Graph Learning for AR/VR Intelligent Assistants

Xin Luna Dong, Meta Reality Lab

9/2022

This talk does not represent the company’s point of view
Everyone Deserves An Assistant
What is A Virtual Intelligent Assistant?

Respond to commands

“Hey Siri, set a timer to 7pm”

“Ok, added to today’s reminders”
What is A Virtual Intelligent Assistant?

Control devices

“Hey Alexa, turn off bedroom lights”
What is A Virtual Intelligent Assistant?

Provide information

“Hey, Google, when does summer end?”

“Summer ends on Thursday, Sept 22”
Meta’s Assistant
Empowering connection to people and experiences in your life

- **Facebook Portal**
  - “Hey Portal”
  - Hop on a call hands-free
  - Get help with music, timers, alarms, weather, show photos from your Facebook profile, and more.

- **Meta Quest 2**
  - “Hey Facebook” (double press the button on your controller)
  - “Who’s online?”—meet up with friends
  - “Open Beat Saber”—jump straight in the game, and more.

- **Ray-Ban Stories**
  - “Hey Facebook, take a picture” -- capture moments hands-free
  - “Hey Facebook”--call friends on Messenger, manage device settings, and more.
What is An Ideal Virtual Intelligent Assistant?

An intelligent assistant should be an agent that knows you and the world, can receive your requests or predict your needs, and provide you the right services at the right time with your permission.
Technologies to Support Current Intelligent Assistants
Two Types of Commands to Intelligent Assistants

- **Task driven**
  - E.g., Set a timer
  - E.g., Turn on lights
  - E.g., When is Easter?

- **Info driven**
  - E.g., Set a timer
  - E.g., Turn on lights
  - E.g., When is Easter?
Intelligent Assistant Is Essentially a Conversation System

ASR (Automatic Speech Recognizer) → NLU (Natural Language Understanding) → Dialog State Tracker

NLU

Turning on lights

{Intent: DEVICE_TURNON; Device: Lights #1}

{ActType: TURRON; Device: Lights #1; Status: Off}

{Status: SUCCESS}

TTS (Text-To-Speech Synthesizer) → NLG (Natural Language Generator) → Dialog Policy

One round: Confirmation

Devices

Apps

KG/Web
Int. Asst. Is Essentially a Conversation System

Rnd 2. “Bedroom lights”
Rnd 3. “Dim the lights”
Rnd 4. “How many people does it take to change its bulb?”

One round: Confirmation
Two rounds: Clarification questions
Multi rounds: Follow-up requests

Intelligent Assistant

ASR (Automatic Speech Recognizer)
NLU (Natural Language Understanding)
Dialog State Tracker
NLU (Natural Language Generator)
Dialog Policy

TTS (Text-To-Speech Synthesizer)

Apps
Devices
KG/Web
Challenges and Initial Solutions to AR/VR Assistants
What Is Different for An AR/VR Assistant?

You see through it

You wear it everywhere

May not have connection

You wear for a long time
From Voice-Only to Multi-Modal

“How tall is Empire State Building?”

“What’s the name of this building and how tall is it?”
From Context-Agnostic to Context-Aware

“Show my shopping list”

“Remember to buy apples and bananas at the grocery store around the corner”
"What's the weather today?"

"Today is sunny, 70 degree. Would you like to play your favorite morning music?"
From Server-Side to On-Device

May not have connection

+ Privacy!!!
MultiModal, Context-Aware Assistant

Intelligent Assistant

Understanding
- ASR (Automatic Speech Recognizer)
- NLU (Natural Lang. Understanding)

Planning
- Dialog State Tracker
- Dialog Policy

Responding
- TTS (Text-To-Speech Synthesizer)
- NLG (Natural Language Generator)

Inputs:
- Apps
- Devices
- KG/Web
MultiModal, Context-Aware Assistant

Understanding
- ASR
- CV (Comp. Vision)
- Multi-modal Understanding

Responding
- TTS
- Others
- NLG
- Image Gen

Planning
- Dialog State Tracker
- Dialog Policy

Apps
- Devices
- KG/Web

Intelligent Assistant

You see through it
MultiModal, Context-Aware Assistant

Intelligent Assistant

Understanding
- ASR
- CV
- Multi-modal & Context-aware Understanding

Context

Responding
- TTS
- Others
- NLG
- Image Gen

Dialog State Tracker

Context-Aware Planning
- Dialog Policy

Apps

Devices

KG/Web
MultiModal, Context-Aware Assistant

Intelligent Assistant

Understanding
- ASR
- CV
- Context
- TTS
- Others

Responding

Context-Aware Planning
- Dialog State Tracker
- Dialog Policy

Multi-modal & Context-aware Understanding
- Proactive Trigger

Devices
- Apps
- KG/Web

You wear for a long time
Roles of Graph Learning in Enabling AR/VR Assistants
Assistant Recommendation

Challenges: Leveraging vast volume of public knowledge for reactive recommendation w/o sacrificing latency
4-1. Conv. Recom. w. Public KG

(a) Dialog
1. Can you recommend any classic books like *Catcher in the Rye*?

2. Do you prefer books by the same *author* or same *genre*?

3. I am interested in reading classic examples of *American literature*.

4. *Literary realism* is a common genre in classic *American literature*.

5. Do you prefer *First-person* or *Third-person narrative*?

6. I mostly prefer *third-person narrative*.

7. Consider reading the *Scarlet Letter*: a novel by *Nathaniel Hawthorne*.

(b) KG

- **Genre**
- **The Catcher in the Rye**
- **Writer**
- **American Literature**
- **1st-person**
- **JD Salinger**
- **New York City**
- **Film Location**
- **Literacy Realism**
- **The Scarlet Letter**
- **3rd-person**

Find paths to follow when making suggestions or providing information.

Seungwhan Moon, Pararth Shah, Anuj Kumar, Rajen Subba. OpenDialKG: Explainable Conversational Reasoning w. Attention-based Walks over Knowledge Graphs. ACL, 2019
4-1. Conv. Recom. w. Public KG

```
for each
Romance
DiCaprio
...

"... recommend me ..."

```

Initial KG Nodes

```
Input Encoder
(Section 2.2)
```

Path Generation via DialKG Walk
(Section 2.3)

Entity Re-ranker

```
Eq. (3) Modality Attention

Eq. (6) DialKG Walker

Eq. (8) Re-ranking

```

Prune unattended paths to reduce search space

Seungwhan Moon, Pararth Shah, Anuj Kumar, Rajen Subba. OpenDialKG: Explainable Conversational Reasoning w. Attention-based Walks over Knowledge Graphs. ACL, 2019
Assistant Recommendation (Cont’)

Challenges: Leveraging personal action log, preferences, routines, etc., to improve context-aware recommendation w/o sacrificing privacy.
4-2. Conv. Recom. w. Public+Personal KG

(a) Conversational Recommendations

1. Hello, I’m looking for a good place to eat.
2. Is Downtown Phoenix a good place to start your search as you’ve been there a few times?
3. Yes, I’d like something there please.
4. You ordered American food a few times, do you want something similar or feeling adventurous today?
5. No, I’m in the mood for a BBQ today.
6. Bear’s BBQ got some affordable yet great BBQ. They also serve wine.

(b) Memory Graph (MG)

Public KG

Personal KG

4-2. Conv. Recom. w. Public+Personal KG

Figure 5: Visualization of item-level conversational reasoning, given an example dialog. Darker color indicates more salient items for recommendation at each given turn (row), predicted by our UMGR model.
4-3. Federated Learning

- Push data to model → Push models to data

Figure 1: Personalized Document Model in FL.
Future of Graph Learning for AR/VR Assistant

- Multi-modal
- Contextual AI
- Proactive

Knowledge

Social

Behavior

Runtime
On-device
Privacy
My Everyday Worries When Working on Devices

- Memory
- Battery
- Thermal
Take-Aways

- An intelligent assistant should be an agent that *knows you and the world*, can *receive your requests* or *predict your needs*, and provide you *the right services at the right time* with your permission.
- An intelligent assistant is essentially a *conversation system*, *task-driven or information-driven*.
- Next-generation AR/VR assistants require new research on graph learning to support multi-modal & context-aware recommendations.
Thank You

Q&A?