Heurisztikusan gyorsított Fuzzy szabály-interpoláció alapú Q-tanulás

Heuristically Accelerated Fuzzy Rule Interpolation-based Q-learning

TOMPA Tamás[1], KOVÁCS Szilveszter[2]
[1]PhD, egyetemi adjunktus, [2]Prof, egyetemi tanár

[1,2]Miskolci Egyetem, Általános Informatikai Intézeti Tanszék
Miskolc-Egyetemváros, Egyetem út 1. H-3515
[1]email: tompa@iit.uni-miskolc.hu
[2]email: szkovacs@iit.uni-miskolc.hu

Abstract

The learning phase of the conventional reinforcement learning methods (e.g. Q-learning, SARSA, and Fuzzy Q-learning) starts with an empty knowledge base, which the system gradually builds during the learning process based on the feedback from the environment. However, if partially knowledge base is available and can be integrated into the learning phase, this can have a positive effect on learning performance. The aim of the paper is to introduce a Fuzzy Rule-Interpolation method ('FIVE') based Q-learning method which is suitable for incorporating external expert knowledge into the learning process and capable of fine-tuning the initially imprecisely defined expert knowledge during the learning process, thereby correcting (optimizing) it.

Keywords: reinforcement learning, Q-learning, Fuzzy Rule Interpolation, expert knowledge, knowledge optimization

Kivonat

A klasszikus megerősítéses tanulási módszerek (Q-tanulás, SARSA, Fuzzy Q-tanulás), tudásbázisa kezdetben ismeretlen, a rendszer ezt a tanulási folyamat során alakítja ki a környezet visszajelzései alapján. Azonban, ha rendelkezésre áll egy részlegesen ismert tudásbázis, amely beilleszthető a tanulási folyamatba, akkor ez által a tanulás hatékonysága növelhető. A cikk bemutat egy olyan fuzzy szabály-interpolációs módszeren ('FIVE') alapúló Q-tanulási módszert, amely alkalmas külső szakértői tudásbázis injektálására a tanulási folyamatba és alkalmas továbbá a pontatlanul megadott kezdeti szakértői tudásbázis finomhangolására a tanulási folyamat során, így pontosítva (optimalizálva) azt.

Kulcsszavak: megerősítéses tanulás, Q-tanulás, Fuzzy Szabály-Interpoláció, szakértői tudásbázis, tudásbázis hangolás

1. BEVEZETÉS

A megerősítéses tanulás (Reinforcement Learning - RL) [8] a gépi tanulás egyre inkább népszerűbb kutatási területe és az élet számos területén már alkalmazzák [5]. Ezen módszerek (pl. Q-learning [19], SARSA [7], Deep Q-learning [2] és ezek Fuzzy modell-alapú kiterjesztései [1]) működése a környezet által adott visszajelzéseken (megerősítéseken) alapszik. Egy megfelelően definiált jutalomfüggvény által az ágens az adott állapotban a végrehajtott cselekvésre jutalmat vagy büntetést (negatív jutalmat) kap a környezettől, amely alapján igyekszik jövőbeli cselekvéseit megválasztani, majd a szerzett jutalmakat hosszútávon maximalizálni. Ezen módszereket próbálkozás (trial and error) típusú módszereknek is nevezik, mert a megoldást (és az azt leíró Q-függvényt) számtalan próbálkozás (különböző akciók kipróbálása) útján keresik meg csupán a kapott megerősítések ismeretében. Egyik nagy előnyük tehát, hogy a modell előzetes ismerete nélkül képesek megkeresni a problémát leíró megoldás tudásbázisát csupán az elérendő cél meghatározása által.

A RL algoritmusok általában nem rendelkeznek semmilyen előzetes tudásbázissal, hanem a tanulási folyamat során, számtalan iteráció alatt hozzák létre azt. A tudásbázis reprezentálásának módja algoritmusonként eltérő, Q-learning esetében egy Q-tábla (állapot-akció-érték párok) írja ezt le, fuzzy modell-alapú módszerek esetében pedig egy fuzzy „ha-akkor" típusú szabályokat tartalmazó szabálybázis. Szakértő által meghatározott tudásbázis beépítése az RL módszerek tanulási folyamatába jelentősen javíthat a rendszer konvergencia sebességén, tanulási hatékonyságán [9].

A cikk célja egy olyan fuzzy szabály-interpolációs módszeren alapúló Q-learning módszer bemutatása, amely lehetőséget ad szakértői által definiált tudásbázis beépítésére a rendszerbe és mindemellett alkalmas az esetlegesen pontatlanul megadott szakértői szabályrendszer finomhangolására és validálására.

2. HEURISZTIKUSAN GYORSÍTOTT FUZZY SZABÁLY-INTERPOLÁCIÓ ALAPÚ Q-TANULÁS

A HFRIQ-learning (Heuristically Accelerated Fuzzy Rule-Interpolation based Q-learning) [13] a FRIQ-learning (Fuzzy Rule Interpolation-based Q-learning) [17] rendszer kiterjesztése, amely által külső emberi (szakértői) tudásbázis beépíthető (és hangolható) a rendszer tanulási folyamatába. A rendszer tudásbázisa egy ritka (fuzzy szabály-interpolált) fuzzy szabálybázis által leírt, ahol az m méretű, R szabálybázis egy rendszer által létrehozott r_i ($i \in [1, m]$) szabályának formátuma a következő [17]:

$$r_i\text{: } \boldsymbol{If}\ s_1 \text{ is } S_1^i\ \boldsymbol{And}\ s_2 \text{ is } S_2^i\ \boldsymbol{And}\ \dots\ \boldsymbol{And}\ s_n \text{ is } S_n^i\ \ \boldsymbol{And}\ a \text{ is } A^i\ \boldsymbol{Then}\ \tilde{Q}(\boldsymbol{s}, a) = q^i \qquad (1)$$

ahol S_j^i az i-edik ($i \in [1, m]$) szabály j-edik ($j \in [1, n]$) állapot dimenziójának fuzzy halmaza az n-dimenziós $\boldsymbol{S}$ állapottérben, $\boldsymbol{s} \in \boldsymbol{S}$ az n-dimenziós állapot megfigyelés, s_j a j-edik dimenziója az $\boldsymbol{s}$ állapot megfigyelésnek, A^i az i-edik szabály egydimenziós akció univerzumának (U) fuzzy halmaza, $a \in U$ az akció, $\tilde{Q}(\boldsymbol{s}, a)$ a FIVE FRI [4] által becsült Q-függvény, q^i pedig az i-edik szabály konzekvense (Q-értéke).

A HFRIQ-learning rendszer esetében a R_{expert} szakértői szabálybázis $\hat{r}_i$ ($i \in [1, \hat{m}]$) szakértői szabályai állapot-akció formátumban definiálhatók, ahol az $\hat{S}_n^i$ állapot a szabályantecedens az $\hat{A}^i$ akció pedig az ehhez az állapothoz tartozó konzekvens:

$$\hat{r}_i\text{: } \textbf{If } s_1 \boldsymbol{is}\ \hat{S}_1^i\ \boldsymbol{And}\ s_2 \boldsymbol{is}\ \hat{S}_2^i\ \boldsymbol{And}\ \dots \boldsymbol{And}\ s_n \boldsymbol{is}\ \hat{S}_n^i\ \boldsymbol{Then}\ \text{a} = \hat{A}^i \qquad (2)$$

ahol $\hat{r}_i$ az i-edik ($i \in [1, \hat{m}]$) szakértői szabály az $\hat{m}$-méretű R_{expert} szabálybázisban, $\hat{S}_n^i = [\hat{S}_1^i, \hat{S}_2^i, \dots \hat{S}_n^i]$ az i-edik szakértői szabály n-dimenziós állapot megfigyelése, $\hat{A}^i$ az ehhez az $\hat{S}_n^i$ állapot megfigyeléshez tartozó akció, i ($i \in [1, \hat{m}]$) pedig a szabály indexe.

A rendszer a tanulási fázis elején az R_{expert} szakértői szabálybázist injektálja a tanulási folyamatba. Az injektálás során az $\hat{r}$ szakértői szabályok (2) formátuma átalakításra kerül az (1) formátumúra, aminek következményeképp minden egyes szakértői szabályra meghatározásra kerül egy kezdeti Q-érték $(\tilde{Q}_{init})$, a szintén szakértő által meghatározott g_{max} maximálisan adható megerősítés alapján [10]. Az így módon létrejött szakértői szabályok új antecedense az állapot-akció, konzekvense pedig a $\tilde{Q}_{init}$ érték lesz. Abban az esetben, ha a valamely $\hat{r}_i$ szakértői szabály pontosan illeszkedne a 2^{n+1} (n: állapotváltozók száma) darabszámú szabállyal rendelkező sarokponti szabálybázis valamely $r_i^{\square}$ sarokponti szabályára (amely így ellentmondáshoz vezet az azonos antecedens de eltérő konzekvens miatt), akkor az ellentmondás feloldása érdekében az $r_i^{\square}$ sarokponti szabály 0 Q-értéke lecserélésre kerül az $\hat{r}_i$ szakértő szabály $\tilde{Q}_{init}$ Q-értékére. Az így létrejött, ellenmondó szabályokat már nem tartalmazó kezdeti szabálybázis fog kiegészülni a tanulási folyamat során a rendszer által létrehozott új r_i szabályokkal [17].

Egy új, rendszer által létrehozott r_i szabály akkor kerül hozzáadásra a szabálybázisba (az aktuális állapot-akció pontban), ha a $\Delta\tilde{Q}$ Q-frissítés értéke nagyobb, mint egy ε_Q küszöbérték ($\Delta\tilde{Q} > \varepsilon_Q$) és a megfigyeléshez legközelebbi, már létező szabály is távolinak tekinthető. A legközelebbi szabály meghatározásának alapja a szabályok közötti, dimenziónként meghatározott távolságküszöbök [12][13]. Ellenkező esetben, ha a $\Delta\tilde{Q}$ relatívan kicsi ($\Delta\tilde{Q} < \varepsilon_Q$) és a legközelebbi szabály is távolinak számít, akkor a teljes szabálybázis konzekvense (Q-értéke) frissülni fog a következő módon [17]:

$$\tilde{Q}^{k+1}(\boldsymbol{s}, a) = \tilde{Q}^k(\boldsymbol{s}, a) + \Delta\tilde{Q}^{k+1}(\boldsymbol{s}, a) \qquad (3)$$

$$\Delta\tilde{Q}^{k+1}(\boldsymbol{s}, a) = \alpha * \left(g(\boldsymbol{s}, a, \boldsymbol{s}') + \gamma * \max_{a' \epsilon U} \tilde{Q}^k(\boldsymbol{s}', a') - \tilde{Q}^k(\boldsymbol{s}, a)\right) \qquad (4)$$

ahol q_i^{k+1} az i-edik szabály konzekvense a $(k+1)$-edik iterációban, $\alpha \in [0,1]$ a tanulási ráta, $\gamma \in [0,1]$ a leszámítolási tényező. Az új megfigyelt állapot $\boldsymbol{s}'$, $g(\boldsymbol{s}, a, \boldsymbol{s}')$ a megfigyelt jutalom az $\boldsymbol{s} \to \boldsymbol{s}'$ állapot átmenetre, a pedig az $\boldsymbol{s}$ állapotban végrehajtott akció, $\tilde{Q}^k$ és $\tilde{Q}^{k+1}$ pedig a k-adik és a $(k+1)$-edik iteráció

FIVE FRI módszer által becsült Q-értéke [17]:

$$\tilde{Q}(\boldsymbol{s},a)=\begin{cases} q^i & \text{ha } (\boldsymbol{s},a)=(\boldsymbol{s}^i,a^i) \\ \sum_{i=1}^{m}\left(\left({q^i}/{(\delta_v^i)^\lambda}\right)\Big/\left(\sum_{j=1}^{m}{1}/{(\delta_v^j)^\lambda}\right)\right) & \text{valamennyi } i-re, \\ & \text{egyébként} \end{cases} \tag{5}$$

ahol $(\boldsymbol{s},a)$ az állapot-akció megfigyelés, q^i az i-edik ($i\in[1,m]$) szabály konzekvense, δ_v^i a skálázott távolság [4] az $(\boldsymbol{s},a)$ megfigyelés és az i-edik szabály $(\boldsymbol{s}^i,a^i)$ antecedense között, λ a Shepard paraméter, m pedig a szabályok száma a szabálybázisban.

Ha a $\Delta\tilde{Q}$ értéke kicsi ($\Delta\tilde{Q}<\varepsilon_Q$) de van létező szabály az állapot-akció megfigyelés közelében (a dimenziónként meghatározott távolságok és távolságküszöbök alapján [12][13]), akkor a legközelebbi szabálypont hangolása valósul meg egy gradiens-módszer alapú optimalizálási eljárással. Ekkor az állapot-akció megfigyeléshez legközelebb elhelyezkedő szabály állapot-akció antecedense és konzekvens Q-értéke fog frissülni (és ennek következtében elmozdulni a térben) a következő összefüggés alapján [13]:

$$\boldsymbol{s}_{k+1}=\boldsymbol{s}_k-\left(2*TDerror*\frac{\partial\tilde{Q}(\boldsymbol{s},a)}{\partial\boldsymbol{s}}\right)*\alpha \tag{6}$$

$$a_{k+1}=a_k-\left(2*TDerror*\frac{\partial\tilde{Q}(\boldsymbol{s},a)}{\partial a}\right)*\alpha \tag{7}$$

$$q_{k+1}=q_k-\left(2*TDerror*\frac{\partial\tilde{Q}(\boldsymbol{s},a)}{\partial q}\right)*\alpha \tag{8}$$

$$\text{ahol } TDerror=g(\boldsymbol{s},a,\boldsymbol{s}')+\gamma*\max_{a'\epsilon U}\tilde{Q}^k(\boldsymbol{s}',a')-\tilde{Q}^k(\boldsymbol{s},a) \tag{9}$$

ahol $\boldsymbol{s}_k,a_k,q_k$ a szabálypont régi állapot, akció és Q-értéke, a $\boldsymbol{s}_{k+1},a_{k+1},q_{k+1}$ a gradiens-módszer által meghatározott új állapot, akció és Q-érték, $TDerror$ a TD-hiba, α a gradiens-módszer tanulási rátája, $\frac{\partial\tilde{Q}(\boldsymbol{s},a)}{\partial\boldsymbol{s}},\frac{\partial\tilde{Q}(\boldsymbol{s},a)}{\partial a},\frac{\partial\tilde{Q}(\boldsymbol{s},a)}{\partial q}$ pedig a Q-függvény állapot, akció és Q-érték szerinti parciális deriváltjai.

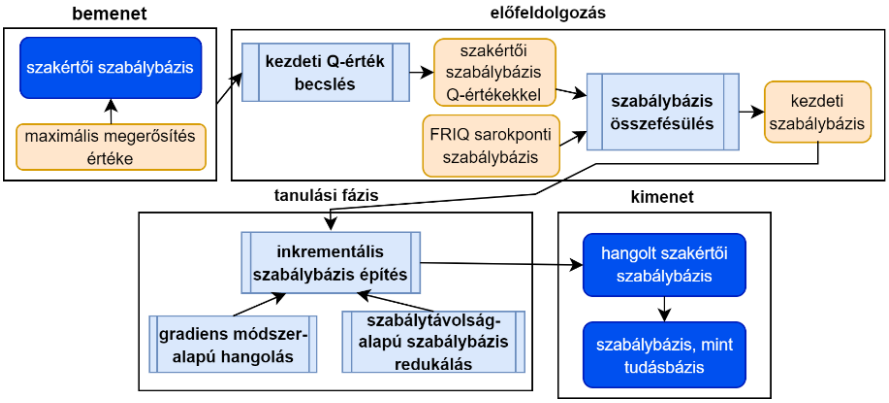
A graidens módszer-alapú hangolási folyamat következtében előfordulhat, hogy néhány szabálypont közel kerül egymáshoz. Az egymáshoz közel kerülő szabály nagyon hasonló információt írnak le, így ezek a szabályszám (és így a rendszer komplexitásának) csökkentése érdekében egyesíthetők egyetlen szabállyá. A tanulási folyamat során alkalmazott szabálytávolság alapú szabálybázis redukálási módszerről részletesebb információk a [12][13] és [14] hivatkozásokban találhatók. A tanulási folyamat után opcionálisan alkalmazható további szabálybázis-csökkentési módszereket mutatnak be a [11][16] és [18] hivatkozások.

A HFRIQ-learning algoritmus (azaz a tanulási folyamat) akkor ér véget és áll elő a behangolt, végleges szabálybázis (tudásbázis), ha már nem kerül új szabály létrehozásra a rendszer által, a $\Delta\tilde{Q}$ értéke elenyészően kicsi és a szabálypontok már nem mozdulnak el jelentősen (azaz stabilizálódtak).

A tanulási folyamat előtt megadott szakértői szabálybázis, majd a tanulási (hangolási) folyamat után visszaolvasott szabályrendszer szabályainak eltérései alapján következtetni lehet a kezdetben megadott szakértői tudásbázis helyességére az által, hogy a rendszer milyen mértékben módosította (hangolta) az egyes szabálypontokat.

A szakértői szabályrendszer validálására vonatkozó mérőszám (helyességmérték) kidolgozása jövőbeli kutatási terv, ahogyan az állapot-akció formátumú szakértői szabályok emberi szem által olvashatóbb formában, FBDL (Fuzzy Behavior Declarative Language) fuzzy viselkedést leíró nyelv [6] által történő megadása (és visszaolvasása) is.

A HFRIQ-learning felépítését a következő ábra szemlélteti:



1. ábra. *A HFRIQ-learning megerősítéses tanulási módszer blokkvázlata*

2.

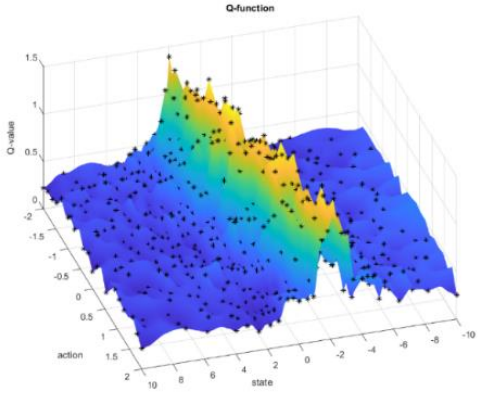
3. SZIMULÁCIÓS MINTAPÉLDA

A szakértői tudásbázis hangolása a HFRIQ-learning rendszer által egy egyszerű, egyetlen állapot- és akcióváltozóval rendelkező mintapéldán keresztül kerül bemutatásra, amely által a kapott irányítási felület (Q-függvény) jól vizualizálható (3 dimenzió). A szimulációs példa paraméterei és azok értékei a következők:

- állapotváltozó: $s_1 \in [-10,10]$
- akcióváltozó: $a \in [-2,2]$
- tanulási ráta: $\alpha = 0.5$
- leszámítolási tényező: $\gamma = 0.4$
- ε-greedy politika véletlen akció választásának valószínűsége: $\varepsilon = 0.5$
- epizódok száma: 1
- iterációk száma (az egyetlen epizódon belül): 10000

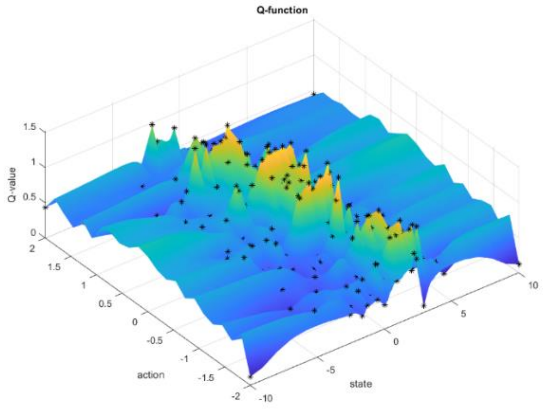
A jutalomfüggvény úgy definiált, hogy a környezet $+1$ jutalmat ad az ágensnek, ha s_1 változó pillanatnyi értéke -1 és $+1$ közötti, ellenkező esetben a megerősítés (büntetés) értéke -1.

A kidolgozott mintapélda mind a FRIQ-learning és mind a HFRIQ-learning rendszerben futtatásra került. A kapott futási eredmények összehasonlításának alapja az eredeti FRIQ-learning [17] rendszer alkalmazása esetében kapott futási eredmények, tehát a szakértői heurisztika nélküli, illetve a bemutatott hangolási- és szabályszám csökkenetési módszerek alkalmazása nélküli verzió. Ebben a futási esetben (I.) a Q-függvényt 530 darab szabályból álló szabálybázis írta le, amely felületét a következő 2. ábra szemlélteti:



3. ábra. *A FRIQ-learning alkalmazása esetében (I. futási eset) kapott Q-függvény felülete (530 darab szabály)*

A II. futási esetben alkalmazásra került a bemutatott HFRIQ-learning rendszer, és egyetlen szakértői szabály került megadásra, amely a 0 állapotban a 0 értékű akciót definiálja, a $g_{max} = 1$ maximálisan adható megerősítés mellett. A tanulási folyamat során a szabályok között megengedett minimális szabálytávolság az állapot- és akcióuniverzum 200-ad része ($\mathrm{dR}_S = \mathrm{dR}_U = 200$), a szabályszám redukálás (szabályegyesítés) távolságküszöbeinek értékei univerzumonként pedig $\mathrm{dR}_S = 45$, $\mathrm{dR}_U = 45$ és $\mathrm{dR}_q = 100$. Ebben a futási esetben az 1. ábrán bemutatott (referenciaként szolgáló) Q-függvényt 327 darab szabály írta le (az 530 darab helyett), amely felületét a 3. ábra szemlélteti:



4. ábra. *A HFRIQ-learning alkalmazása esetében (II. futási eset) kapott Q-függvény felülete (327 darab szabály)*

Ebben a futási esetben (II.) a Q-függvény alakja a gradiens módszer-alapú hangolási eljárás miatt kisimultabb és a szabálymozgások következtében a szabálytávolság alapú szabályszám csökkentési módszer összevont egymáshoz közeli szabályokat, így 327 darab szabály írja le ugyanazon Q-függvényt, melyet a 2. ábra szemléltet. A szabályegyesítések során az új (összevont) szabály típusa a forrásszabályok típusa (új, szakértői vagy sarokponti) alapján kerül meghatározásra [13].

A következő táblázat a kezdeti szakértői szabályt hasonlítja össze a hangolási (tanulási) folyamat végeztével kapott (behangolt) szakértői szabállyal:

A szakértői szabály a tanulási folyamat előtt és a hangolási folyamat után 1. táblázat

Szakértői szabály	**állapot**	**akció**	**Q-érték**
eredetileg (tanulási folyamat előtt) megadott	0	0	0.1
hangolási folyamat után kapott	0.06	0	0.59

Mivel a tanulási folyamat során a szakértői szabály állapot-akció pontja és Q-értéke csak kismértékben került hangolásra a HFRIQ-learning által, így a megadott szakértői szabály helyesnek tekinthető. A szabálypont hangolása után előállt pozitív Q-érték alapján is a szabály helyesnek tekinthető, a Q-érték változása csak a szabály hasznosságára vonatkozó becslést pontosította.

4. ÖSSZEFOGLALÁS

A cikkben bemutatásra került egy olyan megerősítéses tanulási módszer, amely által emberi szakértői tudás beépíthető a rendszerbe. A bemutatott HFRIQ-learning módszer lehetőséget ad a tanulási folyamat hatékonyságának növelésére az szakértői tudásbázis injektálása által, amely következtében csökkenhet az állapottér bejárásának mértéke. A tanulási folyamat hatékonyságának növekedésén túl, az injektált szakértői szabályok visszakeresésével a tanulási és optimalizálási folyamat után, az eredeti verziójukkal történő összehasonlításával kiértékelhető (validálható) a kezdeti szakértői szabályok minősége, helyessége. A kezdeti szakértői szabályok kinyerése a tanulási folyamat után lehetővé teszi a korábban megszerzett és a rendszer által finomhangolt tudás újra felhasználását más rendszerekben (pl. transfer learning [15]). A bemutatott HFRIQ-learning rendszer szintén hasznos lehet olyan modellek optimalizálásában, ahol a szakértői tudásbázis fuzzy szabályok formájában leírható, például az etológiai indíttatású, robotok viselkedésmodellezését megvalósító rendszerek esetében [3].

KÖSZÖNETNYILVÁNÍTÁS

„A Kulturális és Innovációs Minisztérium ÚNKP-23-4-I-ME/5 kódszámú Új Nemzeti Kiválóság Programjának a Nemzeti Kutatási, Fejlesztési és Innovációs Alapból finanszírozott szakmai támogatásával készült.”

IRODALMI HIVATKOZÁSOK

[1] Berenji, Hamid R. "Fuzzy Q-learning for generalization of reinforcement learning." Proceedings of IEEE 5th International Fuzzy Systems. Vol. 3. IEEE, 1996.
[2] Fan, Jianqing, et al. "A theoretical analysis of deep Q-learning." Learning for Dynamics and Control. PMLR, 2020.
[3] Kovács, S., Vincze, D., Gácsi, M., Miklósi, Á., & Korondi, P. (2011, May). Ethologically inspired robot behavior implementation. In 2011 4th International Conference on Human System Interactions, HSI 2011 (pp. 64-69). IEEE
[4] Kovács, Szilveszter. "Extending the fuzzy rule interpolation" FIVE" by fuzzy observation." Computational Intelligence, Theory and Applications. Springer, Berlin, Heidelberg, 2006. 485-497.
[5] Naeem, M., Rizvi, S. T. H., & Coronato, A. (2020). A gentle introduction to reinforcement learning and its application in different fields. IEEE access, 8, 209320-209344.
[6] Piller, Imre, and Szilveszter Kovács. "FBDL: A Declarative Language for Interpolative Fuzzy Behavior Modeling." 2019 IEEE 23rd International Conference on Intelligent Engineering Systems (INES). IEEE, 2019.
[7] Rummery, Gavin A., and Mahesan Niranjan. On-line Q-learning using connectionist systems. Vol. 37. Cambridge, UK: University of Cambridge, Department of Engineering, 1994.
[8] Sutton, Richard S., and Andrew G. Barto. Reinforcement learning: An introduction. MIT press, 2018.
[9] Tompa, T., Kovács, S., Vincze, D., & Niitsuma, M. (2021, January). Demonstration of expert knowledge injection in Fuzzy Rule Interpolation based Q-learning. In 2021 IEEE/SICE International Symposium on System Integration (SII) (pp. 843-844). IEEE.
[10] Tompa, Tamás, and Szilveszter Kovács. "Applying Expert Heuristic as an a Priori Knowledge for FRIQ-Learning." Acta Polytechnica Hungarica 17.4 (2020).
[11] Tompa, Tamás, and Szilveszter Kovács. "Clustering-based fuzzy knowledgebase reduction in the FRIQ-learning." 2017 IEEE 15th International Symposium on Applied Machine Intelligence and Informatics (SAMI). IEEE, 2017.
[12] Tompa, Tamás, and Szilveszter Kovács. "Determining the minimally allowed rule-distance for the incremental rule-base contruction phase of the FRIQ-learning." 2018 19th International Carpathian Control Conference (ICCC). IEEE, 2018.
[13] Tompa, Tamás, and Szilveszter Kovács. "Heuristically accelerated FRIQ-learning." 20th Jubilee International Symposium on Intelligent Systems and Informatics (SISY 2022). IEEE, 2022.
[14] Tompa, Tamás, and Szilveszter Kovács. "Tudásbázis redukálás a heurisztikusan gyorsított FRIQ-learning rendszerben." Production Systems and Information Engineering 11.2 (2023): 1-12.
[15] Torrey, Lisa, and Jude Shavlik. "Transfer learning." Handbook of research on machine learning applications and trends: algorithms, methods, and techniques. IGI global, 2010. 242-264.
[16] Vincze, Dávid, Alex Tóth, and Mihoko Niitsuma. "Antecedent redundancy exploitation in fuzzy rule interpolation-based reinforcement learning." 2020 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM).
[17] Vincze, Dávid, and Szilveszter Kovács. "Fuzzy rule interpolation-based Q-learning." 2009 5th International Symposium on Applied Computational Intelligence and Informatics. IEEE, 2009.
[18] Vincze, Dávid, and Szilveszter Kovács. "Rule-base reduction in Fuzzy Rule Interpolation-based Q-learning." Recent Innovations in Mechatronics 2.1-2. (2015): 1-6.
[19] Watkins, Christopher JCH, and Peter Dayan. "Q-learning." Machine learning 8.3 (1992): 279-292.