*Multidiszciplináris tudományok, 12. kötet, 1. sz. (2022), pp. 90-102* **https://doi.org/10.35925/j.multi.2022.1.8**

# SZABÁLYTÁVOLSÁG ALAPÚ SZABÁLYBÁZIS-REDUKCIÓ A SZAKÉRTŐI TUDÁSBÁZISSAL BŐVÍTETT FRIQ-LEARNING KÖRNYEZETBEN

**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 szakértői tudásbázissal bővített fuzzy szabály-interpoláció alapú Q-learning (FRIQ-learning) megerősítéses tanulási rendszerben a szakértői által definiált tudásbázis a tanulási folyamat során kiegészül a rendszer által létrehozott fuzzy szabályokkal. A rendszer a tanulási folyamat során hangolja (és optimalizálja) a tudásbázist leíró fuzzy szabályok antecedensét és konzekvensét is, amely következtében előfordulhat olyan eset, hogy több szabály közel kerül egymáshoz. Az egymáshoz közel kerülő szabályok összevonásával (redukálásával) a rendszer tudásbázisának a mérete csökkenthető. A tudásbázis redukálási módszer a szabályok közötti távolságot (és így a szabályközelség mértékét) az antecedens (állapot-akció) dimenzióban határozza meg. Ez azonban problémát okozhat olyan esetekben mikor a szabályok antecedens univerzumokban közelinek számítanak, azonban a konzekvensükben (Q-érték) nagy eltérés mutatkozik, tehát az általuk leírt Q-függvényben meredek lejtő található. Jelen cikk célja szakértői heurisztikával bővített FRIQ-learning tudásbázis redukálási módszerének kiegészítése (finomítása) olyan módon, hogy az a szabályközelség (és egyben a szabálytávolság) meghatározásánál a szabályok konzekvens (Q-érték) dimenzióját is figyelembe vegye.*

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

***Abstract***

*In the expert knowledge-included Fuzzy Rule Interpolation-based Q-learning (expert knowledge-included FRIQ-learning) reinforcement learning system by the expert defined knowledgebase will be extended by the system created fuzzy rules during the learning phase. The system tune (and optimize) the antecedent and consequent parts of fuzzy rules during the learning iterations, due to this method can be the case when any rules will be close to each other. The size of the knowledgebase of the system can be reduced by merging the closing rules. Based on the knowledgebase reduction methodology, the distance between the rules (therefore the measure of rule closure) is determined only in antecedent (state-action) universes. However, it can be problematic in the cases when the rules can be determined as closing rules in the antecedent universes but there is a large difference in the consequent dimension, thus there is a steep slope of the Q-function. The main contribution of the paper is to introduce a modification of the knowledge reduction methodology that will extend the distance determination to the consequent (Q-value) universe as well.*

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

## 1. Bevezetés

A megerősítéses tanulás (Reinforcement Learning – RL) (Sutton et al., 1998) olyan módszerek összessége, melyek által egy ágens (tanuló entitás) a környezetből érkező megerősítési információk alapján képes tanulni, azaz feltérképezni egy adott probléma megoldásához vezető utat.

Az ágens a végrehajtott cselekedetei (akciói) alapján a környezettől megerősítéseket (jutalmat vagy büntetést) kap melyek következtében megváltozik az állapota. Az ágens célja, hogy hosszútávon maximalizálja a pozitív megerősítéseit (jutalmait), melyek által egyre nagyobb jutalommal járó akciókat igyekszik végrehajtani. A jutalmak vagy büntetések numerikus értékek, melyek értékének meghatározása a jutalomfüggvény által történik. A lehetséges állapotokban végrehajtható akciók minőségét a Q-függvény (állapot-akció függvény) írja le, amely egyben a rendszer tudásbázisát reprezentálja.

A klasszikus RL-módszerek (Q-learning, SARSA, Fuzzy Q-learning) mindegyike a tanulási folyamat elején egy teljesen üres tudásbázissal rendelkezik, majd ez az üres kezdeti tudásbázis bővül iterációról-iterációra a környezet megerősítési információ alapján. Ezen elterjedtebb megerősítés tanulási módszerekről a (Kaelbling et al., 1996) irodalom ad bővebb áttekintést.

A Q-függvény leírási módja RL módszerenként eltérő lehet, a diszkrét állapot- és akciótér felontással rendelkező Q-learning (Watkins, 1989) illetve SARSA (Rummery et al., 1994) esetében egy Q-tábla (lookup tábla), fuzzy modellel történő folytonos állapot- és akciótérre való kiterjesztés (Glorennec et al., 1997; Berenji, 1996) esetében pedig egy fuzzy szabályrendszer (szabálybázis) írja le. A tudásbázist reprezentáló Q-tábla vagy Fuzzy Q-learning módszerek esetében a fuzzy szabálybázis mérete jelentősen függ az adott probléma méretétől, azaz az állapot- és akcióváltozók számától (dimenziószámától). Ezen algoritmusok csak adott méretű probléma (kb. 10 000 állapot) esetében alkalmazhatók hatékonyan, tudásbázisukat reprezentáló Q-tábla vagy fuzzy szabálybázis mérete a probléma dimenziószámával exponenciálisan növekszik (Kóczy et al., 1997). Különféle módszerek alkalmazhatók ezen negatívum kiküszöbölésére, ilyen például a Fuzzy Q-learning módszerek esetében alkalmazható fuzzy szabály-interpolációs eljárások (Fuzzy Rule Interpolation – FRI) amelyek által a fuzzy szabálybázis mérete csökkenthető. Az elterjedtebb FRI-eljárásokról a (Johanyák et al., 2006) irodalom ad bővebb áttekintést. Azonban az FRI-modellt alkalmazó megerősítéses tanulási módszerek tudásbázisában előfordulhatnak olyan fuzzy szabályok, melyek kiadódhatnak más szabályokból. Ezen redundáns szabályok elhagyásával a tudásbázis mérete tovább csökkenthető. A fuzzy szabály-interpoláció alapú Q-learning (Fuzzy Rule Interpolation based Q-learning – FRIQ-learning) (Vincze et al., 2009) módszerben különféle szabálybázis redukálási stratégiák alkalmazhatók ezen kiadódó (és így elhagyható) szabályok keresésére (Vincze et al., 2009) (Tompa et al., 2017).

A szakértői tudásbázissal bővített FRIQ-learning rendszerben (Tompa et al., 2020) a szabálybázis redukálása (csökkentése) oly módon történik, hogy az egymáshoz közeli szabályok egyesítésre kerülnek már a tanulási folyamat során (Tompa et al., 2021), így a tanulási folyamat végeztével már egy minimális szabályszámú tudásbázis keletkezhet. Az egymáshoz közeli szabályok egyesítésének alapja a szabályok között lévő távolság, illetve az ezek alapján meghatározott távolságküszöbök (Tompa et al., 2018). Azonban a szabályok közötti távolság meghatározása csak az antecedens (állapot-akció) univerzumok alapján történik, a konzekvens (Q-érték) univerzum figyelmen kívül hagyásával, amely bizonyos esetekben problémákat okozhat (például meredek lejtővel rendelkező Q-függvény esetében).

Jelen cikk célja szakértői tudásbázissal bővített FRIQ-learning rendszerben, a tanulási fázis során alkalmazott szabálybázis redukálási (szabály összevonási) módszer kibővítése, pontosítása oly módon, hogy a szabályok közötti távolságok meghatározása (és így a szabályösszevonás) a fuzzy szabályok konzekvens univerzumára is kiterjesztésre kerüljön.

## 2. A szakértői heurisztikával bővített FRIQ-learning

A szakértői tudásbázissal bővített fuzzy szabály-interpoláció alapú Q-tanulás (expert knowledge-included Fuzzy Rule Interpolation-based Q-learning) (Tompa et al., 2020) egy olyan fuzzy szabály-interpolációs eljárást (Fuzzy Rule Interpolation – FRI) alkalmazó, folytonos állapot-akció térre kiterjesztett Q-learning módszer, amely alkalmas szakértői által megadott előzetes tudásbázis rendszerbe történő injektálására.

A módszeralapja az FRIQ-learning (Vincze et al., 2009) algoritmus, amely a FIVE (Fuzzy Rule Interpolation based on Vague Environment) FRI-eljárást (Kovács et al., 1997) alkalmazza az univerzumok folytonos térre való kiterjesztéséhez. A FIVE egy kis számításigényű (Vincze, 2018; Vincze et al., 2010), gyakorlatorientált alkalmazásokban jól használható (Kovács et al., 2011; Vincze et al., 2008) FRI-módszer.

Az FRIQ-learning rendszer tudásbázisát ($R$) leíró ($r_i, i \in [1, m]$ fuzzy szabályok formája a következő:

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

ahol $r_i \in R$ ($i \in [1, m]$) 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, $A^i$ az $i$-edik szabály fuzzy halmaza az egydimenziós $U$ akciótérben, $a$ ($a \in U$) pedig a végrehajtott akció.

Szakértői által megadott tudásbázis leírására szintén fuzzy szabályok formájában van lehetőség, melyek definiálják, hogy az adott állapotban mely akció végrehajtása preferált. Egy $\hat{r}_i$ ($i \in [1, \hat{m}]$) szakértő által definiált szabály formája a következő (Tompa et al., 2020):

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

ahol $\hat{A}^i$ az $i$-edik ($i \in [1, \hat{m}]$) szakértői szabály akciója (konzekvense), $\hat{S}_n^i = \left[\hat{S}_1^i, \hat{S}_2^i, \dots \hat{S}_n^i\right]$ az $n$-dimenziós állapot megfigyelés, $\hat{m}$ a szakértői szabályok száma az $R_{expert}$ szakértői szabályrendszerben, $\hat{r}_i \in R_{expert}$ pedig az $i$-edik szakértői szabály. Ezen szabályok formája megegyezik az (1) által definiált szabályokéval, azzal az eltéréssel, hogy ebben az esetben a szabályantecedens az állapot, a konzekvens pedig az ebben az állapotban preferált akció. A rendszerbe történő injektálás során ezen szabályok antecedense az állapot-akció, konzekvense pedig egy kezdeti becsült Q-érték lesz. A kezdeti Q-érték meghatározása a következő összefüggés alapján történik (Tompa T. et al., 2020):

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

ahol $\tilde{Q}_{init}$ a számított kezdeti Q-érték, $g_{max}$ a környezet által adható maximális megerősítés, $\gamma$ a leszámítolási tényező, $\eta \in [0,1]$ pedig a $\tilde{Q}_{init}$ értékre vonatkozó skálatényező.

A kezdeti Q-érték meghatározását követően a szakértői szabályok formátuma az (1)-el megegyező lesz, azzal az eltéréssel, hogy a konzekvensük a $\tilde{Q}_{init}$ érték, antecedensük meg a megadott állapot-akció érték lesz. A szakértői szabályrendszer injektálást követően a módszer kiegészíti (majd hangolja) ezen szabályrendszert a 0 konzekvens értékkel rendelkező $2^{n+1}$ ($n$ az állapotdimenziók száma) darabszámú

sarokponti szabállyal. Ha valamely szakértői szabály antecedense illeszkedne valamely sarokponti szabályra (sarokponti szabálypontot talál el), akkor a sarokponti szabály $q^i = 0$ konzekvense lecserélésre kerül a szakértői szabály $q^i = \tilde{Q}_{init}$ konzekvensére, azaz a sarokponti szabály lecserélésre kerül a szakértői szabályra.

A tanulási folyamat során a szabályrendszer konzekvensei iterációról-iterációra a következő frissítési formula szerint változnak:

$$q_i^{k+1} = \begin{cases} q_i^k + \Delta\tilde{Q}^{k+1}(\boldsymbol{s}, a) & \text{ha } (\boldsymbol{s}, a) = (\boldsymbol{s}^i, a^i) \text{ valamennyi } i\text{-re,} \\ q_i^k + \Delta\tilde{Q}^{k+1}(\boldsymbol{s}, a) * \left(1/\delta_{v,i}^{\lambda}\right) \Big/ \left(\sum_{i=1}^{m} 1/\delta_{v,i}^{\lambda}\right) & \text{egyébként} \end{cases} \quad (4)$$

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:

$$\Delta\tilde{Q}^{k+1}(\boldsymbol{s}, a) = \alpha * \left(g(\boldsymbol{s}, a, \boldsymbol{s}') + \gamma * \max_{a' \epsilon U} \tilde{Q}^k(\boldsymbol{s}', a') - \tilde{Q}^k(\boldsymbol{s}, a)\right) \quad (5)$$

ahol $\alpha \in [0,1]$ a tanulási ráta, $\gamma$ a leértékelési tényező, $g(\boldsymbol{s}, a, \boldsymbol{s}')$ a megerősítés értéke az $\boldsymbol{s} \to \boldsymbol{s}'$ 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 fázis során a kezdeti szabályrendszer új szabályok beszúrásával inkrementálisan növekszik. Új szabály akkor kerül felvételre a szabályrendszerbe, ha a megfigyeléshez legközelebb lévő szabály távolinak tekinthető, azaz a távolsága nagyobb, mint egy küszöbérték (Tompa et al., 2018), illetve a Q-frissítés értéke is nagyobb, mint egy előre definiált küszöbérték (Vincze et al., 2010). Ellenkező esetben, ha a megfigyelés közelében található már létező szabály, akkor a teljes szabályrendszer konzekvense frissítésre kerül. A tanulási folyamat akkor ér véget, ha a szabálybázisba már nem kerül új szabály beszúrásra és a Q-frissítés értéke is elenyészően kicsi. A tanulási folyamat végeztével előállt inkrementális szabálybázis tartalmazhat olyan szabályokat melyeknek csak a tanulási fázis során volt jelentőségük és emiatt redundánsak, kiadódhatnak más szabályokból. Ezen szabályok törlésére a szabálybázis redukálási módszerek alkalmazhatók, melyek által a szabálybázis mérete csökkenthető (Vincze et al., 2009; Vincze et al., 2020). Ha azonban az elhagyható szabályok keresése már a tanulási folyamat során megvalósul olyan módon, hogy ha a tanulási fázis során két (esetleg több) szabály közel kerül egymáshoz akkor ezen szabályok egyesítésével a szabálybázis mérete tovább csökkenthető és már a tanulási folyamat végeztével egy minimális szabályszámú szabálybázis keletkezhet (Tompa et al., 2021).

## 3. Szabályközelség mértékének meghatározása

Az egymáshoz közeli szabályok egyesítéséhez, illetve új szabály felvétele esetében is szükséges a szabályok közötti távolság meghatározása. Két szabály akkor vonható össze egyelten szabállyá (azaz redukálható), ha azok nagyon közel kerülnek egymáshoz, illetve új szabály akkor kerül beszúrásra a szabálybázisba, ha a hozzá legközelebb eső szabály is távol van, azaz távolsága nagyobb, mint egy meghatározott távolságmérték.

A szabályok közeliségének meghatározása egy antecedens dimenziónként számított $\boldsymbol{dtr} = [dtr_1, dtr_2, \dots, dtr_n, dtrU]$ távolságküszöb alapján történik, ahol $dtr_1, dtr_2, \dots, dtr_n$ az állapotdimenzióra, $dtrU$ pedig az akció dimenzióra számított távolságküszöb. Ez által a $\boldsymbol{dtr}$ vektor elemszáma megegyezik az antecedens dimenziók számával, azaz $(n+1)$. A távolságküszöbök számításának alapja

a szintén ancetedens dimenziónként meghatározott $d_j$ távolság, ahol a $j$ az antecedens univerzumok számát ($j \in [1, n+1]$) jelöli.

Ha a szabálybázis egy szabályának távolsága az aktuális megfigyeléstől (állapot-akció ponttól) minden egyes állapot-akció dimenzióban kisebb, mint az adott dimenziókra számított $dtr_j$ ($j \in [1, n+1]$) távolságküszöb értéke, akkor a szabálypont közelinek tekinthető az adott megfigyeléshez (Tompa et al., 2018):

$$\underset{t,p\in[1,m+\hat{m}]}{\exists} t,p \;\; ahol \underset{j\in[1,n+1]}{\forall} (d_j(t,p) < dtr_j) \tag{6}$$

Azaz két szabály közelinek tekinthető, ha létezik olyan $t$ és $p$ szabálysorszám az $m+\hat{m}$ méretű szabályrendszerben ($m$ számosságú FRIQ szabályok + $\hat{m}$ számosságú szakértői szabályok), amire igaz, hogy minden egyes $j$-edik ($j \in [1, n+1]$) antecedens dimenzióban (az $n+1$ dimenziószámú állapot-akció térben) az adott szabály $d_j(t,p)$ távolsága kisebb, mint a $dtr_j$ távolságküszöbök értéke. A $d_j(t,p)$ a $t$ és $p$ indexű szabályok ($t,p \in [1, m+\hat{m}]$) közötti távolságot jelöli a $j$-edik antecedens dimenzióban ($j \in [1, n+1]$):

$$\begin{aligned} d_j(t,p) &= \left| s_j^t - s_j^p \right| && j \in [1,n] \\ d_j(t,p) &= |a^t - a^p| && j = n+1 \end{aligned} \tag{7}$$

ahol $s_j^t$ a $j$-edik antecedens fuzzy halmaza a $t$-edik indexű szabálynak, $s_j^p$ a $j$-edik antecedens fuzzy halmaza a $p$-edik indexű szabálynak, $a^t$ $t$-edik indexű szabály $a^p$ pedig a $p$-edik indexű szabály akciója, $|\,|$ az abszolútértéket jelöli. A $t$-edik és $p$-edik indexű szabály közötti távolság tehát nem egyetlen szám, hanem egy vektor ($[d_1(t,p), d_2(t,p), \dots, d_n(t,p), d_{n+1}(t,p)]$), amely elemszáma megegyezik az antecedens dimenziók számával ($n+1$) és dimenziónként tartalmazza az adott szabályok antecedensei közötti távolságokat.

Az egymáshoz közeli szabályok meghatározásához a szabálybázis összes szabálya közötti távolság meghatározására szükség van. Egy többdimenziós $D$ távolságmátrix tartalmazza a szabálybázis mindegyik szabálya közötti távolságot minden egyes antecedens dimenzióban. Mivel mindegyik szabály önmagától számított távolsága 0, így a többdimenziós mátrix főátlójában csupa nullák szereplenek, illetve a $(t,p)$ indexű szabályok közötti távolság az ugyanaz, mint a $(p,t)$ indexűek közötti ($d_j(t,p) = d_j(p,t)$), így a számításoknál a főátló alatti elemeket (alsóháromszög mátrix) szükséges figyelembe venni.

Az akció és állapot dimenziókra számított távolságküszöbök értéke az adott dimenzió hossza elosztva egy úgynevezett közelségaránnyal (*distance rate – dR*), amely a szakértő által definiált (akár univerzumonként eltérő) $\mathrm{dR}_S$ és $\mathrm{dR}_U$ numerikus konstans értékek. Tehát a távolságküszöbök értéke az adott dimenziók hosszának valamekkora része, amelyek ennek következtében a szabályok között megengedett minimális távolságot határozzák meg:

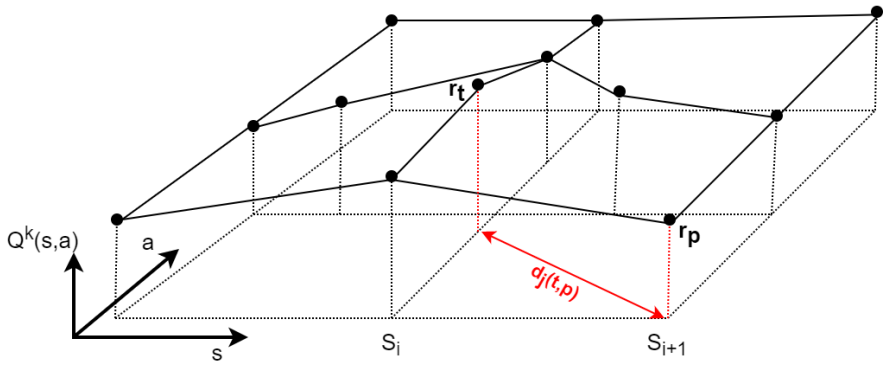
$$\begin{aligned} dtr_j &= \frac{length(S_j)}{\mathrm{dR}_S} && j \in [1,n] \\ dtrU &= \frac{length(U)}{\mathrm{dR}_U} && j = n+1 \end{aligned} \tag{8}$$

Ahol $length(S_j)$ az állapotuniverzum, $length(U)$ pedig az akció dimenzió legkisebb és legnagyobb elemei közötti különbség abszolútértéke, azaz az értelmezési tartományuk hossza:

$$length(s_j) = \left|\max(S_j) - \min(S_j)\right|$$
$$length(U) = \left|\max(U) - \min(U)\right| \tag{9}$$

Ahol $\max(S_j)$ a maximum (legnagyobb), $\min(S_j)$ pedig a minimum (legkisebb) eleme a $j$-edik ($j \in [1, n]$) $S_j$ állapot és $U$ akció univerzumnak.

A következő ábra a $t$ és a $p$ indexű szabályok közötti $d_j(t,p)$ antecedens dimenzióbeli távolságot szemlélteti:




***1. ábra.*** *A $t$ és a $p$ indexű szabályok közötti $d_j(t,p)$ távolság az antecedens dimenzióban*

Összegezve tehát a (6) összefüggésnek eleget tevő $t$ és a $p$ indexű szabálypárok tekinthetők egymáshoz közeli szabályoknak. Az algoritmus pszeudokódja az alábbi (Tompa et al., 2018):

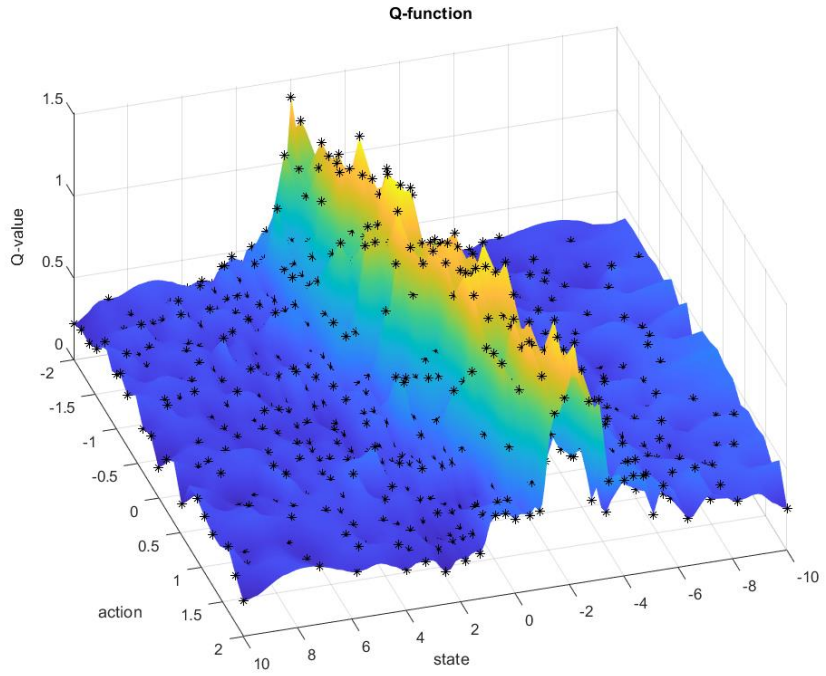
```
Input: the actual observation and rules of the rule-base
Output: true or false (the observation is a close rule or not)

1. initialize the distance matrix (D)
2. compute the rule distances (d) from the actual observation
3. compute the distance thresholds (dtr) each antecedent dimensions
4. examine the distances (d) of given rule in the rule-base
5. If ∃ t,p∈[1,m+m̂] t,p where ∀ j∈[1,n+1] (d_j(t,p) < dtr_j) ,j ∈ [1,n+1]
     a. true: the observation can be considered as a close rule
     b. false: the observation cannot be considered as a close rule
```

## 4. Szabályközelség mértékének kiterjesztése a konzekvens univerzumra

A tanulási fázis során a szabálybázis hangolási eljárása (Tompa et al., 2020) következtében a szabályok állapot-akció pontja (nem csak a konzekvense) elmozdulhat, amely következtében előfordulhat olyan eset, mikor két vagy esetleg több szabály közel kerül egymáshoz. Az egymáshoz nagyon közel kerülő szabályok közel ugyanazt az információt írják le, így ezek egyesítésével (összevonásával) a szabálybázis mérte csökkenthető, redukálható.

Az egymáshoz közel kerülő szabályok összevonásának alapja az előző fejezetben bemutatott $\boldsymbol{dtr}$ távolságmérték. Ez az antecedens dimenziókra meghatározott $\boldsymbol{dtr}$ távolságküszöb azonban nem minden

esetben lehet elegendő közelség mértékének definiálásához, mert előfordulhat olyat eset mikor két szabály ezen távolságküszöb alapján (azaz az antecedens dimenzióban) egymáshoz közelinek számít a (6) összefüggés alapján, de a konzekvensükben (Q-értékükben) nagy eltérés található, így a konzekvens univerzumban távoliknak tekinthetők. Ebben az esetben nem célszerű a két (forrás) szabályt egymáshoz képest közelinek tekinteni, majd ez alapján összevonni őket és egyetlen szabállyá redukálni mert az általuk leírt függvényben ennek következtében meredek lejtő vagy emelkedő lehetséges. Egy ilyen lehetséges meredek „hegyoldallal" rendelkező függvényt a következő ábra szemléltet, ahol a * (csillag) jel által jelölt pozíciókban a szabálypontok találhatók:




***2. 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 közelinek, ha azok a távolságmértékek 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:

$$d_Q(t,p) = |Q^t - Q^p| \tag{10}$$

A konzekvens univerzumbeli közelségmérték meghatározása szintén egy távolságküszöb alapján történik, emiatt a $\boldsymbol{dtr}$ vektor kiegészül a konzekvens dimenzióra is vonatkozó $dtrQ$ távolságküszöbbel, így $\boldsymbol{dtr} = [dtr_1, dtr_2, \dots, dtr_n, dtrU, dtrQ]$. Ennek értéke a teljes (éppen aktuális) Q-érték tartomány

valamekkora része, azaz a legnagyobb és a legkisebb Q-érték különbségének (a teljes tartomány hosszának) valamekkora, a szakértő által definiált $dR_q$ része:

$$length(Q) = |\max(Q) - \min(Q)| \tag{11}$$

$$dtrQ = \frac{length(Q)}{dR_q} \tag{12}$$

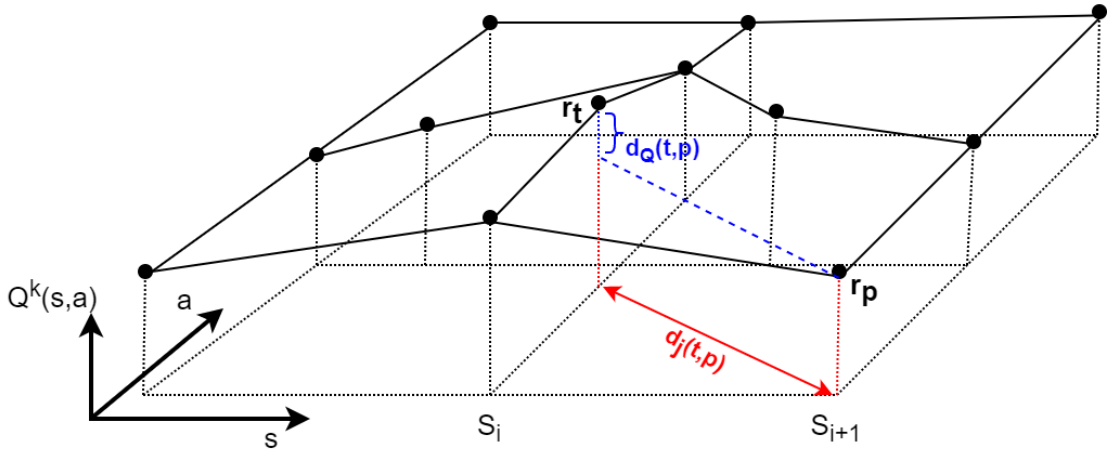
Mivel a Q-értékek a tanulási fázis során iterációnként változnak (a módszer által hangolásra kerülnek), így a $length(Q)$ értéke minden egyes olyan iterációban újraszámításra kerül, ahol szabálybázis-redukálás történik.

Összegezve, a szabálybázis-redukálás során akkor tekintjük a szabálybázis két $t$ és $p$ indexű $r_t$ és $r_p$ szabályát egymáshoz közelinek és ennek következtében akkor kerülnek összevonásra (redukálásra) egyetlen szabályként, ha a (6) összefüggés teljesül rájuk és még a két szabály Q-értékében (konzekvensében) 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 $dtrQ$ küszöbérték:

$$\underset{t,p\in[1,m+\hat{m}]}{\exists} t,p \text{ hogy } d_Q(t,p) < dtrQ \tag{13}$$

Ellenkező esetben, ha az antecedens dimenzióban közelinek tekinthetőek, de a konzekvens dimenzióban távolinak, akkor nem kerül összevonásra a $t$ indexű $r_t$ és $p$ indexű $r_p$ szabály egyetlen szabállyá.

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




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

A szabálybázis-redukálás során az új, egyesített szabály állapot-akció pontjának, illetve konzekvensének meghatározása a két ($t$ és $p$ indexű) forrásszabály antecedens és konzekvens értékeinek az átlagolásával történik.

A kifejlesztett, tanulási folyamat közben alkalmazható szabályösszevonás (szabálybázis-redukálás) algoritmusának pszeudokódja az alábbi:

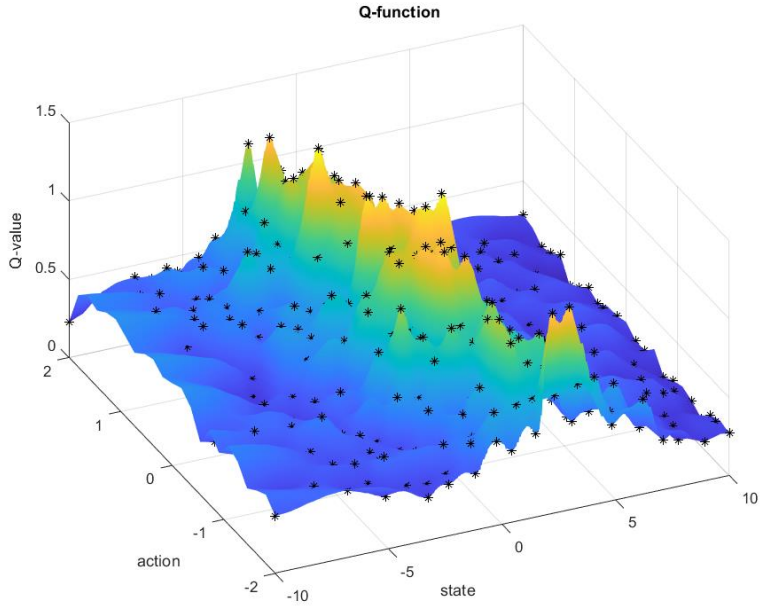
```
Input: the rule-base, dR parameters
Output: the reduced rule-base

1. initialize the distance matrix (D)
2. compute the rule distances (d) between each rule
3. compute the distance thresholds (dtr) each antecedent and the consequent uni-
   verses based on the dR input parameters
4. examine the distances (d) of each rule pair in the rule-base
5. If ∃ t,p where ∀ (d_j(t,p) < dtr_j) and (d_Q(t,p) < dtrQ) ,j ∈ [1,n+1]
      a. true: the r_t and the r_p rule pair can be merged into one rule
                  i. add the merged rule to the reduced rule-base
6. return the reduced rule-base
```

## 5. Futási eredmények

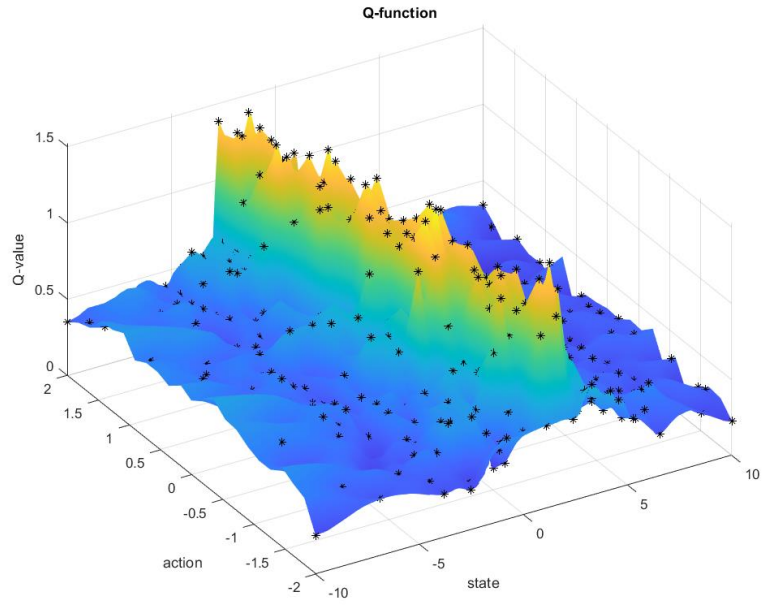
Alapvetően 3 futási eset által vizsgáltuk a bemutatott módszer hatékonyságát. Az 1. eset a szabálybázis-redukálás nélküli, a 2. eset mikor a szabálybázis-redukálás (szabályegyesítés) a csak az antecedens dimenzióra alkalmazott szabálytávolságokkal és távolságküszöbök által történik, a 3. eset pedig mikor a szabálytávolságok és távolságküszöbök a konzekvens univerzumra is alkalmazásra kerültek.

Az összehasonlítás alapja a 2. ábra. *Meredek töréspontot tartalmazó Q-függvény* látható, szabálybázis-redukálás nélküli, 1 állapot- és 1 akcióváltozóval rendelkező („referencia") Q-függvény, ahol a szabályok között megengedett távolság az univerzumok hosszának a 100-ad része. A futás során kapott szabálybázis ebben az esetben 530 darab szabályt tartalmaz. A 2. futási esetben a futás a szabálybázis-redukálási módszer alkalmazásával történik, de a szabálytávolságok csak az antecedens dimenzió által meghatározottak. Ebben az esetben, ha két (vagy több) szabály az antcedens univerzumokban meghatározott távolságküszöbök által közelinek számít, akkor azok összevonásra kerülnek egyetlen szabállyá. Ez abban az esetekben okozhat problémát, ha a két forrásszabály konzekvensében (Q-értékében) nagy eltérés van, azaz a konzekvens univerzumban távolinak tekinthetők, de mégis egyesítésre kerülnek egyetlen szabállyá. Ekkor a forrásszabályok egyesítése során, az antecedens és konzekvens értékeiknek átlagolása következtében, az általuk leírt szabálypontban a Q-függvény alakja rossz irányban módosulhat, mert a szabályok Q-értékében nagy eltérés volt. Ez által elromolhat az általuk leírt Q-függvény alakja, hamis információt adva a rendszer szabálybázisába. Azonban, ha ebben az esetben nem kerül összevonásra a két forrásszabály, mert a konzekvensükben távolinak tekinthetők, akkor ez kiküszöböli az említett problémás esetet.

A következő ábrák ezen a futási eseteket szemléltetik. A 4. ábrán a szabályegyesítés csak az antecedens univerzumban történő szabálytávolságok és távolságküszöbök alkalmazásával, az 5. ábrán pedig ezek a konzekvens univerzumra történő kiterjesztésével. Mindkét futási esetben a szakértő által megadott $dR$ paraméterek értéke 50 volt.




***4. ábra.*** *A 2. futási esetben keletkezett Q-függvény*




***5. ábra.*** *A 3. futási esetben keletkezett Q-függvény*

Az egyes futási esetek eredményeit a következő táblázat foglalja össze:

***1. táblázat.*** *Futási eredmények*

| # | Futás eset | Szabályszám |
|---|---|---|
| 1. | szabálybázis-redukálás nélkül | 530 |
| 2. | szabálybázis-redukálás csak az antecedens univerzumra alkalmazott szabálytávolságokkal és távolságküszöbökkel | 334 |
| 3. | szabálybázis-redukálás a szabálytávolságok és távolságküszöbök konzekvens univerzumra történő kiterjesztésével | 403 |

A 2. futási esetben kevesebb szabályt tartalmaz a szabálybázis, de a 4. ábrán látható, hogy a függvény gerince nem alakult ki teljesen (a 2. ábrához viszonyítva), tartalmaz több alacsonyabb csúcsot, mert összevonásra kerültek olyan szabályok is melyek Q-értékében nagy volt az eltérés és az átlagolás következtében ezen csúcsok alacsonyabbak lettek. Ebben az esetben a kapott összejutalom értéke is kisebb. A 3. futási esetben a szabálybázis több szabályt tartalmaz, mint a 2. futási esetben, de az 5. ábrán látható, hogy a függvény gerince kialakult, a távolságok és távolságküszöbök konzekvens univerzumra történő kiterjesztése miatt nem kerültek olyan szabályok összevonásra, melyek Q-értékében nagy az eltérés, így ez által nem romlott el a Q-függvény formája.

## 6. Összefoglalás

Bemutatásra került egy olyan módszer továbbfejlesztése, amely által a szakértői tudásbázissal bővített FRIQ-learning rendszerben az egymáshoz közel kerülő fuzzy szabályok összevonásával a fuzzy szabálybázis mérete már a tanulási fázis közben csökkenthető. Az egymáshoz közeli szabályok egyesítésének (illetve új szabály felvételének) az alapja a szabályok között lévő távolság, illetve távolságküszöbök meghatározása. A szabálytávolságok meghatározása az antecedens (állapot-akció) univerzumokban történik, a távoságküszöbök értékek pedig az univerzumok hosszának a szakértő által definiált valamekkora része.

A távolságküszöbök konzekvens univerzumra történő kiterjesztésének következtében a szabálybázis redukálása során nem kerülnek olyan szabályok összevonásra, melyek az antecedens univerzumban közelinek, de a konzekvens dimenzióban azonban távolinak tekinthetők. Ha a távolságküszöbök csak az antecedens univerzumokra lennének meghatározva, akkor összevonásra kerülnének olyan szabályok, melyek az antecedens dimenzióban közelinek számítanak, de a konzekvensükben (Q-értékükben) nagy eltérés található, ami helytelen információt vinne a fuzzy szabályrendszer által leírt Q-függvénybe.

A bemutatásra került szabályegyesítés módszerének továbbfejlesztése által a szabálytávolságok, illetve távolságküszöbök meghatározásakor a konzekvens (Q-érték) univerzum is figyelembevételre kerül, amely által csak a ténylegesen egymáshoz (antecedens és konzekvens univerzumban is) közel lévő szabályok kerülnek egyesítésre.

A távolságküszöb értékek jelenleg a teljes tanulási fázis során állandók, értékük az univerzumok hosszának a szakértő által meghatározott $dR$ része. A jövőben célszerű lehet a módszer további finomítása olyan módon, hogy a $dR$ paraméterek és így a távolságküszöb értékek a tanulási folyamat során történő hangolása, optimalizálása.

## Irodalom

[1] Berenji, H. R.: *Fuzzy Q-learning for generalization of reinforcement learning.* 1996 Proceedings of IEEE 5th International Fuzzy Systems, vol. 3, pp. 2208–2214. **https://doi.org/10.1109/FUZZY.1996.553542**

[2] Kóczy, L. T., Hirota, K.: *Size reduction by interpolation in fuzzy rule bases*. 1997 IEEE Transactions on Systems, Man, and Cybernetics, vol. 27, pp. 14–25. **https://doi.org/10.1109/3477.552182**

[3] 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. 1997 Proceedings of the *7th International Fuzzy Systems Association World Congress*, Prague, Czech Republic, pp. 144–149.

[4] Kovács, Sz., Kóczy, L. T. (1997). 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, 12, pp. 169–181.

[5] Kovács, Sz., Vincze, D., Gácsi, M., Miklósi, Á., Korondi, P.: Ethologically inspired robot behavior implementation. 2011 Proceedings of the *4th International Conference on Human System Interaction* (HSI 2011), Keio University, Yokohama, Japan, pp. 64–69. **https://doi.org/10.1109/HSI.2011.5937344**

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

[7] Kovács, Sz.: New aspects of interpolative reasoning. 1996 Proceedings of the *6th International Conference* on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Granada, Spain, pp. 477–482.

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

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

[10] Tompa, T., Kovács, Sz. (2020). Expert heuristic tuning design for the FRIQ-learning. *Multidiszciplináris Tudományok*, 10 (4), 119–125. **https://doi.org/10.35925/j.multi.2020.4.15**

[11] Tompa, T., Kovács, Sz. (2019). Szakértői heurisztika alkalmazása a FRIQ-learning megerősítéses tanulási módszerben. *Multidiszciplináris Tudományok*, 9 (4), pp. 356–368. **https://doi.org/10.35925/j.multi.2019.4.35**

[12] Tompa, T., Kovács, Sz. (2020). Applying Expert Heuristic as an a Priori Knowledge for FRIQ-Learning. *Acta Polytechnica Hungarica*, 17 (4). **https://doi.org/10.12700/APH.17.4.2020.4.2**

[13] 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. **https://doi.org/10.1109/SAMI.2017.7880302**

[14] 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. **https://doi.org/10.1109/CarpathianCC.2018.8399677**

[15] Vincze, D., 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, pp. 79–84. **https://doi.org/10.1109/ICCCYB.2008.4721383**

[16] Vincze, D., Kovacs, Sz.: Performance issues of the implemented FRI 'FIVE'. Proc. 2010 *11th International Symposium on Computational Intelligence and Informatics (CINTI)*. IEEE, pp. 131–136. **https://doi.org/10.1109/CINTI.2010.5672259**

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

[18] Vincze, D., Kovács, Sz. (2010). Incremental rule base creation with fuzzy rule interpolation-based Q-learning. In I. J. Rudas et al. (Eds.), *Computational Intelligence in Engineering, Studies in Computational Intelligence* (volume 313/2010, pp. 191-203). Springer-Verlag, Berlin Heilderberg. **https://doi.org/10.1007/978-3-642-15220-7_16**

[19] Vincze, D., Kovács, Sz.: Reduced rule base in fuzzy rule interpolation-based Q-learning. 2009 Proceedings of the *10th International Symposium of Hungarian Researchers on Computational Intelligence and Informatics*, CINTI 2009, Budapest Tech, Budapest, pp. 533–544. **https://doi.org/10.1109/SACI.2009.5136311**

[20] Vincze, D., Tóth A., Niitsuma, M.: Antecedent redundancy exploitation in fuzzy rule interpolation-based reinforcement learning. 2020 IEEE/ASME *International Conference on Advanced Intelligent Mechatronics (AIM)*. IEEE. **https://doi.org/10.1109/AIM43001.2020.9158875**

[21] Watkins, C. J. C. H. (1989). *Learning from delayed rewards*. Ph.D. thesis, Cambridge University, Cambridge, England.

[22] Kaelbling, L. P., Littman, M. L., Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, 4, pp. 237–285. **https://doi.org/10.1613/jair.301**

[23] Johanyák, Zs. Cs., Kovács, Sz. (2006). A brief survey and comparison on various interpolation based fuzzy reasoning methods. *Acta Polytechnica Hungarica*, 3 (1), pp. 91–105.

[24] Tompa, T., Kovács, Sz. (2021). Tudásbázis redukció a szakértői szabályrendszerrel bővített FRIQ-learning módszerben. *Multidiszciplináris Tudományok*, 11 (4), pp. 70–80. **https://doi.org/10.35925/j.multi.2021.4.8**

[25] Glorennec, P. Y., Jouffe, L. (1997). Fuzzy Q-learning. In Proceedings of *6th International Fuzzy Systems Conference*, Barcelona, Spain. **https://doi.org/10.1109/FUZZY.1997.622790**