Návrh bakalářské práce

Emergentní koalice v multi-agentním posilovaném učení

Moje bakalářská práce se zaměří na studium toho, zda explicitní koaliční bonus α za simultánní akce dvou či více agentů vede k lepší koordinaci a prostorové kohezi oproti standardnímu individuálnímu odměňování.

Simulační prostředí. V Python-Gym (Gymnasium) prostředí `ArenaEnv` vytvořím malý diskretní grid (10×10) se sbíratelnými zdroji (dřevo, ruda) a nepřátelskými moby a finálním bossem. Nasadím několik „přátelských" agentů, kteří mají akce `move`, `gather`, `attack`, `idle` a `craft`. Agenti mají plnou observabilitu; odlišujeme pouze jejich odměňování.

Sdílený týmový inventář.

- Všechny agenty spojuje jediný slovník `team_resources` = {wood : 0, ore : 0}.
- Jakýkoli agent, který provede `gather`, přidá buď dřevo, nebo rudu do této sdílené zásoby.
- Akce `craft` odebere z `team_resources` určité množství (dřeva i rudy) a buffuje zvoleného agenta na omezenou dobu.
- Tím se (teoreticky) podpoří přirozené rozdělení rolí (někdo sbírá, někdo bojuje).

Sběr a craftování (bez detailních čísel).

- *Rudné node & stromy:*
  - *Solo sběr* (1 agent) je pomalejší a dává menší množství suroviny.
  - *Joint sběr* (dvě a více agentů současně) proběhne rychleji a přinese bonus α oběma zúčastněným.
  - Solo i joint sběr dává odměnu za získanou surovinu; joint akce navíc vyvolají koaliční bonus α.
- *Craftování a buffy:*
  - Craftovací stanice (bench) jsou náhodně rozmístěny po mapě.
  - Pro vyvolání buffu agent spotřebuje určitý počet dřeva a rudy ze `team_resources`.
  - Buff dočasně zvyšuje HP a sílu útoku agenta na předem určenou dobu.
  - Odměna za craft je vyšší; v kooperativní politice existuje navíc proximity-bonus γ, pokud buffovaný agent má ve svém okolí další spolupracovníky.

Nepřátelští mobové & Boss.

- *Easy mobové:*
  - Lze je zabít v několika útocích; solo zabití přináší základní odměnu a šanci na drop surovin.
  - *Joint* zabití proběhne rychleji, každý z útočníků dostane odměnu $+\alpha$ navíc.

- *Boss:*
  - Spawnuje se na pevné centrální pozici na začátku epizody. Začne být agresivní za určitý počet kroků od startu, nebo pokud na něj jeden z agentů zaútočí.
  - Solo buffovaný agent ho může porazit (možná), ale je to riskantní a trvá výrazně déle.
  - Joint útok dvou či více buffovaných agentů vede k rychlejšímu zabití s menším rizikem. Každý účastník navíc získá α za každý krok, kdy na bosse útočí.

Průběh epizody & stop-condition.

1. **Teoretický průběh epizody:**
   - Na začátku se spawnou agenti, suroviny, mobové a boss.
   - Agenti podle své politiky (kooperativní / konkurenční / random) vybírají akce $\in\{$`move`, `gather`, `attack`, `idle`, `craft`$\}$.
   - Během sběru se u jednotlivých politik rozhodují, zda agenti půjdou sólo nebo společně.
   - Po craftu (dočasném buffu) se agenti shromáždí a pokoušejí se zabít bosse, buď sólo, nebo společně.
2. **Ukončení epizody:**
   - *Victory*, pokud bosse zabije alespoň jeden agent a alespoň jeden přežije.
   - *Defeat*, pokud všichni agenti padnou nebo překročíme maximum kroků.
3. **Metriky:**
   - `time_to_boss_kill` = počet kroků do zabití bosse (nebo max přetrvání).
   - `time_to_first_buff` = kdy proběhl první craft.
   - `num_joint_mining_events` = kolikrát $\geq$2 agenti sbírali ten samý node.
   - `num_joint_attack_events_boss` = kolikrát $\geq$2 agenti zároveň útočili na bosse.
   - `num_crafts` = počet craft akcí.
   - `survival_rate` = poměr přeživších agentů.
   - `average_distance_between_agents` = průměrná vzdálenost dvojic agentů.
   - ...

Politiky a jejich rozlišení

Kooperativní ($\alpha > 0,\ \beta > 0,\ \gamma > 0$) Agenti získávají bonus α za simultánní „joint-gather“ a „joint-attack“, dále β za udržování blízkosti ostatních a γ při provedení craftu ve skupině. Odměny za individuální (solo) akce jsou oproti tomu nižší. Předpokládáme vyšší četnost společných akcí, zkrácené `time_to_boss_kill` a zvýšenou `survival_rate`.

Konkurenční ($\alpha = 0,\ \beta = 0,\ \gamma = 0$) Žádné koaliční odměny; agenti jednají samostatně a soustředí se na vlastní zisk. Roztáhnou se po mapě, sbírají odděleně, craftují bez vzájemné spolupráce a postupně útočí na bosse. Očekáváme menší počet společných akcí, delší čas k zabití bosse a nižší míru přežití.

Random baseline Akce jsou vybírány náhodně z množiny {`move`, `gather`, `attack`, `idle`}. Agenti provádějí akce nezávisle na kontextu (stav ostatních agentů ani okolního prostředí). Výsledkem je téměř nulová kooperace a vysoká pravděpodobnost neúspěchu.

Očekávané přínosy & možné výsledky.

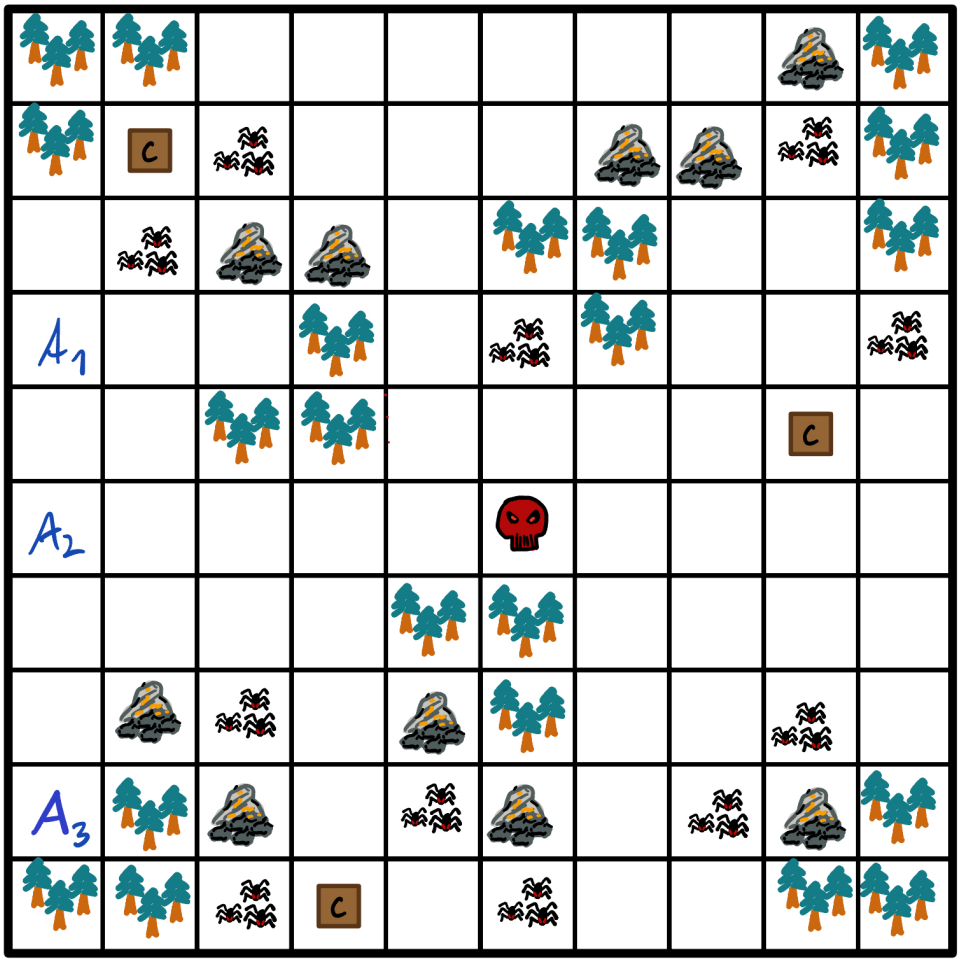
- *Hypotéza:* kooperativní shaping zkrátí `time_to_boss_kill`, zvýší `survival_rate` a počet společných akcí oproti konkurenční politice.
- Pokud obě politiky nakonec dosáhnou stejné úspěšnosti, zaměřím se na *sample-efficiency*: rychlost konvergence a stabilitu výkonu.
- V případě odporu hypotézy popíšu příčiny (např. ztráta času na hledání se navzájem) a navrhnu další směry (např. omezená observabilita, jiná hustota zdrojů).

Technologický stack

- **Jazyk & RL:**
  - Python 3.10
  - Gymnasium (vlastní ArenaEnv – NumPy)
  - Stable-Baselines3 (PPO + parameter-sharing, PyTorch + CUDA)
- **Vizualizace:**
  - Matplotlib + FuncAnimation (rychlé animace, MP4/GIF)
  - Streamlit (interaktivní replay-viewer: slider, play/pause, heatmapy)
- **Analytika & Statistika:**
  - Pandas / SciPy (zpracování dat, t-test, Mann–Whitney U, ANOVA, Tukey)
  - Matplotlib / Seaborn (boxploty, heatmapy, learning-curve)

Prostor pro rozšíření.

- **Částečná observabilita** nebo jednoduchý komunikační kanál mezi agenty.
- **Heterogenní role** (např. tank, healer, DPS) s role-conditioned vstupy a případně individuálními inventáři.
- **Porovnání alternativních algoritmů** (COMA, QMIX/VDN, MAPPO) s koaličními bonusy i bez nich.
- **Přechod na spojité prostředí** – agenti se pohybují v $\mathbb{R}^2$ se spojitými akcemi.



Obrázek 1: Náčrt herního prostředí ArenaEnv (10×10 grid, zdroje, mobové, boss, bench)