Dialogue Systems

(Focused on Task-Oriented Systems)

Cristina España-Bonet

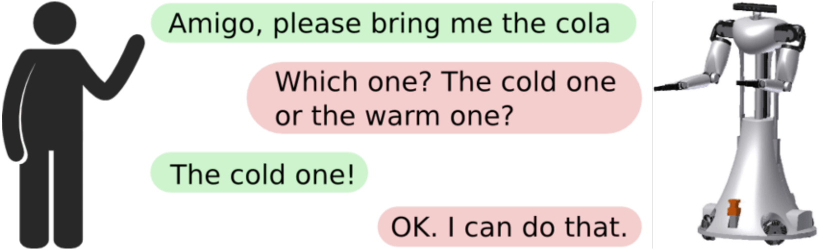
UdS & DFKI, Saarbrücken, Germany

Artificial Intelligence with Deep Learning

2nd May 2019

Motivation

Let a Robot Work for you! That's lazy...



[Perzylo *et al.*, 2015]

Motivation

They can Help or Amuse you

Motivation

More than Dialogue...

http://www.aliz-e.org/

https://vimeo.com/111724122

Motivation

More than Dialogue... An Example

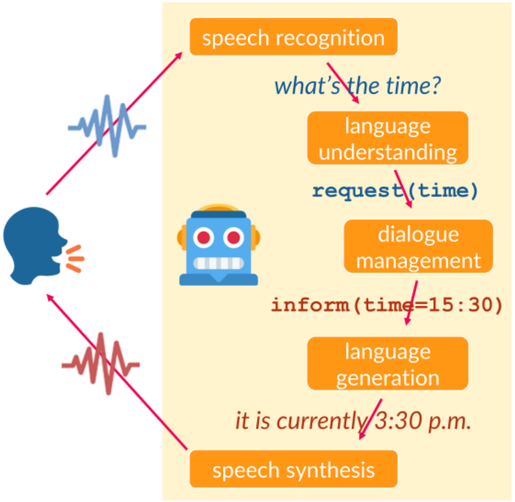
The ALIZ-E scientific and technological goals

- Prolonged human-robot interaction over a range of days instead of in the here and now
- Long-term memory and self-sustained long-term interaction, personalised adaptive memory storing experiences and interaction episodes
- Analysis and synthesis of emotion and affect in human-robot interaction
- Pervasive machine learning and adaptation. Learning experiences will be unstructured
- ...

A *(spoken)* dialogue system is a computer system designed to **interact with users** in *(spoken)* **natural language**

Motivation

A Grosso Modo...



https://ufal.mff.cuni.cz/courses/npfl123

Outline

Introduction

Task-Oriented

- Where can I find the closest Peruan restaurant?
- Which bus do I need to take to go the theater?

Introduction

Broad Distinction

Task-Oriented

- Where can I find the closest Peruan restaurant?
- Which bus do I need to take to go the theater?

Non-Task-Oriented

- chitchat, chat bots
- How do you feel today?

As always, from rules to neurons:

- Rule-based
- Information retrieval
- Neural generative

Non-Task-Oriented: Chatbots

As always, from rules to neurons:

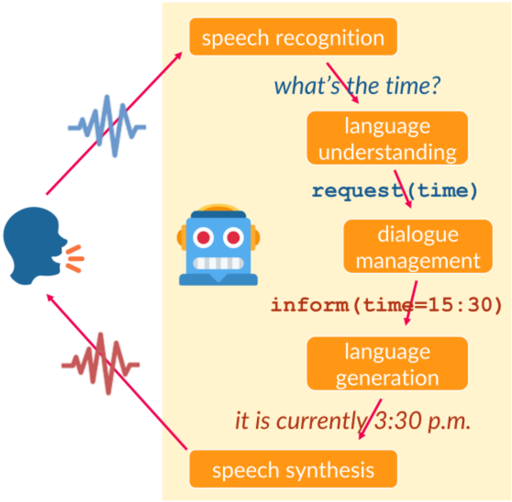
- Rule-based
- Information retrieval
- Neural generative

But you'll have to wait a bit...

Specific lecture by Carlos Segura in the audio block

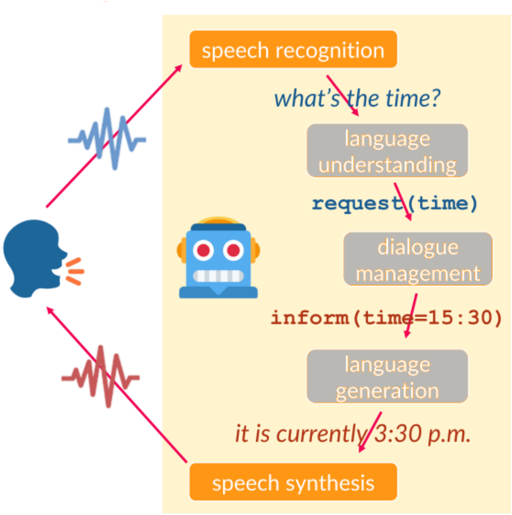
Classical Modular Architecture

Task-Oriented Systems



Classical Modular Architecture

Task-Oriented Systems



Classical Modular Architecture

Task-Oriented Systems, Pipeline

Natural Language Understanding

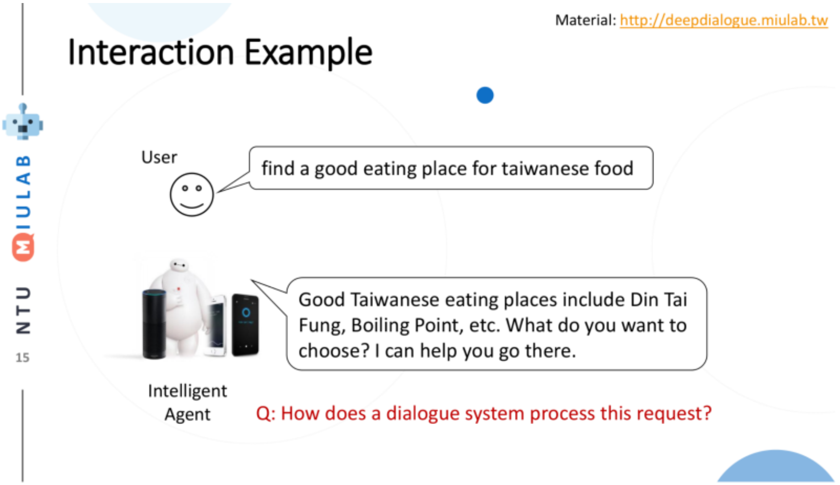
Extract the meaning from the user utterance and convert it into a structured semantic representation (filling semantic slots)

Main steps:

1. Domain Identification (if needed)
2. User Intent Detection
3. Slot Filling

Classical Modular Architecture

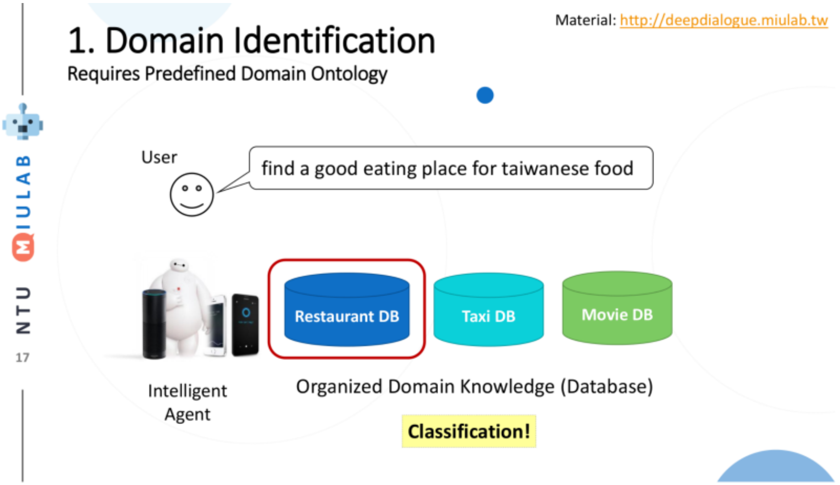
Language Understanding



(Yun-Nung Chen slides)

Classical Modular Architecture

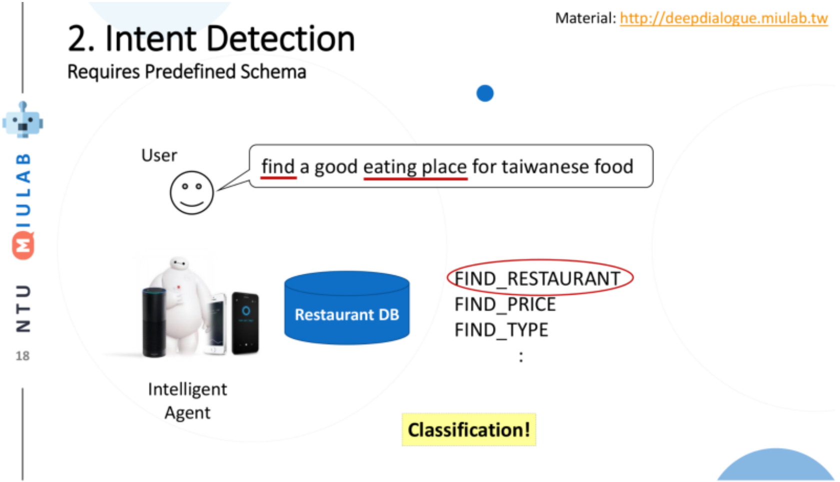
Language Understanding



(Yun-Nung Chen slides)

Classical Modular Architecture

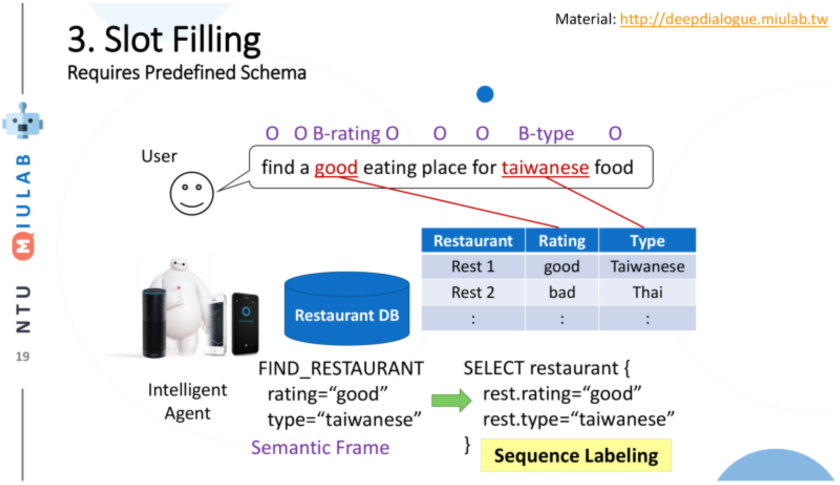
Language Understanding



(Yun-Nung Chen slides)

Classical Modular Architecture

Language Understanding



(Yun-Nung Chen slides)

Classical Modular Architecture

Language Understanding: Implementations

Classification

- Handcrafted
  keyword spotting, regular expressions, grammars

- Machine learning
  intent classifiers+slot/value extraction (BOW, tf-idf);
  RNNs (GRU, LSTM), CNNs (1D)...

Sequence Labeling

- Machine learning
  CRFs, DBNs, RNNs...

Main steps:

1. Domain Identification (if needed)
2. User Intent Detection
3. Slot Filling

General dialogue acts:

- act type/intent (inform, request, confirm)
- slot/attribute (price, time, quality...)
- value (11:34, good, city center...)

Classical Modular Architecture

Language Understanding

Main steps:

1. Domain Identification (if needed)
2. User Intent Detection
3. Slot Filling

General dialogue acts:

- act type/intent (inform, request, confirm)
- slot/attribute (price, time, quality...)
- value (11:34, good, city center...)

inform(food=Taiwanese, quality=good)

Classical Modular Architecture

Task-Oriented Systems, Pipeline

Dialogue Management

Decide next action given the previous dialogue and the output of the NLU module

Main tasks:

- Keeping track of what has been said
- Interaction with backend
- Deciding the action to perform

Classical Modular Architecture

Task-Oriented Systems, Pipeline

Dialogue Management

Decide next action given the previous dialogue and the output of the NLU module

Main tasks:

- Keeping track of what has been said
- Interaction with backend
- Deciding the action to perform

Modules:

1. Dialogue State Tracking
2. Dialogue Policy

Dialogue State Tracking

Estimates the user goal at each turn

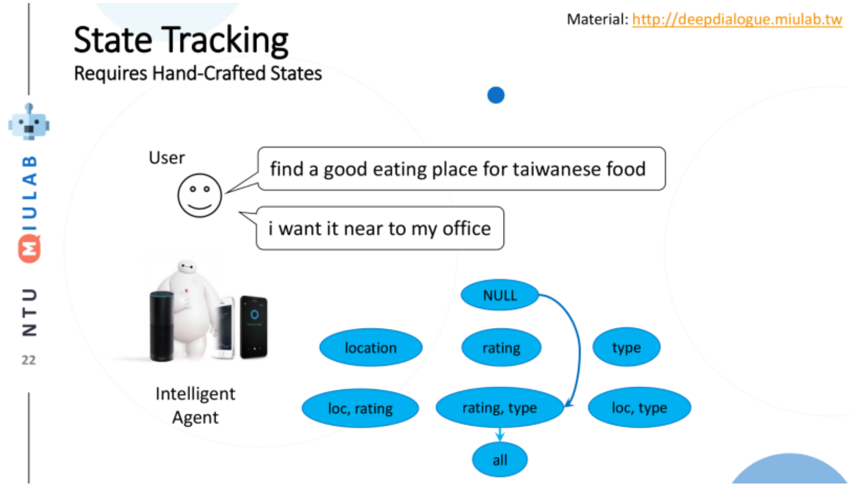
Dialogue state includes:

- User requests and the information they provided so far
- Information requested and provided by the system
- User preferences

Maintain a probabilistic distribution for robustness

Classical Modular Architecture

Dialogue Management



(Yun-Nung Chen slides)

Dialogue State Tracking

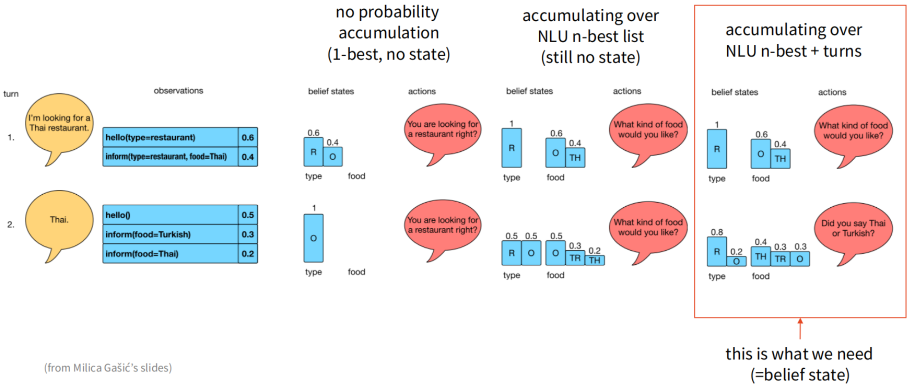
Estimates the user goal at each turn

Maintain a probabilistic distribution for robustness:

$$b'\underbrace{(p', a'_u, s'_d)}_{\substack{h' \\ \text{new} \\ \text{hypothesis}}} = k \cdot \underbrace{P(o'|a'_u)}_{\substack{\text{observation} \\ \text{model}}} \underbrace{P(a'_u|p', a_m)}_{\substack{\text{user} \\ \text{action} \\ \text{model}}} \underbrace{P(p'|p)}_{\substack{\text{user goal} \\ \text{model}}} \sum_{s_d} \underbrace{P(s'_d|p', a'_u, s_d, a_m)}_{\substack{\text{dialogue history} \\ \text{model}}} b\underbrace{(p, a_u, s_d)}_{\substack{h \\ \text{old} \\ \text{hypothesis}}},$$

belief state

Belief State



(from Milica Gašić's slides)

Classical Modular Architecture

Dialogue Management

Dialogue Policy

Given a state by the DST selects an action

- Predefined actions (or dialog acts)
- Rule-based: if-then-else clauses
- Data-based: Probability distribution (robustness again) over the predefined system action set
- Machine learning
  RNNs, reinforcement learning...

Example, after turn 1: request(location) or confirm(GPSloc)

Natural Language Generation

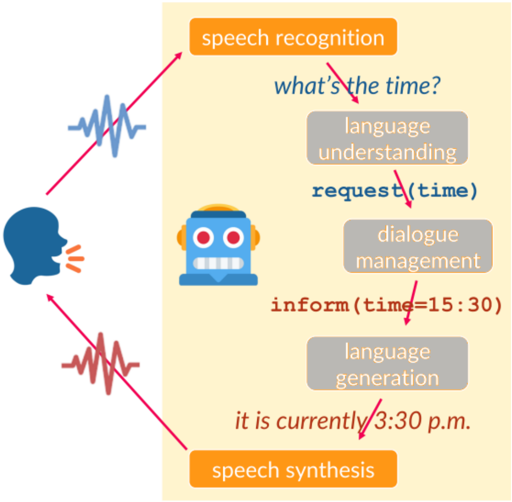
Convert the selected action into natural language

Methods:

1. Template filling
2. Grammar-based (action to tree to text)
3. (Recurrent) Neural networks

Classical Modular Architecture

So, Task-Oriented Systems



Neural Network Architectures

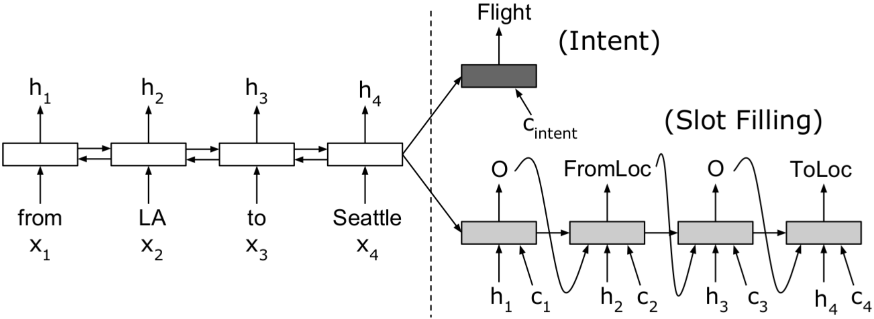
Neural Network Architectures

Modular Pipeline

We'll see neural ideas for:

- Intent classifier and slot tagger (NLU)
- Context (turns) for slot tagger (NLU)
- Neural belief tracker (DM)
- Joint LU and policy learning (DM)
- Seq2Seq for language generation (NLG)

Attention-Based Recurrent Neural Network Models (enc-dec!)

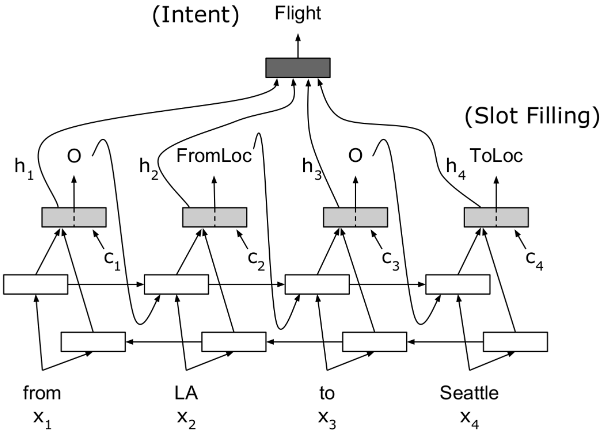


[Liu and Lane, 2016]

Neural Network Architectures

NLU: Intent Classifier and Slot Tagger

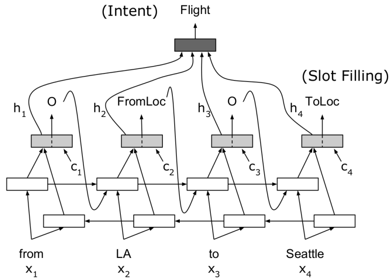
With attention [Liu and Lane, 2016]



Neural Network Architectures

NLU: Intent Classifier and Slot Tagger

With attention [Liu and Lane, 2016]



O	O	B-fromloc. city_name	O	B-toloc. city_name	O	O	B-depart_time. time_relative	B-depart_time. period_of_day
flight	from	cleveland	to	dallas	that	leves	before	noon

Neural Network Architectures

NLU: Intent Classifier and Slot Tagger

- Joint learning improves separate learning both with classical approaches (boosting, CRFs...) and NNs

Neural Network Architectures

NLU: Intent Classifier and Slot Tagger

- Joint learning improves separate learning both with classical approaches (boosting, CRFs...) and NNs
- And BERT comes again!

Models	Snips			ATIS		
	Intent	Slot	Sent	Intent	Slot	Sent
RNN-LSTM (Hakkani-Tür et al., 2016)	96.9	87.3	73.2	92.6	94.3	80.7
Atten.-BiRNN (Liu and Lane, 2016)	96.7	87.8	74.1	91.1	94.2	78.9
Slot-Gated (Goo et al., 2018)	97.0	88.8	75.5	94.1	95.2	82.6
Joint BERT	**98.6**	**97.0**	**92.8**	97.5	**96.1**	88.2
Joint BERT + CRF	98.4	96.7	92.6	**97.9**	96.0	**88.6**

Table 2: NLU performance on Snips and ATIS datasets. The metrics are intent classification accuracy, slot filling F1, and sentence-level semantic frame accuracy (%).

[Chen et al. 2019]

Neural Network Architectures

NLU: Intent Classifier and Slot Tagger

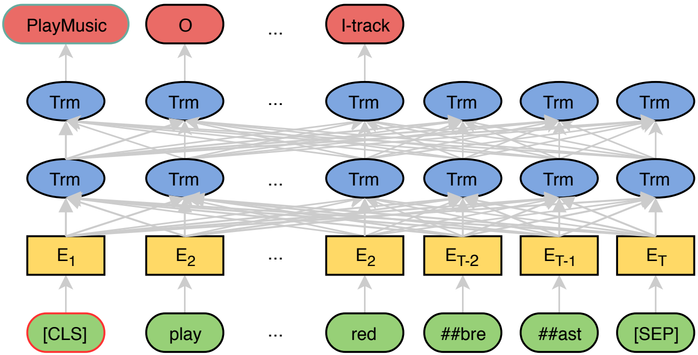


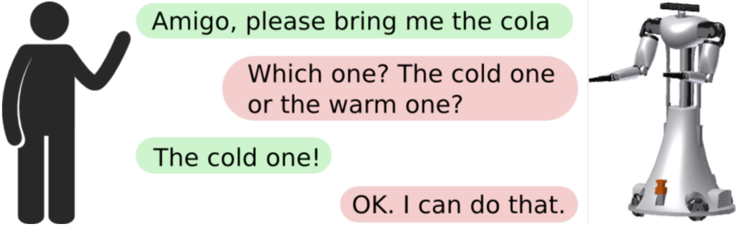
Figure 1: A high-level view of the proposed model. The input query is "play the song little robin redbreast".

[Chen et al. 2019]

Neural Network Architectures

Intents, Slots...

...but remember that a dialogue has turns! And the info must be remembered from turn to turn.



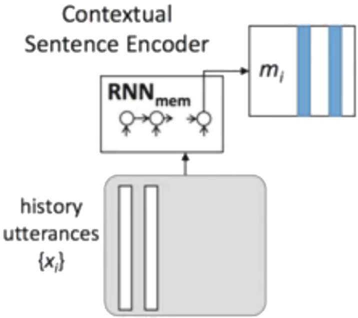
[Perzylo *et al.*, 2015]

Memory Networks for Multi-Turn NLU

Storage in explicit memory cells

- Explicitly model long-term memory
- Helps for reasoning
- Can introduce external knowledge
- Can visualize memory access
- Solves memory selection via attention

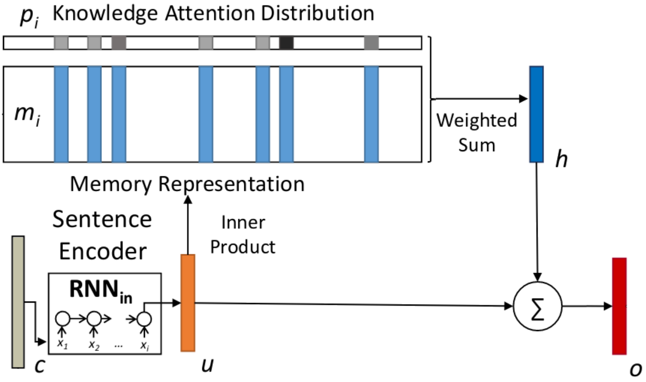
1. Encode each sentence as a memory (dense) vector



$$m_i = \mathrm{RNN}_{\mathrm{mem}}(x_i)$$

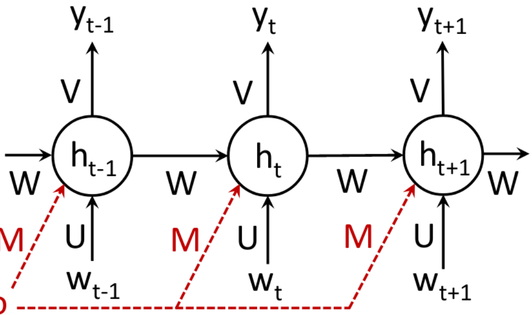
[Chen *et al.*, 2016]

2. Match the representation of a new utterance with the memory



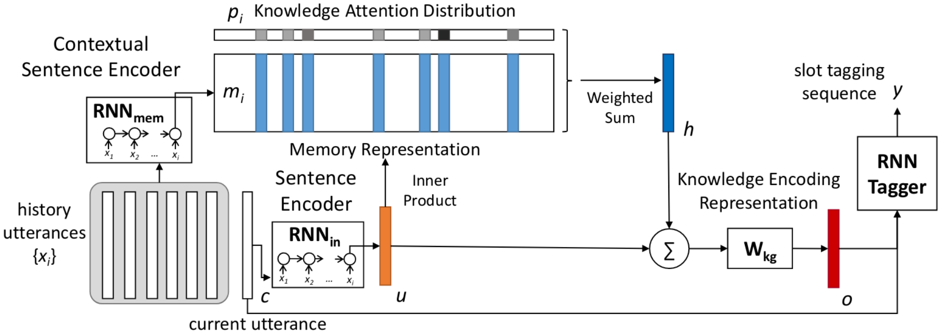
[Chen *et al.*, 2016]

3. The composition is feeded into the final RNN for tagging at each time step



[Chen *et al.*, 2016]

3. The composition is feeded into the final RNN for tagging



[Chen *et al.*, 2016]

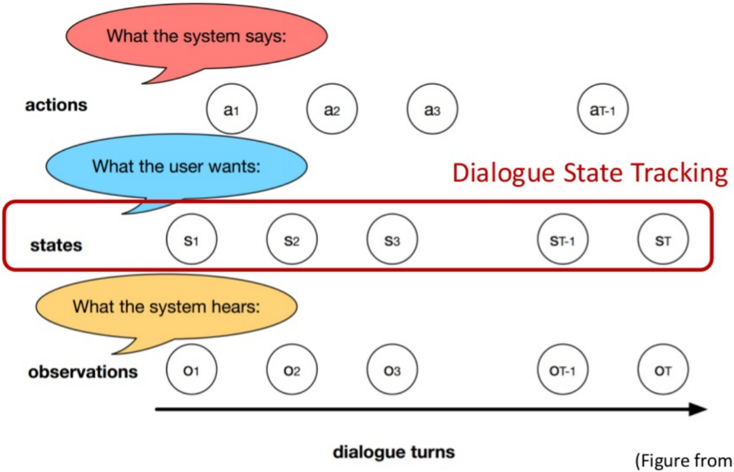
And is it good for slot tagging?

- F1 measure (%) on a multi-turn dataset (dialogues collected from Microsoft Cortana with #turns $\geqslant 5$)

	First Turn	Other turns	Overall
RNN	55.8	45.7	47.4
+Memory	73.2	65.7	67.1

Neural Network Architectures

DM: State Tracking

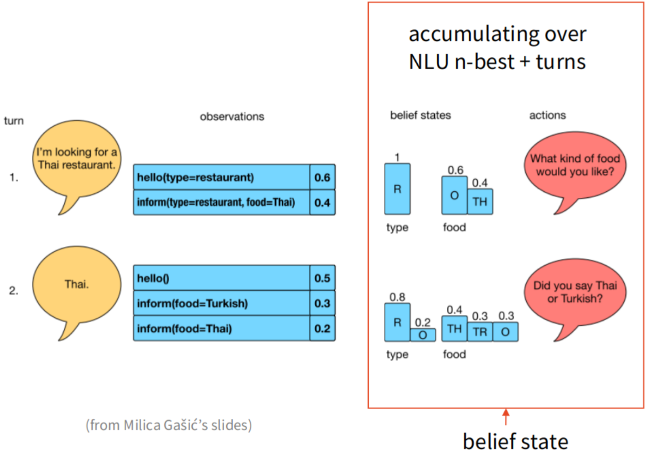


(Figure from Gašić)

Neural Network Architectures

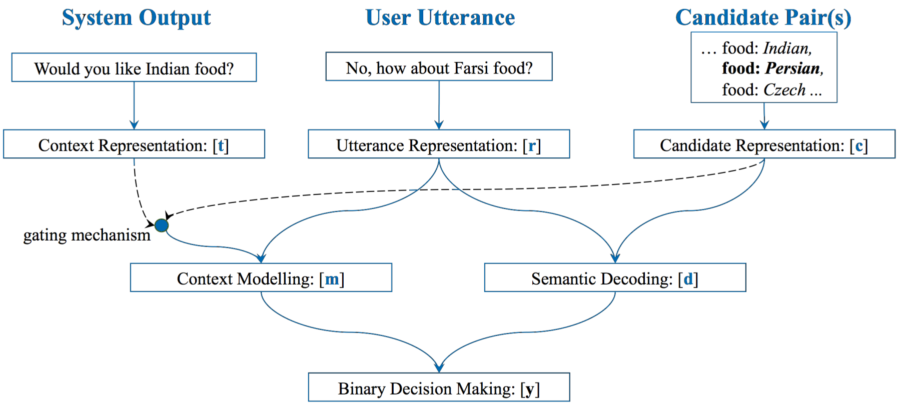
DM: Neural Belief States

Remember...



(from Milica Gašić's slides)

Joint LU and State Tracking



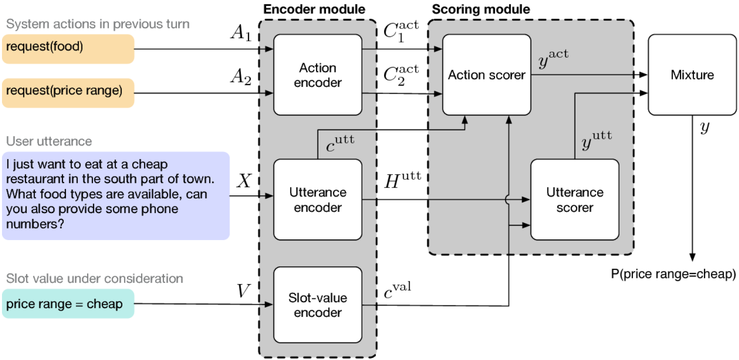
[Mrkšić *et al.*, 2016]

Neural Network Architectures

DM: Neural State Tracker

Global-Locally Self-Attentive DST [Zhong *et al.*, 2018]

- Global modules share parameters for all slots
- Local modules learn slot-specific feature representations

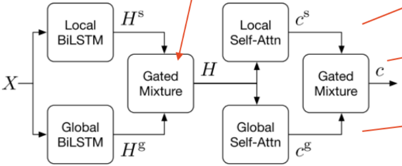


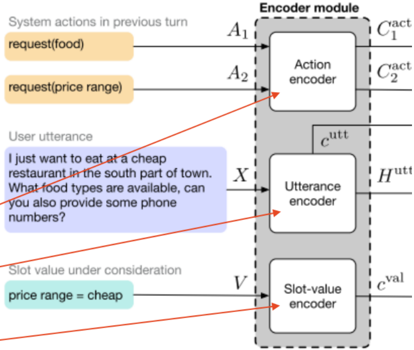
Neural Network Architectures

DM: Neural State Tracker

$\beta \cdot \text{global} + (1-\beta) \cdot \text{local}$

encoders shape:



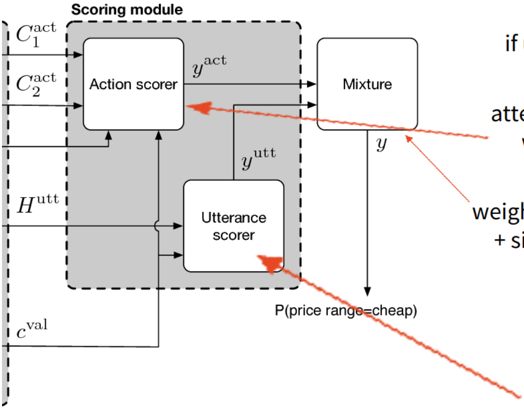


local = per-slot, global = shared among slots

[Zhong *et al.*, 2018]

Neural Network Architectures

DM: Neural State Tracker



if utterance refers to previous system actions

attention over prev. system actions w. r. t. current user utterance

weighted sum + sigmoid

does the utterance specify this slot-value pair? attention over utterance w. r. t. slot-value pair

[Zhong *et al.*, 2018]

Neural Network Architectures

DM: State Tracking SotA

Second dialogue state tracking challenge

For goal-oriented dialogue, the dataset of the second Dialogue Systems Technology Challenges (DSTC2) is a common evaluation dataset. The DSTC2 focuses on the restaurant search domain. Models are evaluated based on accuracy on both individual and joint slot tracking.

Model	Request	Area	Food	Price	Joint	Paper / Source
Zhong et al. (2018)	97.5	-	-	-	74.5	Global-locally Self-attentive Dialogue State Tracker
Liu et al. (2018)	-	90	84	92	72	Dialogue Learning with Human Teaching and Feedback in End-to-End Trainable Task-Oriented Dialogue Systems
Neural belief tracker (Mrkšić et al., 2017)	96.5	90	84	94	73.4	Neural Belief Tracker: Data-Driven Dialogue State Tracking
RNN (Henderson et al., 2014)	95.7	92	86	86	69	Robust dialog state tracking using delexicalised recurrent neural networks and unsupervised gate

https://github.com/sebastianruder/NLP-progress

Wizard-of-Oz

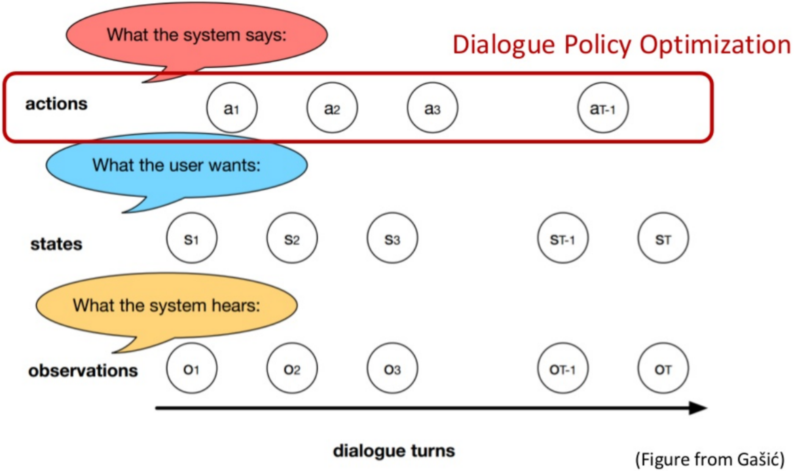
The WoZ 2.0 dataset is a newer dialogue state tracking dataset whose evaluation is detached from the noisy output of speech recognition systems. Similar to DSTC2, it covers the restaurant search domain and has identical evaluation.

Model	Request	Joint	Paper / Source
Zhong et al. (2018)	97.1	88.1	Global-locally Self-attentive Dialogue State Tracker
Neural belief tracker (Mrkšić et al., 2017)	96.5	84.4	Neural Belief Tracker: Data-Driven Dialogue State Tracking
RNN (Henderson et al., 2014)	87.1	70.8	Robust dialog state tracking using delexicalised recurrent neural networks and unsupervised gate

https://github.com/sebastianruder/NLP-progress

Neural Network Architectures

DM: Policy Learning



(Figure from Gašić)

Rules

- Given the output of NLU state tracker write hand crafted rules

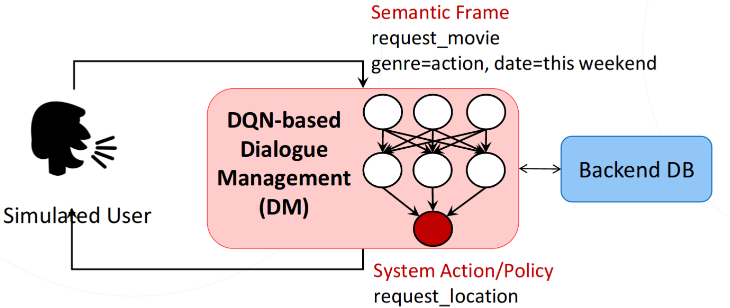
Supervised learning

- Train on large amount of expert-labeled dialogues

Reinforcement learning

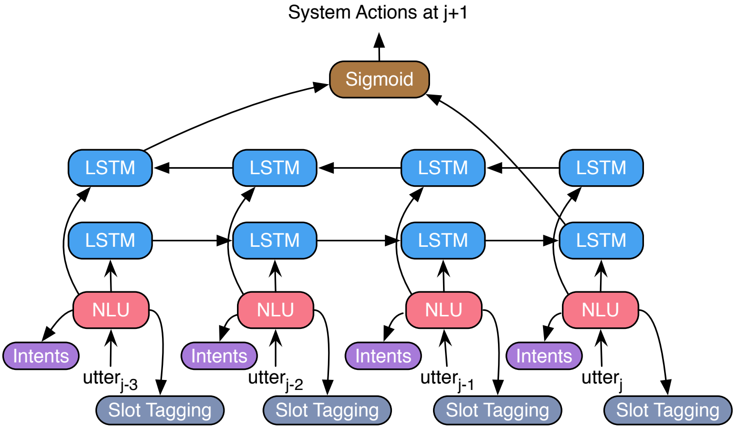
- Train against a reward signal –possibly given by simulated users

Policy with Reinforcement Learning (Deep Q-Networks)



(Yun-Nung Chen slides)

Joint LU and Policy Learning (without State Tracking)



[Yang *et al.*, 2016]

Neural Network Architectures

NLG: Generation

- As standard NLG + specific semantic information given by the dialogue policy
- NNs1: generation output are delexicalized utterances with variable names for slots, which are then replaced with actual values as part of post-processing

Neural Network Architectures

NLG: Generation

- As standard NLG + specific semantic information given by the dialogue policy

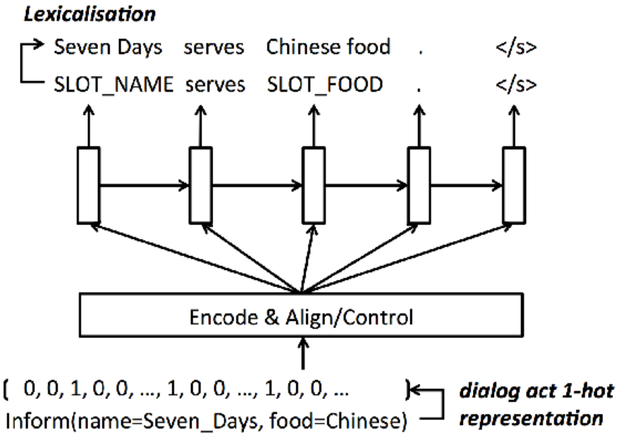
- NNs1: generation output are delexicalized utterances with variable names for slots, which are then replaced with actual values as part of post-processing

- NNs2: the content of the slots is important!

#	`recommend(restaurant_name= Au Midi, neighborhood = midtown, cuisine = french`
1	Au Midi is in Midtown and serves French food.
	`recommend(restaurant_name= Loch Fyne, neighborhood = city centre, cuisine = seafood)`
4	Loch Fyne is in **the** City Center and serves seafood food.

Neural Network Architectures

NLU: Seq2Seq for Generation



[Wen *et al.*, 2015]

Neural Network Architectures

Modular vs. End-to-End

Modular

- Lot of domain-specific handcrafting
- User's feedback hard to propagate (backwards)
- System's errors easy to propagate (forward)

End-to-End

- Lots of data needed
- Joint learning of different tasks
- Difficulty to query external databases

References

References

Learn More

- Dialogue Systems Course at Charles University
  `https://ufal.mff.cuni.cz/courses/npfl123`

- Deep Learning for Dialogue Systems – Tutorial by Yun-Nung (Vivian) Chen
  `https://www.csie.ntu.edu.tw/~yvchen/doc/KAIST19_Tutorial.pdf`

- Hongshen Chen, Xiaorui Liu, Dawei Yin and Jiliang Tang. 2017. **A Survey on Dialogue Systems: Recent Advances and New Frontiers**. SIGKDD Explor. Newsl. 19, 2. Pages 25–35.

References

References Used

- Bing Liu and Ian Lane. 2016. **Attention-Based Recurrent Neural Network Models for Joint Intent Detectionand Slot Filling**. *Proceedings of the 17th Annual Meeting of the International Speech Communication Association (INTERSPEECH)*.
- Yun-Nung Vivian Chen, Dilek Hakkani-Tür, Gokhan Tur, Jianfeng Gao, Li Deng. 2016. **End-to-End Memory Networks with Knowledge Carryover for Multi-Turn Spoken Language Understanding**. *Proceedings of the 17th Annual Meeting of the International Speech Communication Association INTERSPEECH*.
- Qian Chen, Zhu Zhuo and Wen Wang. 2019. **BERT for Joint Intent Classification and Slot Filling**. *https://arxiv.org/pdf/1902.10909.pdf*.

References Used

- Nikola Mrkšić, Diarmuid Ó Séaghdha, Tsung-Hsien Wen, Blaise Thomson and Steve Young. 2016. **Neural Belief Tracker: Data-Driven Dialogue State Tracking**. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL)*. Pages 1777–1788.

- Victor Zhong, Caiming Xiong and Richard Socher. 2018. **Global-Locally Self-Attentive Encoder for Dialogue State Tracking**. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL)*. Pages 1458–1467.

- Xuesong Yang, Yun-Nung Chen, Dilek Hakkani-Tur, Paul Crook, Xiujun Li, Jianfeng Gao and Li Deng. 2016. **End-to-End Joint Learning of Natural Language Understanding and Dialogue Manager**. *Proceedings of the 42nd IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2017)*.

References

References Used

- Xiujun Li, Yun-Nung Chen, Lihong Li, Jianfeng Gao, Asli Celikyilmaz. 2017. **End-to-End Task-Completion Neural Dialogue Systems**. *Proceedings of the 8th International Joint Conference on Natural Language Processing (IJCNLP)*. Pages 733–743.

- Tsung-Hsien Wen, Milica Gasic, Nikola Mrkšić, Pei-hao Su, David Vandyke and Steve J. Young. 2015. **Semantically Conditioned LSTM-based Natural Language Generation for Spoken Dialogue Systems**. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Pages 1711–1721.

Over and Out

Questions?

Dialogue Systems

(Focused on Task-Oriented Systems)

Cristina España-Bonet

UdS & DFKI, Saarbrücken, Germany

Artificial Intelligence with Deep Learning

2nd May 2019