

Uczenie ze wzmocnieniem w środowisku Robocode

Patryk Czuchnowski





Obserwacja

```
private static final int DISTANCE_BUCKETS = 5;
2 usages
private static final int ENERGY_BUCKETS = 3;
2 usages
private static final int HEADING_BUCKETS = 6;
2 usages
private static final int GUN_HEADING_BUCKETS = 10;
2 usages
private static final int ENEMY_BEARING_BUCKETS = 6;

1 usage
public Observation(double enemyDistance, double energy, double heading, double relativeGunHeading, double enemyBearing,
    boolean enemyEnergyLoss) {
    this.enemyDistance = bucket( min: 0, max: 1, DISTANCE_BUCKETS, enemyDistance);
    this.energy = bucket( min: 0, max: 100, ENERGY_BUCKETS, energy);
    this.heading = bucket( min: 0, max: 360, HEADING_BUCKETS, heading);
    this.relativeGunHeading = bucket( min: -180, max: 180, GUN_HEADING_BUCKETS, relativeGunHeading);
    this.enemyBearing = bucket( min: -180, max: 180, ENEMY_BEARING_BUCKETS, enemyBearing);
    this.enemyEnergyLoss = enemyEnergyLoss;
    this.zeroBearing = Math.abs(relativeGunHeading) < 3;
}
```



Losowe akcje

```
if (obs.getZeroBearing() && random.nextDouble() < 0.5) {  
    return new Fire( value: 1 + random.nextInt( bound: 4) * 0.5);  
}  
  
if (random.nextDouble() < experimentRate || rewards == null || rewards.isEmpty()) {  
    //out.println("random");  
    return switch (random.nextInt( bound: 7)) {  
        case 0 -> new GoAhead( value: random.nextInt( bound: 10) * 5);  
        case 1 -> new TurnGunLeft( value: random.nextInt( bound: 9) * 5);  
        case 2 -> new TurnGunRight( value: random.nextInt( bound: 9) * 5);  
        case 3 -> new TurnLeft( value: random.nextInt( bound: 9) * 5);  
        case 4 -> new TurnRight( value: random.nextInt( bound: 9) * 5);  
        case 5 -> new GoBack( value: random.nextInt( bound: 10) * 5);  
        case 6 -> new Fire( value: 1 + random.nextInt( bound: 4) * 0.5);  
        default -> null;  
    };  
}
```



Nagrody

- Nagroda za zadanie obrażeń
- Kara za otrzymanie obrażeń
- Duża nagroda za wygraną
- Kara za przegraną
- Nagroda za przeciwnika w polu widzenia
- Kara za brak przeciwnika w polu widzenia
- Kąt jaki musi wykonać działko by trafić w przeciwnika - im mniejszy tym większa nagroda
- Kara za liczbę rund



Trening vs CrazyBot

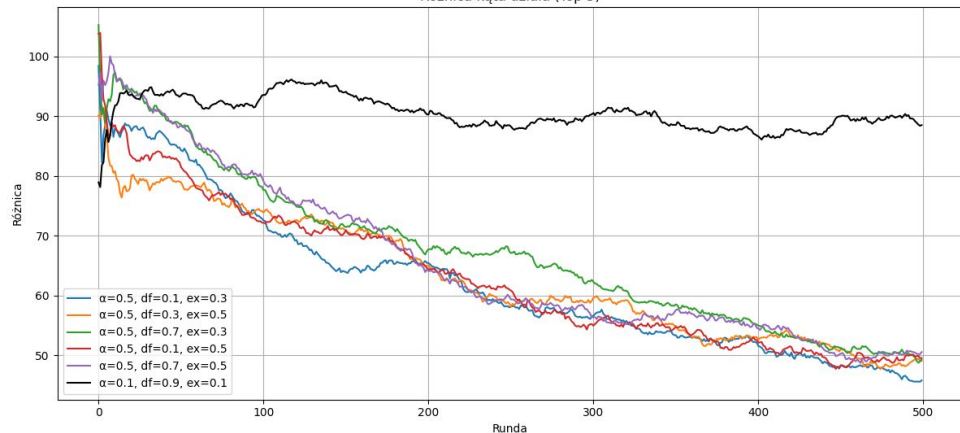
- rozmiar planszy 400x400
- na wykresach przedstawiono średnią kroczącą
- losowość crazy bota pozwala sprawdzić wiele różnych przypadków

Tuning hiperparametrów

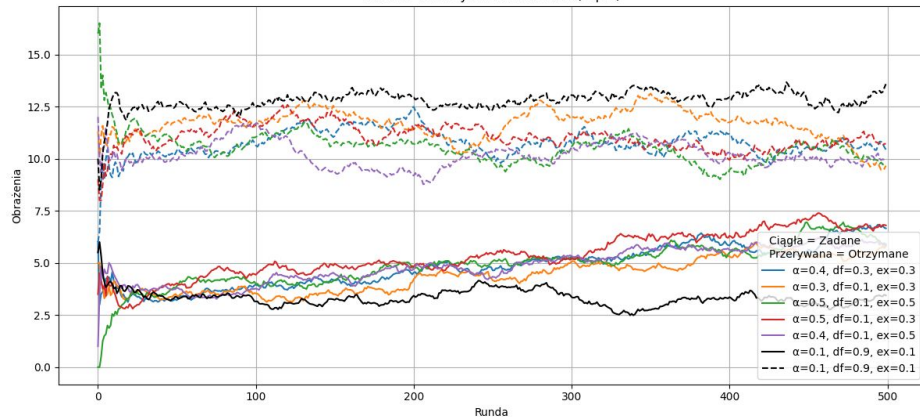
GridSearch:

- $\alpha \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$
- $\text{discountFactor} \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$
- $\text{experimentRate} \in \{0.1, 0.3, 0.5, 0.7\}$

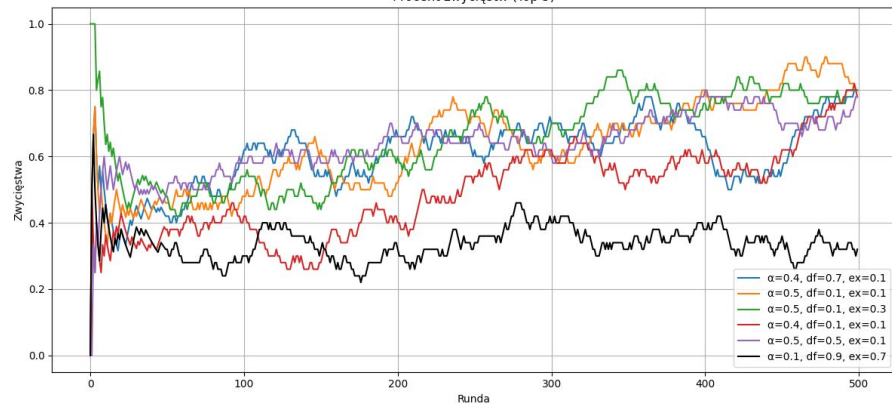
Tuning hiperparametrów vs Crazy Robot
Różnica kąta działu (Top 5)



Tuning hiperparametrów vs Crazy Robot
Zadane vs otrzymane obrazy (Top 5)



Tuning hiperparametrów vs Crazy Robot
Procent zwycięstw (Top 5)

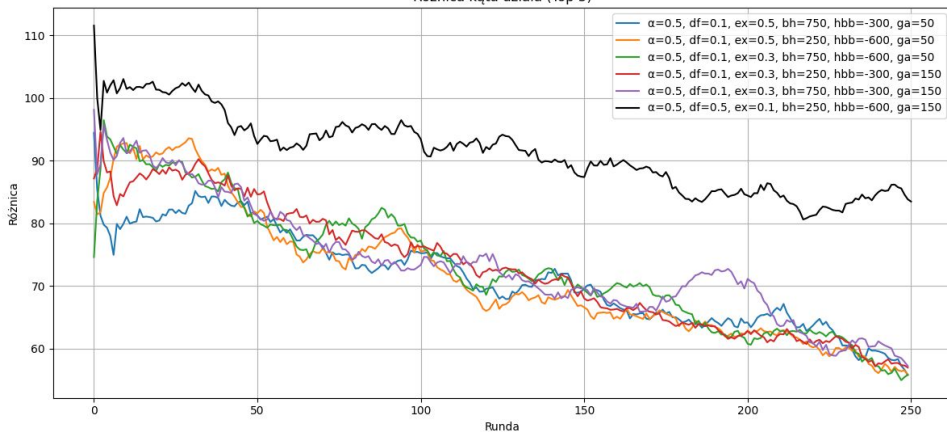


Tuning hiperparametrów

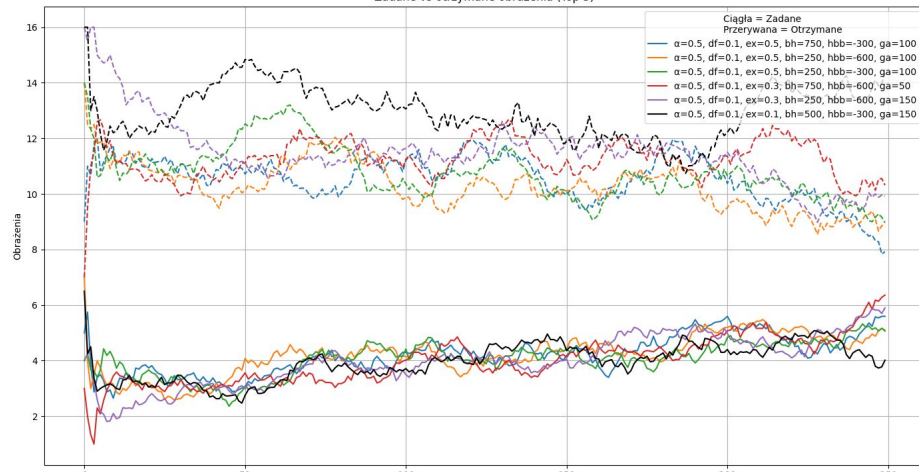
GridSearch:

- bulletHitReward $\in \{250, 500, 750\}$
- hitByBulletPenalty $\in \{-300, -600\}$
- gunAlignmentReward $\in \{50, 100, 150\}$

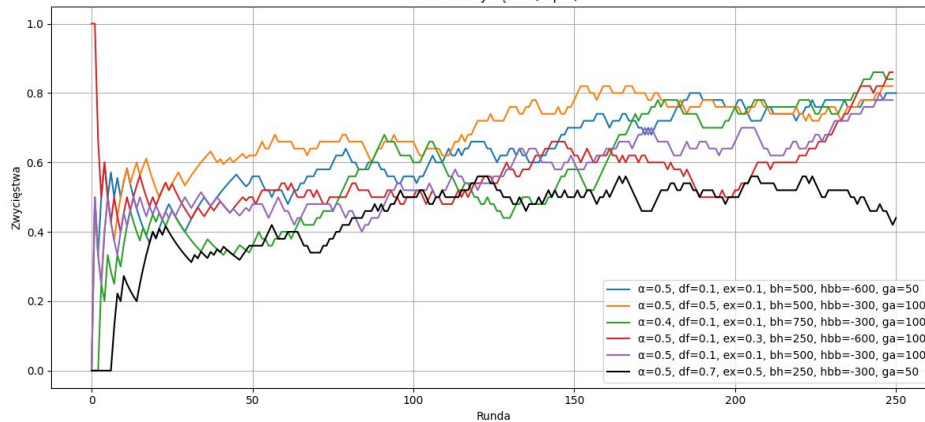
Tuning hiperparametrów vs Crazy Robot
Różnica kąta działła (Top 5)



Tuning hiperparametrów vs Crazy Robot
Zadane vs otrzymane obrazy (Top 5)



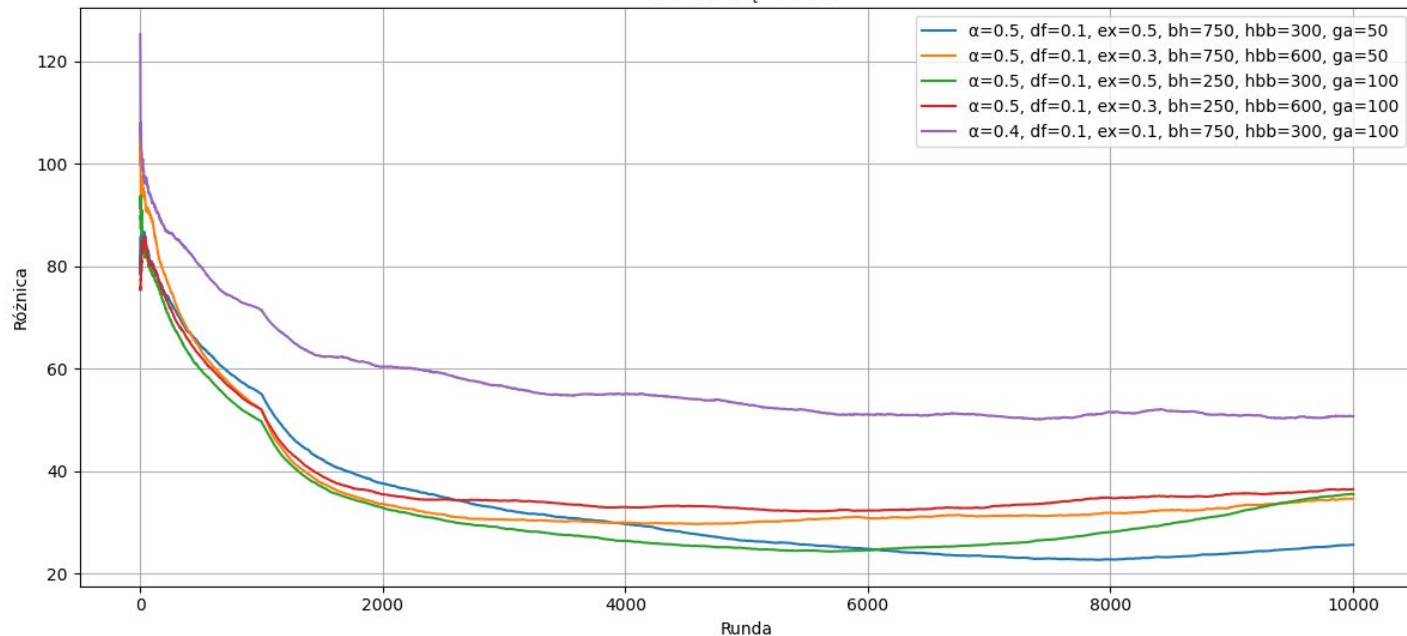
Tuning hiperparametrów vs Crazy Robot
Procent zwycięstw (Top 5)





Trening vs CrazyBot

Tuning hiperparametrów vs Crazy Robot
Różnica kąta działa

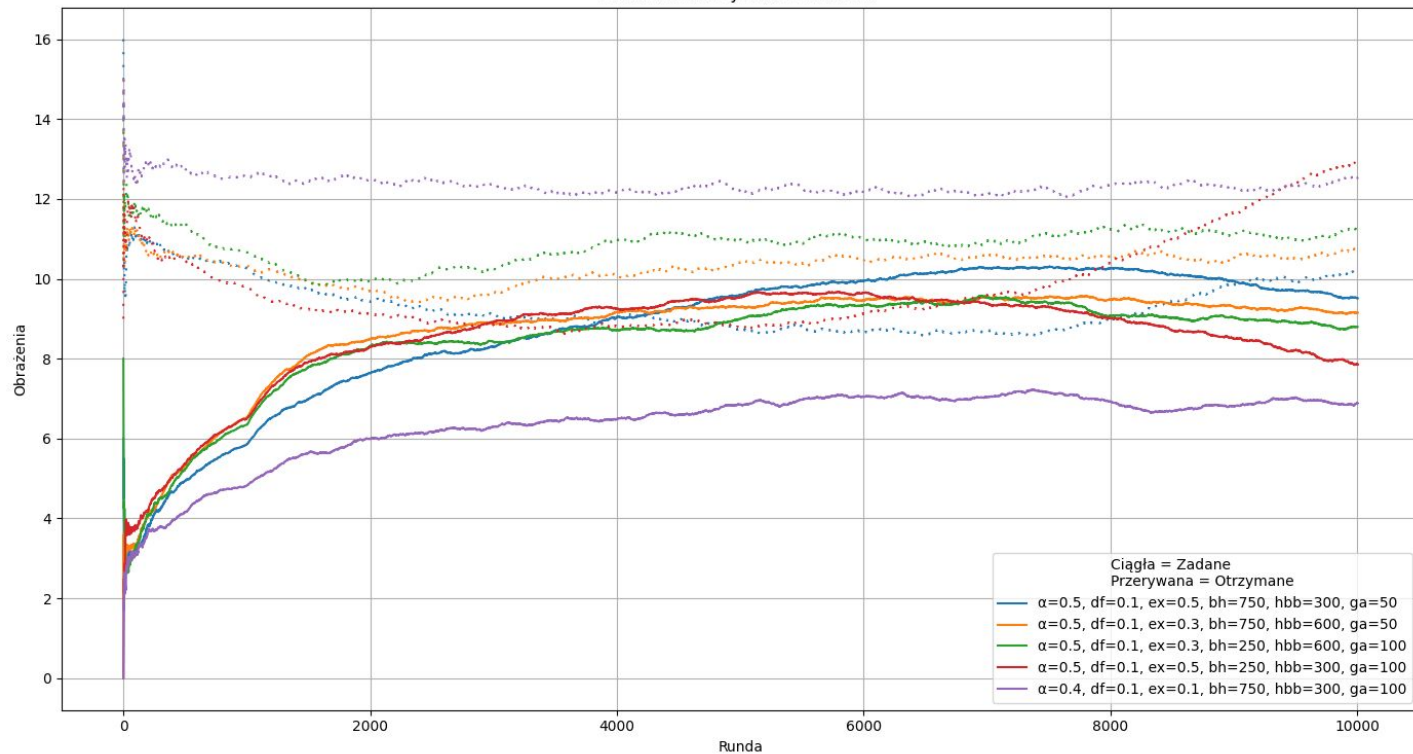


Najlepszy zestaw hiperparametrów niebieska linia
 $\alpha=0.5$, $df=0.1$, $ex=0.5$, $bh=750$, $hbb=300$, $ga=50$

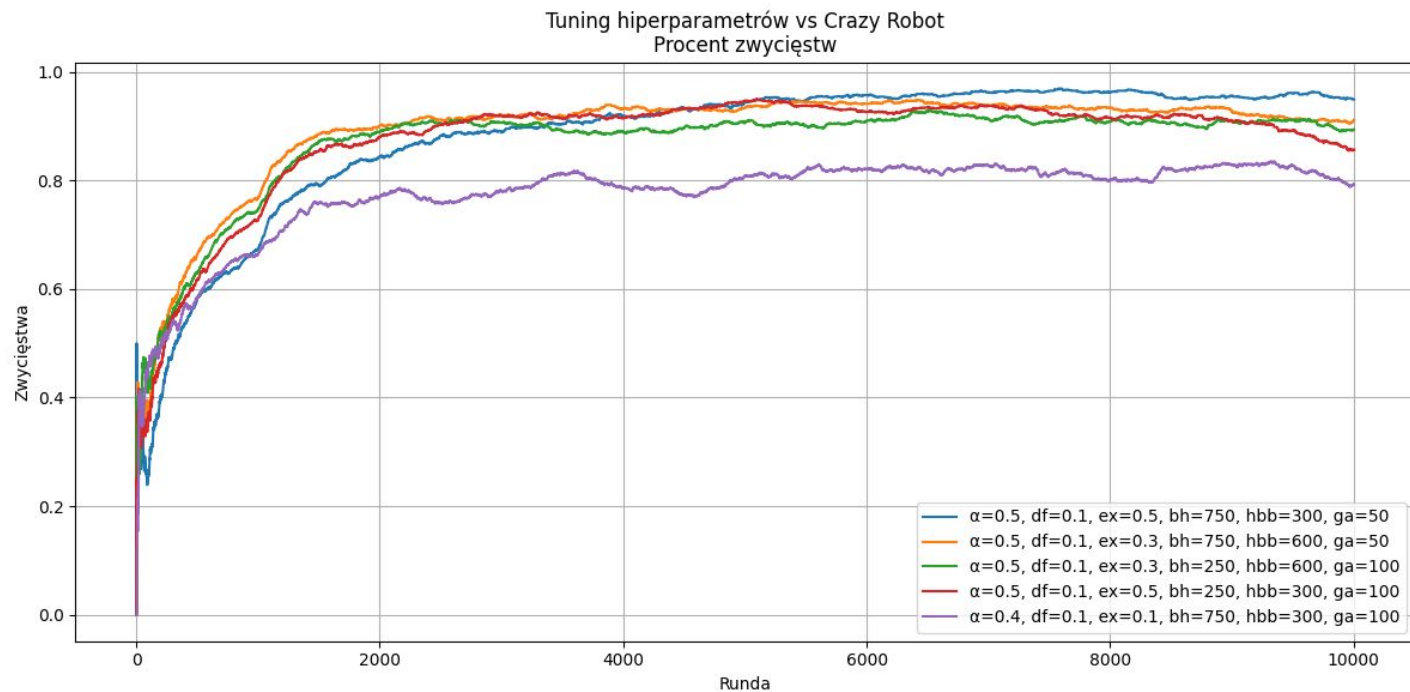


Trening vs CrazyBot

Tuning hiperparametrów vs Crazy Robot
Zadane vs otrzymane obrażenia



Trening vs CrazyBot



Końcowy procent zwycięstw około 95%



Demo

Wyniki bitew

vs Crazy Robot - 949/1000 wygranych

Rank	Robot Name	Total Score	Survival	Surv Bonus	Bullet Dmg	Bullet Bonus	Ram Dmg * 2	Ram Bonus	1sts	2nds
1st	agh.reinforced.ReinforcedLearningRobotLoa...	114366 (68%)	47350	9470	49537	7703	270	35	949	51
2nd	sample.Crazy	53505 (32%)	2550	510	44287	668	5318	171	53	947

vs Corners Robot - 352/1000 wygranych

Rank	Robot Name	Total Score	Survival	Surv Bonus	Bullet Dmg	Bullet Bonus	Ram Dmg * 2	Ram Bonus	1sts	2nds
1st	sample.Corners	160138 (58%)	32400	6480	106235	14958	65	0	648	352
2nd	agh.reinforced.ReinforcedLearningRobotLoa...	116052 (42%)	17600	3520	85421	9410	101	0	352	648

vs Fire Robot - 300/1000 wygranych

Rank	Robot Name	Total Score	Survival	Surv Bonus	Bullet Dmg	Bullet Bonus	Ram Dmg * 2	Ram Bonus	1sts	2nds
1st	sample.Fire	164295 (57%)	35000	7000	106096	16196	2	0	701	299
2nd	agh.reinforced.ReinforcedLearningRobotLoa...	121985 (43%)	14950	2990	95547	8432	66	0	300	700

vs Veloci Robot - 36/1000 wygranych

Rank	Robot Name	Total Score	Survival	Surv Bonus	Bullet Dmg	Bullet Bonus	Ram Dmg * 2	Ram Bonus	1sts	2nds
1st	sample.VelociRobot	181190 (83%)	48200	9640	84818	16724	19572	2237	964	36
2nd	agh.reinforced.ReinforcedLearningRobotLoa...	37071 (17%)	1800	360	33779	525	606	0	36	964

vs Spin Bot - 1/1000 wygranych

Rank	Robot Name	Total Score	Survival	Surv Bonus	Bullet Dmg	Bullet Bonus	Ram Dmg * 2	Ram Bonus	1sts	2nds
1st	sample.SpinBot	177739 (87%)	49950	9990	97109	19472	1159	59	999	1
2nd	agh.reinforced.ReinforcedLearningRobotLoa...	25581 (13%)	50	10	25359	27	134	0	1	999

Dziękuję za uwagę

