# Szakértői tudásbázis validálás a HFRIQ-learning rendszerben

# Expert knowledge base validation in the HFRIQ-learning

***Dr. TOMPA Tamás***
*Miskolci Egyetem, Gépészmérnöki és Informatikai kar*
*Általános Informatikai Intézeti Tanszék*
*3515, Magyarország, Miskolc-Egyetemváros, Telefon: +36-46-565-333*

## Abstract

*The HFRIQ-learning is a reinforcement learning method that starts with a non-empty knowledge defined by expert heuristics. The expert defined a priori rules can be imprecise or inconsistent, but the system automatically optimizes them during learning. The final knowledge base after the learning can differ from the initial expert-defined one, therefore validating the initial expert heuristics is crucial the applicability of the learned model. The paper introduces a method for comparing the expert knowledge base before and after learning, thereby enabling the evaluation of the correctness of the initial expert heuristics through the proposed validation approach.*

**Keywords:** reinforcement learning, expert knowledge injection, knowledge base validation, fuzzy knowledge, Q-learning, fuzzy rule interpolation

## Kivonat

*A HFRIQ-learning egy megerősítéses tanulási módszer, amely tudásbázisa a tanulási folyamat kezdetén nem üres, hanem emberi szakértő által megadott fuzzy szabályokat tartalmaz. A módszer képes a pontatlan vagy nem teljesen konzisztens szakértői szabályokat automatikusan finomítani, bővíteni és optimalizálni. A tanulási folyamat végére kialakult tudásbázis eltérhet az eredetileg megadottól, ezért kulcsfontosságú a kezdeti szakértői heurisztika validálása. A cikk célja egy olyan módszer bemutatása, amely által a kezdeti szakértői heurisztika és a tanulási folyamat végeztével előállt tudásbázis összehasonlítható, validálható, azaz annak helyességére vonatkozó következtetések levonhatók.*

**Kulcsszavak:** megerősítéses tanulás, szakértői tudásbázis injektálás, tudásbázis validálás, fuzzy tudásbázis, Q-tanulás, fuzzy szabályinterpoláció

## 1. BEVEZETÉS

A mesterséges intelligencia térnyerésével egyre nagyobb figyelmet kapnak a gépi tanulási módszerek, közöttük a megerősítéses tanulás (Reinforcement Learning - RL) is [9]. Ezek az algoritmusok a környezettől érkező megerősítési információk (jutalmak, büntetések) alapján keresik a megoldáshoz vezető optimális stratégiát. Az ágens a tapasztalatai által, azaz az egyes akciók végrehajtása során kapott jutalmak és büntetések alapján tanul. A Q-learning [18] és a Fuzzy Q-learning [1] elterjedt algoritmusai ezen RL megközelítésnek, melyek üres tudásbázissal indítják a tanulási folyamatot. A Q-learning Q-táblát (többdimenziós mátrix), míg a Fuzzy Q-learning fuzzy szabálybázist alkalmaz a tudásbázis leírására, amelyet iteratív módon épít (és frissít) a tanulási fázis során a környezeti visszacsatolások alapján.

A tanítóminták szükségessége, az emberi tudás integrálása egyre fontosabb szerepet kap a gépi tanulási rendszerekben és így a fuzzy logikán alapuló megközelítésekben is. A [8] tanulmány részletesen bemutatja, hogyan épülnek fel fuzzy rendszerek és hogy milyen szerepet játszik a szakértői szabályok minősége a rendszer teljesítményére vonatkozóan. A [4] tanulmány egy olyan megerősítéses tanulási keretrendszert (RL-KS) mutat be, amely lehetővé teszi a korábban tanult feladatokból származó tudás transzferjét új tanulási folyamatokba. A [7] hivatkozás a KEEL nevű módszert mutatja be, amely fuzzy tudás és megfigyelési adatok együttes felhasználásával végez ok-okozati felfedezést. A tanulmány kiemeli, hogy a fuzzy tudás integrálása javítja a kauzális felfedezés robusztusságát és pontosságát, kis mintaszámú és nagy dimenziójú adatok esetében is. A Hierarchical Reasoning Model (HRM) [17] nevű agyi inspirációjú architektúra, amely két szinten (absztrakt tervezés majd részletes számítás) végez következtetést. A modell előzetesen megadott tanítómintákból (~1000

minta), előtanítás nélkül képes olyan komplex problémák megoldására, mint például a Sudoku, a labirintusnavigáció és az ARC-AGI [2][3] feladatok.

A heurisztikusan gyorsított Fuzzy szabály-interpoláció alapú Q-tanulás (Heuristically Accelerated Fuzzy Rule-Interpolation based Q-learning - HFRIQ-learning) [12] egy olyan Q-tanulás módszer, amely által a tanulási folyamatba külső szakértői tudás injektálható, majd ez a kezdeti tudásbázis a tanulási folyamat során hangolható, optimalizálható. Abban az esetben, ha rendelkezésre áll részleges információ a megoldás mikéntjére vonatkozóan, akkor ennek a megerősítéses tanuló rendszerekbe történő beépítésével felgyorsítható a teljes tanulási folyamat. A tanulási folyamat végeztével előállt tudásbázisban (fuzzy szabálybázisban) megtalálhatók a már behangolt (optimalizált) kezdeti szakértői szabályok.

Jelen cikk célja egy olyan módszer bemutatása, amely által a kezdetben megadott szakértői tudásbázis és a tanulási folyamat végeztével előállt tudásbázis összehasonlítható és validálható, azaz a kezdeti szakértői heurisztika helyességére utaló következtetések levonhatók.

## 2. A HFRIQ-LEARNING MÓDSZER

A heurisztikusan gyorsított Fuzzy szabály-interpoláció alapú Q-tanulás (Heuristically Accelerated Fuzzy Rule-Interpolation based Q-learning - HFRIQ-learning) [18] egy olyan megerősítéses tanulási algoritmus, amely lehetőséget ad külső szakértő által megadott tudásbázis tanulási folyamatba történő injektálására és hangolására [10][13]. A HFRIQ-learning módszer a FRIQ-learning (Fuzzy Rule-Interpolation based Q-learning) [14][16] algoritmus kiterjesztése, amely az alkalmazott „FIVE" (Fuzzy Rule Interpolation based on Vague Environment) [5][6] fuzzy szabály-interpolációs eljárás következtében folytonos állapot- és akció térrel rendelkezik. A módszer tudásbázisát ha-akkor típusú fuzzy szabályok írják le. Egy $r_i$ $(i \in [1,m])$ szabály formátuma az $m$ méretű $R$ szabálybázisban a következő [14][16]:

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

ahol $S_j^i$ az $i$-edik $(i \in [1,m])$ szabály $j$-edik $(j \in [1,n])$ állapot dimenziójának fuzzy halmaza az $n$-dimenziós $\boldsymbol{S}$ állapottérben, $\boldsymbol{s} \in \boldsymbol{S}$ az $n$-dimenziós állapot megfigyelés, $s_j$ a $j$-edik dimenziója az $\boldsymbol{s}$ állapot megfigyelésnek, $A^i$ az $i$-edik szabály egydimenziós akció univerzumának $(U)$ fuzzy halmaza, $a \in U$ az akció, $\tilde{Q}(\boldsymbol{s},a)$ a FIVE módszer [5][6] által becsült Q-függvény, $q^i$ pedig az $i$-edik szabály konzekvense (Q-értéke).
A szakértői szabályok formátuma hasonló az (1) formulában lévő fuzzy szabályokéhoz, de ekkor az $\hat{r}$ szakértői szabályok antecedense az állapot, a konzekvense pedig az ebben az állapotban végrehajtandó akció [10]:

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

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

A szakértői szabályrendszer tanulási folyamatba történő injektálása megköveteli a szakértői szabályok átalakítását „állapot-akció-Q-érték" formátumra. Az új antecedens az állapot-akció, a konzekvens pedig egy kezdeti Q-érték lesz $(\tilde{Q}_{init})$, amelyet egy Q-érték becslési módszer határoz meg [10]. A tanulási folyamat egy olyan fuzzy szabálybázissal indul, amely tartalmazza az átalakított formátumú szakértői szabályokat és az állapottér sarokpontjaihoz rendelt, 0 konzekvens értékkel rendelkező sarokponti szabályokat. Az esetlegesen ellentmondó szabályokat (azonos antecedens de eltérő konzekvens) a módszer előzetesen kiszűri. A tanulási folyamat során új fuzzy szabályok jönnek létre, a már meglévők finomhangolódnak, illetve az egymáshoz közeli szabályok összevonásra kerülnek. Új szabály akkor jön létre, ha a Q-érték frissítése meghalad egy előre definiált ε küszöböt és nincs közeli szabály. Ha a frissítés mértéke elhanyagolható, a meglévő szabályok konzekvensei módosulnak az alábbi frissítési összefüggés szerint [14][16]:

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

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

ahol $\alpha \in [0,1]$ a tanulási ráta, $\gamma \in [0,1]$ a leszámítolási tényező, $q_i^{k+1}$ az $i$-edik szabály konklúziója a $(k+1)$-edik iterációban, $a$ az $\boldsymbol{s}$-ben végrehajtott akció. Az új megfigyelt állapot $\boldsymbol{s}'$, $g(\boldsymbol{s},a,\boldsymbol{s}')$ a megfigyelt jutalom az $\boldsymbol{s} \to \boldsymbol{s}'$ állapot átmenetre, $\tilde{Q}^k$ és $\tilde{Q}^{k+1}$ pedig a $k$-adik és a $(k+1)$-edik iteráció FIVE módszer [5][6] módszer által becsült Q-értéke.

Ha a Q-érték frissítése nem haladja meg az ε küszöböt és létezik az aktuális ($\boldsymbol{s}$, a) megfigyeléshez közeli szabály, akkor a rendszer gradiens alapú hangolást alkalmaz [12][13]. Ez a módszer a legközelebbi szabálypont antecedensét (állapot-akció pontját) és konzekvensét (Q-értékét) módosítja az alábbi összefüggések alapján [12][13]:

$$\boldsymbol{s}_{k+1} = \boldsymbol{s}_k - \left(2 * TDerror * \frac{\partial \tilde{Q}(\boldsymbol{s}, a)}{\partial \boldsymbol{s}}\right) * \alpha \tag{5}$$

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

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

ahol a $\boldsymbol{s}_{k+1}, a_{k+1}, q_{k+1}$ a gradiens alapú hangolás által meghatározott új állapot, akció és Q-érték, $\boldsymbol{s}_k, a_k, q_k$ a régi állapot, akció és Q-érték, $\alpha$ a gradiens-módszer tanulási rátája, $\frac{\partial \tilde{Q}(\boldsymbol{s},a)}{\partial \boldsymbol{s}}, \frac{\partial \tilde{Q}(\boldsymbol{s},a)}{\partial a}, \frac{\partial \tilde{Q}(\boldsymbol{s},a)}{\partial q}$ a Q-függvény állapot, akció és Q-érték szerinti parciális deriváltjai, a $TDerror$ (Temporal Difference error) pedig a Q-becslés hibája.

A gradiens módszer alapú szabályhangolás során előfordulhat, hogy egyes szabályok egymáshoz közel kerülnek a szabálypontok vándorlása következtében. Ebben az esetben ezeket a szabályokat a rendszer összevonja, ezáltal csökkentve szabálybázis méretét [12][13]. A tanulás után további opcionális szabálybázis-redukciós módszerek is alkalmazhatók, melyeket a [15] foglal össze. A tanulási folyamat akkor ér véget, amikor már nem jönnek létre új szabályok, a Q-értékek frissítése és szabálypontok változása is minimális és nem történik újabb szabályösszevonás [12].

# 3. TUDÁSBÁZIS VALIDÁLÁS

A tanulási folyamat során a rendszer az előzetesen definiált szakértői szabálybázist hangolja, kiegészíti új szabályokkal, illetve összevonja azokat, ha a szabálypontok hangolása következtében több szabály közel [11] kerülne egymáshoz. Ennek következtében a szabályok változásának nyomonkövetéséhez szükséges egy módszer kidolgozása, amely által a tanulási folyamat végén megállapítható a szabályok hangolásának mértéke.

A rendszerben 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 szabálypontok mozgása és azok esetleges egyesítése következtében minden egyes szabály egyedi azonosítása szükséges, hogy nyomonkövethetőek legyenek a tanulási fázis során. Ennek következtében mindegyik szabályt célszerű egy egyedi azonosítóval (*ID*-vel) ellátni. Az egyedi szabályazonosítók meghatározásának alapja szabálytípusonként az alábbi:

- sarokponti szabály ($r^{\square}$) azonosító: 1000-1999
- szakértői szabály ($\hat{r}$) azonosító: 2000-2999
- új, rendszer által felvett szabály ($r$) azonosító: 3000-3999

A szabálycímkéző algoritmus a tanulási folyamat előtt minden egyes szakértői és sarokponti szabályra véletlenszerűen meghatároz egy ID-t az adott tartományban. Ha a tanulás során a rendszer új szabályt hoz létre akkor minden egyes új szabálynak is sorsol egy ID-t a 3000-3999 közötti tartományban. Az ID generálása során ellenőrzi, hogy az éppen generált azonosító egyedi-e és ha nem (azaz már létező) akkor generál egy új ID-t és addig ismétli ezt a folyamatot amíg az adott generált azonosító egyedi nem lesz a teljes szabálybázisban.

A hangolási folyamat következtében a rendszer az esetlegesen egymáshoz közel kerülő szabályokat egy távolságküszöb [11] alapján egyesíti. Annak következtében, hogy a rendszerben 3 szabálytípus különböztethető meg így az összevont (redukált) szabály típusát is szükséges meghatározni, aminek alapja az egyes forrásszabályok típusa. Mivel a szakértő által meghatározott 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 és szabályazonosító 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, szabályazonosítója pedig a szakértői szabály azonosítója lesz. Ha a két szabály közül mindkét szakértői szabály volt, akkor az új összevont szabály is szakértői szabály lesz, szabályazonosítója pedig a 2 azonosító közül az lesz, amelyik a legkisebb ID értékű volt. Ha a két szabály közül mindkét szabály a rendszer által újonnan felvett szabály, akkor az új egyesített szabály is újonnan beszúrt szabályként lesz jelölve, azonosítója pedig szintén az lesz, amelyik a 2 azonosító közül a legkisebb értékű volt. 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, mert a sarokponti szabály antecedense nem változhat.

Egy, a szakértői által definiált szabályt jelöljünk $\hat{r}$-el $\left(\hat{r} \in R_{expert}\right)$, egy a rendszer által felvett szabályt $r$-el $(r \in R)$, egy sarokponti szabályt pedig $r^{\square}$-el $(r^{\square} \in R)$. A (8) összefüggés az adott típusú, $r_t$ és $r_p$ indexű szabályok egyesítését követően létrejött új $r_{red}$ szabály típusát és annak azonosítóját szemlélteti. A „⊔" operátor a két szabály egyesítését, a „→" operátor pedig a szabályegyesítés eredményét, azaz a szabályösszevonás során létrejött új $r_{red}$ (redukált) szabály típusát jelöli:

| $r_t$ | | $r_p$ | | $r_{red}$ | Szabályazonosító | Megjegyzés |
|---|---|---|---|---|---|---|
| $r$ | ⊔ | $\hat{r}$ | → | $\hat{r}$ | a szakértői szabályé | szakértői szabály prioritás |
| $\hat{r}$ | ⊔ | $\hat{r}$ | → | $\hat{r}$ | a kisebb ID értékű | mindkettő szakértői szabály |
| $r$ | ⊔ | $r$ | → | $r$ | a kisebb ID értékű | mindkettő új szabály |
| $r^{\square}$ | ⊔ | $r$ | → | $r^{\square}$ | a sarokponti szabályé | sarokponti szabály prioritás |
| $r^{\square}$ | ⊔ | $\hat{r}$ | → | $r^{\square}, \hat{r}$ | mindkettő megmarad | sarokponti szabály prioritás |

(8)

A szabályegyesítés során nem csak az új szabály típusa, hanem az új szabály antecedens és konzekvens értékeit is szükséges meghatározni. Ennek alapja a forrásszabályok antecedens és konzekvens értékei, tehát az új szabálypont a forrásszabályok átlaga lesz. Az 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:

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

| Szabály | $s_1$ állapot | $s_2$ állapot | $a$ akció | $Q$-érték | ID |
|---|---|---|---|---|---|
| $\hat{r}$ | 1 | 2 | 4 | 123.45 | 2001 |
| $r$ | 2 | 2 | 5 | 145.23 | 3001 |
| $\hat{r} \sqcup r \rightarrow \hat{r}$ | 1.5 | 2 | 4.5 | 134.34 | 2001 |

Az egyedi azonosítók és a redukált szabály létrehozási folyamat által a szabályok hangolása és egyesítése is nyomon követhető és a tanulási folyamat végén kinyerhető. Ezáltal utólagosan meghatározható, hogy a hangolási, optimalizálási folyamat során a szakértő által megadott szabályok milyen mértékben változtak.

# 4. FORDÍTOTT INGA MINTAPÉLDA

A javasolt tudásbázis validálási módszer működése a klasszikus „Cart-Pole" szimuláción keresztül kerül vizsgálatra, ahol az ágens (kisautó) célja az autó középén elhelyezkedő rúd függőleges pozícióban való tartása. A „Cart-Pole" probléma négy állapotváltozóval ($s_1$: ágens pozíciója, $s_2$: ágens sebessége, $s_3$: inga szöge, $s_4$: inga szögsebessége) és egy akcióváltozóval ($a$: irányított elmozdulás) írható le.

A szakértői tudásbázis 7 darab állapot-akció formátumú helytelen szabályt tartalmaz, amelyeket a következő táblázat tartalmaz:

*Az előzetesen megadott szakértői tudásbázis szabályai 2. táblázat*

| R# | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| $s_1$ | 1 | 1 | 1 | 1 | -1 | -1 | 1 |
| $s_2$ | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| $s_3$ | 0 | -0.0524 | 0 | -0.0524 | -0.2094 | -0.2094 | 0.2094 |
| $s_4$ | 1 | -1 | -1 | 1 | -1 | -1 | 1 |
| $a$ | -1 | 1 | -1 | -1 | 0.8 | 0.4 | 1 |

A tanulási folyamat során ezen 7 darab kezdeti szakértői szabály hangolása valósul meg, amelyet a rendszer új szabályokkal egészít ki és adott esetekben közeli szabályok összevonásával optimalizál. Annak következtében, hogy a megadott szakértői szabályok részben helytelenek, így a rendszer a tanulási folyamat során hangolja, módosítja azok állapot-akció pontját és Q-értékét. A tanulás után előállt, optimalizált szakértői tudásbázis szabályait a következő táblázat szemlélteti:

*A tanulási folyamat során behangolt szakértői tudásbázis szabályai 3. táblázat*

| R# | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| $s_1$ | 0.9340 | 1.0115 | 1.0095 | 0.9630 | -1 | -1 | |
| $s_2$ | -1.3334 | 0.5245 | 0.1166 | -0.3350 | 0 | 1 | |
| $s_3$ | 0.0169 | -0.0173 | -0.0156 | 0.0819 | -0.2094 | -0.2094 | x |
| $s_4$ | 1.520 | -0.7900 | -0.1966 | 0.6808 | -1 | -1 | |
| $a$ | -0.9475 | 0.9692 | -0.9108 | 0.975 | 0.8 | 0.4 | |

A 3. táblázat alapján megállapítható, hogy a hét hibásan megadott szakértői szabály közül négyet (1.–4.) a rendszer finomhangolt, kettőt (5.–6.) változatlanul hagyott, míg a hetedik szabályt egy másikhoz való közelsége miatt összevonta és törölte. A tanulási folyamat ebben az esetben 92 epizódot vett igénybe, amelynek eredményeként egy 95 szabályt tartalmazó végső szabálybázis alakult ki.

A bemutatott szakértői tudásbázis validálási módszer által az előzetesen definiált szakértői szabályok változása a tanulási folyamat során nyomonkövethető, majd a tanulási folyamat végeztével pedig visszaolvasható, következtetve ezáltal azok helyességének mértékére. A kapott hangolt szabálybázis alapján elmondható, hogy a szakértői szabályok helytelensége alátámasztható, hiszen a rendszer a tanulási folyamat során nagymértékben módosította, hangolta azokat. A szakértői szabályok módosulásának mértéke közvetlenül tükrözi, hogy a kezdeti heurisztika mennyire volt konzisztens és releváns az adott problématérben. A tanulás eredményeként létrejött hangolt szabálybázis alapján megállapítható, hogy a szakértői szabályok jelentős része nem volt pontos, hiszen a rendszer azokat nagymértékben módosította azokat, ami alátámasztja a validálási módszer gyakorlati hasznosságát.

# 5. ÖSSZEFOGLALÁS

A cikkben bemutatásra került egy olyan módszer amely által a HFRIQ-learning megerősítéses tanulási módszer bemeneteként megadott előzetes szakértői tudás helyessége megállapítható, azaz validálható. A tanulási folyamat során a rendszer a szakértői szabályokat automatikusan finomítja, bővíti és optimalizálja, így a tanulási folyamat végeztével előállt végső tudásbázis eltérhetett a kezdetben megadottól. A bemutatott módszerrel az eredeti és a hangolást követően kinyert szakértői szabálybázis összevethető, a kezdeti szakértői heurisztika helyessége ellenőrizhető. A hangolás előtt megadott és a hangolás utáni előállt szakértői tudásbázis összehasonlításával következtetni lehet a szakértői szabályok helyességének mértékére. A hangolást követő kismértékű eltérés igazolhatja a szakértői szabályok helyességét, nagyobb mértékű eltérés értelmezhető a kezdeti heurisztika pontosításaként, a jelentős eltérések, vagy az eredeti szabályrendszernek ellentmondó produkciós szabályok pedig utalhatnak a kezdeti heurisztika egyes részeinek helytelenségére. A szabálybázis redukciója során eltűnő szabályok a szakértői heurisztika redundanciájára utalhatnak.

További kutatási cél egy metrika kidolgozása, amely a validálási algoritmus alapján számszerűsíti a kezdeti szakértői szabálybázis (heurisztika) minőségét és robusztusságát a tanulási folyamat során bekövetkező változások által. Ez a metrika nem csupán a szabályok számosságát és módosulását vizsgálná, hanem a szabályok strukturális és paraméteres stabilitását is figyelembe venné, 0 és 1 közötti számmal jellemezve a kezdeti helyességet.

# IRODALMI HIVATKOZÁSOK


[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] Chollet, F., Knoop, M., Kamradt, G., Landers, B., & Pinkard, H. (2025). Arc-agi-2: A new challenge for frontier ai reasoning systems. *arXiv preprint arXiv:2505.11831*.

[3] Chollet, François. "On the measure of intelligence." arXiv preprint arXiv:1911.01547 (2019).

[4] Gao, X., Si, J., & Huang, H. (2023). Reinforcement learning control with knowledge shaping. *IEEE Transactions on Neural Networks and Learning Systems*, *35*(3), 3156-3167.

[5] 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.

[6] Kovács, Szilveszter. "Extending the fuzzy rule interpolation" five" by fuzzy observation." Computational Intelligence, Theory and Applications: International Conference 9th Fuzzy Days in Dortmund, Germany, Sept. 18–20, 2006 Proceedings. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006.

[7] Li, W., Zhang, W., Zhang, Q., Zhang, X., & Wang, X. (2024). Weakly-supervised causal discovery based on fuzzy knowledge and complex data complementarity. *IEEE Transactions on Fuzzy Systems*.

[8] Lilly, J. H. (2011). *Fuzzy control and identification*. John Wiley & Sons.

[9] Sutton, R. S., Barto, A. G.: Reinforcement Learning: An Introduction, MIT Press, Cambridge (1998)

[10] Tompa, Tamás, and Szilveszter Kovács. "Applying Expert Heuristic as an a Priori Knowledge for FRIQ-Learning." Acta Polytechnica Hungarica 17.4 (2020).

[11] Tompa, Tamás, and Szilveszter Kovács. "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.

[12] Tompa, Tamás, and Szilveszter Kovács. "Heuristically accelerated FRIQ-learning." 20th Jubilee International Symposium on Intelligent Systems and Informatics (SISY 2022). IEEE, 2022.

[13] Tompa, Tamás, and Szilveszter Kovács. "Knowledge Base Optimization of the HFRIQ-Learning." Acta Polytechnica Hungarica 21.10 (2024).

[14] Vincze, Dávid, and Szilveszter Kovács. "Fuzzy rule interpolation-based Q-learning." 2009 5th International Symposium on Applied Computational Intelligence and Informatics. IEEE, 2009

[15] 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.

[16] Vincze, Dávid.: "Fuzzy Rule Interpolation-based Q-learning." PhD dissertation, 2013.

[17] Wang, G., Li, J., Sun, Y., Chen, X., Liu, C., Wu, Y., ... & Yadkori, Y. A. (2025). Hierarchical Reasoning Model. *arXiv preprint arXiv:2506.21734*.

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