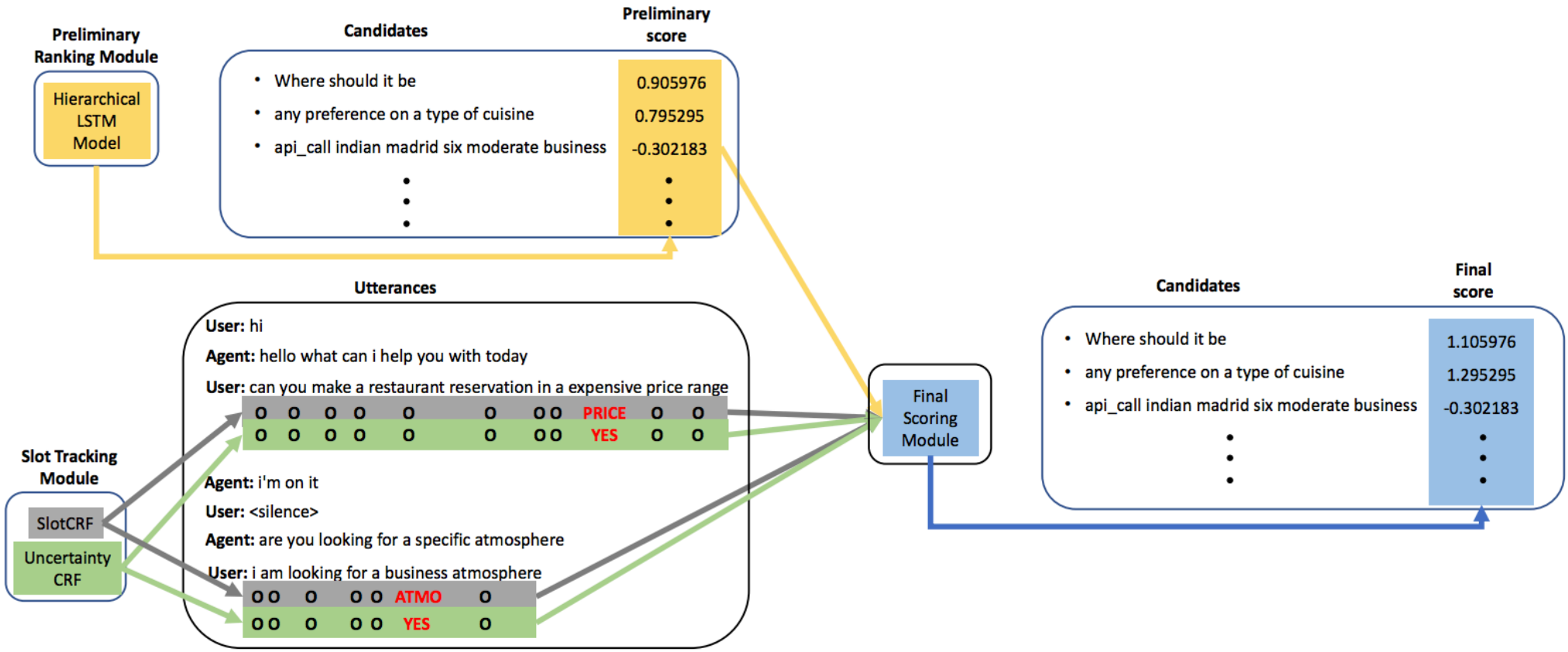
- **2 challenge participants achieved 100% on all tasks:**
  - Extended Hybrid Code Networks for DSTC6 (*Ham et al., 2017*)



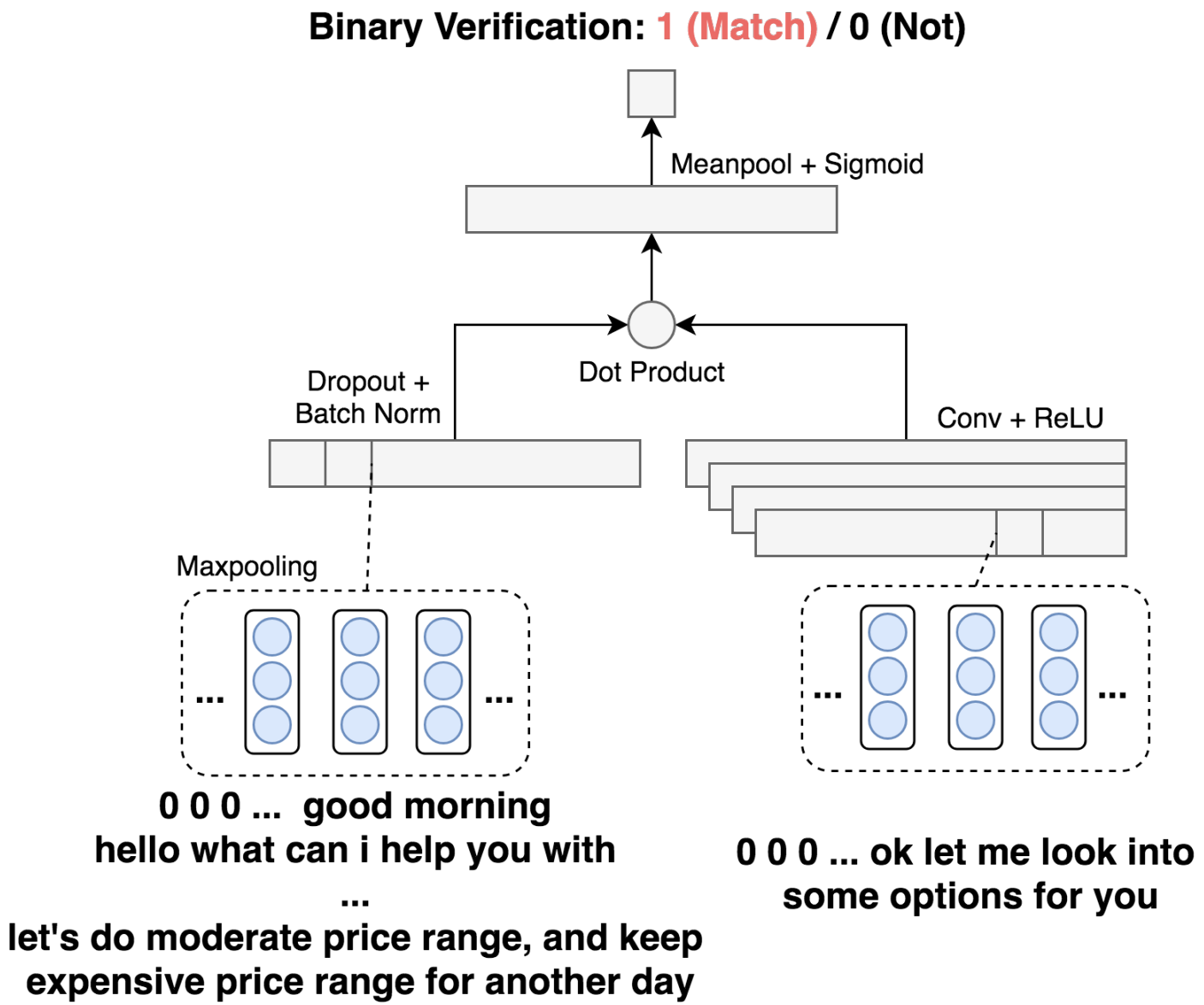
  - Modeling Conversations to Learn Responding Policies (*Bai et al., 2017*)

DSTC6: Related Work

- **2 challenge participants achieved 100% on all tasks:**
  - Extended Hybrid Code Networks for DSTC6 (*Ham et al., 2017*)
  - Modeling Conversations to Learn Responding Policies (*Bai et al., 2017*)



DSTC6: Our Binary CNN Baseline



DSTC6: Our Full CNN Architecture

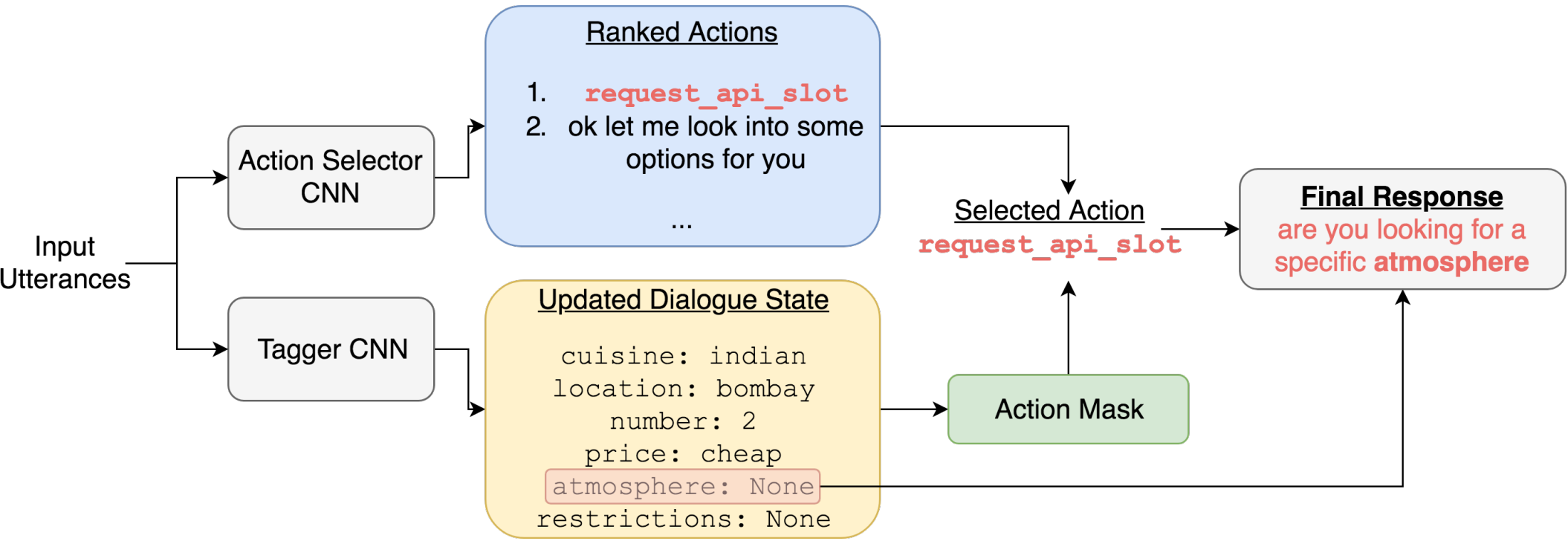
Approach:

1. Select action template with CNN.
2. Populate action template with CNN-predicted semantic tags.

DSTC6: Our Full CNN Architecture

Approach:

1. Select action template with CNN.
2. Populate action template with CNN-predicted semantic tags.



Step 1: Semantic Tagging

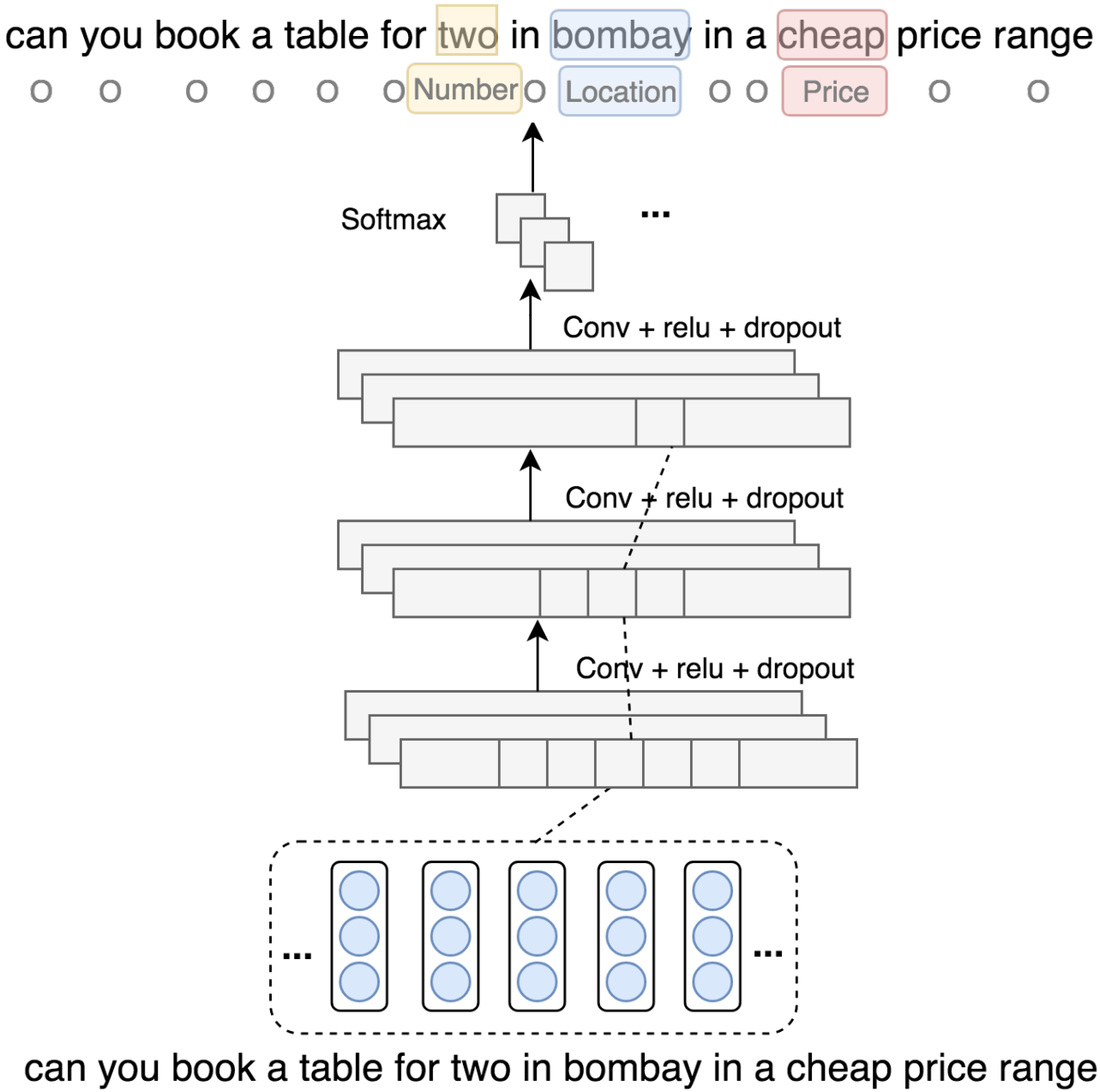
(Korpusik et al., SLT 2014)
(Korpusik et al., ICASSP 2016)
(Korpusik et al., Trans. ASLP 2017)

can	you	book	a	table	for	two	in	bombay	in	a	cheap	price	range
O	O	O	O	O	O	Number	O	Location	O	O	Price	O	O

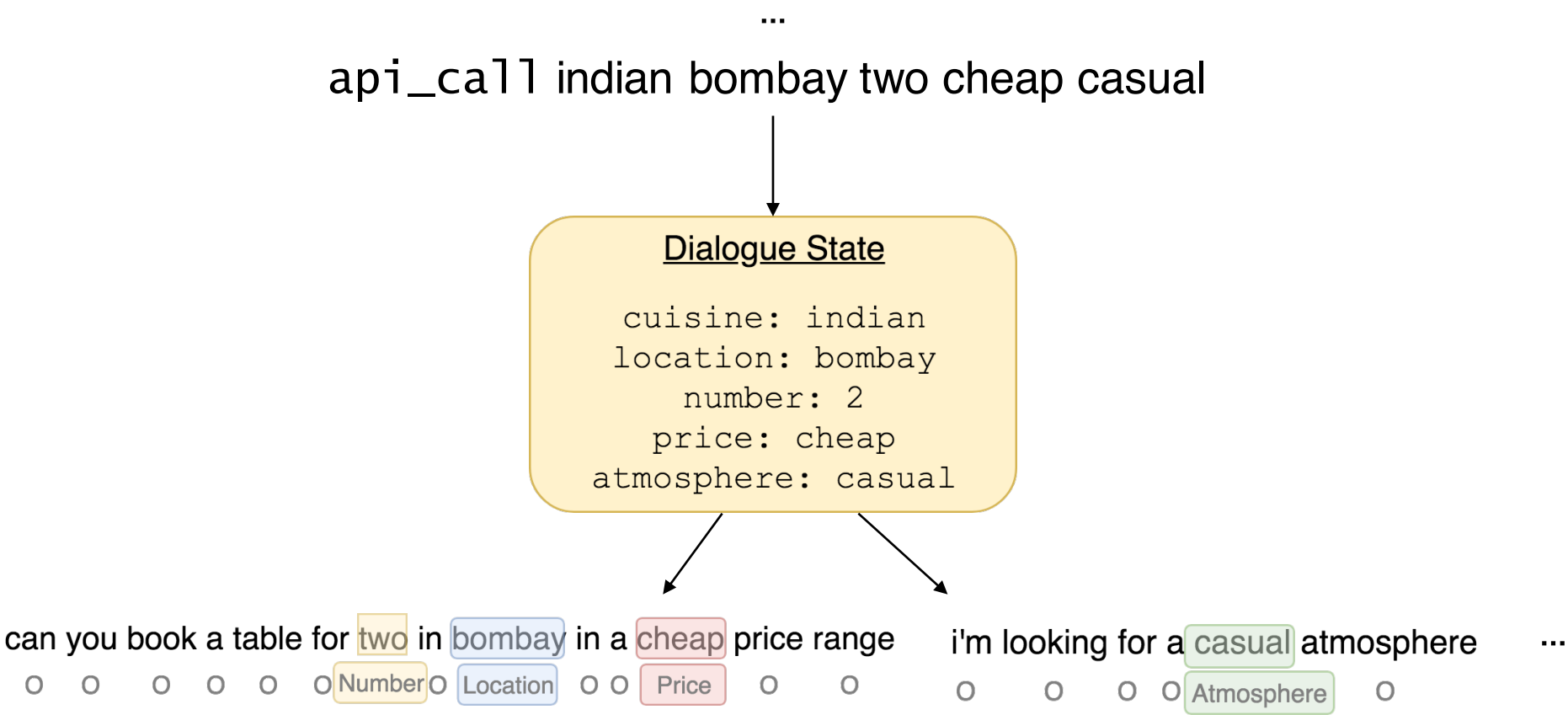
Problem: prior state-of-the-art Conditional Random Field (CRF) model requires **hand-crafted features**.

Solution: use a neural network to **automatically learn features** during training.

Step 1: Semantic Tagging



Step 1: Generating Tagging Data



Step 1: Semantic Tagging Results

Semantic Tag	Precision	Recall	F-score
Cuisine	100	96.9	98.4
Location	100	95.9	97.9
Number	100	100	100
Price	96.9	96.5	96.7
Atmosphere	100	100	100
All	99.8	99.8	99.8

Filter	Top-3 Highest Activation Tokens
19	french, spanish, italian
52	two, six, four
63	bombay, london, paris

Price

expensive is tempting but cheap may be more reasonable

Predicted
O (p = .66)
Price (p = .30)

Location

let me check if london or bombay would work

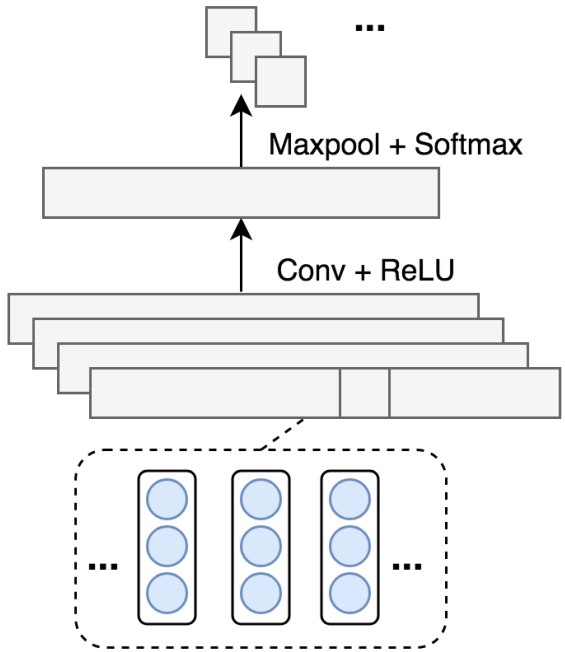
Predicted
O (p = .87)
Location (p = .12)

Step 2: Action Template Selection

Action Template
ok let me look into some options for you
`api_call`
i'm on it
hello what can i help you with today
sure is there anything else to update
you're welcome
what do you think of this option:
great let me do the reservation
sure let me find another option for you
here it is
whenever you're ready
the option was
i am sorry i don't have an answer to that question
is there anything i can help you with
`request_api_slot`

Ranked Actions: 1) `request_api_slot`

2) ok let me look into some options for you



0 0 0 ... good morning

hello what can i help you with

...

let's do moderate price range, and keep expensive price range for another day

Step 3: Final Response Generation

1) Action mask

`api_call`: masked out if any slots are still unspecified.

`request_api_slot`: masked out if all slots are specified.

2) Use dialogue state

`api_call`: populate slots with values in dialogue state.

Dialogue State

```
cuisine: indian
location: bombay
number: 2
price: cheap
atmosphere: casual
```

→ `api_call` indian bombay two cheap casual

`request_api_slot`: select the next slot missing a value.

Dialogue State

```
cuisine: None
location: bombay
number: 2
price: cheap
atmosphere: None
```

Slot	System Response
Cuisine	any preference on a type of cuisine
Location	where should it be
Number	how many people would be in your party
Price	which price range are you looking for
Atmosphere	are you looking for a specific atmosphere

DSTC6: Test Results

100% precision on both tasks

	Task 1			Task 2		
Model	P@1	P@2	P@5	P@1	P@2	P@5
Random	10.2	20.4	50.9	0.95	19.5	46.7
TFIDF	21.0	29.9	52.2	36.7	47.4	66.9
SVM	81.3	81.6	83.0	74.5	76.4	78.9
LSTM	84.3	90.6	98.5	77.8	84.0	97.8
Hier. LSTM	88.6	94.1	99.9	81.7	92.6	100
Bai et al.	99.8	100	100	99.7	100	100
Ham et al.	100	100	100	100	100	100
Binary CNN	78.9	88.9	99.7	69.0	79.3	99.6
Our Model	100	100	100	100	100	100

(Korpusik et al., SLT 2018)

From Simulated to Real Data: WOZ 2.0

Task: predict all the user's requested and informable slots at each turn in a restaurant booking dialogue.

Slot	Type	Num Values
Food	Informable, Requestable	75
Area	Informable, Requestable	7
Pricerange	Informable, Requestable	4
Name	Requestable	N/A
Address	Requestable	N/A
Phone	Requestable	N/A
Postcode	Requestable	N/A
Signature	Requestable	N/A

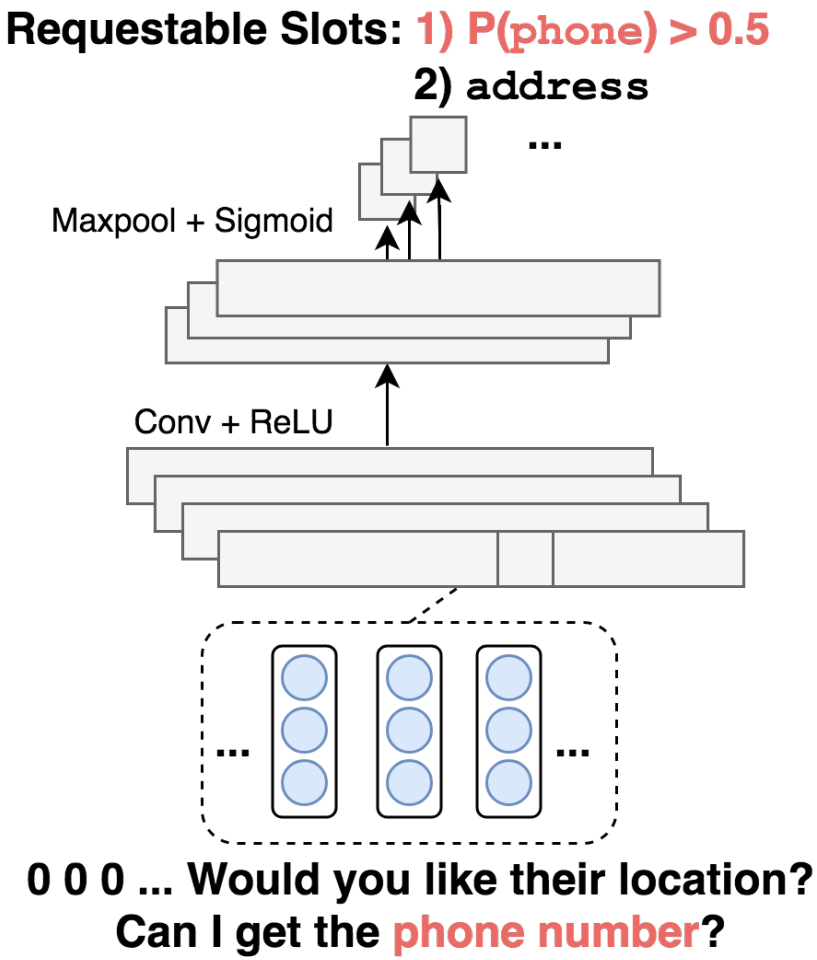
(Henderson et al., SLT 2014)
Neural Belief Tracker: (Mrksic et al., ACL 2017)
GLAD: (Zhong et al., ACL 2018)

User: Is there any place here in the centre that serves corsica food? `food = corsica; area = centre`
System: What price range are you looking for? **User:** Any price range will do. `food = corsica; area = centre; price = dontcare`
System: There are no restaurants available matching your criteria. Would you like to try a different area, price range, or food type? **User:** Are there any restaurants in the centre that serves North American type of food? `food = north_american; area = centre; price = dontcare`
System: Yes. The gourmet burger kitchen serves north american food and is located in the centre part of the city. Would you like their location? **User:** Can I get the phone number? `food = north_american; area = centre; price = dontcare; requested = phone`
System: The phone number for gourmet burger kitchen is 01223 312598. **User:** Thank you. Good bye. `food = north_american; area = centre; price = dontcare`

(Korpusik et al., SLT 2018)

From Simulated to Real Data: WOZ 2.0

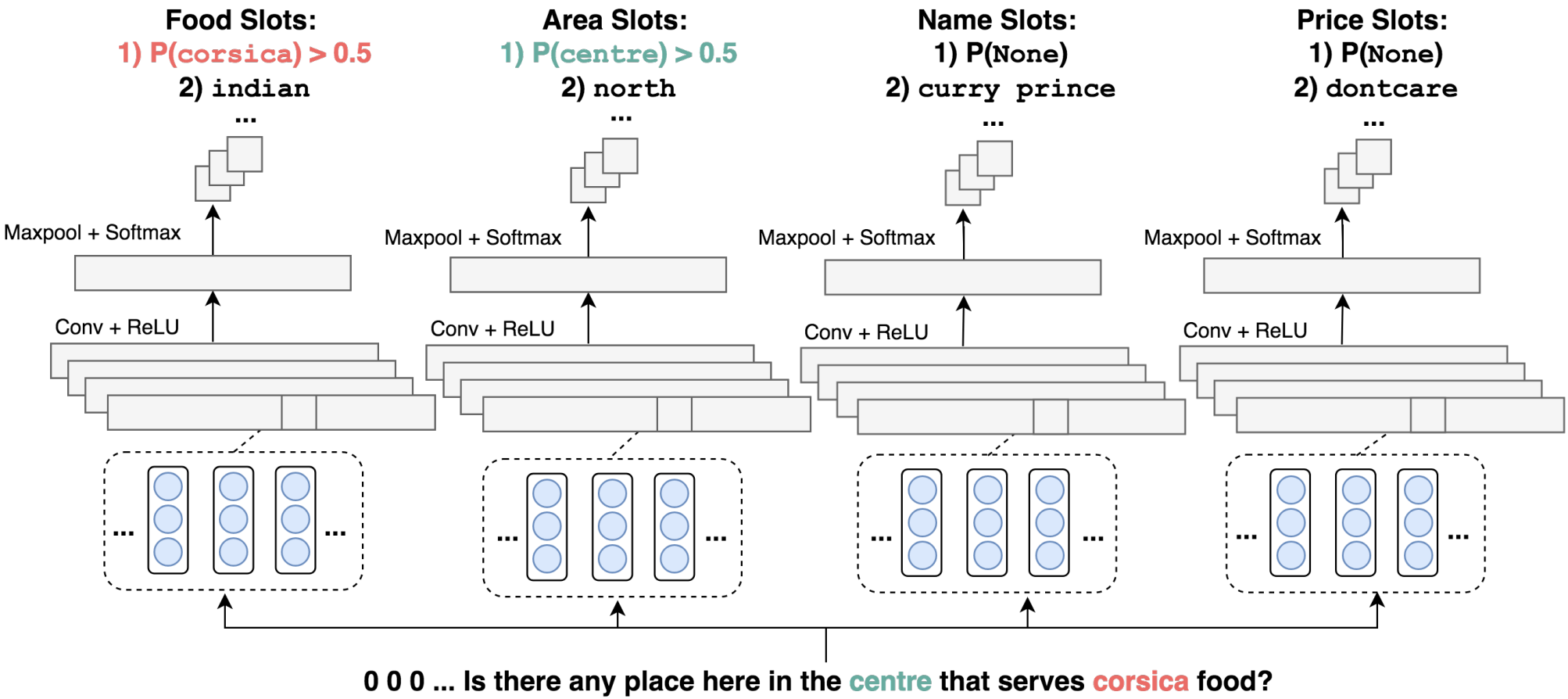
Our Goal: Show our NN models generalize to real dialogues.



(Korpusik et al., SLT 2018)

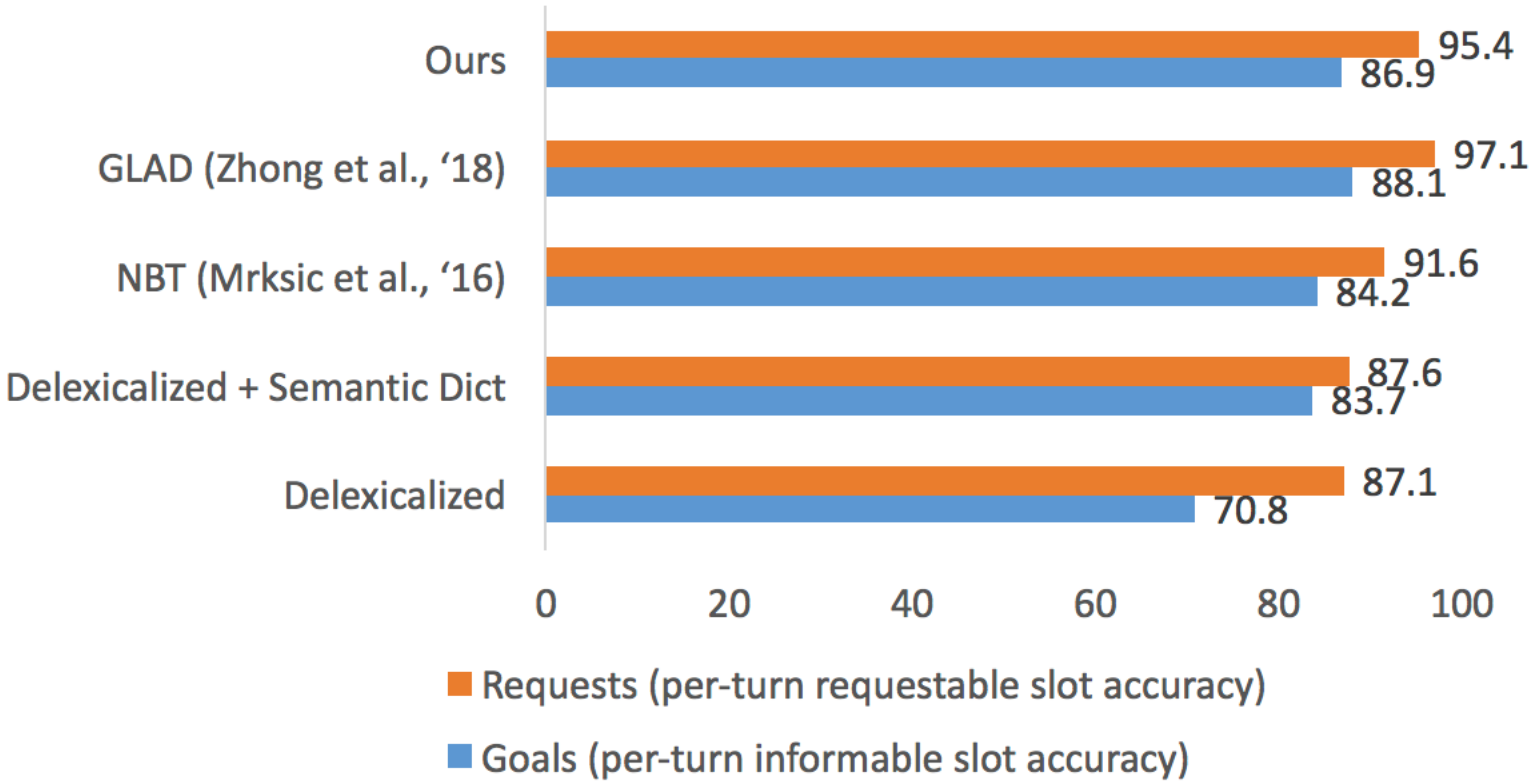
From Simulated to Real Data: WOZ 2.0

Separately trained informable slot models.



From Simulated to Real Data: WOZ 2.0

CNN is competitive with state-of-the-art, without requiring semantic dictionaries or pre-trained word vectors.



Conclusion

Demonstrated our neural network models' ability to do dialogue state tracking in several domains.

Future Work:

- Experiment on the remaining DSTC6 subtasks.
- Jointly train tagger and action selector as end-to-end model.
- Automatically learn action mask by adding a feature to action selector model indicating whether all slots have values.
- **Apply these techniques to the nutrition domain!**