Toward Affective Dialogue Modeling using POMDP

Trung H. Bui
Job Zwiers
Mannes Poel
Anton Nijholt
University of Twente

Bremen, 19 June 2006
Workshop Emotion & Computing

Outline

- Motivation
- POMDP & dialogue management
- Affective dialogue modeling
- Example
- Conclusions and future work

Motivation

- Affective dialogue management (ADM) model is a dialogue management which is able to take into account some aspects of the emotional state and acts appropriately

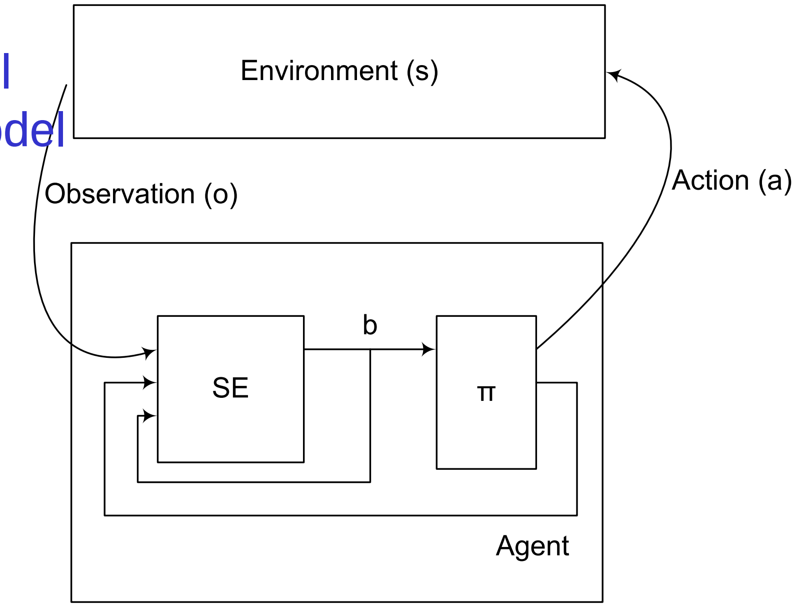
- Scope of the ADM we are focusing on
  - human-computer interaction using multimodal input/output
  - acting appropriately given knowing the user's emotional state and the user action with uncertainty (not emotion recognition, dialogue-act recognition)

- POMDP provides an elegant framework for this type of dialogue models

Partially Observable Markov Decision Process

- <S, A,Z,T,O,R>

S = state set	T = transition model
A = action set	O = observation model
Z = observation set	R = reward model



- Related notations
  - b is the agent's belief state b
  - π is the agent's policy to select the action
- Two main tasks
  - Computing the belief state
  - Finding the optimal policy

Human Media Interaction

Example (Roy et al. 2000)

- <S, A,Z,T,O,R>

output of ASR

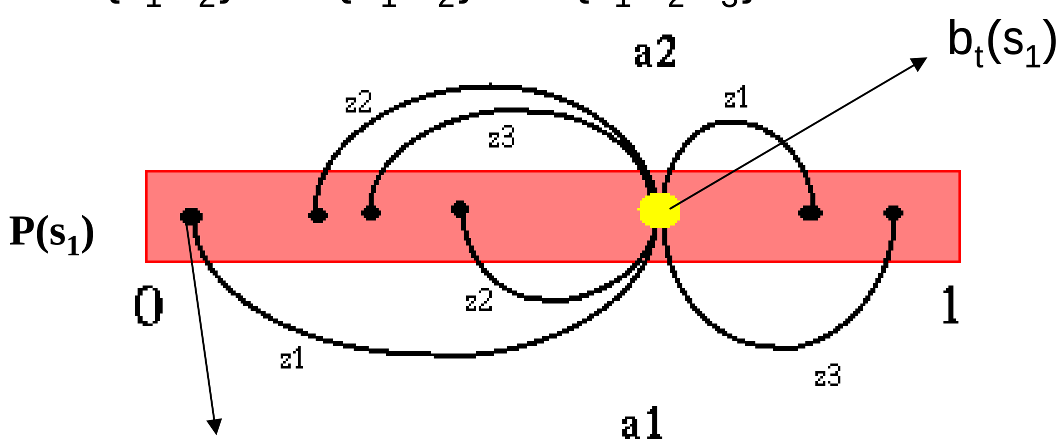
Observation	True State	Belief Entropy	Action	Reward
flo hello	request_begun	0.406	say_hello	100
flo what is like	start_meds	2.735	ask_repeat	-100
flo what time is it for will the	want_time	0.490	say_time	100
flo was on abc	want_tv	1.176	ask_which_station	-1
flo was on abc	want_abc	0.886	say_abc	100
flo what is on nbc	want_nbc	1.375	confirm_channel_nbc	-1
flo yes	want_nbc	0.062	say_nbc	100
flo go to the that pretty good what	send_robot	0.864	ask_robot_where	-1
flo that that hello be	send_robot_bedroom	1.839	confirm_robot_place	-1
flo the bedroom any i	send_robot_bedroom	0.194	go_to_bedroom	100
flo go it eight a hello	send_robot	1.110	ask_robot_where	-1
flo the kitchen hello	send_robot_kitchen	1.184	go_to_kitchen	100

Computing the belief state

$$b_{t+1}(s_{t+1}) = \alpha O(s_{t+1}, a_t, o_{t+1}) \sum_{s_t \in S} T(s_t, a_t, s_{t+1}) b_t(s_t)$$

new belief | normalizing constant | observation model | transition model | old belief

Example: S = $\{s_1, s_2\}$, A=$\{a_1, a_2\}$, Z=$\{z_1, z_2, z_3\}$



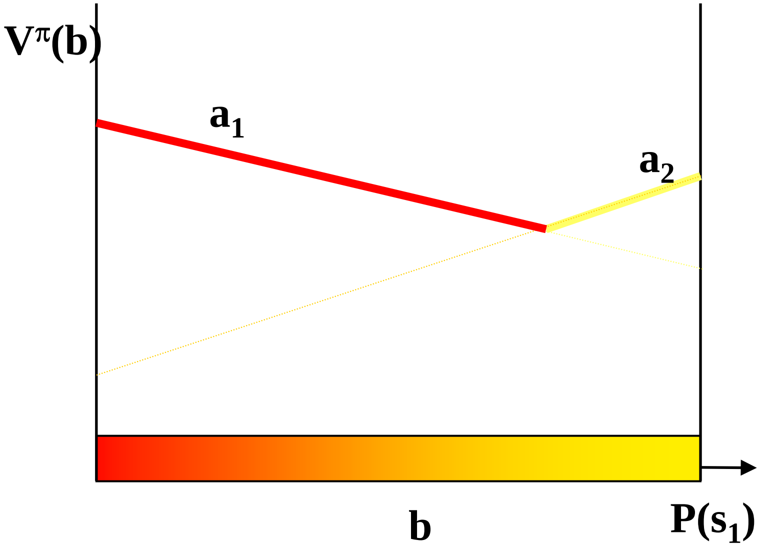
$b_{t+1}(s_1)$ given $a_t=a_1, s_{t+1}=s_1, o_{t+1}=z_1$

Finding the optimal policy

- $V^{\pi}(b)$ = expected total discounted future reward starting from b for a policy π

$$V^{\pi}(b) = \max_{a \in A}\left[R(b,a) + \gamma \sum_{b' \in B} T(b,a,b') V^{\pi}(b') \right]$$

γ is the discount factor



- The optimal policy:

$$\pi^* = \arg\max_{\pi} E[V^{\pi}(b)]$$

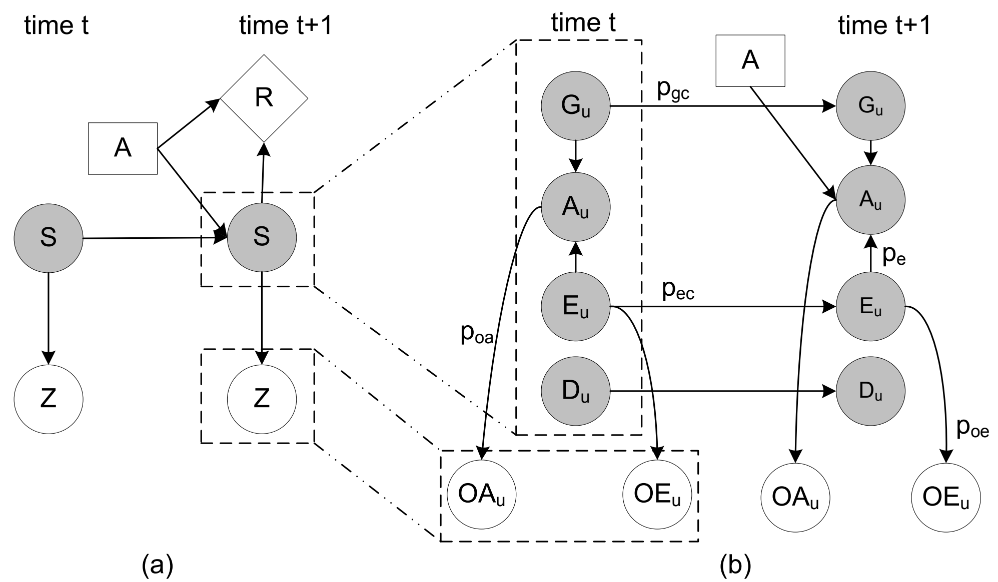
POMDP Dialogue management

Works	Slots, states, actions, obs.	Algorithm, strategy	Reward model
Nursing home robot (Roy et al. 2000)	4,13,20,16	Augmented MDP	Each action labeled as Correct(+100), OK(-1) or Wrong(-100)
Tours guide (Zhang et al. 2001)	30,40,18,25	QMDP,FIB,Grid-based appro.	positive reward when the answer matches user's request negative reward if mismatch occurs
Robot interface domain (Pineau & Thrun 2001)	3,10,15,16	Incremental Prunning	Computation time in seconds
Travel booking (Williams & Young 2005)	2,36,5,5	Perseus	-1 if ask slot not stated, -3 if confirm slot not stated -2 if ask slot stated, -1 if confirm slot stated -3 if ask slot confirmed, -2 if confirm slot confirmed +50 if dialogue goal ends successfully, -50 otherwise

→ Focus on spoken dialogue management, noisy environment

Proposed POMDP Affective dialogue model

- Using a two time-slice Dynamic Bayesian network of factored POMDP
- State set and observation set are composed of 6 features
  - State set: user's goal (G_u), user's emotional state (E_u), user's action (A_u), & user's dialogue state (D_u)
  - Observation set: observed user's action (OA_u) & observed user's emotional state (OE_u)

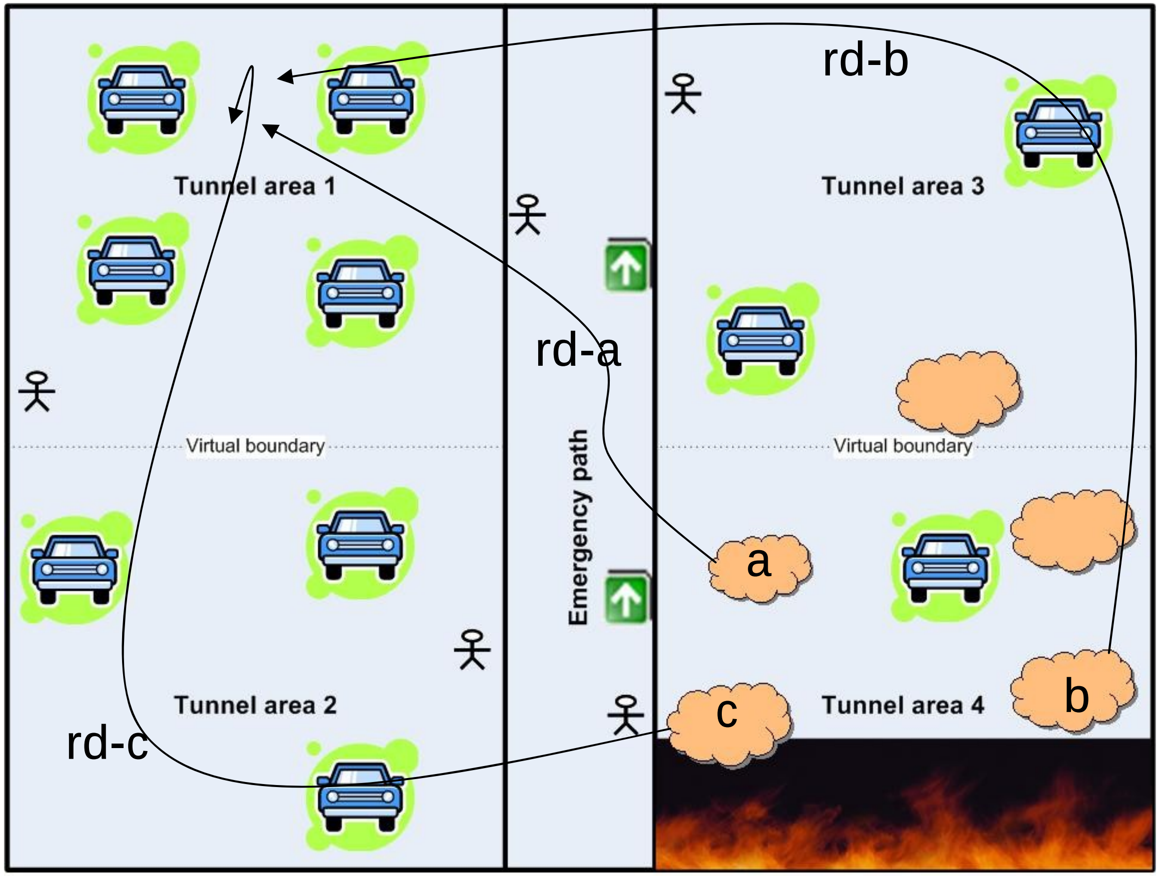


Transition model & observation model

- No data available → Use parameters
  - p_{gc} & p_{ec} are the probability the user's goal & emotion change
  - p_e is the probability of the user's action error being induced by emotion
  - p_{oa} & p_{oe} are the probabilities of the observed action & observed emotional state errors

- Partial or full data available → construct and adjust the model from the collected data

Example: Simulated Route navigation in the unsafe tunnel

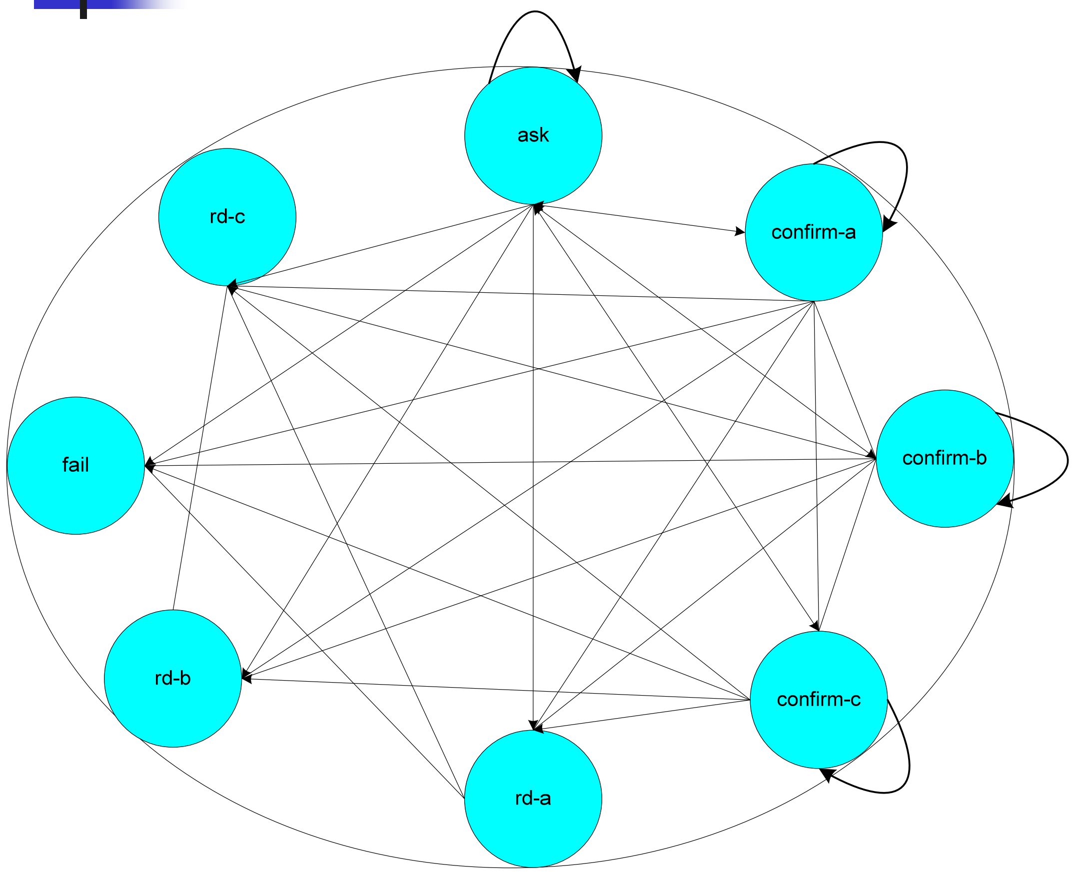


Model specification

- State space (including an absorbing *end* state)
  - G_u = {a,b,c}
  - E_u = {stress, no-stress}
  - A_u = {a,b,c,yes,no}
  - D_u = {1=location-specified, 2=location-not-specified}
- System action
  - *A = {ask, confirm-a, confirm-b, confirm-c, rd*-a, rd-b,rd-c,fail}*
- Observation
  - OE_u = {stress, nostress}
  - OA_u = {a,b,c,yes,no}
- Reward
  - *Confirms before the location is specified → reward = -2*
  - *Fail action → reward = -5*
  - *rd-x with gu=x → 10 otherwise -10*
  - *The reward for any action taken in end state is 0*
  - *The reward for other action is -1*

(rd means give route description)

Possible dialogue strategies



$\text{ask} \xrightarrow{a} \text{rd-a}$

$\text{ask} \xrightarrow{a} \text{confirm-a} \xrightarrow{yes} \text{rd-a}$

$\text{ask} \xrightarrow{a} \text{ask} \xrightarrow{a} \text{confirm-a} \xrightarrow{yes} \text{rd-a}$

…

Some of them are useful. Which ones are optimal?

Optimal policy (Using Standard PBVI Algorithm 27.83s)

- Test case:

$$p_{gc} = p_{ec} = p_{oa} = p_{oe} = 0,$$

$$p_e = 0.1\,(if\;\; ask)\,\& = 0\,(otherwise)$$

- Reformulated model

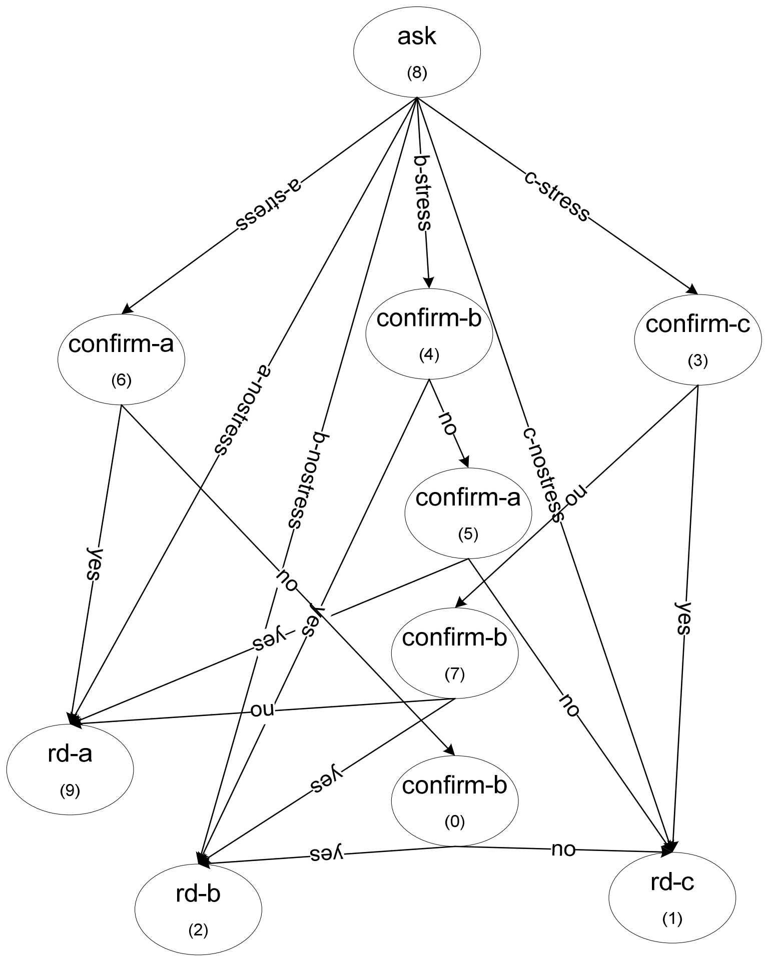
$$S = Gu \times Du + end = \{a1, a0, b1, b0, c1, c0, end\}$$

$$Z = OAu \times OEu$$

$$= \{\mathrm{a-stress}, \mathrm{a-nostress}, \mathrm{b-stress}, \ldots, \mathrm{no-nostress}\}$$

$$\gamma = 0.95$$

$$b_0 = < \tfrac{1}{3}, 0, \tfrac{1}{3}, 0, \tfrac{1}{3}, 0, 0 >$$



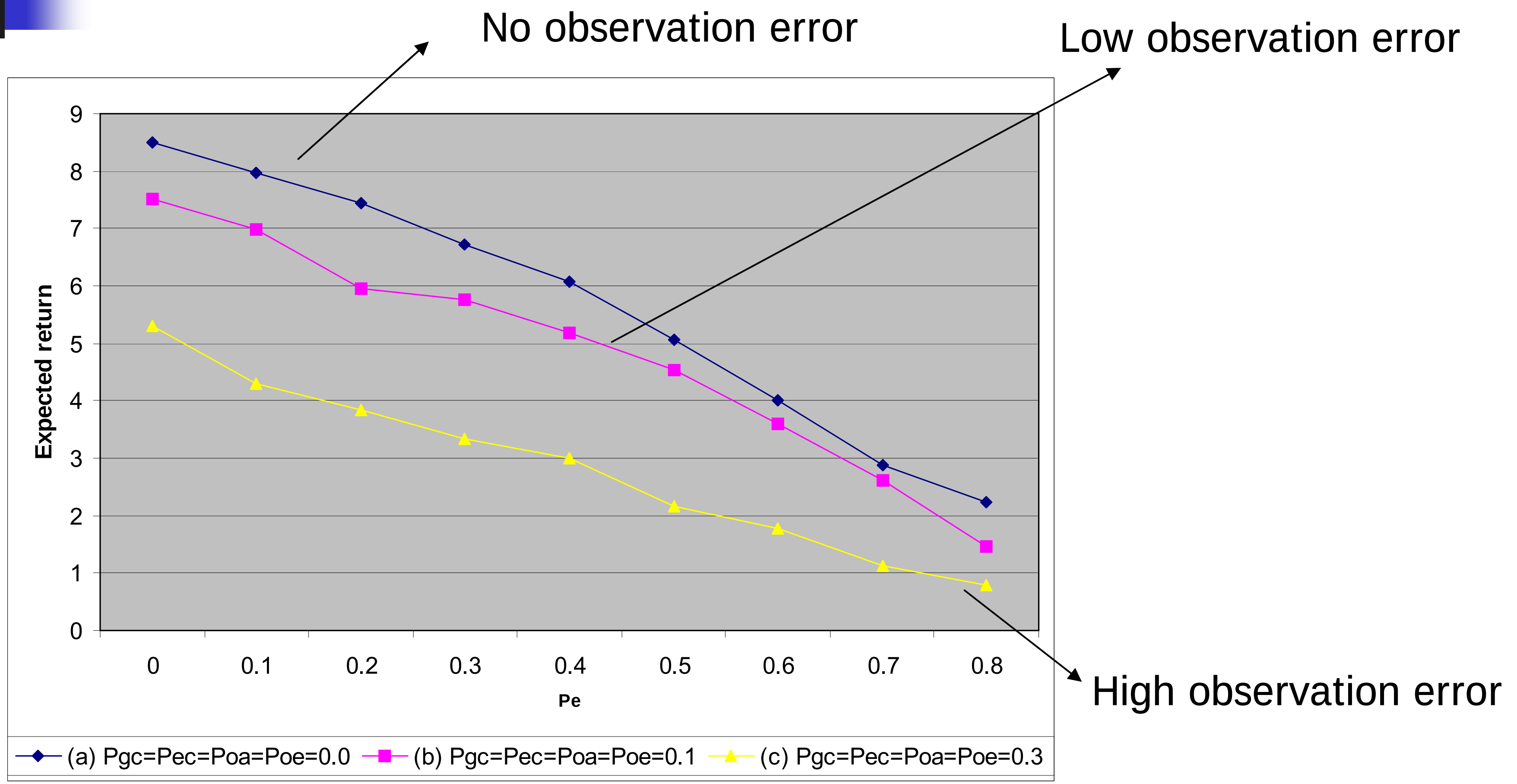
Value function table

Node#	Action	a1	a0	b1	b0	c1	c0	end
0	confirm-b	-12.5	-10.5	6.5	8.5	8.5	8.5	0
1	rd-c	-10.0	-10.0	-10.0	-10.0	10.0	10.0	0
2	rd-b	-10.0	-10.0	10.0	10.0	-10.0	-10.0	0
3	confirm-c	5.0	7.0	5.0	7.0	6.5	8.5	0
4	confirm-b	5.0	7.0	6.5	8.5	7.0	7.0	0
5	confirm-a	6.5	8.5	-12.5	-10.5	6.5	8.5	0
6	confirm-a	6.5	8.5	5.0	7.0	5.0	7.0	0
7	confirm-b	6.5	8.5	6.5	8.5	-10.5	-10.5	0
8	ask	7.7	7.7	7.7	7.7	7.7	7.7	0
9	rd-a	10.0	10.0	-10.0	-10.0	-10.0	-10.0	0

the optimal the action should start given the initial belief:

$$b_0 = < \tfrac{1}{3}, 0, \tfrac{1}{3}, 0, \tfrac{1}{3}, 0, 0 >$$

Expected return vs. user's action error being induced by stress (p_e)



Test results were carried out using Perseus algorithm on full POMDP model (61 states, 8 actions, 10 observations)

Conclusions

- The optimal dialogue strategy depends on the correlation between the user's emotion state & action
- 2TBN of factored POMDP allows integrating the features of states, actions, & observations in a *flexible* way
- But!!!
  - Computational complexity in finding the optimal policy using both exact and some approximate algorithms except small, toy dialogue problems
- Recent advances in approximate POMDP techniques plus heuristics in dialogue model design are expected to solve real-world dialogue applications

Future work

- Scaling up the model with larger state, action, & observation sets for real-world dialogue management problems
- Extending the model representation, e.g. correlations between user’s emotion & goal
- Collecting & generating both real & artificial data to build and train the model

Questions?