Multidiszciplináris tudományok, 11. kötet. (2021) 4 sz. pp. 70-80 **https://doi.org/10.35925/j.multi.2021.4.8**

TUDÁSBÁZIS REDUKCIÓ A SZAKÉRTŐI SZABÁLYRENDSZERREL BŐVÍTETT FRIQ-LEARNING MÓDSZERBEN

Tompa Tamás
tanársegéd, Miskolci Egyetem, Informatikai Intézet, Általános Informatikai Intézeti Tanszék
3515 Miskolc, Miskolc-Egyetemváros-s, e-mail: tompa@iit.uni-miskolc.hu

Kovács Szilveszter
egyetemi tanár, Miskolci Egyetem, Informatikai Intézet, Általános Informatikai Intézeti Tanszék
3515 Miskolc, Miskolc-Egyetemváros, e-mail: szkovacs@iit.uni-miskolc.hu

Absztrakt

A megerősítéses tanulási módszerek tudásábrázolási formája eltérő, a klasszikus Q-learning algoritmus Q-táblát, a fuzzy szabályalapú megerősítéses tanuló rendszerek 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 előfordulhat olyan eset mikor ezek mérete igen nagy is lehet. A fuzzy szabály interpoláció alapú megerősítéses tanulási rendszerekben a rendszer végleges működtető tudásbázis méretének csökkentésére szabálybázis redukálási (csökkentési) módszerek alkalmazhatók. A FRIQ-learning rendszerben a tudásbázist leíró szabálybázis méretének csökkentésére, azaz az elhagyható szabályok keresésére a tanulási fázis után van lehetőség opcionálisan. A szakértői tudásbázissal bővített FRIQ-learning rendszerben a tudásbázis építési módszer működéséből adódóan elfordulhatnak olyan esetek mikor több szabály közel kerülhet egymáshoz. Célszerű lehet ezen szabályokat valamilyen stratégia alapján összevonni, csökkentve ezáltal a szabálybázis méretét. Jelen cikk célja a szakértői tudásbázissal bővített FRIQ-learning rendszerben egy olyan tudásbázis redukálási módszer bemutatása, amely már a tanulási fázis közben megvalósítja a rendszer tudásbázisának redukálását, olyan módon, hogy egyesíti a hasonló tudást leíró fuzzy szabályokat.

Kulcsszavak: *megerősítéses tanulás, heurisztikusan gyorsított megerősítéses tanulás, szakértői tudásbázis, tudásbázis csökkentés, Q-learning, fuzzy Q-learning*

Abstract

The knowledge representation of reinforcement learning (RL) methods can be different, in case of the conventional Q-learning method it is a Q-table and in case of fuzzy-based RL systems it is a fuzzy rule-base. The size of the final knowledgebase (number of elements in the Q-table, number of rules in the fuzzy rule-base) to be depend on the complexity of the problem (and the dimension size), thus there may be cases when the number of elements in the final knowledge can be considered high. In the fuzzy rule-based RL systems the rule-base reduction methods can be applied to reduce the size of the complete rule-base. In the Fuzzy Rule Interpolation based Q-learning (FRIQ-learning) the rule-base reduction can be performed optionally after the learning phase. In the expert knowledge-included FRIQ-learning, due to the knowledge building method, there can be cases when rules can get close to each other. Merging those rules, which are close to each other, could significantly reduce the size of the final rule-base. The main goal of this paper is to introduce a rule-base reduction strategy of the expert knowledge-included FRIQ-learning, which is able to reduce the rule-base size during the construction (learning) phase.

Keywords: *reinforcement learning, heuristically accelerated reinforcement learning, expert knowledgebase, knowledgebase reduction, Q-learning, fuzzy Q-learning*

1. Bevezetés

A megerősítéses tanuló (Reinforcement Learning – RL) [15] rendszerek működésének alapja az ágens-környezet modell. A tanuló entitás (ágens) kapcsolatba kerül környezetével, amelyben a végrehajtott cselekvései (akciói) hatására új állapotokba kerül, majd ezen akciók sikerességétől függően a környezettől megerősítéseket kap. A megerősítések numerikus értékek, amelyek lehetnek jutalmak (pozitív megerősítés) és büntetések (negatív megerősítés) egyaránt. Az ágens célja, hogy hosszútávon maximalizálja a gyűjtött jutalmakat, azaz olyan cselekvések végrehajtására törekedjen, amelyek minél nagyobb jutalomértékkel kecsegtetnek. Az elérendő cél magában a jutalomfüggvényben van definiálva, amely alapján adott értékű jutalmat vagy büntetést kap a végrehajtott cselekvésre. Az egyes állapotokban végrehajtott akciók minőségét az állapot-akció-érték függvény (Q-függvény) írja le, amely az ágens állapot átmenetei szempontjából irányítási felületnek is tekinthető.

Elterjedtebb megerősítéses tanuló algoritmusok a diszkrét felbontással rendelkező Q-learning [35] és SARSA [14], illetve ezek különböző, akár folytonos térre kiterjesztett változatai. Ezen módszerek közös jellemzője, hogy a tanulási fázis üres tudásbázissal indul (hiszen az a céljuk, hogy ezt a tudásbázist automatikusan feltérképezzék) majd iterációról-iterációra, számos epizódon keresztül folyamatosan bővítik azt a környezettől érkező megerősítési információk alapján. A tanulási folyamat végeztével áll elő az adott problémát megoldó, a rendszert működtető végleges tudásbázis.

Ezen módszerek tudásleírási formája az alkalmazott modelltől függően eltérő lehet. A klasszikus Q-learning (és SARSA) módszerek esetében a rendszer tudásbázisa egy Q-tábla (look-up tábla). Ezen Q-tábla tartalmazza a lehetséges állapot-akció kombinációkhoz tartozó jóság értékeket (Q-értékeket). Az adott probléma méretétől (dimenziószámától) függően a Q-tábla mérete exponenciálisan növekszik, amely következtében ezen módszerek csak bizonyos problémaméret (nagyjából 10000 állapot) esetében tudnak jó eredményt adni. A fuzzy logika használatával folytonos állapot-akció térre kiterjesztett Q-learning variánsok [1][5][7][29] esetében a rendszer tudásbázisát egy fuzzy szabályrendszer reprezentálja. Ebben az esetben a tudásbázis méretét a probléma megoldását leíró fuzzy szabályrendszer szabályainak száma határozza meg. A klasszikus fuzzy logikát alkalmazó rendszerekben a szabálybázisnak teljesnek (fedő jellegű) kell lennie, azaz minden egyes lehetséges állapot-akció kombinációra kell, hogy létezzen valamilyen mértékben illeszkedő szabály, ellenkező esetben előfordulhat, hogy a rendszer valamely állapot-akció esetben nem szolgáltat Q-értéket.

További negatívum a problémaméret (dimenziószám) növekedésével a szabályszám exponenciális növekedése [8]. Ezen problémák kiküszöbölhetők fuzzy szabályinterpolációs (Fuzzy Rule Interpolation, FRI) modell alkalmazásával, mely ritka (nem teljes) szabálybázis esetén is alkalmas következtetésre. Az FRI modell alkalmazásának további előnye, hogy egyes esetekben a szabálybázis mérete tovább csökkenthető. Az alkalmazott tudásbázis építési módszertől függően előforduló redundáns, más szabályokból kiadó szabályok elhagyhatók. Szabálybázis redukciós módszerek alkalmazásával, a FRIQ-learning [29][34] tanulási fázis után a Q-függvényt leíró szabálybázis mérete csökkenthető.

Jelen cikk célja egy olyan szabálybázis redukciós módszer bemutatása, amely a szakértői tudásbázissal bővített FRIQ-learning rendszerben a tanulási folyamat közben vizsgálja a tudásbázis kiadódó, elhagyható szabályait, közvetlenül törekedve egy minimális szabályszámú tudásbázis létrehozására.

2. Szakértői tudásbázissal bővített FRIQ-learning

A Fuzzy szabály-interpoláció alapú Q-tanulás (Fuzzy Rule Interpolation based Q-learning, FRIQ-learning) [29][34] egy fuzzy szabály-interpolációs eljárást (Fuzzy Rule Interpolation, FRI) alkalmazó Q-learning módszer, amely a FRI modellnek köszönhetően folytonos állapot-akció térrel rendelkezik. Az alkalmazott interpolációs módszer, a FIVE (Fuzzy Rule Interpolation based on Vague Environment) [9][10][13], egy többdimenziós FRI módszer, amely viszonylag kis számításigényének [23][26][28] következtében jól használható valós idejű beágyazott rendszerekben [2][4], illetve robotikához kapcsolódó gyakorlati alkalmazásokban (pl. Robot navigáció [3][24][27], viselkedés modellek leírása [11][12][22][25]).
A FRIQ-learning rendszer tudásbázisát fuzzy szabályrendszer (R) írja le, amelyben egy szabály ($r_i, r_i \in R$) formája a következő [29]:

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

ahol r_i ($i \in [1, m]$) az i-edik szabály az m méretű R szabálybázisban, $\tilde{Q}(\boldsymbol{s}, a)$ a FIVE FRI által közelített Q-függvény, q^i az i-edik szabály konzekvense, S_j^i ($j \in [1, n]$) az i-edik szabály fuzzy halmaza a j-edik antecedens dimenzióban, $\boldsymbol{S}$ az n-dimenziós megfigyelés ($s_1, s_2 \dots s_n \in \boldsymbol{S}$), s_j a j-edik dimenziója az n-dimenziós $\boldsymbol{S}$ állapot megfigyelésnek, A^i i-edik szabály fuzzy halmaza az egydimenziós U akciótérben, a ($a \in U$) pedig a végrehajtott akció.

Külső tudásbázis beágyazására a szakértői tudásbázissal bővített FRIQ-learning rendszer [17][18] ad lehetőséget, emberi szakértő által definiált szakértői szabályrendszerrel (R_{expert}). Ezesetben ha rendelkezésre áll a megoldást leírásának egy része, akkor az a tanulási fázis kezdete előtt szakértői szabályrendszer formájában a rendszerbe ágyazható. A szakértő a szabályrendszer szabályai határozzák meg a rendszer adott állapotaiban előnyben részesített akciókat (mint heurisztikus politikamódosító [6]). Egy $\hat{r}_i$ ($\hat{r}_i \in R_{expert}$) szakértő által megadott szabály formája a következő [18]:

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

ahol $\hat{A}^i$ az i^{th} ($i \in [1, \hat{m}]$) szakértői szabály akciója (azaz konzekvense), $\hat{S}_n^i = [\hat{S}_1^i, \hat{S}_2^i, \dots \hat{S}_n^i]$ az n-dimenziós állapot megfigyelés, $\hat{m}$ a szakértői szabályok száma R_{expert} szakértői szabályrendszerben, $\hat{r}_i$ pedig az i-edik szakértői szabály. Ezen szabályrendszer formája hasonló az (1) formátumú szabályokhoz, azzal a különbséggel, hogy itt az antecedens az állapot, a konzekvens pedig az ebben az állapotban preferált akció. Ezen szabályok FRIQ-learning rendszerbe történő injektáláshoz szükség van azok kezdeti Q-értékenek meghatározására is, hogy a szabályok (1) formátumra alakíthatók legyenek. A szakértői szabályrendszer szabályaira az előzetes Q-érték (a szabály minősége) az alábbi összefüggés alapján határozható meg [18]:

$$\tilde{Q}_{init} = \eta * \frac{g_{max}}{1-\gamma} \text{ if } \gamma < 1 \tag{3}$$

ahol $\tilde{Q}_{init}$ a számított kezdeti Q-érték, g_{max} a környezet által adható lehetséges maximális megerősítés, γ a leszámítolási tényező, η pedig a $\tilde{Q}_{init}$ értékre vonatkozó skála tényező (további részletekért lésd [18]). Az előzetes Q-érték számítási módszer alkalmazását követően a szakértői szabályok formája a következő lesz:

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

A szakértői szabályrendszer beillesztése során a szakértő által definiált szabályrendszer állapot-akció formájú szabályainak akció konzekvense antecedensre módosul, majd az új konzekvens pedig a korábban meghatározott $\tilde{Q}_{init}$-érték lesz. Az így kapott (4) formátumú szabályok pedig már beágyazhatók a szabályrendszerbe. Ezt követően a tanulási fázis az előzőleg definiált szakértői szabályrendszert 2^{n+1} (ahol n az állapotdimenziók száma) darab $q^i = 0$ konzekvensű sarokponti szabállyal ($r_i^{\square}$) egészítve indul. Abban az esetben, ha valamely szakértői szabály sarokponti szabályra illeszkedik (állapot-akció egyezés) akkor a módszer a sarokponti szabály $q^i = 0$ konzekvensét lecseréli a szakértői szabály konzekvensére, azaz $q^i = \tilde{Q}_{init}$ lesz.

A tanulási fázisban a szabályrendszer iterációról-iterációra a következő frissítési összefüggés szerint változik:

$$q_i^{k+1} = \begin{cases} q_i^k + \Delta\tilde{Q}^{k+1}(\boldsymbol{s}, a) & \text{ha } (\boldsymbol{s}, a) = (\boldsymbol{s}^i, a^i) \text{ minden } i\text{-re,} \\ q_i^k + \Delta\tilde{Q}^{k+1}(\boldsymbol{s}, a) * \left(1/\delta_{v,i}^{\lambda}\right) \Big/ \left(\sum_{i=1}^{m} 1/\delta_{v,i}^{\lambda}\right) & \text{egyébként} \end{cases} \tag{5}$$

ahol q_i^k az i-edik szabály konzekvense a k-adik iterációban, $(\boldsymbol{s}, a)$ az adott állapot-akció pont, $\delta_{v,i}^{\lambda}$ a skálázott távolság az aktuális megfigyelés és az i-edik szabály között, $\Delta\tilde{Q}^{k+1}(\boldsymbol{s}, a)$ pedig a következő:

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

ahol α a tanulási ráta, γ a leértékelési tényező, $g(\boldsymbol{s}, a, \boldsymbol{s}')$ a megerősítés értéke az $\boldsymbol{s} \rightarrow \boldsymbol{s}'$ állapot átmenetre, $\tilde{Q}^k$ a k-adik, $\tilde{Q}^{k+1}$ pedig a $(k+1)$-edik iterációban a FIVE FRI alapján számított konzekvens érték.

A tanulási fázisban a szabályrendszer a hangolás mellett új szabályok beszúrásával inkrementálisan is változik. Új fuzzy szabály akkor kerül beszúrásra a rendszer tudásbázisát leíró szabályrendszerbe, ha az állapot-akció térben már létező legközelebbi szabály is távol van (az aktuális megfigyeléshez képest) [20] és a $\Delta\tilde{Q}$ Q-frissítési érték nagyobb, mint az előre meghatározott ε_Q küszöbérték, azaz $\Delta\tilde{Q} > \varepsilon_Q$. Ellenkező esetben, ha az aktuális megfigyelés közelében található már szabálypont és teljesül, hogy $\Delta\tilde{Q} < \varepsilon_Q$ akkor a teljes szabályrendszer szabály konzekvensei kerülnek frissítésre [30]. Az inkrementális szabálybázis építési fázis (azaz a tanulási fázis) véget ér, ha a $\Delta\tilde{Q}$ értékeke számos epizódon keresztül viszonylag kicsi és már nem kerül beszúrásra új szabály. Ez az inkrementálisan felépített szabálybázis tartalmazhat olyan szabályokat, amelyek kiadódnak más szabályokból. A redundáns szabályok keresésére a tanulási fázis után alkalmazható szabálybázis redukálási módszerek [19][21][31][32][33] adnak lehetőséget, amelyek használatával a szabálybázis mérete csökkenthető.

3. Szabályredukálási módszerek a FRIQ-learning rendszerben

A szakértői heurisztikával bővített FRIQ-learning rendszerben a tanulási fázis végeztével előállt tudásbázis méretének csökkentésére számos szabálybázis redukálási módszer alkalmazható. 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 redukciós 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).

Az I. redukálási stratégia [31][32] azon szabályokat törli a teljes szabálybázisból, amelyeknek abszolútértékben alacsony a Q-értékük (konzekvens értékük). Minden egyes szabály törlése után megvizsgálja, hogy az adott szabályt elhagyva a probléma még megoldható-e és ha igen, akkor folytatja a folyamatot. Ezt addig teszi, amig az adott szabály törlése után kapott eredmény nem tér el lényegesen az azt megelőzőtől. Ha lényegesen eltér, azaz a feladat már nem oldható meg, akkor a törölt szabályt visszahelyezi a szabálybázisba és fontos szabályként jelöli meg. Ellenkező esetben azonban véglegesen törli azt a szabálybázisból.

A II. redukálási módszer [31][32] hasonló, mint az I., de azzal a különbséggel, hogy ebben az esetben a legnagyobb Q-értékkel rendelkező szabályok kerülnek vizsgálatra, feltételezve, hogy a nagyobb Q-értékkel rendelkező szabályok jelentősebb befolyással bírnak.

A III. módszer [31][32] nem egyesével vizsgálja a szabályokat, hanem szabálycsoportokat alakít ki, majd ezeket távoltja el. A szabálycsoportok kialakítása szintén Q-érték alapján történik, a módszer meghatározza a Q-értékek teljes tartományát (legkisebb és legnagyobb érték közötti értérést), majd ezen tartomány alapján hoz létre két szabálycsoportot, úgy hogy a tartomány fele lesz a tűréshatár. Ezt követően a nagyobb Q-értékkel rendelkező szabálycsoport kerül kiértékelésre. Ha ezzel a probléma még sikeresen megoldható, akkor az ebből a megmaradt szabálycsoportból kiindulva ismétlődik az eljárás. Ha nem oldható meg sikeresen, akkor a törölt szabályok visszakerülnek a szabályrendszerbe, de a vizsgált Q-érték tartomány újra megfeleződik. Ez addig ismétlődik, amíg a tűréshatár értéke olyan nem lesz, hogy az adott szabálycsoport már eltávolítható. Abban az esetben, ha a tűréshatás alapján csak egyetlen szabály marad a csoportban és a probléma így sem oldható meg, akkor ez a szabály fontos (állandó) jelölést kap, majd a továbbiakban ezen állandónak jelölt szabályokat már nem vizsgálja.

A IV. szabálybázis redukálási stratégia [19] klaszterezési módszeren alapszik. A teljes szabályrendszerből kiválasztásra kerül két pivot elem (p_1, p_2), amely a szabálybázis azon két szabálya amelyek egymástól mert távolsága a legnagyobb. A távolságok meghatározása a FIVE módszer által alkalmazott skálázott távolságon alapszik. Egy meghatározott távolságküszöb (ε) alapján, a két pivot elem által létrehozott klaszterhez hozzárendelődnek a szabályok. Az így létrejött két klaszterből (left-right branch) a legkisebb és legnagyobb Q-értékkel rendelkező szabály kerül kiválasztásra, mint lényegi szabály majd ellenőrzésre kerül, hogy a kiválasztott szabályok által a probléma megoldható-e. Ha nem, akkor a folyamat megismétlődik az így megmaradt két klaszterre, azaz ezen két klaszter további két alklaszterre (hierarchikus felosztó klaszterezés) oszlik, amelyekből ismét a legnagyobb és legkisebb Q-értékkel rendelkező szabály kerül megjelölésre fontos szabályként. A fontos szabályként megjelölt szabályok kikerülnek a klaszterekből és bekerülnek a fontos szabályok halmazába és megvizsgálja, hogy ezen szabályokkal a probléma megoldható-e. Ez a folyamat addig ismétlődik, amíg a létrejött fontos szabályok halmaza sikeresen meg nem oldja az adott problémát. Tehát kezdetben a fontos szabályok halmaza (redukált

szabálybázis) üres, majd az eljárás kettesével ad hozzá a távolságküszöb alapján a klaszterekből kiemelt lényegi szabályokat.
A szabálybázis redukció egy másik lehetséges megközelítése nem teljes szabályok elhagyása, vagy összevonása, hanem csak a szabályok egyes részeinek elhagyása. A fuzzy szabály interpoláció használatának köszönhetően egy szabály egyes antecedensei is függetlennek jelölhetőek. Az inkrementális szabálybázis építő módszer kezdeti szabálybázisa és az újonnan hozzáadott szabályaiban minden antecedens dimenzió megjelenik. Azonban olyan esetek is elfordulhatnak, ahol egy-egy antecedens dimenziót elhagyva is működőképes marad a rendszer. Azaz az érintett szabály ezen antecedens dimenziótól független. Több szabályból is elhagyva a független antecedens dimenziókat a szabálybázis tovább egyszerűsíthető. A független antecedensek feltárására mutat be módszereket [21][33].

4. A javasolt szabálybázis redukciós módszer

A szakértői tudásbázissal bővített FRIQ-learning rendszerben [17][18] a tanulási folyamat a szakértő által definiált szabályrendszer injektálásával 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 beszúrásra az éppen aktuális megfigyelés pozícióijába, ha a legközelebbi szabálypont is távol van a megfigyeléstől és a Q-frissítési érték nagyobb, mint az előre meghatározott ε_Q küszöbérték ($\Delta\tilde{Q} > \varepsilon_Q$). 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 (konzekvensük 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 redukciós (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 redukciós módszer jellemzői a következők:
- szabályok közötti távolság, azaz a „közelség” meghatározása történjen a Q függvény leírásánál is alkalmazott FIVE FRI fuzzy szabály távolság [20] alapján,
- a közel ugyanazt az információt leíró (egymáshoz közel kerülő) fuzzy szabályok összevonása, egyesítése történjen a szabálybázis építése és a hangolási folyamat során,
- a szabályok összevonásánál legyen lehetőség a vizsgált szabályok típusának figyelembevételére, azaz a szakértői szabályok súlyának (fontosságának) beszámítására,
- a javasolt szabálycsökkentési módszer használatával a 3. fejezetben bemutatott I.-IV. szabálybázis redukciós módszerek utólagos alkalmazása egyes esetekben el is hagyható.

4.1. A szabályközelség meghatározása

Az egymáshoz közeli szabályok egyesítéséhez szükséges a szabályok távolság mértékének meghatározása. Két szabály akkor vonható össze egyelten szabállyá, ha nagyon közel kerülnek egymáshoz. A szabályok közeliségének meghatározása egy dimenziónkénti (állapot és akció) távolságküszöb (dtr_j)

alapján történik. Ha az adott szabály távolsága az aktuális megfigyeléstől (vagy adott szabálytól) kisebb, mint ez a dimenziónként számított távolságküszöb (minden egyes antecedens dimenzióban), akkor a szabálypont közelinek tekinthető [20]:

$$\underset{i,j\in[1,r]}{\exists} i,j \;\; ahol \;\; \underset{n\in[1,N]}{\forall} (d_n(i,j) < dtr_n) \tag{7}$$

Azaz két szabály közeli, ha létezik olyan i és j szabálysorszám az $m + \hat{m}$ (FRIQ szabályok + szakértői szabályok) méretű szabályrendszerben, amire igaz, hogy minden egyes n-edik dimenzióban (az N dimenziószámú állapot-akció térben) az adott szabály $d_n(i,j)$ távolsága kisebb, mint a dtr_n távolságküszöb. A $d_n(i,j)$ az i és j szabályok közötti távolság az n-edik dimenzióban, amelyeket a D távolságmátrix tárol ($d_n(i,j) \in D$ és $i \in [1,m], j \in [1,N]$):

$$\begin{aligned} D_{[i,j]} &= d_n(i,j) = \left|s_j^i - S^i\right| \\ D_{[i,j+1]} &= d_n(i,j+1) = \left|a^i - A^i\right| \end{aligned} \tag{8}$$

Az akció és állapot dimenziókra számított távolságküszöbök értéke az adott dimenzió hossza elosztva a dimenziókban lévő elemek számával:

$$\begin{aligned} dtr_j(s_j) &= length(s_j)/(na_j + 1) \\ dtr_{j+1}(U) &= length(U)/(na_{j+1} + 1) \end{aligned} \tag{9}$$

Ahol $length(s_j)$ és $length(U)$ az állapot és akció dimenziók legkisebb és legnagyobb elemei közötti különbség abszolútértéke, azaz azok hossza:

$$\begin{aligned} length(s_j) &= \left|\max(S_j) - \min(S_j)\right| \\ length(U) &= \left|\max(U) - \min(U)\right| \end{aligned} \tag{10}$$

Összegezve tehát a (7) összefüggésnek eleget tevő szabályok tekinthetők egymáshoz közeli szabályoknak.

4.2. Közeli szabályok egyesítése

A tanulási fázis szabálybázis hangolási eljárása [16] során (szabályvándorlás), az egymáshoz közel kerülő szabályok egyesítésével (összevonásával) az adott iterációban a szabálybázis mérte csökkenthető.

A szakértői tudásbázissal kiegészített FRIQ-learning rendszerben a szabályok két típusa különböztethető meg: a szakértő által megadott szabályrendszer (R_{expert}) és a rendszer által létrehozott szabályrendszer (R). A rendszer által létrehozott szabályrendszer tartalmazza az interpolációs eljárás miatt szükséges 2^{n+1} darabszámú sarokponti szabályt, valamint lehetnek benne újonnan beszúrt (felvett) szabályok. Mivel a szakértői szabályok feltételezhetően helyesek, így azokat a szabályegyesítés során nagyobb súllyal célszerű kezelni.

Egy a szakértői által definiált szabályt jelöljünk $\hat{r}$-el ($\hat{r} \in R_{expert}$), egy a rendszer által felvett szabályt r-el ($r \in R$), egy sarokponti szabályt pedig $r^{\square}$-el ($r^{\square} \in R$).

Két egymáshoz közeli szabály egyesítésének az alapötlete a következő: ha a két szabály közül az egyik szakértői a másik pedig a rendszer által felvett szabály volt, akkor az új egyesített szabály maradjon a szakértői szabály. Ha a két szabály közül mindkét szakértői szabály, akkor az új összevont szabály is

legyen szakértői szabály. Ha pedig a két szabály közül mindkét szabály a rendszer által újonnan felvett szabály akkor az új egyesített szabály is legyen újonnan beszúrt szabályként jelölve. A (11) ezt összegzi, ahol a „⊔" operátor két szabály egyesítését (concatenation), a „→" operátor pedig az így előállt művelet eredményét szemlélteti.

$$\begin{aligned} \hat{r} \sqcup r &\to \hat{r} \\ \hat{r} \sqcup \hat{r} &\to \hat{r} \\ r \sqcup r &\to r \end{aligned} \tag{11}$$

Az új szabálypont, azaz az egyesítés után létrejövő szabály antecedensének és konzekvensének meghatározása az egyesített két szabály antecedensének és konzekvensének átlagolásával történik. Mivel a szakértő által definiált szabályrendszer formátuma (4) a FRIQ-learning-be történő beágyazása után megegyezik a rendszer szabályaival (1), így az új szabály antecedensének és konzekvensének számítása a rendszer szabályainak egyesítésével azonos módon történhet. Az alapötlet az, hogy az új szabály antecedense (állapot-akció értéke) és következménye (Q-értéke) legyen az egyesített szabályok antecedens és konzekvens értékeinek az átlaga. Az alábbi táblázat egy lehetséges szabályegyesítési példát szemléltet, ahol $\hat{r}$ egy szakértői szabály, r pedig egy rendszer által beszúrt 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 $\to \hat{r}$ 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	s_2	a	q
$\hat{r}$	1	2	4	123.45
r	2	2	5	145.23
$\hat{r} \sqcup r \to \hat{r}$	1.5	2	4.5	134.34

5. Összegzés

Kifejlesztésre került egy olyan módszer, amely a szakértői tudásbázissal bővített FRIQ-learning megerősítéses tanulási rendszerben alkalmas lehet a tanulási folyamat során a rendszer tudásbázisát leíró fuzzy szabályrendszer méretének csökkentésére. A javasolt módszer alapötlete, hogy az inkrementális szabálybázis építés és hangolás során az egymáshoz közel kerülő szabályokat. amennyiben azok hasonló következtetésre vonatkoznak, egyesíteni kell. A javasolt szabályegyesítési módszer figyelembe veszi az egyesítendő szabályok típusát és külön kezeli azokat az eseteket, amikor valamelyik egyesítendő szabály szakértői szabály. Ezesetben az egyesített szabály is szakértői szabály típusú lesz. Így a szakértői szabályok hangolás során történő változása, esetleg összeolvadása nyomonkövethető.

6. Köszönetnyilvánítás

„A cikkben ismertetett kutató munka az EFOP-3.6.1-16-2016-00011 jelű „Fiatalodó és Megújuló Egyetem – Innovatív Tudásváros – a Miskolci Egyetem intelligens szakosodást szolgáló intézményi fejlesztése" projekt részeként – a Széchenyi 2020 keretében – az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg."

Irodalomjegyzék

[1] Appl, M.: Model-based Reinforcement Learning in Continuous Environments. Ph.D. thesis, Technical University of München, München, Germany, dissertation.de, Verlag im Internet (2000)

[2] Bartók, R., Vásárhelyi, J.: "A fuzzy rule interpolation base algorithm implementation on different platforms." *Proceedings of the 2015 16th International Carpathian Control Conference (ICCC)*. IEEE, 2015. **https://doi.org/10.1109/CarpathianCC.2015.7145041**

[3] Bartók, R., Vásárhelyi, J.: "Fuzzy Rule Interpolation Based Object Tracking and Navigation for Social Robot." *Vehicle and Automotive Engineering*. Springer, Cham, 2018. **https://doi.org/10.1007/978-3-319-75677-6_31**

[4] Bartók, R., Vásárhelyi, J.: "Parallelization of FIVE method on multicore embedded system." *2018 19th International Carpathian Control Conference (ICCC)*. IEEE, 2018. **https://doi.org/10.1109/CarpathianCC.2018.8399663**

[5] Berenji, H.R.: Fuzzy Q-Learning for Generalization of Reinforcement Learning. Proc. of the 5th IEEE International Conference on Fuzzy Systems (1996) pp. 2208-2214.

[6] 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. **https://doi.org/10.1007/s10732-007-9031-5**

[7] Glorennec, P. Y., & Jouffe, L. (1997, July). Fuzzy Q-learning. In Proceedings of 6th international fuzzy systems conference (Vol. 2, pp. 659-662). IEEE.

[8] 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. **https://doi.org/10.1109/3477.552182**

[9] 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.

[10] 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.

[11] Kovács, Sz., Vincze, D., Gácsi, M., Miklósi, Á., Korondi, P.: Ethologically Inspired Robot Behavior Implementation, in Proceedings of the 4th International Conference on Human System Interaction (HSI 2011), Keio University, Yokohama, Japan, pp. 64-69, May 19-21, 2011. **https://doi.org/10.1109/HSI.2011.5937344**

[12] Kovács, Sz., Vincze, D., Gácsi, M., Miklósi, Á., Korondi, P.: Fuzzy automaton based Human-Robot Interaction, in Proc. of the 8th IEEE International Symposium on Applied Machine Intelligence and Informatics (SAMI), 28-30 Jan 2010, pp. 165-169. **https://doi.org/10.1109/SAMI.2010.5423746**

[13] 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.

[14] Rummery, G. A., Niranjan, M.: On-line Q-learning using connectionist systems. CUED/F-INFENG/TR 166, Cambridge University, UK., 1994

[15] Sutton, R. S., Barto, A. G.: Reinforcement Learning: An Introduction, MIT Press, Cambridge (1998) **https://doi.org/10.1109/TNN.1998.712192**

[16] Tompa, T., Kovács, Sz.: "Expert heuristic tuning design for the FRIQ-learning." Multidiszciplináris Tudományok 10.4 (2020): 119-125. **https://doi.org/10.35925/j.multi.2020.4.15**

[17] Tompa, T., Kovács, Sz.: "Szakértői heurisztika alkalmazása a FRIQ-learning megerősítéses tanulási módszerben." Multidiszciplináris Tudományok 9.4 (2019): 356-368. **https://doi.org/10.35925/j.multi.2019.4.35**

[18] Tompa, T., Kovács, Sz.: "Applying Expert Heuristic as an a Priori Knowledge for FRIQ-Learning." *Acta Polytechnica Hungarica* 17.4 (2020). **https://doi.org/10.12700/APH.17.4.2020.4.2**

[19] Tompa, T., Kovács, Sz.: "Clustering-based fuzzy knowledge-base reduction in the FRIQ-learning." *Applied Machine Intelligence and Informatics (SAMI), 2017 IEEE 15th International Symposium on*. IEEE, 2017. **https://doi.org/10.1109/SAMI.2017.7880302**

[20] Tompa, T., Kovács, Sz.: "Determining the minimally allowed rule-distance for the incremental rule-base contruction phase of the FRIQ-learning." *2018 19th International Carpathian Control Conference (ICCC)*. IEEE, 2018. **https://doi.org/10.1109/CarpathianCC.2018.8399677**

[21] Tóth, A., Vincze, D.: Antecedens redundancia feltárása a fuzzy szabály interpoláció alapú megerősítéses tanulási módszerben, Multidiszciplináris Tudományok, Évf. 9. szám 4. (2019), pp. 523-534. **https://doi.org/10.35925/j.multi.2019.4.56**

[22] Tóth, A., Vincze, D.: Futball szimuláció megvalósítása fuzzy szabály interpoláció alapú fuzzy automatával, Multidiszciplináris Tudományok, Évf. 9. szám 1. (2019), pp. 12-22. **https://doi.org/10.35925/j.multi.2019.1.2**

[23] Vincze, D. “Parallelization by vectorization in Fuzzy Rule Interpolation adapted to FRIQ-Learning”. Proc. *2018 World Symposium on Digital Intelligence for Systems and Machines (DISA)*. IEEE, 2018. pp. 131-136. **https://doi.org/10.1109/DISA.2018.8490614**

[24] Vincze, D. and Kovacs, Sz., "Using fuzzy rule interpolation based automata for controlling navigation and collision avoidance behaviour of a robot," 2008 IEEE International Conference on Computational Cybernetics, Stara Lesna, 2008, pp. 79-84, **https://doi.org/10.1109/ICCCYB.2008.4721383**

[25] Vincze, D., Tóth, A. and Niitsuma, M., "Football Simulation Modeling with Fuzzy Rule Interpolation-based Fuzzy Automaton," 2020 17th International Conference on Ubiquitous Robots (UR), Kyoto, Japan, 2020, pp. 87-92, **https://doi.org/10.1109/UR49135.2020.9144752**

[26] Vincze, D., and Kovács Sz. "Performance Optimization of the Fuzzy Rule Interpolation Method" FIVE"." *JACIII* 15.3 (2011): 313-320. **https://doi.org/10.20965/jaciii.2011.p0313**

[27] Vincze, D., Kovacs, Sz. „Behaviour Based Control with Fuzzy Automaton in Vehicle Navigation”. *Production Systems and Information Engineering*, 2009, vol. 5., pp. 151-166.

[28] Vincze, D., Kovacs, Sz., “Performance issues of the implemented FRI ‘FIVE’”, Proc. *2010 11th International Symposium on Computational Intelligence and Informatics (CINTI)*. IEEE, 2010. p. 131-136. **https://doi.org/10.1109/CINTI.2010.5672259**

[29] Vincze, D., Kovács, Sz.: "Fuzzy rule interpolation-based Q-learning." Applied Computational Intelligence and Informatics, 2009. SACI'09. 5th International Symposium on. IEEE, 2009. **https://doi.org/10.1109/SACI.2009.5136311**

[30] 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 Heilderberg, 2010, pp. 191-203. **https://doi.org/10.1007/978-3-642-15220-7_16**

[31] 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. **https://doi.org/10.1109/SACI.2009.5136311**

[32] 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. https://doi.org/10.17667/riim.2015.1-2/10.

[33] Vincze, D., Tóth A., and Niitsuma, M., "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**

[34] Vincze, D.: Fuzzy Rule Interpolation and Reinforcement Learning, 15th International Symposium on Applied Machine Intelligence and Informatics (SAMI 2017), Herl'any, Slovakia, pp. 173–178. **https://doi.org/10.1109/SAMI.2017.7880298**

[35] Watkins, C. J. C. H.: Learning from Delayed Rewards. Ph.D. thesis, Cambridge University, Cambridge, England (1989)