



TUDÁSBÁZIS REDUKÁLÁS A HEURISZTIKUSAN GYORSÍTOTT FRIQ- LEARNING RENDSZERBEN

TOMPA TAMÁS

Miskolci Egyetem

Informatikai Intézet

Általános Informatikai Intézeti Tanszék

tompa@iit.uni-miskolc.hu

KOVÁCS SZILVESZTER

Miskolci Egyetem

Informatikai Intézet

Általános Informatikai Intézeti Tanszék

szkovacs@iit.uni-miskolc.hu

Absztrakt. A megerősítéssel tanuló módszerek tudásbázis leírás formája eltérő. A heurisztikusan gyorsított Fuzzy szabály-interpoláció alapú Q-tanulás (Heuristically Accelerated Fuzzy Rule-Interpolation based Q-learning - HFRIQ-learning) rendszerben a tanulási folyamat során létrejövő tudásbázis egy ritka fuzzy szabálybázis által reprezentált, amely az állapot-akció formátumú szakértői szabályok megadása következtében külső információkat tartalmazhat a probléma megoldására vonatkozóan. A HFRIQ-learning módszer szabálybázis építési és hangolási eljárása következtében előfordulhat olyan eset, hogy több szabálypont egymáshoz közel kerül a fuzzy szabálybázis hangolása (szabálypontok vándorlása) miatt. Az állapot-akció-Q-érték univerzumban egymáshoz közel elhelyezkedő szabályok nagyon hasonló információt írnak le, amely következtében az alapötlet, hogy ezen szabályok összevonásával (egyesítésével) a szabályrendszer mérete csökkenthető. Jelen cikk célja egy olyan javasolt fuzzy szabálybázis csökkentési módszer bemutatása, amely alkalmas lehet a fuzzy szabályok között értelmezhető távolság alapján történő szabálybázis redukálására (szabály összevonásra) a tanulási folyamat során.

Kulcsszavak: megerősítéssel tanulás, heurisztikusan gyorsított megerősítéssel tanulás, Q-learning, Fuzzy Q-learning, szakértői tudásbázis, tudásbázis redukálás

1. Bevezetés

A megerősítéssel tanuló módszerek (Reinforcement Learning - RL) [5] tudásbázis leírás formája eltérő, a klasszikus Q-learning [18] algoritmus Q-táblát (többdimenziós mátrix), a fuzzy szabályalapú megerősítéssel tanuló rendszerek [1][2][3] pedig fuzzy szabálybázist alkalmaznak a rendszer működtető tudásbázisának leírására. A végső tudásbázis mérete, azaz a Q-tábla elemeinek

száma, a fuzzy szabályrendszer szabályainak száma függ az adott probléma méretétől, dimenzióinak számától, így előfordulhatnak olyan esetek mikor ezek mérete igen nagy. A fuzzy szabály interpoláció alapú megerősítéses tanulási rendszerekben a rendszer végleges működtető tudásbázis méretének csökkentésére szabálybázis redukálási (csökkentési) módszerek alkalmazhatók.

A FRIQ-learning (Fuzzy Rule-Interpolation based Q-learning) rendszerben [14][17] a tudásbázist leíró szabálybázis méretének csökkentésére a tanulási fázis után van lehetőség. A tanulási folyamat során létrejött, inkrementálisan felépített szabálybázisban az elhagyható szabályok keresését szabálybázis redukálási módszerek [8][12][16] valósítják meg. A redukációs módszerek közös jellemzője, hogy a szabályok konzekvens értékét, azaz a Q-értéket vizsgálja. Az I.-III. jelölésű redukálási módszerek [16] dekrementálisak, azaz a végső redukált szabálybázis a tanulási fázis végén kapott teljes szabálybázis egyes szabályainak elhagyásával jön létre, fokozatosan csökkentve annak méretét. A IV., hierarchikus klaszterezési eljárás alapuló redukálási módszer [8] 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 szabálybázis közel ugyanazt a Q-függvényt (irányítási felületet) írja le, mint a redukálás előtti esetben, de kevesebb szabállyal (azaz interpolációs tartóponttal).

A heurisztikusan gyorsított FRIQ-learning (HFRIQ-learning) [7][10][11] rendszerben a tudásbázis építési módszer működéséből adódóan a szabály antecedensek hangolása [10] során elfordulhatnak olyan esetek mikor több szabály közel kerül egymáshoz. Ha az egymáshoz közel lévő szabályok közel ugyanazt az információt írják le, akkor célszerű azokat valamilyen stratégia alapján egyetlen szabállyá egyesíteni (összevonni), csökkentve ez által a szabálybázis méretét.

Jelen cikk célja a heurisztikusan gyorsított FRIQ-learning rendszerben egy olyan javasolt tudásbázis redukálási módszer bemutatása, amely már a tanulási fázis közben a közel hasonló információt leíró fuzzy szabályok összevonásával csökkenti a szabálybázis méretét.

2. Heurisztikusan gyorsított FRIQ-learning (HFRIQ-learning)

A HFRIQ-learning (Heuristically Accelerated Fuzzy Rule-Interpolation based Q-learning - HFRIQ-learning) [10] egy olyan fuzzy szabályinterpoláció alapú Q-tanulási módszer, amely a FRIQ-learning módszer [17] kibővítése olyan módon, hogy lehetőséget ad szakértő által definiált előzetes (a priori) tudásbázis rendszerbe történő beillesztésére [7] és optimalizálására [10][11]. A módszer a Q-függvény leírására a „FIVE” FRI [4] fuzzy szabályinterpolációs eljárást alkalmazza, amely következtében a működtető tudásbázist egy ritka fuzzy szabálybázis írja le. Ezen az m méretű R szabálybázisban egy r_i ($i \in [1, m]$) szabály formátuma a következő [15][17]:

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

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

konzekvense (Q-értéke).

Szakértői által meghatározott és ez által az ágens viselkedését leíró információ injektálására szintén fuzzy szabályok formájában van lehetőség, melyek által az adott állapotban preferált akció végrehajtása definiálható a következő formában [7]:

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

ahol \hat{r}_i az i -edik ($i \in [1, \hat{m}]$) szakértői szabályt jelöli az 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 sorszáma az \hat{m} méretű szakértői szabálybázisban. Ezen szabályrendszer szabálynak formátuma hasonló az (1) által meghatározott szabályok formátumához, azzal a különbséggel, hogy ebben az esetben az antecedens az állapot, a konzekvens pedig az ebben az állapotban preferált akció. A FIVE FRI modell alkalmazása (és így a ritka szabálybázis) következtében tetszőleges számú szakértői szabály definiálható a tanulási folyamat előtt.

A tanulási fázis a szakértő által definiált \hat{r}_i [7] illetve az n -dimenziós hiperkocka sarkaiban elhelyezkedő 2^{n+1} (n : állapotdimenziók száma) darabszámú r_i^\square sarokponti szabály [15] összefésülésével kezdődik. A 0 konzekvens értékkel rendelkező sarokponti szabályok formátuma a következő:

$$r_i^\square: \text{If } s_1 \text{ is } S_1^{\square i} \text{ And } s_2 \text{ is } S_2^{\square i} \text{ And ... And } s_n \text{ is } S_n^{\square i} \text{ And } a \text{ is } A^{\square i} \text{ Then } \tilde{Q}(s, a) = 0 \quad (3)$$

A két szabályrendszer a rendszer tanulási fázisába összefésülve kerül injektálásra, olyan módon, hogy az állapot-akció szakértői szabályok (2) formátuma az (1) formátumnak megfelelően módosul, azaz a szakértői szabályok akció konzekvense antecedensre módosul, majd az új konzekvensük egy becsült \tilde{Q}_{init} érték lesz, amelyet egy Q-érték becslési módszer [7] határoz meg.

$$\hat{r}_i: \text{If } s_1 \text{ is } \hat{S}_1^i \text{ And } s_2 \text{ is } \hat{S}_2^i \text{ And ... And } s_n \text{ is } \hat{S}_n^i \text{ And } a = \hat{A}^i \text{ Then } \tilde{Q}(s, a) = \tilde{Q}_{init} \quad (4)$$

Ha a két szabálybázis összefésülése során előfordul olyan szakértői szabály, amely sarokponti szabályra illeszkedik, azaz ennek következtében ellentmondás alakul ki, akkor ennek megszüntetése érdekében a sarokponti szabály 0 értékű konzekvense lecserélésre kerül a szakértői szabály \tilde{Q}_{init} értékű konzekvensére, majd a sarokponti szabály pedig eltávolításra kerül a szabálybázisból.

Az így előállt kezdeti szabálybázis (kezdeti szakértői tudásbázis) fog kiegészülni inkrementálisan a tanulási folyamat során a rendszer által létrehozott új szabályokkal [15]. Új szabály akkor kerül beszúrára az aktuális megfigyelés pontjába, ha a $\Delta\tilde{Q}$ Q-frissítés értéke nagyobb, mint az ε_Q küszöbérték ($\Delta\tilde{Q} > \varepsilon_Q$) és az aktuális megfigyeléshez legközelebb elhelyezkedő szabály is távolinak tekinthető [10]. A szabályok közötti távolság (és így a legközelebbi szabály) meghatározása egy dimenzióként számított távolságon illetve távolságküszöb értékeken alapszik [9][10]. Ha a Q-frissítés értéke is kicsi ($\Delta\tilde{Q} < \varepsilon_Q$), akkor a már létező szabályok konzekvense (Q-értéke) kerül frissítésre az alábbi összefüggés által [15][17]:

$$\tilde{Q}^{k+1}(\mathbf{s}, a) = \tilde{Q}^k(\mathbf{s}, a) + \Delta\tilde{Q}^{k+1}(\mathbf{s}, a) \quad (5)$$

$$\Delta\tilde{Q}^{k+1}(\mathbf{s}, a) = \alpha * \left(g(\mathbf{s}, a, \mathbf{s}') + \gamma * \max_{a' \in U} \tilde{Q}^k(\mathbf{s}', a') - \tilde{Q}^k(\mathbf{s}, a) \right) \quad (6)$$

ahol $\gamma \in [0,1]$ a leszámítolási tényező, $\alpha \in [0,1]$ a tanulási ráta, q_i^{k+1} az i -edik szabály konklúziója a $(k+1)$ -edik iterációban, a pedig az \mathbf{s} -ben végrehajtott akció. Az új megfigyelt állapot \mathbf{s}' , $g(\mathbf{s}, a, \mathbf{s}')$ a megfigyelt jutalom az $\mathbf{s} \rightarrow \mathbf{s}'$ állapot átmenetre, \tilde{Q}^k és \tilde{Q}^{k+1} pedig a k -edik és a $(k+1)$ -edik iteráció FIVE FRI módszer által becsült Q-értéke [13][17]:

$$\tilde{Q}(\mathbf{s}, a) = \begin{cases} \sum_{i=1}^m \left(\left(\frac{q^i}{(\delta_v^i)^\lambda} \right) / \left(\sum_{j=1}^m 1 / (\delta_v^j)^\lambda \right) \right) & \text{ha } (\mathbf{s}, a) = (\mathbf{s}^i, a^i) \\ & \text{valamennyi } i - re, \\ & \text{egyébként} \end{cases} \quad (7)$$

ahol q^i az i -edik ($i \in [1, m]$) szabály konklúziója, (\mathbf{s}, a) a megfigyelés, δ_v^i a skálázott távolság [4] az (\mathbf{s}, a) megfigyelés és az i -edik szabály (\mathbf{s}^i, a^i) antecedense között, λ a Shepard paraméter, m pedig a szabályok száma (további információk az [4], [7] és [9] sorszámú hivatkozásokban).

Ha a Q-frissítés értéke kicsi és a megfigyelés közelében már van létező (szakértői vagy új) szabály akkor a legközelebbi szabály állapot-akció pontja és konzekvens értéke kerül hangolásra a gradiens-módszer alkalmazásával, amely frissítési formulája a láncszabály következtében a következő [10][11]:

$$\mathbf{s}_{k+1} = \mathbf{s}_k - \left(2 * TDerror * \frac{\partial \tilde{Q}(\mathbf{s}, a)}{\partial \mathbf{s}} \right) * \alpha \quad (8)$$

$$a_{k+1} = a_k - \left(2 * TDerror * \frac{\partial \tilde{Q}(\mathbf{s}, a)}{\partial a} \right) * \alpha \quad (9)$$

$$q_{k+1} = q_k - \left(2 * TDerror * \frac{\partial \tilde{Q}(\mathbf{s}, a)}{\partial q} \right) * \alpha \quad (10)$$

ahol a $\mathbf{s}_{k+1}, a_{k+1}, q_{k+1}$ a gradiens-módszer által meghatározott új állapot, akció és Q-értékek, \mathbf{s}_k, a_k, q_k a régi állapot, akció és Q-értékek, α a gradiens-módszer tanulási rátája, $\frac{\partial \tilde{Q}(\mathbf{s}, a)}{\partial \mathbf{s}}, \frac{\partial \tilde{Q}(\mathbf{s}, a)}{\partial a}, \frac{\partial \tilde{Q}(\mathbf{s}, a)}{\partial q}$ a Q-függvény állapot, akció és Q-érték szerinti parciális deriváltjai, a $TDerror$ értéke pedig a következő [10][11]:

$$TDerror = g(\mathbf{s}, a, \mathbf{s}') + \gamma * \max_{a' \in U} \tilde{Q}^k(\mathbf{s}', a') - \tilde{Q}^k(\mathbf{s}, a) \quad (11)$$

A tanulási folyamat akkor ér véget, hogy már nem kerül új szabály beillesztésre a szabálybázisba és a Q-frissítés értéke nem változik jelentősen. Az I-IV. szabálybázis redukálási módszerek [8][12][16][17] a teljes szabálybázisból elhagyható szabályok keresése érdekében alkalmazhatóak a tanulási folyamat végeztével, amelyek következtében a Q-függvényt reprezentáló szabálybázis mérete (szabályainak száma) csökkenthető.

3. Tudásbázis redukálás a HFRIQ-learning rendszerben

Az előzőekben bemutatott HFRIQ-learning rendszerben [7][10] a tanulási folyamat a szakértő által definiált szabályrendszer beépítésével indul. Ezt követően a tanulási fázis során epizódról-epizódra számos új szabály kerül beillesztésre, inkrementálisan bővítve a tudásbázist. A szakértői szabályrendszer bármennyi, a szakértő által meghatározott szabályt tartalmazhat, amelyek tetszőleges (a szakértő által meghatározott) szabálypontokban helyezkedhetnek el. A szabálybázis bővítése során új szabály akkor kerül beszurásra az éppen aktuális megfigyelés pozíciójába, ha a legközelebbi szabálypont is távol van a megfigyeléstől és a Q-frissítési érték nagyobb, mint az előre meghatározott ε_Q küszöbérték ($\Delta Q > \varepsilon_Q$). A szabálybázis hangolása során előfordulhat olyan eset, mikor kettő vagy több szabály közel kerül egymáshoz. Amennyiben ezek a hasonló antecedenssel rendelkező szabályok közel ugyanazt az információt írják le (q konzekvens értékük is hasonló), úgy a tanulási fázis után alkalmazható szabálybázis redukálási módszerek ezen szabályok valamelyikét valószínűleg el fogja távolítani a szabályrendszerből. Azonban ha már a tanulási folyamat közben megállapítható, hogy egyes szabályok az állapot-akció térben (antecedens) közel kerülnek egymáshoz, akkor ezen szabályok egyesítésére, azaz valamilyen módszer alapján történő összevonására, már a tanulás fázisban is sor kerülhet. Azaz egy olyan szabálybázis redukációs (módosított tudásbázis építési) módszer fejleszthető, amely már a tanulási fázis közben ellenőrzi a közel ugyanazt az információt leíró szabályok előfordulását, majd valamilyen stratégia alapján egyiküket elhagyja, vagy összevonja (egyesíti) azokat.

A javasolt szabálybázis redukációs módszer elvárt jellemzői a következők:

- A közel ugyanazt az információt leíró (egymáshoz hasonló) fuzzy szabályok összevonása (redukálása) egyetlen szabályként történjen a szabálybázis építési folyamat, azaz tanulási fázis közben a szabályok közötti távolság alapján.
- A szabályok közötti távolság és távolságküszöbök meghatározása történjen univerzumonként. Akkor tekinthető két szabály egymáshoz közelinek (hasonlónak), ha minden egyes dimenzióban (a konzekvens dimenzióban is) közelinek számít.
- A szabályok összevonása során legyen lehetőség a vizsgált szabályok típusának figyelembevételére, azaz a szakértői szabályok súlyának (fontosságának) beszámítására.
- A javasolt szabálycsökkentési módszer használatával az I.-IV. szabálybázis redukációs módszerek utólagos alkalmazása egyes esetekben el is hagyható.

3.1. A közeli szabályok egyesítése

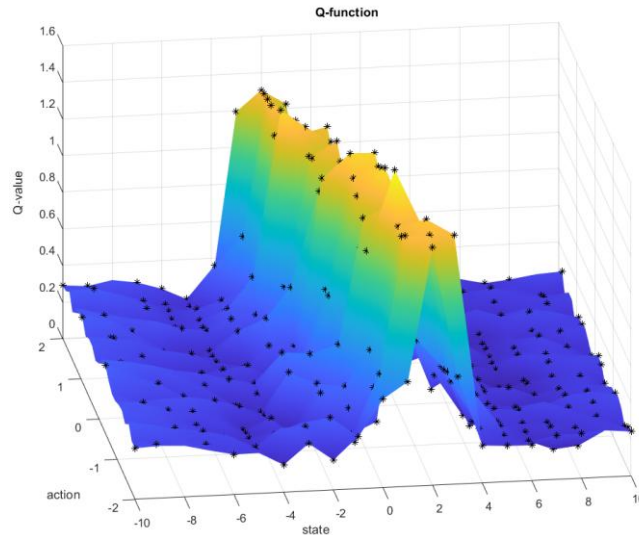
A tanulási fázis során a szabálybázis hangolási eljárása [10][11] következtében a szabályok állapot-akció pontja elmozdulhat (nem csak a konzekvens változhat), amely következtében előfordulhat olyan eset, mikor két vagy esetleg több szabály közel kerül egymáshoz. Ha az egymáshoz nagyon közel kerülő szabályoknak hasonló a konzekvensé, akkor közel ugyanazt az információt írják le. Az ilyen hasonló szabályok egyesítésével (összevonásával) a szabálybázis mérete csökkenthető, redukálható.

Az egymáshoz közel kerülő szabályok összevonásának alapja egy *dtr* távolság küszöbérték [6][10]. Ez az antecedens dimenziókra meghatározott *dtr*

távolságküszöb az alábbi:

$$\exists_{t,p \in [1, m + \hat{m}]} t, p \text{ ahol } \forall_{j \in [1, n+1]} (d_j(t, p) < dtr_j) \quad (12)$$

ahol t, p szabálysorszám, $d_j(t, p)$ távolság a j -edik dimenzióban a t és a p indexű szabályok között, dtr_j pedig a j -edik dimenzió távolságküszöb értéke. Azonban ez nem elegendő hasonlóságának definiálásához, mert előfordulhat olyan eset mikor két szabály ezen távolságküszöb alapján (azaz az antecedens dimenzióban) egymáshoz közelinek számít a (12) összefüggés alapján, de a konzekvensükben (Q-értékükben) nagy az eltérés, ezért nem hasonló szabályok. Ebben az esetben nem célszerű a két (forrás) szabályt összevonni és egyetlen szabállyá redukálni mert lehet, hogy a közeli antecedensek és különböző konzekvenssek az általuk leírt Q-függvényben egy meredek lejtőt vagy emelkedőt jelentenek. Egy ilyen meredek „hegyoldallal” rendelkező Q-függvényt szemléltet az 1. ábra, ahol a „*” karakter a szabálypontokat jelöli.



1. ábra Meredek töréspontot tartalmazó Q-függvény

Ennek következtében a konzekvens (Q-érték) dimenzióra is szükséges közelségmérték definiálása, amely által csak akkor tekinthető két szabály egymáshoz hasonlóknak, ha azok a távolságértékek alapján az antecedens és a konzekvens univerzumokban is közelinek számítanak. A konzekvens dimenzióbeli távolságot d_Q jelöli, amely a két szabály Q-érték különbségének az abszolútértéke. A $d_Q(t, p)$ a t és p indexű szabályok ($t, p \in [1, m + \hat{m}]$) közötti távolság a konzekvens dimenzióban a következőképpen írható fel:

$$d_Q(t, p) = |Q^t - Q^p| \quad (13)$$

A konzekvens univerzumbeli közelség (és az értékének) meghatározása szintén egy távolságküszöb alapján történik, emiatt a dtr vektor kiegészül a konzekvens dimenzióra is vonatkozó dtr_Q távolságküszöbvel, így $dtr = [dtr_1, dtr_2, \dots, dtr_n, dtr_U, dtr_Q]$. Ennek értéke a teljes (éppen aktuális) Q-érték tartomány valamekkora része, azaz a legnagyobb és a legkisebb Q-érték különbségének (a teljes tartomány hosszának) a szakértő által definiált dR_q része:

$$length(Q) = |\max(Q) - \min(Q)| \quad (14)$$

$$dtrQ = \frac{length(Q)}{dR_q} \quad (15)$$

Mivel a Q -értékek a tanulási fázis során iterációnként változnak (a módszer által hangolásra kerülnek), így a $length(Q)$ értéke minden egyes olyan iterációban újraszámításra kerül ahol szabálybázis redukálás történik.

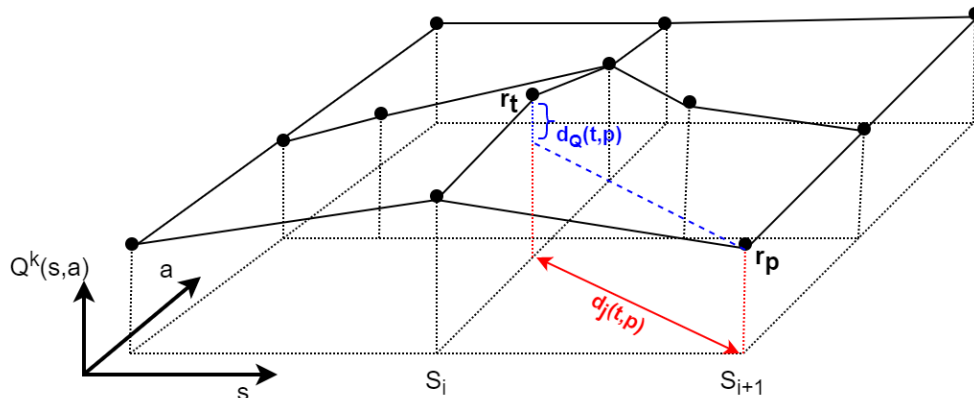
Összegezve, a szabálybázis redukálás során akkor tekintjük a szabálybázis két t és p indexű r_t és r_p szabályát egymáshoz hasonlóknak és ennek következtében akkor kerülnek összevonásra (redukálásra) egyetlen r_{red} szabályként, ha a (12) összefüggés teljesül rájuk és még az r_t és r_p szabályok konzekvensében (Q -értékében) nincs nagy eltérés. Tehát minden egyes antecedens dimenzióban a távolságuk egymáshoz képest közelinek számít ($d_j(t,p) < dtr_j$) és a két szabály Q -érték különbségének abszolútértéke kisebb, mint a dtr_Q küszöbérték:

$$\exists_{t,p \in [1,m+\hat{m}]} t,p \text{ hogy } d_Q(t,p) < dtrQ \quad (16)$$

a t indexű r_t és p indexű r_p szabály összevonásra kerül egyetlen szabállyá.

Akkor azonban, ha az antecedens dimenzióban közelinek tekinthetők, de a konzekvensük a dtr_Q küszöbértéknél nagyobb, akkor nem kerülnek összevonásra.

A 2. ábra a t és p indexű szabályok közötti konzekvens dimenzióbeli $d_Q(t,p)$ és az antecedens dimenzióbeli $d_j(t,p)$ távolságokat szemlélteti:



2. ábra A p és t indexű szabályok közötti $d_j(t,p)$ antecedens dimenzióbeli és $d_Q(t,p)$ konzekvens univerzumbeli távolság

Az új szabálypont, azaz az egyesítés után létrejövő új r_{red} szabály antecedensének és konzekvensének meghatározása a két forrásszabály antecedens és konzekvens értékeinek az átlagolásával történik. Mivel a szakértő által definiált produkciós szabályrendszer formátuma (2) a FRIQ-learning rendszerbe történő beillesztésük után a fuzzy Q szabály formára (4) változik, így az új szabály antecedensének és konzekvensének számítása az eredetileg szakértő által definiált szabályok esetén is a rendszer szabályainak egyesítésével azonos módon történhet. A javasolt megoldás az, hogy az új szabály állapot-akció értéke (antecedense) és Q -értéke (következménye) legyen a forrás szabályok antecedens és konzekvens értékeinek az átlaga.

A javasolt tanulási fázis közben alkalmazható szabályösszevonás (szabálybázis redukálás) algoritmusának pszeudokódja az alábbi [6][10]:

Algoritmus: distBasedRBreduce(R, dR)

Input: the rule-base, dR parameters

Output: the reduced rule-base

initialize the reduced rule-base

initialize the r_{red} reduced rule

initialize the distance matrix (D)

compute the distance thresholds (dtr) for each antecedent and consequent universe according to the dR

Loop (for each rule in the rule-base)

 compute the rule distances (d) between each rule ($d \in D$)

 check the distances (d) of each rule pair in the rule-base

 If $\exists_{t,p \in [1,m+\hat{m}]} t,p$ where $\forall_{j \in [1,n+1]} (d_j(t,p) < dtr_j)$ and $(d_Q(t,p) < dtr_Q)$, $j \in [1, n + 1]$

 the r_t and the r_p rule pair can be merged into a r_{red} rule

 compute the antecedent and consequent of r_{red} using the average of r_t and r_p source rules

 add the r_{red} merged rule to the reduced rule-base

 end

return the reduced rule-base

Továbbfejlesztésként célszerű lehet majd egy olyan módszer kidolgozása, amely alkalmas lehet a szakértő által meghatározott konstans dR paraméterek (és így a távolság alapú küszöbértékek) hangolására, optimalizálására.

3.2. Az összevont szabály típusának meghatározása

A HFRIQ-learning rendszerben [10] a fuzzy szabályoknak három típusa különböztethető meg: a szakértő által definiált szabályrendszer (R_{expert}), a rendszer által létrehozott szabályrendszer (R) és a rendszer által létrehozott sarokponti szabályok ($r^{\square} \in R$). A sarokponti szabályrendszer az interpolációs eljárás miatt szükséges 2^{n+1} darabszámú sarokponti szabályt, a rendszer által létrehozott szabályrendszer pedig az újonnan létrehozott szabályokat tartalmazza. A szabályok különböző típusa miatt az egyesítés (redukálás) során előálló új szabály típusát célszerű a forrásszabályok típusa alapján meghatározni. Mivel a szakértő által definiált szabályok feltételezhetően helyes tudást írnak le a rendszer működésére vonatkozóan, így azok nagyobb fontossággal (súllyal) kerülnek figyelembe vételre. Ezen fontosság (vagy súly) határozza meg az új szabály típusát illetve azt, hogy a forrás szabályok milyen módon kerülnek, vagy éppen nem kerülnek összeolvasztásra egyetlen szabállyá.

Két egymáshoz közel kerülő szabály esetében az új, egyesített szabálytípus meghatározásának módja a következő: ha a két szabály közül az egyik szakértői a másik pedig a rendszer által felvett új szabály, akkor az egyesített szabály szakértői szabály lesz. Ha a két szabály közül mindkét szakértői szabály volt, akkor az új összevont szabály is szakértői szabály lesz. Ha a két szabály közül mindkét szabály a rendszer által újonnan felvett szabály, akkor az új egyesített szabály is újonnan beszűrt szabályként lesz jelölve.

Speciális eset mikor az egymáshoz közeli szabályok közül az egyik sarokponti típusú a másik pedig attól különböző, azaz szakértői vagy újonnan létrehozott szabály, tehát a sarokpontot leíró szabály közelébe kerül egy attól különböző típusú szabály. Annak következtében, hogy a sarokponti szabályok a FIVE interpolációs

módszer alkalmazása miatt fontos szerepet töltenek be a szabálybázis építés folyamatában, így ezek eltérő fontossági súllyal kerülnek figyelembevételre a szabályegyesítés során. Ha a közeli szabály a rendszer által létrehozott szabály, akkor az törlésre kerül a sarokponti szabály közeléből. Ha a közeli szabály szakértői szabály volt, akkor az változatlanul megmarad. Ezekben az esetekben így nem történik szabályösszevonás (a sarokponti szabály antecedense nem változhat), hanem a rendszer által létrehozott közeli szabály esetében az törlésre kerül, szakértői szabály esetében pedig változatlanul megmarad.

A szabályok típusának meghatározásához illetve az egyes szabályok változásának (hangolásának) nyomonkövetése céljából a szabályrendszer minden egyes szabálya egy egyedi azonosítóval (ID) rendelkezik. Ezen egyedi azonosítóval a szabályok hangolása és egyesítése nyomonkövethető, a hangolási folyamat végén kinyerhető. Utólagosan meghatározható, hogy a hangolási, optimalizálási folyamat során a szakértő által megadott szabályok milyen mértékben változtak.

Egy, a szakértői által definiált szabályt jelöljük \hat{r} -el ($\hat{r} \in R_{expert}$), egy a rendszer által felvett szabályt r -el ($r \in R$), egy sarokponti szabályt pedig r^\square -el ($r^\square \in R$). A (17) összefüggés az adott típusú, r_t és r_p indexű szabályok egyesítését követően létrejött új r_{red} szabály típusát szemlélteti. A „ \sqcup ” operátor a két szabály egyesítését, a „ \rightarrow ” operátor pedig a szabályegyesítés eredményét, azaz a szabályösszevonás során létrejött új r_{red} szabály típusát jelöli:

$$\begin{array}{ccccc}
 r_t & & r_p & & r_{red} \\
 \hline
 r & \sqcup & \hat{r} & \rightarrow & \hat{r} \\
 \hat{r} & \sqcup & \hat{r} & \rightarrow & \hat{r} \\
 r & \sqcup & r & \rightarrow & r \\
 r^\square & \sqcup & r & \rightarrow & r^\square \\
 r^\square & \sqcup & \hat{r} & \rightarrow & r^\square, \hat{r}
 \end{array} \tag{17}$$

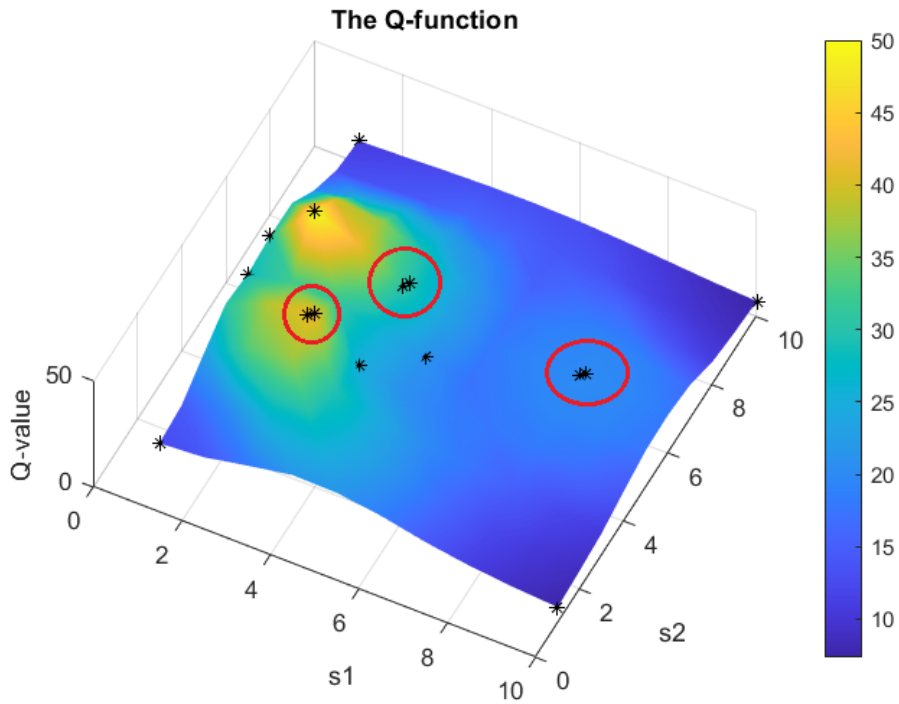
A 1. táblázat egy lehetséges szabályegyesítési példát szemléltet, ahol \hat{r} egy szakértői szabály, r pedig egy rendszer által beszúrt (új) szabály. Feltételezzük, hogy r és \hat{r} távolsága közelinek tekinthető, továbbá s_1, s_2 az állapot dimenziók, a az akcióérték, q a Q-érték, $\hat{r} \sqcup r \rightarrow \hat{r}$ pedig az egyesített új szabály, amely szakértői szabályként kerül megjelölésre:

1. táblázat Szabályegyesítés egy szakértői és egy rendszer által felvett szabály esetében

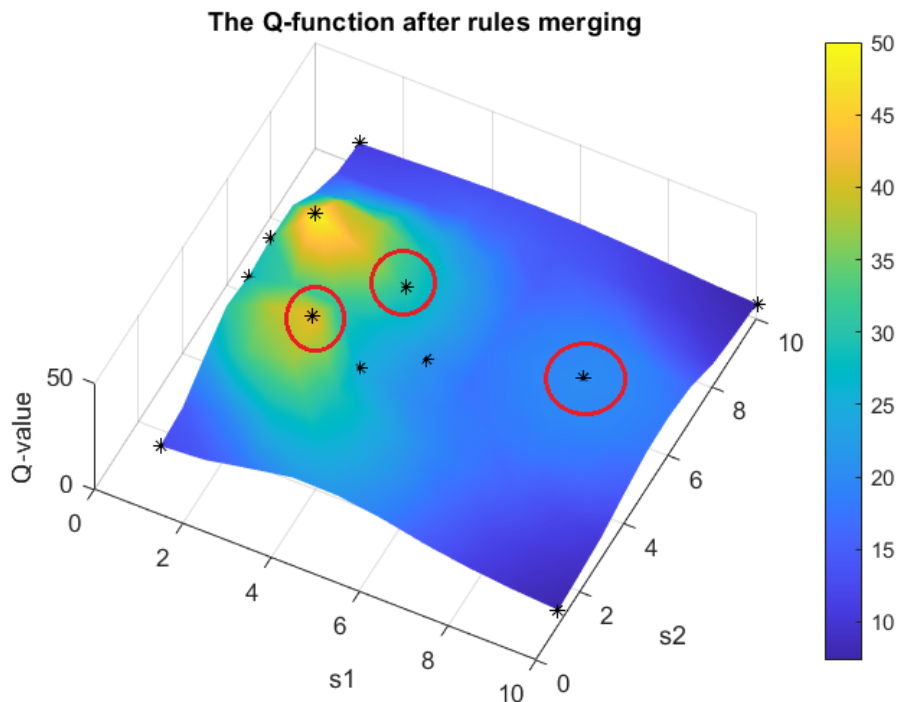
Szabály	s_1 állapot	s_2 állapot	a akció	q -érték
\hat{r}	1	2	4	123.45
r	2	2	5	145.23
$\hat{r} \sqcup r \rightarrow \hat{r}$	1.5	2	4.5	134.34

A 3. és 4. ábrák egy mintapéldán keresztül vizuálisan is szemléltetik a szabályegyesítés folyamatát. Az ábrákon egy Q-függvény felülete látható, melyet a „*” által jelölt fuzzy szabályok (tartópontok) írják le. A függvény két állapotdimenzióval (s_1 és s_2) és egy akciódimenzióval rendelkezik. A Q-függvény megjelenítésénél a legjobb akció melletti Q-értékek kerültek kirajzolásra az s_1 és s_2 univerzumok mentén, az akciódimenziót a megjelenítésből. A 3. ábra a szabályegyesítés (redukálás) előtti állapot, mikor 2-2-2 darab (3 pár) piros

karikázással jelölt egymáshoz közeli szabály található, a 4. ábra pedig a szabályegyesítés utáni állapotot szemlélteti, miután ezen közeli szabályok összevonásra kerültek. Így a 15 darab szabály helyett már csak 12 darab „*” által jelölt tartópont (fuzzy szabály) írja le a függvény felületét:



3. ábra Q -függvény felülete szabályegyesítés előtt (15 darab „*”-al jelölt fuzzy szabállyal)



4. ábra Q -függvény felülete szabályegyesítés után (12 darab „*”-al jelölt fuzzy szabállyal)

A szabályösszevonások eredménye táblázatban szemléltetve:

2. táblázat A 3. és 4. ábrákon látható szabályegyesítések eredménye

Szabály	s_1	s_2	a	q
r_1	4	6	1	10
r_2	4.1	6.1	1	10.2
$r_1 \sqcup r_2 \rightarrow r$	4.05	6.05	1	10.1
r_3	3	4	1	42
r_4	2.9	3.9	1	41.8
$r_3 \sqcup r_4 \rightarrow r$	2.95	3.95	1	41.9
r_5	8	6	1	20
r_6	8.1	6.1	1	20.3
$r_5 \sqcup r_6 \rightarrow r$	8.05	6.05	1	20.15

4. Összefoglalás

Bemutatásra került egy olyan javasolt módszer, amely alkalmas tanulási fázis során, a hangolási folyamat következtében egymáshoz közel kerülő fuzzy szabályok egyesítésére (összevonására) a tudásbázist leíró fuzzy szabályrendszer méretének csökkentése céljából. Az egymáshoz közel elhelyezkedő fuzzy szabályok hasonló információt írnak le, így ezek összevonásával a szabálybázis mérete a bemutatott módszer alkalmazásával csökkenthető. A hasonlóság meghatározásának alapja az univerzumként meghatározott szabálytávolságok, amelyek alapján két szabály akkor tekinthető egymáshoz közelinek, ha azok minden egyes univerzumban (állapot, akció, Q-érték) közelinek mondhatók, azaz a távolságuk minden dimenzióban kisebbek, mint az adott univerzum távolságküszöb értéke. Bemutatásra került továbbá egy olyan módszer, amely által az összevont (új) szabály típusa meghatározható illetve antecedens és konzekvens értéke számítható.

Irodalomjegyzék

- [1] Berenji, Hamid R. "Fuzzy Q-learning for generalization of reinforcement learning." Proceedings of IEEE 5th International Fuzzy Systems. Vol. 3. IEEE, 1996. <https://doi.org/10.1109/FUZZY.1996.553542>
- [2] 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.
- [3] Glorennec, P. Y., & Jouffe, L. (1997, July). Fuzzy Q-learning. In Proceedings of 6th international fuzzy systems conference (Vol. 2, pp. 659-662). IEEE.
- [4] Kovács, Szilveszter. "Extending the fuzzy rule interpolation" FIVE" by fuzzy observation." Computational Intelligence, Theory and Applications. Springer, Berlin, Heidelberg, 2006. 485-497. https://doi.org/10.1007/3-540-34783-6_48
- [5] Sutton, Richard S., and Andrew G. Barto. Reinforcement learning: An introduction. MIT press, 2018.
- [6] Tamás, Tompa, and Kovács Szilveszter. "Szabálytávolság alapú szabálybázis redukció

- a szakértői tudásbázissal bővített FRIQ-learning környezetben." *Multidiszciplináris Tudományok* 12.1 (2022): 90-102. <https://doi.org/10.35925/j.multi.2022.1.8>
- [7] Tompa, Tamás, and Szilveszter Kovács. "Applying Expert Heuristic as an a Priori Knowledge for FRIQ-Learning." *Acta Polytechnica Hungarica* 17.4 (2020).
- [8] 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. <https://doi.org/10.1109/SAMI.2017.7880302>
- [9] 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. <https://doi.org/10.1109/CarpathianCC.2018.8399677>
- [10] Tompa, Tamás, and Szilveszter Kovács. "Heuristically accelerated FRIQ-learning." 20th Jubilee International Symposium on Intelligent Systems and Informatics (SISY 2022). IEEE, 2022.
- [11] Tompa, Tamás, and Szilveszter Kovács. "TUDÁSBÁZIS HANGOLÁSA A FRIQ-LEARNING MEGERŐSÍTÉSES TANULÁSI RENDSZERBEN." *Production Systems and Information Engineering* 10.4 (2022): 32-41. <https://doi.org/10.32968/psaie.2022.4.4>
- [12] 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). IEEE, 2020. <https://doi.org/10.1109/AIM43001.2020.9158875>
- [13] Vincze, Dávid, and Szilveszter Kovács. "Fuzzy rule interpolation-based Q-learning." 2009 5th International Symposium on Applied Computational Intelligence and Informatics. IEEE, 2009. <https://doi.org/10.1109/SACI.2009.5136311>
- [14] Vincze, Dávid, and Szilveszter Kovács. "Fuzzy rule interpolation-based Q-learning." 2009 5th International Symposium on Applied Computational Intelligence and Informatics. IEEE, 2009. <https://doi.org/10.1109/SACI.2009.5136311>
- [15] Vincze, Dávid, and Szilveszter Kovács. "Incremental rule base creation with fuzzy rule interpolation-based Q-learning." *Computational Intelligence in Engineering*. Springer, Berlin, Heidelberg, 2010. 191-203. https://doi.org/10.1007/978-3-642-15220-7_16
- [16] 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. <https://doi.org/10.17667/riim.2015.1-2/10>.
- [17] Vincze, Dávid.: "Fuzzy Rule Interpolation-based Q-learning." PhD dissertation, 2013. <https://doi.org/10.14750/ME.2014.003>
- [18] Watkins, Christopher JCH, and Peter Dayan. "Q-learning." *Machine learning* 8.3 (1992): 279-292.