Deep Learning for Dialogue Systems

PROF. YUN-NUNG (VIVIAN) CHEN 陳縈儂

GTC 2018
Mar 28th, 2018

HTTP://VIVIANCHEN.IDV.TW
Best Poster Award @ GTC 2017

Thanks NVIDIA!!!
Future Life – Intelligent Assistant
Introduction & Background
Language Empowering Intelligent Assistant

Apple Siri (2011)
Google Now (2012)
Microsoft Cortana (2014)

Google Assistant (2016)

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**.
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.
A bot is responsible for a “single” domain, similar to an app.
Task-Oriented Dialogue System (Young, 2000)

Speech Signal

Hypothesis
are there any action movies to see this weekend

Speech Recognition

Text Input
Are there any action movies to see this weekend?

Natural Language Generation (NLG)

Text response
Where are you located?

Language Understanding (LU)
- Domain Identification
- User Intent Detection
- Slot Filling

Semantic Frame
request_movie
genre=action, date=this weekend

Dialogue Management (DM)
- Dialogue State Tracking (DST)
- Dialogue Policy

System Action/Policy
request_location

Backend Action / Knowledge Providers
Interaction Example

Q: How does a dialogue system process this request?

User

find a good eating place for Taiwanese food

Intelligent Agent

Good Taiwanese eating places include Din Tai Fung, Boiling Point, etc. What do you want to choose? I can help you go there.
Task-Oriented Dialogue System (Young, 2000)

Speech Signal

Hypothesis
are there any action movies to see this weekend

Text Input
Are there any action movies to see this weekend?

Speech Recognition

Semantic Frame
request_movie
genre=action, date=this weekend

Language Understanding (LU)
• Domain Identification
• User Intent Detection
• Slot Filling

Natural Language Generation (NLG)

Text response
Where are you located?

System Action/Policy
request_location

Dialogue Management (DM)
• Dialogue State Tracking (DST)
• Dialogue Policy

Backend Action / Knowledge Providers
1. Domain Identification
Requires Predefined Domain Ontology

User

find a good eating place for Taiwanese food

Intelligent Agent

Organized Domain Knowledge (Database)

Restaurant DB

Taxi DB

Movie DB

Classification!
2. Intent Detection
Requires Predefined Schema

User: **find a good eating place for Taiwanese food**

Intelligent Agent

Restaurant DB

**FIND_RESTAURANT**
**FIND_PRICE**
**FIND_TYPE**

Classification!
3. Slot Filling
Requires Predefined Schema

User:

Restaurant DB:

<table>
<thead>
<tr>
<th>Restaurant</th>
<th>Rating</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rest 1</td>
<td>good</td>
<td>Taiwanese</td>
</tr>
<tr>
<td>Rest 2</td>
<td>bad</td>
<td>Thai</td>
</tr>
</tbody>
</table>

Semantic Frame:

FIND_RESTAURANT
rating="good" type="taiwanese"

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

Sequence Labeling:

User: find a **good** eating place for **taiwanese** food

Intelligent Agent:
Task-Oriented Dialogue System (Young, 2000)

- **Speech Recognition**
  - Text Input: "Are there any action movies to see this weekend?"
  - Hypothesis: "Are there any action movies to see this weekend?"

- **Language Understanding (LU)**
  - Domain Identification
  - User Intent Detection
  - Slot Filling
  - Semantic Frame: request_movie, genre=action, date=this weekend

- **Dialogue Management (DM)**
  - Dialogue State Tracking (DST)
  - Dialogue Policy
  - System Action/Policy: request_location

- **Natural Language Generation (NLG)**
  - Text response: "Where are you located?"

- **Backend Action / Knowledge Providers**
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

location, rating
rating, type
loc, rating
all
NULL
rating
type
loc, type
State Tracking
Requires Hand-Crafted States

User
- find a good eating place for Taiwanese food
- I want it near to my office

Intelligent Agent

- location
- rating
- type
- loc, rating
- rating, type
- loc, type
- all
- NULL
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

location
rating
loc, rating
rating, type
loc, type
all

rating="good",
type="thai"

rating="good",
type="taiwanese"

rating="good"
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)

Language Understanding (LU)
- Domain Identification
- User Intent Detection
- Slot Filling

Semantic Frame:
request_movie
genre=action, date=this weekend

Dialogue Management (DM)
- Dialogue State Tracking (DST)
- Dialogue Policy

Backend Action / Knowledge Providers

Text Input:
Are there any action movies to see this weekend?

Hypothesis:
are there any action movies to see this weekend

Speech Signal

Text response
Where are you located?

System Action/Policy
request_location

Natural Language Generation (NLG)
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
  \[ f(\text{声波}) = \text{“你好 (Hello)”} \]

- Image Recognition
  \[ f(\text{猫}) = \text{cat} \]

- Go Playing
  \[ f(\text{棋子}) = \text{5-5 (next move)} \]

- Chat Bot
  \[ f(\text{“Where is GTC?”}) = \text{“The address is...”} \]
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 can only handle binary classification.

\[ f : \mathbb{R}^N \rightarrow \mathbb{R}^M \]

\[ \begin{cases} \text{is "2"} & y \geq 0.5 \\ \text{not "2"} & y < 0.5 \end{cases} \]
A Layer of Neurons

- Handwriting digit classification

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

\[ f : \mathbb{R}^N \rightarrow \mathbb{R}^M \]

10 neurons/10 classes

Which one is max?
Deep Neural Networks (DNN)

- Fully connected feedforward network

\[ f : \mathbb{R}^N \rightarrow \mathbb{R}^M \]

Deep NN: multiple hidden layers
Recurrent Neural Network (RNN)

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

\[ o_t = \text{softmax}(V s_t) \]

RNN can learn accumulated sequential information (time-series)

Deep Learning for LU

- IOB Sequence Labeling for Slot Filling

- Intent Classification
Joint Semantic Frame Parsing

- Slot filling and intent prediction in the same output sequence

- Intent prediction and slot filling are performed in two branches

33

Contextual LU

Domain Identification \rightarrow Intent Prediction \rightarrow Slot Filling

$D$ communication \quad $I$ send_email

$U$ just sent email to bob about fishing this weekend

$S$

O O O O O

B-contact_name B-subject l-subject l-subject

$\Rightarrow$ send_email(contact_name="bob", subject="fishing this weekend")

$U_1$ send email to bob

$S_1$

B-contact_name

$\Rightarrow$ send_email(contact_name="bob")

$U_2$ are we going to fish this weekend

$S_2$

B-message l-message l-message l-message l-message

$\Rightarrow$ send_email(message="are we going to fish this weekend")
Supervised v.s. Reinforcement

- **Supervised**
  - Learning from teacher
  - "Hello"
  - "Bye bye"
  - Say "Hi"
  - Say "Good bye"

- **Reinforcement**
  - Learning from critics
  - "......"
  - Hello 😊
  - "......"
  - "......"
  - "OXX???!
  - Bad
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)

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

Natural Language

Dialogue Policy

Knowledge Database

User Goal

User Model

User Simulation

Are there any action movies to see this weekend?

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

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.

https://arxiv.org/abs/1703.01008
Recent Trends on Learning Dialogues
Evolution Roadmap

- **Dialogue breadth (coverage):**
  - Single domain systems
  - Extended systems
  - Multi-domain systems
  - Open domain systems

- **Dialogue depth (complexity):**
  - I’ve got a cold what do I do?
  - Tell me a joke.
  - What is influenza?
  - I feel sad...
Intent Expansion (Chen et al., 2016)

- 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

CDSSM

Training Data
- `<change_note>`
  - “adjust my note”
- `<change_setting>`
  - “volume turn down”

New Intent
- `<change_calender>`

postpone my meeting to five pm

Intent Representation

Embedding Generation

K
K+1
K+2
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

Dialogue breadth (coverage)

Dialogue depth (complexity)

Knowledge based system

I feel sad...

Empathetic systems

I’ve got a cold what do I do?

Common sense system

Tell me a joke.

What is influenza?

Knowledge based system
App Behavior for Understanding

- Task: user intent prediction
- Challenge: language ambiguity

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

- 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)

- High-level intention may span several domains

Users can interact via high-level descriptions and the system learns how to plan the dialogues.
Empathy in Dialogue System (Fung et al., 2016)

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

https://arxiv.org/abs/1605.04072
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 scarce 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

Speech Recognition

Hypothesis
are there any action movies to see this weekend

Language Understanding (LU)
- Domain Identification
- User Intent Detection
- Slot Filling

Semantic Frame
request_movie
genre=action, date=this weekend

Dialogue Management (DM)
- Dialogue State Tracking (DST)
- Dialogue Policy

System Action/Policy
request_location

Backend Action / Knowledge Providers

Natural Language Generation (NLG)

Text response
Where are you located?

Text Input
Are there any action movies to see this weekend?
Thanks for Your Attention!

Q & A

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