

MISKOLCI EGYETEM
GÉPÉSZMÉRNÖKI ÉS INFORMATIKAI KAR



**Fuzzy szabály-interpolációs állapotgép modellek
és hangolási eljárásaik**

Tézisfüzet

Készítette:

Tompa Tamás

okleveles villamosmérnök

Hatvany József Informatikai Tudományok Doktori Iskola

Doktori iskola vezető

Prof. Dr. Szigeti Jenő

egyetemi tanár

Témavezető

Prof. Dr. Kovács Szilveszter

egyetemi tanár

Miskolc

2023

TARTALOMJEGYZÉK

1	BEVEZETÉS.....	- 1 -
1.1	A KUTATÁS CÉLKITŰZÉSEI.....	- 2 -
2	ELŐZMÉNYEK	- 4 -
2.1	A MEGERŐSÍTÉSES TANULÁS.....	- 4 -
2.2	A FRIQ-LEARNING MÓDSZER	- 5 -
3	ÚJ TUDOMÁNYOS EREDMÉNYEK	- 9 -
3.1	SZAKÉRTŐI TUDÁSBÁZIS BEÉPÍTÉSE.....	- 9 -
3.1.1	<i>Tézis I.</i>	- 9 -
3.2	A FRI Q-FÜGGVÉNYT LEÍRÓ SZABÁLYBÁZIS HANGOLÁSA	- 10 -
3.2.1	<i>Tézis II.</i>	- 11 -
3.3	SZABÁLYBÁZIS REDUKCIÓ	- 11 -
3.3.1	<i>Tézis III.</i>	- 12 -
3.4	A HFRIQ-LEARNING	- 13 -
4	ÖSSZEFOGLALÁS.....	- 15 -
5	SUMMARY	- 17 -
5.1	THESIS I.	- 18 -
5.2	THESIS II.....	- 18 -
5.3	THESIS III.	- 19 -
6	AZ ÉRTEKEZÉS IRODALOMJEGYZÉKE	- 20 -
7	SAJÁT PUBLIKÁCIÓK	- 29 -

1 BEVEZETÉS

Az egyre növekvő gépi számítási kapacitás és ennek a hétköznapi eszközökben történő megjelenése következtében a mesterséges intelligencia [80], a gépi tanulás [17][68] témaköre illetve az ezen módszerek nyújtotta lehetőségek kiaknázására való törekvés egyre inkább aktuálisává válik, egyre nagyobb jelentőséggel bír. A gépi tanulás olyan módszerek összessége, amelyek tapasztalatszerzés útján tanulnak, ezáltal lépésről-lépésre építve fel a rendszer működtető tudásbázisát. Több típusa elterjedt, vannak olyan algoritmusok amelyek külső mintaadatok (példa-halmazok, tanítóminták) alapján igyekeznek törvényszerűségeket feltárni a rendszer működésére vonatkozóan, majd ezek alapján a még „nem látott” ismeretlen helyzetekre is „helyes” döntést hozni. Másik típusa mikor nem áll rendelkezésre külső tanítóminta majd a rendszer próbálkozások és a próbálkozásokra kapott válaszok (megerősítések) alapján térképezi fel a megoldás mikéntjét, hozza létre a működtető tudásbázist. Ez az úgynevezett - jelenleg is népszerű és egyre jobban kutatott tudományterület - megerősítéssel tanulás.

A megerősítéssel tanuló módszereken alapuló rendszerek hasonló módon tanulnak, mint az ember is teszi a gyermekkorától kezdve. A környezettel való kölcsönhatásba lépés során az abból érkező megerősítési információk (amelyek lehetnek jutalmak vagy büntetések) alapján igyekszik lehetséges cselekedetei, döntései közül a legmegfelelőbbet végrehajtani, megfigyelni a környezet arra adott reakcióját (megerősítését), tapasztalatait ennek megfelelően bővíteni, hogy elérje a kívánt célt. Ezen tanulási algoritmusok összessége általában üres tudásbázissal indítja a tanulási folyamatot, ahogyan az ember is gyermek korában (kezdetben semmit sem ismer környezetből, majd törekszik felfedezni azt), majd a cselekedetekre kapott megerősítések következtében lépésről-lépésre igyekszik gyarapítani tudását, tapasztalatait. Ennek következtében ezen módszerek jól használhatók olyan rendszerekben ahol a működés egzakt folyamata nem ismert, az elérendő cél definiálása után a megerősítések alapján térképezik fel a megoldásul szolgáló működtető tudásbázist, modellt. A megoldás keresésének folyamatát és annak hosszúságát nagymértékben befolyásolja a definiált állapot-cselekvési dimenziók és a jutalomfüggvény, ezek megfelelő meghatározása kulcsfontosságú lépés. Abban az esetben ha rendelkezésre áll részleges információ a megoldás mikéntjére vonatkozóan,

akkor ennek a megerősítéses tanuló rendszerekbe történő beépítésével felgyorsítható a teljes tanulási folyamat.

A kutatás (jelen doktori értekezés) célja egy olyan fuzzy szabály-interpoláción alapuló megerősítéses tanulási módszer továbbfejlesztése és kidolgozása, amely alkalmas emberi szakértő által megadott előzetes (*a priori*) tudásbázis (mint heurisztika) injektálására a rendszerbe, a rendszerbe adoptált szakértői heurisztika hangolására (optimalizálására), illetve a rendszer tudásbázisát leíró fuzzy szabálybázis méretének tanulási folyamat közbeni csökkentése (redukálása). A kutatás alapjául a „Fuzzy szabály-interpoláció alapú Q-learning” [97][98] (*Fuzzy Rule Interpolation-based Q-learning - FRIQ-learning*, D. Vincze, Sz. Kovács, 2009) rendszer szolgál, a kutatás ezen Fuzzy szabályinterpoláció alapú Q-learning módszer szakértői heurisztikával való kiterjesztésére, hangolására és tudásbázisának redukálására irányul.

1.1 A KUTATÁS CÉLKITŰZÉSEI

Elsődleges cél a „Fuzzy szabály-interpoláció alapú Q-learning” (*Fuzzy Rule Interpolation-based Q-learning - FRIQ-learning*) megerősítéses tanulási módszer továbbfejlesztése oly módon, hogy alkalmas legyen egy előre megadott (*a priori*) szakértői tudásbázis (mint heurisztika) befogadására úgy, hogy az a későbbiekben, szükség esetén a tudásbázis többi részével együtt hangolható legyen. Ez olyan módszerek kifejlesztését jelenti, amelyek lehetővé teszik a szakértői tudásbázis valamilyen (magasabb szintű) formában történő leírását és a rendszerbe történő injektálását, és képesek a szakértői szabályrendszer hangolására és validálására (helyességének ellenőrzésére) is. Ezek alapján a kutatás célkitűzései a következőképpen fogalmazhatók meg:

- Olyan előzetes tudásbázis, szakértői szabályrendszer leírási forma kidolgozása, melyben az előzetes szakértői heurisztika megadható és a FRIQ-learning rendszerbe építhető.
- Kezdeti Q-érték becslési módszer kialakítása, amely lehetővé teszi a szakértői produkciós szabályrendszer kezdeti Q-függvényt leíró fuzzy szabályokká való alakítását és így azok FRIQ-learning rendszer tanulási folyamatába való injektálását.
- Hangolási eljárás kidolgozása, amely lehetővé teszi az előzetesen megadott szakértői tudásbázis hangolását, optimalizálását. A hangolási eljárásnak

alkalmasnak kell lennie a nem feltétlenül helyes szakértői heurisztika (szabályrendszer) negatív hatásainak kompenzálására, a téves vagy nem teljesen helyes szakértői szabályok korrekciójára.

- A Q-függvényt leíró fuzzy szabálybázis szabályszámaának csökkentésére (redukálására) alkalmas módszer kidolgozása, amely a tanulási folyamat közben, a közel ugyanazon információt leíró szabályok összevonásával csökkenti a szabálybázis méretét és így a rendszer komplexitását.
- Módszer kidolgozása a Q-függvényt leíró fuzzy szabálybázis hangolása és redukciója során a szakértői szabályok követésére és hangolást követő kinyerésére. A módszerrel az eredeti és a hangolást követően kinyert szakértői szabálybázis összevethető, a kezdeti szakértői heurisztika helyessége ellenőrizhető, validálható. A hangolás előtt megadott és a hangolás utáni előállt szakértői tudásbázis összehasonlításával következtetni lehet a szakértői szabályok helyességének mértékére. A hangolást követő kismértékű eltérés igazolhatja a szakértői szabályok helyességét, nagyobb mértékű eltérés értelmezhető a kezdeti heurisztika pontosításaként, a jelentős eltérések, vagy az eredeti szabályrendszernek ellentmondó produkciós szabályok pedig utalhatnak a kezdeti heurisztika egyes részeinek helytelenségére. A szabálybázis redukciója során eltűnő szabályok a szakértői heurisztika redundanciájára utalhatnak.

A kutatási téma célkitűzése tehát kettős. Egyrészt olyan fuzzy interpolációs állapotgép viselkedésmodell kidolgozása, amelyben az a priori tudás viszonylag egyszerű módon implementálható, másrészt olyan automatikus hangolási eljárás kidolgozása, amellyel ezen a priori elemeket is tartalmazó modell hiányos minta alapján hangolható. Az így kialakítandó modell és módszer jelentősége amellet, hogy egy nyelvi leírási formából kiindulva (pl. etológiai modell, mint a priori tudás) működtető modell kialakítására alkalmas, megfelelő teljesítmény mérték választása és minták megléte esetén akár etológiai modell hangolására és akár annak validálására is lehetőséget nyújthat.

2 ELŐZMÉNYEK

Ezen fejezet a tudományos előzményeket, azaz az gépi megerősítéses tanulás alapjait és a kutatás alapjául szolgáló FRIQ-learning módszert foglalja össze.

2.1 A MEGERŐSÍTÉSES TANULÁS

A gépi tanulás [17][68] témaköre alapvetően három, nem teljesen élesen elkülöníthető csoportba sorolható, melyek a felügyelt tanulás, a felügyelet nélküli tanulás illetve a megerősítéses tanulás.

A megerősítéses tanulás (*reinforcement learning* - *RL*) [84] esetében a rendszer megerősítési információk alapján alakítja ki a viselkedést, minden lépésben kap visszajelzést, úgynevezett megerősítést (jutalmat vagy büntetést) az adott döntés vagy döntések végrehajtását követően, de ezekből arra nem lehet következtetni, hogy ezt mely döntéssorozatának köszönhetően kapta (nincs külső tanár, aki minden esetben adna visszajelzést arról, hogy mi volt a helyes cselekvés). A megerősítési információk egy jutalomfüggvényből származnak, amely definiálja a rendszer számára, hogy mely esetekben jutalmazhat vagy büntethet. Ezen módszerek próbálkozás típusú (*trial-end-error*) algoritmusok összessége, melyek a megoldásra vonatkozó ismeret nélkül, a környezettől kapott megerősítési információk alapján térképezik fel a rendszer elvárt viselkedését. A jutalomfüggvény által van definiálva az elérendő cél, hogy mely esetben jár nagy jutalom az adott cselekvésért (vagy cselekvéssorozatért), mely döntés (döntéssorozat) milyen mértékben volt helyes, de a megoldás mikéntjére vonatkozóan nincs információ. A tanulási módszer alapötlete tehát, hogy a visszajelzéseket ne csak az ágens (azaz a tanuló entitás) jelenlegi cselekvéseinek kialakítására használják fel, hanem arra is, hogy javítsa a jövőbeli döntésekre irányuló képességet, tehát a tanulás során lépésről-lépésre egyre helyesebben oldja meg az adott feladatot.

2.2 A FRIQ-LEARNING MÓDSZER

A „fuzzy szabály-interpoláció alapú Q-tanulás” (*Fuzzy Rule Interpolation-based Q-learning - FRIQ-learning*) [97][98] a „FIVE” fuzzy szabály-interpolációs módszert alkalmazó megerősítéses tanulási algoritmus. A „FIVE” FRI (*Fuzzy Interpolation in the Vague Environment, bizonytalan környezet alapú fuzzy szabály-interpoláció*) modell alkalmazása következtében a módszer folytonos állapot-akció (és Q-érték) dimenziókkal rendelkezik, a folytonos és interpolált $\tilde{Q}(s, a)$ függvényt a tanulási folyamat közben létrejött fuzzy szabályrendszer írja le, melyben a szabályok a $\tilde{Q}(s, a)$ függvény tartópontjai. A módszer Q-függvény reprezentációja a „FIVE” FRI (*Fuzzy Rule Interpolation, Fuzzy szabály-interpoláció*) modell alkalmazása következtében így kisebb, mint a klasszikus Q-learning módszerek esetében [S11], amely által a FRIQ-learning módszer hatékonyan alkalmazható a klasszikus megerősítéses tanulási alkalmazási példák vagy akár a Pong játék [S14] esetében is.

A rendszer működtető tudásbázisát (R szabálybázis) leíró i -edik ($i \in [1, m]$, m a szabálysám) r_i fuzzy szabály alakja a következő [97][98]:

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

Ahol $\tilde{Q}(s, a)$ a közelített Q-függvény, q^i az i -edik szabály konzekvense. S_j^i ($j \in [1, n]$) a fuzzy halmaza az i -edik szabálynak a j -edik állapot univerzumban, $s \in S$ az n dimenziós állapot megfigyelés az n dimenziós állapot térben, s_j a j -edik ($j \in [1, n]$) dimenziója az állapot megfigyelés s -nek. A^i a fuzzy halmaza az i -edik szabálynak az egydimenziós U akciótérben, $a, a \in U$ pedig a végrehajtott akció. A rendszer állapot-akció univerzuma ($n + 1$) dimenziós, ahol n az állapot dimenziók száma, a további dimenzió pedig az akciótérrel jelöli. Az R szabálybázis m darab $r_i \in R$ ($i \in [1, m]$) fuzzy szabályt tartalmaz.

A „FIVE” FRI modellel közelített $\tilde{Q}(s, a)$ függvény i -edik fuzzy szabályának konzekvense a $(k + 1)$ -edik iterációban a következő [97][98]:

$$q_i^{k+1} = \begin{cases} q_i^k + \Delta\tilde{Q}^{k+1}(\mathbf{s}, a) & \text{ha } (\mathbf{s}, a) = (\mathbf{s}^i, a^i) \\ & \text{valamennyi } i\text{-re,} \\ q_i^k + \Delta\tilde{Q}^{k+1}(\mathbf{s}, a) * (1/\delta_{v,i}^\lambda) / \left(\sum_{i=1}^m 1/\delta_{v,i}^\lambda \right) & \text{egyébként} \end{cases} \quad (2)$$

Ahol $\Delta\tilde{Q}^{k+1}(\mathbf{s}, a)$ a Q -függvény $(k+1)$ -edik iterációbeli frissítési értéke (\mathbf{s}, a) -ban amely a következő módon határozható meg [97][98]:

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

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

Ahol γ a leszámítolási tényező, $\alpha \in [0,1]$ a tanulási ráta, q_i^{k+1} az i -edik szabály singleton konklúziója a $(k+1)$ -edik iterációban, a a végrehajtott akció \mathbf{s} -ben, \mathbf{s}' az új állapot megfigyelés, $g(\mathbf{s}, a, \mathbf{s}')$ a jutalom értéke az $\mathbf{s} \rightarrow \mathbf{s}'$ állapotátmenetre, \tilde{Q}^k a k -adik, \tilde{Q}^{k+1} pedig a $k+1$ -edik iteráció becslt konklúziója a „FIVE” FRI által.

Összegezve, a $\tilde{Q}(\mathbf{s}, a)$ függvény alakja a következőképpen írható fel a „FIVE” FRI modellbe való behelyettesítésével [97][98]:

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

Ahol q^i az i -edik ($i \in [1, m]$) szabály konzekvensé, (\mathbf{s}, a) a crisp megfigyelés, λ a Shepard paraméter, amely értéke jelen esetben megegyezik az antecedens dimenziók számával [90], m pedig a szabálybázisban lévő szabályok száma. A δ_v^i a skálázott (súlyozott) távolság az (\mathbf{s}, a) állapot-akció megfigyelés és az i -edik szabály (\mathbf{s}^i, a^i) állapot-akció antecedense között, amely a következőképpen fejezhető ki [56]:

$$\delta_v^i = \delta_v((s, a), (s^i, a^i)) = \left[\sum_{j=1}^n \left(\int_{s_j^i}^{s_j} v_j(s_j) ds_j \right)^2 + \left(\int_{a^i}^a v(a) da \right)^2 \right]^{1/2} \quad (6)$$

Ahol (s, a) az állapot-akció megfigyelés, (s^i, a^i) az i -edik szabály állapot-akció antecedense, s_j a j -edik ($j \in [1, n]$) dimenziója az n -dimenziós állapot-tér univerzumnak, s_j^i az i -edik szabály j -edik állapot dimenziója, a^i az i -edik szabály akció univerzuma, $v_j(s_j)$ az s_j állapot univerzum skálafüggvénye, a $v(a)$ pedig az U akció univerzum skálafüggvénye. A skálafüggvények által az adott fuzzy halmazok alakja jellemezhető [47][56].

A tanulási fázis kezdetben 2^{n+1} darabszámú, 0 konzekvens értékkel rendelkező fuzzy szabállyal indul, amit az inkrementális szabálybázis építési módszer [94] iterációról-iterációra bővít, illetve hangol. Ezek a kezdeti vagy úgynevezett sarokponti szabályok az $(n + 1)$ -dimenziós hiperkocka sarkaiban, azaz az univerzumok határain helyezkednek el. A továbbiakban a módszer ezt a kezdeti szabálybázist bővíti új szabályokkal vagy azok konzekvensét (Q-értékét) hangolja a (2) formula által, attól függően, hogy új szabály beillesztésére vagy csak a meglévő szabálybázis Q-értékének a frissítésére van-e szükség.

Új szabály szabálybázisba történő beszúrása az ágens környezetéből érkező megerősítési információk és a Q-függvény frissítési értékei ($\Delta\tilde{Q}$) alapján történik. Ha $\Delta\tilde{Q}$ értéke magasabb, mint egy előre meghatározott Q-frissítési limit ($\Delta\tilde{Q} > \varepsilon_Q$) és a létező legközelebbi szabály is távol van az éppen beszúrandó szabály pozíciójához képest, akkor új szabály felvétele történik az adott lehetséges szabálypozícióba. A lehetséges szabálypozíciók meghatározása egy állapot-akció tér rácsháló által történik, mely által az állapot-akció tér csak adott $(s_{k+1} = s_k, \forall k > i, s_{i+1} = \frac{s_i + s_{i+2}}{2})$ pontjaiba illeszthetők be az új szabályok [94]. Abban az esetben, ha $\Delta\tilde{Q}$ értéke kisebb, mint az előre meghatározott Q-frissítési limit ($\Delta\tilde{Q} < \varepsilon_Q$), akkor nem történik új szabály beszúrása a szabálybázisba. Ebben az esetben a teljes szabálybázis konzekvensének, azaz Q-értékének a frissítése (hangolása) valósul meg. Az említett lépések minden egyes iterációban végrehajtásra kerülnek addig, amíg a tanulási fázis (azaz az inkrementális szabálybázis építési fázis) véget nem ér. Akkor áll elő a rendszert működtető, végleges tudásbázis (fuzzy szabályrendszer) és ér véget a tanulási

folyamat, ha már nem kerül új szabály beszúrásra a szabálybázisba és a $\Delta\tilde{Q}$ frissítési értékek relatívan kicsik maradnak.

Az inkrementális szabálybázis építési fázisban létrejött szabálybázis tartalmazhat redundáns szabályokat vagy olyan szabályokat melyeknek kiadódhatnak más, már létező szabályokból. Ezen szabályok a szabálybázisból való végleges törlésével a szabálybázis mérete csökkenthető, csökkentve ez által a szabálybázis komplexitását és a működtető végleges tudásbázis méretét [95][96]. Az elhagyható szabályok megállapítására eredetileg 3 dekrementális tudásbázis redukálási módszerrel (I., II., III.) [95][96] rendelkezik a FRIQ-learning rendszer, amelyek az inkrementális szabálybázis építési fázis után alkalmazhatók opcionálisan. Kifejlesztettem egy további inkrementális szabálybázis redukálási módszert (IV.) [S8], amely klaszterezési eljárásan alapszik.

Ezen módszerek mindegyike a tanulási folyamat végén előállt teljes szabályrendszer szabályait vizsgálja, hogy az egyes szabályok lényegiek (kardinális), vagy kiadódók (redundáns). A szabálybázis redukálási módszerek eltávolítják a redundáns szabályokat a szabályrendszerből, így az eredetivel közel azonos az információt hordozó szabályrendszert alkotnak a lényegi szabályokból. A redukációs módszerek közös jellemzője, hogy a szabályok konzekvens értékét, azaz a Q-értéket vizsgálja. Az I.-III. redukálási módszerek dekrementálisak, azaz a végső redukált szabálybázis a tanulási fázis végén kapott teljes szabálybázis egyes szabályainak elhagyásával jön létre, fokozatosan csökkentve annak méretét. A IV. redukálási módszer inkrementális, azaz a végső redukált szabálybázis a tanulási fázis végén kapott teljes szabálybázisból a feltételezett lényegi szabályok kiemelésével keletkezik. Az egyes szabálybázis redukálási módszerekkel kapott csökkentett méretű szabálybázisok közel ugyanazt a Q-függvényt (irányítási felületet) írják le, mint a redukálás előtti esetben, de kevesebb szabállyal (azaz interpolációs tartóponttal).

A FRIQ-learning módszer működése tehát 2 fő lépésre bontható. Az első fázisban az inkrementális szabálybázis építési módszer [94] iterációról-iterációra bővíti, majd a tanulási fázis végeztével létrehozza a rendszert működtető végleges tudásbázist. A második fázisban a tanulási folyamat végén előállt szabálybázis méretének (szabályainak számának) csökkentésére van lehetőség az adott dekrementális szabálybázis csökkentési módszerek [95][96][S8] (I., II., III., és IV.) opcionális alkalmazása által.

3 ÚJ TUDOMÁNYOS EREDMÉNYEK

Ebben a fejezetben az elért új tudományos eredmények, a szakértői tudásbázissal bővített (heurisztikusan gyorsított) FRIQ-learning módszer (HFRIQ-learning) és a hozzá kapcsolódó tézisek kerülnek bemutatásra.

3.1 SZAKÉRTŐI TUDÁSBÁZIS BEÉPÍTÉSE

Első tézisként javaslatot tettem a szakértő által definiálható előzetes tudásbázis leírási módjára, valamint az a priori tudásbázist leíró fuzzy szabályrendszerre történő kezdeti Q -érték meghatározási módszerre, amely által a kezdeti szakértői tudásbázis beépíthető az FRIQ-learning módszer tanulási folyamatába. A javasolt módszerek alkalmazása által, a „Cart-Pole” és a „Mountain car” megerősítéses tanulási mintapéldák alkalmazásával vizsgáltam az injektált szakértői tudásbázis FRIQ-learning rendszer tanulási hatékonyságára gyakorolt hatását.

A kapott eredmények alapján megállapítható, hogy a fejlesztett módszerek által külső szakértői tudásbázis fuzzy produkciós szabályok formájában injektálható a FRIQ-learning rendszerbe, továbbá egy helyesen megadott, a rendszer szempontjából külső szakértői tudásbázis nagymértékben javíthatja a FRIQ-learning rendszer hatékonyságát, pozitív módon befolyásolja a tanulási fázis konvergencia sebességét. Bebizonyosodott továbbá, hogy az általam javasolt módszerek által külső szakértői heurisztika fuzzy produkciós szabályok formájában megfelelő módon injektálható a FRIQ-learning rendszerbe.

3.1.1 Tézis I.

A FRIQ-learning megerősítéses tanulási rendszer konvergencia sebessége javítható a kezdeti Q -érték szabálybázisba illesztett helyes szakértői produkciós szabályokból képzett Q fuzzy szabályokkal, ahol ezen beillesztett szabályok kezdeti konzekvens Q -értéke a környezet által adható maximális megerősítés érték alapján becsülhető.

I.1. Altézis: *A konvergencia sebesség az esetben is javulhat, ha a felhasznált helyes szakértői produkciós szabályok csak részben fedik le a teljes állapotteret.*

I.2. Altézis: *Amennyiben a felhasznált szakértői produkciós szabályok helytelen szabályokat is tartalmaznak, azaz egyes szabályok esetén az érintett állapotban javasolt akció választása csökkentené a várható megerősítés értékét, a teljes FRIQ-learning rendszer konvergencia sebessége romolhat.*

Az I. tézishez kapcsolódó saját publikációk: [S2], [S4], [S6], [S7], [S15]

3.2 A FRI Q-FÜGGVÉNYT LEÍRÓ SZABÁLYBÁZIS HANGOLÁSA

Második tézisként kidolgoztam egy gradiens módszeren alapuló hangolási (optimalizálási) eljárást, amely alkalmas a tanulási folyamat közben, a tudásbázist leíró Q-függvény tartópontjainak hangolására, azaz a fuzzy szabályrendszer antecedens (állapot-akció) és konzekvens (Q-érték) értékeinek pontosítására. Meghatároztam továbbá a gradiens módszer alkalmazása következtében szükséges, a „FIVE” FRI modellel leírt Q-függvény parciális deriváltjait.

Megállapítottam, hogy az összes szabálypont hangolásával a Q-függvény függvény elromlik, nem alakul ki a függvény elvárt felülete. Ennek oka az lehet, hogy az állapot-akció tér azon pontjainak antecedense és Q-értéke, melyek ritkán kerülnek bejárásra és így frissítésre (a felderítés-kiaknázás következtében), az összes szabálypont hangolása miatt elromlik. Az összes szabálypont egyidejű hangolása ezért nem járható út. A megoldást azt jelentette, hogy csak azon szabályok kerültek hangolásra, amely az éppen aktuális állapot-akció megfigyelési pont közelében található, azaz a legközelebb helyezkednek el ahhoz. A javasolt módszer megoldotta a fentebb említett problémát és helyes irányban hangolja a Q-függvény szabálypontjait, nem változtatva azon szabályokat melyek a aktuális állapot-akció megfigyelési ponttól távol, a kevésbé látogatott területekre esnek.

A rendszerbe történő új szabályok felvétele (majd a későbbi redukálása) következtében meghatároztam továbbá a szabályok között definiálható távolságot és ez alapján egyfajta szabályközelséget, amelynek alapja az egyes dimenziókra meghatározott távolságkülbségek.

A javasolt módszerek hatékonyságának vizsgálatát gyakorlati mintapéldák és az így kapott Q -függvény felületek segítségével vizsgálatam.

3.2.1 Tézis II.

A HFRIQ-learning megerősítéses tanulási rendszer inkrementális szabálybázis építési fázisában a Q -függvényét leíró fuzzy szabálybázis szabályainak (fuzzy Q -szabályok) antecedensei és konzekvenszei gradiens módszerrel optimalizálhatók, hangolhatók. Az aktuális állapot-akció pontban egy új szabály beillesztésének feltétele a már meglévő szabályoktól vett távolsága és a Q -függvény frissítésének mértéke alapján meghatározható.

II.1. Altézis: *Amennyiben nincs olyan fuzzy Q -szabály, melynek antecedense valamennyi antecedens dimenzióban vett távolsága kisebb az egyes dimenziókra meghatározott távolságkülönbélnél és a Q -függvény frissítésének mértéke nagyobb, mint egy küszöbérték, úgy az aktuális állapot-akció pontba egy új szabály kerül beillesztésre. Ellenkező esetben a meglévő fuzzy Q -szabályok kerülnek hangolásra.*

II.2. Altézis: *Abban az esetben, ha szabálybázisba illesztett kezdeti szakértői szabályrendszer helytelen szakértői produktív szabályokat is tartalmaz, akkor azok a tanulási fázisban a szabálybázis többi szabályával együtt hangolhatók, korrigálhatók.*

II.3. Altézis: *A HFRIQ-learning megerősítéses tanulási rendszer tudásbázisának hangolása során az állapot-akció tér ritkán bejárt területein lévő fuzzy Q -szabályok elhangolódása csökkenthető, ha az összes fuzzy szabálypont egyidejű hangolása helyett, csak azon szabályok kerülnek hangolásra, amely az éppen aktuális állapot-akció megfigyelési pont közelében található.*

A II. tézishez kapcsolódó saját publikációk: [S1], [S9], [S10], [S12]

3.3 SZABÁLYBÁZIS REDUKCIÓ

Harmadik tézisként kidolgoztam egy olyan tudásbázis csökkentési (szabálybázis redukálási) módszert, amely a rendszer tanulási folyamata közben alkalmazható. A második tézisként javasolt szabálybázis hangolása során a

szabálypontok vándorlása miatt előfordulhat olyan eset, amikor több szabály kerül egymáshoz közel az $(n + 2)$ -dimenziós térben. Abban az esetben, ha az egymáshoz közel kerülő szabályok nagyon hasonló információt írnak le (azaz ilyen szempontból redundánsak), a szabályok egyesíthetők, a szabályrendszer mérete csökkenthető. A javasolt szabálybázis redukálási módszer a rendszer tanulási fázisa közben a fuzzy szabályok közötti távolságok (és távolságküszöbök) alapján egyesíti az egymáshoz hasonló szabályokat, csökkentve a rendszer tudásbázisának méretét. A javasolt távolság alapú szabálybázis redukálási módszer hatékonyságát különböző futási esetek felhasználásával vizsgáltam.

Kidolgoztam egy olyan módszert is amely által az egyes szakértői szabályok változása nyomon követhető a tanulási folyamat során, így a tanulási folyamat végeztél visszaolvasott szakértői szabályok helyességére ezáltal következtetni lehet.

Kidolgozásra került továbbá egy olyan tudásbázis (szabálybázis) redukálási módszer is, amely hierarchikus klaszterezési eljárással a tanulási folyamatot követően alkalmas a Q-függvényt leíró szabálybázis méretét csökkenteni.

3.3.1 Tézis III.

A HFRIQ-learning megerősítéses tanulási módszer tudásbázisának mérete a tanulási folyamat során csökkenthető a hasonló fuzzy Q-szabályok összevonásával. A fuzzy szabályok hasonlósága becsülhető antecedenseik és konzekvenseik távolságával.

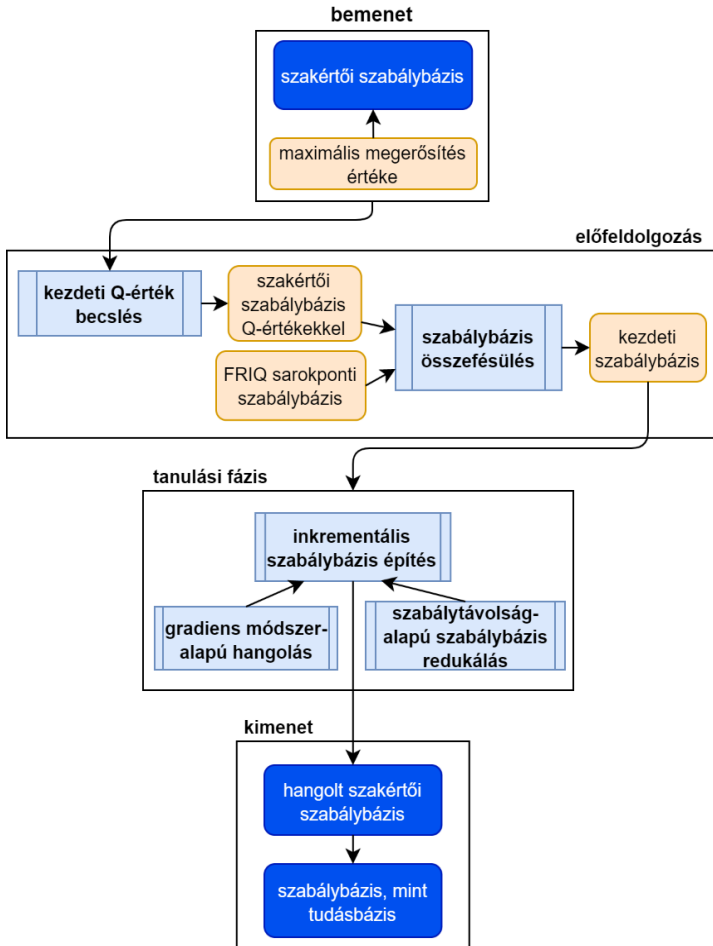
III.1. Altézis: *A szabályok hangolása és összevonása során követhető a szabályok típusa, így a tanulási folyamat végeztével a HFRIQ-learning Q fuzzy szabályrendszeréből a kezdeti szakértői szabályok visszanyerhetők. Az eredetileg megadott és a hangolást követően visszanyert szakértői szabályok összevetésével következtetni lehet a kezdeti szakértői szabályok helyességére.*

III.2. Altézis: *A hierarchikus klaszterezési módszer alkalmas a HFRIQ-learning tanulási fázisának végeztével előállt Q-függvényt leíró fuzzy szabályrendszer méretének csökkentésére.*

A III. tézishez kapcsolódó saját publikációk: [S3], [S5], [S8], [S13]

3.4 A HFRIQ-LEARNING

A fejlesztett, heurisztikusan gyorsított FRIQ-learning (HFRIQ-learning) rendszer felépítésének blokkvázlatát az alábbi 1. ábra szemlélteti.



1. ábra: A heurisztikusan gyorsított FRIQ-learning (HFRIQ-learning) rendszer felépítése

A javasolt HFRIQ-learning módszer működésének hatékonyságát egy egyetlen állapot- és egyetlen akcióváltozóval rendelkező mintapélda és a „Mountain car” illetve a „Cart-Pole” klasszikus megerősítéses tanulási alkalmazáspéldákon keresztül vizsgáltam.

A kapott futási eredmények alapján megállapítható, hogy a javasolt (és fejlesztett) módszerek következtében a rendszer tanulási fázisának konvergencia sebessége javulhat, de csak olyan esetekben mikor a szakértői heurisztika helyes. Ellenkező esetben, mikor helytelen szakértői szabályok is injektálásra kerülnek, a tanulási módszer továbbra is konvergál, de a szabályok hangolásához (optimalizálásához) több epizódra (és így több iterációra) van szükség.

4 ÖSSZEFOGLALÁS

A disszertáció tárgya a gépi tanuláshoz, azon belül a megerősítéses tanulás témaköréhez kapcsolódik, amely egy Fuzzy szabály-interpoláción alapú Q-learning (FRIQ-learning) módszer szakértői tudásbázissal való kibővítését, a rendszerbe injektált szakértői tudásbázis hangolását (optimalizálását), valamint a tanulási folyamat közben létrejött tudásbázis méretének a csökkentését (redukálását) takarja.

Az értekezés első fejezetében bemutatásra került a kutatás célkitűzése, a 2.-3. fejezetekben részletezésre került a Fuzzy logika és a megerősítéses tanulás témaköre, valamint a kutatás alapjául szolgáló Fuzzy szabály-interpoláción alapú Q-learning (FRIQ-learning) módszer. A 4. fejezetben került bemutatásra a javasolt „Heurisztikusan gyorsított FRIQ-learning” (HFRIQ-learning) módszer, a hozzá kapcsolódó kutatás eredményeit összefoglaló tézisek és saját publikációk.

Első tézisként javaslatot tettem a szakértő által definiálható előzetes tudásbázis leírási módjára, valamint az a priori tudásbázist leíró fuzzy szabályrendszerre történő kezdeti Q-érték meghatározási módszerre, amely által a kezdeti szakértői tudásbázis beépíthető az FRIQ-learning módszer tanulási folyamatába. A javasolt módszerek alkalmazása által vizsgáltam továbbá az injektált szakértői tudásbázis FRIQ-learning rendszer tanulási hatékonyságára gyakorolt hatását.

Második tézisként kidolgoztam egy gradiens módszeren alapuló hangolási (optimalizálási) eljárást, amely alkalmas a tanulási folyamat közben, a tudásbázist leíró Q-függvény tartópontjainak hangolására, azaz a fuzzy szabályrendszer antecedens (állapot-akció) és konzekvens (Q-érték) értékeinek pontosítására.

Harmadik tézisként kidolgoztam egy olyan tudásbázis csökkentési (szabálybázis redukálási) módszert, amely a rendszer tanulási folyamata közben alkalmazható. A második tézisként javasolt szabálybázis hangolása során a szabálypontok vándorlása miatt előfordulhat olyan eset, amikor több szabály kerül egymáshoz közel az $(n + 2)$ -dimenziós térben. Abban az esetben, ha az egymáshoz közel kerülő szabályok nagyon hasonló információt írnak le (azaz ilyen szempontból redundánsak), a szabályok egyesíthetők, a szabályrendszer mérete csökkenthető. A javasolt szabálybázis redukálási módszer a rendszer

tanulási fázisa közben a fuzzy szabályok közötti távolságok (és távolságkülbszöbök) alapján egyesíti az egymáshoz hasonló szabályokat, csökkentve a rendszer tudásbázisának méretét. Kidolgozásra került továbbá egy olyan tudásbázis redukálási módszer is, amely hierarchikus klaszterezési eljárással a tanulási folyamatot követően képes a Q-függvényt leíró szabálybázis méretét csökkenteni.

További kutatási terv egy olyan módszer kidolgozása, amely alkalmas lehet a szakértő által megadott kezdeti, állapot-akció típusú fuzzy szabályok és a tanulási (hangolási) folyamat végeztével előállt optimalizált szakértői szabályok összehasonlítására, információt adhat arról, hogy a kezdeti szakértői szabályrendszer milyen mértékben voltak helyesek. Azaz alkalmas a kezdeti szakértői heurisztika helyességének igazolására validálására. További fejlesztési terv az előzetes szakértői tudásbázis definiálásának egyszerűsítése egy fuzzy viselkedésleíró nyelv [74] alkalmazásával. Az így kialakítandó modellek és módszerek jelentősége amellet, hogy egy nyelvi leírási formából kiindulva (például etológiai modell [85], mint a priori tudás) valamilyen rendszert közvetlenül működtető modelliként használhatók, megfelelő teljesítmény mérték választása és minták megléte esetén a kezdeti szakértői heurisztika validálására is lehetőséget nyújthatnak.

5 SUMMARY

The subject of the dissertation is related to the machine learning area, especially to the reinforcement learning, which includes extending the Fuzzy Rule Interpolation-based Q-learning (FRIQ-learning) method by embedding a priori expert knowledge, tuning the injected expert knowledge (optimization), and reducing the size of the rule-base representing the knowledge base during and after the incremental learning process.

In the first section of the dissertation, the main goal of the study is summarized and the related literature is reviewed. The next two chapters briefly introduce the fuzzy logic, the reinforcement learning and the Fuzzy Rule Interpolation-based Q-learning (FRIQ-learning). The fourth section introduces the suggested "Heuristically Accelerated FRIQ-learning" method, the related theses and our papers supporting them.

In the first thesis, I proposed a way to describe the preliminary (a priori) knowledge base which is defined by a human expert as a set of production (state-action) fuzzy rules. I also suggest a way for embedding the expert defined rules to the initial rule-base of the FRIQ-learning, by reformulating the state-action fuzzy rules to state-action-Q-value fuzzy rules required by the FRIQ-learning. The way as the required initial Q-values of the rules are determined is also part of the first thesis. Furthermore, I investigated the effect of the injected expert knowledge base on the learning efficiency of the FRIQ-learning system applying the proposed methods.

My second thesis, is a gradient descent-based optimization method for optimizing the antecedents (state-action) and consequent (Q-value) parameters of the fuzzy rule-base describing the Q-function.

My third thesis, is a Q-function fuzzy rule-base reduction method, based on merging the similar rules during the learning phase. During the learning phase, rule antecedents could shift to be close to each other, and in the case if the rule consequents are also close, the rules are similar. This case the similar rules can be merged to reduce the rule-base size with a negligible change in the Q-function described. This thesis also describes a method based on hierarchical clustering for reducing the Q-function fuzzy rule-base size after the learning phase.

Further research will focus on the development of a method for validating the preliminary expert knowledge base by comparing the initial expert fuzzy rules with the rules fetched back from the optimized Q-function fuzzy rule-base after the learning phase. Another future goal is to apply a fuzzy behavior description language (FBDL, introduced in [74]), for simplifying and standardizing the form as the initial state-action fuzzy expert rules are specified. These models and methods could support the validation of practical heuristic expert models like ethological models [85].

5.1 THESIS I.

The convergence speed of the FRIQ-learning reinforcement learning system can be improved by embedding Q fuzzy rules created from valid expert production rules into the initial Q-value rule-base, where the initial consequent Q-values of the embedded rules can be estimated based on the maximum reinforcement value that can be provided by the environment.

I.1. Subthesis: *The convergence speed can be improved even if the embedded valid expert production rules are only partly covering the complete state space.*

I.2. Subthesis: *If the expert production rules are also containing invalid rules, i.e., in some rules, the action suggested in the affected state would decrease the expected reinforcement value, it may negatively effects the convergence speed of the entire FRIQ-learning system.*

My publications related to Thesis I: [S2], [S4], [S6], [S7], [S15]

5.2 THESIS II.

In the incremental rule-base construction phase of the HFRIQ-learning reinforcement learning system, the antecedents and consequents of the rules of the fuzzy rule-base describing the Q-function (fuzzy Q-rules) can be optimized and tuned using a gradient based method. The conditions for inserting a new rule at the current state-action point can be determined based on the distance from the existing rules and the extent of the Q-function update.

II.1. Subthesis: *If there is no fuzzy Q -rule, whose distances in all antecedent dimensions are smaller than the distance thresholds determined for each dimensions and the Q -function update is also greater than a threshold value, then a new rule is inserted to the current state-action position. Otherwise, the existing fuzzy Q -rules are tuned.*

II.2. Subthesis: *If the initial expert rule system, which is embedded into the rule-base also contains incorrect expert production rules, then they can be tuned and corrected together with the other rules of the rule-base during the learning phase.*

II.3. Subthesis: *During the tuning of the knowledgebase of the HFRIQ-learning reinforcement learning system, the detuning of fuzzy Q -rules in the sparsely explored areas of the state-action space can be reduced by tuning the rules that are located close to the current state-action observation point only, instead of the simultaneous tuning of all the fuzzy rule points.*

My publications related to Thesis II: [S1], [S9], [S10], [S12]

5.3 THESIS III.

During the learning process the size of the knowledgebase of the HFRIQ-learning reinforcement learning method can be reduced by merging the similar fuzzy Q -rules. The similarity of fuzzy rules can be estimated by the distance between their antecedents and consequents.

III.1. Subthesis: *During the tuning and merging of rules, the type of the rules can be tracked, thus at the end of the learning process from the Q fuzzy rule system of the HFRIQ-learning, the initial expert rules can be recovered. By comparing the originally defined and the recovered expert rules after tuning, it is possible to evaluate the correctness of the initial expert rules.*

III.2. Subthesis: *At the end of the HFRIQ-learning phase, the hierarchical clustering method is suitable for reducing the size of the fuzzy rule-base describing the Q -function.*

My publications related to Thesis III: [S3], [S5], [S8], [S13]

6 AZ ÉRTEKEZÉS IRODALOMJEGYZÉKE

- [1] Almadi, A. I., Al Mamlook, R. E., Almarhabi, Y., Ullah, I., Jamal, A., & Bandara, N. (2022). A fuzzy-logic approach based on driver decision-making behavior modeling and simulation. *Sustainability*, 14(14), 8874.
- [2] Appl, M.: Model-based Reinforcement Learning in Continuous Environments. Ph.D. thesis, Technical University of München, München, Germany, dissertation.de, Verlag im Internet (2000)
- [3] Arulkumaran, Kai, et al. "Deep reinforcement learning: A brief survey." *IEEE Signal Processing Magazine* 34.6 (2017): 26-38.
- [4] Baranyi, P., Kóczy, L. T., Gedeon, T. D.: A Generalized Concept for Fuzzy Rule Interpolation, *IEEE Trans. on Fuzzy Systems*, vol. 12, No. 6, 2004, pp 820-837.
- [5] Baranyi, P., Kóczy, L. T.: A General and Specialised Solid Cutting Method for Fuzzy Rule Interpolation, in *Journal BUSEFAL, URA-CNRS, Vol. 66., Toulouse, France, 1996*, pp. 13-22.
- [6] Baranyi, P., Mizik, S., Kóczy, L.T., Gedeon, T. and Nagy, I.: Fuzzy Rule Base Interpolation Based on Semantic Revision, in *Proceedings of the IEEE International Conference on System Man and Cybernetics (IEEE SMC'98)*, San Diego, USA, 1998, pp.1306-1311.
- [7] Bartók, Roland, and József Vásárhelyi. "A fuzzy rule interpolation base algorithm implementation on different platforms." *Proceedings of the 2015 16th International Carpathian Control Conference (ICCC)*. IEEE, 2015.
- [8] Bartók, Roland, and József Vásárhelyi. "Fuzzy Rule Interpolation Based Object Tracking and Navigation for Social Robot." *Vehicle and Automotive Engineering*. Springer, Cham, 2018.
- [9] Bartók, Roland, and József Vásárhelyi. "Parallelization of FIVE method on multicore embedded system." *2018 19th International Carpathian Control Conference (ICCC)*. IEEE, 2018.
- [10] Bellman, R. (1957). A Markovian decision process. *Journal of mathematics and mechanics*, 679-684.
- [11] Bellman, R. E.: *Dynamic Programming*. Princeton University Press, Princeton, NJ 1957

- [12] Bellman, Richard, and Robert Kalaba. "On the role of dynamic programming in statistical communication theory." *IRE Transactions on Information Theory* 3.3 (1957): 197-203.
- [13] Berenji, H.R.: Fuzzy Q-Learning for Generalization of Reinforcement Learning. Proc. of the 5th IEEE International Conference on Fuzzy Systems, pp. 2208-2214., 1996
- [14] Bianchi, Reinaldo AC, Carlos HC Ribeiro, and Anna Helena Reali Costa. "Heuristically Accelerated Reinforcement Learning: Theoretical and Experimental Results." *ECAI*. 2012.
- [15] Bianchi, Reinaldo AC, Carlos HC Ribeiro, and Anna HR Costa. "Accelerating autonomous learning by using heuristic selection of actions." *Journal of Heuristics* 14.2 (2008): 135-168.
- [16] Bianchi, Reinaldo AC, Carlos HC Ribeiro, and Anna HR Costa. "Heuristically Accelerated Q-Learning: a new approach to speed up Reinforcement Learning." *Brazilian Symposium on Artificial Intelligence*. Springer, Berlin, Heidelberg, 2004.
- [17] Bonaccorso, Giuseppe. *Machine learning algorithms*. Packt Publishing Ltd, 2017.
- [18] Bonarini, A.: Delayed Reinforcement, Fuzzy Q-Learning and Fuzzy Logic Controllers. In Herrera, F., Verdegay, J. L. (Eds.) *Genetic Algorithms and Soft Computing*, (Studies in Fuzziness, 8), Physica-Verlag, Berlin, D, (1996), pp. 447-466.
- [19] Broekens, Joost, Koen Hindriks, and Pascal Wiggers. "Reinforcement learning as heuristic for action-rule preferences." *International Workshop on Programming Multi-Agent Systems*. Springer Berlin Heidelberg, 2010.
- [20] Brunton, Steven L., and J. Nathan Kutz. *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press, 2019.
- [21] Brys, Tim. *Reinforcement Learning with Heuristic Information*. Diss. PhD thesis, PhD thesis, Vrije Universitet Brussel, 2016.
- [22] Chaudhari, Swati R., and Manoj E. Patil. "Comparative analysis of fuzzy inference systems for air conditioner." *International Journal of Advanced computer research* 4.4 (2014): 922.
- [23] Chiou, C. B., Chiou, C. H., Chu, C. M., & Lin, S. L. (2009). The application of fuzzy control on energy saving for multi-unit room air-conditioners. *Applied thermal engineering*, 29(2-3), 310-316.

- [24] Csaba, Johanyák Zsolt, and Kovács Szilveszter. "A fuzzy tagsági függvény megválasztásáról." *A GAMF közleményei, Kecskemét, XIX. évfolyam*, ISSN: 0230-6182.
- [25] D. Shepard, "A two dimensional interpolation function for irregularly spaced data", Proc. 23rd ACM Internat. Conf., 1968, pp. 517-524.
- [26] Dubios, D., Ostasiewicz, W., Prade, H.: Fuzzy Sets: History and Basic Notions, in: Fundamentals of Fuzzy Sets, ISBN 978-0-7923-7732-0, Kluwer Academic, 2000
- [27] Duchi, John, Elad Hazan, and Yoram Singer. "Adaptive subgradient methods for online learning and stochastic optimization." *Journal of machine learning research* 12.7 (2011).
- [28] FIVE FRI MATLAB Toolbox: <http://fri.uni-miskolc.hu/>
- [29] François-Lavet, Vincent, et al. "An introduction to deep reinforcement learning." *arXiv preprint arXiv:1811.12560* (2018).
- [30] Fuchida, Takayasu, Kathy Thi Aung, and Atsushi Sakuragi. "A study of Q-learning considering negative rewards." *Artificial Life and Robotics* 15.3 (2010): 351-354.
- [31] G. Tesauro et al. Temporal difference learning and td-gammon. *Communications of the ACM*, 38(3):58–68, 1995.
- [32] Glorennec, P. Y., & Jouffe, L. (1997, July). Fuzzy Q-learning. In *Proceedings of 6th international fuzzy systems conference* (Vol. 2, pp. 659-662). IEEE.
- [33] Hailu, Getachew, and Gerald Sommer. "Embedding knowledge in reinforcement learning." *International Conference on Artificial Neural Networks*. Springer, London, 1998.
- [34] Haji, Saad Hikmat, and Adnan Mohsin Abdulazeez. "Comparison of optimization techniques based on gradient descent algorithm: A review." *PalArch's Journal of Archaeology of Egypt/Egyptology* 18.4 (2021): 2715-2743.
- [35] Hindriks, K. V., De Boer, F. S., Van Der Hoek, W., & Meyer, J. J. C. (2000, July). Agent programming with declarative goals. In *International Workshop on Agent Theories, Architectures, and Languages* (pp. 228-243). Springer, Berlin, Heidelberg.
- [36] Horiuchi, T., Fujino, A., Katai, O., & Sawaragi, T. (1996, September). Fuzzy interpolation-based Q-learning with continuous states and actions. In

- Proceedings of IEEE 5th International Fuzzy Systems* (Vol. 1, pp. 594-600). IEEE.
- [37] J. Dombi. A general class of fuzzy operators, the De Morgan class of fuzzy operator and fuzziness measures induced by fuzzy operators. *Fuzzy Sets and Systems*, 8(2):149–163, 1982.
- [38] Jenei, S.: Interpolation and Extrapolation of Fuzzy Quantities revisited - (I). An Axiomatic Approach, in *Soft Computing*, ISSN: 1432-7643, Vol. 5, 2001, pp. 179-193.
- [39] Johanyák, Zs. Cs. and Kovács, Sz.: Fuzzy Rule Interpolation Based on Polar Cuts, in *Computational Intelligence, Theory and Applications*, Springer Berlin Heidelberg, 2006, pp. 499-511.
- [40] Johanyák, Zs. Cs. and Kovács, Sz.: Fuzzy Rule Interpolation by the Least Squares Method, in *Proceedings of the 7th International Symposium of Hungarian Researchers on Computational Intelligence (HUCI 2006)*, Budapest, Hungary, 2006, pp. 495-506.
- [41] Johanyák, Zs. Cs. and Kovács, Sz.: Vague Environment-based Two-step Fuzzy Rule Interpolation Method, in *Proceedings of the 5th Slovakian-Hungarian Joint Symposium on Applied Machine Intelligence and Informatics (SAMI 2007)*, Poprad, Slovakia, 2007, pp. 189-200.
- [42] Johanyák, Zsolt Csaba, and Szilveszter Kovács. "A brief survey and comparison on various interpolation based fuzzy reasoning methods." *Acta Polytechnica Hungarica* 3.1 (2006): 91-105.
- [43] Jose Antonio Martin H. PhD, *Software Tools for Reinforcement Learning, Artificial Neural Networks and Robotics (Matlab and Python)*
- [44] Kennedy, James, and Russell Eberhart. "Particle swarm optimization." *Proceedings of ICNN'95-international conference on neural networks*. Vol. 4. IEEE, 1995.
- [45] Kim, M. S., Hong, G. G., & Lee, J. J. (1999, October). Online fuzzy Q-learning with extended rule and interpolation technique. In *Proceedings 1999 IEEE/RSJ International Conference on Intelligent Robots and Systems. Human and Environment Friendly Robots with High Intelligence and Emotional Quotients (Cat. No. 99CH36289)* (Vol. 2, pp. 757-762). IEEE.
- [46] King, P.J., Mamdani, E.H.: The application of fuzzy control systems to industrial processes, *Automatica*, Vol. 13, Issue 3, May 1977, pp. 235–242.

- [47] Klawonn, F.: Fuzzy Sets and Vague Environments, in Fuzzy Sets and Systems, Vol. 66, 1994, pp. 207-221.
- [48] Kochenderfer, Mykel J., and Tim A. Wheeler. *Algorithms for optimization*. Mit Press, 2019.
- [49] Kóczy, L. T. and Hirota, K.: Size reduction by interpolation in fuzzy rule bases, IEEE Transactions on Systems, Man, and Cybernetics, vol. 27, 14 - 25, 1997.
- [50] Kóczy, L. T., Hirota, K., Rule interpolation by α -level sets in fuzzy approximate reasoning, In J. BUSEFAL, Automne, URA-CNRS. Vol. 46. Toulouse, France, 1991, pp. 115-123.
- [51] Kóczy, L. T., Sugeno, M.: Explicit functions of fuzzy control systems, in International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 4:515-535, 1996.
- [52] Kóczy, L.T., Hirota, K. and Gedeon, T. D.: Fuzzy rule interpolation by the conservation of relative fuzziness, in Journal of Advanced Computational Intelligence, Vol. 4/1, 2000, pp. 95-101.
- [53] Kóczy, László T., and Domonkos Tikk. "Fuzzy rendszerek." TypoTEX, Budapest (2000).
- [54] Kovács, Sz., Kóczy, L. T.: Approximate Fuzzy Reasoning Based on Interpolation in the Vague Environment of the Fuzzy Rule base as a Practical Alternative of the Classical CRI. Proceedings of the 7th International Fuzzy Systems Association World Congress, Prague, Czech Republic, 1997, pp. 144-149.
- [55] Kovács, Sz., Kóczy, L. T.: The use of the concept of vague environment in approximate fuzzy reasoning. Fuzzy Set Theory and Applications, Tatra Mountains Mathematical Publications, Mathematical Institute Slovak Academy of Sciences, Bratislava, Slovak Republic, vol.12, 1997, pp. 169-181.
- [56] Kovács, Sz.: Extending the Fuzzy Rule Interpolation 'FIVE' by Fuzzy Observation, Advances in Soft Computing, Computational Intelligence, Theory and Applications, Bernd Reusch (Ed.), Springer Germany, ISBN 3-540-34780-1, 2006, pp. 485-497.
- [57] Kovács, Sz.: New Aspects of Interpolative Reasoning. Proceedings of the 6th. International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Granada, Spain, 1996, pp. 477-482.

- [58] Kovacs, Szilveszter. "Fuzzy Rule Interpolation in Practice." *SCIS & ISIS SCIS & ISIS 2006*. Japan Society for Fuzzy Theory and Intelligent Informatics, 2006.
- [59] Kovács, Szilveszter: Fuzzy logic control, M.Phil. theses, Technical University of Budapest, Faculty of Informatics and Electrical Engineering, Budapest, Branch of Computer Science, p.116, (1993).
- [60] L. T. Kóczy, Computational complexity of various fuzzy inference algorithms, *Annales Univ. Sci. Budapest, Sect. Comp.* 12, pp 151-158, (1991)
- [61] Larsen, P. M.: Industrial application of fuzzy logic control. *Int. J. of Man Machine Studies*, (12) 4, 3-10., 1980
- [62] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." *nature* 521.7553 (2015): 436-444.
- [63] Li, Yuxi. "Deep reinforcement learning: An overview." *arXiv preprint arXiv:1701.07274* (2017).
- [64] Mamdani, E. H., Assilian, S.: An experiment in linguistic synthesis with a fuzzy logic controller. *Int. J. of Man Machine Studies*, (7), 1 -13., 1975
- [65] Matignon, Laëtitia, Guillaume J. Laurent, and Nadine Le Fort-Piat. "Reward function and initial values: better choices for accelerated goal-directed reinforcement learning." *International Conference on Artificial Neural Networks*. Springer, Berlin, Heidelberg, 2006.
- [66] Mazyavkina, N., Sviridov, S., Ivanov, S., & Burnaev, E. (2021). Reinforcement learning for combinatorial optimization: A survey. *Computers & Operations Research*, 134, 105400.
- [67] Mesiar, R.: Triangular Norms - An Overview, *Computational Intelligence in Theory and Practice, Advances in Soft Computing Vol. 8*, 2001, pp. 35-54.
- [68] Mitchell, T. M., & Mitchell, T. M. (1997). *Machine learning* (Vol. 1, No. 9). New York: McGraw-hill.
- [69] Mizik, S., Baranyi, P., Korondi, P. and Kóczy, L.T.: Comparison of fuzzy interpolation techniques, 4th Meeting of the Euro Working Group on Fuzzy Sets and 2nd International Conference on Soft and Intelligent Computing (EUROFUSE-SIC'99), 1999, Budapest, pp.544-549.
- [70] Mizik, S., Szabó, D. and Korondi, P.: Survey on fuzzy interpolation techniques, in *Proceedings of the IEEE International Conference on Intelligent Engineering Systems*, Poprad, Slovakia, 1999, pp. 587–592.

- [71] Murphy, R. R. (2001). Fuzzy logic for fusion of tactical influences on vehicle speed control. Fuzzy logic techniques for autonomous vehicle navigation, 73-96.
- [72] Oh, Chi-Hyon, Tomoharu Nakashima, and Hisao Ishibuchi. "Initialization of Q-values by fuzzy rules for accelerating Q-learning." *1998 IEEE International Joint Conference on Neural Networks Proceedings. IEEE World Congress on Computational Intelligence (Cat. No. 98CH36227)*. Vol. 3. IEEE, 1998.
- [73] Oh, Chi-Hyon, Tomoharu Nakashima, and Hisao Ishibuchi. "Initialization of Q-values by fuzzy rules for accelerating Q-learning." *1998 IEEE International Joint Conference on Neural Networks Proceedings. IEEE World Congress on Computational Intelligence (Cat. No. 98CH36227)*. Vol. 3. IEEE, 1998.
- [74] 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.
- [75] Pourhassan, Mojgan, and Nasser Mozayani. "Incorporating expert knowledge in Q-learning by means of fuzzy rules." *Computational Intelligence for Measurement Systems and Applications, 2009. CIMSA'09. IEEE International Conference on*. IEEE, 2009.
- [76] R. R. Yager. On the measure of fuzziness and negation. part ii: Lattices. *Information and Control*, 44(3):236–260, 1980.
- [77] R. S. Sutton. Learning to predict by the methods of temporal differences. *Machine learning*, 3(1):9–44, 1988
- [78] Ribeiro, Carlos HC. "Embedding a priori knowledge in reinforcement learning." *Journal of Intelligent and Robotic Systems* 21.1 (1998): 51-71.
- [79] Rummery, G. A., Niranjan, M.: On-line Q-learning using connectionist systems. CUED/F-INFENG/TR 166, Cambridge University, UK., 1994
- [80] Russell Stuart, J., & Norvig, P. (2009). *Artificial intelligence: a modern approach*. Prentice Hall.
- [81] S. J. Bradtke and A. G. Barto. Linear least-squares algorithms for temporal difference learning. *Machine learning*, 22(1):33–57, 1996.
- [82] Santra, Santanu, Jun-Wei Hsieh, and Chi-Fang Lin. "Gradient descent effects on differential neural architecture search: A survey." *IEEE Access* 9 (2021): 89602-89618.

- [83] Sugeno, M.: An introductory survey of fuzzy control. *Information Science*, (36), 1985, pp. 59-83.
- [84] Sutton, R. S., Barto, A. G.: *Reinforcement Learning: An Introduction*, MIT Press, Cambridge (1998)
- [85] Sz. Kovács, D. Vincze, M. Gácsi, Á. Miklósi, P. Korondi, "Ethologically inspired robot behavior implementation", *Proc. 4th International Conference on Human System Interaction (HSI 2011)*, Keio University, Yokohama, Japan, 2011, pp. 64–69.
- [86] Takagi, T., Sugeno, M.: Fuzzy identification of systems and its applications to modeling and control. *IEEE Trans. on SMC*, (15), 1985, pp. 116-132.
- [87] Tan, Ming. "Multi-agent reinforcement learning: Independent vs. cooperative agents." *Proceedings of the tenth international conference on machine learning*. 1993.
- [88] Tesauro, G.: Temporal difference learning and TD-Gammon, in *Communications of the ACM*, 1995, 38.3: pp. 58-68.
- [89] Tikk, D. and Baranyi, P.: Comprehensive analysis of a new fuzzy rule interpolation method, in *IEEE Transactions on Fuzzy Systems*, Vol. 8, 2000, pp. 281-296.
- [90] Tikk, D., Joó, I., Kóczy, L., Várlaki, P., Moser, B., & Gedeon, T. D. (2002). Stability of interpolative fuzzy KH controllers. *Fuzzy Sets and Systems*, 125(1), 105-119.
- [91] 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.
- [92] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529, 2015
- [93] Vass, G., Kalmár, L. and Kóczy, L. T.: Extension of the fuzzy rule interpolation method, in *Proceedings of the International Conference on Fuzzy Sets Theory Applications (FSTA '92)*, Liptovsky Mikulas, Czechoslovakia, 1992, pp. 1-6.
- [94] Vincze, D., Kovács, Sz.: Incremental Rule Base Creation with Fuzzy Rule Interpolation-Based Q-Learning, I. J. Rudas et al. (Eds.), *Computational Intelligence in Engineering, Studies in Computational*

- Intelligence, Volume 313/2010, Springer-Verlag, Berlin Heidelberg, 2010, pp. 191-203.
- [95] Vincze, D., Kovács, Sz.: Reduced Rule Base in Fuzzy Rule Interpolation-based Q-learning, Proceedings of the 10th International Symposium of Hungarian Researchers on Computational Intelligence and Informatics, CINTI 2009, November 12-14, 2009, Budapest Tech, Budapest, pp. 533-544.
- [96] Vincze, D., Kovács, Sz.: Rule-Base Reduction in Fuzzy Rule Interpolation-Based Q-Learning, Recent Innovations in Mechatronics (RIiM) Vol. 2. (2015) No. 1-2.
- [97] Vincze, Dávid, and Szilveszter Kovács. "Fuzzy rule interpolation-based Q-learning." 2009 5th International Symposium on Applied Computational Intelligence and Informatics. IEEE, 2009.
- [98] Vincze, Dávid.: "Fuzzy Rule Interpolation-based Q-learning." PhD dissertation, 2013.
- [99] Watkins, C. J. C. H., Dayan, P.: Q-learning, in Machine Learning, Vol. 8 (3/4), 1992., pp. 279-292.
- [100] Wong, K. W., Gedeon, T. D. and Tikk, D.: An improved multidimensional α -cut based fuzzy interpolation technique, In Proc. Int. Conf. Artificial Intelligence in Science and Technology (AISAT'2000) , Hobart, Australia, 2000, pp. 29–32.
- [101] Yan, S., Mizumoto, M. and Qiao, W. Z.: An Improvement to Kóczy and Hirota's Interpolative Reasoning in Sparse Fuzzy Rule Bases, in International Journal of Approximate Reasoning, Vol. 15, 1996, pp. 185-201.
- [102] Zadeh, L. A.: Fuzzy Sets, in Information and Control, Vol. 8, 1965, pp. 338-353.
- [103] Zadeh, L. A.: Outline of a new approach to the analysis of complex systems and decision processes. IEEE Trans. on SMC, (3), 1973, pp. 28-44.

Hivatkozások ellenőrzésének utolsó dátuma: 2023.05.10.

7 SAJÁT PUBLIKÁCIÓK

- [S1] Tamás, Tompa, and Kovács Szilveszter. "Expert heuristic tuning design for the FRIQ-learning." *Multidiszciplináris Tudományok* 10.4 (2020): 119-125.
- [S2] Tamás, Tompa, and Kovács Szilveszter. "Heurisztikusan gyorsított megerősítéses tanulási módszerek-áttekintés." *Multidiszciplináris Tudományok* 10.3 (2020): 394-402.
- [S3] Tamás, Tompa, and Kovács Szilveszter. "Szabálytávolság alapú szabálybázis redukció a szakértői tudásbázissal bővített FRIQ-learning környezetben." *Multidiszciplináris Tudományok* 12.1 (2022): 90-102.
- [S4] Tamás, Tompa, and Kovács Szilveszter. "Szakértői heurisztika alkalmazása a FRIQ-learning megerősítéses tanulási módszerben." *Multidiszciplináris Tudományok* 9.4 (2019): 356-368.
- [S5] Tamás, Tompa, and Kovács Szilveszter. "Tudásbázis redukció a szakértői szabályrendszerrel bővített FRIQ-learning módszerben." *Multidiszciplináris Tudományok* 11.4 (2021): 70-80.
- [S6] 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.
- [S7] Tompa, Tamás, and Szilveszter Kovács. "Applying Expert Heuristic as an a Priori Knowledge for FRIQ-Learning." *Acta Polytechnica Hungarica* 17.4 (2020).
- [S8] 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.
- [S9] Tompa, Tamás, and Szilveszter Kovács. "Determining the minimally allowed rule-distance for the incremental rule-base construction phase of the FRIQ-learning." *2018 19th International Carpathian Control Conference (ICCC)*. IEEE, 2018.

- [S10] Tompa, Tamás, and Szilveszter Kovács. "Heuristically accelerated FRIQ-learning." 2022 IEEE 20th Jubilee International Symposium on Intelligent Systems and Informatics (SISY). IEEE, 2022.
- [S11] Tompa, Tamás, and Szilveszter Kovács. "Q-learning vs. FRIQ-learning in the Maze problem." 2015 6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom). IEEE, 2015.
- [S12] Tompa, Tamás, and Szilveszter Kovács. "TUDÁSBÁZIS HANGOLÁSA A FRIQ-LEARNING MEGERŐSÍTÉSES TANULÁSI RENDSZERBEN." Production Systems and Information Engineering 10.4 (2022): 32-41.
- [S13] Tompa, Tamás, and Szilveszter Kovács. "Tudásbázis redukálás a heurisztikusan gyorsított FRIQ-learning rendszerben." Production Systems and Information Engineering 11.2 (2023): 1-12.
- [S14] Tompa, Tamás, Dávid Vincze, and Szilveszter Kovács. "The Pong game implementation with the FRIQ-learning reinforcement learning algorithm." Proceedings of the 2015 16th International Carpathian Control Conference (ICCC). IEEE, 2015.
- [S15] Tompa, Tamás; Kovács, Szilveszter. „Szakértői tudás alapú FRIQ-learning”, Nemzetközi Energetika-Elektrotechnika Konferencia SzámOkt 2018 XXVIII. Nemzetközi Számítástechnika és Oktatás Konferencia Erdélyi Magyar Műszaki Tudományos Társaság (EMT), (2018) pp. 320-325., 4 p.

Publikációs statisztika (2023.05.10.):

Publikációk száma: 15, Hivatkozások száma: 65 (ebből független: 17)