Intelligens rendszerek egységes tervezése

Kovács Dániel László
II. éves Ph.D. hallgató
Budapesti Műszaki és Gazdaságtudományi Egyetem
Villamosmérnöki és Informatikai Kar
Méréstechnika és Információs Rendszerek Tanszék
dkovacs@mit.bme.hu

Jelen cikk amellett érvel, hogy az ágens-, játék- és evolúciós-elméletek egyesítéséhez, avagy az – intelligens rendszerek egységes tervezését támogató – átfogó rendszer-specifikációs elv kidolgozásához a „racionalitás" fogalma nyújt elsősorban segítséget. Elsőként tehát áttekintjük a fontosabb racionalitás fogalmakat. Ezt követően rátérünk az ágens-elméletre. Ezen belül foglalkozunk az ágensek architektúrájával, környezetével, a megoldandó problémákkal, végül az ágensek racionalitásával, avagy a korlátosan optimális ágensekkel. Ezek után következik a játék-elmélet. Példát adunk egyszerűbb játékokra, majd ezek vizsgálatán keresztül bemutatjuk a fontosabb játék-elméleti racionalitás fogalmakat (pl. dominancia, Nash-egyensúly, Pareto-optimalitás). Rámutatunk a játék- és ágens-elmélet összekapcsolási lehetőségeire, majd bevezetjük a korlátosan optimális játékosok fogalmát. Végül összefoglaljuk a két elmélet összekapcsolásának következményeit. A játék-elméletet az evolúciós-elmélet tárgyalása követi. Az elmélet rövid áttekintése után rátérünk a genetikus algoritmusokra. Az alapfogalmak tisztázását követően rámutatunk az elmélet ágens- és játék-elmélettel való összekapcsolási lehetőségeire. Végül áttekintjük a három elmélet összekapcsolásának következményeit.

Kulcsszavak: ágens, evolúció, játék elmélet, racionalitás, tervkészítés

1. Bevezetés

Az intelligens rendszerek fontos szerepet játszanak mindennapjainkban, kezdve az orvosi diagnosztikától, a katonai tervezésen át egészen az intelligens háztartásokig. Céljuk többek közt az emberi munkavégzés segítése kezelhetetlen bonyolultságú feladatok esetén (pl. nagykomplexitású hálózatok menedzselése), esetleg az emberi munkaerő kiváltása nehezen hozzáférhető környezetekben (mélyűr, tengerfenék, katasztrófa helyszínek, stb). Az intelligens porszívótól, a marsi felderítő járműveken át egészen az Internet-en barangoló szoftver-ágensekig lassan már mindenhol megtalálhatók. Legfontosabb közös jellemzőjük az adaptív feladat-megoldó készség.

Az adaptív feladat-megoldó rendszerek tervezésekor a legfőbb nehézséget egy átfogó rendszer-specifikációs elv hiánya jelenti. Több elmélet is foglalkozik ilyen jellegű rendszerek tervezésével, ám mindegyik másképp közelít a problémához. A mesterséges intelligencia területén oly népszerű ágens-elmélet a rendszer „intelligenciájára", a játék-elmélet a rendszer „racionalitására", az evolúciós-elmélet pedig a rendszer „optimalitására" helyezi a hangsúlyt. Bár az említett fogalmak első hallásra igen különbözőnek tűnhetnek, mégis közös nevezőre hozhatók, lehetőséget adva az elméletek egyesítésére, egy átfogó rendszer-specifikációs elv kidolgozására.

Egy ilyen elv előnye, hogy lehetővé teszi az intelligens rendszerek tervezésekor felmerülő különböző szempontok egységes, ám szemponthoz illeszkedő tárgyalását. Például a rendszer benső felépítésének kialakításához az ágens-elméletet; a rendszer csoporton belüli működésének kialakításához a játék-elméletet; míg rendszerek csoportos viselkedésének kialakításához az evolúciós-elméletet hívhatnánk segítségül, biztosítva a különböző megközelítések közti átjárhatóságot. Ily módon a megtervezett rendszer viselkedése, kommunikációs és együttműködési készsége, hasznossága, robusztussága, vagy más egyéb jellemzői egységes elvek alapján, akár többféle szemszögből is tervezhetők volnának.

2. A racionalitásról általában

Az „intelligens" jelző sok félreértésre ad okot, hiszen rengeteg értelmezést kínál, sőt, talán még igazán pontos definíciója sincs. Amennyiben nem emberekre, hanem mesterséges rendszerekre alkalmazzuk, úgy általában a rendszer önálló, adaptív feladat-felismerő és megoldó készségét, hatékony erőforrás kihasználását és célratörő működését értjük alatta. Tehát ilyen értelemben egy rendszert akkor nevezünk „intelligensnek", ha ésszerű (racionális) a működése; ha emberi intelligenciát igénylő feladatok megoldására is alkalmazható. Az „intelligens" jelző tehát valójában a rendszerek racionalitását takarja.

A filozófusok és a közgazdászok már jóval a mesterséges intelligencia területének létrejötte előtt elkezdték keresni a racionális viselkedés kielégítő meghatározását. Erre azért volt szükség, mert a racionalitás fogalma az etika, az induktív tanulás, a következtetés, a döntés-hozás és a közgazdasági modellek meghatározó tényezője. Több különböző racionalitás-fogalom is napvilágot látott:

- **Tökéletes racionalitás**: a filozófia és a közgazdaságtan klasszikus racionalitás-fogalma. Arisztotelész etikájából fejlődött ki, míg végül a döntés-elméletben [1] nyert formális megfogalmazást. Azt a rendszert nevezzük tökéletesen racionálisnak, mely minden pillanatban olyan döntést hoz tapasztalatai alapján, amivel maximálja várható nyereségét. Mivel a megfelelő döntés meghozatala számításigényes feladat, és ehhez időre van szükség, ezért nem-triviális esetekben nemigen beszélhetünk tökéletesen racionális rendszerekről. A tökéletes racionalitás tehát a gyakorlatban *nem megvalósítható* követelmény.

- **Kalkulatív racionalitás**: a mesterséges intelligencia által vizsgált racionalitás-fogalom. Egy kalkulatíve racionális rendszer olyan döntést hoz, amely számításai megkezdésekor tökéletesen racionális lett volna. Sajnos a gyakorlatban ez a követelmény *sem használható*, mivel egyrészt akár kivárhatatlan sokáig is eltarthat, amíg a rendszer meghozza döntését, másrészt (különösen dinamikus, folytonosan változó környezetekben) a hozott döntés jóságát már semmi sem garantálja. Ezért is fordul elő a gyakorlatban, hogy az intelligens rendszerek tervezői elrugaszkodnak az elvektől, és ad-hoc módon alakítják ki rendszereiket.

- **Metaszintű racionalitás**: a kalkulatív racionalitás által felvetett problémákra adott válasz. Egy metaszinten racionális rendszer a döntések meghozásáért

felelős objektum-szintű számítások felett optimalizál [2]. Tehát kiválasztja az optimális „számítás+döntés” kombinációt, ahol a döntés a számítás eredménye. Sajnos azonban a teljes metaszintű racionalitás igen ritka, hiszen egyrészt már maguk a metaszintű számítások is időt igényelnek, másrészt a metaszintű döntési probléma gyakorta nehezebb, mint az eredeti, objektum-szintű probléma. Ennek ellenére a gyakorlatban hasznosnak bizonyult a metaszintű racionalitás néhány egyszerűbb közelítése. Mindazonáltal ez is csak azt mutatja, hogy az intelligens rendszerek egységes tervezésének mindmáig nincsen elfogadott, átfogó rendszer-specifikációs elve. A gyakorlatban tehát a metaszintű racionalitás *sem elvárható* követelmény.

- **Korlátos racionalitás:** a közgazdaságtan válasza a tökéletes racionalitás problémáira. Amíg tehát a tökéletes racionalitás a rendszer (pl. gazdasági entitás) erőforrásaitól, döntéshozási képességeitől és a döntéshozásra rendelkezésre álló idejétől függetlenül definiálta a racionalitást, addig a korlátosan racionális [3] rendszerek e tényezők, pontosabban döntéshozási korlátaik figyelembevételével maximálják várható nyereségük**.** Magyarán lehetőségeikhez mérten a lehető legjobb eredményt produkálják. A korlátos racionalitás tehát a racionalitás egy *használható* definíciója.

Tehát átértékelődik a racionalitás tárgya. Amíg a tökéletes racionalitás még szigorúan csak a döntések racionalitására fókuszál, addig a metaszintű, illetve a korlátos racionalitás már a rendszer működésére, a döntések meghozásáért felelős mechanizmusokra (is) vonatkozik. Észrevehető a racionalitás definíciójának abszolúttól a relatív irányában történő elmozdulása. Amíg tehát a tökéletes racionalitás gyakorlatilag figyelmen kívül hagyja a rendszer képességeit, addig a kalkulatív racionalitás már közvetve ezekről tesz megállapítást; a metaszintű racionalitás már a rendszer funkcionális felépítését is taglalja; míg végül a korlátos racionalitás már a rendszer egészének viszonylatában definiálja a racionalitást.

Ennek a szemléletbeli változásnak több oka is van. Elmozdulás történt a közgazdaságtanban a tökéletestől a korlátos racionalitás felé; a játék-elméletben a cselekvések kiválasztásától a – cselekvéseket előállító – programok kiválasztása felé; az evolúciós-elméletben a megoldások evolválásától a megoldásokat előállító programok evolválása [4] felé; a filozófiában pedig a cselekedet-utilitarizmus (vagy tett-utilitarizmus) felől a szabály-utilitarizmus, avagy a cselekvések meghozásáért felelős általános mechanizmusok racionalitása felé.

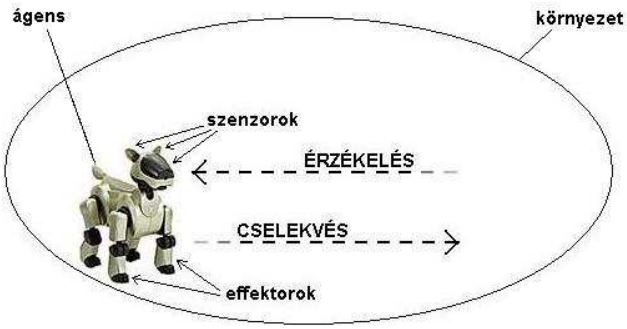
Beláthatjuk, hogy e folyamat gyökerei igen egyszerűen visszavezethetők a racionális rendszerek tervezhetőségére, hiszen így már nem támasztunk velük szemben olyan kritériumokat (pl. nem várunk el olyan döntéseket), melyeket a rendszer egyetlen döntéshozási mechanizmusa se tud teljesíteni. Végtére is a rendszer által véghezvitt döntésekről és számításokról a rendszer tervezője csak közvetve gondoskodik, viszont az őket előállító döntéshozási mechanizmusokról (programokról) már közvetlenül. Ezek a meggondolások vezettek végül a mesterséges intelligencia legújabb racionalitás-fogalmához – a *korlátos optimalitás*hoz [5]. A következőkben ezzel foglalkozunk.

3. Ágens-elméleti racionalitás

A mesterséges intelligenciában az előbb tárgyalt rendszereket (melyek maguk képesek eldönteni, hogy mit kell tenniük ahhoz, hogy teljesítsék a rájuk szabott feladatot) *ágens*nek nevezzük. Az „ágens” jelentését illetően azonban mégsem alakult ki általános konszenzus. Szakterületről szakterületre változik az ágensekkel kapcsolatos jellemzők fontossága. Bizonyos alkalmazásokban például elengedhetetlen, hogy az ágens képes legyen *tanulni*, mígnem másokban szükségtelen. Talán csak az *autonómia*, mint központi megfontolás, tekinthető az ágensek egyetlen, általánosan elfogadott, közös jellemzőjének. Mivel azonban valamiféle definícióra mindenképp szükségünk lesz a későbbiekben, kezdetnek két neves szerző – egymásnak összhangban lévő – javaslatát vesszük alapul.

> *„Egy ágens bármi lehet, amit úgy tekinthetünk, mint ami az érzékelői segítségével érzékeli a környezetét, és beavatkozó szervei segítségével megváltoztatja azt.”* [6]

> *„Ágensnek nevezzük azt a rendszert, mely adott környezetbe helyezve képes autonóm cselekvésre céljai elérése érdekében.”* [7]



1. Ábra: Környezetébe ágyazott ágens

Az 1. ábra alapján tehát az ágensek tekinthetők olyan függvényeknek, melyek érzetek sorozatából képeznek cselekvésekbe. Ennek formális definíciója lehet a következő:

Definíció 1. *Ágens-függvény*:

$$f_{Ágens} : Érzet_történetek \rightarrow Cselekvések$$

Az ágens-függvény specifikálja tehát, hogy egy-egy érzettörténet (azaz érzetek egy-egy sorozatának) hatására mit cselekszik az ágens. Ily módon lehetőség nyílik a *tökéletesen* és a *kalkulatíve racionális ágensek* közti különbségtételre.

Az ágenseket beágyazó környezetet (amely akár más ágenseket is magába foglalhat), röviden *ágens-környezet*nek nevezzük. Az ágens-környezetek többféle osztályozása lehetséges. Jelen cikk szemszögéből a következő szempontok érdekesek:

- **Hozzáférhető, vagy nem hozzáférhető**: A környezet hozzáférhető, ha az ágens érzékelői segítségével hozzáférhet a környezet teljes állapotához. Egy környezet ténylegesen hozzáférhető, ha az ágens érzékelése lefedi a környezet mindazon aspektusait, melyek szükségesek lehetnek egy-egy cselekvés

kiválasztásához. Minél inkább hozzáférhető egy környezet, annál egyszerűbb felépítésű ágensek lesznek elegendőek hozzá. Nem hozzáférhető környezetre lehet példa a fizikai valóság, vagy akár az Internet, stb.

- **Determinisztikus, vagy nem determinisztikus**: A környezet determinisztikus, ha tetszőleges állapotában az ágens által végrehajtott tetszőleges cselekvés egyértelműen meghatározza a cselekvés nyomán előálló következő környezeti állapotot. Egy determinisztikus, ámde nem hozzáférhető környezetet az ágens akár nem-determinisztikusnak is érzékelhet, így gyakran jobb a környezetet a determinizmus szemszögéből vizsgálni.

- **Epizódszerű, vagy nem epizódszerű***:* A környezet epizódszerű, ha benne az ágens tapasztalata „epizódokra" bontható. Ilyen környezetben az ágens haszna a diszkrét számú „epizódon" vett hasznától függ, ahol az egyes „epizódokon" vett hasznok függetlenek egymástól. Az epizódszerű környezetek egyszerűbbek, mint a nem epizódszerűek, mivel az ágensnek csak az adott „epizódon" belül kell előre tekintenie, hiszen az egyes „epizódokban" végrehajtott cselekvések nem befolyásolják a következő „epizód" alakulását.

- **Statikus, vagy dinamikus***:* A környezet statikus, ha csak az ágens cselekvéseinek nyomán módosul az állapota, ám egyébként változatlan. A dinamikus környezetet az ágens hatáskörén kívül eső jelenségek, folyamatok, vagy más egyéb történések is befolyásolhatják, változtathatják. Szemi-dinamikus környezetről beszélünk, ha az idő előrehaladtával a környezet nem változik, de az ágens által elért haszon mértéke igen.

- **Diszkrét, vagy folytonos***:* A környezet diszkrét, ha az ágens különböző lehetséges cselekvéseinek és észleléseinek halmaza véges, egyébként folytonos.

Most pedig térjünk rá az ágens-környezet formális megfogalmazására. Először is tekintsük a környezetet olyannak, amely minden pillanatban valamilyen állapotban van. A környezet lehetséges állapotainak halmaza legyen véges. Ahhoz, hogy kezelni tudjuk azt az esetet, amikor a környezet nem teljesen hozzáférhető az ágens számára, vezessük be az $f_{Érzékelés}$ *függvényt*. Ekkor tehát az ágens egy adott pillanatban – korlátos érzékelőinek köszönhetően – nem a teljes környezeti állapothoz, hanem annak csak egy $f_{Érzékelés}\left(Állapot\right)$ leképződéséhez fér hozzá. Továbbá, az ágens cselekedeteinek környezetre gyakorolt hatását modellezzük egy $f_{Átmenet}$ *függvénnyel*, amely az ágens adott cselekvése, és a környezet aktuális állapota alapján megszabja a környezet következő állapotát. Összefoglalva, az ágens környezetét a következőképp definiálhatjuk:

Definíció 2. *Az ágens környezete*

$$Környezet = \left(Állapotok, f_{Átmenet}, f_{Érzékelés}\right)$$

Látható, hogy az ágens-függvény és a környezet egyértelműen meghatározza a környezet állapotainak – ágens által kialakított – sorozatát: az állapot-történetet.

Jelölje $Hatás\left(f_{Ágens}, Környezet\right)$ azt az állapot-történetet, amit az $f_{Ágens}$ függvény generál az *Környezet*-ben. Vegyük észre, hogy a fentiek alapján a környezetet diszkrétnek és determinisztikusnak definiáltuk. A környezet folytonos és nem-determinisztikus kiterjesztése ennél valamivel komplikáltabb.

Most pedig térjünk rá az ágensek felépítésére. Tekintsük az ágenst egy olyan, benső állapottal rendelkező rendszernek, mely egy architektúra és egy program együttese. Az architektúra felelős a környezet és a program közti kapcsolat biztosításáért, továbbá a program futtatásáért. Minden egyes *Architektúrá*-val kapcsolatban definiáljunk egy $Nyelv_{Architektúra}$ *véges programozási nyelvet*, mely gyakorlatilag az architektúra által futtatható összes $Program \in Nyelv_{Architektúra}$ *ágens-program* halmaza. Az ágens-programok az ágens aktuális *Benső_Állapot*-a és észlelése alapján állítják elő az ágens cselekvéseit, illetve újabb benső állapotát:

Definíció 3. *Az ágens architektúrája egy fix interpreter (értelmező-program) az ágens-programok számára, amely futtatja a programot, frissíti az ágens benső állapotát és generálja cselekvéseit:*

$$Architektúra\left(Program, Benső_Állapot, Észlelés\right) = \left(Benső_Állapot, Cselekvés\right)$$

Most pedig már összefüggésbe hozhatjuk az ágens-programokat és az ágens-függvényeket azáltal, hogy definiáljuk az adott architektúrán futatott ágens-program által *implementált* $Ágens\left(Program, Architektúra\right)$ ágens-függvényt. Ez éppen azon $f_{Ágens}$ függvény, melyre teljesül, hogy tetszőleges környezet esetén:

$$Hatás\left(f_{Ágens}, Környezet\right) = Hatás\left(Ágens\left(Program, Architektúra\right), Környezet\right)$$

Látható, hogy amíg minden architektúrának és programnak megfeleltethető egy-egy ágens-függvény, addig – adott architektúra esetén – már nem minden ágens-függvény implementálható feltétlen. Érdemes tehát bevezetnünk az adott architektúrán *megvalósítható* függvények halmazát:

$$Megvalósítható\left(Architektúra\right) = \left\{ f \middle| \exists Program \in Nyelv_{Architektúra}, f = Ágens\left(Program, Architektúra\right) \right\}$$

Most már tehát rendelkezésünkre áll az ágensek és környezetük teljes definíciója, ám az ágensek által megoldandó problémáról még nem ejtettünk szót. Ez azért szükséges, mert csakis a megoldandó probléma fényében áll módunkban nyilatkozni az ágensek racionalitásáról.

Problémának nevezzük az ágens-környezet *kiinduló állapotának*, *cél-állapotainak* és lehetséges *cselekvéseinek* együttesét. Ekkor a *probléma megoldása* a cselekvések egy olyan sorozata, amelyek végrehajtása a kiinduló állapotból a cél-állapotok valamelyikébe vezet. Az aktuális cselekvés kiválasztása ezért legtöbbször bizonyos szintű „előrelátást” kíván az ágens részéről, aminek során különböző cselekvés-sorozatokat mérlegel, hogy kiválaszthassa közülük a céljainak legmegfelelőbbet. Ezt a folyamatot nevezzük *tervkészítésnek*. *Tervnek* nevezzük ekkor a *lépések* egy halmazát

és a rajtuk értelmezett *kényszerek* és *relációk* összességét. Tehát a terv nem feltétlen azonos annak végrehajtott változatával, hiszen, egyrészt a terv lépései nem feltétlen azonosak a cselekvésekkel; másrészt, amíg a terv tartalmazhat feltételes elemeket is, addig a *végrehajtott terv* már csupán egy cselekvéssorozat, mely szerencsés esetben a környezet kiinduló állapotából a cél-állapotok valamelyikébe vezet.

A cél-állapotok megadása lehet explicit, vagy implicit: Explicit akkor, ha a cél-állapotok halmaza egyértelműen adott, és implicit akkor, ha nem. Az utóbbi esetben többnyire egy *cél-függvény* segítségével dönthetjük el, hogy az adott állapot része-e a cél-állapotok halmazának. A problémákat a következőképp osztályozhatjuk:

- **Egyállapotú problémák** azok, amelyek olyan hozzáférhető (kvázi determinisztikus) környezetet írnak le, ahol az egyes cselekvések kimenetele az ágens számára teljes egészében ismert. Ilyen problémákkal kapcsolatos az *egyállapotú tervkészítés*.

- **Többállapotú problémák** azok, amelyek olyan, nem hozzáférhető (kvázi nem-determinisztikus) környezetet írnak le, ahol az egyes cselekvések lehetséges kimenetelei az ágens számára teljes egészében ismertek. Ilyen problémákkal kapcsolatos a *többállapotú tervkészítés*.

- **Eshetőségi problémák** azok, amelyek olyan, nem hozzáférhető (kvázi nem-determinisztikus) környezetet írnak le, ahol az egyes cselekvések lehetséges kimenetelei az ágens számára nem egészen ismertek. Ilyen problémákkal kapcsolatos az *eshetőségi tervkészítés*.

- **Felderíthetőségi problémák** azok, amelyek olyan, nem hozzáférhető (kvázi nem-determinisztikus) környezetet írnak le, ahol az egyes cselekvések lehetséges kimenetelei az ágens számára (kezdetben) egyáltalán nem ismertek. Ilyen problémákkal kapcsolatos a *felderíthetőségi tervkészítés*.

Definiáljuk tehát az ágensek által megoldandó problémát implicite: adjuk meg az ágensek adott környezetben vett hasznát, s e haszon maximalizálását tekintsük a probléma megoldásának. Vezessünk be egy valós, lehetséges állapot-történetek felett értelmezett *U haszon-függvényt*, aminek maximalizálása megfelel a haszon-függvény és a környezet együttese, egyszóval a *probléma-környezet* által reprezentált probléma megoldásának. Emlékezzünk vissza, hogy az $f_{Ágens}$ függvény által adott *Környezet*-ben generált állapot-történetet $Hatás\left(f_{Ágens}, Környezet\right)$ jelölte. Ennek alapján tehát $f_{Ágens}$ hasznossága a *Környezet*-ben:

$$V\left(f_{Ágens}, Környezet\right) = U\left(Hatás\left(f_{Ágens}, Környezet\right)\right)$$

Hasonlóan, adott *Környezet*-ben adott *Architektúrá*-n futtatott *Program* hasznossága az általa implementált $Ágens\left(Program, Architektúra\right)$ függvény hasznossága:

$$V\left(Program, Architektúra, Környezet\right) = V\left(Ágens\left(Program, Architektúra\right), Környezet\right)$$

Az ágens (avagy az intelligens rendszer) tervezőjének most már csak az a feladata, hogy – a probléma ismeretében – valamilyen elvnek, elvárásnak, követelménynek megfelelően megtervezze ágensét. Ez az elvárás lehet a tökéletes racionalitás. Ekkor

Definíció 4. *Adott probléma-környezetben tökéletesen racionális ágens* f_{opt} *ágens-függvényére teljesül, hogy*

$$f_{opt} = \arg\max_f V\left(f, Környezet\right)$$

A kérdés már csak az, hogy: vajon nem-triviális környezetek esetén is implementálható-e ez a tökéletesen racionális ágens-függvény? Adott *Architektúra* esetén könnyen elképzelhető, hogy $f_{opt} \notin Megvalósítható\left(Architektúra\right)$. Ekkor tehát az ágens-függvény nem implementálható. Node létezhet-e olyan valóságos *Architektúra,* amelyen f_{opt} implementálható? Ha létezne ilyen architektúra, akkor a rajta futó tökéletesen racionális ágens-program, annak érdekében, hogy a következő pillanatban beérkező érzet hatására azon nyomban az optimális cselekvést tudja produkálni, vagy már az érzet beérkezése előtt ki kellene, hogy számítsa az optimális cselekvést, vagy az érzet beérkezésekor nyomban. Az utóbbi eset kizárt, hiszen a számítás a valóságban időt igényel. Az előbbi esetben pedig az ágens megtervezésekor úgy kellene kialakítani programját, hogy már előre tudja a jövőben bekövetkező eseményeket (érzeteket). Ez azonban akauzális (a jövő ismeretét igénylő), nem-triviális környezetek esetén nem elvárható tervezési szempont. Viszont a korlátos optimalitás már reális elvárásokat támaszt az ágens tervezőjével szemben:

Definíció 5. *Adott probléma-környezetben, adott Architektúrá-val rendelkező korlátosan optimális ágens Program-jára teljesül, hogy*

$$Program = \arg\max_{Prg \in Nyelv_{Architektúra}} V\left(Prg, Architektúra, Környezet\right)$$

A korlátos optimalitás tehát már nem az ágens-függvényekre, vagy számításokra, esetleg cselekvésekre tesz megkötést, hanem adott ágens-architektúra mellett az ágens-programokra, s így definíciójából következően *megvalósítható* elvárás az ágenssel (és tervezőjével) szemben. Ha tehát a korlátos optimalitást választjuk a racionalitás mércéjéül, akkor lényegében azt várjuk el, hogy rendszerünk, amennyiben racionális, úgy *képességeihez mérten* a lehető legjobban viselkedjen.

4. Játék-elméleti racionalitás

Az előző fejezetben – a racionalitás fogalmának tárgyalásakor – mindvégig csak egy szereplőre (pl. ágensre, rendszerre) szorítkoztunk, s csak közvetve (pl. az ágens környezetének kimondatlan részeként) érintettünk másokat. A játék-elmélet viszont már közvetlenül is foglalkozik a több szereplő közt létrejövő stratégiai kölcsönhatásokkal. Magyarán a játék-elmélet az ésszerű (racionális) viselkedés elmélete olyan szituációkban, melyekben minden szereplőnek a többi szereplő részéről várható ellenlépések fényében kell meghoznia döntését.

A játék-elmélet a vizsgált szereplőket nem ágenseknek, vagy rendszereknek, hanem játékosoknak tekinti. Számos kiváló matematikus és közgazdász kezdeti munkái után

a játék-elmélet, mint rendszeres tudományos elmélet 1944-ben indult útjára [1]. Eredeti célkitűzése a racionális viselkedés meghatározása – a való életből vett – gazdasági, politikai és társadalmi helyzetekben.

4.1. Játék-elméleti alapfogalmak

A legegyszerűbb – többszereplős döntési problémának megfelelő – játék az, amelyben két játékos két-két stratégiával rendelkezik, s ezek közt kell egyszerre, egymástól függetlenül választania. Egyik játékos számára se ismert, hogy a másik játékos éppen milyen stratégiát választ, viszont mindkettejük haszna (nyeresége, profitja, kifizetése) függ a másik választásától is. Az ilyen kétszereplős játékokat célszerű bimátrix alakban [8] ábrázolni. A 2. ábra néhány bimátrix játékra mutat példát.

1 \ 2	s_1	s_2
s_1	c, c	a, d
s_2	d, a	b, b

1)

1 \ 2	s_1	s_2
s_1	c, c	c, b
s_2	a, c	c, c

2)

1 \ 2	s_1	s_2
s_1	a, b	b, a
s_2	b, a	a, b

3)

2. Ábra: Az *1*-es és a *2*-es játékosok az s_1 és s_2 stratégiák közül választhatnak, amiknek függvényében hasznuk *a*, *b*, *c*, vagy *d* lehet, ahol most $a > b > c > d$.

Kezdetnek vizsgáljuk az 1)-es *játék*ot. Ha az *1*-es *játékos* például az s_1 *stratégiá*t játssza, míg a *2*-es játékos az s_2-t, tömören fogalmazva, ha az *(s_1, s_2) stratégia-kombináció*t játsszák, akkor az *1*-es *játékos haszna a*, míg a *2*-es játékos haszna *d* (lásd. 2. ábrán a második sor harmadik oszlopa), avagy a *játékosok haszna (a, d)*. Látható, hogy az *1*-es játékos s_1 stratégiája a másik játékos tetszőleges stratégia-választása esetén jobb eredményt ad, mint az s_2 stratégia, hiszen $c>d$ és $a>b$. Ekkor azt mondjuk, hogy az *1*-es játékos esetén az s_1 stratégia *domináli*ja az s_2 stratégiát. Hasonló a helyzet a *2*-es játékos esetén is. Az olyan stratégiákat, amelyek minden más stratégiát dominálnak egy adott játékos esetén, a játékos *domináns stratégiá*jának nevezzük. Mivel az 1)-es játékban mindkét játékosnak csupán két-két stratégiája van, ezért mindkettőjük domináns stratégiája s_1.

Ésszerűnek tűnhetne azt mondani, hogy a játékosok mindig a domináns stratégiájukat válasszák, függetlenül attól, hogy a többiek mit döntenek, hiszen másként csak rosszabbul járnának, akármit is döntenek a többiek. Sajnos azonban a legtöbb játékban nincs domináns stratégia (lásd. pl. 2. ábra, 3)-as játék). Mivel tehát a domináns-stratégia választás általában nem elvárható, ezért a játék-elméletben egy ennél valamivel gyengébb elvárást, az egyensúlyra való törekvést szokás racionálisnak tekinteni. A játék-elmélet központi egyensúly-fogalma a Nash-egyensúly [9]. Tömören fogalmazva: *Nash-egyensúly*nak nevezzük azt a stratégia-kombinációt, amely esetén egyik játékosnak se érné meg egyedülálló módon stratégiát váltania.

A Nash-egyensúlyban tehát minden játékos stratégiája legjobb válasz a többiek stratégiájára, s így senkinek se érné meg egyedülálló módon eltérnie az egyensúly

által előírt stratégiától. A Nash-egyensúly tehát bizonyos értelemben „önbeteljesítő”. Viszont több probléma is felmerül vele kapcsolatban.

Egyrészt a Nash-egyensúly *nem kooperatív* meggondolás, hiszen, ha több játékos is stratégiát válthatna egyszerre (pl. közös megegyezés, kommunikáció, bizalom alapján, koalíciókba szerveződve), pontosabban, ha a játékosok valamely, egynél több főből álló csoportja együttműködésben (kooperálva) alakíthatná ki döntéseit, úgy értelmetlenné válna Nash-egyensúlyi stratégiát játszani, hiszen semmit se garantálna.
Ráadásul a Nash-egyensúlyi stratégia még nem kooperatív esetben se jelent feltétlen garanciát, hiszen bizonyos játékosok (akár csak egyedülálló, irracionális módon) eltérhetnek Nash-egyensúlyi stratégiájuktól, s így nem csak ők járnak rosszabbul, hanem esetleg azok is, akik racionálisan a számukra előírt Nash-egyensúlyi stratégiát követték. Az ilyen eseteket *hiteltelen fenyegetés*nek nevezik, hiszen az a játékos, amelyik eltér egyensúlyi stratégiájától, saját magának is árt, ami ugyebár neki sem lehet ésszerű érdeke. A hiteltelen fenyegetéseket viszont kiküszöböli a Nash-egyensúly egy finomítása: az *aljáték-tökéletes egyensúly* [10].

A 2. ábra 1)-es játékában jól láthatóan csak egyetlen Nash-egyensúly van: az (s_1, s_1) stratégia-kombináció. Ez az egyetlen olyan stratégia-kombináció tehát, amely esetén egyik játékosnak se érné meg külön-külön más stratégiát választania. Viszont, ha a játékosok az (s_2, s_2) stratégia-kombinációt játszanák, hasznuk egyaránt magasabb lenne, hiszen $b>c$. Ebből következik, hogy a Nash-egyensúly nem feltétlen „optimális”. Az (s_2, s_2) stratégia-kombináció az 1)-es játék *Pareto-optimum*a, mivel nincs más olyan stratégia-kombináció, amely esetén a játékosok valamely részhalmaza jobban jár, míg a többiek nem járnak rosszabbul. Tehát a Nash-egyensúly bizonyos esetekben (pl. Pareto-értelemben) *szub-optimális*nak adódik.

A 2. ábra 2)-es játékában szemmel láthatóan több, pontosan három különböző Nash-egyensúly is adódik: az (s_2, s_1), (s_1, s_2) és a szub-optimális (s_2, s_2) stratégia-kombináció. A Nash-egyensúly tehát *nem* feltétlen *egyértelmű*. Felmerül a kérdés, hogy egy „racionális” játékos melyiket válassza, illetve, hogy vajon a többi játékos is azt választja-e majd? Az így adódó *egyensúlyválasztási problémá*nak egész irodalma van (lásd. pl. [11]). A játék valamely adott egyensúlyválasztási elv mellett adódó Nash-egyensúlyát a *játék fókusz-pontjá*nak nevezzük.

A 2. ábra 3)-as játékának pedig *nincs Nash-egyensúly*a, hiszen bármely stratégia-kombináció esetén létezik olyan játékos, amelynek megérné más stratégiát választania. Ezt a problémát viszont feloldja a *kevert stratégiá*k, s így a *kevert Nash-egyensúly* fogalmának bevezetése. Az eddigiekben tehát mindvégig olyan *tiszta stratégiá*ról beszéltünk, melyeket a játékosok 1-valószínűséggel követtek. Ha viszont megengedjük, hogy a játékosok egynél kisebb valószínűséggel is választhassák stratégájukat, úgy a kevert stratégiák fogalmához jutunk. Látható, hogy a 2. ábra 3)-as játékának, bár tiszta Nash-egyensúlya nincs, van viszont (egyértelmű) kevert Nash-egyensúlya. Ha mindkét játékos ½ valószínűséggel választja mind az s_1, mind az s_2 stratégiát, azaz, ha mindketten az (½, ½) kevert stratégiát, avagy *lutrit* játsszák, akkor egyiküknek se áll érdekében ezen változtatni. A kevert stratégiák hasznát a játékosok hasznának várható értékeként számítjuk. Egy-egy kevert stratégia-kombináció játszásakor azonban csak valamely tiszta stratégia-kombináció kerül lejátszásra.
A Nash-egyensúly létezésére vonatkozó tétel [9] – nem túl szigorú megkötéseinek köszönhetően – szinte minden játékban garantál legalább egy kevert Nash-egyensúlyt.

Így tehát a kevert stratégiák bevezetésével megoldódik az egyensúly hiányának problémája. Sajnos azonban a kevert Nash-egyensúllyal kapcsolatban is fennáll a tiszta Nash-egyensúly kapcsán említett legtöbb probléma: (1) Kooperatív játékosok esetén nem alkalmas racionalitás definíció. Természetesen a kooperatív játék-elmélet jónéhány alternatív egyensúly-definíciót javasolt (lásd. pl. [12]), ám ezek közös jellemzője, hogy csak igen szigorú, speciális feltételek mellett garantálható a létezésük. (2) Mivel kevert Nash-egyensúlyból is lehet több, ezért továbbra is fennáll az egyensúlyválasztás problémája. (3) A kevert Nash-egyensúly is lehet szub-optimális.

4.2. Ágens- és játék-elmélet összekapcsolása

Jól látható, hogy mindeddig csak olyan racionalitás fogalmakat érintettünk, melyek a játékosok stratégiáira, avagy stratégia-kombinációkra vonatkoztak. Ennek oka egész egyszerűen az, hogy a játék-elméletben ez a szemlélet egyeduralkodó. Viszont, ha megfontoljuk az ágens-elméleti korlátos optimalitás megközelítését, mely szerint nem döntésekre, hanem döntéshozási mechanizmusokra vonatkoztatjuk a racionalitást, akkor kiküszöbölhetővé válik az előbb felsorolt problémák jelentős része.

Tegyünk tehát megkötést a játékosokra, pontosabban a játékosok stratégia-választási mechanizmusára, ne pedig az általuk választott stratégiákra. Érzékeltetésképp gondoljunk csak bele abba, hogy miképpen alakul a 2. ábra 1)-es játéka akkor, ha az egyes játékosok úgy választják meg stratégiájukat, hogy közben azt az elvet követik, mely szerint „azt teszik másnak, amit maguknak is kívánnak”. Mivel a racionális játékosok végeredményben hasznuk növelésére törekszenek, s az 1)-es játékban az (s_2, s_2) stratégia-kombináció adja a legmagasabb hasznot mindkét játékos számára, ezért a fenti stratégia-választási elvet követve, mindketten az s_2 stratégiát választanák. Ha tehát a játékosok döntéshozási mechanizmusa a fenti elv szerint működik, akkor a 2. ábra 1)-es játékában optimálisan cselekszenek. Természetesen ez az elv nem minden játék esetén garantál optimális kimenetelt. Viszont a példa érzékelteti, hogy a játékosok stratégia-választási mechanizmusára tett megkötések képesek lehetnek feloldani a stratégiákra tett racionalitási megkötések szub-optimalitását. Ez tehát jelentős érv amellett, hogy – hasonlóan az ágens-elméleti korlátos optimalitás fogalmához – a játékosok stratégia-választási mechanizmusára vonatkoztassuk a racionalitást, ne pedig a választott stratégiákra.

Ha meggondoljuk, akkor már a többszörös Nash-egyensúlyok okán felmerülő egyensúlyválasztási elv kijelölése is egy implicit megkötés a játékosok stratégia-választási mechanizmusára. Viszont esetünkben nem implicit, hanem explicit megkötésekre van szükség a játékosokkal kapcsolatban ahhoz, hogy – hasonlóan a 3. fejezetben tárgyaltakhoz – elvárható racionalitási követelményeket támaszthassunk velük szemben.

Egyesítsük tehát az ágens és játék-elméletet, s vezessük be az ágensek kapcsán tárgyalt főbb fogalmakat: tekintsük a játékosokat ágensnek. A játékosok stratégiáit terveknek. Az egyes játékosok hasznát tekintsük problémát definiáló haszon-függvénynek (így mindegyik játékos más-más problémával állhat szemben). Minden játékosnak legyen egy-egy architektúrája, mely adott program-nyelv programjait futtatja. Az ágens-programoknak a játékosok úgynevezett *Típus*-át feleltessük meg, s

így az ágensek program-nyelvét, azaz a lehetséges ágens-programok halmazát a játékosok lehetséges típus-halmazának tekintsük (*Típusok*).

A játékosok típusainak bevezetésére eredetileg azért volt szükség, hogy a nem teljes információjú játékokat vissza lehessen vezetni teljes, de nem tökéletes információjú játékokra [13]. Ez lényegében kétfajta megkülönböztetést jelent. *Teljes információjú játékok* esetén a játékosok teljeskörű információval rendelkeznek a játék – normál (vagy extenzív) formájával meghatározott – alapvető matematikai szerkezetéről. *Nem teljes információjú játékok* esetén ez nem teljesül. Tehát itt aszerint osztályozzuk a játékokat, hogy a játékosok milyen mértékben tájékozottak a játék olyan jellemzőiről, amelyek már a lejátszás előtt adottak (játékosok száma, lehetséges stratégiái, haszon-függvényei, stb.). *Tökéletes információjú játékok* alatt pedig olyan játékokat értünk, amelyekben a játékosok teljeskörű információval rendelkeznek a lejátszás során addig bekövetkezett lépésekkel kapcsolatban. Nem *tökéletes információjú játékok* esetén ez nem igaz. Tehát ebben az esetben aszerint osztályozunk, hogy a játékosok milyen mértékben tájékozottak a lejátszás során bekövetkezett eseményekkel kapcsolatban.

A következőkben tegyük fel, hogy a játékos (vagy tervezője) előtt nem pontosan ismert az aktuális játék, viszont minden játékos pontosan tudja, hogy melyik típus képviseli a játékban. Álljon fenn továbbá a közös apriori becslések (common priors assumption) feltevése is [14], miszerint minden játékos ismeri a lehetséges típus-kombinációk apriori valószínűségeit, s ennek ismeretével kölcsönösen is tisztában vannak. Namármost tegyük fel, hogy minden játékos pontosan tisztában van a játék matematika szerkezetével, s ismerik egymás architektúráját. Ekkor már csak az a kérdés, hogy a többi játékosnak mi a típusa – ez jelenti az egyedüli bizonytalanságot. Tehát mindennemű – játékkal kapcsolatos – bizonytalanságot visszavezettünk egy olyan nem tökéletes információjú játékra, melyben csupán a többi játékos típusa nem ismert az egyes játékosok előtt. Ha tehát egy adott játékos haszna egy *Játék*-ban $V(Típus, Játék)$, ahol a *Játék* része a többi játékos típusa és az adott játékost *Típus* képviseli, akkor

Definíció 6. *Adott játékban korlátosan optimális játékos Típus-ára teljesül, hogy*

$$Típus = \arg\max_{Típ \in Típusok} V(Típ, Játék)$$

Tehát egy adott architektúrájú, korlátosan optimális játékosnak olyan a programja, amely adott játék (s így a többi játékos adott típusa) esetén maximálja a játékos várható hasznát. Tehát immár a játék-elméletben is rendelkezésünkre áll egy, a mesterséges intelligencia racionalitás-definíciójával analóg, gyakorlatban is használható racionalitás definíció. Ezen felül, mivel a játékosokat ágenseknek, stratégiáikat pedig terveknek tekintettük, immár lehetőségünk nyílik a játékelméleten belül is felhasználni a tervkészítés elméletének gazdag eredményeit. A játék-elmélet számára tehát előnyös az ágens-elmélettel való összekapcsolás.

Az ágens-elmélet szemszögéből se elhanyagolható a nyereség, amit a játék-elmélettel való összekapcsolás eredményez. Mivel a játék-elméleten belül a típus-reprezentáció módot ad tetszőleges nem teljes információjú játék teljes, ám nem tökéletes információjú játékká alakítására, ezért lényegében a fenti megfontolások a (játékos)ágensek összes bizonytalanságát képesek visszavezetni az előzőleg

bekövetkezett események hiányos ismeretére. Ezáltal formálisan is kezelhetővé válik az eshetőségi és felderíthetőségi tervkészítés (lásd. 3.), ami mind a mai napig megoldatlan feladat. Ráadásul a multi-ágens rendszerekben fellépő társadalmi jelenségek is kiforrott, következetes formális reprezentációt nyerhetnek a játék-elmélet bevezetésével. A játék-elmélet főbb fogalmai az ágens-közösségeken belül is leírhatóvá válnak (pl. egyensúly, kooperáció), hiszen a 3. fejezet nem tett megkötést arra vonatkozólag, hogy az ágens környezete mit tartalmaz. Ily módon akár más ágenseket is tartalmazhat, melyeket immár játék-elméleti eszközökkel modellezhetünk. Az elméletek összekapcsolása tehát nagyban megnöveli leíró-erejük, lehetővé téve a körültekintő, választható tervezési szempontok érvényesítését. Egységes módon válik tervezhető az egyéni és a csoportos működés, hiszen amíg az ágens-elméleti eszköztár az egyén működésére koncentrál, addig a játék-elmélet az egyén csoportviszonylatban vett működésére.

5. Evolúciós-elméleti racionalitás

Az előző fejezeteket a korlátos optimalitás tetőzte be, ám arról egyikben sem ejtettünk szót, hogy a korlátosan optimális megoldás (program, stratégia, stb) megtalálása miképpen lehetséges. Az optimális megoldás keresésére számos módszer ismeretes, melyek közt a természetes evolúció számítási modellje rangos helyet foglal el. Az evolúciós számítási modellek három fajtáját szokták megkülönböztetni aszerint, hogy miként interpretálják a darwini posztulátumot, miszerint a közös erőforrásokért versengő populáció azon egyedei kerülnek előnyösebb helyzetbe, melyek a versenyben előnyt jelentő (racionális) tulajdonságokat hordoznak. A legmagasabb szintű megközelítés *evolúciós programozás* (lásd. [15], [16]) néven vált ismertté, és a *fajok* közti versengést modellezi. Köztes szinten mozognak az *evolúciós stratégiák* [17], melyek az *egyed* szintjén vizsgálják a természetes kiválasztódás darwini folyamatát. A legelemibb megközelítésnek a *genetikus algoritmusok* [18] családjába tartozó módszerek tekinthetők, melyek a *gének* közti versengés szintjén követik nyomon az evolúció folyamatát. Ebben a fejezetben ez utóbbi módszerekkel foglalkozunk, illetve az ágens- és játék-elmélettel való kapcsolatukkal.

5.1. Genetikus Algoritmusok

A genetikus algoritmusok – és változatai – *globálisan konvergens*, sztochasztikus keresőeljárások [19]. Ez azt jelenti, hogy – a véletlenszerűség felhasználásával – végzett (nem randomizált) keresés során bizonyítottan tartanak a keresési tér – jósági mérce által definiált – *globális optimumához.*

Az evolvált *egyed*eket klasszikus esetben bináris bitfüzérek (kromoszómák, avagy génláncok) reprezentálják, melyek pozícióinak (alléljainak) bit-értéke felel meg a géneknek. Az algoritmus lényege, hogy egyedek egy kezdeti, véletlenszerűen előállított *populáció*ját – úgynevezett *genetikus operátorok* felhasználásával – addig-addig módosítja, mígnem a populáció megfelelő jellemzői (pl. legjobb egyed, egyedek átlagos jósága) eleget nem tesznek egy választott *leállási kritériumnak.* Az egyedek jóságát megbecslő jósági mércét szokás *fitness függvénynek* is nevezni. A fitness függvény a populáció egyedeihez rendel egy-egy valós számértéket, mely az adott egyed globális (a populáció viszonylatában vett) *jóság*át szimbolizálja. Alapesetben a következő genetikus operátorokat különböztetjük meg:

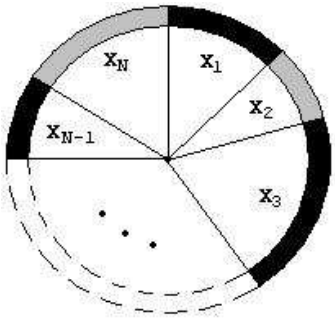
- A *mutáció* adott p_m valószínűséggel, találomra módosítja a génlánc (pl. bitfüzér) alléljain található gének (pl. bitek) értékét (pl. invertálja a bitet).

- A *keresztezés* p_c valószínűséggel kiválaszt egy-egy génláncot, majd megcseréli egy véletlenszerűen választott pozíció után következő részeiket.

- A *szelekció* a populáció egyedeinek jóság szerint történő, véletlenszerű kiválasztása és átörökítése a következő generációba.

A fenti operátorok felhasználásával az algoritmus menete a következő: kezdetben előállítjuk egyedek egy véletlenszerű populációjat. Az egyedek hossza, a populáció mérete, és az operátorok valószínűsége adott. Ezt követően minden egyedhez rendelünk egy-egy jósági értéket, mely alapján kiválasztunk közülük egy – a szelekció típusától függő – mennyiséget. A kiválasztott egyedeket párokba rendezzük. A párokat (szülőket) keresztezzük, majd mutáljuk. Így kapjuk a *következő generáció* egyedeit (a gyerekeket). Ezt ismételjük generációról generációra mindaddig, míg nem teljesül a leállási kritérium.

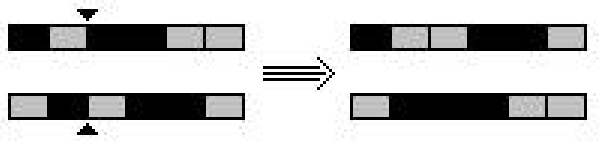
A fentebb említett három alapvető genetikus operátor megvalósítására több lehetőség is kínálkozik. Ezek közül a fontosabbak: A szelekció *rulett kerék* elven működik, ha az egyedeket jósági értékükkel arányos valószínűséggel választja ki a következő generációba. Tegyük fel, hogy a populáció n egyedből áll. Legyen a populáció valamely x_i i egyedének jósága $V(x_i)$ minden $i = 1, 2, \ldots, n$-re. Ekkor (képzeletben) osszunk fel egy egységnyi kerületű „rulett-kereket", avagy kört n darab $V(x_i) \Big/ \sum_{j=1}^{N} V(x_j)$ ívhosszú körcikkre, majd „perdítsük meg a kereket", azaz állítsunk elő egy *[0,1)* tartományba eső, valós véletlen-számot. Ahol a „golyó megáll", azaz amely egyednek megfelelő körcikk íve által meghatározott *[0,1)*-beli zárt rész-intervallumba esik az előállított véletlen-szám, azon egyed kerül kiválasztásra. A 3. ábra szemlélteti a rulett-kerék elven alapuló szelekciót.



3. Ábra: Rulett-kerék elven alapuló szelekció

A *mérkőzés* alapú szelekció is rulett-kerék elven működik: *K*-szor „megpörgetjük a rulett-kereket", majd az így kiválasztott *K* egyed legjobbjával térünk vissza. Az új generáció előállításához ezt ismételjük *N*-szer. A *felső százalék* elve szerint a populáció egyedeinek legjobb *N* százalékából választjuk ki véletlenszerűen a következő generáció egyedeit. A *valahány legjobb* kiválasztás lényege, hogy az egyedeket jóság szerint sorba rendezzük és az első valahányat választjuk ki közülük.

A keresztezésnek is több módja lehetséges. Létezik *egy-pontú* és *több-pontú keresztezés*. A több-pontú keresztezés az egy-pontú keresztezés értelemszerű kiterjesztése. Egy-pontú keresztezésre mutat példát a 4. ábra.



4. Ábra. Egy-pontú keresztezés

A mutációs operátor az egyed valamely véletlenszerűen választott alléjának génjét változtatja meg találomra, így – definit megfogalmazásának köszönhetően – nincsenek változatai. Más struktúrájú gének (pl. nem bitfüzérek, hanem programfák [4]) esetén nyilván más módon kell implementálni, ám elve változatlan marad.

A fentebb említett alap-operátorokon túl természetesen még sok egyéb operátor lehetséges, melyek közül talán még az *elitizmus*t említenénk. Az elitizmus a szelekció egy változata, mely – akár a szelekcióval együttműködésben – egy adott generáció valahány legjobb egyedét örökíti át változtatás nélkül a következő generációba. Újabban bizonyítást nyert, hogy a genetikus algoritmusok általános keresési terekben is globálisan optimálisak, ha használják az *elitizmust*, továbbá a populáció minden egyes egyede *elérhető* bármely más egyedből kereszteződés és mutáció révén [20].

5.2. Játék, ágens és evolúciós-elmélet összekapcsolása

A játék- és ágens-elmélet összekapcsolásának lehetőségeiről már szót ejtettünk (lásd. 4.2). Az evolúciós- és játék-elmélet összekapcsolása is megtörtént már. Az *evolúciós játék-elmélet*ben [21] jellemzően játékosokból alkotott populációk dinamikáját vizsgálják. Tehát a populáció egyedeit tekintik játékosnak, és a fitness függvény a játékosok haszon-függvénye.

A populációs dinamikák leírására legtöbbször a diszkrét dinamikus rendszerek (DDR) módszertanát használják. Az evolúciós játékelmélet központi racionalitás-fogalmának tekinthető *evolúciósan stabil stratégiá*kat (ESS) [22] is ily módon származtatják, mint a populációs dinamika bizonyos „fix-pontjait”. Tömören fogalmazva, ha a populáció egyedei evolúciósan stabil stratégiát játszanak, úgy a populáció nem változik, s ha bárki is eltér a számára előírt ESS stratégiától, csak rosszabbul járhat (mind haszon, mind pedig túlélési esélyek tekintetében). Az ESS stratégiák tehát legtöbbször megegyeznek a Nash-egyensúlyi stratégiákkal, aminek következtében gyengeségeiket is öröklik (lásd. 4.1). Ezért tehát, hasonlóan a játék-elmélethez, más racionalitás definíciót érdemes keresnünk.

Ahhoz, hogy *használható* racionalitás-definícióhoz jussunk az evolúciós-elméleten belül, megoldást jelenthet az ágens-elmélettel való összekapcsolás [23]. Nem elég tehát csupán annak a kijelentése, hogy „az racionális, ami a túléléshez kell”, hanem arra is választ kell tudnunk adni, hogy „mi alapján dől el egy egyed, illetve egyedek egy csoportjának a túlélése – mi határozza meg jóságukat”. A két elmélet összekapcsolása során tehát az ágenseket a populáció egyedeinek, az – ágensek

környezetében fennálló problémát reprezentáló – haszon-függvényt pedig az egyedek jóságát meghatározó fitness függvénynek célszerű tekintenünk. Ekkor lényegében az ágensek probléma-környezete (lásd. 3. fejezet) egy evolúciós optimalizálási problémának felel meg. Hasonlóan, bármely evolúciós optimalizálási probléma triviálisan átfogalmazható ágensek egy megfelelő probléma-környezetévé.

Az ágens- és evolúciós elmélet összekapcsolása az ágens-elmélet szempontjából előnyös, hisz lehetőséget kínál az „emergent”-jellegű, és más egyéb populációs jelenségek evolúciós vizsgálatára. Az evolúciós-elmélet szemszögéből nézve is előnyösnek mondható a két elmélet összekapcsolása, hisz lehetőség nyílik egyrészt az egyedek ágens-elméleti vizsgálatára, mely során az egyedek felépítésébe és működésébe is bepillantást nyerhetünk, másrészt használható módon definiálható az egyedek racionalitása (pl. korlátosan optimális egyed fogalma).

Összességében az evolúciós-, játék- és ágens-elmélet összekapcsolása lehetőséget teremt evolúciós eszközökkel optimalizált, ágens-elméleti léptékkel tárgyalható bonyolultságú komplex-rendszerek játék-elméleti eszközökkel történő vizsgálatára. Végeredményben a három elmélet összekapcsolásával egy – az eddigiekhez képest – átfogó racionalitási elv megfogalmazása válik lehetségessé, mely alapjául szolgálhat az intelligens rendszerek egységes elven történő tervezésének és elemzésének.

6. Összefoglalás

Három főbb témakört érintettünk a cikk során: az ágens-, játék- és evolúciós-elméleteket. Mindhárom esetben a racionalitás fogalmát használtuk az egységes tárgyalás alapjául, hogy rámutassunk az elméletek összekapcsolási lehetőségeire. A három elmélet összekapcsolási lehetőségeinek felvázolásával egy átfogó rendszer-specifikációs elvet körvonalaztunk, amely egységes elméleti alapot teremthet az intelligens rendszerek egységes tervezéséhez és elemzéséhez.

Az elméletek összekapcsolása céljából a mesterséges intelligencia „újkeletű” racionalitás fogalmát, a korlátos optimalitást használtuk. Rendre összevetettük a hagyományos játék-elméleti és evolúciós-elméleti racionalitás-fogalmakkal. Mindkét esetben arra a megállapításra jutottunk, hogy a korlátos optimalitás előnyösebb, mivel használható, a gyakorlatban is elvárható kritériumokat támaszt a tervezett rendszerrel szemben. Ezzel egy olyan átfogó jósági mércéhez jutottunk, melyel a rendszerek hasznossága egyaránt értelmezhető egyéni, csoportos, illetve csoportközi szinten.

Végeredményben tehát egy olyan, az eddigiekhez képes átfogó elvet körvonalaztunk, mely alkalmas alapját képezheti egy, az intelligens rendszerek egységes tervezésére és elemzésére irányuló elméletnek, áthidalva az intelligens rendszerek tervezésével kapcsolatos eddigi főbb nehézségeket.

Irodalomjegyzék

[1] von Neumann, J., & Morgenstern, O. (1944). *Theory of Games and Economic Behavior*, Princeton University Press: Princeton.

[2] Good, I. J. (1971). Twenty-seven principles of rationality. In Godambe, V. P., & Sprott D. A. (Eds.), *Foundations of Statistical Inference,* pp. 108-141. Holt, Rinehart, Winston.: Toronto.

[3] Simon, H. A. (1982). *Models of bounded rationality*, *Volume 2*, MIT Press: Cambridge.

[4] Koza, J. R. (1992). *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, MIT Press: Cambridge.

[5] Russell, S. J., & Subramanian, D. (1995). Provably bounded-optimal agents, *Journal of AI Research*, 2:1–36.

[6] Russell, S. J., & Norvig, P. (1995). *Artificial Intelligence: A Modern Approach,* Prentice Hall.

[7] Wooldridge, M. (1999). Intelligent Agents, In Weiss, G. (Ed.), *Multiagent Systems*, MIT Press.

[8] Lója K. (2003). Játékelméleti módszerek, *Híradástechnika*, 4:29–34.

[9] Nash, J. F. (1951). Non-cooperative games, *Annals of Mathematics*, 54(2):286--295.

[10] Selten, R. (1975). Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games, *International Journal of Game Theory*, 4: 25-55.

[11] Harsányi, J. C., & Selten, R. (1988). *General theory of equilibrium selection in games*. MIT Press.

[12] Aumann, R. (1959). *Acceptable Points in General Cooperative n-Person Games*, in Contributions to the Theory of Games IV, Princeton: Princeton Univ. Press.

[13] Harsányi, J. C. (1967-1968). Games with incomplete information played by Bayesian players I-II-III, *Management Science*, 14:159-182;320-334;486-502.

[14] Fudenberg, D., & Tirole, J. (1991). *Game Theory*, Cambridge, MA: Cambridge University Press.

[15] Fogel, L. J., Owens, A. J., Walsh, M. J. (1966). *Artificial Intelligence through Simulated Evolution*, New York: John Wiley.

[16] Fogel, D. B. (1995). *Evolutionary Computation: Toward a New Philosophy of Machine Intelligence*, Piscataway, NJ: IEEE Press.

[17] Schwefel, H. P. (1995). *Evolution and Optimum Seeking*, Sixth-Generation Computer Technology Series, John Wiley & Sons, Inc., New York.

[18] Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley.

[19] Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology*, Control, and Artificial Intelligence, MIT Press.

[20] Rudolph, G. (1996). Convergence of Evolutionary Algorithms in General Search Spaces, in *Proc. of IEEE International Conference on Evolutionary Computation*, pp. 50–54, Nagoya, Japan.

[21] Weibull, J. W. (1995). *Evolutionary Game Theory*, Cambridge, Mass.: MIT Press.

[22] Maynard-Smith, J. (1982). *Evolution and the theory of games*, Cambridge University Press: Cambridge.

[23] Kovács, D. L. (2003). DINA: Dynamic INtelligent Agents, *Proceedings of IEEE International Symposium on Intelligent Signal Processing*, IEEE.