

Ն.Հ. ՆԵՐՍԻՍՅԱՆ

ԱՆՕԴԱԶՈՒ ԹՈՉՈՂ ՍԱՐՔԻ ՆԵՅՐՈՆԱՅԻՆ ԿԱՐԳԱՎՈՐԻՉԻ ՆԱԽԱԳԾՈՒՄԸ ԱՄՐԱՊԵՂԻՄԱՆ ՈՒՍՈՒՑՄԱՆ ԿԻՐԱՌՄԱՄԲ

Բազմառոտոր ԱԹՍ-ի համար նախագծվել է ներդրանային կարգավորիչ, որի մեքենայական ուսուցումն իրականացվել է ամրապնդման ուսուցմամբ: Matlab Simulink միջավայրում նախագծվել է ԱԹՍ-ի կառավարման համակարգ, որում ներառվել է կարգավորիչը: Իրականացվել է ստացված համակարգի հետազոտություն:

Առանցքային բառեր. անօդաչու թռչող սարք, կառավարման համակարգ, արհեստական բանականություն, կարգավորիչ, ամրապնդման ուսուցում:

Ներածություն: Անօդաչու թռչող սարքերը ներկայումս լայն տարածում են գտել իրենց ցածր գնի, բարձր մանևրելիության, օդաչուի բացակայության և համեմատաբար պարզ կառուցվածքի շնորհիվ: Նրանց լայն կիրառությունը խթանում է՝ ստեղծելու նոր՝ ավելի կատարելագործված կառավարման համակարգեր, որպեսզի ապահովվի կայուն թռիչքը՝ անկախ հնարավոր փոփոխվող արտաքին պայմաններից: Աշխատանքում խնդիր է դրվել նախագծել ԱԹՍ-ի համար ներդրանային կարգավորիչ, որի հիմքում ընկած է արհեստական բանականությունը, որի կարգաբերումը պետք է իրականացվի մեքենայական ուսուցման տեսակ հանդիսացող ամրապնդման ուսուցման մեթոդներով:

ԱԹՍ-ի մաթեմատիկական մոդելը: Ներկայացնենք բազմառոտոր ԱԹՍ-ի դինամիկ հավասարումների համակարգը, որտեղ ϕ, θ, ψ անկյունները էլիքրի անկյուններն են՝ համապատասխանաբար կողաթեքման, երկայնաճոճման և ընթացաշեղման անկյունները, x, y, z -ը համապատասխան առանցքներով տեղափոխություններն են, իսկ m -ը՝ ԱԹՍ-ի զանգվածը [1]:

$$\begin{cases} m \frac{d^2x}{dt^2} = u_z(\cos\phi\sin\theta + \cos\theta\sin\phi\sin\psi) + \beta_x - k_x \frac{dx}{dt}, \\ m \frac{d^2y}{dt^2} = u_z(s\psi s\theta - c\psi c\theta s\phi) + \beta_y - k_y \frac{dy}{dt}, \\ m \frac{d^2z}{dt^2} = u_z(\cos\phi\cos\theta) - mg + \beta_z - k_z \frac{dz}{dt} : \end{cases} \quad (1)$$

$$\begin{cases} I_x \frac{d^2\theta}{dt^2} = u_\theta + \gamma_\theta + (I_y - I_z) \frac{d\phi}{dt} \frac{d\psi}{dt} - J_T \frac{d\phi}{dt} \Omega - k_\theta \frac{d\theta}{dt}, \\ I_y \frac{d^2\phi}{dt^2} = u_\phi + \gamma_\phi + (I_z - I_x) \frac{d\theta}{dt} \frac{d\psi}{dt} - J_T \frac{d\theta}{dt} \Omega - k_\phi \frac{d\phi}{dt}, \\ I_z \frac{d^2\psi}{dt^2} = u_\psi + \gamma_\psi + (I_y - I_x) \frac{d\theta}{dt} \frac{d\phi}{dt} - k_\psi \frac{d\psi}{dt}, \end{cases} \quad (2)$$

որտեղ (1)-ը ԱԹՍ-ի զծային շարժումների համակարգն է, իսկ (2)-ը՝ պտտական շարժումների համակարգը: (1)-ում և (2)-ում u_z -ը ԱԹՍ-ի շարժիչների միջոցով

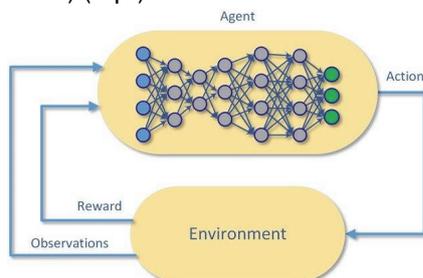
ստեղծված գումարային քարշի ուժն է, u_θ, u_ϕ, u_ψ -ը մուտքային պտտող մոմենտները, I_x, I_y, I_z -ը՝ իրանին կոշտ ամրացված կոորդինատային համակարգի երեք առանցքներով ինեցիայի մոմենտները: $k_x, k_y, k_z, k_\theta, k_\phi, k_\psi$ -ն՝ ճակատային դիմադրության ուժերը, J_T -ն՝ շարժիչների իներցիայի մոմենտները, Ω -ն՝ պտուտակների շարժման ընդհանուր արդյունարար արագությունը, $\gamma_x, \gamma_y, \gamma_z, \gamma_\theta, \gamma_\phi, \gamma_\psi$ -ը՝ արտաքին վրդովող ուժերը և մոմենտները:

Մեքենայական ուսուցում: Ներկայումս մեքենայական ուսուցման երկու ամենակիրառվող մեթոդներն են խոր ուսուցումը և ամրապնդման ուսուցումը [2]: Ի տարբերություն խոր ուսուցման՝ ամրապնդման ուսուցման դեպքում ուսուցման համար անհրաժեշտ տվյալները ստացվում են տեղում՝ համակարգի ուսուցման ընթացքում, սխալի և կրկնման տրամաբանությամբ: Համակարգը ուսուցման ամեն քայլում վերցնում է նախորդ քայլերի արդյունքում ստացված տվյալները և շարունակում է սովորել արդեն այդ տվյալների հիման վրա:

ԱԹՍ-ների կամ ռոբոտների ներդրումային կարգավորիչների կարգաբերումը հարմար է իրականացնել ամրապնդման ուսուցման մեթոդով, քանի որ դժվար է նոր նախագծվող ԱԹՍ-ի դեպքում ունենալ նախօրոք հավաքագրված տվյալների բազա:

Ամրապնդման ուսուցում: Ներկայումս գործում են ամրապնդման ուսուցման մի քանի ալգորիթմներ, դրանցից են՝ արժեքի ֆունկցիան, Q-ուսուցումը, գործող/քննադատ ցանցերը: Վերջինս դիտարկվել է այս աշխատանքում, որի հիման վրա նախագծվել է ներդրումային կարգավորիչը:

Ամրապնդման ուսուցման մեջ ներդրումային կարգավորիչը, որը ներկայացնում է ներդրումային ցանց, կոչվում է գործակալ՝ A (agent): Համակարգը, որի վրա ներդրվում է կարգավորիչը և ենթարկվում է ուսուցման, կոչվում է միջավայր (environment): Ուսուցման ամեն իտերացիայում գործակալն իրականացնում է մեկ գործողություն և որպես հետադարձ կապ ստանում է գործողության արդյունքում ստացված համակարգի նոր վիճակը՝ S (state observation) և որոշակի մրցանակային միավորներ՝ R (reward) (նկ.1):



Նկ. 1. Ամրապնդման ուսուցման սխեման

Այդ ինֆորմացիան նախորդ քայլերում ստացվածների հետ միասին այն օգտագործելու է, որպեսզի լուծի օպտիմալության խնդիր: Այս քայլերի հաջորդականությունը կոչվում է Մարկովյան որոշումների պրոցես (MDP) [3]:

Սահմանենք մրցանակային միավորները: Նախ սահմանենք ակնկալվող մրցանակները որպես ապագա մրցանակների գումար՝

$$G_t = R_{t+1} + R_{t+2} + \dots + R_T: \quad (3)$$

Եթե մեկ ուսուցման իտերացիան անվանենք էպիզոդ կամ դրվագ, ապա՝

$$t = 1, 2 \dots T$$

կլինեն այդ դրվագի ընթացքում քայլերը կամ գործողությունները, իսկ T -ն կլինի դրվագի վերջին գործողությունը կամ ավարտը: Գործակալի հիմնական նպատակն է մաքսիմալացնել ակնկալվող մրցանակը:

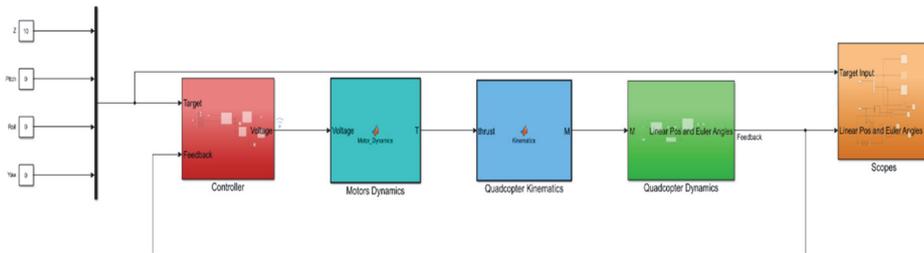
Սահմանենք Բելմանի օպտիմալության հավասարումը ամրապնդման ուսուցման պայմանների դեպքում.

$$q_*(s, a) = E[R_{t+1} + \gamma * \max\{q_*(s', a')\}] \quad (4)$$

Ըստ Բելմանի հավասարման՝ ժամանակի t պահին ցանկացած վիճակ-գործողություն զույգի համար (s, a) ակնկալվող մրցանակը հավասար է այդ քայլում ակնկալվող մրցանակի և հաջորդ քայլում որևէ (s', a') զույգի համար զեղչված մաքսիմում հնարավոր ակնկալվող մրցանակի գումարի մաթեմատիկական սպասմանը: $q_*(s, a)$ կարճ անվանվում է Q -արժեք: Այսպիսով, կարող ենք ասել, որ ցանկացած վիճակ-գործողություն զույգ օպտիմալ է, եթե այն բավարարում է Բելմանի հավասարմանը [4,5]:

Ցանկացած s վիճակի համար օպտիմալ $q(s, a)$ արժեքի հաշվարկը կիրականացնի խոր նեյրոնային ցանցը: Այսպիսով, ցանցի մուտքային նեյրոններին կտրվի համակարգի առկա վիճակը, իսկ ելքում կստացվեն այդ վիճակում բոլոր հնարավոր գործողությունների համար $q(s, a)$ արժեքները: Այնուհետև կհաշվարկվի կորստի ֆունկցիան՝ ինչպես ստանդարտ Q ուսուցման դեպքում: Հաջորդիվ ուսուցման գործընթացն իրականացվում է ինչպես խոր ուսուցման դեպքում, այսինքն՝ օպտիմալության թվային մեթոդներով քայլ առ քայլ կփնտրվեն նեյրոնային ցանցի օպտիմալ քառային գործակիցները, ինչի շնորհիվ ժամանակի ընթացքում նեյրոնային ցանցի ելքում ստացվող $q(s, a)$ արժեքները կմոտենան իրենց օպտիմալ $q_*(s, a)$ արժեքին:

ԱԹՍ-ի նախագծված կառավարման համակարգը: Ներկայացնենք MATLAB SIMULINK միջավայրում ԱԹՍ-ի համար նախագծված կառավարման համակարգը, որը ներառում է նեյրոնային կարգավորիչ (նկ. 2):



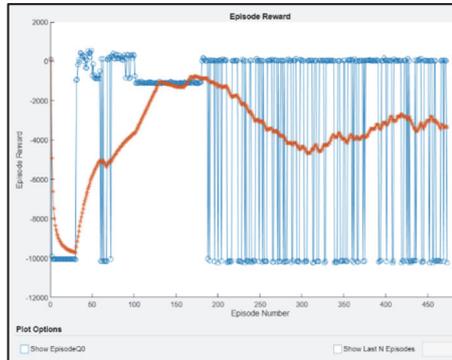
Նկ. 2. Նախագծված կառավարման համակարգը SIMULINK միջավայրում

Կառավարման համակարգը պարունակում է ԱԹՍ-ի մաթեմատիկական մոդելը, որը ներառում է շարժիչների դինամիկան, ԱԹՍ-ի ընդհանուր կինեմատիկան և դինամիկ հավասարումների համակարգը: Այս բլոկները անհրաժեշտ են ԱԹՍ-ի կառավարման համակարգը մոդելավորելու համար: Համակարգը ներկայացնում է բացասական հետադարձ կապով կառավարման համակարգ, որտեղ որպես կարգավորիչի դեր իրականացնում է նեյրոնային կարգավորիչը: Այս համակարգում նեյրոնային կարգավորիչը կիրառվել է ԱԹՍ-ի Z կամ բարձրության առանցքի վրա:

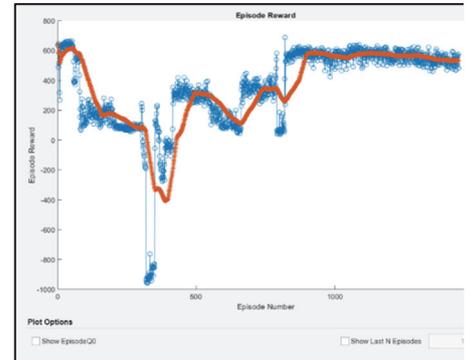
Նեյրոնային կարգավորիչի բլոկը ներառում է նեյրոնային ցանցը, որն ուսուցանվում է ամրապնդման ուսուցման մեթոդներով: Ամրապնդման ուսուցման բլոկի միջոցով իրականացվում է բուն ուսուցումը, ինչպես նաև այդ բլոկի միջոցով տեղի է ունենում մրցանակային միավորների հավաքագրումը, և որպես հետադարձ կապ՝ համակարգի նոր վիճակի ստացումը:

Հետազոտության արդյունքները: Աշխատանքում խնդիր է դրված, որ ամրապնդման ուսուցման շնորհիվ՝ բազմառոտոր ԱԹՍ-ն սովորի թռչել, այսինքն՝ ինքնուրույն գետնից բարձրանա առաջադրված բարձրության վրա և պահի տրված բարձրությունը, լինելով կայուն և չունենալով գերկարգավորում ու ստատիկ սխալանք:

Նկ.3-ում կապույտ կետերով պատկերված են ուսուցման յուրաքանչյուր էպիզոդում ստացված գումարային մրցանակային միավորները, իսկ նարնջագույնով՝ վերջին 100 էպիզոդների միջին մրցանակային միավորը: Բացասական մրցանակային միավորները նշանակում են, որ ԱԹՍ-ն վայր է ընկել ուսուցման ընթացքում կամ ունեցել է մեծ սխալանք: Դրական արժեքներին ձգտող միավորները ստացվում են միայն փոքր սխալանքների դեպքում, այլ կերպ ասած՝ դրական միավորները ստացվում են, երբ ԱԹՍ-ի բարձրությունը մոտ է առաջադրված բարձրությանը:



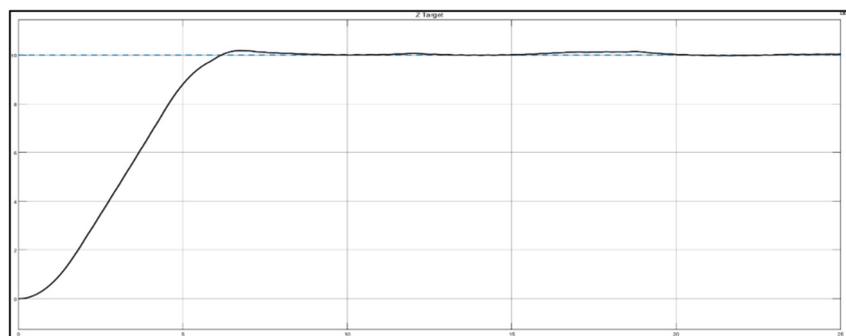
Նկ. 3 ա) Ամրապնդման ուսուցման սկզբնական արդյունքը



Նկ. 3 բ) Ամրապնդման ուսուցման վերջնական արդյունքը

Ինչպես երևում է նկ. 3 ա-ում, նախնական էպիզոդներում ստացվում էին հիմնականում բացասական մրցանակներ, ինչը բնական է, քանի որ նեյրոնային կարգավորիչի գործակիցները դեռևս մոտ չեն օպտիմալ արժեքներին: Ուսուցման 150-րդ և հետագա էպիզոդներում միջին մրցանակային միավորները տատանվում են -3000 արժեքի շուրջ, ինչը դեռևս չի բավարարում ակնկալվող արդյունքին:

Հետագա ուսուցման արդյունքում, ինչպես երևում է նկ. 3բ-ում, միջին մրցանակային միավորն արդեն ձգտում է 500 և ավելի դրական արժեքներին, որոնք ստացվում են այն դեպքում, երբ բարձրության սխալանքը հասցվում է նվազագույնի: Նկ. 3-ում պատկերված է 1500 էպիզոդ ուսուցանված նեյրոնային կարգավորիչով ԱԹՍ-ի բարձրության գրաֆիկը: Ինչպես երևում է նկ.4-ում, համակարգում առկա է 0.1% գերկարգավորում, և առկա չէ ստատիկ սխալանք: ԱԹՍ-ն կայունացած ռեժիմում պահպանում է առաջադրված բարձրությունը, ինչը բավարարում տրված խնդրի պահանջները:



Նկ. 4. ԱԹՍ-ի բարձրության գրաֆիկը՝ նեյրոնային կարգավորիչի կիրառմամբ

Եզրակացություն: Բազմառոտոր ԱԹՍ-ի համար նախագծված նեյրոնային կարգավորիչը, որը ուսուցանվել է արհեստական բանականության ամրապնդման ուսուցման մեթոդներով, բավարարում է առաջադրված պահանջները, կարող է կարգավորել տրված պարամետրերը նվազագույն սխալանքով և գերկարգավորումով: Հետագայում այն կարող է կիրառվել նաև նոր տեսակի ԱԹՍ-ների կառավարման խնդիրների լուծմանն ուղղված աշխատանքներում:

ԳՐԱԿԱՆՈՒԹՅԱՆ ՑԱՆԿ

1. **Gasparyan O.N., Darbinyan H.G.** Adaptive System of Compensation of Motors' Partial Degradations of Multirotor UAVs // In: A.Yuschenko (eds) // Modern Problems of Robotics. MPoR 2020. Communications in Computer and Information Science. - 2020. -Vol 1426. Springer, Cham. https://doi.org/10.1007/978-3-030-88458-1_16. - P. 207-219
2. **Bithas S., Emmanouel T., Nomikos N., Vouyioukas D., Athanasios G.** A Survey on Machine-Learning Techniques for UAV-Based Communications // Sensors. -2019.-No 23: 5170.
3. **Bou-Ammar H., Voos H., Ertel W.** Controller design for quadrotor UAVs using reinforcement learning // 2010 IEEE International Conference on Control Applications. - Yokohama, Japan, 2010. - P. 2130-2135, doi: 10.1109/CCA.2010.5611206.
4. **Yintao Zhang Youmin Zhang and Ziquan Yu.** Path Following Control for UAV Using Deep Reinforcement Learning Approach // Guidance, Navigation and Control.- 2021. -Vol. 01, No. 01, 2150005
5. Multi-Agent Reinforcement Learning Aided Intelligent UAV Swarm for Target Tracking / **Z. Xia, et al** // IEEE Transactions on Vehicular Technology. -Jan. 2022.- Vol. 71, No. 1.- P. 931-945, doi: 10.1109/TVT.2021.3129504.

Н.Г. НЕРСИСЯН

РАЗРАБОТКА НЕЙРОННОГО РЕГУЛЯТОРА БЕСПИЛОТНОГО ЛЕТАТЕЛЬНОГО АППАРАТА С ИСПОЛЬЗОВАНИЕМ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ

Для многороторного беспилотного летательного аппарата (БПЛА) разработан нейронный регулятор, машинное обучение которого осуществлялось с помощью обучения с подкреплением. Спроектирована система управления БПЛА в среде Matlab Simulink, где был включен регулятор. Проведено исследование полученной системы.

Ключевые слова: беспилотный летательный аппарат (БПЛА), система управления, искусственный интеллект, регулятор, обучение с подкреплением.

N.H. NERSISYAN

DESIGN OF A NEURAL CONTROLLER FOR AN UNMANNED AERIAL VEHICLE USING REINFORCEMENT LEARNING

A neural controller was designed for a multi-rotor unmanned aerial vehicle (UAV), the machine learning of which was accomplished using reinforcement learning. A UAV control system was designed in the Matlab Simulink environment, where a neural controller was included. An analysis of the resulting system was carried out.

Keywords: unmanned aerial vehicle, control system, artificial intelligence, controller, reinforcement learning.

ՀՏԴ 62-52

Ձ.Գ. ԽԱՆԱՄԻՐՅԱՆ, Ե.Մ. ՏԵՐՏԵՐՅԱՆ

ՌՌԲՈՏ-ՄԱՆԻԳՈՒԼՅԱՏՈՐԻ ՄՈԴԵԼԱՎՈՐՈՒՄԸ LABVIEW ԾՐԱԳՐԱՅԻՆ ՄԻՋԱՎԱՅՐՈՒՄ

Ռոբոտ-մանիպուլյատորները դարձել են մեր առօրյայի անբաժան մասը: Օր օրի մեծանում է դրանց կիրառությունը արդյունաբերության, կենցաղի, բժշկության և այլ բնագավառներում: Ռոբոտ-մանիպուլյատորների մինչև արտադրական փուլին հասնելը անհրաժեշտ է կատարել դրանց սիմուլյացիոն մոդելի նախագծում և ուսումնասիրում: Այդ նպատակով LabVIEW Robotics միջավայրում մշակվել է ծրագրային հավելված, որը հնարավորություն է տալիս՝ ուսումնասիրելու տարբեր չափսերի ռոբոտ-մանիպուլյատորները և դրանց աշխատանքային տիրույթը:

Առանցքային բառեր. ռոբոտ, ռոբոտատեխնիկա, կինեմատիկա, մանիպուլյատոր, LabVIEW:

Ներածություն: Ռոբոտաշինությունը զբաղվում է այն մեքենաների ուսումնասիրությամբ, որոնք կարող են փոխարինել մարդուն ինչպես ֆիզիկական, այնպես էլ բանական գործողություններ իրականացնելու հարցերում: Վերջին քսան տարիների ընթացքում արդյունաբերության և տեղեկատվական տեխնոլոգիաների կտրուկ աճը զարկ տվեց ռոբոտաշինության զարգացմանը, և այսօր մեր կյանքը պատկերացնել առանց ռոբոտների դարձել է անհնար: Ներկայումս ռոբոտների տեխնոլոգիական հագեցվածությունը թույլ է տալիս փոխարինել մարդկանց այնպիսի ոլորտներում, ինչպիսիք են ծանր արդյունաբերությունը, մեքենաշինությունը, տիեզերագնացությունը և այլն: Արդյունքում սովորական է դարձել տեսնել ռոբոտներ, որոնք օգտագործվում են եռակցման և հոսքագծերի վրա մեքենաների ներկման կամ ստորջրյա և ստորգետնյա միջավայրերում կառույցների ստուգման ու վերանորոգման, սենյակների մաքրության իրականացման կամ խոհանոցում սննդի պատրաստման համար [1]: