Reinforcement Learning for Personalized Dialogue Management

Floris den Hengst, Mark Hoogendoorn, Frank van Harmelen & Joost Bosman

VU VRIJE UNIVERSITEIT AMSTERDAM

ING

Dialogue Systems

Dialogflow

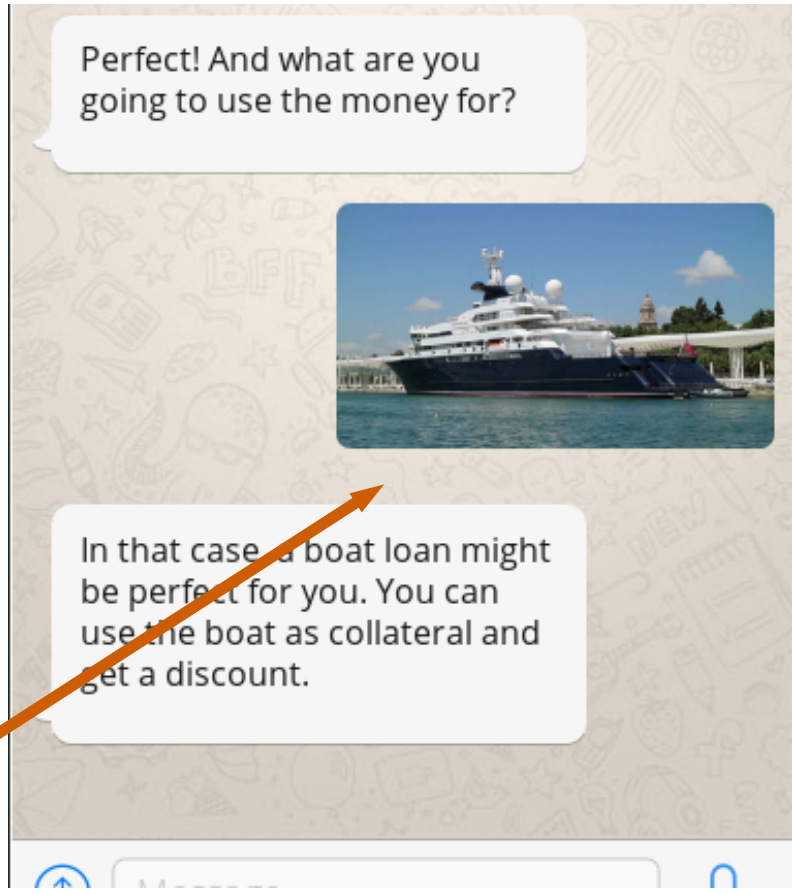
Language and context

[bransford1972]

Intent

Slots:
duration & purpose

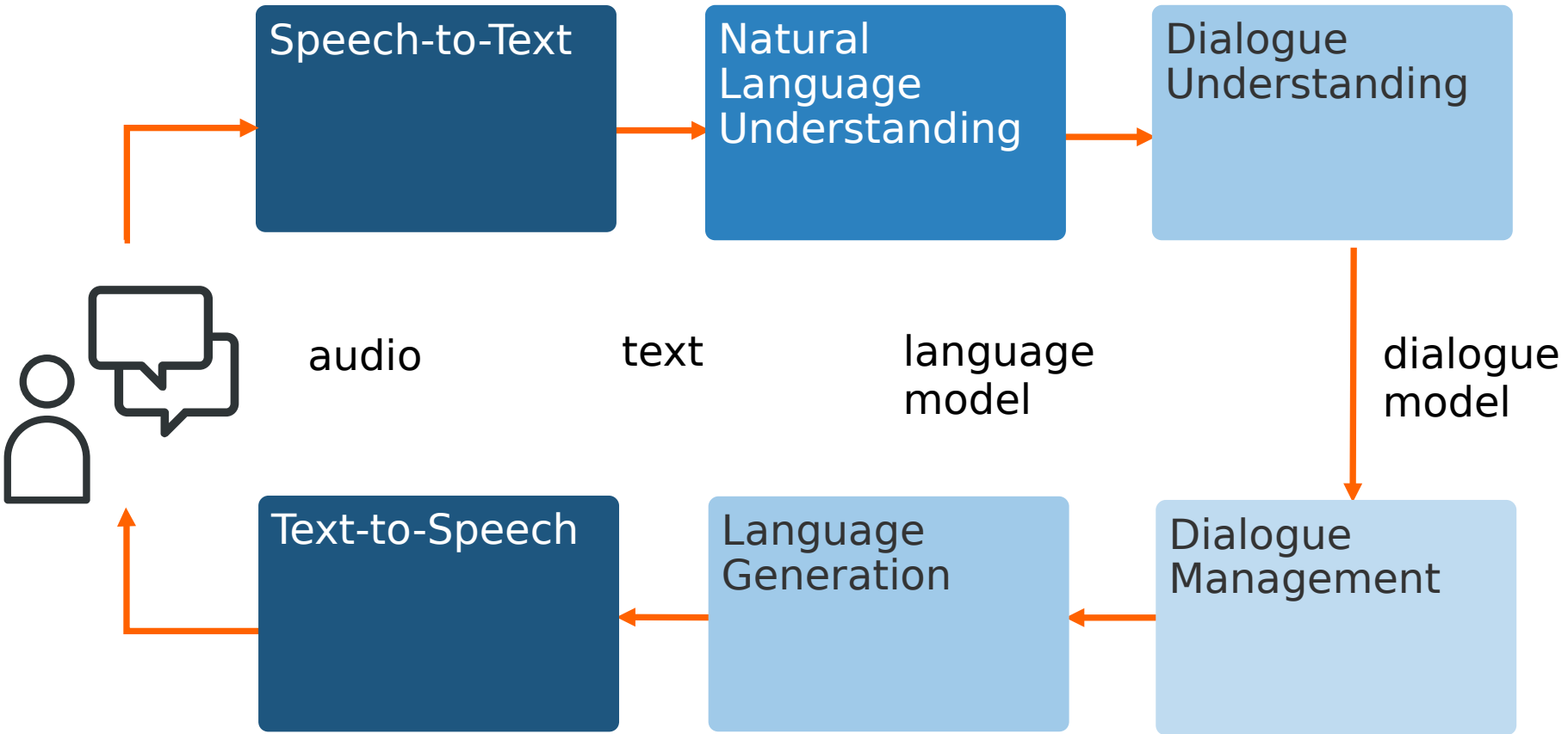
Values:
duration & purpose



VU VRIJE UNIVERSITEIT AMSTERDAM ING

Architecture

[peckham1991]
[ultes2017]



Sequential Model

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

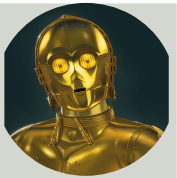
1. Intent: “loan”, purpose:”boat”
2. Duration: “16M”

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, …

….

FSM 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?

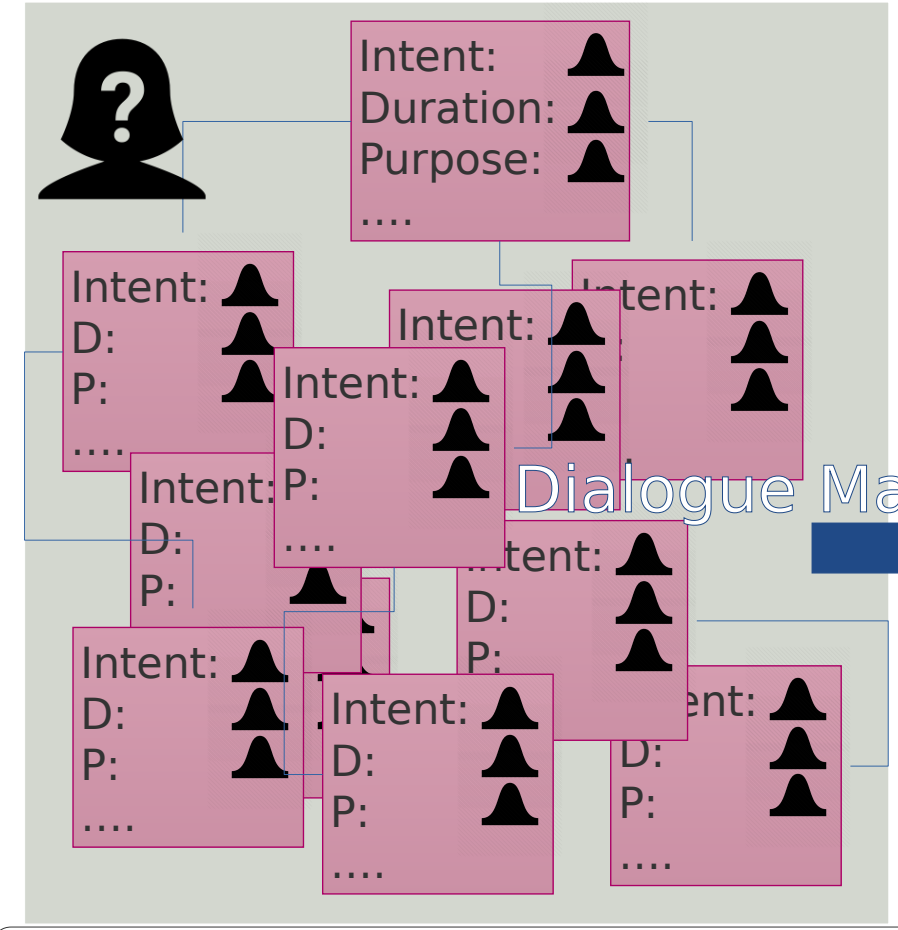
What are you going to use the money for?

....

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

Belief State Model

[young2007]



How may I help you?

When are you going to pay the money back?

What are you going to use the money for?

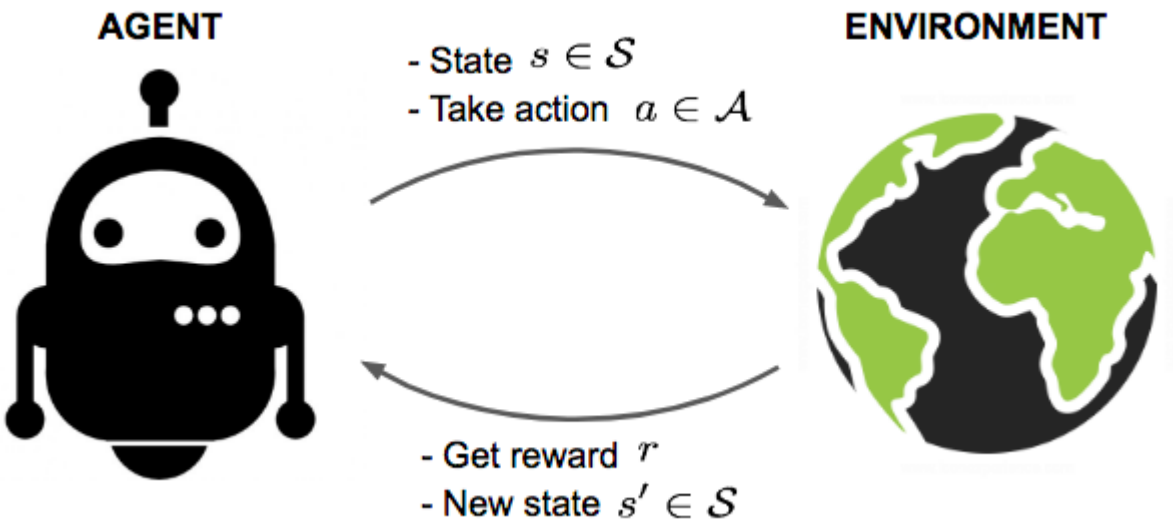
....

Did you say “boat” or “home”

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

VU VRIJE UNIVERSITEIT AMSTERDAM ING

Reinforcement Learning (1/2)

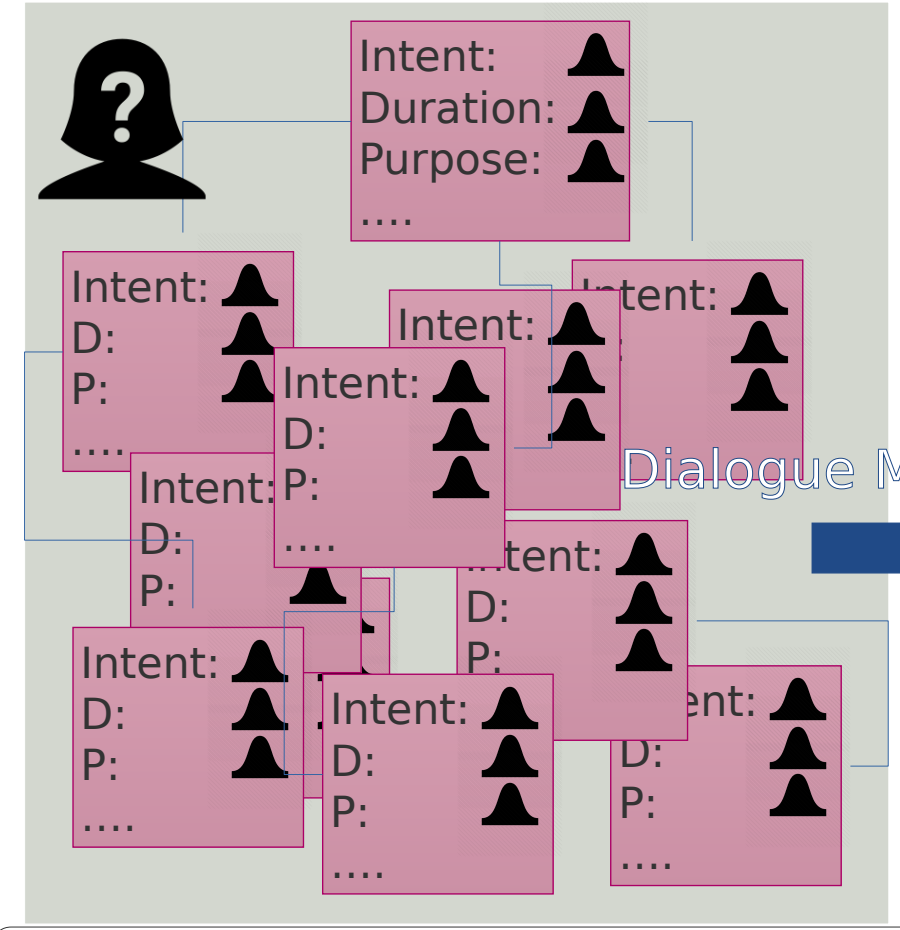


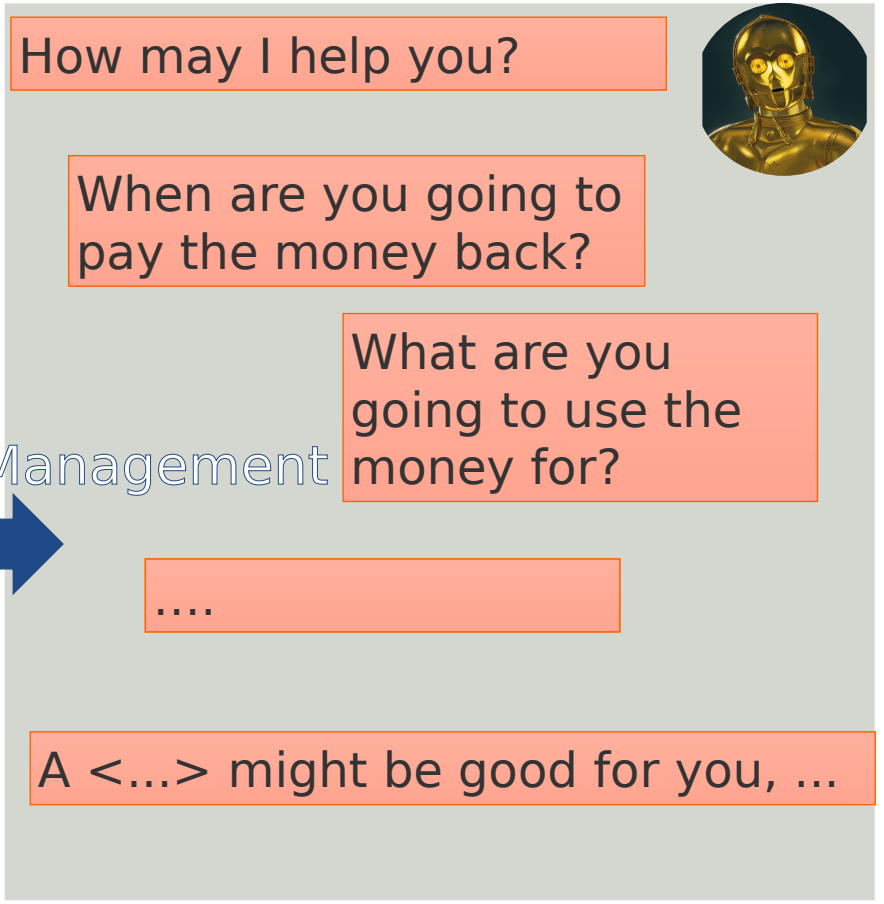
Trajectory $\langle s_0, a_0, r_0, s_1, \ldots, s_T, a_T, r_T \rangle, r_. \in \mathbb{R}$

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

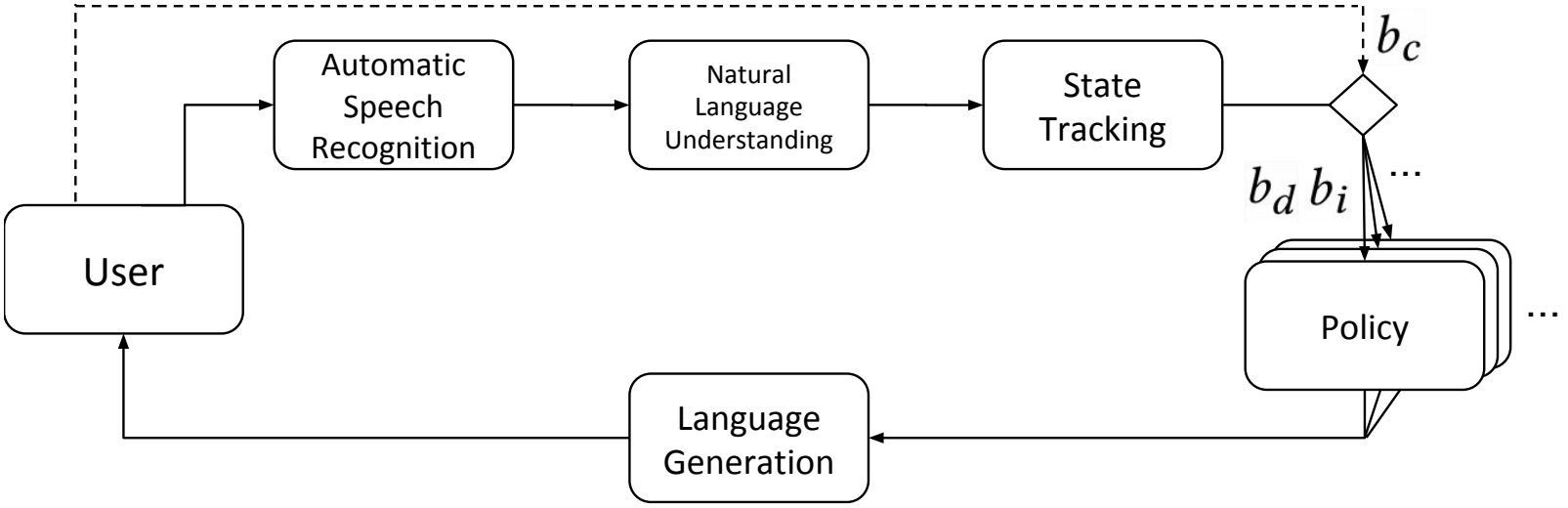
Belief State Representation

[young2007]



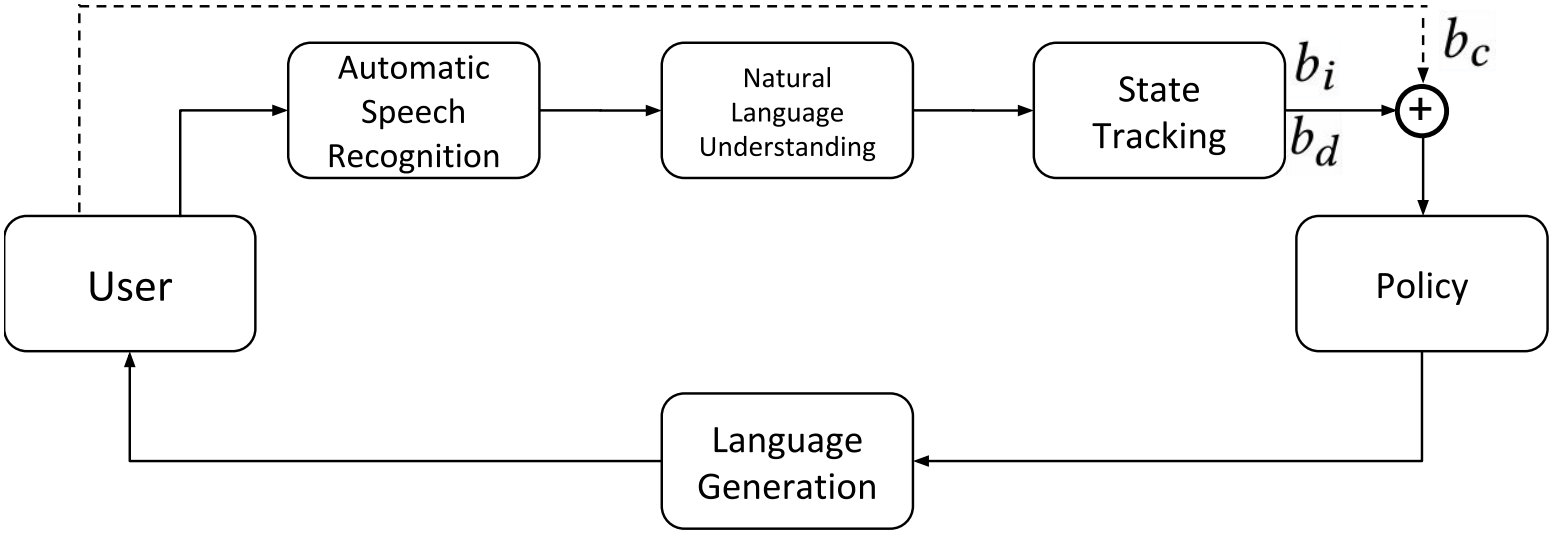


Segmentation-based Personalization



0. Maintain a belief over personal context
1. Segment users based on belief
2. Learn 1 DM policy per segment

Belief State-based Personalization



0. Maintain a belief over personal context
1. Include belief into DM policy input
2. Learn 1 DM policy across all users

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

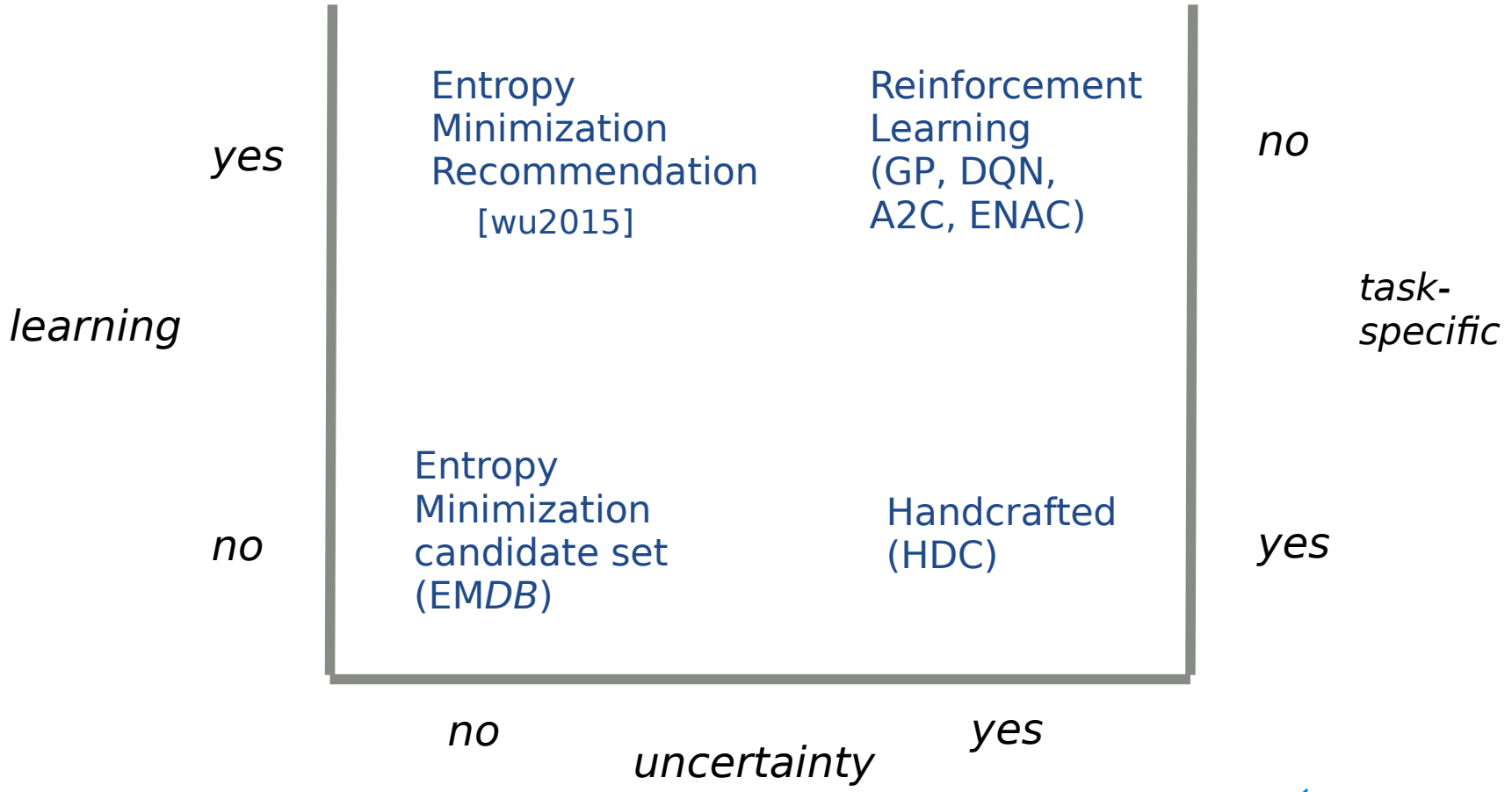
- different user behavior patterns 2
  1. Layperson
  2. Expert
- levels of S2T + NLU error .0, .15, .30
- total number of environments 24

Algorithms varying in

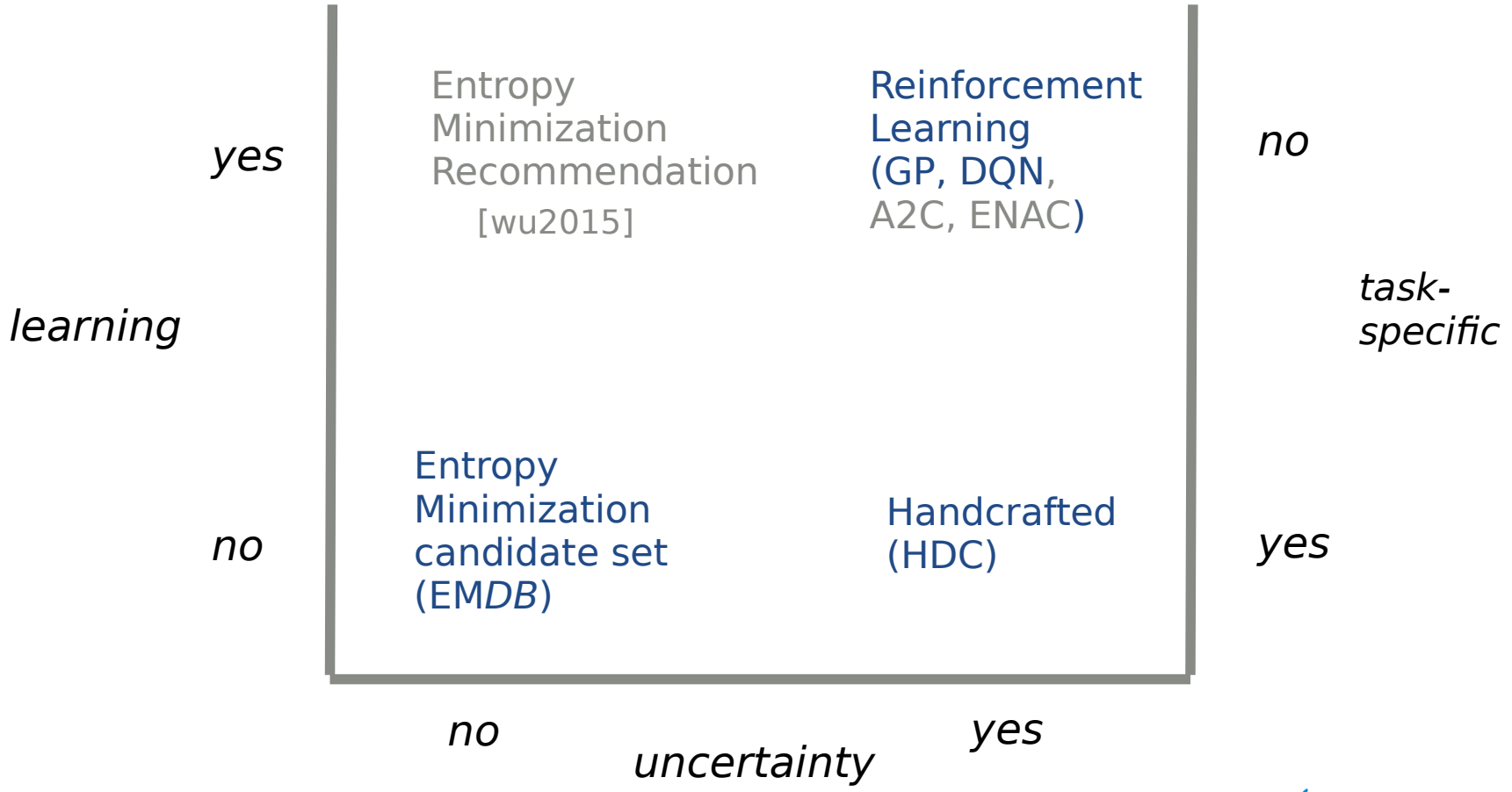
- Taking into account uncertainty
- Ability to learn from experience
- Using task-specific heuristics

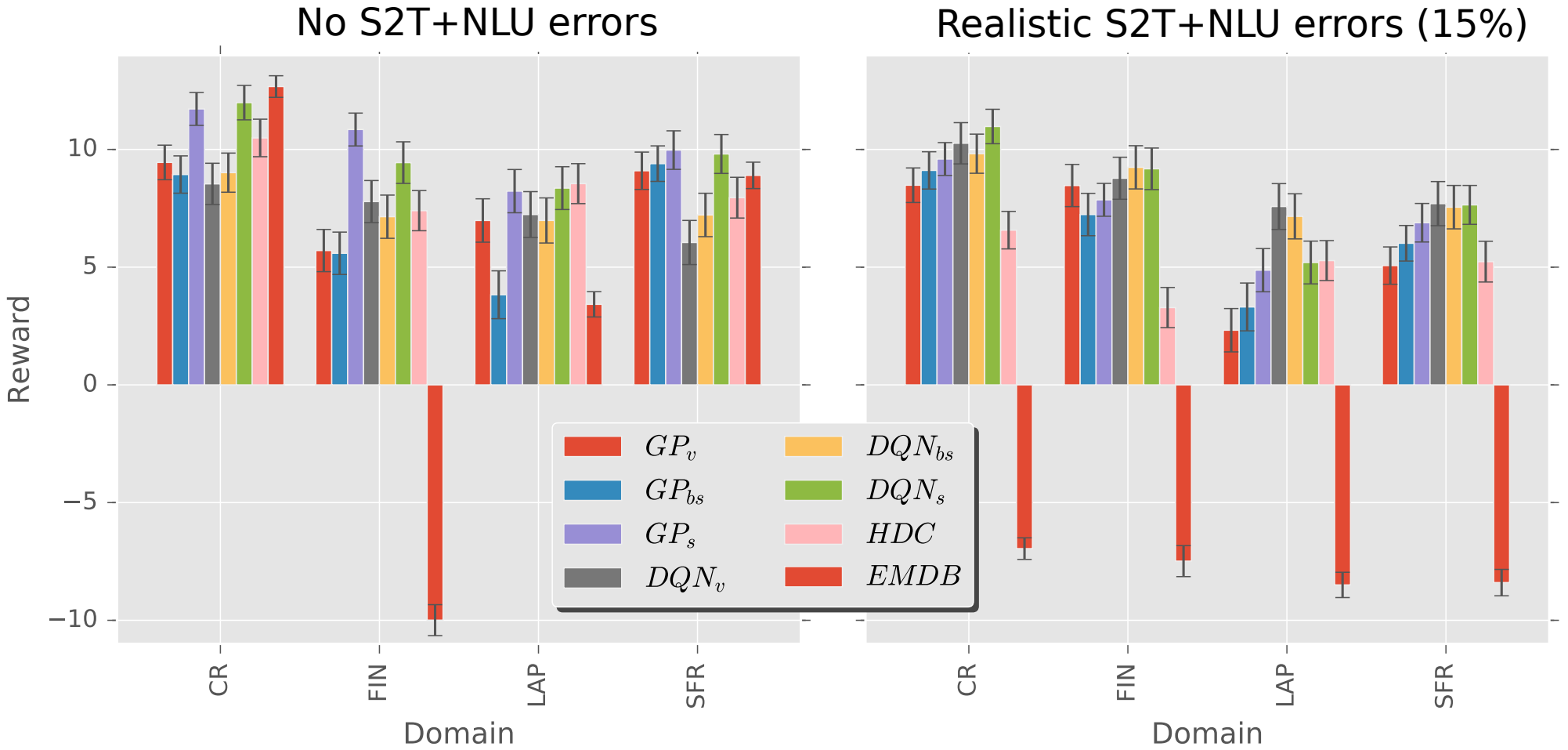
Total environment – algorithm pairs 384

Experimental setup (2/2)



Experimental setup (2/2)





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

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

F.den.hengst@vu.nl

florisdh.nl/presentations/wi-2019.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			✓	✓