# Jutalomfüggvények hasonlóságainak értékelése – avagy egy-egy jobb függvény megtalálásának stratégiája

(Similarity-based evaluation of reward functions – or identification strategy of a best function)

Almásy Márton György, Hörömpő András

Kivonat: A dokumentum rámutat arra, hogy miért érdemes minél nagyobb számosságú megoldási alternatívát készíteni annak érdekében, hogy az ösztönösen, a mindenkori részeredmények alapján intuitívan keletkeztetett nyitó megoldáspopuláció legjobbját célirányosan meg lehessen haladni ezen alternatívák összehasonlítása keretében. Az alacsony számosságú alternatívahalmazok hasonlósági erőterei nem tudnak tetszőleges stabilitás mellett záródni, így bármilyen bonyolult is eljutni egy-egy alternatívához, ezek minimális száma esetén a következő (vélelmezhetően az eddigi legjobbnál is jobb) megoldás már célirányosan (azaz automatizáltan) is feltárható. A dokumentum így rámutat az intuitív fejlesztés korlátozottságából fakadó önkényességi kockázatokra.

Kulcsszavak: genetikai potenciál, kockázatelemzés,

# 1. A projektről

## a. Tumor kezelési eljárás optimalizálása

A daganatos betegségek továbbra is a vezető halálokok között szerepelnek annak ellenére, hogy az utóbbi években és évtizedekben a kezelési módszerek hatékonysága sok esetben nagyságrendileg javult. A jelenleg széleskörűen elérhető terápiák között általánosan alkalmazzák a gyógyszeres kezelést vagy kemoterápiát, amely során a páciens sok esetben hosszútávon is szenved a kezelés mellékhatásaitól, mivel a terápia során az ép és egészséges sejtek egy részére is károsodik. Az általánosan alkalmazott kemoterápiás protokollok alapja jelenleg az, hogy a kezelést a beteg számára még tolerálható maximális dózissal végzik, az adagolt hatóanyag mennyiségével együtt azonban a mellékhatások gyakorisága és súlyosság is növekszik. Épp ezért az elmúlt években több olyan kutatás is indult, amelyek új terápiás protokollok kidolgozását és vizsgálatát tűzték ki célul. Általánosan elmondható, hogy az optimális hatást a lehető legkisebb mennyiségben, de viszonylag gyakran adagolt hatóanyaggal igyekeznek elérni, ezen protokollokkal kapcsolatban számos kérdés azonban jelenleg is nyitott.

Kutatásunkban a kezelés folyamatára intelligens szabályozási eljárásként tekintünk: az alkalmazott hatóanyag beadásának idejéről és a beadott dózis nagyságáról egy virtuális orvosként felfogható öntanuló eljárás, egy úgynevezett ágens hoz döntést. Az ágens által hozott döntések hatását egy matematikai modell segítségével szimulált páciensen vizsgáljuk, az ágens különféle akcióinak következtében a szimulált páciens állapota megváltozik, miközben a tumor reagál a kezelésre (vagy épp annak a hiányára).

## b. Alkalmazott matematikai modell – Drexler modell

Drexler és munkatársai cikkükben egy minimál tumornövekedési modellt mutatnak be, melyben a növekedési dinamikát viszonylag alacsony számú, relatív könnyen identifikálható paraméter, valamint az adagolt hatóanyag együttesen alakítják. Magát a modellt rendszer- és irányításelmélet alapokra fektették, a szabályozás során ügyelve számos, kezeléssel kapcsolatosan felmerülő kérdésre. A modellt formális reakciólépések alapján származtatott differenciálegyenletek segítségével írják fel az alábbi formában:



A paraméterek, és az egyenletben szereplő ismeretlenek az alábbiak: x1: az osztódásra képes tumorsejtek össztérfogata köbmilliméterben x2: a halott tumorsejtek össztérfogata köbmilliméterben x3: a vérben lévő hatóanyag koncentrációja milligramm/kilogrammban u: a befecskendezésre kerülő hatóanyag mennyiségét milligramm/kilogramm ・ nap mértékegységben leíró függvény a: a tumor növekedési rátája n: a tumor nekrotikus rátája (pusztulás mértéke) b: a befecskendezett dózis hatása c: a hatóanyag ürülési rátája (klírensz) w: a nekrotikus szövet kimosódási rátája

ED50: a hatóanyag effektív medián dózis értéke (elméleti érték, ilyen dózis mellett érné el a gyógyszer az 50%-os sejtpusztulást) KB: a hatóanyag Michaelis–Menten-konstansa bk: a hatóanyag felszívódási gyorsaságát jellemző érték

A rendszer kimenete (y) egy olyan, köbmilliméterben kifejezett térfogat, amely az élő és az elpusztult tumorszövet együttes térfogatát adja meg. A rendszer dinamikáját az



egyenlettel lehet leírni, azaz a tumorméretbeli változást közvetlenül a tumornövekedési rátával (a) és a jelenleg "aktív", szaporodásra képes tumorsejtek térfogatával lehet jellemezni. Megállapítható, hogy a rendszer akkor van egyensúlyi állapotban, ha a szaporodásra képes tumorsejtek térfogata (x1) 0 köbmilliméter, továbbá a vérben lévő dózis mennyisége (x3) 0 milligramm/kilogrammal egyezik meg, ha az első egyenletben szereplő a – n érték negatív szám, ami akkor lehet, ha a több tumorsejt pusztul el, mint amennyi újonnan, szaporodó képes állapotban van jelen. Akkor viszont nyeregpontja van a rendszernek, ha ugyanezen kifejezés értéke pozitív, azaz, ha több a szaporodóképes, új tumorsejt, mint az elpusztult. A rendszer irányíthatóságára vonatkozóan akkor tekinthető ez a rendszer irányíthatónak, azaz válik képessé arra a rendszer, hogy a befecskendezett dózis hatására a tumorsejtek térfogata csökkenjen, ha az a − n − b kifejezés negatív.

## c. DQN

A DQN ugyanúgy, ahogy az összes Q-tanulás alapú módszer, ez is jutalom maximalizálásra törekszik, az epizód végére. Ezt ez a megoldás nem egy egyszerű lookup-table megoldásával teszi, hanem itt hívja segítségül a neurális hálót, ami adott state-re választ egy action-t. Ennek a megoldásnak az előnyei szemmel láthatóak a sima Q táblás megoldáshoz képest. Mivel itt egy DNN-el (Deep Neural Network-kel) dolgozunk, a háló képes arra hogy mintákat, összefüggéseket találjon state-ek és döntések (action-ok) között olyan komplexitással, amikre algoritmikailag szinte lehetetlen vagy csak nagyon nehezen lehetne. Ezen felül azért is egy robosztus megoldás a DQN mivel, ha végtelen vagy a végtelenhez közeli hatalmas állapottérrel (state space-el) rendelkezünk, akkor ezt a rengeteg értéket nem tudjuk eltárolni bármekkora adatszerkezetben. Előbb-utóbb ki kellene dobnunk egy részét, tömörítenünk kellene, ami mind-mind a pontosság rovására menne. Ezzel szemben a DQN ezeket az összefüggéseket a neurális háló súlyaiban tanulja és őrzi meg (vö. 1. ábra):



1. ábra: A jutalmazás szerepe/folyamata (forrás: saját ábrázolás)

## d. Mély megerősítéses tanulás

Fontosnak tartjuk azt, hogy legelőször kettéválasszuk a megerősítéses tanulást és a felügyelt tanulást. Felügyelt tanulásra egy konkrét példa, amikor képeket szeretnénk osztályozni. Nincsen konkrétan definiálva a cél, jövőbeli terv vagy bármi, amit el akarnánk érni, csak az, hogy minden kép egy előre meghatározott csoporthoz tartozzon. Ez nagyon absztrakt szinten olyan, mint egy sima függvény, amit, ha meghívunk, akkor szétválogatja a képeket. Ezzel szemben a megerősítéses tanulás sokkal inkább olyan, mint egy ciklus. Iterációkban lépkedünk előrébb és előrébb egy konkrétan definiált cél érdekében. Az algoritmusnak a célt mutatjuk, azt, hogy hova kell eljutnia. És nem mondjuk meg neki, hogy ezt hogyan csinálja, hogy minden egyes lépésben mit tegyen.

Egy konkrét, nagyon egyszerű példán keresztül mutatnánk be mély megerősítéses tanulás folyamatát. Tegyük fel, hogy van egy kiskutyánk (agent). A feladata az, hogy visszahozza a botot, amit eldobunk neki a réten (ez egy ideális episode). Az, hogy a bot hol található, az maga az environment. A kutyánk kap jutalom falatot (reward), ha visszahozza a botot, és nem kap, ha nem sikerül neki. A kiskutya választhat azok közül, hogy felveszi-e a botot vagy nem, vagy hogy melyik irányba indul el a botért (ez az action). Ha a bot ott van a közelében, akkor fogja meg és hozza vissza (ilyen policy-t, irányelvet követ).

## e. A feladat kulcsa: jutalomfüggvények

A jutalomfüggvény, definíció szerint azon része kutató munkánknak, amely osztályozza modellünk döntéseit jóság alapján. A kutatásunk során egyik legnagyobb kihívás elé állított minket, hogy a megfelelő jutalomfüggvényt, jutalomfüggvényeket megtaláljuk, illetve azokat tudjuk effektíven „pontozni” jóság, feladatba illőség, komplexitás szerint. Mindegyik függvényt, mely vizsgálat alá került, több szempont figyelembe vételével vizsgáltuk meg, amik javarészt az alábbiak voltak:

* Mi volt a probléma, amit megoldottak vele, azaz mire is használták pontosan?
	+ megnevezzük magát a betegséget, mely kezelésének optimalizálására készült o megadjuk, hogy sugárterápia, kemoterápia vagy esetlegesen más jellegű kezelést alkalmaztak-e
	+ és milyen eredményekkel járt, miket értek el ezzel a függvénnyel?
* Milyen matematikai modellre, összefüggésre alapozták?
	+ a modellek milyen paraméterekből építkeztek, épültek fel o a hosszú túlélésre törekedtek, vagy inkább a betegség leküzdésére összpontosítottak
	+ megadjuk a függvények felépítését, miszerint hány függvényből állnak, illetve milyen ezeknek a függvényeknek a komplexitása
* Ezeket összevetve hogyan tudna a jelen problémán segíteni?
	+ maga az elképzelés működőképes-e vagy sem
	+ eldöntjük, hogy magukban is megálljak a helyüket vagy módosításra van szükségünk egy-egy helyen
	+ esetlegesen hiányzó paraméterek esetén meg tudnánk-e oldani, hogy ezek a paraméterek valahogy előállíthatóak-e vagy sem

# 2. Jutalomfüggvények

## a. Yazdjerdi-féle jutalomfüggvény

Yazdjerdi és más kutatók által vizsgált „Anti-angiogenic” terápia egy molekuláris szintű terápia, ami kemoterápiával és radioterápiával tarttatja egy bizonyos szinten a rákos daganatot (de nem eliminálja azt teljesen). Ennek is a célja az volt, hogy mikro dózisokkal nagyon finoman szabályozott kezelést kapjon a páciens annak érdekében, hogy az életét minél inkább meg tudjuk hosszabbítani.

A szóban forgó reward function amit alapul vettek:

Ez a függvény egy másik függvényt használ fel, amit hiba függvényként neveztek el. Ez a hiba függvény az elvárt tumor mérettől való távolságot adja vissza abszolútértékben. A jutalomfüggvényben pedig ez a hiba van felhasználva és van összehasonlítva a jelenlegi időpillanatban lévő és az ezt követő időpillanatbéli hibával.

Ebből a működésből levonható pár konklúzió:

* Nincs konkrétan számon tartva, hogy a páciens mennyi ideig élt.
* Nem jutalmazzuk azt az esetet, ha sikerült mégiscsak felszámoljuk a tumort.
* Csak a tumor méretét veszi figyelembe, kemoterápia / sugárterápia dózisokat pedig nem, ami szintén nem elhanyagolható.

## b. Hassani-féle jutalomfüggvény

Hassani és társa is majdnem pont ugyan azt a célt tűzte ki, mint amit mi is kutatunk, mégpedig hogyan lehet minimalizálni a tumor méretét a lehető legkevesebb beadott gyógyszer mellett. Előre haladott stádiumú rákos emberek kezelését optimalizálták, akik a kemoterápiát infúzión keresztül kapták. Ők nem tértek ki a rák pontos típusára, így nem kell attól tartanunk, hogy a megoldásuk és a jutalom függvény rák specifikus lenne. Véleményük szerint a megerősítéses tanuláson alapuló rák kezelés optimalizálás a meglévő megoldások következő lépcsője, és kiválóan használható egyedi karakterisztikákkal rendelkező páciensek pontosabb kezelésére.

A jutalom függvény, amit használtak, meglepően egyszerű:



A tumor sejt populáció változásának a hányadosát majd ennek a tízes alapú logaritmusát vették. Tehát a fókuszban náluk a tumor sejt populáció állt.

Ha nő a tumor sejt populáció, akkor negatívan jutalmazunk, míg ha csökken, akkor pedig pozitív érték kerül megállapításra. A megerősítéses tanuláson alapuló megoldásuk „action”-jei diszkrét normalizált ellenszer adagolásból álltak (0.1, … 1) közötti értékekkel.

A fentebb taglalt függvényt viszonylag egyszerűen be lehet illeszteni a mi megoldásunkba abban az esetben, ha egy egyszerűsítéssel élünk, és azt mondjuk, hogy a tumor sejt populáció egyenlő a tumor méretével a mi esetünkben. Ez a módosítás szükséges, mivel nem tudjuk megmérni a tumor sejt számát, maximum becsülni, viszonylag nagy hibával a tumor méretéből, így nyugodtan dolgozhatunk a tumor méretével egyből. Ez mind szép és jó, viszont a másik legnagyobb problémát, a gyógyszer koncentrációt, nem veszi figyelembe. Ez így önmagában azt eredményezné, hogy rengeteg gyógyszert adnánk a páciensnek, amibe bele is halna. A konklúzió az, hogy ezt a megoldást lehetne alkalmazni némi módosítással, hogy figyelembe vegyük a jutalom kiszámításánál a vérben lévő ellenanyag-koncentrációt is.

## c. Zhao-féle jutalomfüggvény

A Zhao és társai által kutatott terület szintúgy kapcsolódik az általunk vizsgált témához, számukra is a cél az életet közvetlenül veszélyeztető betegségek, mint például a rák elleni küzdelem, melyhez a megerősítéses tanítást használták fel. Igyekeztek a lehető legjobban optimalizált megoldást, stratégiát találni, melyekhez klinikai adatokat használtak fel közvetlenül mindamellett, hogy a megfelelően pontos matematikai modell megtalálására kisebb hangsúlyt fektettek. Megjegyzendő tanulság részükről, hogy „nagy potenciál rejlik” a megerősítéses tanításnak a gyógyászatban történő hasznosítására, amellett is, hogy a rendszer „folyamatosan tanul”, még a „késleltetett hatások” figyelembe vételét is, még akkor is, ha az egyes „cselekvések és eredmények közti kapcsolat nem teljesen ismert”.

Az általuk használt jutalomfüggvények az alábbiak szerint definiálhatóak:



Ezekben Wt az adott időpillanat-beli wellnessértéket mutatja, Mt az adott időpillanat-beli tumorméretet mutatja, Dt pedig az adott időpillanat-beli dózismennyiséget mutatja.

Ahogy azt a rész-jutalomfüggvények is mutatják, amennyiben egy adott határértéken túlmutatnak az egyes paraméterváltozások, úgy azt annak megfelelően plusz jutalommal illeti, ha az a betegség gyógyulása szempontjából pozitív hatással, illetve negatív jutalommal illeti, ha az a gyógyulás szempontjából negatív hatással van a páciensre. Megemlítendő tényező, hogy amennyiben a következő, t+1-edik időpillanatban a tumor mérete beáll egy adott -0.5-ös értékre, úgy abban az esetben eredményeződik a harmadik rész-jóságfüggvényben a legmagasabb jutalom, hogy ezt maximalizálni lehessen, fontos mellé, hogy a páciens wellnesse minél jobb legyen, azaz egy adott időpillanatból ezen t+1-edikbe lépve ne változzon negatív irányba radikálisan.

Kapcsolópontja ezen megközelítésnek, illetve a kutatómunkánkhoz, hogy ezen függvény és megközelítés számára is a túlélés a legfontosabb szempont, melyben döntő szerepe van a besugárzott dózisnak, valamint a tumor méretének is. A függvény lehetőséget ad súlyozásra is, melyik számít az összeredménybe és -jutalomba nagyobb mértékben. Ezen értékek számunkra is adottak és ismertek, így figyelembe vehető, alkalmazható függvény lehet.

## d. Yauney-féle jutalomfüggvény

A Yauney által vezetett kutatás szintén rávilágított arra, hogy megfelelő, jól meghatározott jutalomfüggvény nélkül a megerősítéses tanítás nem működhet kellően hatékonyan olyan területeken, mint az egészségügy és a gyógyítás. Holott ennek meghatározásával klinikai teszteket megelőző dózisadagolásokat lehet meghatározni, melyek később megerősített eredményével akár tényleges gyógyításra lehet fordítani. Fő célpontja, illetve jutalomfüggvényének alapja, hogy az átlagos tumor átmérő (mean tumor diameter – MTD) a lehető legkisebb legyen, miközben a páciens sugárkezelést kap. Természetesen nem minden áron történő csökkentésről van szó, figyelembe veendő a páciensnek adott dózismennyiség is, a túlzottan nagy mennyiség ugyanis károsító, mintsem gyógyító hatással van a kezelt betegre.

A Yauney-féle jutalomfüggvény az alábbiak szerint definiáltható:

A jutalomfüggvényről megállapítható, hogy törekszik a minél kisebb tumorátmérő elérésére, ugyanakkor figyelembe veszi negatív súllyal, ha dózist kapott a kezelt páciens. Adott pillanatban optimális megoldásnak tűnik, ha a tumorméret csökken egy adott mértékben, lehetőség szerint minél jobban, amellett, hogy nem történt dózisadagolás.

Az utolsó epizódban történő jutalomfüggvény-érték a végső eredményt jobban jellemzi, hiszen abban megjelenik, mekkora tumorméret-csökkenést sikerült elérni a folyamat során.



Megállapítható, hogy jelen kutatásban is szerepe lehet ennek a jutalomfüggvénynek, illetve annak implementálásának. Meg tudjuk figyelni ugyanis a tumor méretét, közvetlen paraméterrel rendelkezünk hozzá. Továbbá tudjuk azt is ellenőrizni, mikor történt a páciens számára dózis adagolás. Mivel mindkettő értékkel tudunk dolgozni, fel tudjuk használni a függvény készítéséhez, így megfigyelendő, milyen hatása lehet jelen problémánkra. A paraméterek változtathatóak, így azok finomhangolásával az eredmények javulása várható.

# 3. A feladat megvalósítása

## a. Feladat célja

A feladat fő célja az volt, hogy az általunk megvizsgált és lefuttatott jutalomfüggvények paraméterei alapján vajon készíthető-e a meglevőknél jobb jutalomfüggvény? Ha igen akkor mik kellenek hozzá? Vagy mik nem? Emiatt egy online solvert hívtunk segítségül (<https://miau.my-x.hu/myx-free/> + <https://miau.my-x.hu/myx-free/coco/index.html>).

A függvényépítés szimulált demo-ja itt érhető el: <https://miau.my-x.hu/miau/279/jutalomfuggvenyek.xlsx>

A feladat lépései a következők voltak:

* Meghatároztuk az attribútumok halmazát.
* Kitöltöttük a jutalomfüggvények ismeretében az ezeket jellemző értékeket.
* Meghatároztuk az Y kimeneti értékeket, hogy mi lesz az (pl. keletkezési sorrend)?
* Ezt is kitöltöttük? (vö. pl. háttér-xls)
* Futtattunk egy/több klasszikus solvert (termelési függvénybecslés, pl. idő-múlás modellezése, a programozói intuíciók értékének tesztelése céljából)
* Eredmények értékelése

## b. Paraméterek megnevezése

Ez az alfejezet kifejezetten a paraméterekről fog szólni. Azokról a tulajdonságokról, amik alapján megpróbáltuk kategorizálni az egyes jutalomfüggvényeket. Fontos megjegyezni, hogy két nagy csoportot különítettünk el. Az egyik ilyen csoport inkább élettani kategóriájú volt, az, hogy a tumor kezelésére milyen megoldásokat használ, miket vesz figyelembe, míg a másik nagy kategória inkább matematikai alapú megközelítés volt (milyen matematikai összefüggéseket, függvényeket használnak). Az attribútumok az alábbiak voltak:

* Tumor kezelésre vonatkozó attribútumok
	+ Chemo terápiánál használt-e?
	+ Radio terápiánál használt-e?
	+ Tumor térfogatot figyelembe veszi-e?
	+ Tumor-/más sejtszámot figyelembe vesz-e?
	+ Tumor átmérőt figyelembe vesz-e?
	+ Tumor súlyát figyelembe veszi-e?
	+ Valószínűséggel dolgozik-e?
	+ Van-e wellness érték benne?
	+ Van-e negatív büntetés egy rossz lépés esetén?
* Matematikailag attribútumok
	+ Használ-e másmilyen függvényt még?
	+ Összetett függvény-e?
	+ Definiál-e külön final state-t?
	+ Használ e valamilyen differenciát? (múltbéli különbséget)

## c. Átlag steps értékek jelentése

Felmerülhet bennünk olvasva, hogy fentebb említést tettem a kimeneti (Y) értékéről. Ez a jutalom függvényeink tekintetében az ”átlagos steps” értékek voltak a valóságban – ill. lettek volna a modellezés során, ha lett volna kellő számú függvényhez ilyen következményérték. A háttér-xls megfelelő adatok hiányában Y-ként az egyes függvény-generációkat, vagyis a fejlesztési idő múlását, más szavakkal a fejlesztői intuíció vélelmezett szintjét igyekezett hermeneutikai célokra kiaknázni.

A steps-értékről: Egy-egy kezelésünk 100 napig tartott. Ez az, ami megmutatja azt, hogy átlagosan 100 napból egy páciens meddig élt túl. Így a maximum elérhető pont itt a 100 volt. Fontos azt is megemlíteni, hogy ezek az átlag értékek több ezres kísérletek végeredményei. Az sem elhanyagolható, hogy voltak olyan jutalomfüggvényeink, amik sajnos sehogy vagy nagyon rosszul teljesítettek, így azokat idő előtt elvetettük. Ezek a ”józan paraszti ész alapú” továbbfejlesztésre szorulnak a későbbiekben. Emiatt ez lehet az, ami az objektumlistánkat a későbbiekben bővíteni tudja.

Az alábbi értékeket kaptuk azokra a függvényekre, amiket használtunk:

|  |
| --- |
| ￼Twocompartment csúszó ablakos különbözetű Drexler modell esetén￼  |
| ￼Yauney-féle jutalomfüggvény￼  | 90.785  |
| Hassani-féle jutalomfüggvény  | 86.838  |
| ￼Yazdjerdi-féle jutalomfüggvény￼  | ￼80.7￼  |
| ￼Zhao-féle jutalomfüggvény￼  | ￼17.792￼  |

## d. Eredmények a feladatban (S1)

Alapvetően az a cella tartalmazza azt az értéket, hogy mennyivel jobb jutalomfüggvényt csinálhatnánk, ami mellé az van írva, hogy S1 (vagyis a lépcsős függvények változónkénti 1. szintje – vö. aranyérem ellenértéke). Ennek itt az értéke igen magas: pl. 258323 Ez valószínűleg azért ennyi, mivel elég kevés jutalomfüggvénnyel dolgozunk és emiatt túl nagy a genetikai potenciál – ami nem más, mint az S1 értékek attribútumunkénti összege.

Itt (vagyis az alacsony objektumszám esetén) sajnos nem korlátozza semmi sem az objektumok egy-tényezős összazonosságát. Azt is fontosnak tartjuk megemlíteni, hogy szintén nincs, ami korlátozza itt a maximum elérhető összeget a 100-at. Erre is később valamifajta megoldást kellene találni: vö. háttér-xls.

2. ábra: Kevés elemszámú modellezés korlátozott értékű részeredményei (forrás: saját ábrázolás)

## e. Konklúziók

Összességében elmondható az egyértelműen, hogy ha több jutalomfüggvényünk lenne, akkor nagy valószínűséggel értelmesebb kimenetelt kaphatnánk (vö. háttér-xls módszertani lépései). Emiatt talán az lehetne a következő lépés, hogy bővítjük ezen jutalomfüggvények halmazát, ha sikerül a meglévő rel. ”rosszakat” feltuningolni annyira, hogy használható eredményt produkáljanak. Itt az is kihívást jelent egyébként az átlag steps kialakításában, hogy a neurális háló, amit ilyenkor tanítunk nem mindig ugyanabból a random értékekből áll elő így van egyfajta fluktuáció az elején, ami torzíthatja ezt az átlagot egy picit.



3. ábra: Kevés objektumra alapozó elemzés maximális steps-értékének becslése (forrás: saját ábrázolás)

Tartalomjegyzék

[1. A projektről 1](#_Toc90147524)

[a. Tumor kezelési eljárás optimalizálása 1](#_Toc90147525)

[b. Alkalmazott matematikai modell – Drexler modell 2](#_Toc90147526)

[c. DQN 3](#_Toc90147527)

[d. Mély megerősítéses tanulás 3](#_Toc90147528)

[e. A feladat kulcsa: jutalomfüggvények 4](#_Toc90147529)

[2. Jutalomfüggvények 4](#_Toc90147530)

[a. Yazdjerdi-féle jutalomfüggvény 4](#_Toc90147531)

[b. Hassani-féle jutalomfüggvény 5](#_Toc90147532)

[c. Zhao-féle jutalomfüggvény 6](#_Toc90147533)

[d. Yauney-féle jutalomfüggvény 7](#_Toc90147534)

[3. A feladat megvalósítása 8](#_Toc90147535)

[a. Feladat célja 8](#_Toc90147536)

[b. Paraméterek megnevezése 8](#_Toc90147537)

[c. Átlag steps értékek jelentése 9](#_Toc90147538)

[d. Eredmények a feladatban (S1) 9](#_Toc90147539)

[e. Konklúziók 10](#_Toc90147540)