Conversational AI at a glance

Mohammad Mahdi Samiei

Sharif University of Technology

Oct 15th, 2019

What we’re about to discuss

- Introduction
- Task oriented dialogue systems
- Fully data-driven conversation models (chatbots)
- Conclusions and Research trends
- How to start?

Introduction:

What is a dialogue system?!?

Introduction:

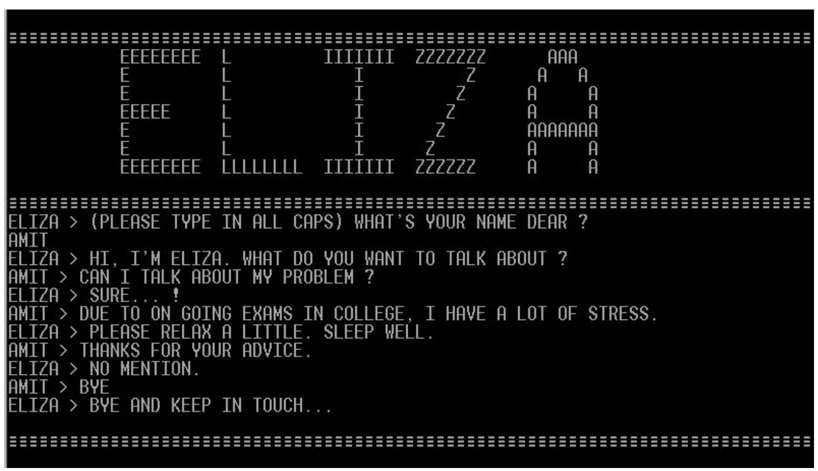
What is a dialogue system?!?

- Emulating human conversation
- Answering questions on a wide range of topics
- Fulfilling complex tasks

Want to create a dialogue system?!?

Want to create a dialogue system?!?

Want to create a dialogue system?!?

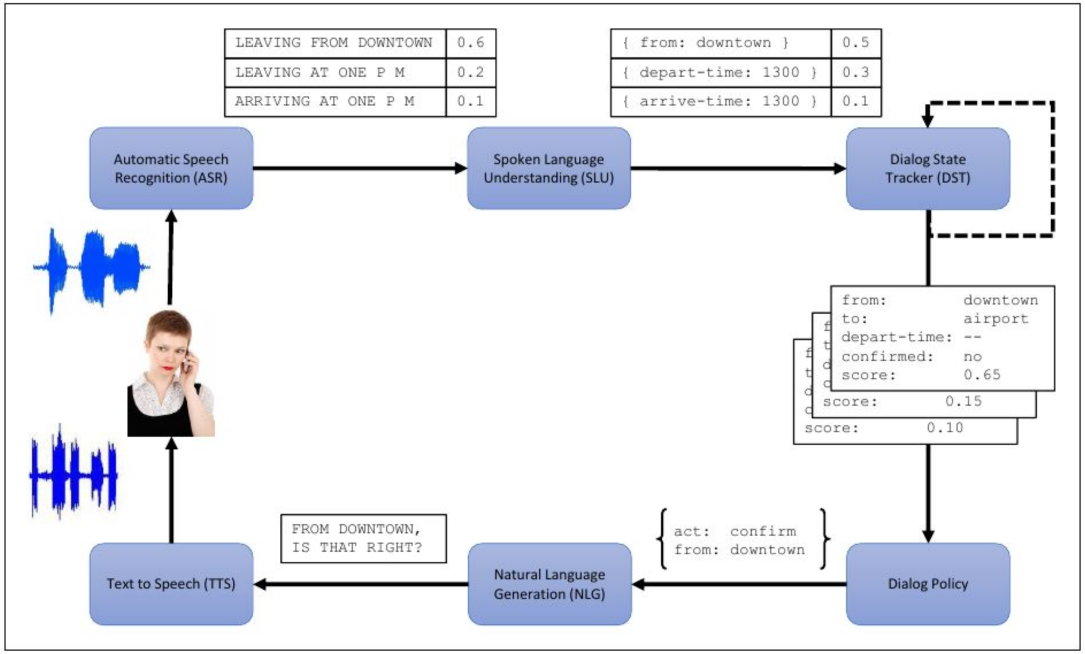


Dialouge: What Kinds of Problems?

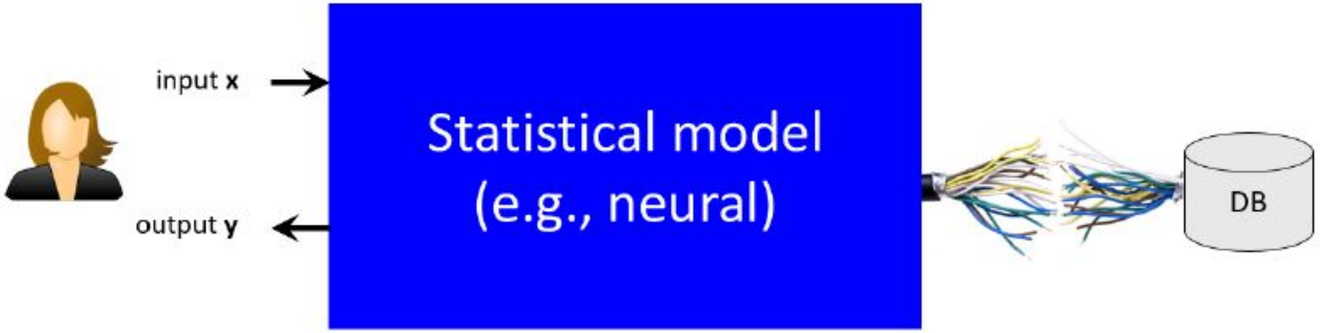
1. Question Answering
2. Task Completion
3. Social Chat

Which architecture do you choose?

Modular



Fully data-driven



E2E

What do you need to learn?!?

- 🔺 NLP
- 🔺 ML
  - 🔻 In particular DL
  - 🔻 RL can help
- 🔺 IR
- 🔺 Speech

Part two:

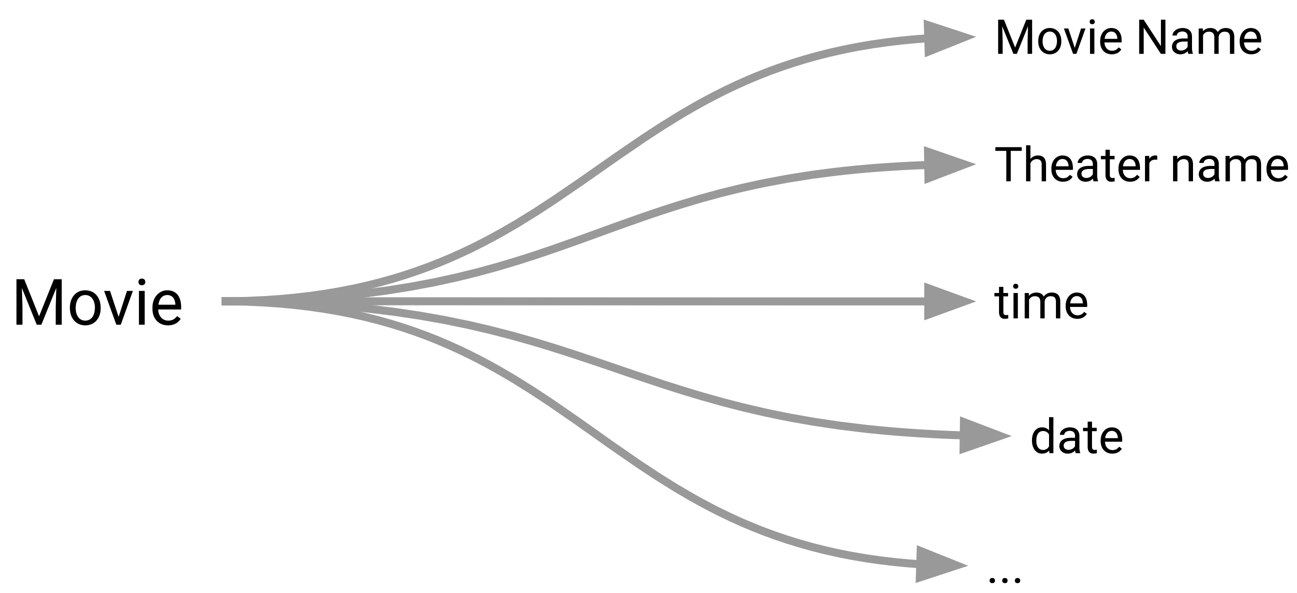
Task-based dialogue system

An Overview

- ↘️ Why task-based dialogue system?
  - ↪️ We are tired!
  - ↪️ Conversation is a tool, not goal maybe.
  - ↪️ The faster the better!!
- ↘️ Example:
  - ↪️ Siri
  - ↪️ Alexa
  - ↪️ Cortana
  - ↪️ Google Now/Home

Slot filling Dialogues

- Simplest model
- Collect necessary information
- For each domain set of **slot** defined by experts

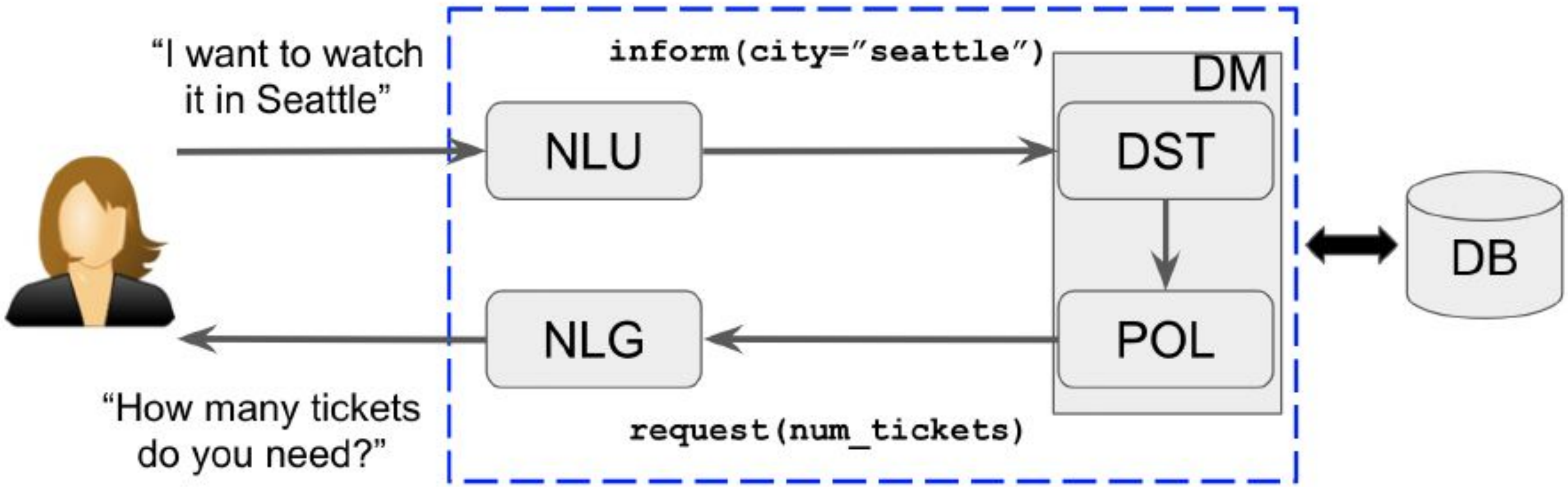


Dialogue Act

- Dialogue Act Theory: Interaction between agent and user
- RL viewpoint
  - Dialogue system is agent
  - Human is environment
  - Utterances are actions that can change state (api calls also can be)
- Dialogue Act: Semantic between Natural Language and Action
- Some dialogue acts may have slot as argument

How many tickets do you need?	request(num_tickets)
I want to watch it in Seattle	infrom(city="seattle")
What is the phone number?	request(phone_number)

Modular Architecture



Natural Language Understanding	Converts the user utterances to dialogue act
Dialogue Manager	Central controller, keep track of dialogue, select next action
Natural Language Generation	Convert the selected dialogue action to natural language

NLU

Three issues:

- domain classification
- intent determination
- slot tagging

Example: *Show me morning flights from Boston to San Francisco on Tuesday*

DOMAIN	AIR-TRAVEL
INTENT	SHOW-FLIGHTS
ORIGIN-CITY	Boston
ORIGIN-DATE	Tuesday
ORIGIN-TIME	morning
DEST-CITY	San Francisco

Dialogue State Tracking

- It contains all information about what users is looking for at the current turn.
- Input for dialogue policy

User: I'm looking for a cheaper restaurant
inform(price=cheap)
System: Sure. What kind - and where?
User: Thai food, somewhere downtown
inform(price=cheap, food=Thai, area=centre)
System: The House serves cheap Thai food
User: Where is it?
inform(price=cheap, food=Thai, area=centre); request(address)
System: The House is at 106 Regent Street

Dialogue Policy

- Decide what action the system should take next
  - Dialogue act
  - Api call

$$\hat{A}_i = \operatorname*{argmax}_{A_i \in A} P(A_i | (A_1, U_1, \ldots, A_{i-1}, U_{i-1}))$$

- Use RL
  - DQN whose entries correspond to all possible (dialogue-act, slot)
- More challenge:
  - Multi-domain dialogue
  - Composite-tasks

Task-based dialogue system

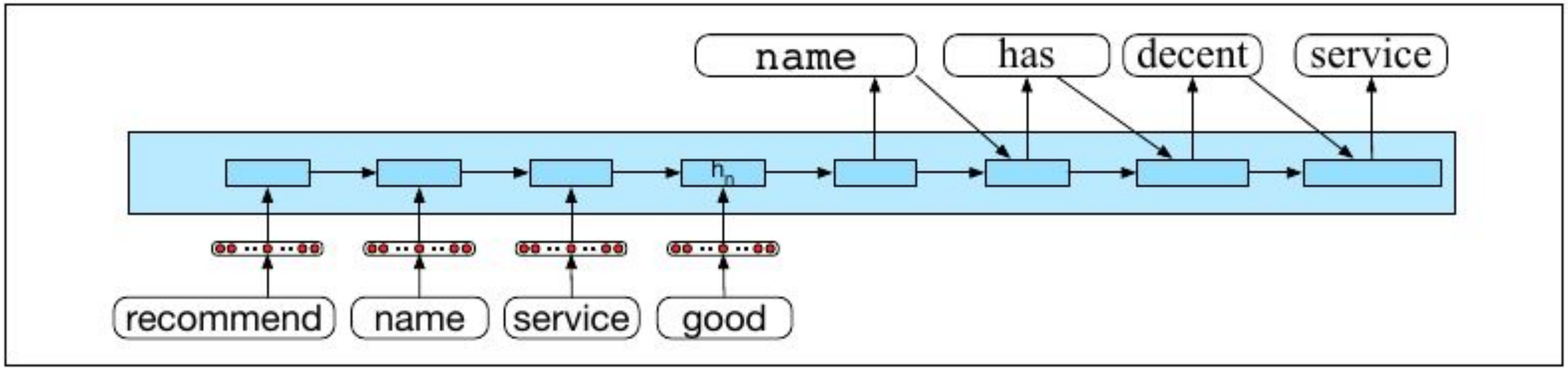
NLG

Rule-based methods:

Delexicalization

recommend(restaurant name= Au Midi, neighborhood = midtown, cuisine = french)
There is a *cuisine* restaurant in *neighborhood* called *restaurant_name*.
There is a *French* restaurant in *Midtown* called *Au Midi*.

Neural Methods:



Evaluating

- We need to evaluate the whole of system
- Evaluation Metrics:
  - Task completion success
  - Time elapsed
- Simulation based evaluation:
  - Simulated user
    - Agenda based
    - Model Based

End to End learning

- Why modularity?
  - Flexibility to build module in independent way
- Why E2E?
  - Complex design of modular
  - Improve in individual module do not necessarily translate into improvement of whole system
- Supervised-Learning
- Reinforcement-Learning
  - Training data may expensive to obtain
  - exploration

Recent Frontier

- Beyond slot filling dialogues
  - Tutorial
  - Transaction
  - Negotiation
  - Multi-modal
- Weaker learning signal
  - Instead of having absolute judgement, one only requires a preferential input that indicates which one of two dialogues is better
- Related Areas
  - User simulation
  - Text games
  - ...

Fully data-driven conversation and chatbots

Fully data-driven conversation

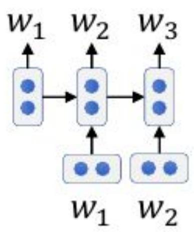
introduction

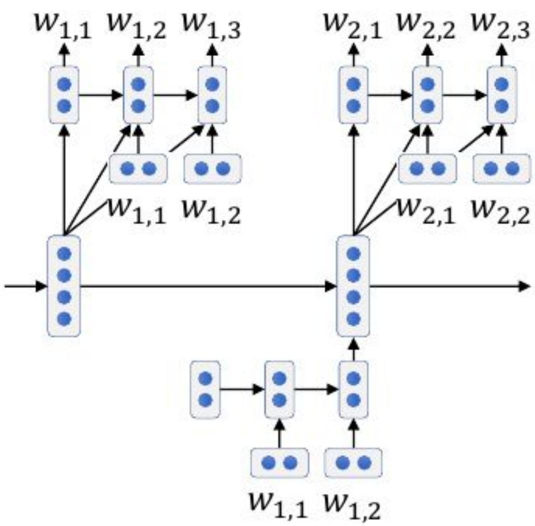
- Trained entirely from data
- Without resorting to any expert knowledge
- Do not rely on four traditional components
- Easily scales to large free-from and open-domain datasets
- Conversatin is a goal!
- Scenarios
  - Chitchat
  - Recommandation
  - ...

Fully data-driven conversation

Conversation Models

- Inspired by statistical machine translation
- Limitation: Bad look at data, contextually appropriate response
- RNN architectures
  - LSTM models
  - HRED models
  - Attention models
  - Pointer-Network models
    - Copy-net





Fully data-driven conversation

A Neural Conversation Model, 2015

Human: *hello !*
Machine: *hello !*
Human: *how are you ?*
Machine: *i 'm good .*
Human: *what 's your name ?*
Machine: *i 'm julia .*
Human: *when were you born ?*
Machine: *july 20th .*
Human: *what year were you born ?*
Machine: *1977 .*
Human: *where are you from ?*
Machine: *i 'm out in the boonies.*
Human: *are you a man or a woman ?*
Machine: *i 'm a woman .*
Human: *why are we here ?*
Machine: *i 'm not sure .*
Human: *okay, bye !*
Machine: *bye .*
Human: *see you later .*
Machine: *bye .*

Human: *my name is david . what is my name ?*
Machine: *david .*
Human: *my name is john . what is my name ?*
Machine: *john .*
Human: *are you a leader or a follower ?*
Machine: *i 'm a leader .*
Human: *are you a follower or a leader ?*
Machine: *i 'm a leader .*

Fully data-driven conversation

A Neural Conversation Model, 2015

Human: *who is skywalker ?*
Machine: *he is a hero .*
Human: *who is bill clinton ?*
Machine: *he 's a billionaire .*
Human: *is sky blue or black ?*
Machine: *blue .*
Human: *does a cat have a tail ?*
Machine: *yes .*
Human: *does a cat have a wing ?*
Machine: *no*
Human: *can a cat fly ?*
Machine: *no .*
Human: *how many legs does a cat have ?*
Machine: *four , i think .*
Human: *how many legs does a spider have ?*
Machine: *three , i think .*
Human: *how many legs does a centipede have ?*
Machine: *eight .*
Human: *what is the color of the sky ?*
Machine: *blue .*
Human: *what is the color of water ?*
Machine: *water .*
Human: *what is the color of blood ?*
Machine: *it is the same as a black eye .*
Human: *what is the usual color of a leaf ?*
Machine: *it is a green one .*
Human: *what is the color of a yellow car ?*
Machine: *yellow .*
Human: *how much is two plus two ?*
Machine: *four .*

Human: *what is the purpose of life ?*
Machine: *to serve the greater good .*
Human: *what is the purpose of living ?*
Machine: *to live forever .*
Human: *what is the purpose of existence ?*
Machine: *to find out what happens when we get to the planet earth .*
Human: *where are you now ?*
Machine: *i 'm in the middle of nowhere .*
Human: *what is the purpose of dying ?*
Machine: *to have a life .*
Human: *what is the purpose of being intelligent ?*
Machine: *to find out what it is .*
Human: *what is the purpose of emotions ?*
Machine: *i don 't know .*

Fully data-driven conversation

A Neural Conversation Model, 2015

Problems begin!!!!

Conversation 7: Job and Personality

Human: *what is your job ?*
Machine: *i 'm a lawyer .*
Human: *what do you do ?*
Machine: *i 'm a doctor .*

Human: *how much is ten minus two ?*
Machine: *seventy - two .*

Fully data-driven conversation

Challenges: Blandness

- Response Blandeness
  - Bland and uninformative
  - “I don’t know”!!! Or “Ok”
- How to solve this problem?

$$p(T|S) = \frac{P(S|T) * P(T)}{P(S)} \longrightarrow \frac{P(S,T)}{P(S) * P(T)}$$

Hard to optimize! So only use in inference time

$$\text{argmax}_T \left\{ (1-\lambda) \log p(T|S) + \lambda \log p(S|T) \right\}.$$

Fully data-driven conversation

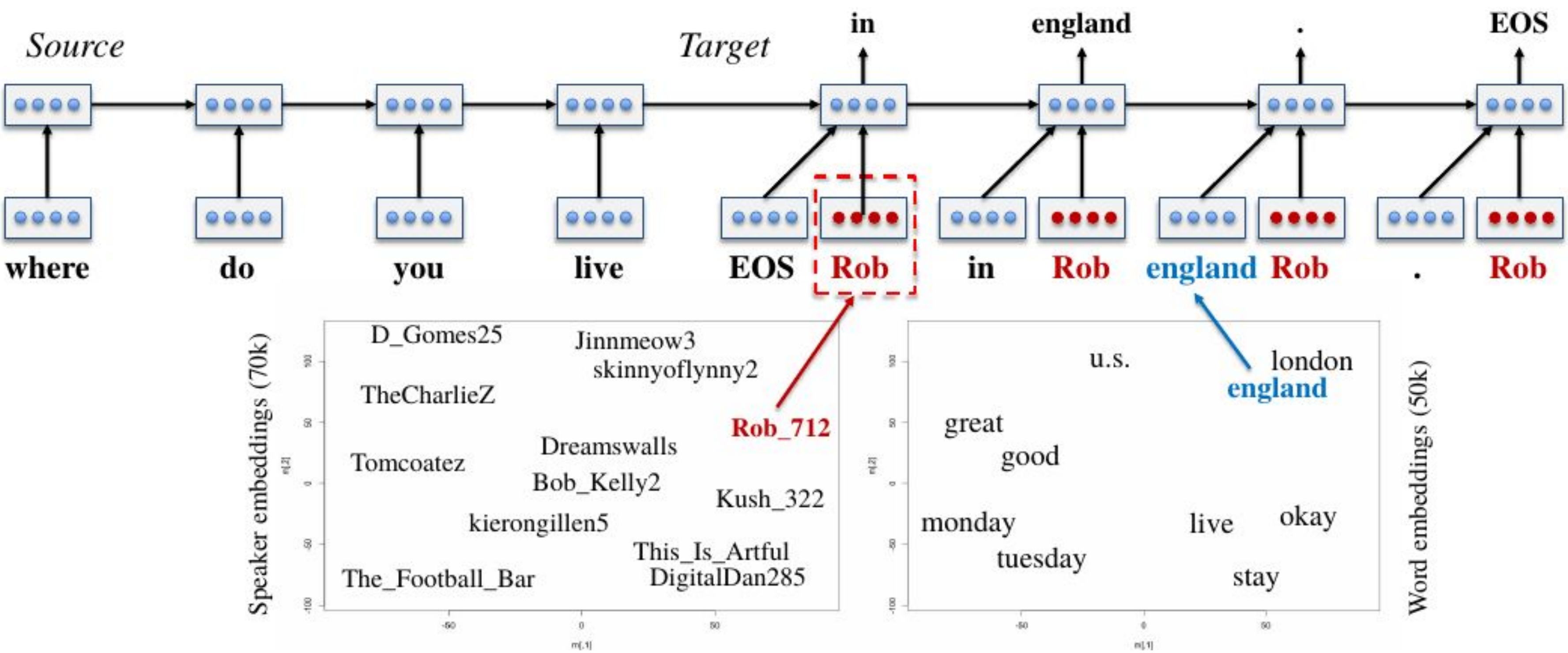
Challenges: Speaker-Consistency

message	Where do you live now?
response	I live in Los Angeles.
message	In which city do you live now?
response	I live in Madrid.
message	In which country do you live now?
response	England, you?
message	Where were you born?
response	I was born in Canada.
message	Where are you from?
response	England, you?
message	Where did you grow up?
response	I grew up in Texas.
message	How old are you?
response	16 and you?
message	What's your age?
response	18.
message	What is your major?
response	I'm majoring in psychology
message	What did you study in college?
response	English lit.

Table 1: Inconsistent responses generated by a 4-layer SEQ2SEQ model trained on 25 million Twitter conversation snippets.

Fully data-driven conversation

Challenges: Speaker-Consistency



Fully data-driven conversation

Challenges: lack of grounding

- Most of chatbots are not grounded in real world!
- This limitation is also inherited from machine translation paradigm
- Some of ideas
  - Grounding system in the persona of speaker
  - Add textual knowledge source
  - Additional input drawn from the user environment, such as image
  - ...

Evaluation

- Evaluation metrics:
  - Blue
  - ROUGE
  - METEOR
  - Delta Blue
- They are not appropriate for dialogue task
- Use GAN to evaluate

Data

- Twitter
- Reddit
- OpenSubtitle
- Ubuntu
- Persona-Chat
- ...

Fully data-driven conversation

Open benchmarks

- Dialogue System Technology Challenge (DSTC)
- ConvAI Competition
- NTCIR STC
- Alexa Prize
- JD Dialogue Challenge
- ...

Conclusion and Research trends

Conclusion and Research Trends

Question answering	SQUAD WikiQA CNN/DailyMail COQA NarrativeQA Wikihop Natural Questions ...
Task-based Dialogue Systems	Other genres Text games E2E learning User simulating Application ...
Chatbots	New architecture Response bladeness Speaker-consistency Knowledge grounded Multi-model ...

How To Start?!

How to start?!

- ⏰ Learn deep learning and nlp!
- ⏰ Learn Pytorch!!
- ⏰ Read Papers about Dialogue
- ⏰ Don’t Ignore Transfer Learning!
- ⏰ Be Up to Date!

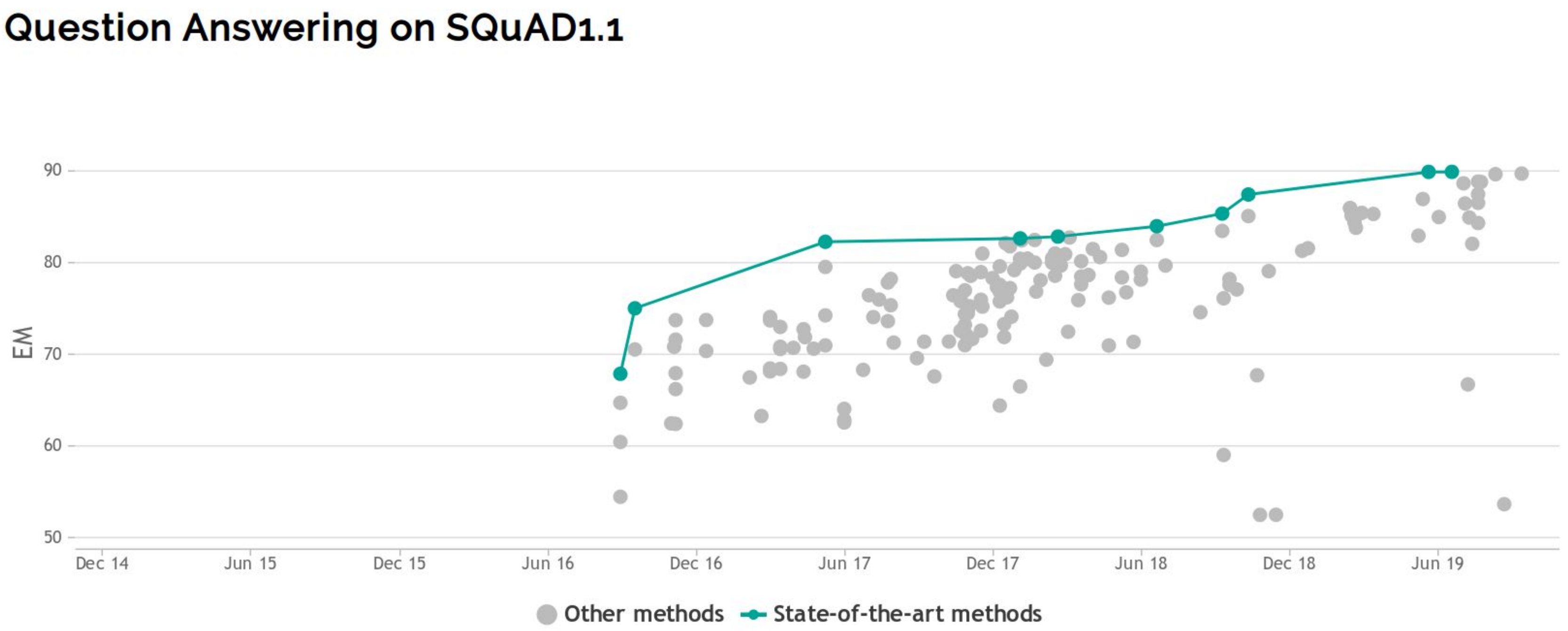
How to start?!

Read paper about dialogue

- A Neural Conversational Model (*ICML Deep Learning Workshop 2015*)
  Oriol Vinyals, Quoc Le
- A Persona-Based Neural Conversation Model (ACL 2016)
  Jiwei Li, Michel Galley, Chris Brockett, Georgios P. Spithourakis, Jianfeng Gao, Bill Dolan
- A Simple, Fast Diverse Decoding Algorithm for Neural Generation (*arXiv* 2017)
  Jiwei Li, Will Monroe, Dan Jurafsky
- Neural Approaches to Conversational AI (*arXiv* 2018)
  Jianfeng Gao, Michel Galley, Lihong Li
- TransferTransfo: A Transfer Learning Approach for Neural Network Based Conversational Agents (*NeurIPS 2018 CAI Workshop*)
  Thomas Wolf, Victor Sanh, Julien Chaumond, Clement Delangue
- Wizard of Wikipedia: Knowledge-Powered Conversational agents (ICLR 2019)
- Learning to Speak and Act in a Fantasy Text Adventure Game (*arXiv* 2019)
  Jack Urbanek, Angela Fan, Siddharth Karamcheti, Saachi Jain, Samuel Humeau, Emily Dinan, Tim Rocktäschel, Douwe Kiela, Arthur Szlam, Jason Weston
- A knowledge-grounded neural conversation model, *Thirty-Second AAAI Conference on Artificial Intelligence*
  Marjan Ghazvininejad, Chris Brockett, Ming-Wei Chang, Bill Dolan, Jianfeng Gao, Wen-tau Yih, Michel Galley

How to start?!

Don't Ignore Transfer Learning!



Transfer Learning

Elmo

Bert

OpenGPT

XLNET

...

How to start?!

Be Up to Date