Skatinamojo mokymosi modelių pritaikymas bendravimui su autizmo spektro sutrikimą turinčiais vaikais

Austėja Tomaševskytė, Asta Slotkienė

Vilniaus Universitetas, Matematikos ir informatikos fakultetas,
Universiteto g. 3, Vilnius, Lietuva
austeja.tomasevskyte@mif.stud.vu.lt

Santrauka. Šiame darbe nagrinėjamos paveikslėlių mainų komunikacijos sistemos (PECS) skaitmenizavimo galimybės, taikant skatinamojo mokymosi metodus. Darbe analizuojami Q šeimos skatinamojo mokymosi algoritmai: Q-mokymosi, Q(λ)-mokymosi ir dvigubo Q-mokymosi – bei jų mokymosi ir veikimo savybės. Tyrimo metu sukurtos modifikuotos FrozenLake eksperimentinės aplinkos su pritaikytais skatinamojo mokymosi metodų algoritmais, turinčios sekančios PECS kortelės parinkimo logiką, ir taikytos skirtingos strategijos algoritmų veikimo optimizavimui. Algoritmų našumas vertintas pagal veiksmų kiekį mokymosi epizodų metu ir po jų bei sukauptą atlygį. Gauti rezultatai rodo, kad skatinamojo mokymosi metodai gali būti efektyviai taikomi PECS sistemų skaitmenizavimui ir adaptyvių komunikacijos sistemų kūrimui.

Raktiniai žodžiai: autizmo spektro sutrikimas, paveikslėlių mainų komunikacijos sistema, skatinamasis mokymasis, Q-mokymasis, Q(λ)-mokymasis, dvigubas Q-mokymasis.

1 Įvadas

Autizmo spektro sutrikimą turintys vaikai dažnai susiduria su bendravimo sunkumais, todėl jiems taikomos alternatyvios ir augmentinės komunikacijos priemonės [1]. Alternatyvioji ir augmentinė komunikacija (AAK) – tai metodų ir priemonių visuma, skirta papildyti arba pakeisti natūralią kalbą, siekiant pagerinti bendravimą asmenims, turintiems kalbos sutrikimų [1].

Viena iš plačiausiai naudojamų alternatyvios komunikacijos sistemų yra paveikslėlių mainų komunikacijos sistema (PECS), kuri leidžia vaikams bendrauti ir išreikšti norus naudojant paveikslėlių korteles.

Tyrimai rodo, kad alternatyvios ir augmentinės komunikacijos priemonės tapo perspektyviomis priemonėmis, padedančiomis pagerinti bendravimą tarp asmenų, kurių verbalinė kalba yra ribota arba jos visai nėra [2]. Nors

DOI: https://doi.org/10.15388/LMITT.2026.31

PECS sistema jau yra skaitmenizuota, jos realizacija neturi veikimo logikos ar adaptacijos galimybių pagal individualius vaiko poreikius. Dėl šių priežasčių yra aktualu ieškoti sprendimų, kaip šias sistemas realizuoti su dirbtinio intelekto modeliais, tokiais kaip skatinamojo mokymosi metodai, kurie leidžia modeliuoti savarankiškai besimokančią interaktyvią sistemą. Skatinamasis mokymasis – tai mašininio mokymosi metodas, kuriame agentas mokosi sąveikaudamas su aplinka ir gaudamas atlygį už atliktus veiksmus. Mokymosi tikslas yra parinkti tokią veiksmų seką, kuri maksimalizuoja sukauptą atlygį ilguoju laikotarpiu [5].

Darbo tikslas - ištirti skatinamojo mokymosi modelių Q mokymosi šeimos algoritmus, siekiant pagerinti alternatyvią ir augmentinę komunikaciją, taikant skaitmenizuotas PECS korteles.

2 Tyrimo metodologija

Skatinamojo mokymosi uždaviniai dažnai aprašomi naudojant Markovo sprendimų procesus, kuriuos sudaro būsenų aibė *S*, veiksmų aibė *A*, perėjimo tikimybių lentelė ir atlygio funkcija. Kiekvienu laiko momentu agentas, būdamas būsenoje, pasirenka veiksmą pagal tam tikrą strategiją, pereina į naują būseną ir gauna atlygį. Mokymosi tikslas yra rasti optimalią strategiją, kuri maksimalizuoja diskontuotą sukauptą atlygį [3].

Šiame tyrime analizuojami be modelio skatinamojo mokymosi algoritmai, priklausantys Q-mokymosi šeimai: Q-mokymosi, Q(λ)-mokymosi ir dvigubo Q-mokymosi algoritmai.

Q-mokymosi algoritmas priklauso algoritmo be modelio klasės algoritmui. Tai reiškia, kad mokymasis vyksta klaidų ir bandymų metodu ir tyrinėjimas nevykdomas pagal strategiją, bet naudojamas koks nors ne strategija veikiantis metodas, šio darbo atveju naudojamas ϵ-godus. Algoritmas naudoja Q-reikšmes kiekvienai būsenos-veiksmo porai [4].

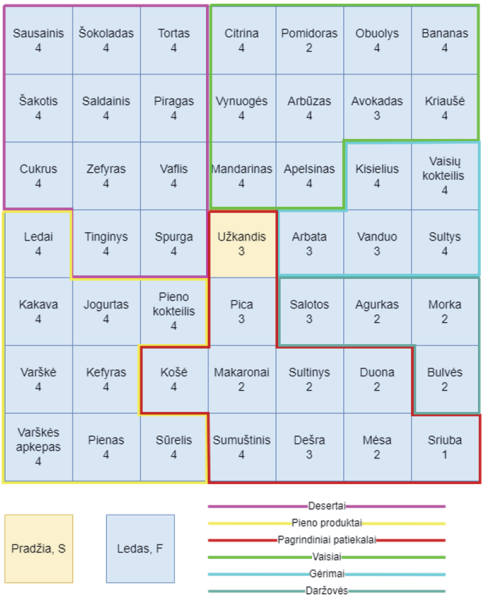
Taikant ϵ-godų metodą, veiksmas, kurio numatoma vertė yra didžiausia, vadinamas godžiu veiksmu, o agentas paprastai išnaudoja savo dabartines žinias pasirinkdamas godų veiksmą. Tačiau agentas taip pat turi ϵ tikimybę tyrinėti atsitiktinai pasirinkdamas vieną iš negodžių veiksmų [6].

Q(λ)-mokymosi algoritmas yra Q-mokymosi algoritmo išplėtimas, kuriame naudojami tinkamumo sekimo mechanizmas. Šis metodas leidžia greičiau perduoti informaciją apie gautą atlygį ankstesnėms būsenoms, todėl mokymasis vyksta greičiau nei naudojant paprastą Q-mokymosi algoritmą. Šiame metode parametras $\lambda \in [0,1]$ nusako, kiek stipriai atlygis yra paskirs-

tomas ankstesnėms būsenoms. Kuo λ reikšmė didesnė, tuo daugiau ankstesnių būsenų įtraukiama į mokymosi procesą, todėl agentas gali greičiau išmokti optimalią strategiją. Šio darbo metu buvo naudojama λ = 0,9.

Dvigubo Q-mokymosi (DQL) algoritmas naudojamas siekiant sumažinti Q reikšmių pervertinimą, kuris atsiranda dėl maksimizavimo operacijos standartiniame Q-mokymosi algoritme. Šiame metode naudojamos dvi Q lentelės, kurios atnaujinamos pakaitomis, todėl gaunami stabilesni rezultatai, o Q lentelė yra struktūra, kurioje saugomos visų galimų būsenos ir veiksmo porų vertės, t. y. strategijos išnaudojimo metu agentas renkasi veiksmą ir būseną su didžiausiu išmoktu įverčiu.

Eksperimentams atlikti buvo naudojama modifikuota *FrozenLake-v1* aplinka iš *OpenAI Gym* bibliotekos, pritaikyta PECS kortelių parinkimo logikai modeliuoti (žiūrėti 1 pav.). Ši aplinka buvo pritaikyta taip, kad kiekviena būsena atitiko tam tikrą maisto produkto PECS kortelę, kurios buvo suskirstytos į platesnes maisto kategorijas. Aplinkoje agentas turi pasirinkti veiksmų



1 pav. FrozenLake modifikuota aplinka su PECS kortelių išdėstymu.

seką, atitinkančią PECS kortelių pasirinkimo seką, kad pasiektų tikslinę būseną. Agentui pasiekus tikslinę būseną skiriamas teigiamas atlygis, o pasirinkus netinkamą veiksmą skiriamas mažesnis arba neigiamas atlygis. Tokiu būdu agentas mokosi parinkti optimalias kortelių sekas. 1 paveikslėlyje vaizduojami kiekvienos būsenos atitinkamos kortelės įvertinimai. Eksperimentų metu agento tikslas buvo rasti „Sriuba“ kortelės būseną ir kiekvienoje būsenoje gaudavo įvestį nuo 1 iki 4, kur 1 reiškė teisingą kortelę, o 4 reiškė visiškai netinkamą kortelę.

Siekiant pagerinti algoritmų veikimą, buvo taikomos kelios strategijos:

1. **Mokymosi greičio α ir diskonto faktoriaus γ reikšmių parinkimo strategija.**

Strategijos tikslas - ištirti α ir γ parametrų įtaką skatinamojo mokymosi algoritmų konvergencijos spartai ir stabilumui. Skirtingos parametrų reikšmės taikomos siekiant nustatyti, kuri parametrų reikšmių kombinacija leidžia agentui efektyviai išmokti optimalią strategiją su mažiausiu mokymosi epizodų kiekiu.

Diskonto faktorius $\gamma \in [0,1]$ nusako būsimo atlygio svarbą lyginant su momentiniu atlygiu. Mažesnė γ reikšmė skatina agentą orientuotis į trumpalaikį atlygį, o didesnė – į ilgalaikį rezultatą. Todėl šio parametro parinkimas turi tiesioginę įtaką mokymosi strategijai ir konvergencijos greičiui [3].

Strategijoje naudotos parametrų reikšmės ir eksperimentų numeriai pavaizduoti 1 lentelėje, kur R yra atitinkamas atlygis įvesčiai nuo 1 iki 4, o ϵ mažinimo koeficientas reiškia, kad agentui buvo skirta $1/\epsilon$ aplinkos tyrinėjimo epizodų.

1 lentelė. Mokymosi greičio α ir diskonto faktoriaus γ reikšmių parinkimo strategijos metrikos ir eksperimentų numeriai.

	α ir γ reikšmių strategija			
	Nr. 1	**Nr. 2**	**Nr. 3**	**Nr. 4**
α	0,9	0,9	0,3	0,3
γ	0,9	0,9	0,6	0,6
ϵ	0,1	0,0625	0,1	0,0625
R	{15; 5; 0; -5}			

2. **Atlygio reikšmių strategija.**

Strategijos tikslas - ištirti skirtingų atlygio reikšmių įtaką agento mokymosi greičiui. Keičiant atlygio reikšmes siekiama nustatyti, su kokia atlygio

kombinacija agentas greičiau išmoksta strategiją, mažinant mokymosi epizodų skaičių.

Strategijos atlygio reikšmės ir eksperimentų numeriai pavaizduoti 2 lentelėje.

2 lentelė. Atlygio reikšmių strategijos metrika.

Atlygio reikšmių strategija			
R1	**R2**	**R3**	**R4**
{15; 5; 0; -5}	{10; 0; -5; -10}	{5; -5; -10; -15}	{20; 10; 5; 0}

3. Lygiagretaus algoritmo strategija.

Strategijos tikslas - pritaikyti lygiagretų algoritmų veikimą, siekiant sumažinti veiksmų kiekį iki norimos kortelės, siūlant po dvi korteles vienu metu.

Lygiagretaus algoritmo strategijoje vienu metu veikia du agentai, kurie siūlo skirtingas korteles iš skirtingų pradinių būsenų. Vartotojas pasirenka labiau tinkamą kortelę, kuri gauna teigiamą atlygį, o kita – neigiamą arba mažesnį atlygį. Tokiu būdu abu agentai mokosi lygiagrečiai, o veiksmų pasirinkimas grindžiamas vartotojo pasirinkimu tarp dviejų pateiktų alternatyvų.

4. Kortelių kategorizavimo strategija.

Strategijos tikslas - pritaikyti visoms kortelėms kategorijas, siekiant sumažinti veiksmų kiekį iki norimos kortelės tyrinėjimo epizodų metu.

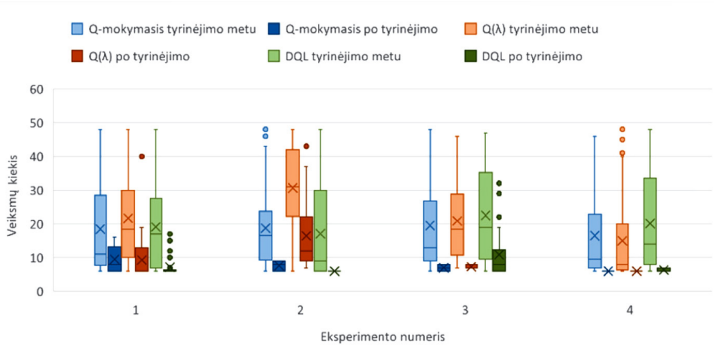
Eksperimentų metu kiekvienas algoritmas buvo vykdomas su skirtingomis strategijomis, o gauti rezultatai buvo lyginami tarpusavyje, siekiant nustatyti, kuris algoritmas ir kuri strategija leidžia pasiekti geriausius rezultatus PECS kortelių parinkimo uždavinyje.

3 Rezultatai

Šiame skyriuje pateikiami eksperimentų rezultatai, gauti taikant skirtingas strategijas Q-mokymosi, Q(λ)-mokymosi ir dvigubo Q-mokymosi algoritmams modifikuotoje *FrozenLake* aplinkoje. Algoritmų efektyvumas buvo vertinamas pagal veiksmų skaičių ir sukauptą atlygį mokymosi epizodų metu ir po jų.

Pirmiausia buvo tiriama mokymosi greičio α ir diskonto faktoriaus γ įtaka algoritmų veikimui. Eksperimentų rezultatai parodė, kad tinkamai parinktos parametrų reikšmės turi didelę įtaką mokymosi stabilumui ir konvergencijos greičiui. Nustatyta, kad Q-mokymosi ir Q(λ)-mokymosi algoritmai sta-

biliausius rezultatus pasiekė su parametrais α = 0,3 ir γ = 0,6, o dvigubo Q-mokymosi algoritmas geriausius rezultatus pasiekė su parametrais α = 0,9 ir γ = 0,9 (žiūrėti 2 pav.).



2 pav. Q-mokymosi, Q(λ)-mokymosi ir DQL algoritmų eksperimentų palyginimas.

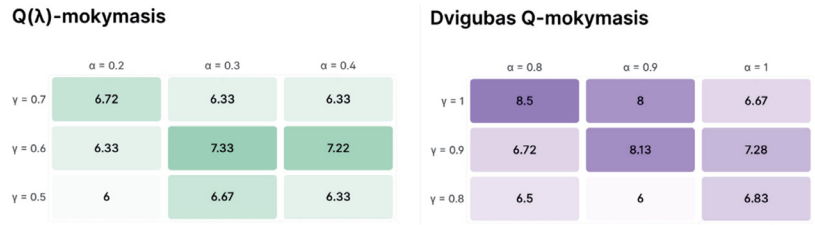
Toliau buvo tiriama, kiek mažiausiai tyrinėjimui skirtų epizodų reikia, kad algoritmai išmoktų optimalią strategiją. Rezultatai parodė, kad Q-mokymosi algoritmui reikia mažiausiai 14 tyrinėjimo epizodų, kad būtų pasiekti stabilūs rezultatai, o Q(λ)-mokymosi ir dvigubo Q-mokymosi algoritmams pakanka 12 epizodų, tačiau sumažinus epizodų skaičių iki 10 pradeda pasireikšti rezultatų nestabilumas (žiūrėti 3 lentelę).

3 lentelė. Metodų rezultatų palyginimas su skirtingomis tyrinėjimų reikšmėmis.

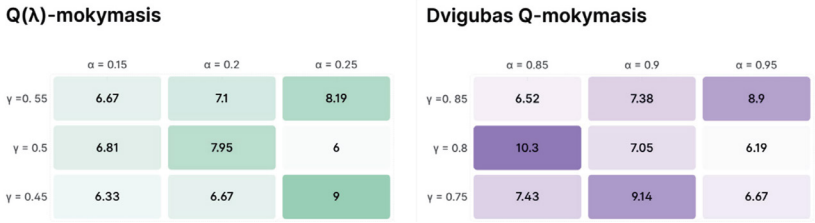
	Tyrinėjimo metu		**Po tyrinėjimo**	
	Vid. veiksmų kiekis	**Vid. atlygis**	**Vid. veiksmų kiekis**	**Vid. atlygis**
tyrinėjimo epizodai = 16				
Q-mokymasis	16,625	-6,979	6	33,333
Q(λ)-mokymasis	15	-1,979	6	26,667
DQL	16,354	-7,708	6	30
tyrinėjimo epizodai = 14				
Q-mokymasis	18,214	-11,548	6	35
Q(λ)-mokymasis	16,405	-9,524	6	30
DQL	21,571	-25,357	6	29,444

	Tyrinėjimo metu		Po tyrinėjimo	
	Vid. veiksmų kiekis	**Vid. atlygis**	**Vid. veiksmų kiekis**	**Vid. atlygis**
tyrinėjimo epizodai = 12				
Q-mokymasis	21,389	-23,75	7	35
Q(λ)-mokymasis	18,056	-9,167	6	35
DQL	17,75	-14,167	6	30,278
tyrinėjimo epizodai = 10				
Q-mokymasis	14,967	-7,667	7,356	21,556
Q(λ)-mokymasis	22,333	-27,333	7,333	31,667
DQL	25,867	-38,333	8,133	21,333

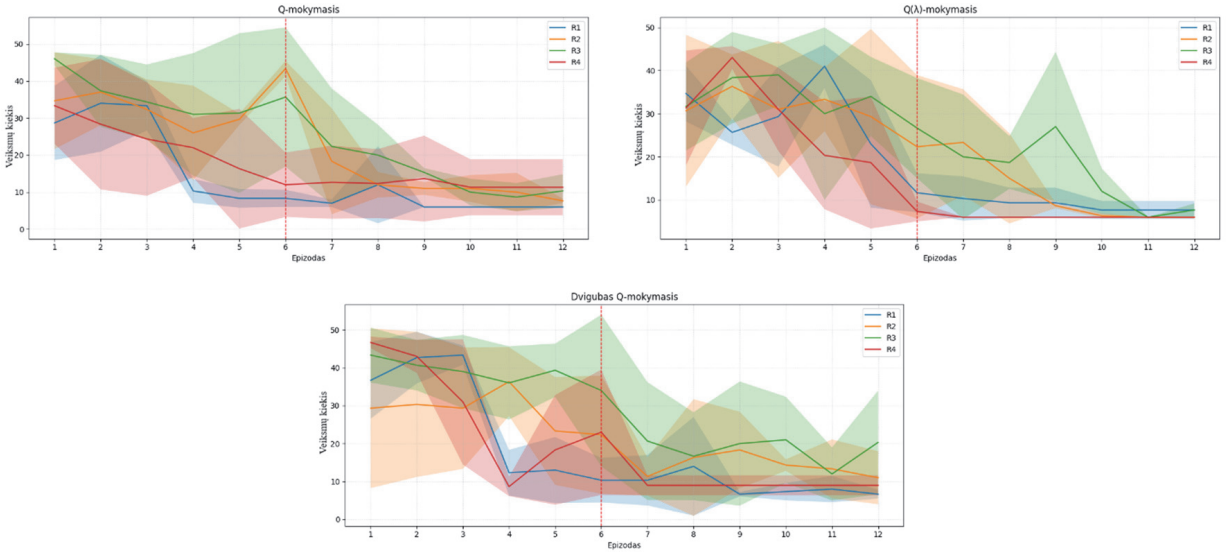
Dėl gautų rezultatų, toliau buvo atlikti eksperimentai Q(λ)-mokymosi ir DQL algoritmų aplinkose, siekiant sumažinti tyrinėjimo epizodus ir rasti tinkamas parametrų reikšmes. Eksperimentais nustatyta, kad DQL algoritmui užtenka 10 tyrinėjimo epizodų, o optimalūs parametrai α = 0.95 ir γ = 0.8, o Q(λ)-mokymosi algoritmo agentui užteko 8 tyrinėjimo epizodų, kai α = 0.25 ir γ = 0.5 (žiūrėti 3 ir 4 pav.).



3 pav. Q(λ)-mokymosi ir DQL parametrų reikšmių įtakos rezultatai (10 tyrinėjimo epizodų).



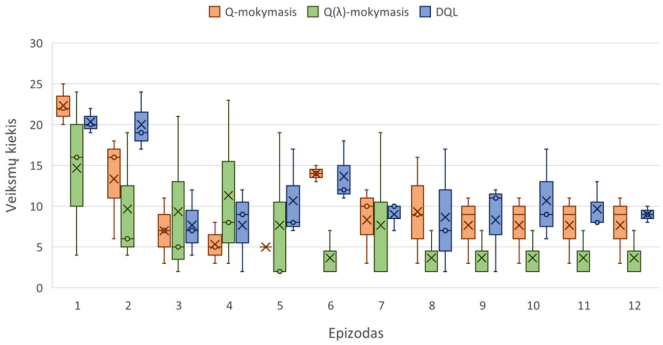
4 pav. Q(λ)-mokymosi ir DQL parametrų reikšmių įtakos rezultatai (8 tyrinėjimo epizodai).



5 pav. Q-mokymosi, Q(λ)-mokymosi ir DQL atlygių reikšmių įtakos rezultatai.

Atlikus eksperimentus su optimaliomis parametrų reikšmėmis, buvo analizuojama atlygio funkcijos įtaka mokymosi rezultatams su 6 tyrinėjimo epizodais. Rezultatai parodė, kad optimaliausią strategiją agentas pasiekė Q-mokymosi metodo aplinkoje su R1 atlygių kombinacija, tačiau šie rezultatai nebuvo stabilūs. Stabiliausi rezultatai buvo gauti Q(λ)-mokymosi aplinkoje, kur visų eksperimentų metu norima kortelė buvo pasiekta per 6 veiksmus, naudojant R4 atlygių kombinaciją. Dvigubo Q-mokymosi algoritmas taip pat parodė gerus rezultatus, tačiau kai kuriais atvejais rezultatai buvo mažiau stabilūs (žiūrėti 5 pav.).

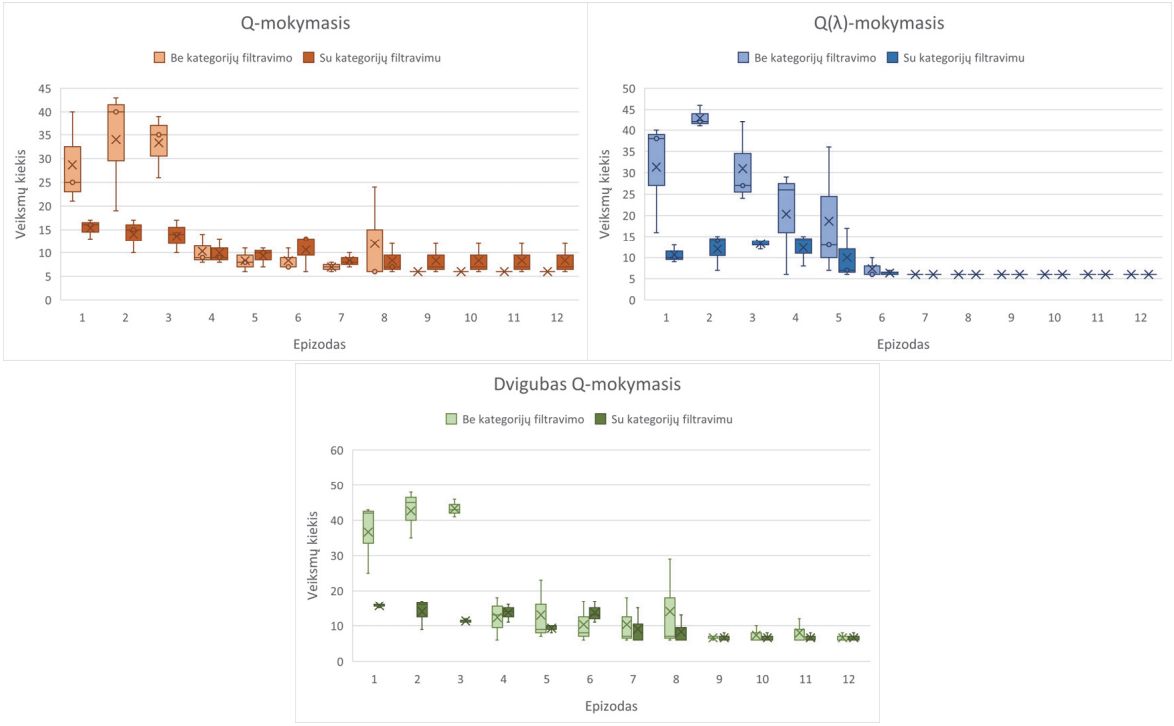
Lygiagretaus algoritmo strategijos rezultatai parodė, kad lygiagrečiai vykdant algoritmą galima ženkliai sumažinti veiksmų kiekį iki norimos kortelės. Eksperimentų metu nustatyta, kad greičiausiai norimą kortelę buvo galima surasti per 2 veiksmus vietoje įprastų 6 veiksmų, tačiau rezultatų stabilumui užtikrinti reikalingas didesnis tyrinėjimo epizodų skaičius (žiūrėti 6 pav.).



6 pav. Lygiagretaus algoritmo veiksmų kiekio skirtingose aplinkose palyginimas.

Kortelių kategorizavimo strategijos rezultatai parodė, kad būsenų sugrupavimas į kategorijas leidžia sumažinti veiksmų skaičių tyrinėjimo metu ir pagreitina mokymosi procesą, nes sumažėja būsenų erdvė ir agentas greičiau išmoksta optimalią strategiją. Rezultatai taip pat parodė, kad ši strategija pagerino algoritmų stabilumą skirtingų epizodų metu (žiūrėti 7 pav.).

Apibendrinant visų strategijų rezultatus galima teigti, kad geriausi rezultatai buvo gauti taikant Q(λ)-mokymosi algoritmą su optimaliomis parametrų reikšmėmis ir pritaikius atlygio funkcijos bei kortelių kategorizavimo



7 pav. Q-mokymosi, Q(λ)-mokymosi ir DQL kortelių kategorizavimo strategijos rezultatai.

strategijas. Šis algoritmas pasiekė stabiliausius rezultatus ir greičiausiai konvergavo, lyginant su Q-mokymosi ir dvigubo Q-mokymosi algoritmais.

4 Išvados

Šiame darbe buvo analizuoti Q šeimos skatinamojo mokymosi algoritmai – Q-mokymosi, Q(λ)-mokymosi ir dvigubo Q-mokymosi – bei jų pritaikymas PECS kortelių skaitmenizacijos uždaviniui. Atlikta analizė parodė, kad šie algoritmai yra tinkami spręsti sprendimų priėmimo uždavinius nežinomoje aplinkoje su nežinomais atlygiais, nes agentas gali savarankiškai išmokti optimalią veiksmų strategiją, remdamasis sąveika su aplinka.

Tyrimo metu buvo sukurta modifikuota *FrozenLake* aplinka, pritaikyta PECS kortelių parinkimo logikai modeliuoti, ir atlikti eksperimentai su skirtingais algoritmais bei strategijomis. Eksperimentų rezultatai parodė, kad algoritmų veikimas labai priklauso nuo parinktų parametrų reikšmių, atlygio funkcijos ir būsenų erdvės mažinimo strategijų.

Nustatyta, kad geriausi rezultatai buvo gauti taikant Q(λ)-mokymosi algoritmą su parinktomis optimaliomis parametrų ir atlygio reikšmėmis, nes šis algoritmas greičiausiai konvergavo ir stabiliausiai pasiekė sprendinį. Taip pat geri rezultatai buvo gauti naudojant dvigubo Q-mokymosi algoritmą, kuris dėl dviejų Q lentelių naudojimo pasižymėjo stabilesniu mokymusi tam tikrose strategijose.

Literatūra

[1] Edgar, T. C., Schlosser, R., & Koulc, R. (2024). Effects of an augmentative and alternative communication intervention package on socio-communicative behaviors between minimally speaking autistic children and their peers. American Journal of Speech-Language Pathology, 33, 1619–1638.

[2] Mavritsakis, D. (2024). Augmentative and alternative communication in autism spectrum disorder: transitioning from letter board to iPad – a case study. Frontiers in Psychiatry, 15, 1–10.

[3] Wiering, M., & Schmidhuber, J. (1998). Speeding up Q(λ)-learning. In Proceedings of the 10th European Conference on Machine Learning (ECML), 1–13.

[4] Meng, Y., Kuppannagari, S., Rajat, R., Srivastava, A., Kannan, R., & Prasanna, V. (2020). Qtaccel: a generic FPGA-based design for Q-table based reinforcement learning accelerators. In IEEE International Parallel and Distributed Processing Symposium Workshops, 107–114.

[5] Sutton, R. S., & Barto, A. G. (2014). Reinforcement learning: an introduction. MIT Press.

[6] Altuntaş, N., İmal, E., Emanet, N., & Öztürk, C. N. (2016). Reinforcement learning-based mobile robot navigation. Turkish Journal of Electrical Engineering and Computer Sciences, 24, 1747–1767.