توضیح کد پروژه RL
نرگس سیفی ۹۹۴۴۲۱۷۳

نکته اول: امکان ویدیو گرفتن از اجرا نبود به دلیل اینکه گرافیکش crash میکرد و فقط صدای بازی رو میتونستم بشنوم. اما اسکرین شات زیر از نتیجه اجرا هست که امتیازی که agent گرفته رو نشون میده.

```
-22.857142857142858
{'score': 33}
-14.047619047619047
{'score': 33}
-14.047619047619047
{'score': 33}
-1000
PS G:\daneshga\darsi\AI\RL>
```

توضیح روش پیاده سازی تابع ها:

- convert_continuous_to_discrete:
فاصله پیوسته افقی و عمودی رو بین ۲۰ تا bin تقسیم میکنیم برای گسسته شدن.

- compute_reward:

اگر done باشه یعنی به یه جایی برخورد داشته و باید یک ریوارد خیلی منفی بگیره (مثلا ۱۰۰۰-)

وگرنه میایم اول فاصله هارو گسسته می‌کنیم با تابع قبلی. هر چقدر فاصله عمودی از نقطه وسط بیشتر باشه چیز بدیه، پس براش پنالتی درنظر گرفتم:

```
-۱۰ * abs(۱۰ - y_d)
```

حالا هرچقدر از نظر افقی به پایپ بعدی نزدیک باشیم طبیعتا باید زودتر به وسط عمودی برسیم که نخوریم به پایپ، پس این رو هم ضریبِ اون پنالتی قبلی می‌کنیم:

```
(1 - ((x_d)/42))
```

در آخر هم که اگر امتیاز جدیدی گرفته باشیم به ازای اون یه مقدار خوبی باید به reward اضافه بشه:

```
reward += 1000 * (new_info['score'] - prev_info['score']) + 5
```

*(اعداد با آزمون و خطا به دست اومدن)*

- Policy:
خب چون روش qlearning هست طبیعتا اون اکشنی باید انتخاب بشه که ماکسیمم q-value رو بهمون بده. پس:

```
self.max_arg(discrete_state)
```

- get_action:
برای انتخاب کردن اکشن میایم به احتمال epsilon یه اکشن رندوم برمیداریم:

```
if utils.flip_coin(self.epsilon):
```

```
return random.choice(self.get_all_actions())
```

و در غیر این صورت (به احتمال epsilon-۱) همون اکشنی که تو `policy` فعلی هست رو برمیگردونیم.

- `maxQ:`

اینجا میایم بین همه اکشن های ممکن (تو این مسئله ۰ و ۱) میگردیم و اونی که بیشترین `qvalue` رو بهمون میده پیدا میکنیم و `q-value` مورد نظر رو برمیگردونیم. یک `for` سادمس.

- `max_arg:`

اینجا هم مشابه تابع قبلی عمل میکنیم با این تفاوت که به جای `q-value`، اون اکشنی که بیشترین `q-value` رو تولید می‌کنه برمیگردونیم.

- `Update:`

اینجا صرفا فرمول `temporal difference` رو اعمال میکنیم برای آپدیت کردن `q-value`.

- `update_epsilon_alpha:`

مقادیر `epsilon` و `alpha` رو هر بار یه مقدار کمی کاهش میدیم.