Intelligent Conversational Bot

Multimodality

Jun 6th, 2017

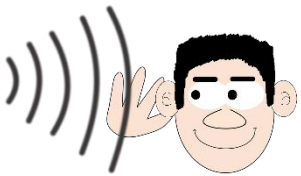
YUN-NUNG (VIVIAN) CHEN

WWW.CSIE.NTU.EDU.TW/~YVCHEN/S105-ICB

國立臺灣大學
National Taiwan University

Multimodality

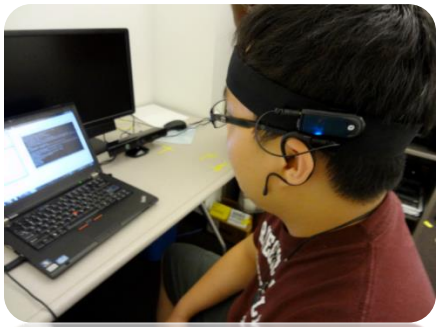
- Definition
  - **Multimodality** describes communication practices in terms of the *textual*, *aural*, *linguistic*, *spatial*, and *visual* resources - or modes - used to compose messages.
  - Where media are concerned, multimodality is the use of several modes (media) to create a single artifact.

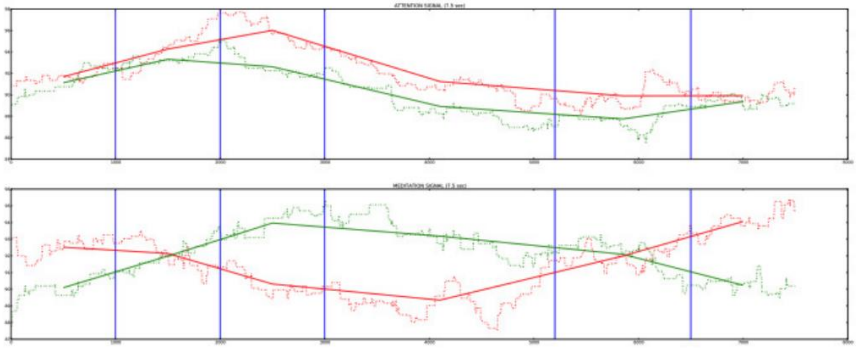



Brain Signal for Understanding

http://dl.acm.org/citation.cfm?id=2388695

- Misunderstanding detection by brain signal
  - Green: listen to the correct answer
  - Red: listen to the wrong answer

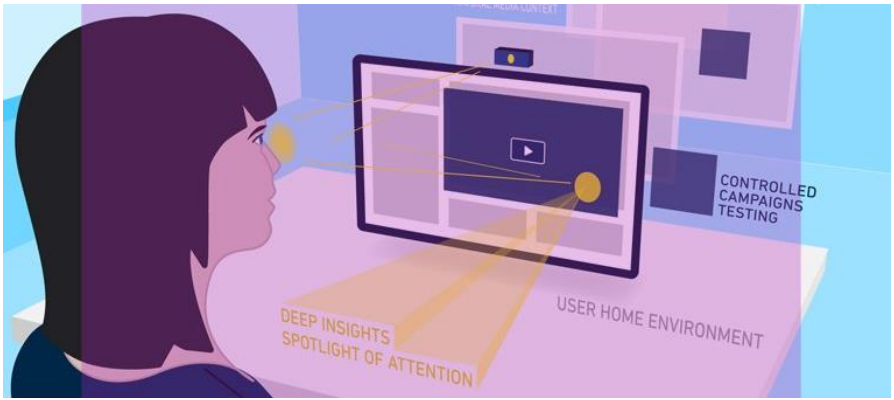




Detecting misunderstanding via brain signal in order to correct the understanding results

Eye Tracking for Understanding

- Better understanding using additional multimodal information



Improving understanding via non-textual signal

Video for Intent Understanding

I want to see a movie on TV!

Intent: turn_on_tv

Proactive (from camera)

Sir, may I turn on the TV for you?

Proactively understanding user intent to initiate the dialogues.

App Behavior for Understanding

- Task: user intent prediction
- Challenge: language ambiguity

v.s.

Email?

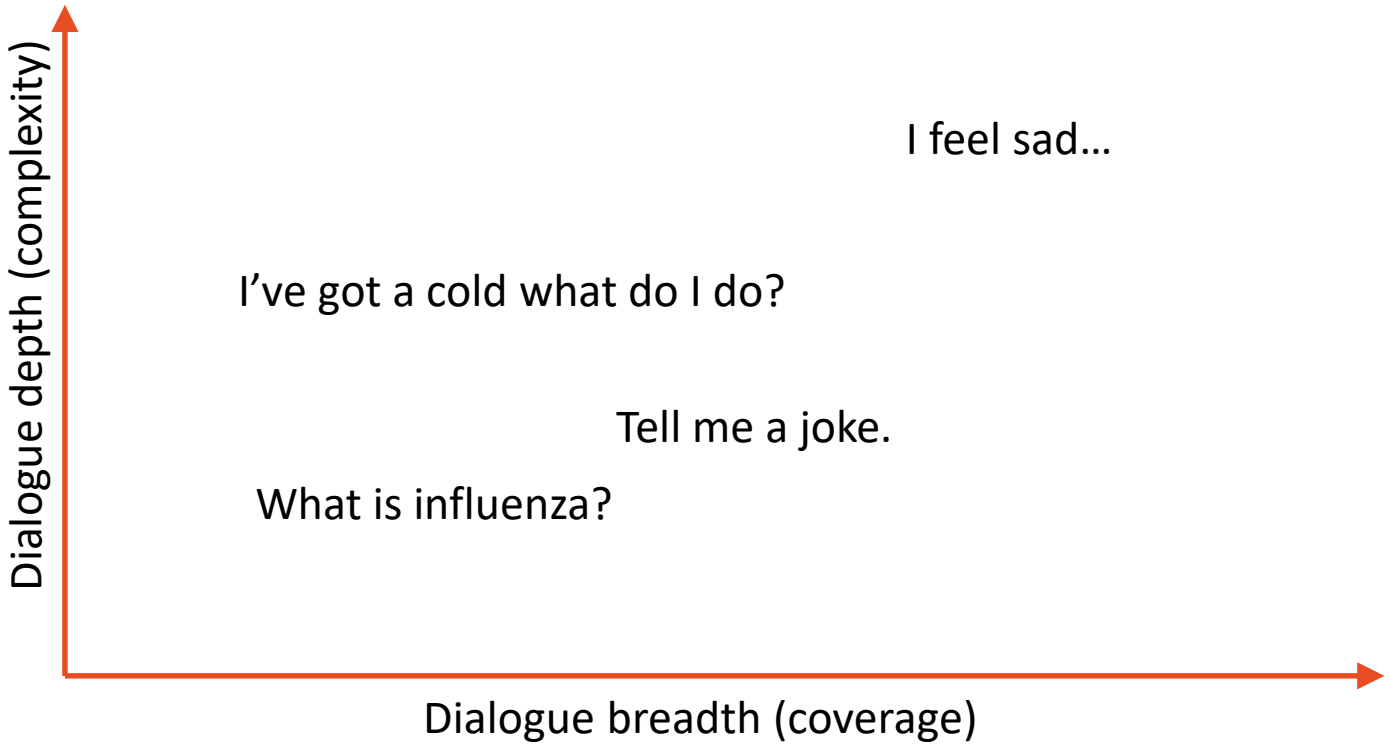
Message?

① **User preference**
- ✓ Some people prefer “Message” to “Email”
- ✓ Some people prefer “Outlook” to “Gmail”

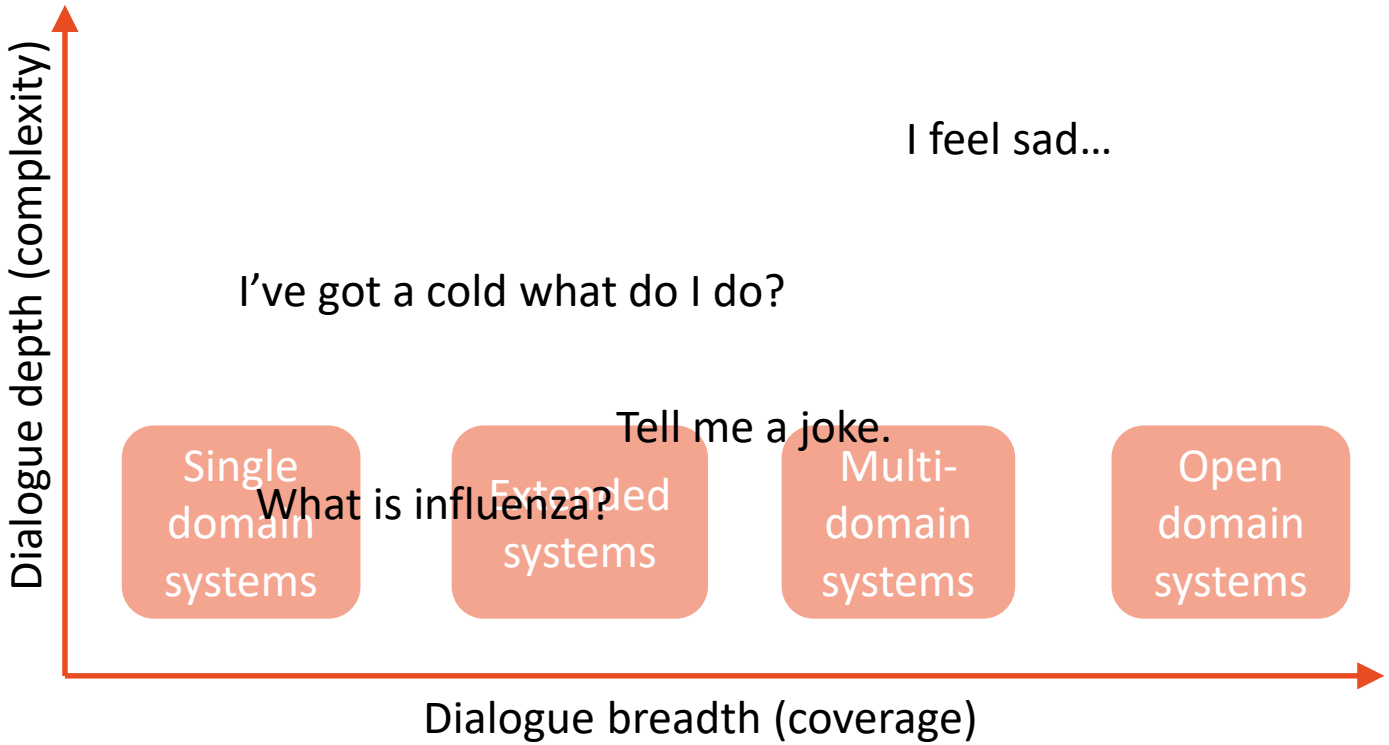
② **App-level contexts**
- ✓ “Message” is more likely to follow “Camera”
- ✓ “Email” is more likely to follow “Excel”

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

Evolution Roadmap



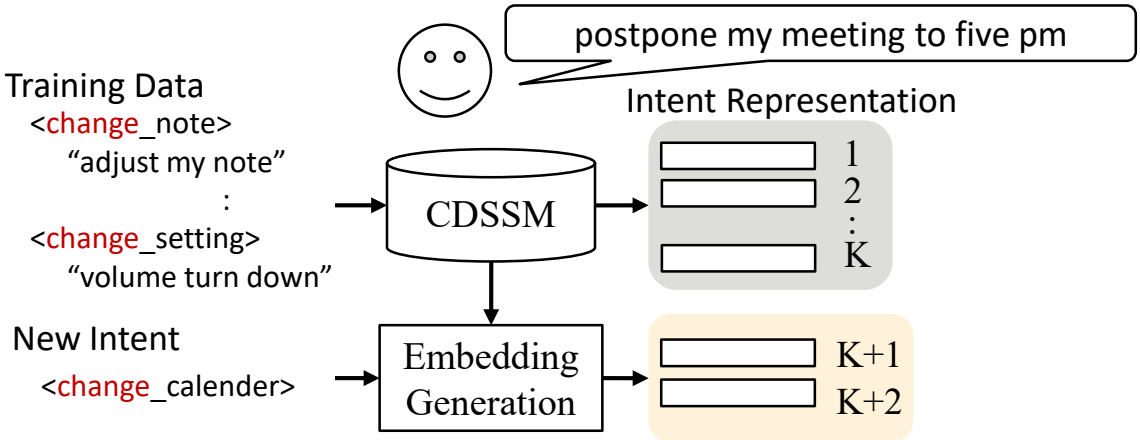
Evolution Roadmap



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

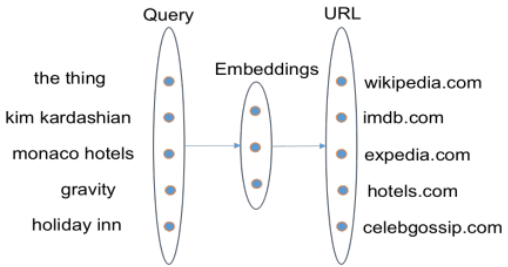


The dialogue act representations can be automatically learned for other domains

Zero-Shot Learning (Daupin et al., 2016)

https://arxiv.org/abs/1401.0509

- Semantic utterance classification
  - Use query click logs to define a task that makes the networks learn the meaning or intent behind the queries



Depiction of the deep network from queries to URLs.

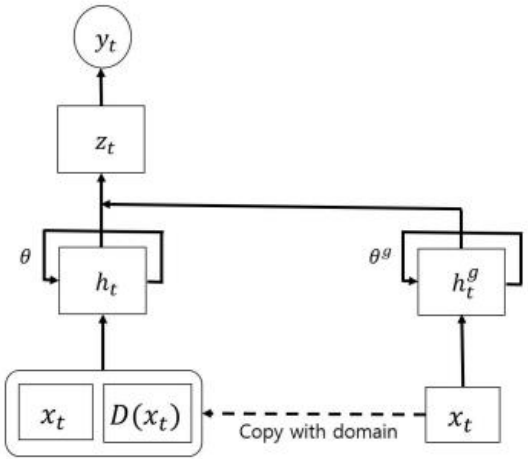
$$\mathcal{L}(X, Y) = -\log P(Y|X) + \lambda H(P(C|X)).$$

  - The semantic features are the last hidden layer of the DNN
  - Use zero-shot discriminative embedding model combines *H* with the minimization of entropy of a zero-shot classifier

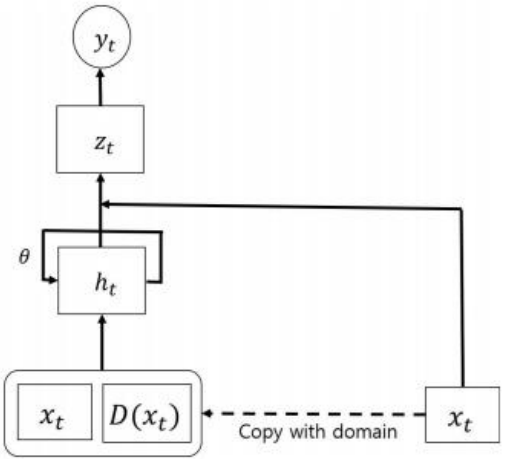
Domain Adaptation for SLU (Kim et al., 2016)

http://www.aclweb.org/anthology/C/C16/C16-1038.pdf

- Frustratingly easy domain adaptation
- Novel neural approaches to domain adaptation
- Improve slot tagging on several domains



(a) 1 domain specific LSTM + generic LSTM

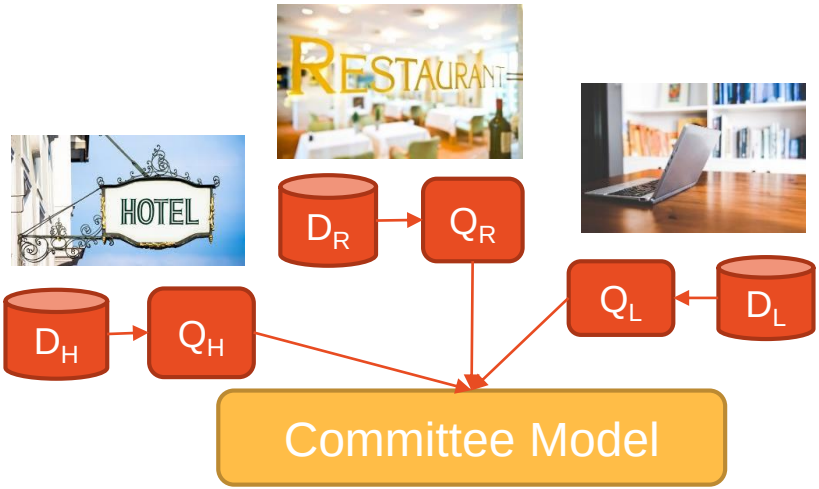


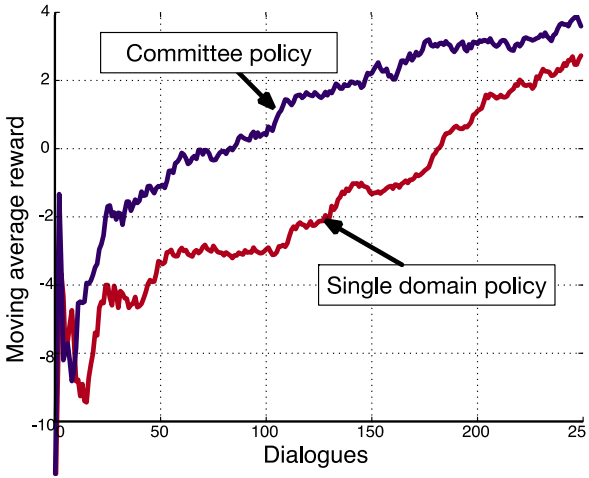
(b) 1 domain specific LSTM + generic embedding

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



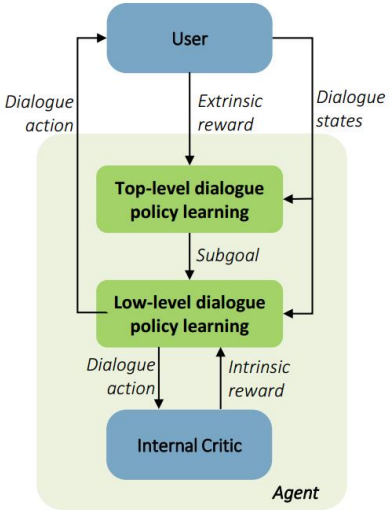


The policy from a new domain can be boosted by the committee policy

Multi-Domain Dialogue System

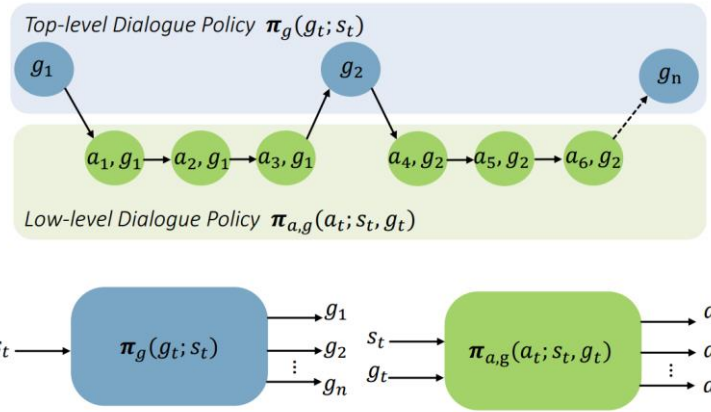
https://arxiv.org/pdf/1704.03084.pdf

- Hierarchical reinforcement learning for DM
  - Meta-controller: select the goal/domain
  - Controller: select the action



Meta-controller

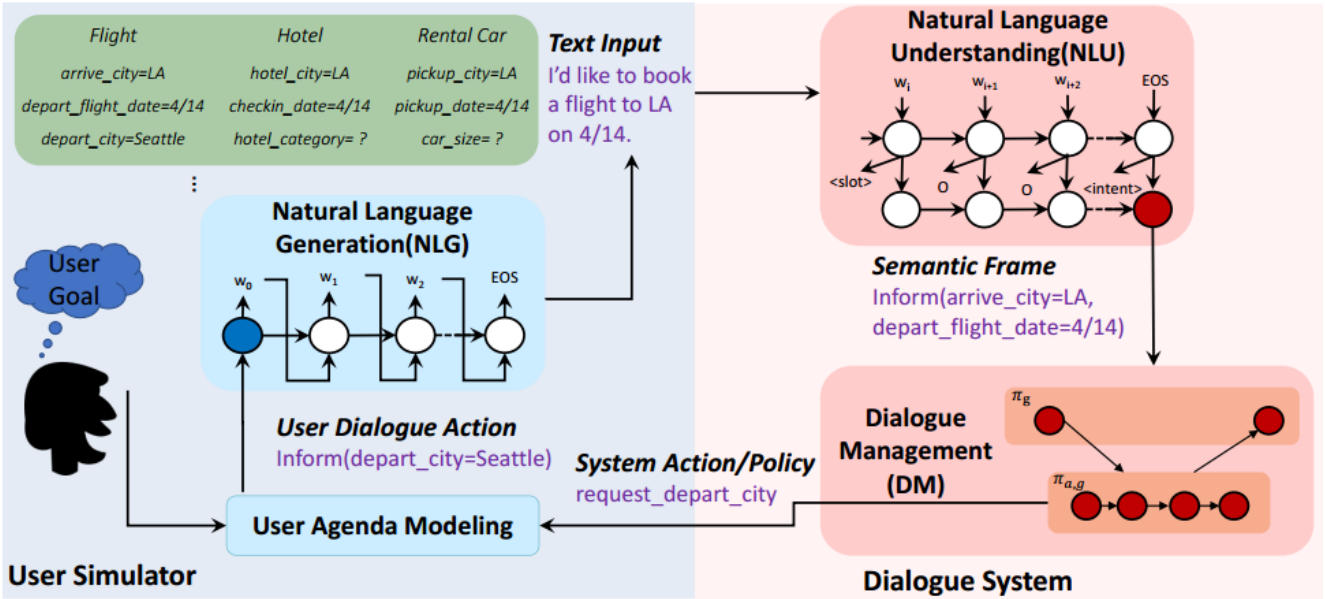
Controller



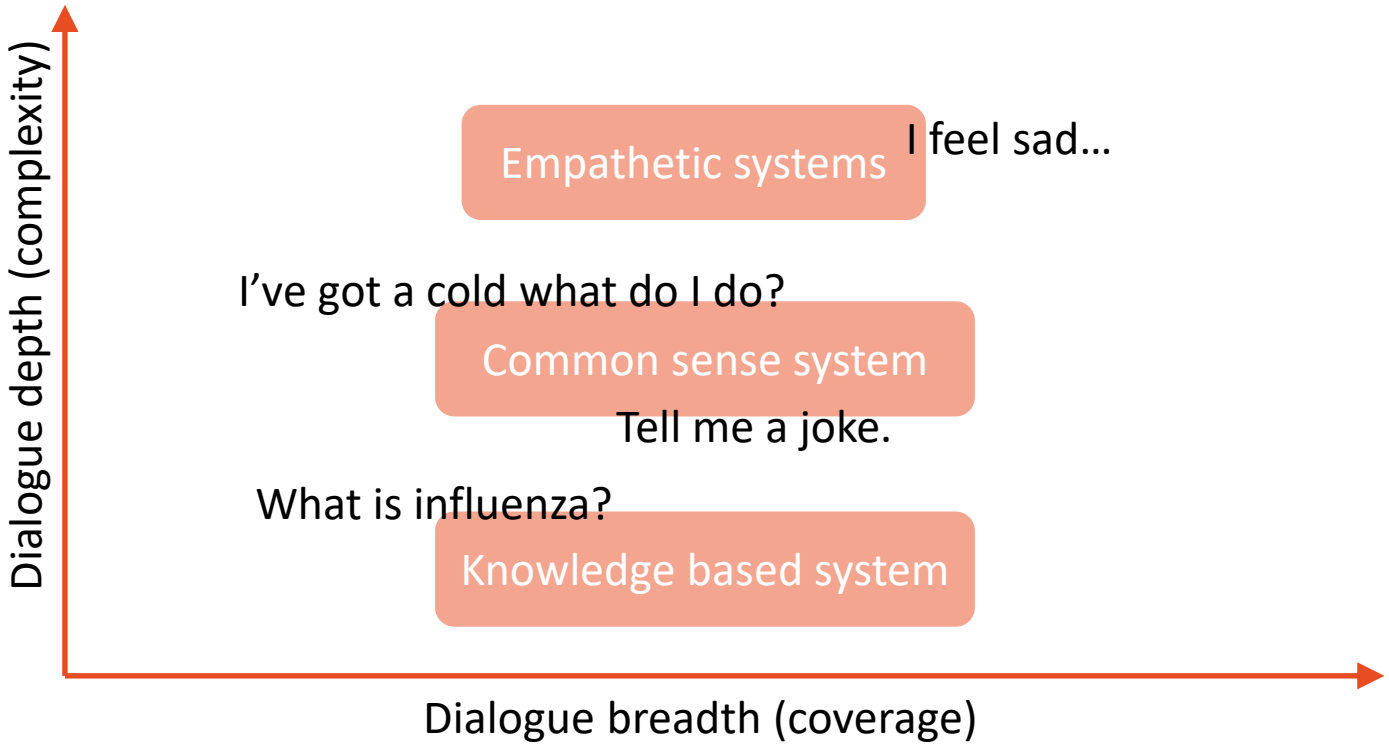
Multi-Domain Dialogue System

https://arxiv.org/pdf/1704.03084.pdf

- Hierarchical reinforcement learning for DM



Evolution Roadmap

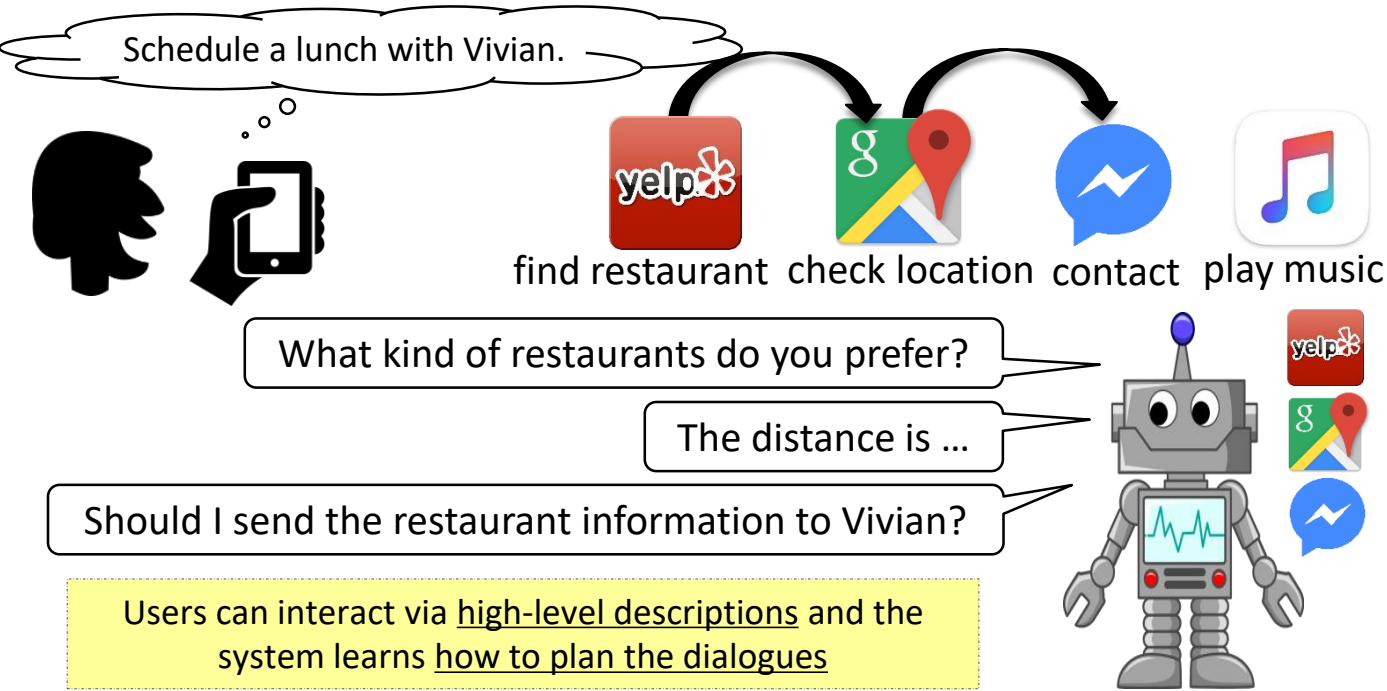


High-Level Intention for Dialogue Planning

(Sun et al., 2016; 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

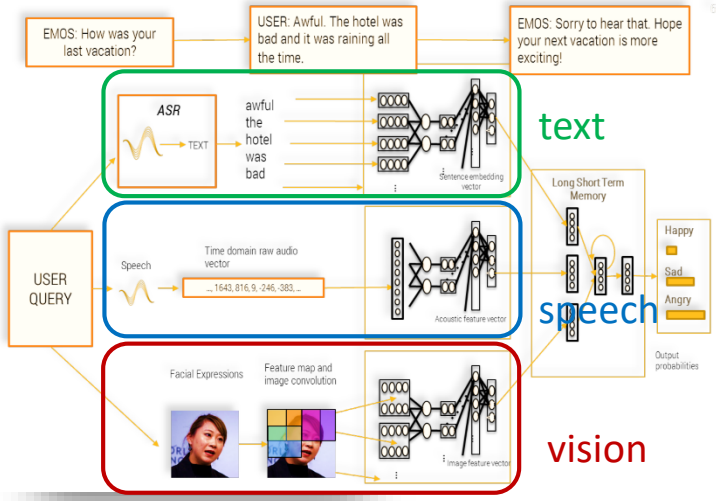


Users can interact via high-level descriptions and the system learns how to plan the dialogues

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

Zara - The Empathetic Supergirl

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"
}
```

Concluding Remarks

- Multimodal signals
  - Detect misunderstanding
  - Benefit understanding
- Dialogue breath
  - Single domain → Extended domain → Multi-domain → Open domain
- Dialogue depth
  - Knowledge-based system → Common sense system → Empathetic system