Learning Interrogation Strategies while Considering Deceptions in Detective Interactive Stories

Guan-Yi Chen, Edward Chao-Chun Kao, and Von-Wun Soo

Institute of Information Systems and Applications, National Tsing Hua University
guanyinu@gmail.com, edkao@cs.nthu.edu.tw, soo@cs.nthu.edu.tw

Abstract

The strategies for interactive characters to select appropriate dialogues remain as an open issue in related research areas. In this paper we propose an approach based on reinforcement learning to learn the strategy of interrogation dialogue from one virtual agent toward another. The emotion variation of the suspect agent is modeled with a hazard function, and the detective agent must learn its interrogation strategies based on the emotion state of the suspect agent. The reinforcement learning reward schemes are evaluated to choose the proper reward in the dialogue. Our contribution is twofold. Firstly, we proposed a new framework of reinforcement learning to model dialogue strategies. Secondly, background knowledge and emotion states of agents are brought into the dialogue strategies. The resulted dialogue strategy in our experiment is sensitive in detecting lies from the suspect, and with it the interrogator may receive more correct answer.

Introduction

As dialogue systems gain importance in the domain of virtual agents as interactive characters, dialogue strategies are receiving more and more attention. In interactive scenarios played by interactive characters, to know when and how to select proper dialogues in a specific social context is not a trivial task. Lies, threats, betrayals, interrogations, and so forth, are the kinds of dialogues will intensify the stories. Here we focus on one type of essential dialogues but with significant importance: interrogation.

While the main purpose of interrogation is to the same as the fundamental ask-inform protocol in agent communication, there is always a probability that the agent being interrogated will attempt to deceive the interrogator. If the interrogator keeps asking questions that touch certain sensitive words or issues, it might cause a suspect to become even more deceive because of being frightened. To verify a reply from a suspect, a detective may ask while knowing the answer. Besides asking while knowing the answer, a detective can also change the subjects temporarily and attempt to pacify the suspect. All of these possible directions of dialogue compared to ask-inform obviously require more intermediate states than just simple knowledge states. The extra speech acts give flavor to interactive stories, but also diverse dialogues and may even be misused if ill-planned. To strike a balance between dramatic effects and communication effectiveness, the detective needs to have proper interrogation strategies, based on the emergent information during dialogues.

Because the dialogue context is rather complex, it is hard to implement the proper dialogue knowledge a priori. To allow an agent to learn strategies from background contexts, we propose a reinforcement learning scheme to learn interrogation strategies. Our goal of study is to learn the proper interrogation dialogue strategies with background knowledge by using reinforcement learning. However, to conduct the reinforcement learning, the reward scheme plays an important role. Providing proper rewards (or punishment) in right contexts will actually determine a policy function of an interrogation dialogue to be properly acquired for a given social context. For this reason, we also have to simulate varying contexts, especially the mental states of the suspect, to ensure the reward scheme is properly designed by various experiments in terms of actual performances.

The rest of this paper is organized as follows. Firstly, we describe related work to our learning method. Secondly, we show our method of applying Q-learning to emotion model in details. Thirdly, we present our experiment settings, results, and discussions. Finally, we conclude our research.

Related Work

Machine Learning in Form-filling Dialogues

Many applications of reinforcement learning for finding dialogue strategies have been done in the framework of slot-filling tasks such as negotiation dialogue (Selfridge and Heeman 2011), restaurant recommendations (Jurčíček, Thomson, and Young 2012) and virtual museum guides (Misul et al. 2012). In these dialogue systems, the agent asks the user for the value of a slot, and the state representation of slot-filler is straight-forward. For each slot in the form, variables are used to track back the system knowledge from the slot. For a given value of the state variables, it does not matter how the system got to that point, as the cost to the end of the dialogue only depends on the system knowledge of the slots. With such a state representation, good policies can be learned. Thus, the variables needed for the Reinforcement Learning state are mainly action-decision variables and few if any bookkeeping variables are needed.

Machine Learning in Unstructured Dialogues

Once we move away from form-filling dialogues, characterizing the state becomes more difficult. Walker used reinforcement learning to learn the dialogue policy of a system that helps a user to read email (Walker 2000). Reinforcement learning is used to learn the choices for a dialogue strategy at an utterance that is effective. The choice is then memorized as part of the reinforcement learning state that further constrains the system's subsequent behavior.

A dialogue strategy can be also acquired using Markov decision processes (MDPs) and the reinforcement learning algorithms (Biermann and Long 1996; Levin, Pieraccini, and Eckert 1997; Walker, Fromer, and Narayanan 1998; Singh et. al 1999; Litman et. al 2000; Cuayáhuitl 2011). In order to reduce the rich search space of reinforcement learning with updating rules, Heeman proposed combing reinforcement learning and information state to generate complex dialogues between the system and the simulated user (Heeman 2007). In the flight information dialogue, it needs to quickly get the information of the origin of a flight, the airline of the flight, the departure time, etc. They do not take possibility of lying into consideration. In this type of dialogue, the less dialogue steps taken it is better. Hence they need to consider the cost of dialogue steps. Recently, there have been studies on learning negotiation policies (Heeman 2009; Georgila and Traum 2011a; Georgila and Traum 2011b), in which the system and the user need to exchange information in order to decide on a compromised good solution. The information though, unlike in a slot-filling dialogue, is not part of the final solution. However, the superintendent agent in our detective story scenario not only needs to efficiently get the information about the crime, but also needs to make sure answer is true since there is a possibility that the suspect may lie.

Method

System Architecture

The system architecture that consists of two agents: the agent1 represents the superintendent (the detective), and agent2 the crime the suspect. Interactions between the two agents are shown in Figure 1. The superintendent agent contains two parts: *Cognition module* and *Learning module*, whereas the suspect agent contains *Emotion simulation* instead of *Learning module*. Both of their *Cognition modules* consist of an *Understanding module*, *social contexts* and *speech acts*.

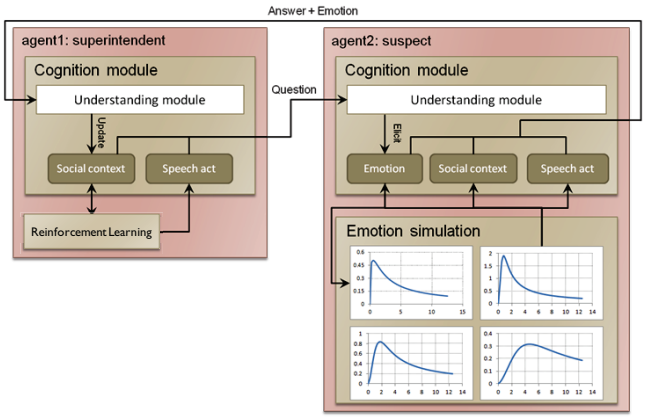

Figure 1: System Architecture

The superintendent uses its *Understanding module* to interpret an answer and emotion which feedback from the suspect. When the superintendent receives an answer, it becomes a known predicate in the *social context* knowledge of the superintendent. The superintendent will obtain a reward depending on the correctness and relevance of the criminal case in terms of a reward table, or reward function. After getting reward, the superintendent update its interrogation policy and social contexts accordingly based on the method in Reinforcement Learning.

When initialized, the suspect is given known social context facts and relation as prior social context knowledge and can use its *Understanding module* to interpret the question requested by superintendent. If the question mentions a sensitive keyword during interrogation, it could trigger the fear emotion of the suspect, and through *Emotion simulation* the suspect decides to tell a lie or the truth.

To simplify the problem without losing the generality for automated simulation, we assume both understanding modules for agents can simply interpret the predicates implemented in the social context regardless their truth values are known or unknown.

Q-learning

The Q-learning algorithm use a Q-function to calculate the quality score of a state-action combination:

$$Q{:}S\times A\rightarrow R \quad (1)$$

, where S is current state, A is next action.

Before Q-learning, Q-function is assigned a very low fixed value initially. Then with each step of actions, superintendent is received a reward, a new score value is calculated for each combination of a current state from S, and a next action from A. The core of the algorithm is a simple value-update iteration. It updates a new value of Q based on its old value and the new reward information after the state transition of executing an action.

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) +$$

$$\alpha_t(S_t, A_t) \times \left[R_{t+1} + \gamma \max_{a_{t+1}} Q(S_{t+1}, A_{t+1}) - Q(S_t, A_t) \right] \quad (2)$$

, where R_{t+1} is the reward observed after performing A_t in S_t, $\alpha(S, A)$ ($0 < \alpha \leq 1$) is the learning rate. The discount factor γ is such that $0 \leq \gamma < 1$.

In the transitional Q-learning, the environment moves to a new state S_{t+1} and the reward R_{t+1} associated with the transition (S_t, A_t, S_{t+1}) is determined.

In reinforcement learning, agent selects the action to optimize Q based on equation (3).

$$\arg\max_{a_t \in A} Q^*(s_t, a_t) \quad (3)$$

, where the Q-function specifies the cumulative rewards for each state-action pair. In contrast to the previous reinforcement learning methods, the learning goal is to make all unknown of the superintendent become known that is different from the previous work whose goal state is in its own state space.

Simulating the Suspect's Emotional Response

The major reason for why suspects will lie is fear, and fear can be treated as an emotion arousal concept in general. Here we follow psychological studies that the emotion arousal could be represented by a Hazard Distribution (Avinadav and Raz. 2008) (Verduyn et. al 2009)(Verduyn, Mechelen, and Tuerlinckx, 2011), and the emotion changes of the suspect is modeled with it in equation (4):

$$h(t) = k\rho \frac{(\rho t)^{k-1}}{1 + (\rho t)^k} \quad (4)$$

The definition of hazard function $h(t)$ is the event of the lying event where t is the time to accumulate fear emotion, ρ is failure rate (hazard rate) and k is dimensionless parameter. As time grows, the larger is ρ the sooner (smaller t) will the suspect lie, while the smaller ρ, *the later will* the suspect lie. Using the hazard function with different parameter values, we could model the personality of a suspect in terms of his or her tendency to lie in face of fear. We use two different values of ρ in hazard function Equation (4) to simulate the different personalities of a suspect to lie in response to fear. We use a fixed value k as 2.464 and two values of ρ as 2.232 and 2.993 to represent two kinds of personalities. As in Figure 2, x-axis indicates fear emotion value while y-axis indicates the hazard function value that determines the tendency of lying behavior for an interactive character. The two curves in Figure 2 represent two different lying personality curves in terms of fear emotion accumulation. To simplify the emotion simulation, we take the emotion value of maximum probability to lie as the lying threshold of fear emotion. When the suspect's fear exceeds the threshold, she will always lie.

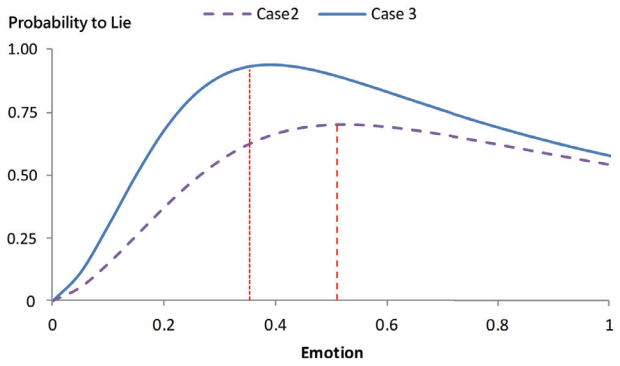

Figure 2: The simulation of two kinds of personalities

Emotion detection for the superintendent

The superintendent needs to conduct a learning algorithm according to the suspect's emotion variation and social context and learn a proper dialogue strategy. We divide the suspect's emotion value of fear into five discrete and rough levels shown in Table 2 can be sensed by superintendent by observation. Nevertheless, the superintendent is unable to determine whether the suspect will lie by observation.

Table 1: Emotion levels for learning

Emotion Level	Emotion Value
level 1: normal	$0 \leq E \leq 0.2$
level 2: worried	$0.2 < E \leq 0.4$
level 3: afraid	$0.4 < E \leq 0.6$
level 4: scared	$0.6 < E \leq 0.8$
level 5: terrified	$0.8 < E \leq 1$

There are two reasons for why we divide emotion value into five levels: to simplify the differences of Q-table with respect to different values of emotion, and to limit the observation ability of the superintendent. Each emotion level has its own Q-table, and it has transitions between the state-action of each Q-table as shown in Figure 3.

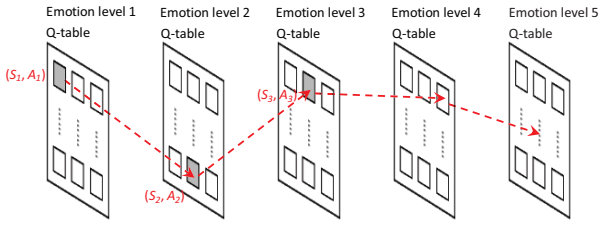

Figure 3: The relation between emotion level and emotion state transition

Experimentation

Experiment Settings

Background Knowledge for Dialogue Agents

We adopt a scenario in the detective novel *Cards on the Table* written by Agatha Christie that was first published in 1936. The main characters are Jim Wheeler as the superintendent, Miss Meredith as one of suspects, and Mr. Shaitana as the victim. In the story, there is a scene of interrogation dialogues between the superintendent Jim Wheeler and suspect Miss Meredith. In the original scenario, Miss Meredith tell lies because sensing intensified fear during the interrogation. The goal of our experiment is to make Wheeler learn an interrogation strategy, which can keep Miss Meredith from lying, with minimal amount of steps during interrogation.

We assume 7 kinds of background social contexts as facts or relations to be known by a detective. *Person (p)* indicates *p* is a person. *Object (o)* indicates *o* is an object in the crime scene. *Location (l)* indicates l is location around the place where the crime happened. *Action (act)* indicates *act* is an action related to the crime scene. *Where-Object (o, l)* indicates *Object (o)* is at *Location (l)*. *Where-Person* (w, p) indicates *Person (p)* is at *Location (l)*. *Behavior (w, act, l)* indicates *Person (p)* performs *Action (act)* at *Location (l)*. *Relation* (w_1, w_2) indicates w_1 and w_2 are friends.

In the social context for the case study, we totally have 84 possible states for predicates above. The background knowledge for the superintendent can be initially assigned as context nodes in either known or unknown categories. The ultimate goal for the superintendent agent is to interrogate the suspect and ensure all unknown facts and relations to become known ones with correct values.

Speech Acts for Interrogation

As shown in table 1, three kinds of speech acts are defined in interrogation strategies of the superintendent to search for facts. For each unknown predicate state, the superintendent can use ASK to interrogate the suspect. The superintendent uses CONFIRM only when he asks for a known predicate state. When superintendent detects the suspect's fear emotion, superintendent can use PACIFY to calm down his emotion.

Table 2: Three Speech acts of the Superintendent

Speech Act	Description
ASK	The general form of ask-if and ask-ref. To ask the suspect a question.
CONFIRM	To verify whether the suspect tells a lie or the truth.
PACIFY	To calm down a suspect's emotion.

5000 dialogue sentences with different order and lengths of speech acts are used to learn an interrogation strategy. At every dialogue, we totally have 84 states and 3 kinds of speech acts of ASK, CONFIRM, and PACIFY to be selected in the reinforcement learning. The total size of two-dimensional Q-table is (3*84)*(3*84).

Sensitive words related to the murder case are defined in advance, and the fear emotion of the suspect will be aroused if any of these keywords in table 3 are mentioned during interrogation.

Table 3: Sensitive keywords in the scenario

Sensitive keywords		
The fireplace	The empire chair	The table near fireplace
The stiletto	The door	The table near door

Each interrogation dialogue is generated according the exploration procedure in Table 4.

Table 4: The exploration procedure of reinforcement learning

```
1   CountEmpty(unknown) = N;
2   while (CountEmpty(unknown) != 0)
3   {
4     Obtain answer and emotion from the suspect;
5     Update unknown table;
6     Set emotion level and update its own Q-table;
7       Random choose a next state and action;
8   }
```

Reward Scheme

A reward scheme of each type of speech acts in the reinforcement learning is described in Table 5. The basic reward for both ASK and CONFIRM is -500 as a communication cost. However, the reward for successfully detecting a lie using CONFIRM is 1000, and the superintendent will record the suspect is telling a lie. The PACIFY speech act is to calm down the fear of the suspect. The sole reward of PACIFY is -250, but this act will also decrease the fear emotion of the suspect. If the fear emotion changes, the reward is given as the emotional level change multiplied by 2000. According to the Hazard function, the range of emotion change is between 0.03 and 0.06, so the change in terms of reward is about 60~120.

Table 5: Reward scheme for reinforcement learning

Category	Description		Reward
Speech act	ASK	Interrogate the suspect	-500
	CONFIRM	Detect lying	1000
		Does not detect lying	-500
	PACIFY	Calm down the suspect emotion	-250
Social context	Behavior	Know Where did suspect go to	200
		Know What did victim do	
		Know What did suspect do	
	Relation	Know The relation between object and location	
Emotion	Fear Up	Emotion fear raises up.	-2000 * emotion deviation
	Fear Down	Emotion fear goes down.	

Results and Discussions

We compare the rate of correct answer and the cost of dialogue steps based on the performance in all 5 cases as shown in Table 6.

Table 6: The comparison of 5 different cases

	Emotion Change	Reward Scheme	Lying threshold	Correct information rate	Number of dialogue steps
Case 1	Without detecting emotion		0.52	0.41	61
Case 2	$\rho = 2.232$	Shown in Table 5	0.52	1	103
Case 3	$\rho = 2.993$	Shown in Table 5	0.39	0.79	132
Case 4	$\rho = 2.232$	10x emotion change value as that in case 2	0.52	1	132
Case 5	$\rho = 2.232$	The same as Case 2, except reward of CONFIRM = 200	0.52	0.79	81

Baseline: Case 1

Case 1 stands as a baseline in our experiment, as the superintendent in case 1 learns an interrogation strategy without the ability to detect fear. The resulted dialogues are shown in section A, B and C of Table 7. Note that the Lie and Truth signals are not directly sent to the superintendent.

In section A in Table 7, the superintendent (Supt Jim Wheeler) asks whether the victim (Mr. Shaitana) went to the bridge table or not. The suspect (Miss Meredith) replies yes. The next question the superintendent asks what the victim did at the bridge table. This time since the question is not related to the murder, the fear emotion of the suspect is stable. The suspect replies with the truth.

In section B, when questions containing sensitive keywords **fireplace** and **empire chair** are asked, the suspect's fear emotion raises up. By exhaustively asking unknown predicate states, the suspect replies that the victim was at the empire chair.

Table 7: The description of section A, B and C

Section A	
Wheeler:	[ASK] if: Where-Person (mr-Shaitana, bridge-table)
Meredith:	yes [True, Emotion: 0]
Wheeler:	[ASK] ref: Behavior (mr-Shaitana, ?, bridge-table)
Meredith:	Behavior (mr-Shaitana, see-bridge, bridge-table) [True, Emotion: 0]
	.....
Section B	
Wheeler:	[ASK] if: Where-Person (mr-Shaitana, **fireplace**)
Meredith:	yes [True, Emotion: 0.09]
Wheeler:	[ASK] if: Where-Person (mr-Shaitana, **empire-chair**)
Meredith:	yes [True, Emotion: 0.27]
Wheeler:	[ASK] if: Where-Person (ms-Meridith, **empire-chair**)
Meredith:	yes [True, Emotion: 0.36]
Wheeler:	[ASK] ref: Behavior (ms-Meridith, ?, **empire-chair**)
Meredith:	found-mr-Shaitana-dead [True, Emotion: 0.45]
	.....
Section C	
Wheeler:	[ASK] if: Where-Person (ms-Meridith, **table-near-fireplace**)
Meredith:	no [Lie, Emotion: 0.72]

However, the suspect begins to lie in section C, which is shown in Figure 4, resulting in low overall correct information rate.

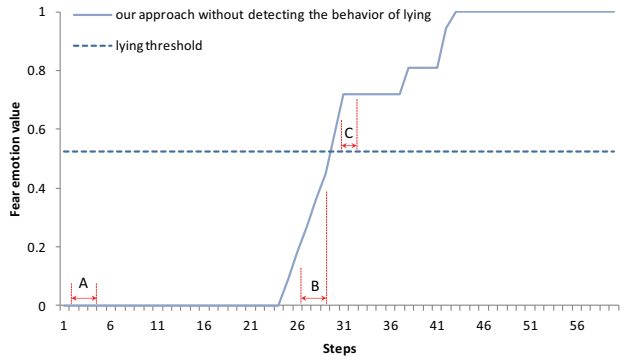

Figure 4: The emotion variation in case 1

Interrogate Different Personalities: Case 2 and Case 3

In case 2, we set ρ = 2.232 to model a rather bold Miss Meredith. In this case, her fear emotion accumulated slowly. Before reaching the lying threshold, her fear emotion is lowered by PACIFY effectively. We can see clearly that the suspect did not tell a lie from the beginning till the end, which is shown in Figure 5.

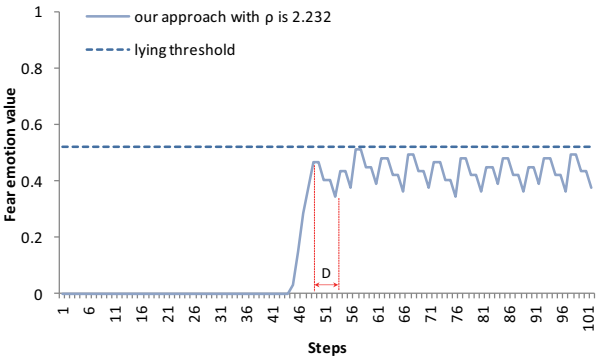

Figure 5: The emotion variation in case 2

The dialogue result in section D is shown in Table 8. The general order of speech acts is CONFIRM → PACIFY → ASK. Since the fear emotion value lays in level 3, every time when fear emotion value of the suspect reach the lying threshold the superintendent use CONFIRM to check whether the suspect lies or not. At level 3, when detecting the fear emotion change, the superintendent uses PACIFY to lower down the fear emotion, and then uses ASK to continue on asking questions.

Table 8: The description of section D

Section D	
Wheeler:	[ASK] ref: Behavior (mr-Shaitana, ?, **table-near-fireplace**)
Meredith:	drink [True, Emotion: 0.51]
Wheeler:	[CONFIRM] if: Behavior (ms-Meredith, leave, bridge-seat)
Meredith:	yes [True, Emotion: 0.51]
Wheeler:	[PACIFY] (Do not be afraid, take a deep breath. Say clearly.)
Meredith:	(Silence) [True, Emotion: 0.45]
Wheeler:	[ASK] if: Where-Person (ms-Meredith, **fireplace**)
Meredith:	yes [True, Emotion: 0.48]
Wheeler:	[PACIFY] (Do not be afraid, take a deep breath. Say clearly.)
Meredith:	(Silence) [True, Emotion: 0.42]

In case 3, we set ρ is 2.993 to model a cowardly Miss Meredith. In contrast to case 2, the fear emotion of the suspect accumulated relatively faster to reach the threshold to lie. Similar to case 2, the superintendent learned a policy with detection of lying behavior. The result is shown in Figure 6. It turns out that the superintendent detects the lying behavior in emotion level 3 (emotion value: 0.4 ~ 0.6) and begins to use CONFIRM and PACIFY.

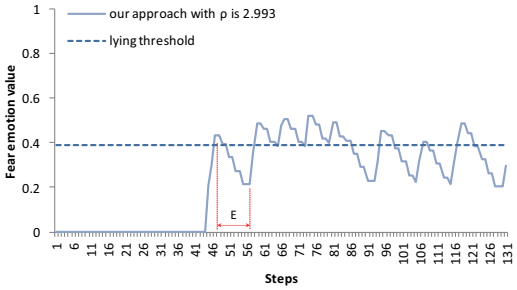

Figure 6: The emotion variation in case 3

The resulted dialogues of section E is described in Table 9.

Table 9: The description of section E

Section E	
Wheeler:	[ASK] if: Where-Person (ms-Meredith, **table-near-fireplace**)
Meredith:	no (Lie, Emotion: 0.52)
Wheeler:	[CONFIRM] if: Behavior (ms-Meredith, leave, bridge-seat)
Meredith:	no (Lie, Emotion: 0.52)
Wheeler:	[PACIFY] (Do not be afraid, take a deep breath. Say clearly)
Meredith:	(Silence) (Emotion: 0.49)
	.....
Wheeler:	[ASK] ref: Behavior (ms-Meredith, ?, **table-near-fireplace**)
Meredith:	drink (True, Emotion: 0.22)

At the end of session E, the fear emotion of the suspect is reduced to a level that causes the superintendent to believe that the suspect may tell the truth, therefore the strategy of speech act of the superintendent changes to ASK again.

The results of case 2 and case 3 show that the superintendent may comfort the suspect in time to decrease the fear emotion of the suspect before lying. Even when interrogating a cowardly suspect, the superintendent will still perform PACIFY as soon as after detecting a lie.

Different Reward Values: Case 4 and Case 5

In case 4, we change the reward on emotion change with tenfold value where the reward value for emotion change is among 600~1200, resulting in a "merciful" superintendent who uses PACIFY after asking every important question. The result is shown in Fig 7.

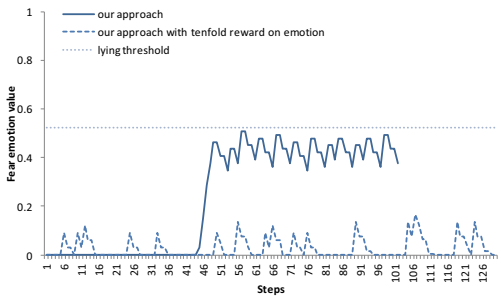

Figure 7: The emotion variation with tenfold reward on emotion changes in case 4

The reason we only change the reward on emotion change is to observe the difference of using PACIFY speech act. And the result of this case it shows that the reward 2000*(emotion difference) is rather proper.

In case 5, the importance of CONFIRM is lowered by setting its reward value to 200. The superintendent learned a different interrogation strategy compared to that in case 2.

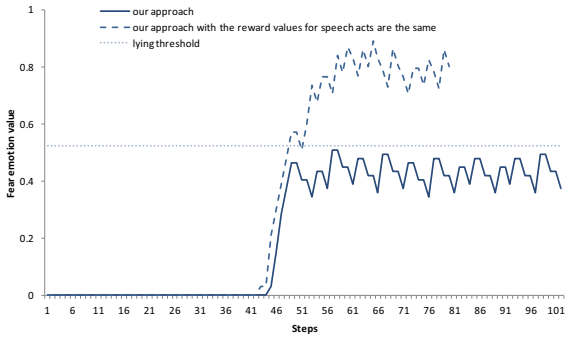

Figure 8: The emotion variation of case 5, in contrast to case 2

In Figure 8, the performance of detecting lying by using the reward values for emotion change to be the same as 200 is not good shown as the dashed curve in contrast to the solid curve of case 2. The superintendent may detect lying at the wrong emotion level and cannot calm down the suspect's fear emotion beyond the lying threshold.

Conclusion

In interactive stories, the detective who is conducting interrogation dialogue must ensure to obtain the true information in face of possibility of lying under fear by the suspect. Previous work of learning from dialogue aim to use reinforcement learning to get answers from the user quickly in terms of cost of dialogue steps. They do not take lying into consideration. We propose a framework of

reinforcement learning by a reward scheme based on speech acts and social contexts to learn the best strategy in interrogation dialogue, where formal rules could not be given in advance. While both the suspect and the superintendent are modeled with agents, the suspect may lie if its fear is arisen during interrogation, and its fear emotion is simulated with a hazard function. The result of learned interrogation strategies is very sensitive in detecting lies of the suspect, allowing the superintendent to elicit more correct answers during interrogation with less dialogue steps.

This work can be extended to acquire different interrogation strategies for a superintendent to deal with suspects with different personalities. It can also be augmented in future work by versatile types of speech acts and reward schemes to investigate more elaborate dialogue strategies to be adopted by dialogue agents in an interactive scenario.

References

Avinadav T., and Raz, T. 2008. A New Inverted U-Shape Hazard Function. *IEEE Transactions on Reliability* 57(1): 32-40.

Biermann, A.W., and Long, P. M. 1996. The composition of messages in speech-graphics interactive systems. In *Proceedings of the 1996 International Symposium on Spoken Dialogue*, 97-100.

Cuayáhuitl, H. 2011. Learning Dialogue Agents with Bayesian Relational State Representations. In *Proceedings of the IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems (IJCAI-KRPDS)*, 9-15, Barcelona, Spain: International Joint Conferences on Artificial Intelligence, Inc.

Georgila, K., and Traum, D. 2011a. Learning Culture-Specific Dialogue Models from Non Culture-Specific Data. In *Proceedings of the 6th International Conference on Universal Access in Human-Computer Interaction: Users Diversity - Volume Part II (UAHCI'11)*, Stephanidis, C. (Ed.), Vol. Part II, 440-449. Berlin, Heidelberg: Springer-Verlag.

Georgila, K., and Traum, D. 2011b. Reinforcement Learning of Argumentation Dialogue Policies in Negotiation. In *Proceedings of INTERSPEECH 2011, 12th Annual Conference of the International Speech Communication Association*, 2073-2076. Florence, Italy: ISCA.

Heeman, P. A. 2007. Combining Reinforcement Learning with Information-State Update Rules. In *Proceedings of the North American Chapter of the Association for Computational Linguistics Annual Meeting*, 268-275. Stroudsburg, PA: Association for Computational Linguistics.

Heeman, P. A. 2009. Representing the Reinforcement Learning State in a Negotiation Dialogue. In *Proceedings of the IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, 450-455. Merano, Italy: IEEE.

Jurčíček, F., Thomson, B., and Young, S. 2012. Reinforcement Learning for Parameter Estimation in Statistical Spoken Dialogue Systems. *Computer Speech and Language* 26(3):168-192.

Levin, E., Pieraccini, R., and Eckert, W. 2000. A Stochastic Model of Human-Machine Interaction for Learning Dialog Strategies. *IEEE Transactions on Speech and Audio Processing* 8(1):11–23.

Litman, D. J., Kearns, M. S., Singh, S. P., and Walker, M. A. 2000. Automatic Optimization of Dialogue Management. In Proceedings of the 18th conference on Computational linguistics - Volume 1 (COLING '00), Vol. 1, 502-508. Stroudsburg, PA: Association for Computational Linguistics.

Misu1, T., Georgila, K., Leuski, A., and Traum, D. 2012. Reinforcement Learning of Question-Answering Dialogue Policies for Virtual Museum Guides. In *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 84-93, Stroudsburg, PA: Association for Computational Linguistics.

Selfridge, E. O., and Heeman, P. A. 2011. Learning Turn, Attention, and Utterance Decisions in a Negotiative Slot-Filling Domain, Technical Report, CSLU-11-005, Center for Spoken Language Understanding, Oregon Health & Science University, Portland, OR.

Singh, S. P., Kearns, M. J., Litman, D. J., and Walker, M. A. 1999. Reinforcement Learning for Spoken Dialogue Systems. In *Proceedings of NIPS 1999*, 956-962. Cambridge, MA: The MIT Press.

Verduyn, P., Delvaux, E., Coillie, H. V., Tuerlinckx, F., and Mechelen, I. V. 2009. Predicting the Duration of Emotional Experience: Two Experience Sampling Studies. *Emotion* 9:83-91.

Verduyn, P., Mechelen, I. V., and Tuerlinckx, F. 2011. The Relation between Event Processing and the Duration of Emotional Experience. *Emotion* 11:20-28.

Walker, M. A., Fromer, J. C., and Narayanan, S. 1998. Learning Optimal Dialogue Strategies: A Case Study of a Spoken Dialogue Agent for Email. In *Proceedings of the 36th Annual Meeting of the Association of Computational Linguistics*, 1345-1352. Stroudsburg, PA: Association for Computational Linguistics.

Walker, M. A. 2000. An Application of Reinforcement Learning to Dialog Strategy Selection in a Spoken Dialogue System for Email. *Journal of Artificial Intelligence Research* 12:387-416.