Deep Learning for Dialogue Systems

國立臺灣大學
National Taiwan University

PROF. YUN-NUNG (VIVIAN) CHEN 陳縕儂

HTTP://VIVIANCHEN.IDV.TW

GTC 2018
Mar 28th, 2018

Best Poster Award @ GTC 2017

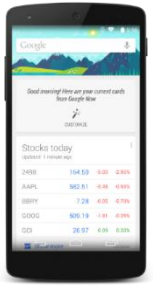
Thanks NVIDIA!!!

Future Life – Intelligent Assistant

Introduction & Background

Language Empowering Intelligent Assistant

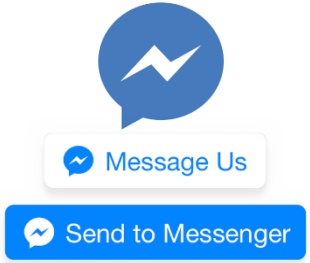
Apple Siri (2011)



Google Now (2012)
Google Assistant (2016)

Microsoft Cortana (2014)

Amazon Alexa/Echo (2014)



Facebook M & Bot (2015)

Google Home (2016)

Apple HomePod (2017)

Why We Need?

- Get things done
  - E.g. set up alarm/reminder, take note
- Easy access to structured data, services and apps
  - E.g. find docs/photos/restaurants
- Assist your daily schedule and routine
  - E.g. commute alerts to/from work
- Be more productive in managing your work and personal life

“Hey Assistant”

Why Natural Language?

- Global Digital Statistics (2017 January)

Total Population
7.48B

Internet Users
3.77B

Active Social Media Users
2.79B

Unique Mobile Users
4.92B

Active Mobile Social Users
2.55B

The more **natural** and **convenient** input of devices evolves towards speech.

Dialogue System

- Spoken dialogue systems are intelligent agents that are able to help users finish tasks more efficiently via spoken interactions.
- Spoken dialogue systems are being incorporated into various devices (smart-phones, smart TVs, in-car navigating system, etc).

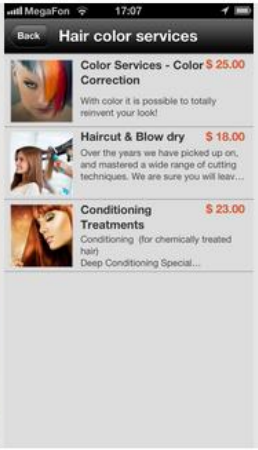
JARVIS – Iron Man's Personal Assistant

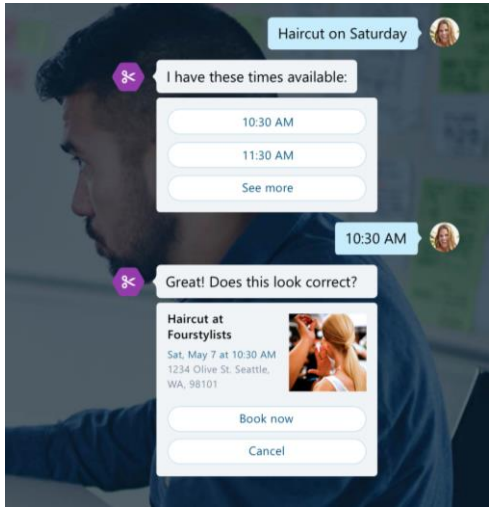
Baymax – Personal Healthcare Companion

Good dialogue systems assist users to access information conveniently and finish tasks efficiently.

App → Bot

- A bot is responsible for a “single” domain, similar to an app

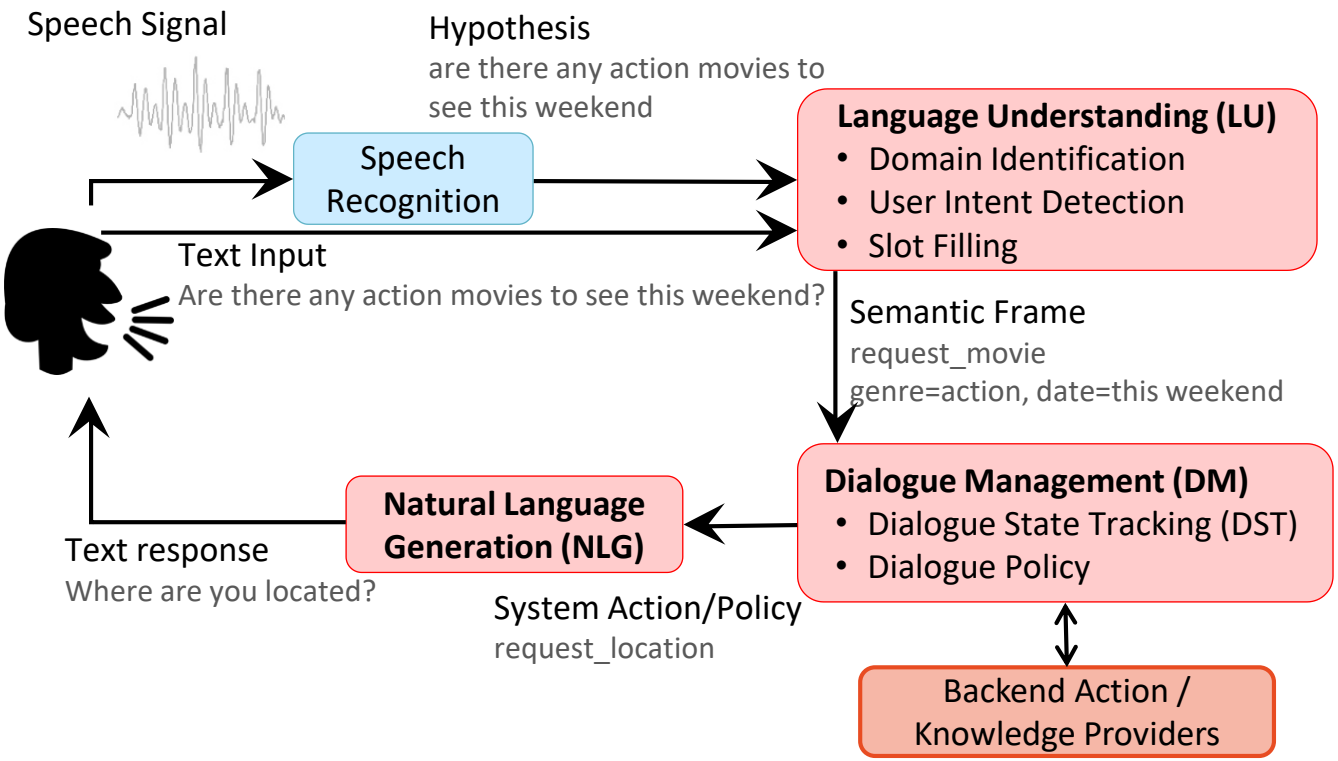




Users can initiate dialogues instead of following the GUI design

Task-Oriented Dialogue System (Young, 2000)

http://rsta.royalsocietypublishing.org/content/358/1769/1389.short



Interaction Example

User

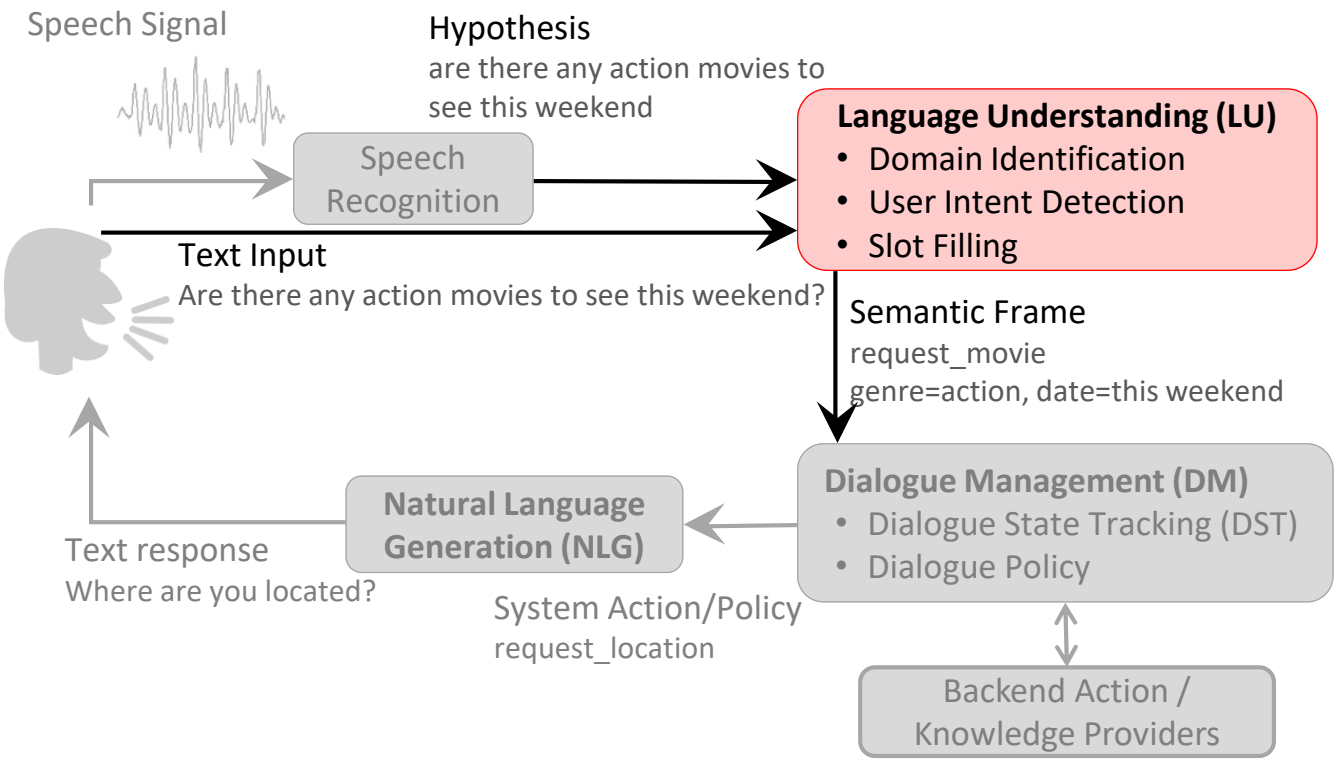
find a good eating place for taiwanese food

Good Taiwanese eating places include Din Tai Fung, Boiling Point, etc. What do you want to choose? I can help you go there.

Intelligent Agent

Q: How does a dialogue system process this request?

Task-Oriented Dialogue System (Young, 2000)

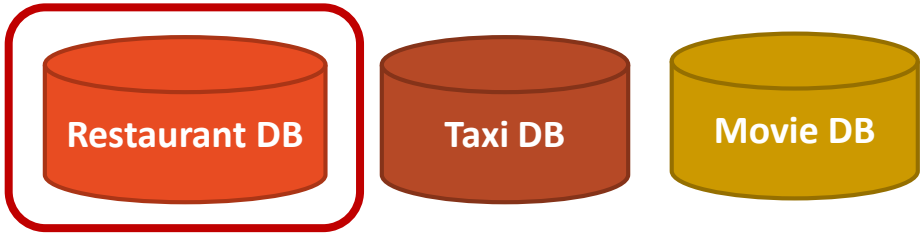


1. Domain Identification

Requires Predefined Domain Ontology

User

find a good eating place for taiwanese food



Intelligent Agent

Organized Domain Knowledge (Database)

Classification!

2. Intent Detection

Requires Predefined Schema

User

find a good eating place for taiwanese food

Intelligent Agent

FIND_RESTAURANT
FIND_PRICE
FIND_TYPE
:

Classification!

3. Slot Filling

Requires Predefined Schema

O O B-rating O O O B-type O

User

find a good eating place for taiwanese food

Restaurant	Rating	Type
Rest 1	good	Taiwanese
Rest 2	bad	Thai
:	:	:

Intelligent Agent

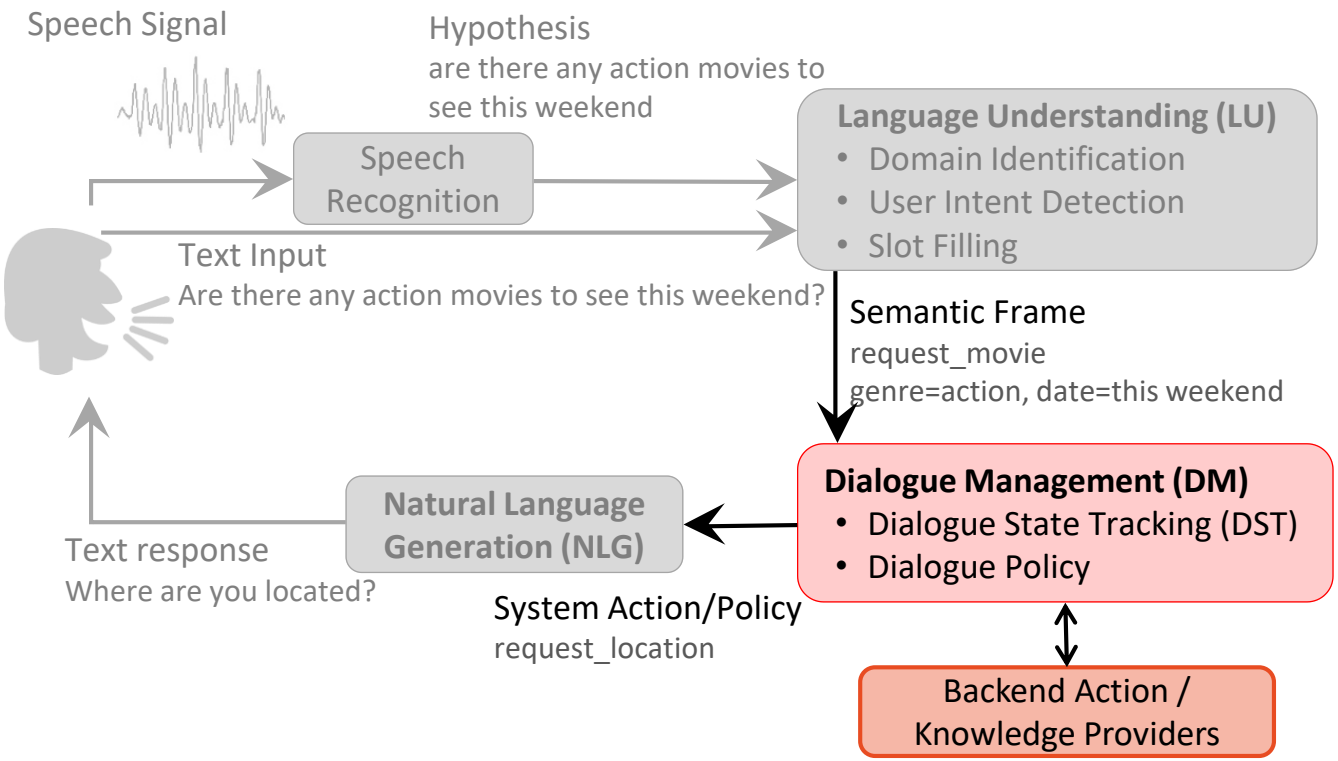
FIND_RESTAURANT
rating="good"
type="taiwanese"

Semantic Frame

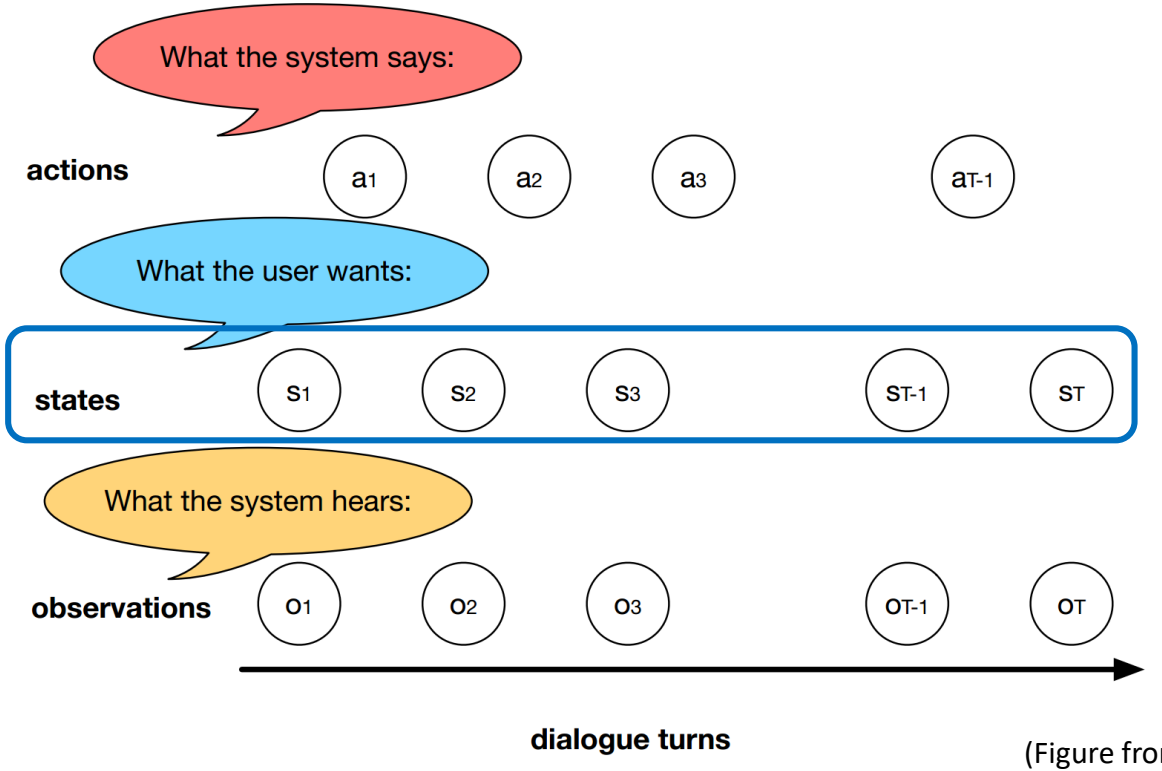
SELECT restaurant {
rest.rating="good"
rest.type="taiwanese"
}

Sequence Labeling

Task-Oriented Dialogue System (Young, 2000)



Elements of Dialogue Management



(Figure from Gašić)

State Tracking

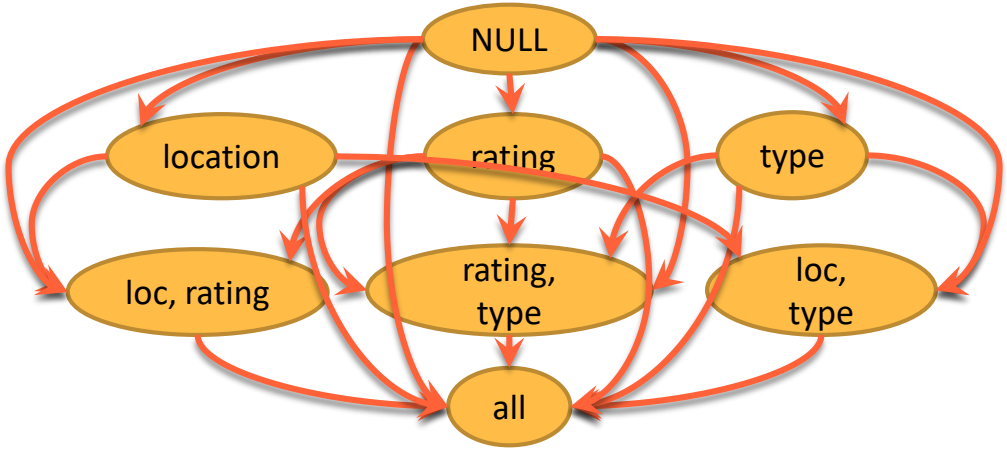
Requires Hand-Crafted States

User

find a good eating place for taiwanese food

i want it near to my office

Intelligent Agent



State Tracking

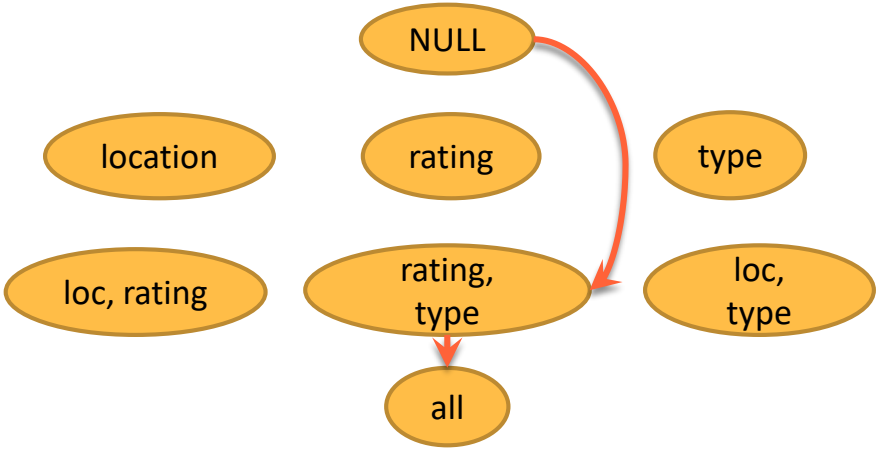
Requires Hand-Crafted States

User

find a good eating place for taiwanese food

i want it near to my office

Intelligent Agent



State Tracking

Handling Errors and Confidence

User

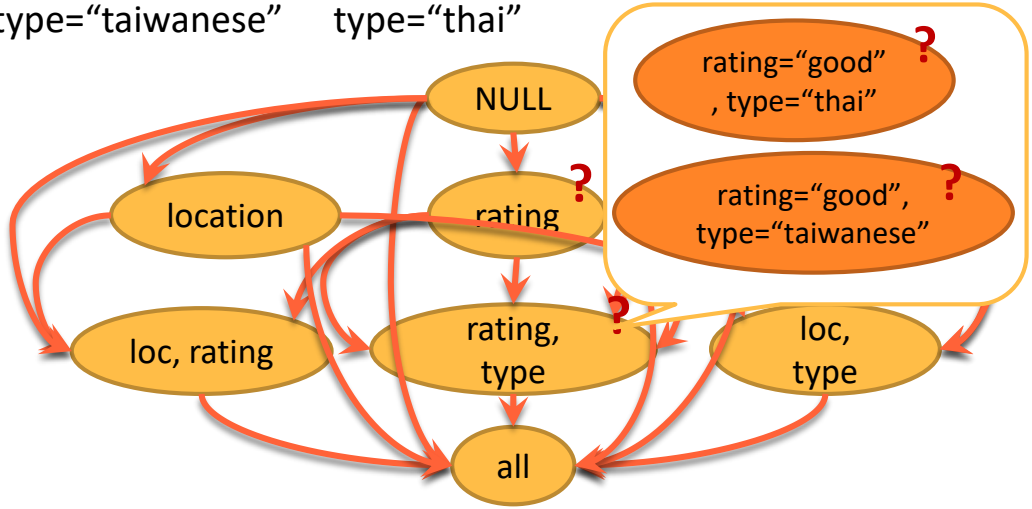
find a good eating place for taixxxx food

FIND_RESTAURANT
rating="good"
type="taiwanese"

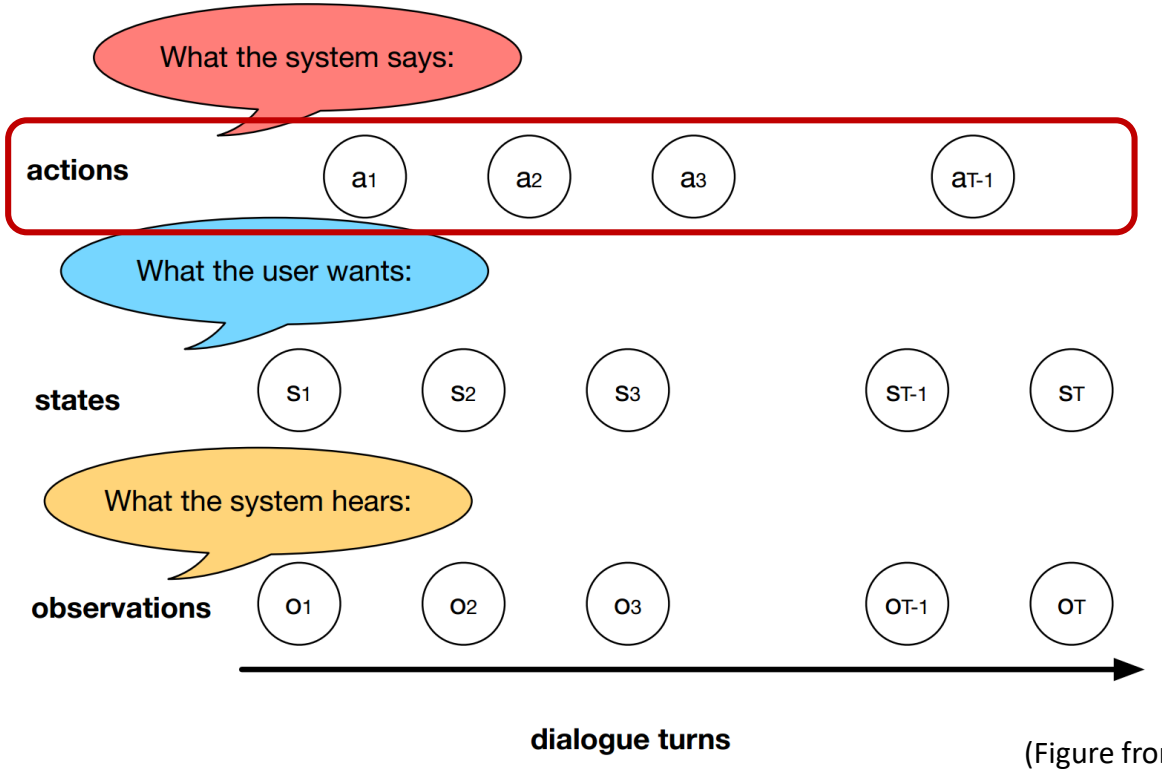
FIND_RESTAURANT
rating="good"
type="thai"

FIND_RESTAURANT
rating="good"

Intelligent Agent



Elements of Dialogue Management

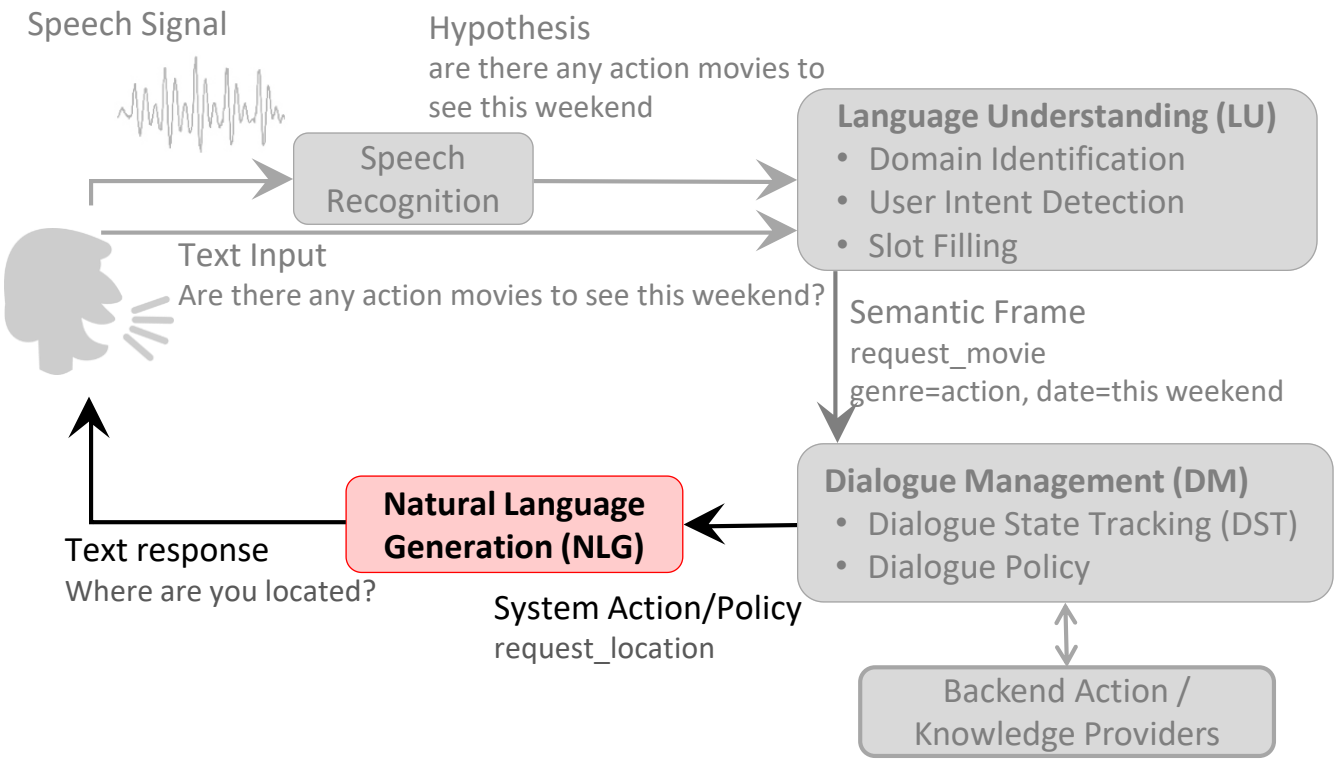


(Figure from Gašić)

Dialogue Policy for Agent Action

- Inform(location=“Taipei 101”)
  - “The nearest one is at Taipei 101”
- Request(location)
  - “Where is your home?”
- Confirm(type=“taiwanese”)
  - “Did you want Taiwanese food?”

Task-Oriented Dialogue System (Young, 2000)



Output / Natural Language Generation

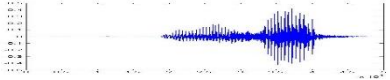
- Goal: generate natural language or GUI given the selected dialogue action for interactions

- Inform(location=“Taipei 101”)
  - “The nearest one is at Taipei 101” v.s.
- Request(location)
  - “Where is your home?” v.s.
- Confirm(type=“taiwanese”)
  - “Did you want Taiwanese food?” v.s.

Deep Learning for Dialogue Systems

Machine Learning ≈ Looking for a Function

- Speech Recognition

$f($  $)=$ “你好 (Hello) ”

- Image Recognition

$f($  $)=$ cat

- Go Playing

$f($  $)=$ 5-5 (next move)

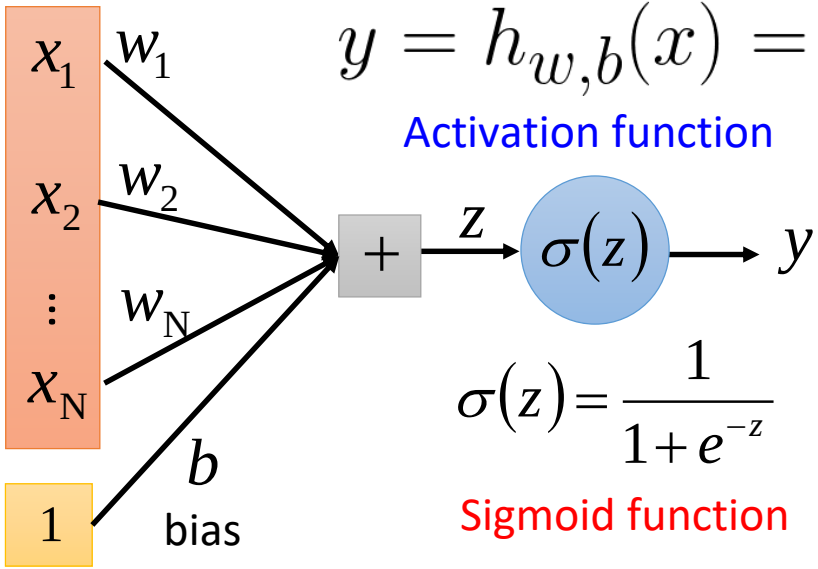
- Chat Bot

$f($ “Where is GTC?” $)=$ “The address is...”

A Single Neuron

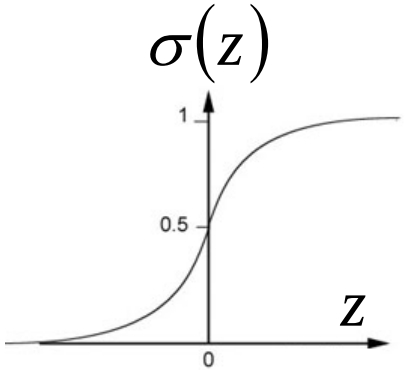
$$y = h_{w,b}(x) = \sigma(w^T x + b)$$

Activation function



$$\sigma(z) = \frac{1}{1+e^{-z}}$$

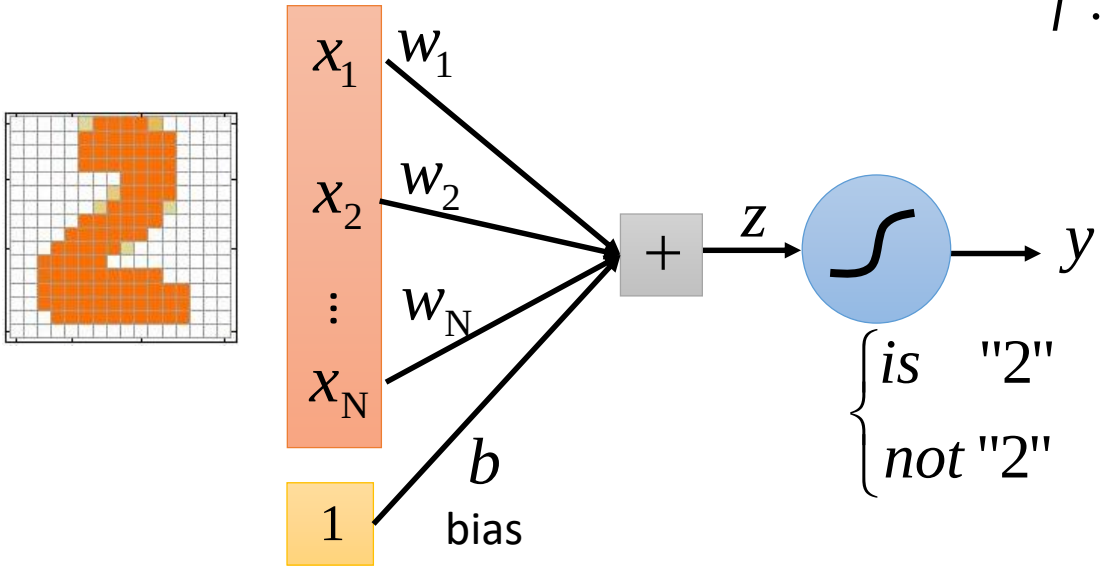
Sigmoid function



w, b are the parameters of this neuron

A Single Neuron

$$f : R^N \rightarrow R^M$$



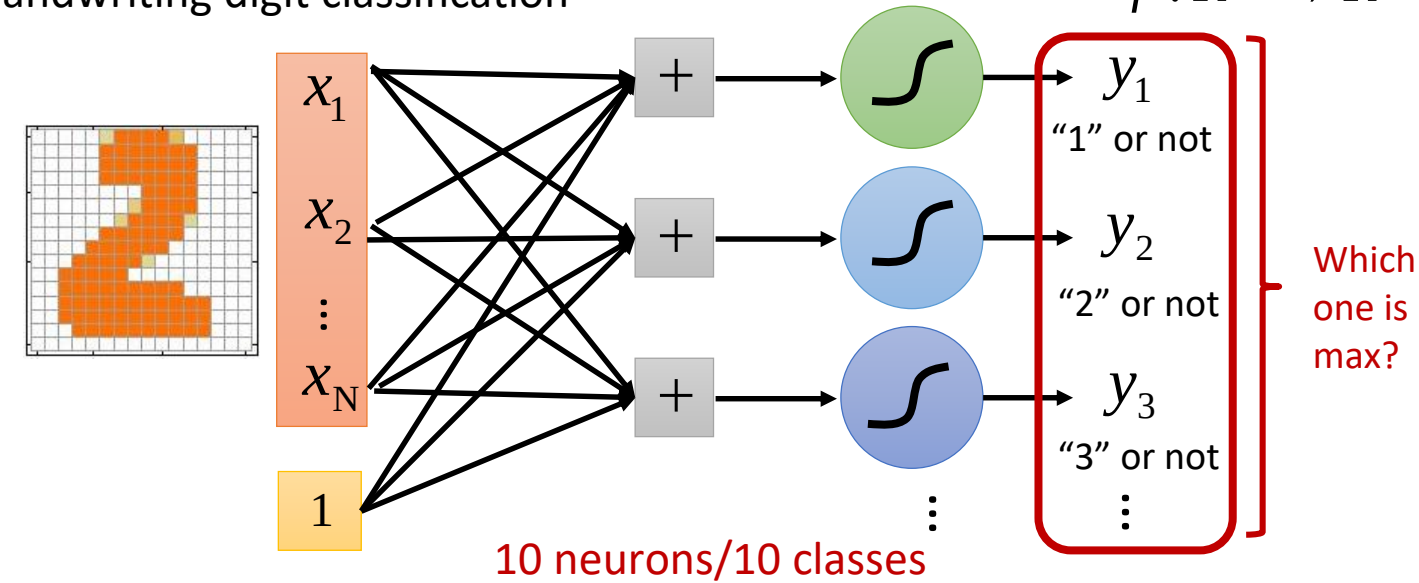
$$\begin{cases} is \quad "2" & y \geq 0.5 \\ not\ "2" & y < 0.5 \end{cases}$$

A single neuron can only handle binary classification

A Layer of Neurons

- Handwriting digit classification

$$f : R^N \rightarrow R^M$$

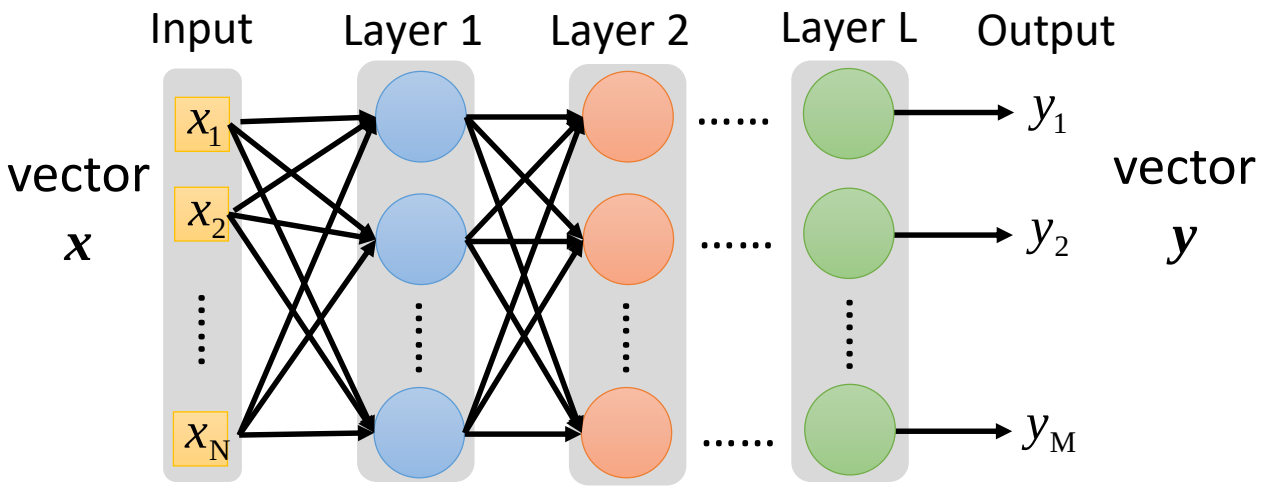


A layer of neurons can handle multiple possible output, and the result depends on the max one

Deep Neural Networks (DNN)

- Fully connected feedforward network

$$f : R^N \rightarrow R^M$$

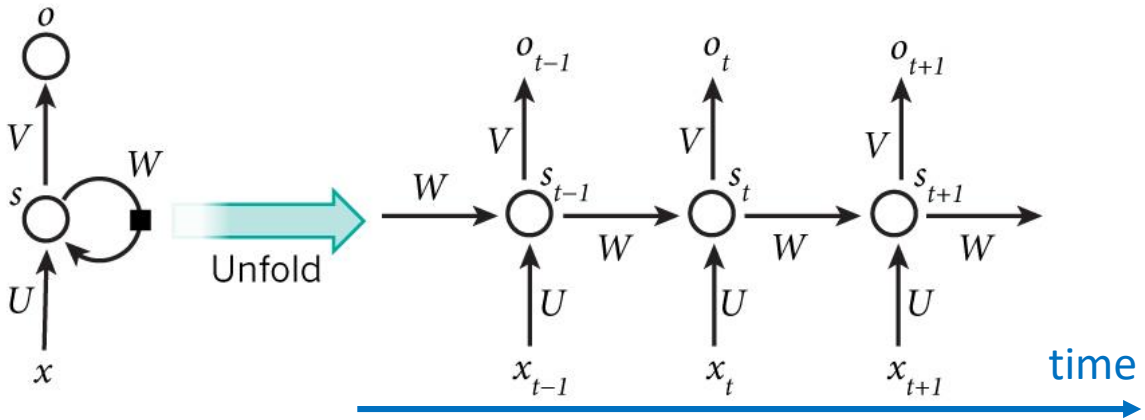


Deep NN: multiple hidden layers

Recurrent Neural Network (RNN)

$$s_t = \sigma(W s_{t-1} + U x_t) \qquad \sigma(\cdot)\text{: tanh, ReLU}$$

$$o_t = \mathrm{softmax}(V s_t)$$

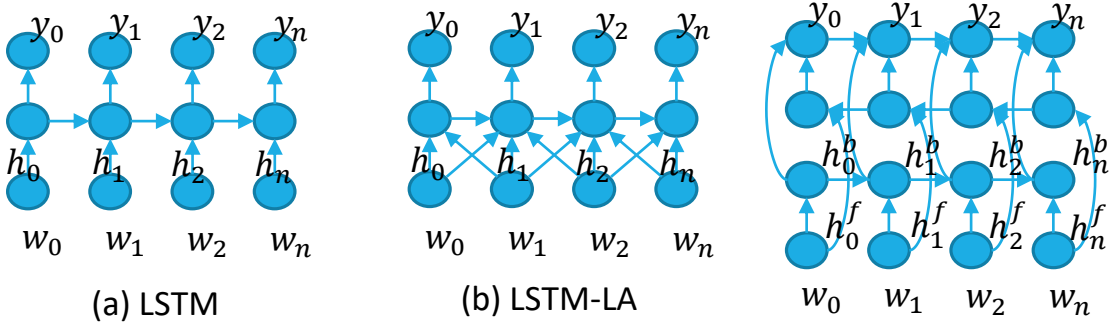


RNN can learn accumulated sequential information (time-series)

http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/

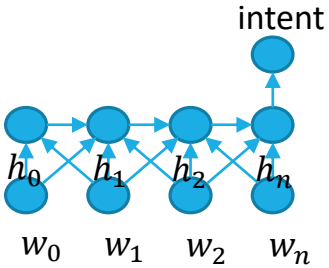
Deep Learning for LU

- IOB Sequence Labeling for Slot Filling



(a) LSTM (b) LSTM-LA (c) bLSTM

- Intent Classification



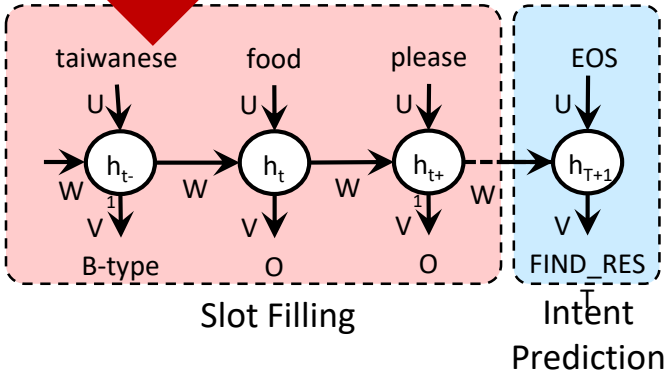
(d) Intent LSTM

Joint Semantic Frame Parsing

https://www.microsoft.com/en-us/research/wp-content/uploads/2016/06/IS16_MultiJoint.pdf; https://arxiv.org/abs/1609.01454

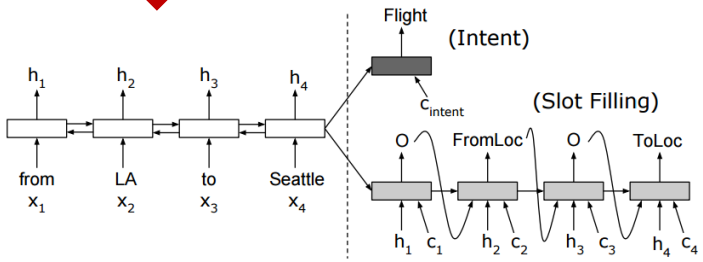
Sequence-based (Hakkani-Tur et al., 2016)

- Slot filling and intent prediction in the same output sequence



Parallel (Liu and Lane, 2016)

- Intent prediction and slot filling are performed in two branches



Contextual LU

Domain Identification → Intent Prediction → Slot Filling

D communication ***I*** send_email

Single Turn

U	just	sent	email	to	bob	about	fishing	this	weekend
S	O	O	O	O	B-contact_name	O	B-subject	I-subject	I-subject

→ send_email(contact_name="bob", subject="fishing this weekend")

Multi-Turn

U_1 send email to bob

S_1 B-contact_name

→ send_email(contact_name="bob")

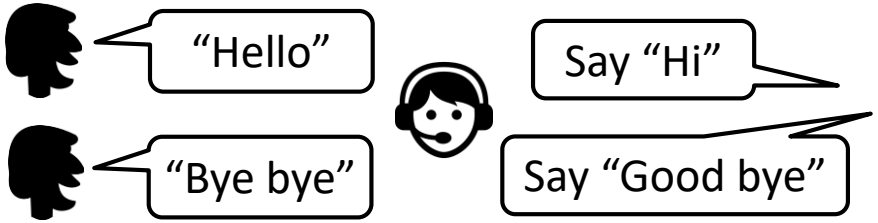
U_2	are	we	going	to	fish	this	weekend
S_2	B-message	I-message	I-message	I-message	I-message	I-message	I-message

→ send_email(message="are we going to fish this weekend")

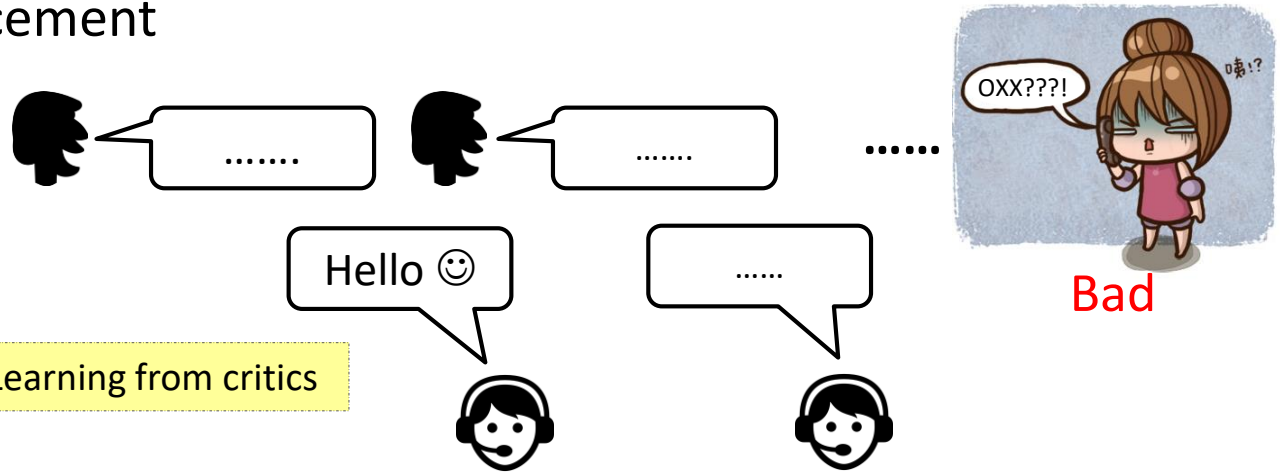
Supervised v.s. Reinforcement

- Supervised

Learning from teacher



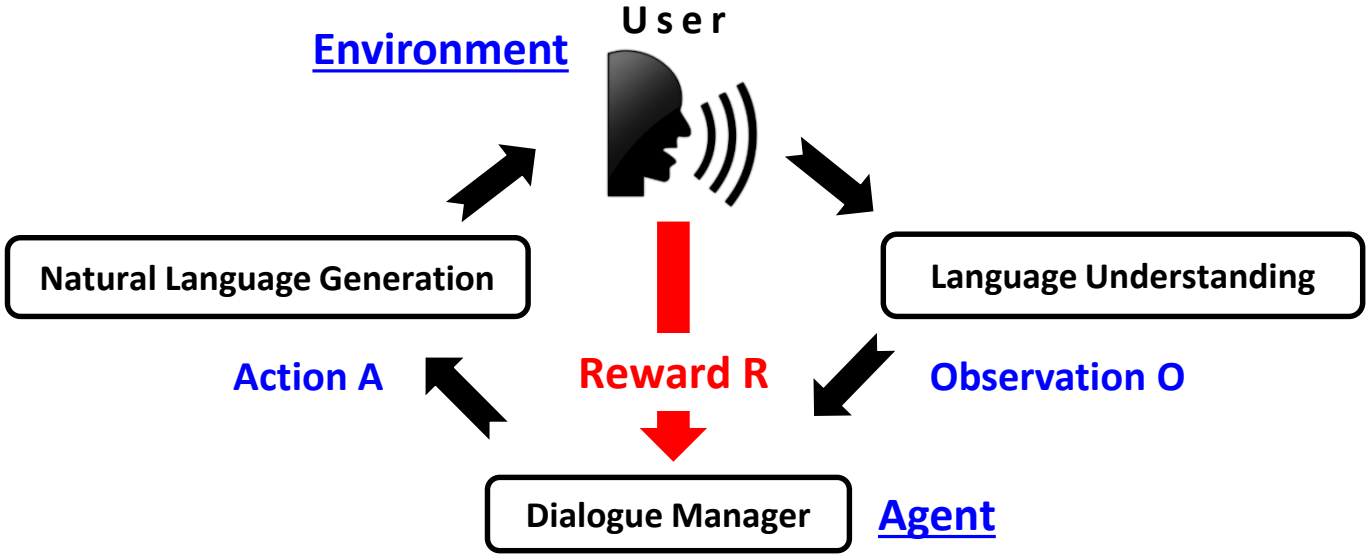
- Reinforcement



Learning from critics

Dialogue Policy Optimization

- Dialogue management in a RL framework



The optimized dialogue policy selects the best action that **maximizes the future reward**

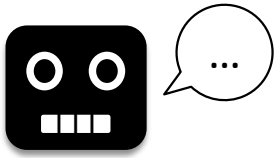
Dialogue Reinforcement Learning Signal

Typical reward function

- -1 for per turn penalty
- Large reward at completion if successful

Typically requires domain knowledge

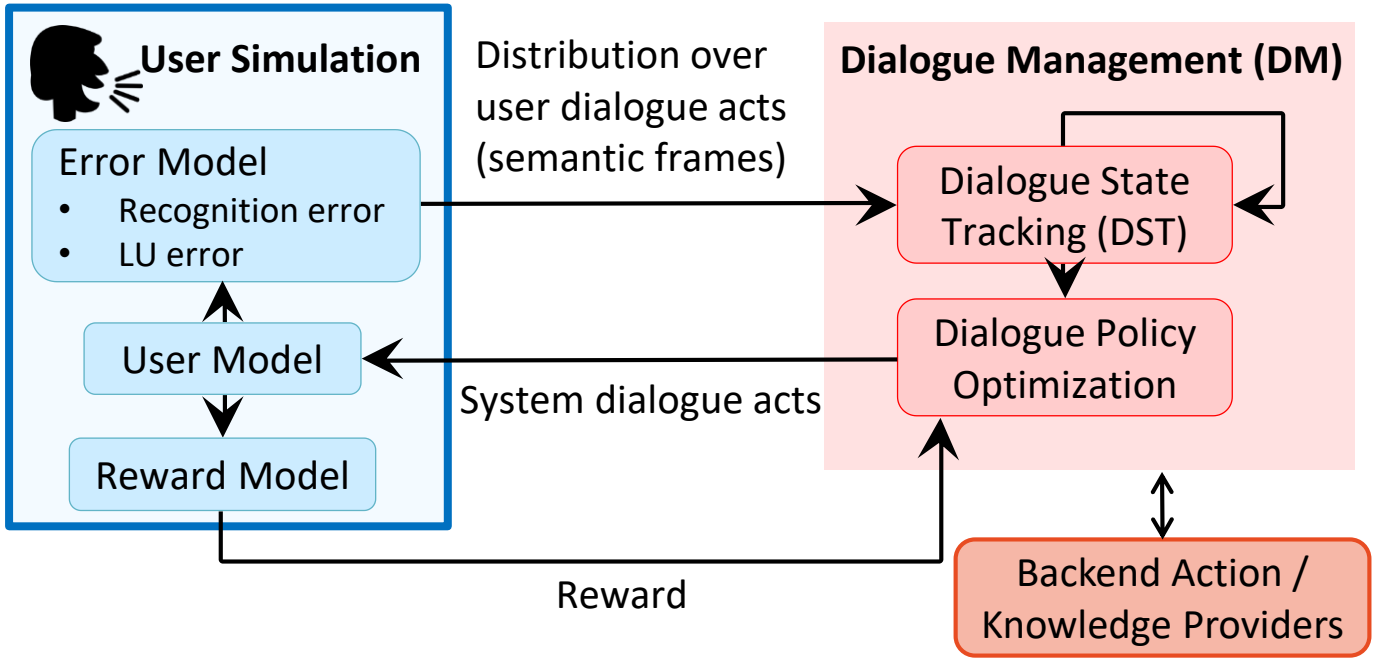
- ✔ Simulated user
- ✔✔ Paid users (Amazon Mechanical Turk)
- ✘ Real users



The user simulator is usually required for dialogue system training before deployment

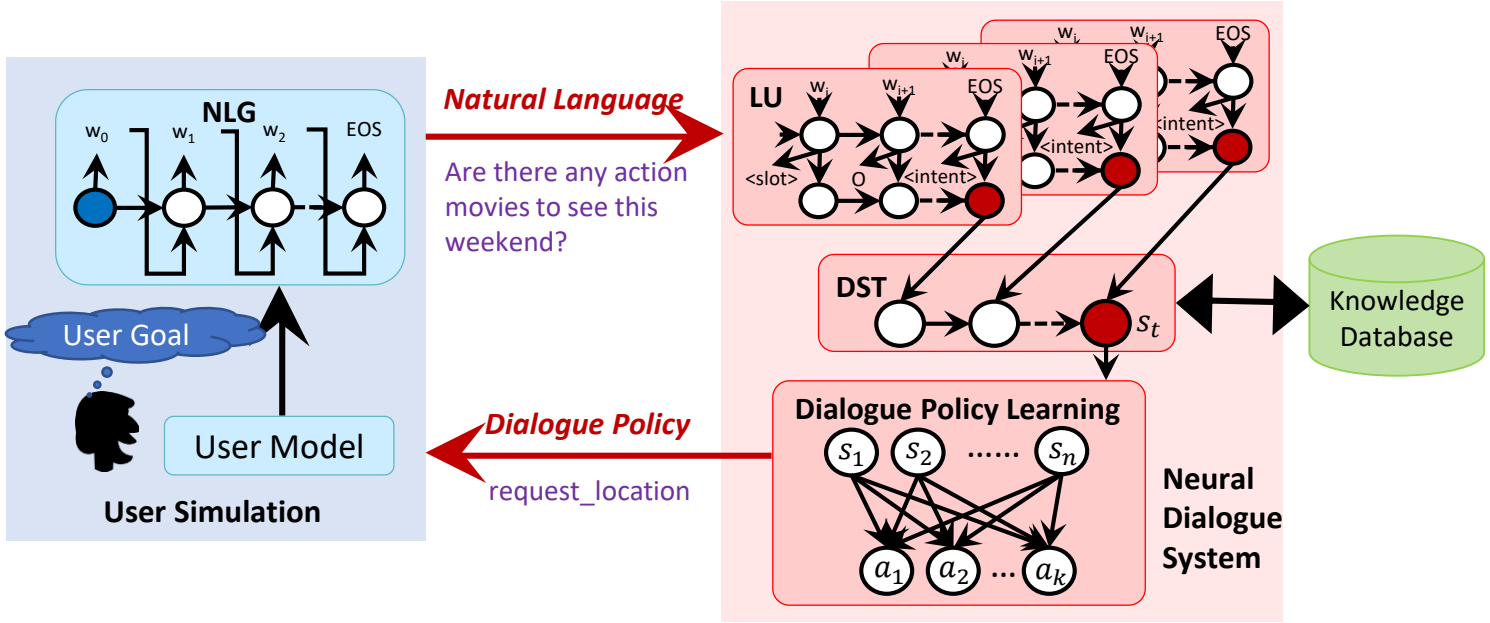
Learning from Environments

- Solution: learn from a simulated user



E2E Task-Completion Bot (TC-Bot) (Li et al., 2017)

https://arxiv.org/abs/1703.01008



Idea: SL for each component and RL for end-to-end training the neural dialogue system

E2E Task-Completion Bot (TC-Bot) (Li et al., 2017)

https://arxiv.org/abs/1703.01008

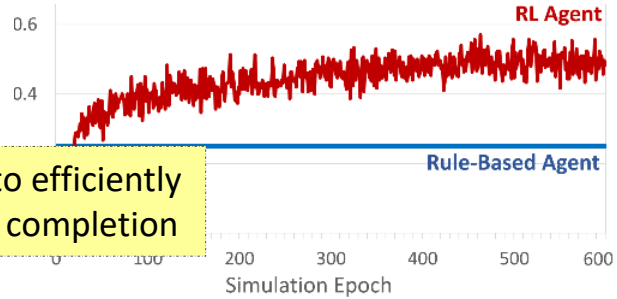
- User goal: ***Two tickets for "the witch" tomorrow 9:30 PM at regal meridian 16, Seattle.***

RULE BASED SYSTEM

User: Can I get 2 tickets for the witch?
Agent: Which theater do you want?
User: Regal meridian 16
Agent: What time would you like to see it?
User: 9:30 pm
Agent: What date would you like to see it?
User: Tomorrow!
Agent: How many people are you?
User: Just two adults.
Agent: What is your location?
User: Seattle!
Agent: What kind of video format do you like?
User: Oh, any kind is okay.
Agent: Do you have any price constraint?
User: No, I don't care.
Agent: Great - I was able to purchase 2 tickets for you to see the witch tomorrow at regal meridian 16 theater in Seattle at 9:30 pm.
User: Thank you.
Agent: Thank you!

REINFORCEMENT LEARNING SYSTEM

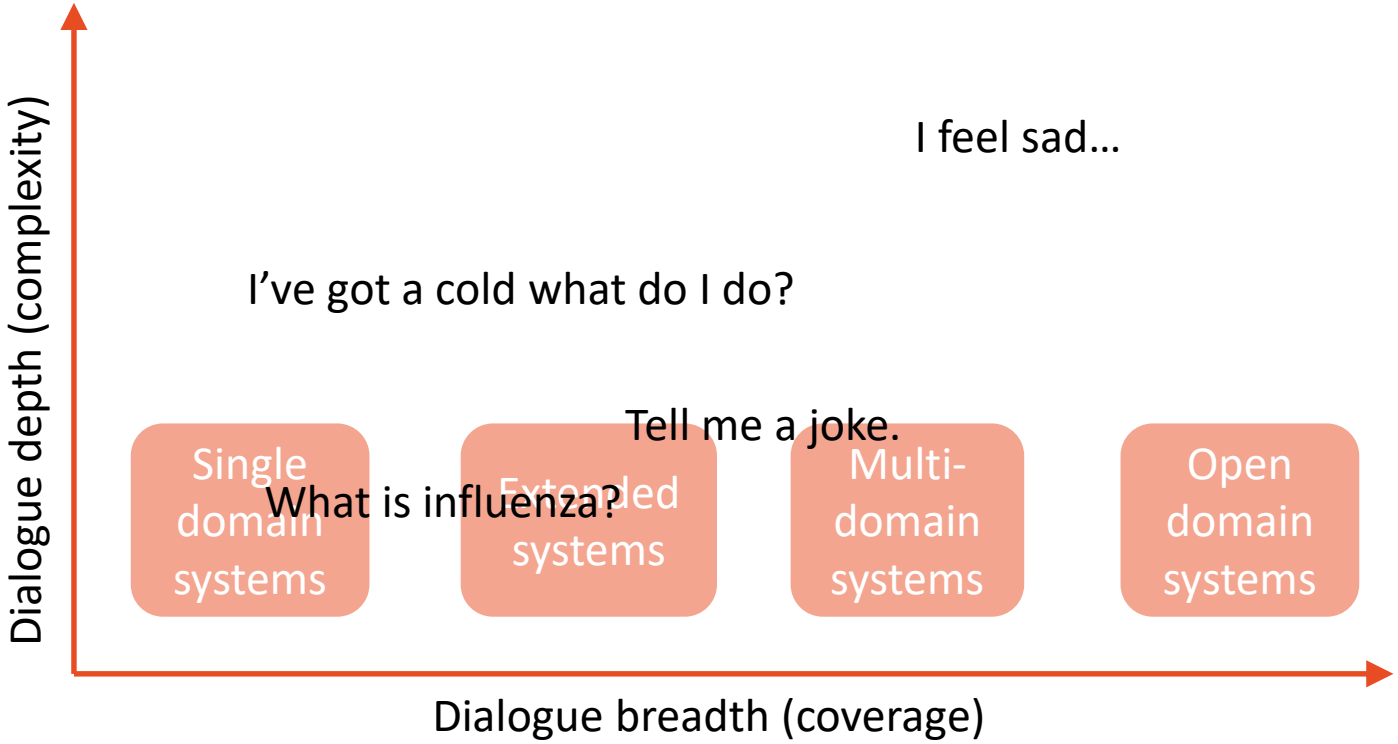
User: Can I get 2 tickets for the witch?
Agent: What date would you like to see it?
User: Tomorrow
Agent: Which theater do you prefer?
User: Regal meridian 16
Agent: What is your location?
User: Seattle
Agent: What time would you like to see it?
User: 9:30 pm.
Agent: Great - I was able to purchase 2 tickets for you to see the witch tomorrow at regal meridian 16 theater in Seattle at 9:30 pm.
User: Thanks.
Agent: Thanks!



The system can learn how to efficiently interact with users for task completion

Recent Trends on Learning Dialogues

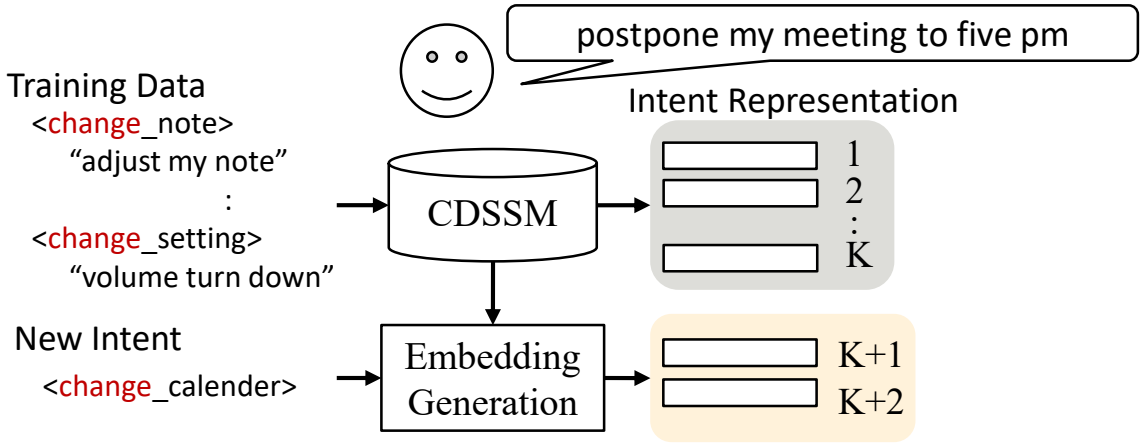
Evolution Roadmap



Intent Expansion (Chen et al., 2016)

http://ieeexplore.ieee.org/abstract/document/7472838/

- Transfer dialogue acts across domains
  - Dialogue acts are similar for multiple domains
  - Learning new intents by information from other domains

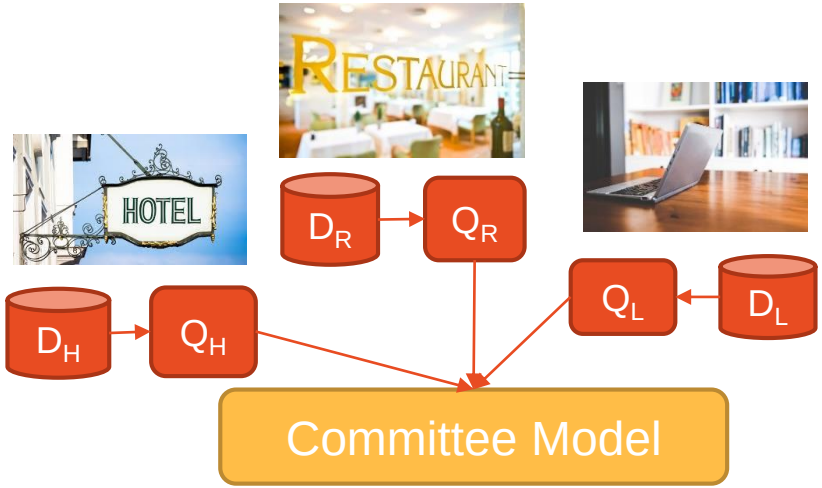


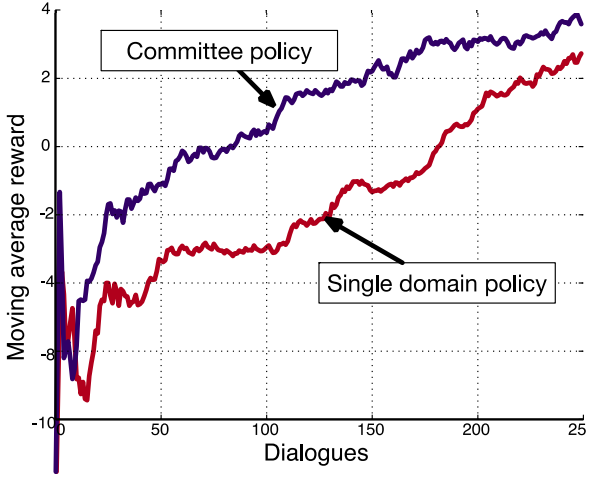
The dialogue act representations can be automatically learned for other domains

Policy for Domain Adaptation (Gašić et al., 2015)

http://ieeexplore.ieee.org/abstract/document/7404871/

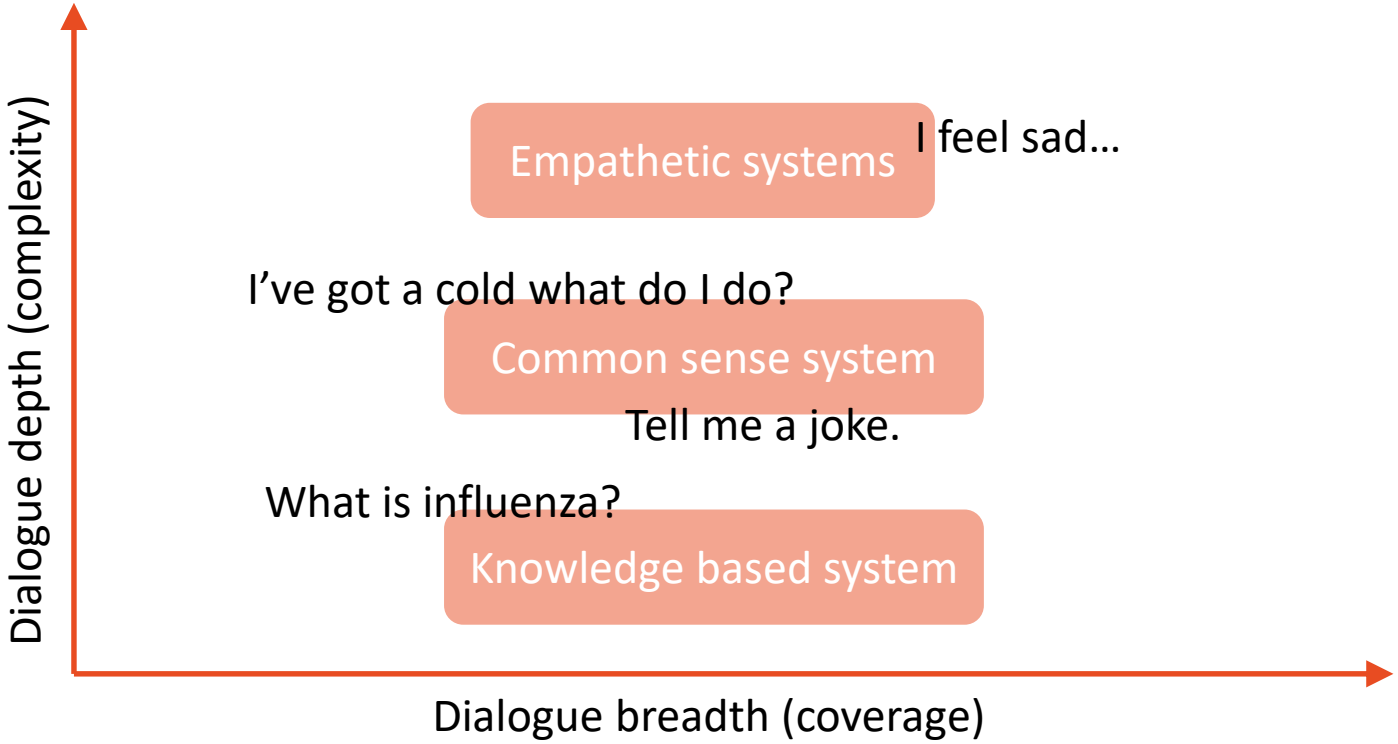
- Bayesian committee machine (BCM) enables estimated Q-function to share knowledge across domains





The policy from a new domain can be boosted by the committee policy

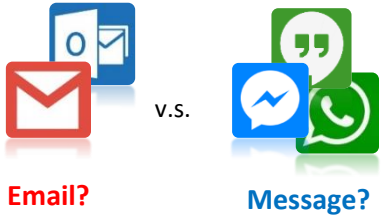
Evolution Roadmap



App Behavior for Understanding

http://dl.acm.org/citation.cfm?id=2820781

- Task: user intent prediction
- Challenge: language ambiguity



① **User preference**
- ✓ Some people prefer “Message” to “Email”
- ✓ Some people prefer “Ping” to “Text”

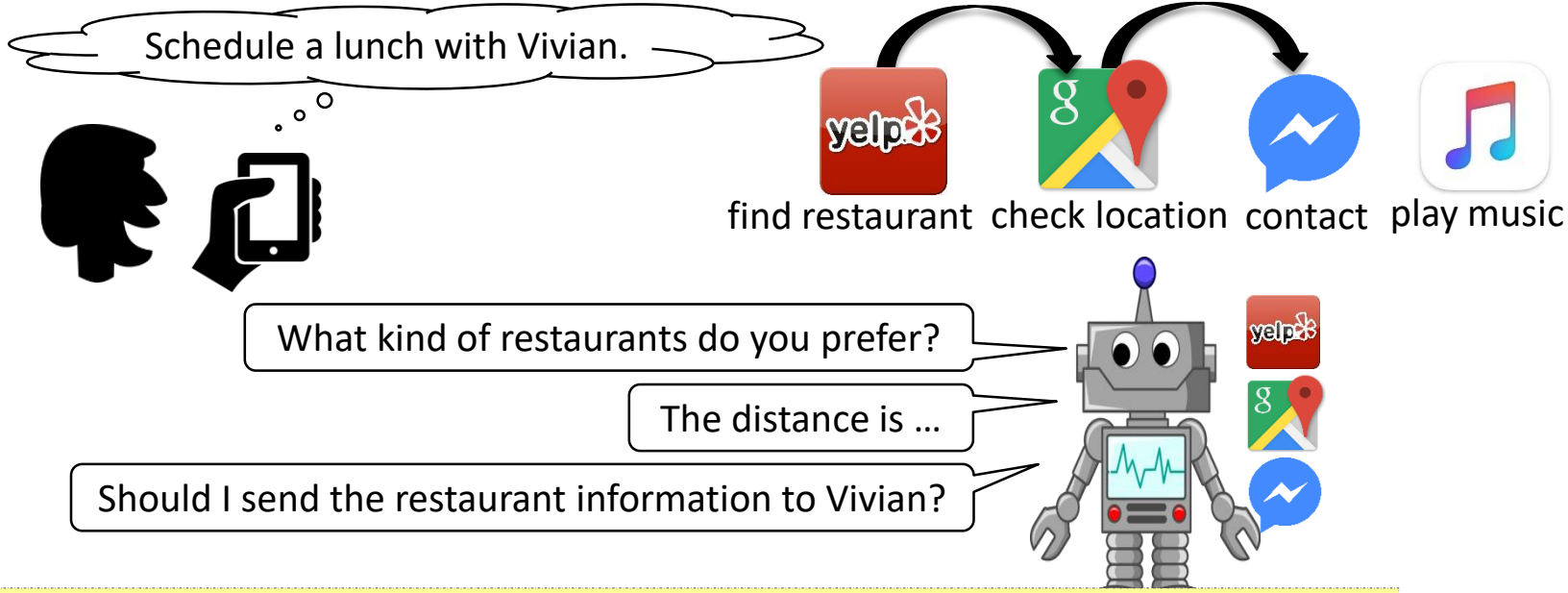
② **App-level contexts**
- ✓ “Message” is more likely to follow “Camera”
- ✓ “Email” is more likely to follow “Excel”

Considering behavioral patterns in history to model understanding for intent prediction.

High-Level Intention for Dialogue Planning (Sun et al., 2016)

http://dl.acm.org/citation.cfm?id=2856818; http://www.lrec-conf.org/proceedings/lrec2016/pdf/75_Paper.pdf

- High-level intention may span several domains

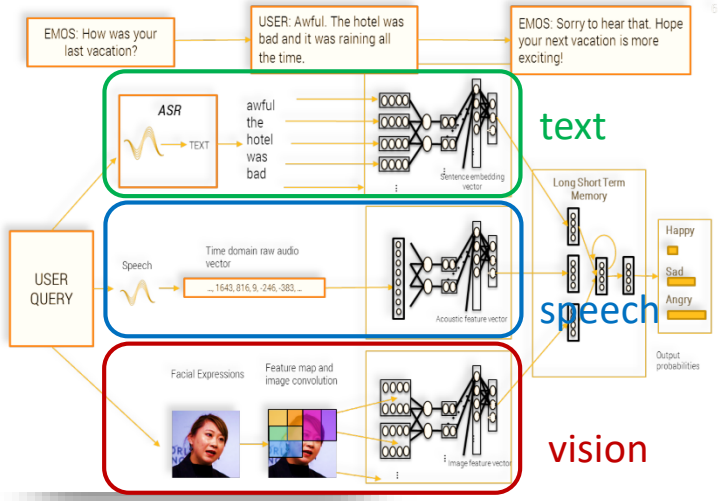


Users can interact via <u>high-level descriptions</u> and the system learns <u>how to plan the dialogues</u>

Empathy in Dialogue System (Fung et al., 2016)

https://arxiv.org/abs/1605.04072

- Embed an empathy module
  - Recognize emotion using multimodality
  - Generate emotion-aware responses



Emotion Recognizer


```
Face recognition output                                  (index):1728
                                                         (index):1729
 {
  "recognition": "Race: Asian Confidence: 65.42750000000001 Smiling:
3.95896 Gender: Female Confidence: 88.9369",
  "race": "Asian",
  "race_confidence": "65.42750000000001",
  "smiling": "3.95896",
  "gender": "Female",
  "gender_confidence": "88.9369"
 }
```

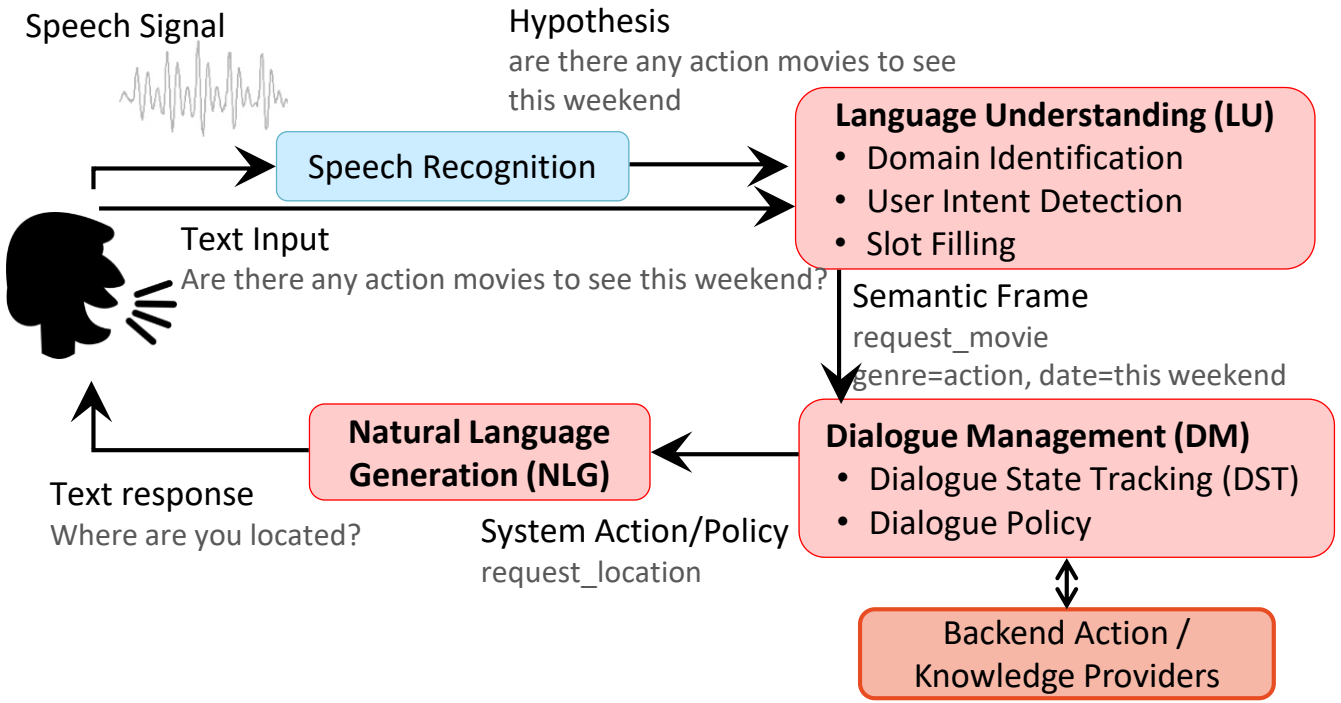
Challenges and Conclusions

Challenge Summary

- The human-machine interface is a hot topic but several components must be integrated!
- Most state-of-the-art technologies are based on DNN
  - Requires huge amounts of labeled data
  - Several frameworks/models are available
- Fast domain adaptation with scarse data + re-use of rules/knowledge
- Handling reasoning
- Data collection and analysis from un-structured data
- Complex-cascade systems requires high accuracy for working good as a whole

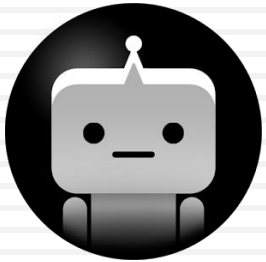
Concluding Remarks

- Modular dialogue system



Thanks for Your Attention!

Q & A



Yun-Nung (Vivian) Chen
Assistant Professor
National Taiwan University
y.v.chen@ieee.org / http://vivianchen.idv.tw