

Vágásgenerálás az egészértékű programozásban gépi tanulással

Önálló projekt II

Becsó Gergely

Témavezető: Lukács András, Madarasi Péter

2023. június

Megerősítéssel tanulás / Formális játék

- $\{\mathbf{S}, \mathbf{D}, \mathbf{T}, \mathbf{R}\}$ környezet ($\mathbf{T} : \mathbf{S} \times \mathbf{D} \rightarrow \mathbf{S}, \mathbf{R} : \mathbf{S} \rightarrow \mathbb{R}$)
- trajektória: $\tau = (s_i, d_i, r_i)$ sorozat,
- policy függvények: $\pi(\mathbf{S}, \theta) \rightarrow D$
- jutalomfüggvény, $R(\tau) = \sum_{i \in \tau} r_i$

Policy gradient

Q -segédfüggvény:

$$Q^\pi(s, d) = \max_{\theta} \mathbb{E}_{\pi} \left[\sum_{i=0}^{\infty} r_i \gamma^i \right]$$

Közismert gradiens:

$$\nabla_{\theta} J(\theta) = \mathbb{E}_{\pi} \left[\sum_{t=0}^{\infty} r_t \nabla_{\theta} \log p(s_t, d_t) \right]$$

A Columbiái egyetem tananyagából:

$$\nabla_{\theta} J(\theta) = \mathbb{E}_{\pi} \left[\sum_{t=0}^{\infty} Q(s_t, d_t) \nabla_{\pi} \log p_{\pi}(s_t, d_t) \right]$$

$$\begin{aligned} Ax &\leq b \\ \min(cx) \end{aligned}$$

- $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$ és $x \in \mathbb{Z}^n$.
- Additív egészértékűségi rés: $cx_{LP}^* - cx_{IP}^*$

$$\begin{aligned} Ax + Is &= b \\ x, s &\geq 0 \\ \min(cx) \end{aligned}$$

Ahol $I \in \mathbb{Q}^{m \times m}$ egységmátrix és $s \in \mathbb{R}^m$

Gomory vágások

Gomory vágások: $(-A_i^* + \lfloor A_i^* \rfloor)^T x \leq -b_i^* + \lfloor b_i^* \rfloor$

$$e^T x + r^T s \leq d \quad (1)$$

$$r^T Ax + r^T s = r^T b \quad (2)$$

$$(e^T - r^T A)x \leq d - r^T b \quad (3)$$

Az IP-játék

- állapotok: IP feladatok leírásai
- akciók: adott Gomory vágás kiválasztása
- átmenet: a kiválasztott vágás hozzávétele
- jutalomfüggvény: az additív egészértékűségi rés csökkentése

- a feltételeket és a lehetséges vágásokat LSTM hálózat segítségével 64 dimenziós vektorokba ágyazzuk
- a beágyazott feltételeken attention jellegű mechanizmust alkalmazunk

$$P = \text{soft max}\left(\sum_{\text{sorok}} V \cdot F^T\right)$$

A félév eredményei

SCIP, Gurobi solver

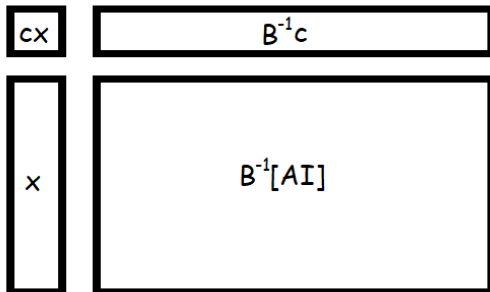
Újonnan követett adatok:

- százalékos javítás
- kiválasztott akció követése
- MV heurisztika összehasonlításként

Problémák:

- Vágás számolás
- Súly-frissítés

Vágás számolás



Súly frissítés

$$\nabla_{\theta} J(\theta) = \mathbb{E}_{\pi} [\sum_{t=0}^{\infty} Q(s_t, d_t) \nabla_{\pi} \log p_{\pi}(s_t, d_t)]$$

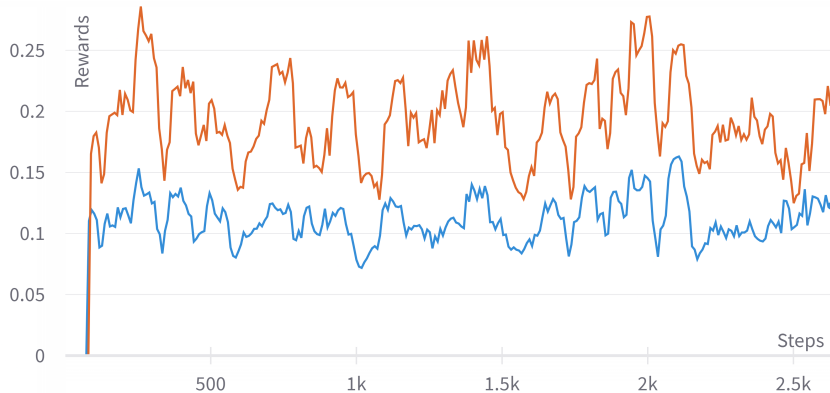
- Szükség van a p_{π} eloszlásra a kiszámolásához, amit a hálóból kaphatunk.
- Eredetileg nem megjegyeztük az eloszlást, hanem a bemeneteket jegyeztük fel és abból kiszámoltuk újra.
- Frissített súlyú háló azonban nem ugyanazt az eloszlást adja.

Tervek

Tanulási formák: rendes célfüggvény, REINFORCE, Actor-Critic módszer, TRPO, PPO

Egyéb funkciók: ritka mátrixok, egyéb háló-architektúrák,

Network reward vs Max Violation



Köszönöm a figyelmet!