Sustiprinto mokymosi modelio jautrumo parametrams ir stabilumo tyrimas: PECS kortelių atvejis

Andrius Lukas Maslovas, Asta Slotkienė

Vilniaus universitetas, Matematikos ir informatikos fakultetas,
Universiteto g. 3, Vilnius, Lietuva
Andrius.Maslovas@mif.stud.vu.lt

Santrauka. Tyrime analizuojamas „CardFinder“ algoritmas, skirtas apmokyti sustiprinto mokymosi modelį, gebantį surasti tinkamą PECS kortelę. Šio algoritmo pagrindu sukurtas „CardFinder+ Sense“ algoritmas, kuriuo siekiama pagerinti modelio veikimą ir jo pritaikomumą autizmo spektro sutrikimą turintiems vaikams. Atliktos įvairios modifikacijos, kuriomis siekta sumažinti žingsnių skaičių, per kuriuos modelis suranda tinkamą kortelę. Nustatyti hiperparametrai, darantys didžiausią įtaką mokymosi stabilumui, ir parinktas jų rinkinys, užtikrinantis kuo pastovesnius rezultatus ieškant teisingos kortelės. Algoritmas papildytas funkcija, leidžiančia pasirinkti pradinės kortelės vietą mokymosi ir testavimo procesų metu. Ši funkcija leido „CardFinder+ Sense“ algoritmui surasti teisingą kortelę per mažesnį žingsnių skaičių, lyginant su pirminiu „CardFinder“ algoritmu.

Raktiniai žodžiai: PECS, sustiprintas mokymasis.

1. Įvadas

Asmenų komunikaciniams gebėjimams lavinti dažnai pasitelkiamos specializuotos komunikacinės sistemos, pavyzdžiui, paveikslėlių apsikeitimo komunikacijos sistema (toliau – PECS), kuri yra viena iš populiaresnių ir efektyvesnių tokio tipo sistemų [1]. Kortelėmis galima naudotis įvairiais būdais, pavyzdžiui, vaikas gali paduoti ar parodyti vieną kortelę tėvams ar prižiūrėtojui, rodančią vaiko poreikį, arba, labiau pažengusių arba užaugusių vaikų atveju, jie gali sudaryti ištisą sakinį, nusakantį veiksmą, objektą ar vietą. Kadangi skirtingi asmenys turi skirtingus komunikacinius gebėjimus arba poreikius, o atskirų kortelių naudojimo dažnis kiekvienam individui skiriasi, aktualu integruoti PECS sistemas su mašininio mokymosi metodais, siekiant individualizuoti jų naudojimą ir pagerinti komunikacinius gebėjimus.

DOI: https://doi.org/10.15388/LMITT.2026.17

Vaikams, turintiems autizmo spektro sutrikimą (toliau – ASS), yra būdingi mokymosi gebėjimų skirtumai, todėl yra kuriamos prie jų prisitaikančios sistemos arba sistemų variacijos, tinkančios įvairaus būdo vaikams. „Kinems“ platforma siūlo įvairius kompiuterinius žaidimus vaikams, turintiems mokymosi sutrikimų. Žaidimai yra atrenkami pagal kiekvieno vaiko pageidavimus, o jų sudėtingumas gali būti keičiamas, siekiant išlaikyti vaiko motyvaciją žaisti. Nagrinėto tyrimo metu buvo tiriama sistemos nauda ASS turintiems vaikams ir nustatyta, kad „Kinems“ platformos žaidimai padarė teigiamą įtaką vaikų judrumui [2].

Sustiprintas mokymasis – tai mašininio mokymosi kryptis, grindžiama principu, kad teigiamas atoveiksmis arba situacijos pagerėjimas, sekantis po konkretaus veiksmo, sustiprina polinkį tą veiksmą kartoti [3]. Sustiprinto mokymosi algoritmai apmokomi per daugybę veiksmų iteracijų, kurias atitinkamoje aplinkoje atlieka vadinamasis agentas, siekdamas rasti trumpiausią ir rezultatyviausią kelią iki tikslo. Ši savybė leidžia generuoti duomenis sąveikaujant su aplinka ir rasti optimaliausią strategiją nesiremiant aplinkos modeliu.

ASS yra būdingi kognityvinių įgūdžių skirtumai, todėl yra siekiama individualizuoti mokymosi procesą pagal kiekvieno mokinio gebėjimus. 2020 metų tyrime yra siūloma mokymo sistema, apjungianti sustiprinto mokymosi algoritmus, kad pasiūlytų individualius klausimus, pagal kiekvieno ASS turinčio studento sugebėjimus [4].

Sustiprinto mokymosi strategijų taikymas alternatyviosios augmentinės komunikacijos srityje tampa ypatingai naudingas tuomet, kai reikia individualizuoti sistemos panaudojamumą, kai atitinkamų duomenų rinkinių modeliui apmokyti nėra arba negali būti. PECS kortelių atveju, dalis kortelių gali būti naudojamos retai arba visai nenaudojamos, dėl to tinkamos kortelės paieška kaladėje užtrunka ilgiau ir gali trikdyti arba kitaip neigiamai paveikti komunikaciją.

Tyrimo metu buvo analizuojamas sustiprinto mokymosi algoritmas „CardFinder“, sukurtas autoriaus Augusto Mikulėno, kuris yra skirtas pasiūlyti PECS korteles [5]. Tyrimo metu yra pastebėti trūkumai, dėl kurių „CardFinder“ algoritmas negali būti deramai pritaikytas ASS turinčių vaikų panaudojimui. Kiekvieno apmokymo metu yra pastebimas rezultatų nepastovumas. Maždaug 13 % apmokymo atvejų, modelis kortelę randa lėčiau, nei optimalus algoritmo rezultatas. Tai reiškia, kad bent 13 % atvejų jis gali veikti prasčiau negu numatyta. Geriausias modelio pasiekiamas rezultatas

yra 5 žingsniai, algoritmas pirmiausia parodys 4 neteisingas PECS korteles ir tik 5 kartą teisingąją, tai gali lemti neigiamą patirtį ir sistemos atmetimą, jei vaikui būtų rodoma pernelyg daug netinkamų kortelių.

Šiuo tyrimu yra siekiama pasiūlyti PECS kortelių panaudojamumo strategijas, grindžiamas sustiprinto mokymosi principais, ir užtikrinti modelio veikimo stabilumą.

2. Tyrimo metodologija

Tyrimui atlikti naudoti Augusto Mikulėno „CardFinder" algoritmo [5] metodai, skirti agentui išmokti surasti teisingą kortelę iš 81 kortelės kaladėje (toliau – aplinkoje). Algoritmas veikia pagal konkretų sustiprinto mokymosi metodą – Q-mokymą. Šis metodas suteikia agentams galimybę išmokti optimaliai veikti Markovo grandinėse, patiriant veiksmų pasekmes ir nereikalaujant išankstinio aplinkos modelio. [6].

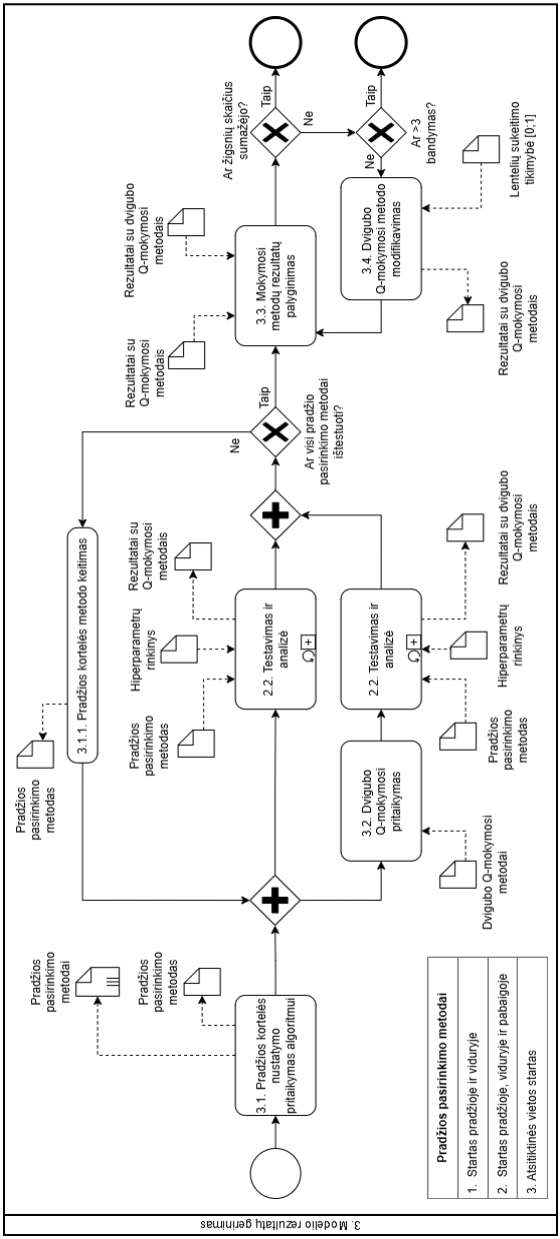
Q-mokymasis nesiremia optimalia strategija ir nėra grindžiamas aplinkos modeliu – jo tikslas yra gauti maksimalią Q-reikšmę (sudėtinį atlygį) [7].

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left[r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t) \right] \quad (1)$$

$Q(s_t, a_t)$ yra Q-reikšmė vykdant veiksmą a_t būsenoje s_t, α yra mokymosi greitis, o $[r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t)]$ nustato skirtumą tarp esamų ir naujų verčių. Q-mokymosi metodai yra tinkami individualizuotos komunikacijos sistemos problemai spręsti, nes optimali Q-reikšmė atnaujinama pagal geriausią galimą veiksmą, užtikrinant didžiausią lankstumą ir tikslumą mokantis iš įvairių veiksmų, nepriklausomai nuo ankstesnių sprendimų.

Siekiant efektyviau surasti teisingą kortelę kaladėje, buvo sudaryta dviejų etapų strategija: pirmiausia nustatyti algoritmo hiperparametrų rinkinį, leidžiantį gauti stabilesnius rezultatus, o tuomet taikyti modelio rezultatų gerinimo metodus. Siekiant pagerinti rezultatus, buvo tiriama pradinės kortelės vietos keitimo įtaka algoritmo veikimui, taip pat lyginama dvigubo Q-mokymosi ir standartinio Q-mokymosi metodų įtaka rezultatams. (1pav.).

Prieš eksperimentų vykdymą „CardFinder+ Sense" algoritmas buvo papildytas funkcija, leidžiančia pasirinkti pradinės kortelės vietą mokymosi ir testavimo procesų metu. Sudaryti trys skirtingi pradinės kortelės vietos parinkimo metodai, kurie valdomi atitinkamu algoritmo hiperparametru. Tuomet Q-mokymosi ir dvigubo Q-mokymosi metodų eksperimentai buvo vykdomi lygiagrečiai. Pirmojo eksperimento metu atliekamas testavimas ir



1 pav. Modelio rezultato gerinimo metodologija.

analizė taikant standartinius Q-mokymosi metodus. Antrojo eksperimento metu tas pats testavimas ir analizė vykdomi su dvigubo Q-mokymosi funkcionalumu papildytu algoritmu. Abiejų eksperimentų metu modelio mokymosi procesai vykdomi su hiperparametrų rinkiniu, gautu modelio rezultatų stabilizavimo etape.

Lyginant rezultatus, jei dvigubo Q-mokymosi metodu apmokyto modelio žingsnių skaičius testavimo metu nėra mažesnis nei standartinio Q-mokymosi modelio, dvigubo Q-mokymosi lentelių keitimo dažnis koreguojamas intervale [0;1]. Tuomet modelis apmokomas iš naujo ir gauti rezultatai lyginami pakartotinai. Jei mažesnis žingsnių skaičius neužfiksuojamas per tris bandymus, laikoma, kad dvigubo Q-mokymosi metodai šiame kontekste reikšmingos įtakos neturi.

Pagrindinis skirtumas tarp Q-mokymosi ir dvigubo Q-mokymosi yra tai, kad dvigubas Q-mokymasis naudoja dvi nepriklausomas Q-reikšmių lenteles (Q^A ir Q^B).

$$Q^A(s,a) \leftarrow Q^A(s,a) + \alpha\left[r + \gamma Q^B\left(s', \arg\max_{a'} Q^A(s',a')\right) - Q^A(s,a)\right] \quad (2)$$

$$Q^B(s,a) \leftarrow Q^B(s,a) + \alpha\left[r + \gamma Q^A\left(s', \arg\max_{a'} Q^B(s',a')\right) - Q^B(s,a)\right] \quad (3)$$

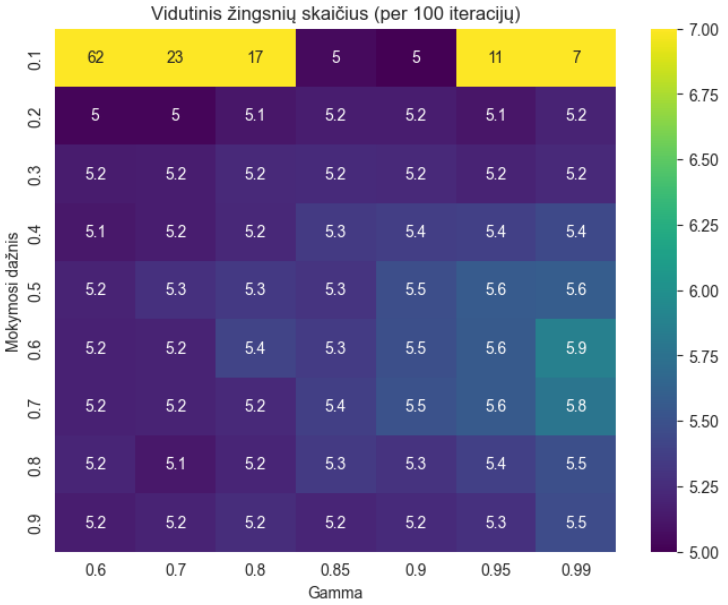
Q-verčių atnaujinimo metu yra atnaujinama lentelė Q^A, remiantis lentelėje Q^B saugomomis Q-reikšmėmis. Šis principas taikomas abiem lentelėms modelio apmokymo metu, o atnaujinama lentelė pasirenkama atsitiktinai su 50 % tikimybe. Kai agentas pasirenka naudoti optimalų veiksmą, jis yra atrenkamas atsižvelgiant į abiejų Q-reikšmių lentelių duomenis.

Siekiant stabilizuoti žingsnių skaičių, per kurį surandama teisinga PECS kortelė kiekvieno vykdomo mokymosi proceso metu, buvo modifikuojami algoritmo hiperparametrai ir sudaromi jų poveikio rezultatui šilumos žemėlapiai. Šie žemėlapiai parodė, su kokiomis hiperparametrų reikšmėmis per 100 modelio apmokymo iteracijų gaunami geriausi rezultatai.

Modelio rezultatų gerinimo eksperimentų metu algoritmui sukurti metodai, leidžiantys nustatyti pradinę kortelę į iš anksto apibrėžtą arba atsitiktinę vietą aplinkoje. Buvo tiriami trys atvejai: pirmajame pradinė kortelė nustatoma atsitiktinai – aplinkos pradžioje arba viduryje; antrajame – pradžioje, viduryje arba pabaigoje; trečiajame – visiškai atsitiktinai bet kurioje aplinkos vietoje. Gauti ir palyginti Q-mokymosi ir dvigubo Q-mokymosi algoritmų rezultatai.

3. Rezultatai

Modelio stabilizavimo eksperimentai buvo vykdomi vienoje iš anksto apibrėžtoje aplinkoje. Eksperimentų metu buvo varijuojama anksčiau minėtais algoritmo hiperparametrais ir stebima jų įtaka modelio mokymosi stabilumui per daugelį apmokymo kartų. Šiam tikslui, algoritmui buvo sukurta automatizuoto testavimo funkcija, kuri nuosekliai tikrina įvairias hiperparametrų reikšmes ir fiksuoja gaunamą vidutinį žingsnių skaičių. Remiantis surinktais duomenimis, buvo sudaryti šilumos žemėlapiai, vaizduojantys hiperparametrų įtaką žingsnių skaičiui (2 pav.), taip siekiant išvengti atsitiktinio hiperparametrų keitimo tikintis rezultato pagerėjimo. Remiantis šio proceso rezultatais, nustatyta, kad didžiausią įtaką mokymosi stabilumui sudaro mokymosi dažnis α ir nuolaidos dydis γ. Taip pat nustatyta, kokiame reikšmių intervale naudinga išlaikyti tyrinėjimo tikimybės mažinimo koeficiento ir minimalios tyrinėjimo tikimybės parametrus. Tyrimo metu sėkmingai pavyko stabilizuoti apmokyto modelio rezultatus ir išvengti reikšmingų nuokrypių pakartotinai apmokant modelį identiškoje aplinkoje.



2 pav. Mokymosi dažnio ir nuolaidos dydžio šilumos žemėlapis.

„CardFinder+ Sense“ rezultatų gerinimo eksperimentams buvo naudojamas stabilizavimo etape gautas hiperparametrų rinkinys, užtikrinantis rezultatų stabilumą. Siekiant įvertinti pradinės kortelės vietos keitimo naudą apmokymo ir testavimo procesų metu, buvo vykdomi trys eksperimentai, kurių metu pradinė kortelė buvo priskiriama skirtingoms pozicijoms.

Pirmojo eksperimento metu, kai kortelė buvo nustatoma aplinkos pradžioje arba viduryje, modelis vidutiniškai surado kortelę per 4,5 žingsnio per 100 iteracijų, o žemiausias pasiektas rezultatas – 3 žingsniai. Antrojo eksperimento metu, kai kortelė buvo nustatoma pradžioje, viduryje arba pabaigoje, rezultatas buvo prastesnis – papildoma galima pradinė būsena prailgino teisingos kortelės paieškos procesą. Geriausius rezultatus parodė trečiasis eksperimentas – atsitiktinė pradinės kortelės pozicija dažnai leido agentui surasti kortelę greičiau nei kitų eksperimentų metu, vidutiniškai per 4 žingsnius per 100 iteracijų.

Siekiant toliau mažinti žingsnių skaičių, „CardFinder+ Sense“ algoritmui pritaikyti dvigubo Q-mokymosi metodai. Atlikus tuos pačius tris eksperimentus su dvigubo Q-mokymosi metodais, pastebėta, kad modelį apmokant per 15 epochų žingsnių skaičius pasižymi reikšmingu nepastovumu. Tikėtina, kad 15 epochų nepakanka abiem Q-reikšmių lentelėms sukaupti pakankamai duomenų, dėl ko modelis veikia netinkamai. Testavimo metu tokie modeliai teisingą kortelę suranda tik dalį atvejų. Padidinus epochų skaičių iki 100, rezultatai tampa panašūs į standartinio Q-mokymosi rezultatus, tačiau jokio pagerėjimo nepastebima. Keičiant mokymosi greičius atskiroms Q-reikšmių lentelėms arba keičiant lentelių atnaujinimo dažnį, rezultatai reikšmingai nesikeičia ir mažesnio žingsnių skaičiaus pasiekti nepavyksta, lyginant su standartinio Q-mokymosi modelio rezultatais.

Tyrimo ir eksperimentų metu sukurtas „CardFinder+ Sense“ algoritmas, gebantis išmokti siūlyti PECS korteles, kurios galėtų būti tinkamos vaikų, turinčių autizmo spektro sutrikimą, komunikacijai. Algoritmas grindžiamas sustiprinto mokymosi principais, taikant Q-mokymosi metodus.

4. Išvados

Išanalizavus sustiprinto mokymosi algoritmo hiperparametrų įtaką ir parinkus tinkamą jų rinkinį, modelio apmokymo rezultatai tampa stabilesni – skirtingų apmokymo procesų metu kortelė surandama per optimalų žingsnių skaičių. Apmokant modelį per 15 epochų, kortelė surandama ne per optimalų žingsnių skaičių tik 1 % atvejų, palyginti su ankstesniais 13 % atvejų.

Pradinės kortelės nustatymas į skirtingas pozicijas mokymosi ir veikimo procesų metu įgalina sustiprinto mokymosi modelį surasti teisingą kortelę per mažesnį žingsnių skaičių, sumažindamas žingsnių skaičių mažiausiai vienu žingsniu.

Atlikus eksperimentus su dvigubo Q-mokymosi metodais nustatyta, kad dvigubo Q-mokymosi metodai nepagerino algoritmo veikimo: apmokant modelį per 15 epochų, teisinga kortelė nerandama arba randama nedaugiau kaip 66 % atvejų, o didinant epochų skaičių rezultatai nesiskiria nuo standartinio Q-mokymosi metodais gautų rezultatų.

Literatūra

[1] A. Lerna, D. Esposito, M. Conson, and A. Massagli. Long-term effects of PECS on social–communicative skills of children with autism spectrum disorders: a follow-up study. International Journal of Language & Communication Disorders, 49:478–485, 2014.

[2] E. Farsari and C. Nitsiou. The implementation of an interactive educational intervention program using the Kinems learning games platform to improve gross motor skills in children with ASD. Preschool and Primary Education, 13:29–49, 2025.

[3] A. G. Barto and T. G. Dietterich. Reinforcement learning and its relationship to supervised learning. In Handbook of Learning and Approximate Dynamic Programming, volume 10. Wiley-IEEE Press, 2004.

[4] S. Milani, Z. Fan, S. Gulati, T. Nguyen, F. Fang, and A. Yadav. Intelligent tutoring strategies for students with autism spectrum disorder: a reinforcement learning approach. In The 2020 CMU Symposium on Artificial Intelligence and Social Good. 2020.

[5] A. Mikulėnas. Mašininiu mokymusi grindžiamas naudotojo sąsajos kūrimas specialiųjų poreikių turintiems naudotojams. Master's thesis. Vilniaus Gedimino Technikos Universitetas, 2022.

[6] C. J. C. H. Watkins and P. Dayan. Q-learning. Machine Learning, 8:279–292, 1992.

[7] H. KungHsiang. Introduction to various reinforcement learning algorithms. Part I (Q-Learning, SARSA, DQN, DDPG). Medium, 2018. URL: https://medium.com/data-science/introduction-to-various-reinforcement-learning-algorithms-i-q-learning-sarsa-dqn-ddpg-72a5e0cb6287.