Dialog Management with MDP and POMDP

Max Friedrich

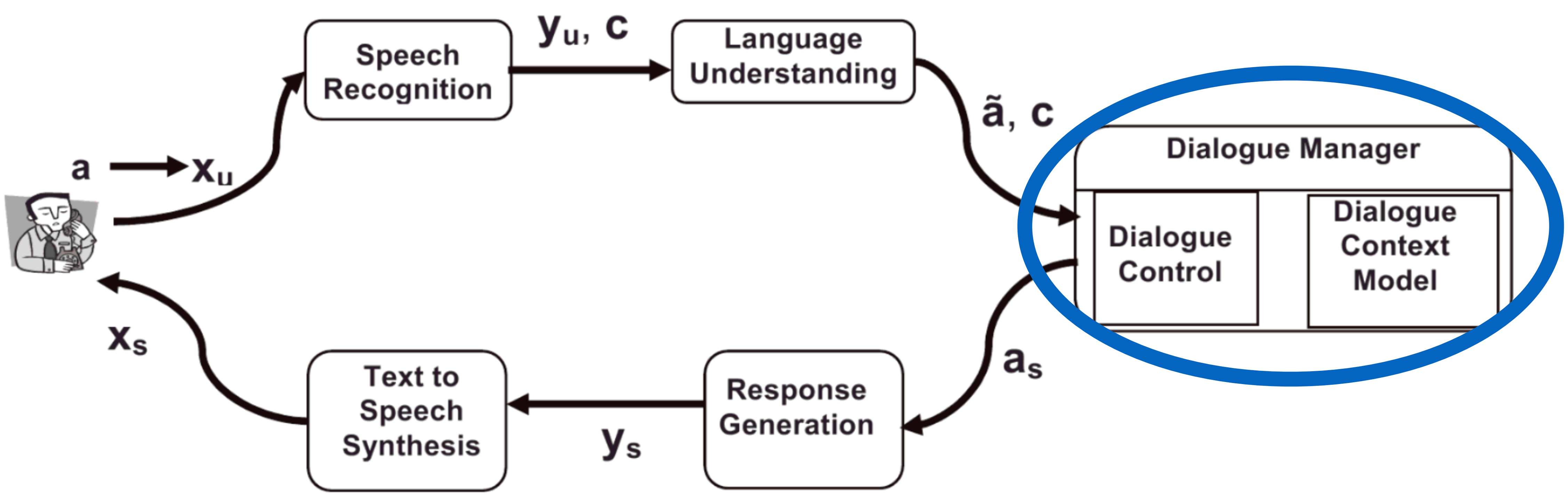
Seminar Speech Technology, SS 2016

June 15, 2016

Outline

1. **Recap: Dialog Management, Reinforcement Learning**
2. Dialog Management with MDP
3. Dialog Management with POMDP
4. Conclusion

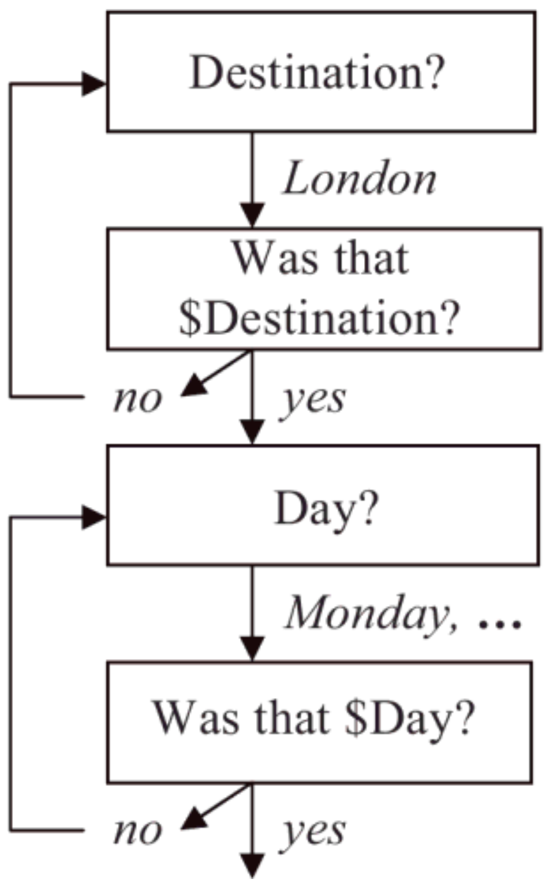
Dialog Management



[Jokinen and McTear, 2009]

Dialog Control

Classical approaches: graphs and frames

Graphs	Frames
	(1) destination: unknown date: unknown time of departure: unknown (2) destination: London date: unknown time of departure: 9

[Jokinen and McTear, 2009]

Dialog Control

- Handcrafted rules are hard to create
  - Error handling, when to ask for confirmation, …
- Alternative: Statistical approaches, including Reinforcement Learning

Recap: Reinforcement Learning (RL)

- "Goal directed learning from interaction"
- Finds a policy π that maximizes reward
- Between supervised and unsupervised learning
- Problems often specified as Markov Decision Processes (MDP)

[Sutton and Barto, 2008; Marsland, 2009]

Outline

1. Recap: Dialog Management, Reinforcement Learning
2. **Dialog Management with MDP**
3. Dialog Management with POMDP
4. Conclusion

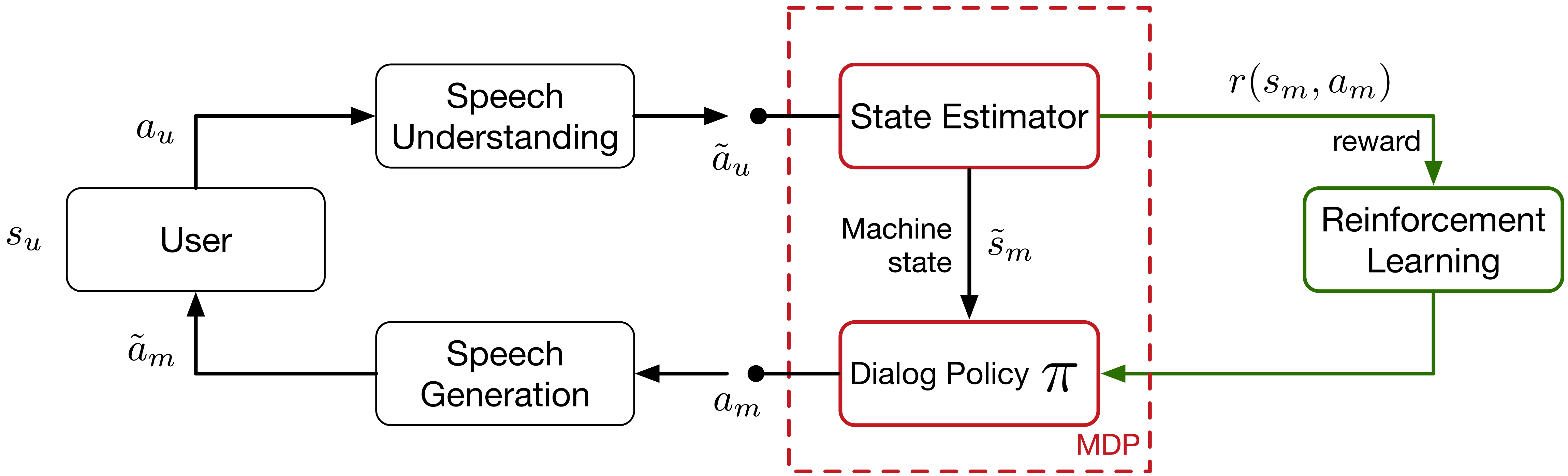
Markov Decision Process (MDP)

S	set of system states
A	set of actions that the system can take
T	transition probabilities $P_T(s_t \mid s_{t-1}, a_{t-1})$
R	immediate rewards $R : S \times A \rightarrow \mathbb{R}$

Policy $\pi : S \rightarrow A$

[Jokinen and McTear, 2009; Levin et al., 2000]

MDP Dialog Management

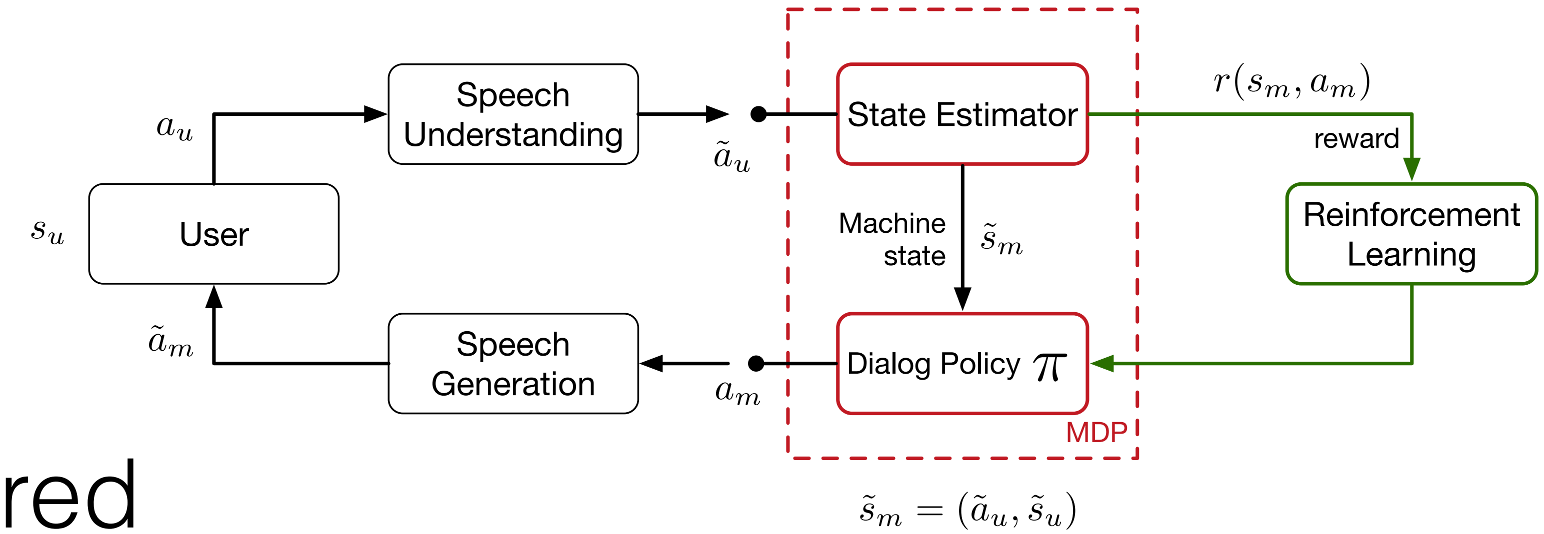


$$\tilde{s}_m = (\tilde{a}_u, \tilde{s}_u)$$

What are problems with MDP Dialog Management?
(small groups, 3 minutes)

redrawn after [Young, 2006]

Limitations of MDP DM



- Huge state space
- Much training required
- What is a good reward function?
- Dialog state is estimated – Single dialog hypothesis
- “Handcrafted vs Machine Learning”

Single Hypothesis Problem

Example: Dialog system that offers travel booking and coffee making

U	“Please make me a cup of coffee” [40% “I want to go to Berlin”, 20% “Please make me a cup of coffee”]
M	*When do you want to go?* – Dialog state: user wants to go to Berlin
U	“Coffee!” [70% “Coffee”, 20% “Tonight”] – how to proceed?

Outline

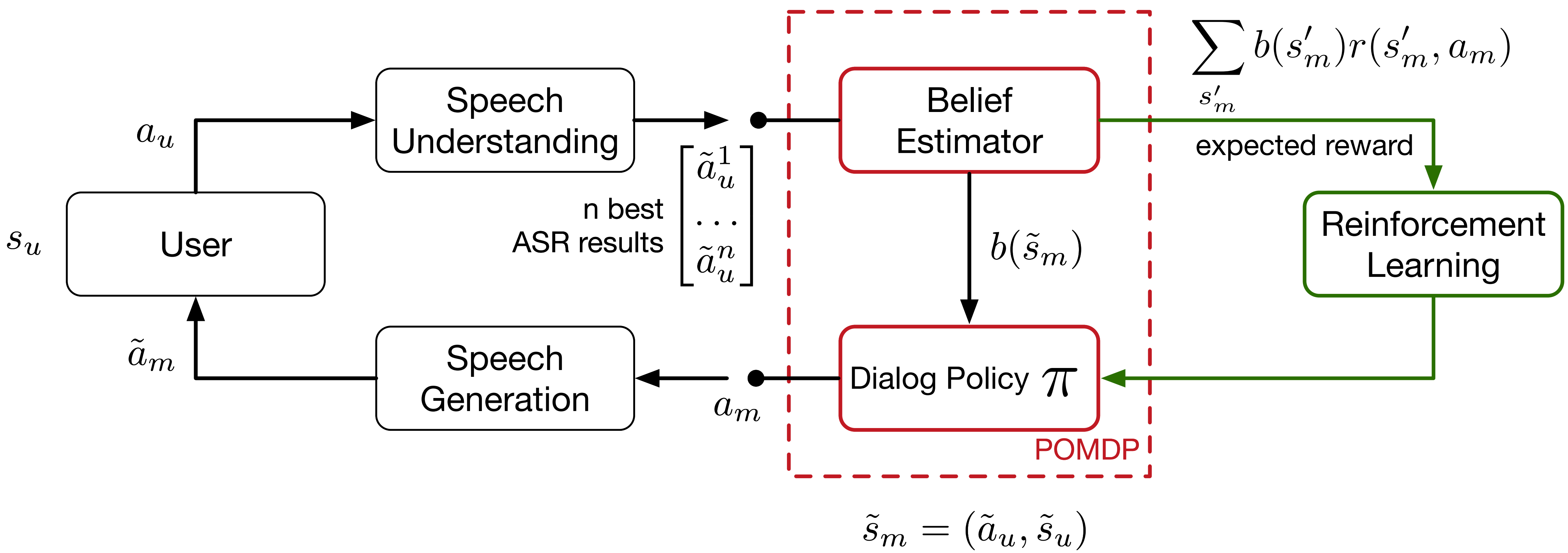
1. Recap: Dialog Management, Reinforcement Learning
2. Dialog Management with MDP
3. **Dialog Management with POMDP**
4. Conclusion

Partially Observable Markov Decision Process (**PO**MDP)

S	set of system states
A	set of actions that the system can take
T	transition probabilities $P_T(s_t \mid s_{t-1}, a_{t-1})$
R	immediate rewards $R : S \times A \times \boldsymbol{O} \rightarrow \mathbb{R}$
$\boldsymbol{O}$	**observations**
$\boldsymbol{Z}$	**observation probabilities** $\boldsymbol{P_Z(s_t / s_{t-1}, a_{t-1})}$

[Jokinen and McTear, 2009; Williams and Young, 2007]

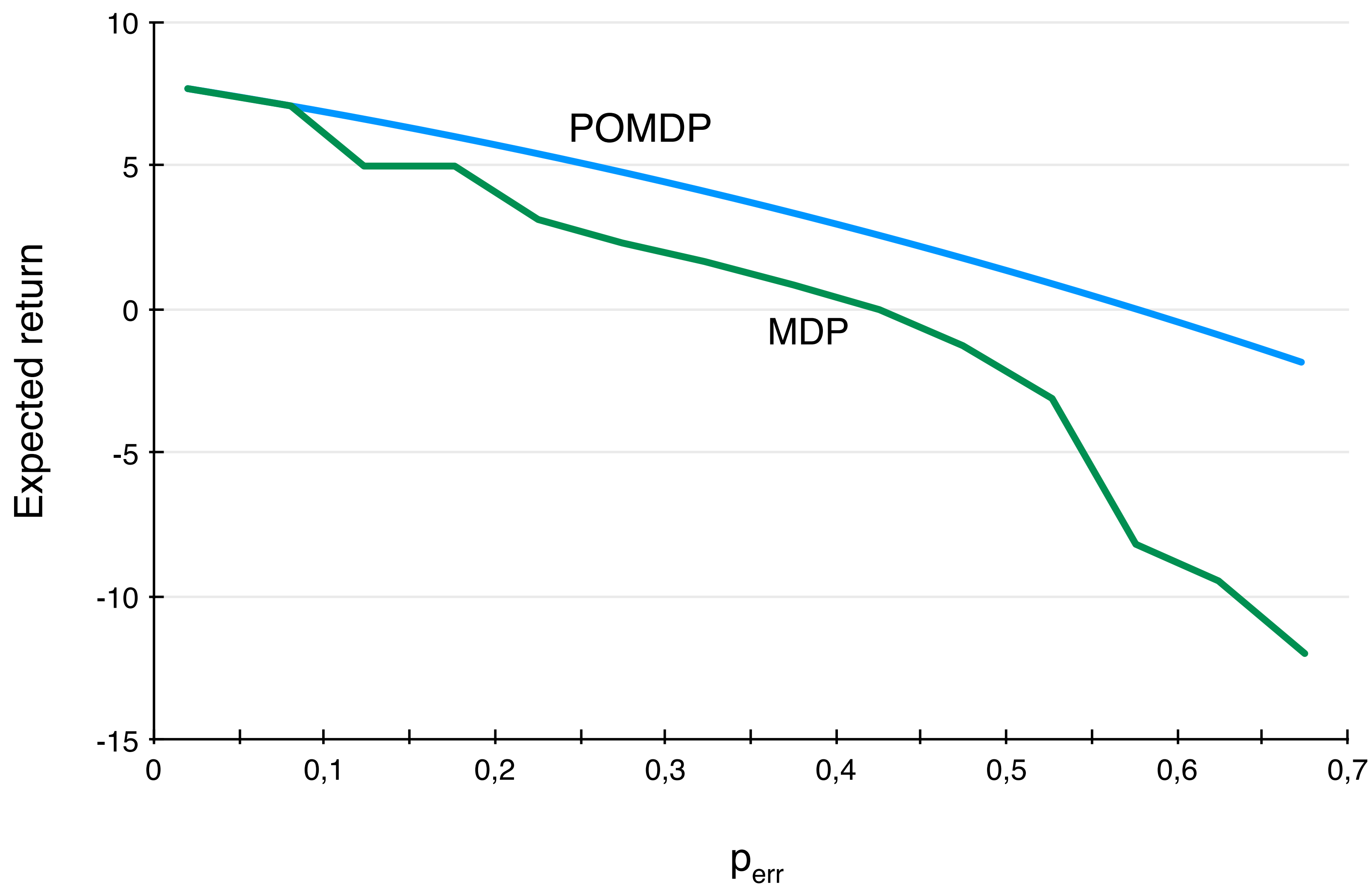
POMDP Dialog Management



redrawn after [Young, 2006]

POMDP Dialog Management

Travel domain (3 cities)



redrawn after [Williams et al., 2005]

Limitations of POMDP DM

- Even bigger state space, very hard to scale. Approximations are needed
- Training, reward function, “Handcrafted vs Machine Learning”: same as MDP
- „Hidden Information State“ [Young et al., 2010] addresses state space problems

Outline

1. Recap: Dialog Management, Reinforcement Learning
2. Dialog Management with MDP
3. Dialog Management with POMDP
4. **Conclusion**

Conclusion

- Dialog Control with Graphs and Frames
- MDP Dialog Management
- POMDP Dialog Management is more robust, explicitly models uncertainty
- State space problem
- “Handcrafted vs. Machine Learning” problem

Bibliography (1)

[Jokinen and McTear, 2009] Kristiina Jokinen and Michael McTear. Spoken Dialogue Systems, volume 2. Morgan & Claypool Publishers, 2009.

[Levin et al., 2009] Esther Levin, Roberto Pieraccini, and Wieland Eckert. *A stochastic model of human-machine interaction for learning dialog strategies*. IEEE Transactions on Speech and Audio Processing, 8(1):11–23, 2000.

[Marsland, 2009] Stephen Marsland. Machine Learning: An Algorithmic Perspective. Chapman & Hall/CRC, 1st edition, 2009.

[Sutton and Barto, 1998] Richard S Sutton and Andrew G Barto. Reinforcement Learning: An introduction. MIT press, 1998.

Bibliography (2)

[Williams et al., 2005] Jason D Williams, Pascal Poupart, and Steve Young. *Factored Partially Observable Markov Decision Processes for Dialogue Management*. In 4th Workshop on Knowledge and Reasoning in Practical Dialog Systems, International Joint Conference on Artificial Intelligence (IJCAI), pages 76–82, 2005.

[Williams, 2007] Jason D Williams. *Applying POMDPs to dialog systems in the troubleshooting domain*. In Proceedings of the Workshop on Bridging the Gap: Academic and Industrial Research in Dialog Technologies, pages 1–8. Association for Computational Linguistics, 2007.

[Young, 2006] Steve Young. *Using POMDPs for Dialog Management*. Presentation slides. http://mi.eng.cam.ac.uk/research/dialogue/slt06_sjy-talk.pdf

[Young et al., 2010] Steve Young, Milica Gašić, Simon Keizer, François Mairesse, Jost Schatzmann, Blaise Thomson, and Kai Yu. *The hidden information state model: A practical framework for POMDP-based spoken dialogue management*. Computer Speech & Language, 24(2):150–174, 2010.