# Reinforcement Learning With Safety Education

Liu Leilei
College of Intelligence and Computing
Tianjin University
TJ, 300350 PRC
liuleilei@tju.edu.cn

## Abstract

*Reinforcement learning trains agent in the specific environment through the feedback, mainly rewards from the environment. If agent takes a correct action in the current state, it will receive a positive reward, so agent will associate the state with the action and that helps it make a right choice when facing the state again next time. This process is similar with education, we can naturally think of applying RL in the safety education. As we all know, children lack of social experience and self-protection awareness, they are easily deceived by strangers. If we consider a child as an learning agent, we can train him/her in our designed environment. In this paper, we build a platform called "Protect Yourself" to improve children's self-protection awareness in RL ways.*

## 1. Introduction

In recent days, people's representative proposed a motion that trafficking in women and children can be sentenced to death, which caused intense discussion in China. According to incomplete statistics, missing children is about 200,000 every year in China, but only 0.1% of them could be retrieved. Protecting children from harm has been a serious problem, government needs to make more severe punishment and parents needs to spend more time to guide their children. More importantly, children should learn how to protect themselves.

Based on that, we build a platform to help children learn some useful skills to improve their self-protection ability. In the platform, we create several different scenes, all of these scenes are maximumly consistent with the real situation. For example, in the "Candy Allure" environment, villains use delicious candies as a bait to attract children; In the "Friend Trap" environment, villains usually pretend to be a friend of parents to gain the trust of children. These two environments are showed in Figure 1 and Figure 2.


**Figure 1. "Candy Allure" environment**


**Figure 2. "Frend Trap" environment**

## 2. Task and model

Our environment is a finite horizon and deterministic Markov decision process, in every environment, state space and action space are very different. If agent choose the right action in the current state, it will receive a positive reward. In contrast, agent will be punished for the wrong choice. Taking the "Candy Allure"as an example.

**State:** State is the agent's observation and is mainly composed of four parts : time, location, stranger, words. So, we can represent our state by a quad, S = (T, L, Str, W). In our "Candy Allure"environment, S = (morning, park, strange middle-aged woman, "I have candies"). We could see that the environment is complex and deceptive, the setting of time and location is easy to let agent relax; and the presence of "candy noise"maybe leads to our agent choose a wrong action.

**Action:** In every state we have some different action choices for our agent, some are right and the others are wrong. In the previous state, we have 3 action choices, like the Figure 3 shows:


**Figure 3. Available Actions**

**Transition:** According to agent's action choice, environment will update the state. In the above choices, if we take the first choice, the next state will be Figure 4:


**Figure 4. New State**

**Reward:** Our agents are mainly children aged from 5 to 15, a numerical score is abstract, so we need to design another type of rewards. In China, Monkey-King Sun Wukong is a superstar among children, and they want to get recognition of idols. So reward function is designed to be associated with Sun WuKong. For example, the reward of previous action is showed like Figure 5:


**Figure 5. Reward Function**

## 3. Experiments

There are no children aged from 5 to 15 around me, I have to look for some children with my laptop in the street. And I find two amazing facts, one is that children have a strong self-protection awareness, when I asked if they wanted to play an interesting game, they all refused. So the performance of our platform is unknown.

The second amazing fact is that children aged form 5 to 15 have less interest in this platform, but college students show great enthusiasm, my roommate indulges in it.

## 4. Conclusion and Future Work

Our platform is designed in the game form, which is entertaining and learning, it can arouse users'interest. Not only for children, college students also show great enthusiasm. So I will perfect the platform in the future, and design some more complex scenes, such as "Online loan trap".