Reinforcement Learning for Personalized Dialogue Management

Floris den Hengst, AI for Fintech, ING

VU VRIJE UNIVERSITEIT AMSTERDAM

ING

Reinforcement Learning for Personalized Dialogue Management

Floris den Hengst
ING Groep NV
Floris.den.Hengst@ing.com

Mark Hoogendoorn
Vrije Universiteit Amsterdam
M.Hoogendoorn@vu.nl

Frank van Harmelen
Vrije Universiteit Amsterdam
Frank.van.Harmelen@vu.nl

Joost Bosman
ING Groep NV
Joost.Bosman@ing.com

ABSTRACT

Language systems have been of great interest to the research community and have recently reached the mass market through various assistant platforms on the web. Reinforcement Learning methods that optimize dialogue policies have seen successes in past years and have recently been extended into methods that *personalize* the dialogue, e.g. take the personal context of users into account. These works, however, are limited to personalization to a single user with whom they require multiple interactions and do not generalize the usage of context across users. This work introduces a problem where a generalized usage of context is relevant and proposes two Reinforcement Learning (RL)-based approaches to this problem,

1 INTRODUCTION

The use of language by machines has been one of the central challenges in Artificial Intelligence since its initiation as a field of research [30] [19]. Decades of research have advanced the state of art to such an extent that major consumer-facing web platforms currently offer text- and voice-based 'assistant' capabilities, such as Tencent's WeChat, Microsoft's Cortana, Google's Assistant etc. These platforms have made access to the web through dialogue ordinary. Although such platforms offer high-quality Automatic Speech Recognition (ASR), Natural Language Understanding (NLU) and audio synthesis modules, Dialogue Management (DM) modules are typically handcrafted and require many non-trivial decisions in

https://arxiv.org/abs/1908.00286

VU VRIJE UNIVERSITEIT AMSTERDAM ING

Dialogue Systems

Dialogflow

Floris den Hengst, AI for Fintech ICAI lab, Floris.den.Hengst@ing.com

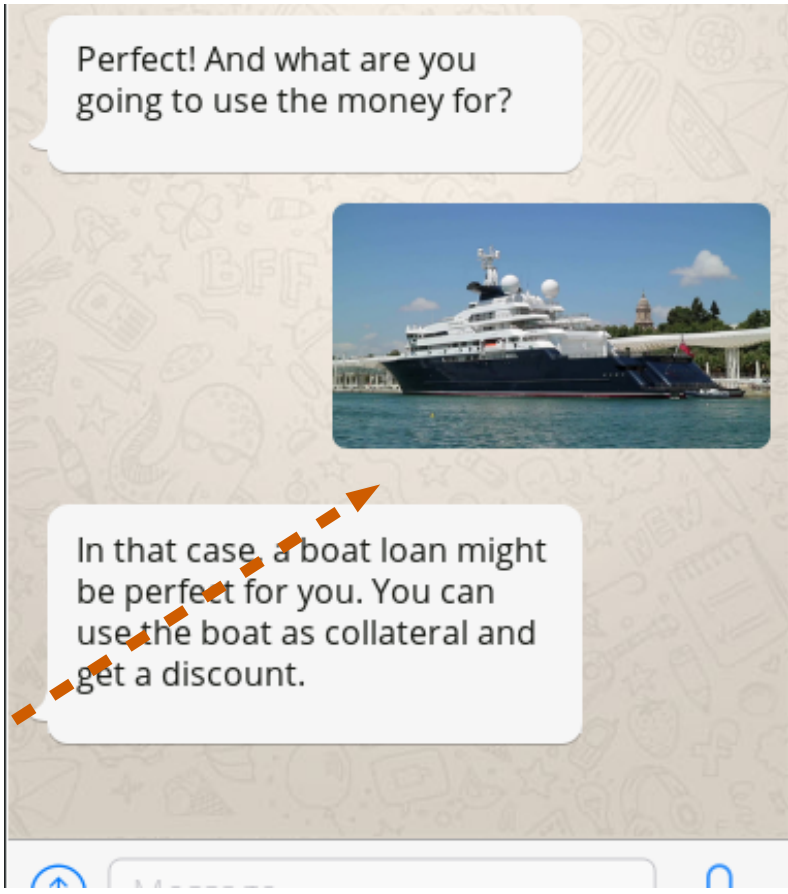
Language and context

[bransford1972]

Intent

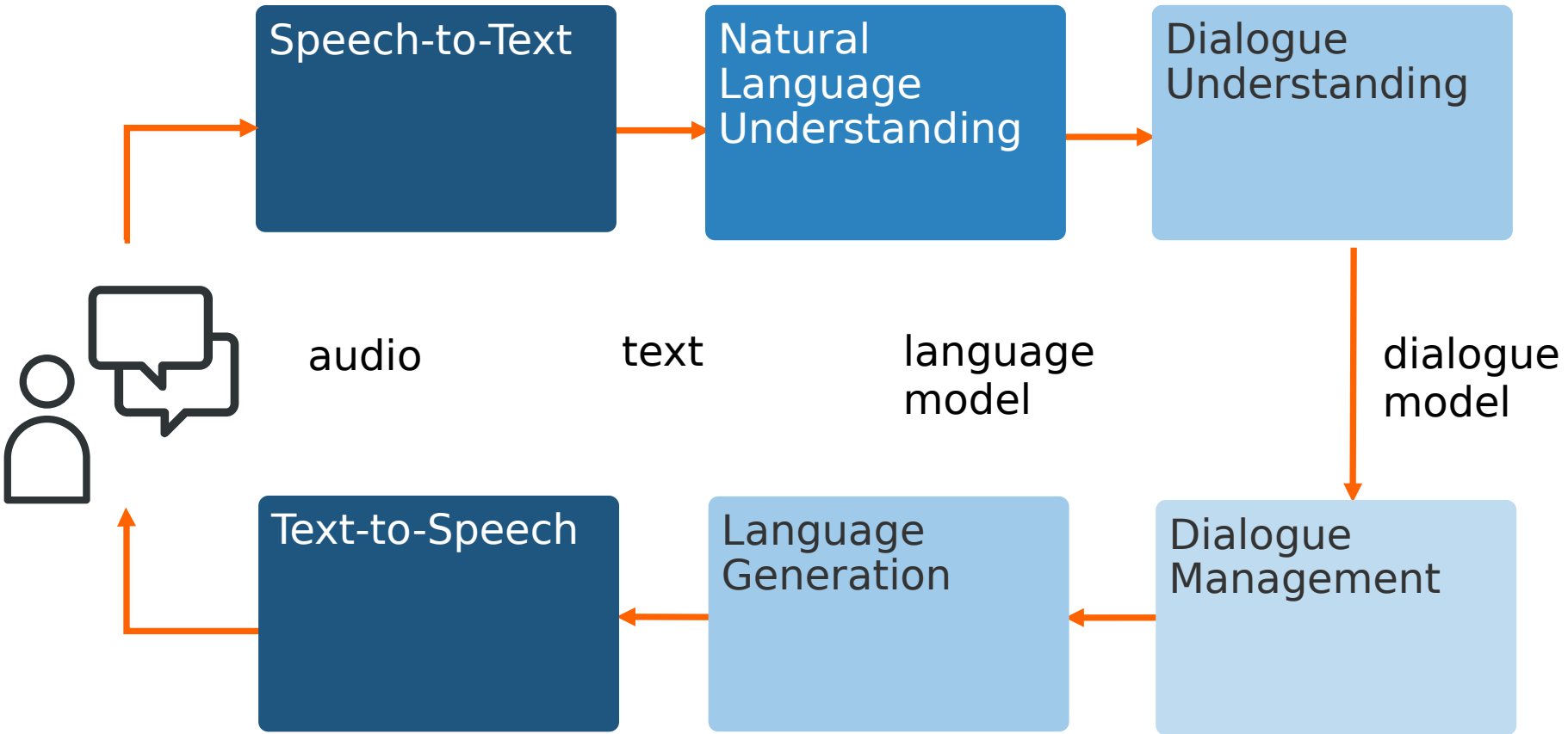
Slots:
duration & purpose

Values:
duration & purpose



Architecture

[peckham1991]
[ultes2017]



Sequential Model

1. Intent: “loan”
2. Purpose: “boat”
3. Duration: “16M”

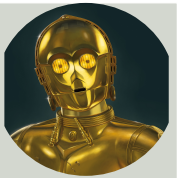
1. Intent: “loan”, purpose:”boat”
2. Duration: “16M”

1. Intent: “loan”, p: ”boat”, d: ”16M”

…..

Dialogue Management

How may I help you?



When are you going to pay the money back?

What are you going to use the money for?

A <…> might be good for you, …

….

Floris den Hengst, AI for Fintech ICAI lab, Floris.den.Hengst@ing.com

Finite State Machine Model

Intent: loan
Duration: ?
Purpose: ?
....

Intent: loan
D: 12 M
P: ?
....

Intent: loan
D: ?
P: boat
....

Intent: loan
D: ?
P: home
....

Intent: loan
D: 12 M
P: boat
....

Dialogue Management

How may I help you?

When are you going to pay the money back?

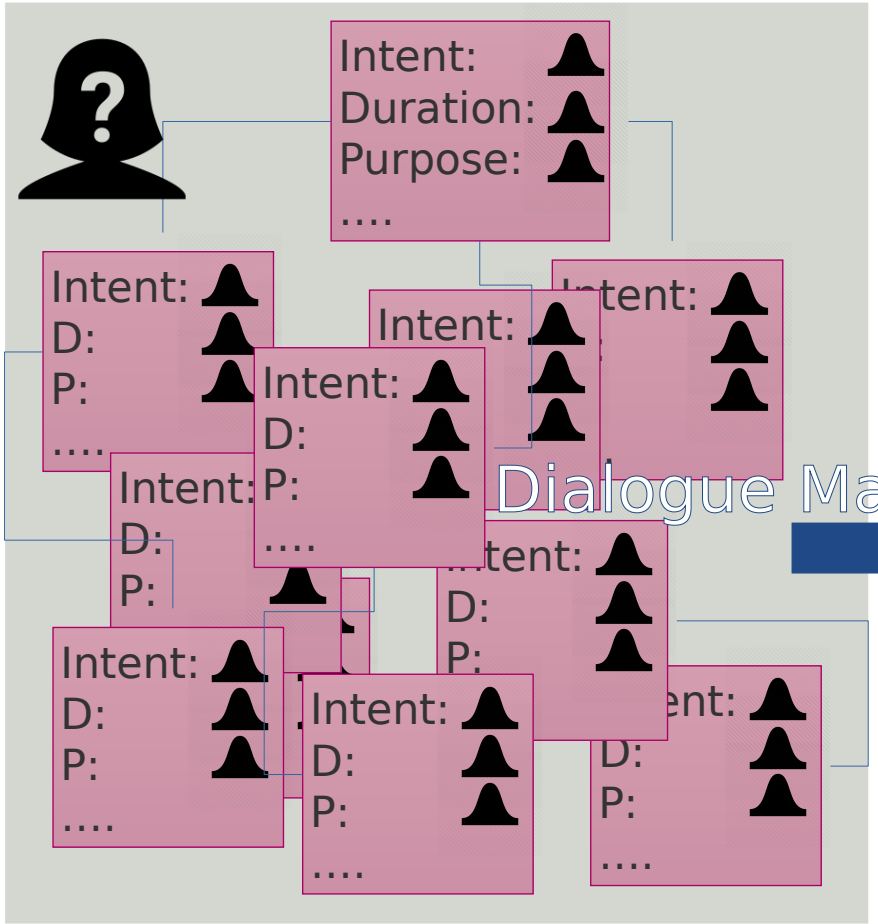
What are you going to use the money for?

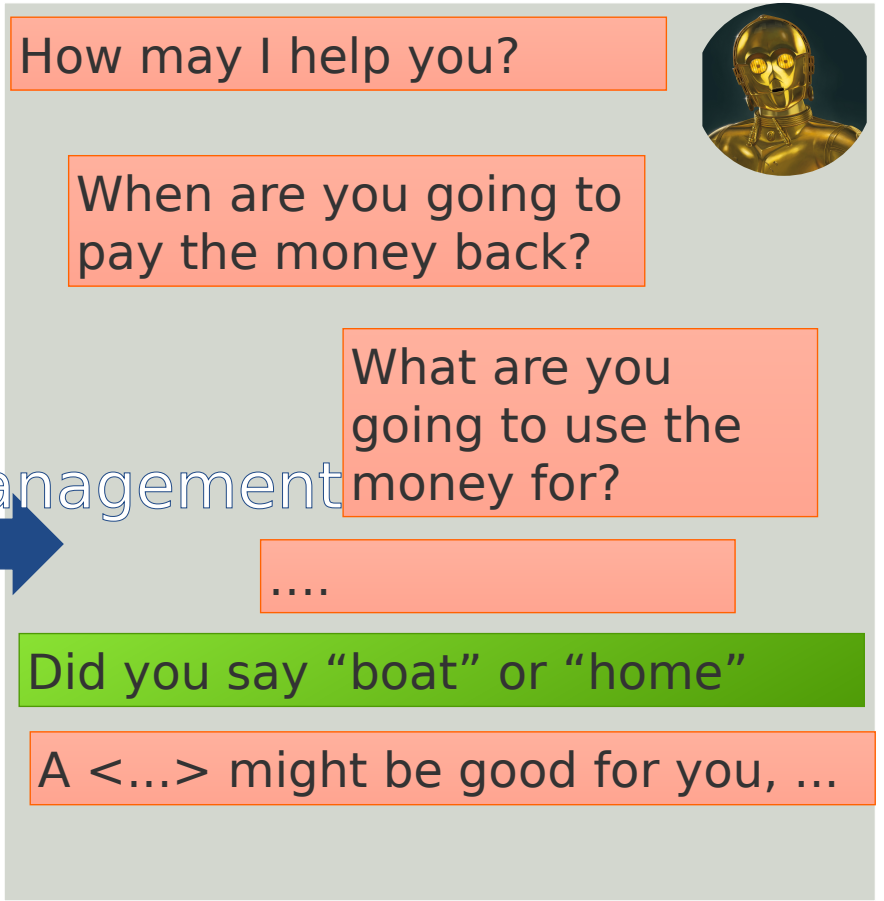
....

A <...> might be good for you, ...

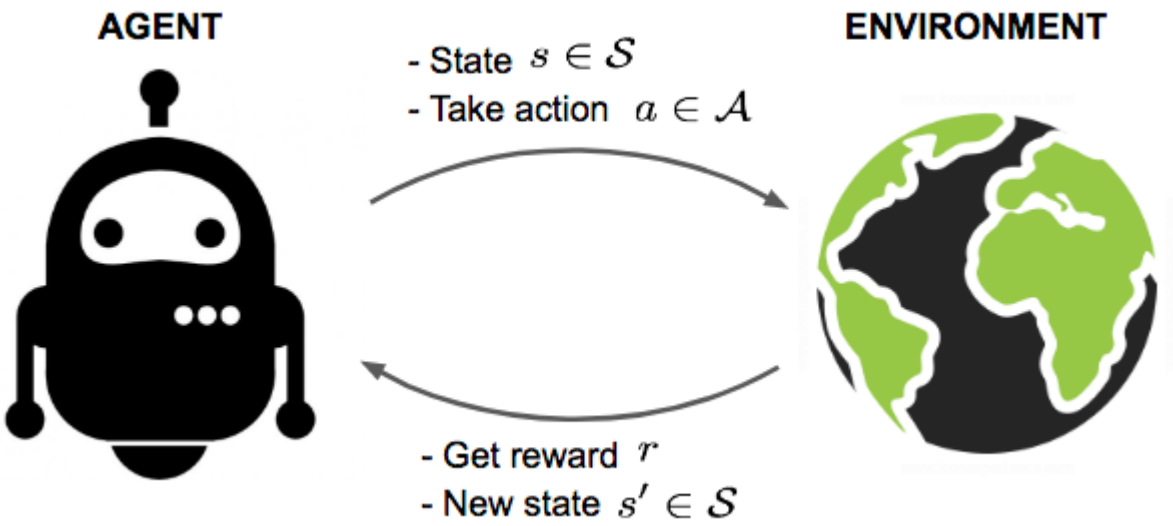
Belief State Model

[young2007]





Reinforcement Learning (1/2)



Trajectory $\langle s_0, a_0, r_0, s_1, \ldots, s_T, a_T, r_T \rangle, r_. \in \mathbb{R}$

$$\text{Maximize} \sum_{t=0}^{t=T} \gamma^{t+1} r_t, \gamma \in [0, 1]$$

Reinforcement Learning (2/2)

$A \in \{a_1, \ldots, a_n\}, S \in \{s_1, \ldots, s_m\}$

Maximize $\sum_{t=0}^{t=T} \gamma^{t+1} r_t$

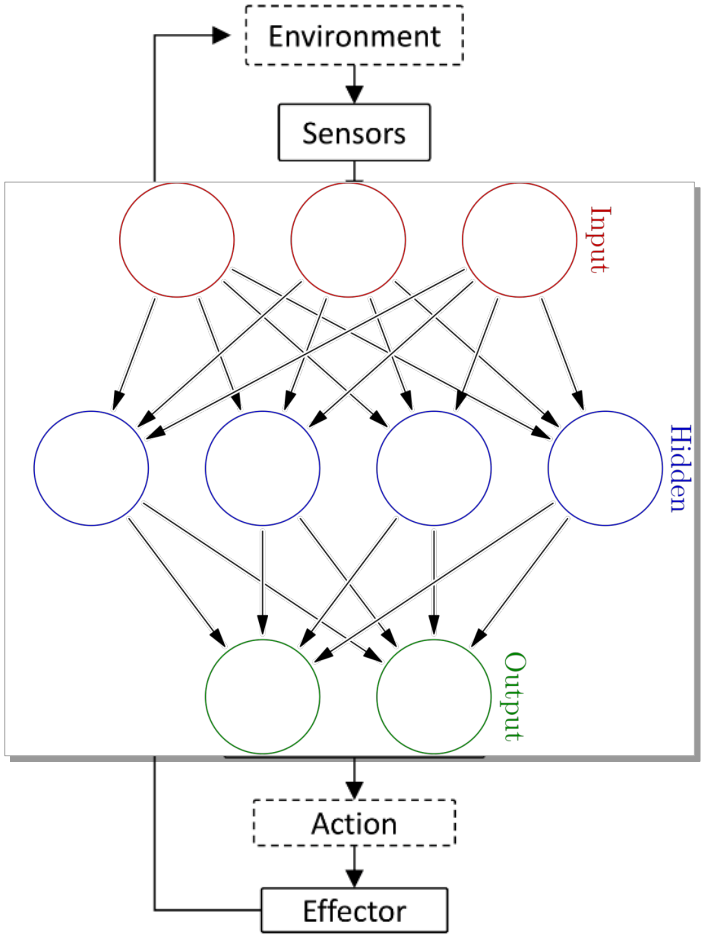
Trajectory $\langle s_0, a_0, r_0, s_1, \ldots, s_T, a_T, r_T \rangle$

$\pi^x \in \Pi : S \rightarrow A$

$$Q^{\pi^x}(s,a) = E_{\pi^x}\{\sum_{k=0}^{k=T} \gamma^k r_{t+k+1} | s_t = s, a_t = a\}$$

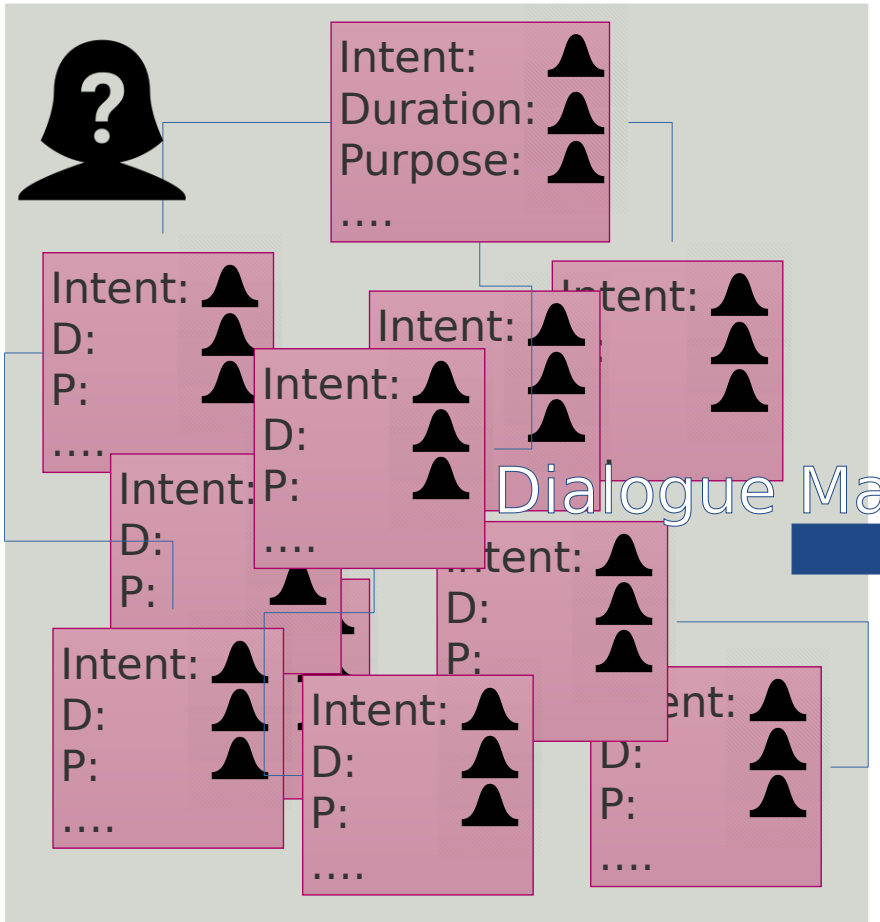
$$\pi^*(s) = \arg\max_a Q^{\pi^*}(s,a), \forall s \in S, a \in A$$

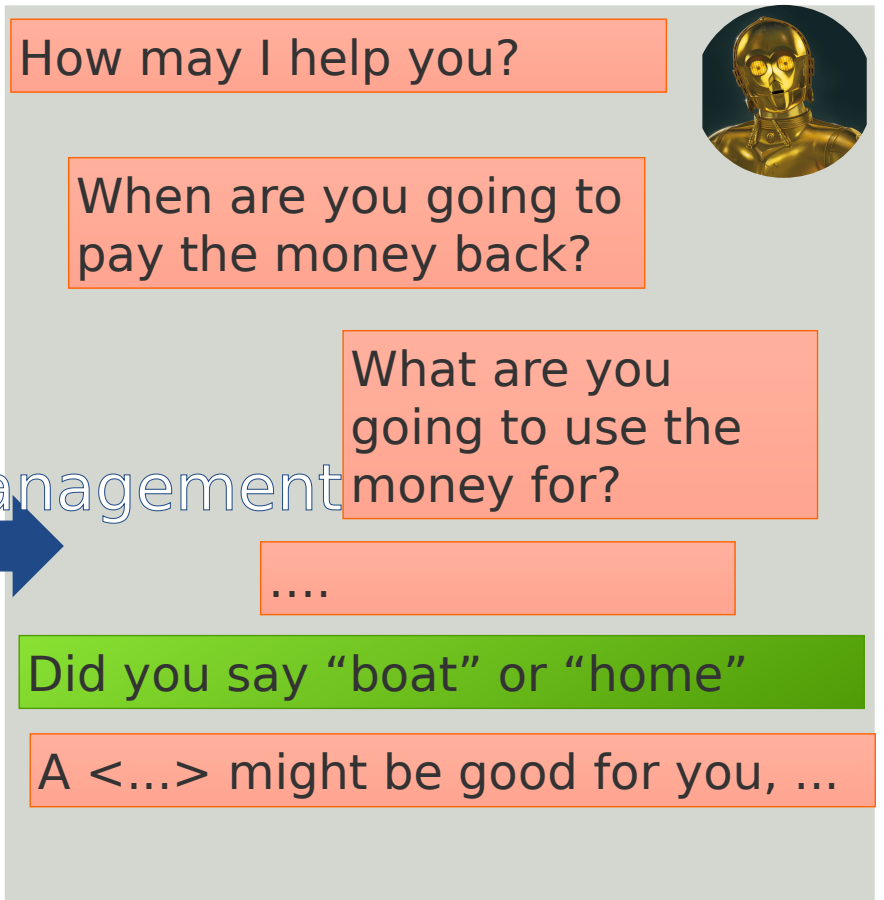
The pipeline



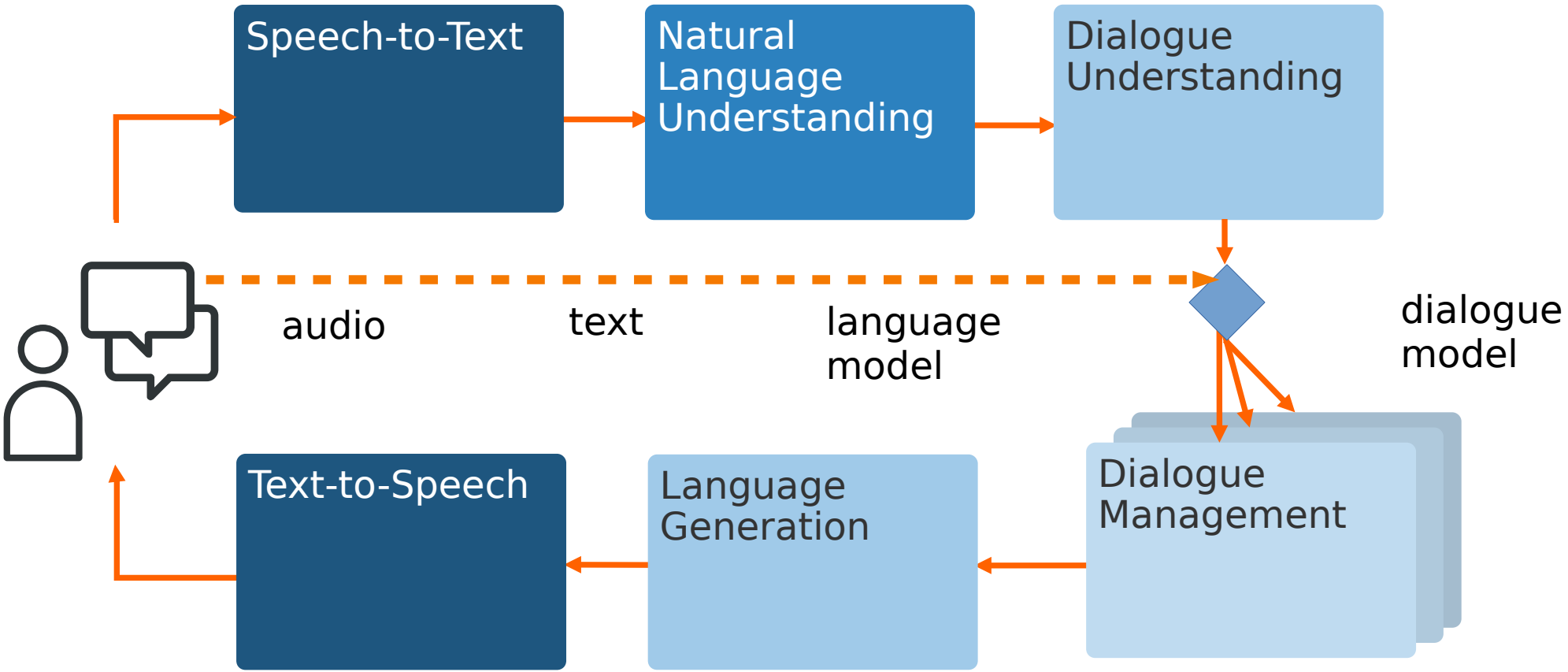
Belief State Model

[young2007]

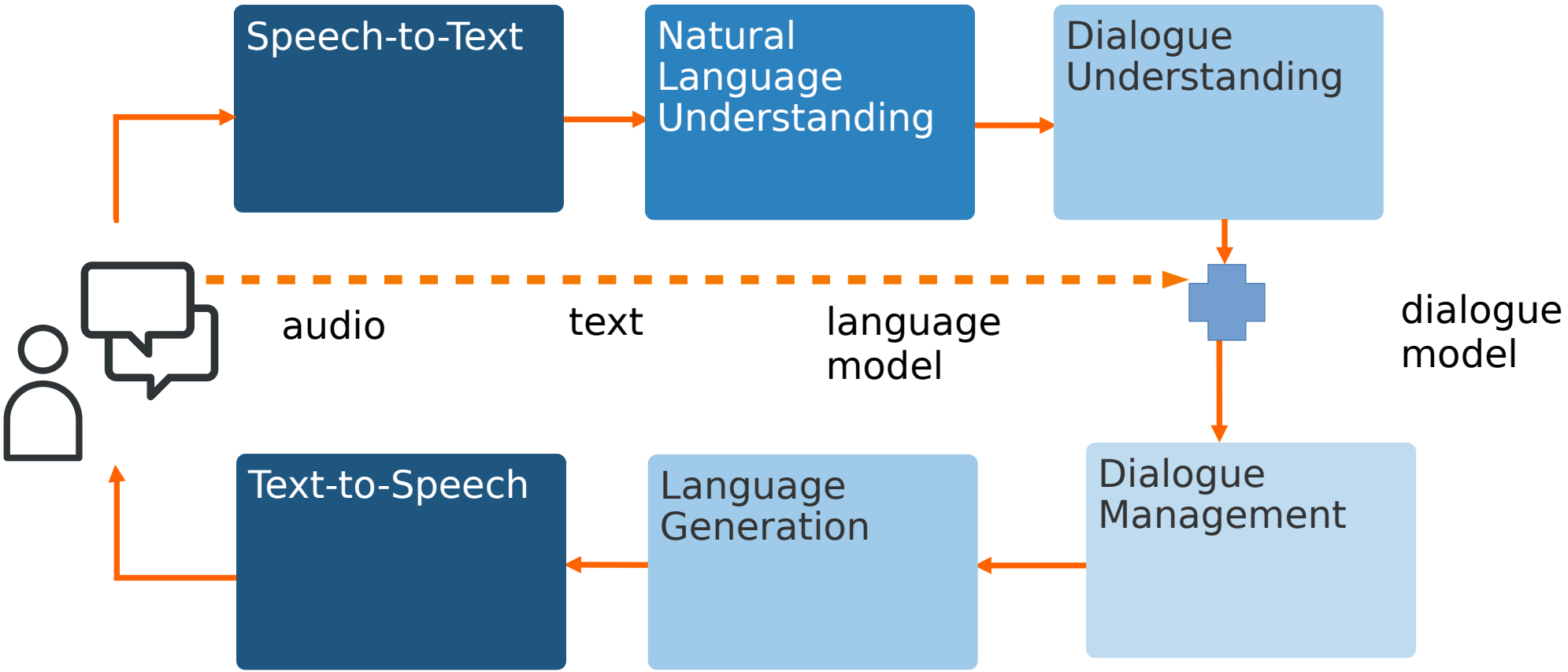




Segmentation-based Personalization



Belief-state-based Personalization



Floris den Hengst, AI for Fintech ICAI lab, Floris.den.Hengst@ing.com

Based on [ultes2017]

Experimental setup (1/2)

Recommendation scenarios

1. Restaurant 1
2. Restaurant 2
3. Laptop
4. Financial products

Reward based on task completion and # turns

Simulation

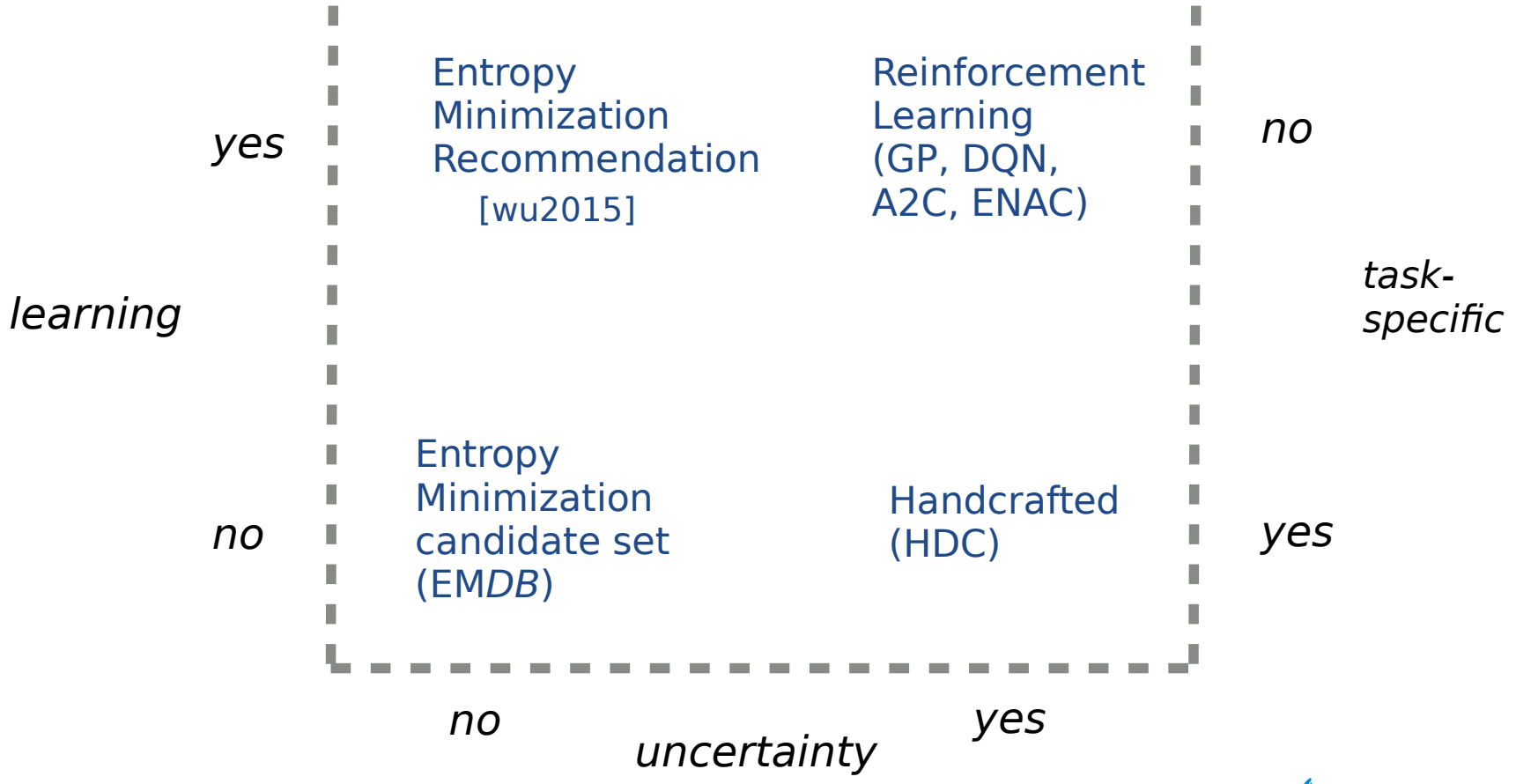
- different user behavior patterns 2
  1. Layperson
  2. Expert
- levels of S2T + NLU error .0, .15, .30
- total number of environments 24

Algorithms varying in

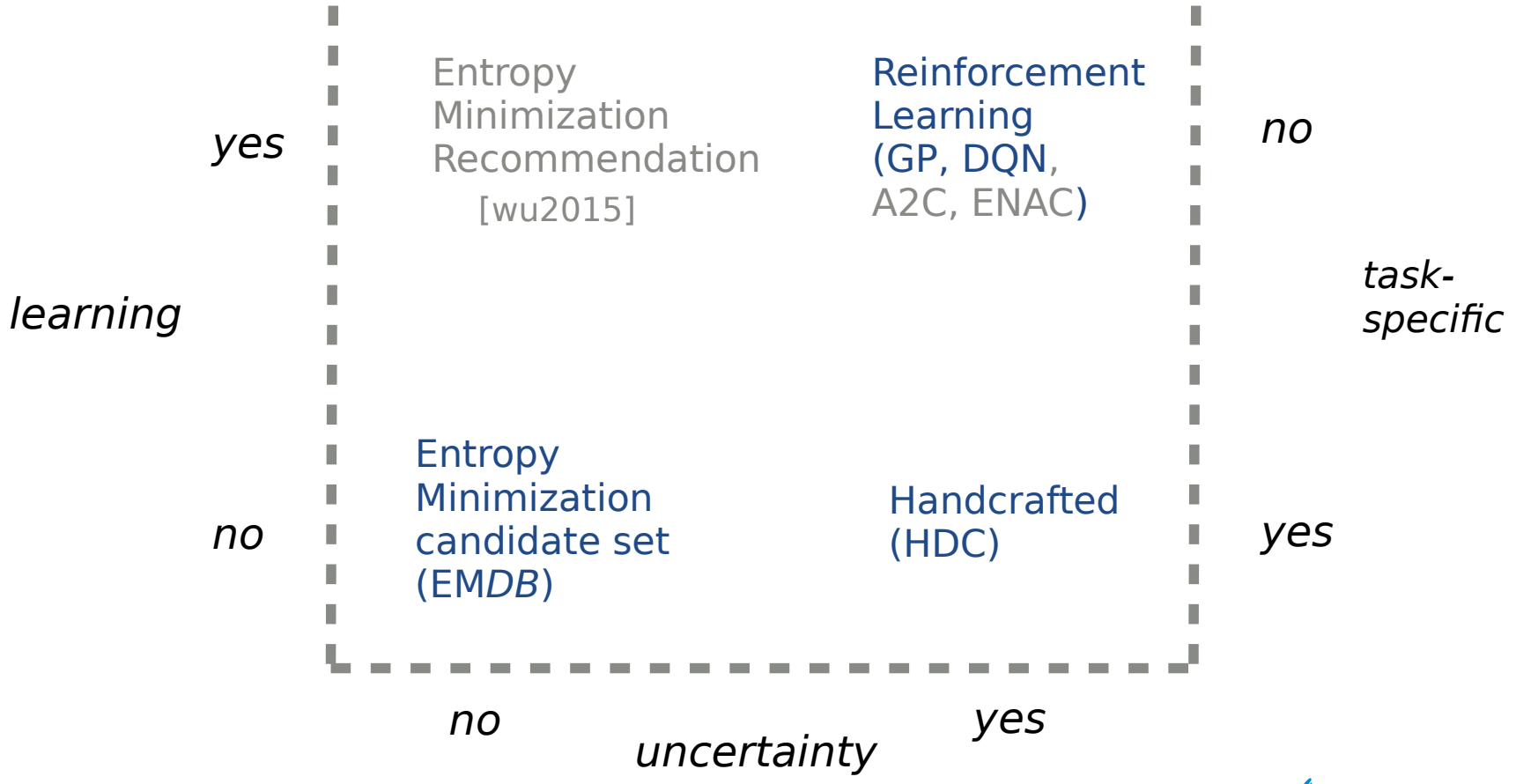
- Taking into account uncertainty
- Ability to learn from experience
- Using task-specific heuristics

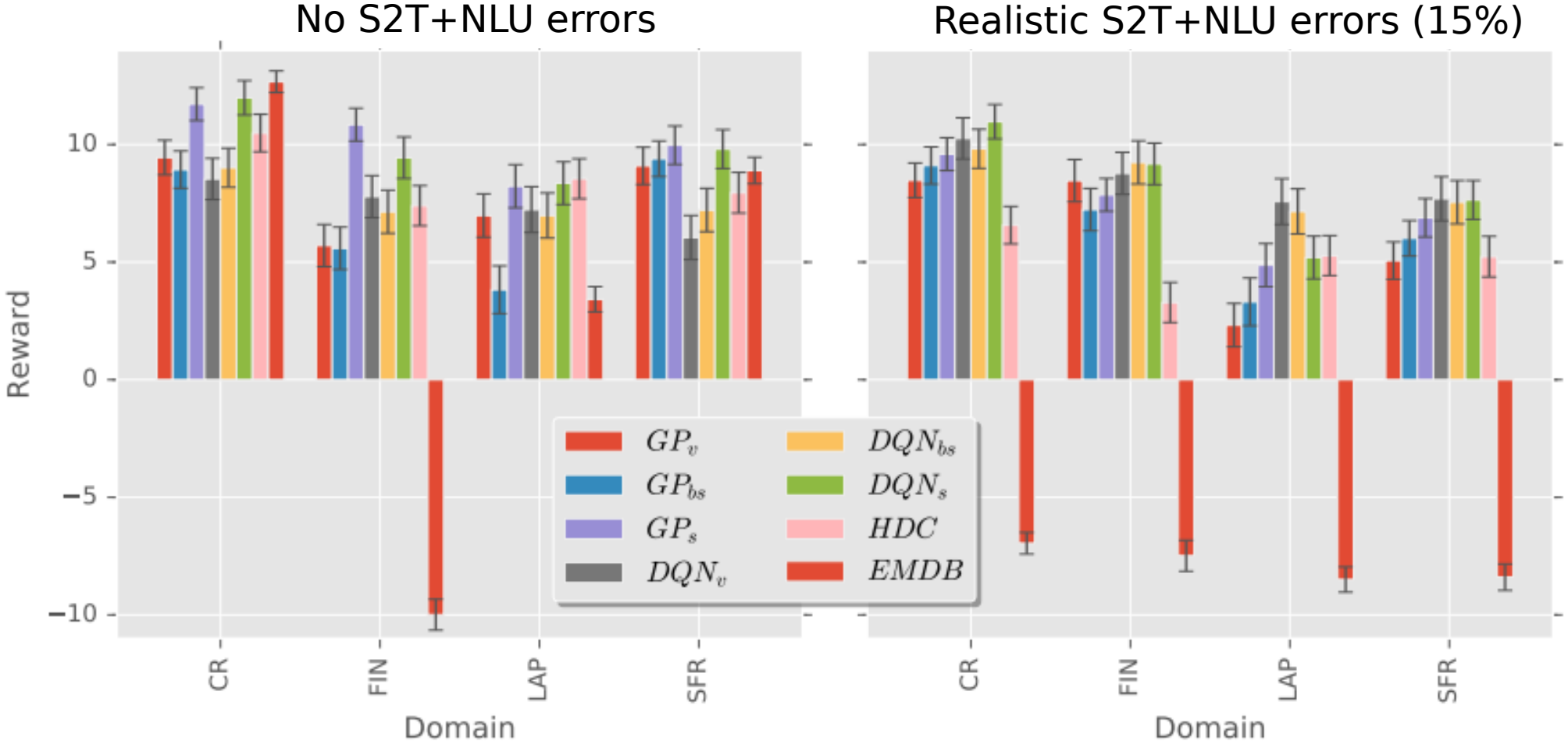
Total environment – algorithm pairs 384

Experimental setup (2/2)



Experimental setup (2/2)




No S2T+NLU errors
Realistic S2T+NLU errors (15%)
Reward
10
5
0
−5
−10
CR
FIN
LAP
SFR
Domain
GP_v
GP_{bs}
GP_s
DQN_v
DQN_{bs}
DQN_s
HDC
$EMDB$

Conclusions & Discussion

Take uncertainty into account

Learning approaches most robust to
- novel domain
- personalization setting

Personalized ≥ gold-standard handcrafted approach

Performance personalized approaches varies with
- environment
- algorithm
- available data

Next Steps (1/2)

Collecting High-Quality Dialogue User Satisfaction Ratings with Third-Party Annotators

Mickey van Zeelt
ING Bank N.V.
mickeyvanzeelt@hotmail.com

Floris den Hengst
ING Bank N.V.
florisdenhengst@gmail.com

Seyyed Hadi Hashemi
University of Amsterdam
hashemi@uva.nl

ABSTRACT

The design, evaluation and adaptation of conversational information systems are typically guided by ratings from third-party, i.e. non-user, annotators. Interfaces used in gathering such ratings are designed in an ad-hoc fashion as it has not yet been investigated which design yields high-quality ratings. This work describes how to design user interfaces for gathering high-quality ratings with third-party annotators. In a user study, we compare a base interface that consolidates best practices from literature, an interface with clear definitions and an interface in which tasks are separated visually. We find that these interfaces yield annotations of high quality and separation of tasks. We find no significant improvements in quality between UIs. This work can serve as a starting point for researchers and practitioners interested in collecting high-quality dialogue user satisfaction ratings using third-party annotators.

CCS CONCEPTS

• Human-centered computing → Natural language interfaces; User interface design; Empirical studies in HCI; • Information

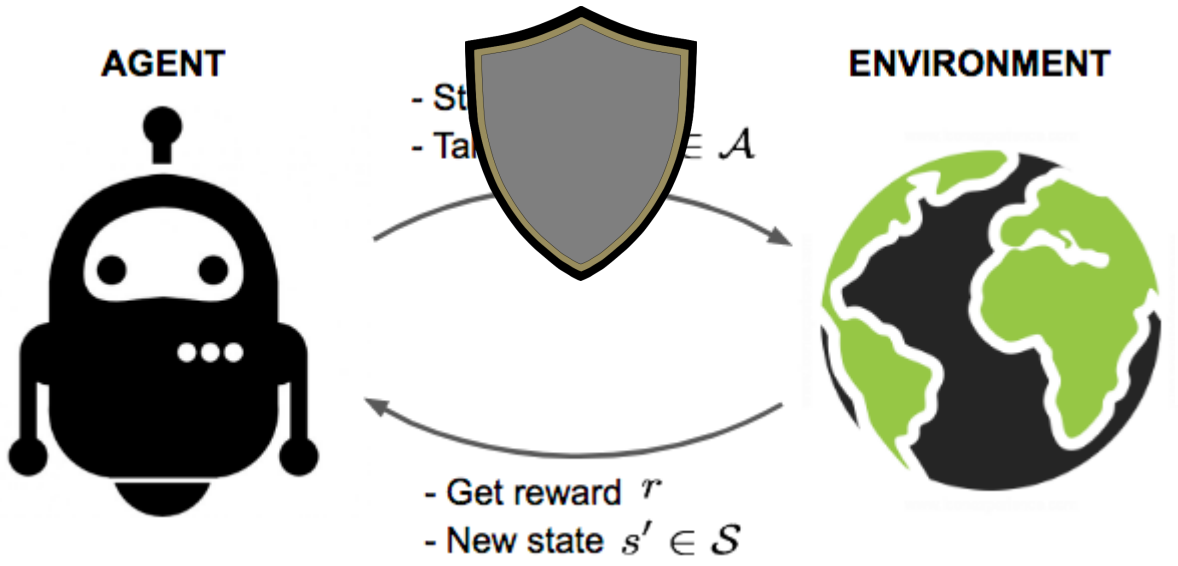
for subtasks such as converting text to speech and extracting keywords from utterances, the overall quality of the entire interface typically needs to be evaluated per system deployment. Furthermore, quality ratings may be required to adapt and personalize the interface in an online fashion [6].

Although various signals for capturing the quality of conversational interfaces have been studied, user satisfaction is typically the ultimate metric to optimize for. User satisfaction is a subjective measure of the quality of an interaction [15] and a rating of user satisfaction can be acquired from users directly or from third-party annotators. These two types of ratings are considered complementary [27]. Third-party ratings come with the inherent challenge that the annotator does not know the intent of the user. In contrast to user ratings though, they can be acquired at manageable costs in a controlled environment.

Acquiring third-party ratings of high quality remains challenging, though. Previous research frequently gathered third-party ratings with low reliability or agreement between raters [1, 10, 12, 13, 25, 29]. The reasons for this have been studied to some ex-

VU VRIJE UNIVERSITEIT AMSTERDAM ING

Next Steps (2/2)



References

[bransford1972]
Bransford, J.D., & Johnson, M.K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. Journal of Verbal Learning and Verbal Behavior, 11, 717-726.

[peckham1991]
Peckham, Jeremy. "Speech Understanding and Dialogue over the telephone: an overview of the ESPRIT SUNDIAL project." Speech and Natural Language: Proceedings of a Workshop Held at Pacific Grove, California, February 19-22, 1991. 1991.

[ultes2017]
Ultes, Stefan, et al. "Pydial: A multi-domain statistical dialogue system toolkit." Proceedings of ACL 2017, System Demonstrations. 2017.

[young2007]
Young, Steve, et al. "The hidden information state approach to dialog management." 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07. Vol. 4. IEEE, 2007.

[casanueva2015]
Casanueva, Inigo, et al. "Knowledge transfer between speakers for personalised dialogue management." Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue. 2015.

[mo2018]
Mo, Kaixiang, et al. "Personalizing a dialogue system with transfer reinforcement learning." Thirty-Second AAAI Conference on Artificial Intelligence. 2018.

[genevay2016]
Genevay, Aude, and Romain Laroche. "Transfer learning for user adaptation in spoken dialogue systems." Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems. International Foundation for Autonomous Agents and Multiagent Systems, 2016.

[wu2015]
Ji Wu, Miao Li, and Chin-Hui Lee. An entropy minimization framework for goaldriven dialogue management. In Sixteenth Annual Conference of the InternationSpeech Communication Association, 2015.

Thank you

Floris.den.Hengst@ing.com

florisdh.nl/presentations/AIRlab_2020.pdf

[ultes2017]

	hidden layer 1	hidden layer 2	ϵ
DQN	300	100	.5
A2C	200	75	.5
eNAC	130	50	.5

Personalizing DM

	[casanueva 2015]	[mo2018]	[genevay 2016]	This talk: segmentation based	This talk: belief-state based
Assumes pre-existing interactions with user		✓			
Assumes user similarity metric	✓			✓	
Small number of users			✓		
Assumes existing information on user	= personal context			✓	✓