Reinforcement learning

Lukáš Hromadník

May 28, 2018

1 Řešení

1. Navrhněte tři možné netriviální možnosti, jakým způsobem definovat stav ve hře.

   (a) Součet karet, které má hráč v ruce, a hodnota karty, kterou obdržel dealer.

   (b) Kombinace hodnot karet, které má hráč v ruce, a hodnota karty, kterou obdržel dealer.

   (c) Pouze součet karet, které má hráč v ruce.

2. Pro každou reprezentaci z bodu 1 určete celkový počet stavů.

   (a) Hráč se v nejhorším případě dostane na hodnotu 30 (v ruce má karty s celkovou hodnotou 20 a přijde karta s hodnotou 10). Dealer dostane kartu, jejíž hodnota bude 2 až 11 (eso), tedy 10 možností.

   Pronásobením těchto dvou čísel dostaneme 300 stavů. Některé z těchto stavů jsou nepotřebné.

   (b) U kombinací hodnot jednotlivých karet, které hráč během hry obdrží, je situace komplikovanější. Nejméně může v ruce mít 2 karty a nejvíce 11. Dle zdrojů na Stack Overflow (https://stackoverflow.com/a/39113661) se počet kombinací, kterými lze dosáhnout 21, rovná 186 184 kombinacím. Pokud toto číslo vezmeme a vynásobíme ho počtem možností u dealera, dostaneme 1 861 840 kombinací. K tomu musíme přičíst dalších minimálně 10, kterými budeme reprezentovat hodnoty karet, které překročily hodnotu 21. Ve výsledku tedy máme 1 861 850 stavů.

   (c) Pokud budeme uvažovat pouze dosažitelné stavy pro tuto reprezentaci, tzn. hodnoty hráčovi ruky nemůže být 0, 1, 2 nebo 3, a pokud pro všechny hodnoty, které překročí hodnotu 21, zavedeme jeden stav, dostaneme se na číslo 19 stavů.

3. Vyberte jednu z reprezentací uvedených v bodě 1. Vysvětlete, proč ji považujete za nejlepší a odpovězte na následující otázky.

   Za nejvhodnější reprezentaci považuji 1a. Tato reprezentace není příliš paměťová složitá a vyžaduje tak méně iterací pro naučení.

- Zachytává tato reprezentace všechny informace, které má agent k dispozici, pro rozhodnutí?

  Nezachytává přesnou kombinaci karet, které má hráč k dispozici. Jinak využívá všech dostupných informací.

- Lze tuto reprezentaci zjednodušit?

  Reprezentaci lze zjednodušit tak, že se stavy, kde hráčovi karty mají hodnotu větší než 21, sloučí do jednoho.

- Pokud ano, ovlivní zjednodušení nějak výsledek?

  Zjednodušení neovlivňuje výsledek.

- Můžeme použít exaktní metody pro řešení této hry? Pokud ano, jak? Pokud ne, proč?

  Jelikož Value Iteration využívá při výpočtu pravděpodobnostní model prostředí, tak by mělo být možné použít exaktní metody i pro řešení této hry.

4. Porovnejte úspěšnost jednotlivých strategií.

   - Jaké jsou jejich očekávané utility?
     - Random: -0.403
     - Dealer: -0.0848
     - TD: -0.0797
     - SARSA: -0.3124

     Výsledek u Dealera a TD je očekávaný, jelikož oba agenti následovaly stejnou policy. Výsledek u SARSY je na druhé straně zvláštní. Očekával bych, že bude velice podobný jako u TD, avšak výsledek je více podobný náhodnému agentu.

   - Byla naučená utilita v rozporu s intuicí?

     V případě TD odpovídá naučená utilita intuici, kde malá hodnoty na začátku měly malou záporonou utilitu, ta se postupně klesala, až se dostala do bodu (kolem hodnoty součtu 17), kde se velice rychle přehoupla do kladných čísel a nejvyšší utilitu měl stav číslo 21.

     V případě SARSY je výsledek podobný, avšak jednotlivé stavy mají menší rozdíly mezi utilitami.

   - Jaká je utilita při obdržení křížové devítky, kárového kluka (spodka) a pikové dvojky do ruky hráče a dealer odbržel křížovou čtyřku?
     - TD: 1.86
     - SARSA: 0.98

   - Jaká je utilita při obdržení kárového esa a pikové pětky, pokud má dealer pikové eso? Je lepší v této situaci požádat o další kartu?
     - TD: -1.24
     - SARSA: -1.99

Vzhledem k tomu, že policy u TD má řádově vyšší utilitu, tak je určitě lepší v tomto případě si vzít další kartu.

- Konvergovali hodnoty utilit?
  Dle mých pozorování vypadá, že jednotlivé hodnoty konvergují.