Közösségi döntések implementációjának új megközelítése

KOVÁCS DÁNIEL LÁSZLÓ

BME, Villamosmérnöki és Informatikai Kar, Méréstechnika és Információs Rendszerek Tanszék
dkovacs@mit.bme.hu

Reviewed

Kulcsszavak: intelligens ágensek, korlátos optimalizálás, implementációs elmélet

Az intelligens rendszerek jelentős szerepet játszanak mindennapjainkban. Ennek ellenére mindmáig nincsen olyan átfogó rendszerspecifikációs elv, amely lehetővé tenné e rendszerek egységes tervezését, és elemzését. Az intelligens rendszerek tervezése, és elemzése tehát mind a mai napig esetleges, megoldandó feladathoz igazított, többnyire ad-hoc módon történik.

1. Bevezetés

A fent említett problémára keres megoldást a játék, ágens, és evolúciós elméletek egyesítése [1]. Az egyesítés kulcsa, pontosabban az intelligens rendszerek jóságának általános mércéje a racionalitás egy újfajta definícióján, a korlátos optimalitáson alapszik. Egy intelligens rendszert (ágenst) akkor tekintünk korlátosan optimálisnak, ha a környezetében kivitelezett cselekvéseit egy olyan program szerint választja meg, amelynél nincs jobb azok között, amelyeket futtatni képes [2]. Ahhoz tehát, hogy módunkban álljon ilyen rendszerekről érdemben beszélni, szükségünk lesz az ágensek programjának egy használható, absztrakt modelljére. Ebből a célból kerül bevezetésre a virtuális haszon fogalma, mint az ágens-programok modelljének egy központi összetevője.

Az intelligens rendszerek döntési mechanizmusának, avagy az ágensek programjának ily módon történő modellezése lehetővé teszi az ágensekből alkotott közösségek működésének (pontosabban a közösségi döntések implementációjának [3]) egy újfajta, az eddigieknél hatékonyabb megközelítését.

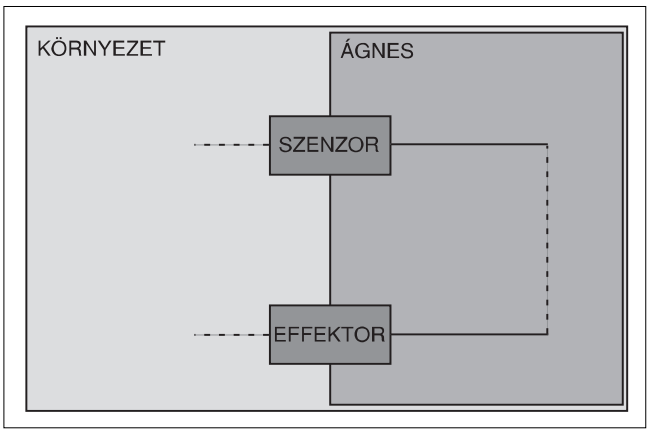
2. Hogyan modellezzük az intelligens rendszereket?

Tegyük fel, hogy az intelligens rendszerek modellezhetők ágensként *(1. ábra)*. Egy *ágens „bármi lehet, amit úgy tekintünk, mint ami szenzorai segítségével érzékeli környezetét, és effektorai segítségével megváltoztatja azt”* [4]. Ennek a föltételezésnek az adja a létjogosultságát, hogy – túl azon, hogy intuitív, és kellőképp általános – lehetővé teszi az intelligens rendszerek környezetének, szenzorainak, és effektorainak konkrét megfeleltetését. Azaz tetszőleges intelligens rendszer esetén megadható a környezet, az érzékelő szervek, és a beavatkozó szervek konkrét megfeleltetése.

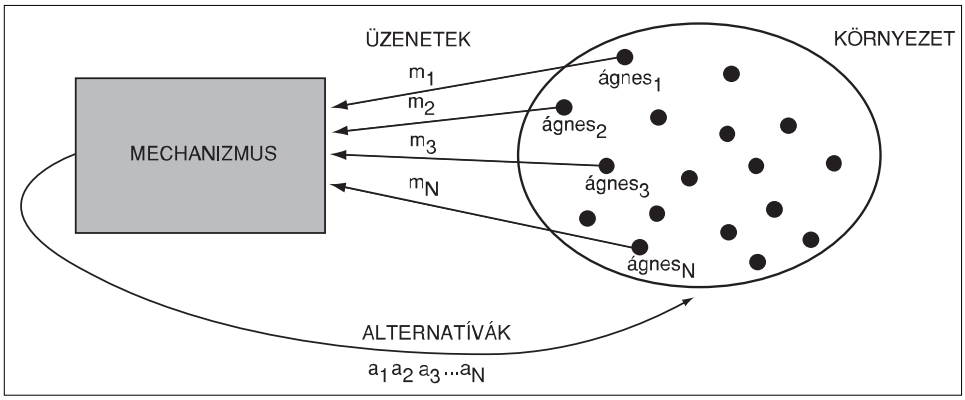
A környezetébe ágyazott ágens minden pillanatban a következő cselekvés kiválasztásának problémájával szembesül (ahol magát a tétlenséget is egyfajta cselekvésnek tekinthetjük). Nem-triviális környezetek esetén *tervkészítésre* van szükség e döntések hatékony meghozásához. Több-ágenses környezetben az egyes ágensnek ráadásul még a többi ágens viselkedését is figyelembe kell vennie ahhoz, hogy hatékony cselekvési tervet készíthessen. E helyzetek modellezésére nyújt alkalmas keretet a játékelmélet [5], amely az ágensek egymásra gyakorolt hatását játékosok közt fellépő stratégiai kölcsönhatásoknak tekinti egy *játékban*, ahol az ágensek a *játékosok*, terveik pedig a *játékosok stratégiái* [6]. Mindazonáltal a játékelmélet csak az egyed szempontjából, nem pedig közösségi szinten vizsgálja a döntéshozás kérdését.

Jelenleg az implementációs elmélet (mint a játékelmélet egyik legújabb ága) foglalkozik közösségi döntési helyzetek modellezésével. Az ágenseket együttesen *közösségnek* tekinti, melynek céljai egy *közösségi döntési szabály* (KDSZ) formájában összegezhetők, azaz egy olyan leképzés formájában, amely a releváns rejtett paraméterek alapján előállítja a végkimeneteleket. *Magyarán az ágens-közösséget úgy tekinti, mint ami – kollektív entitásként – egy adott KDSZ-nek megfelelően cselekszik.* A KDSZ tehát úgynevezett közösségi alternatívákat (például végkimeneteleket) állít elő a közösségen belüli ágensek privát információja (például egyéni – végkimenetelek felett értelmezett – preferen-

1. ábra Intelligens ágensek általános felépítése



ciái) alapján. Az egy-értékű KDSZ-t szokás *közösségi döntési függvénynek* (KDF) is nevezni. Az implementáció problémája ekkor a következőképp foglalható össze: *Adható-e olyan mechanizmus, amely az ágensek adott elv szerint hozott döntései mellett a közösségi optimumot implementálja?* [3]



2. ábra Az implementáció problémája

A *2. ábra* valamivel részletesebben szemlélteti az implementáció problémáját: adott tehát egy Tervező, akinek a feladata az, hogy – az ágensek adott viselkedését feltételezve – olyan mechanizmust hozzon létre, amely implementál egy adott KDSZ-t. Pontosan fogalmazva: a cél egy olyan mechanizmus létrehozása, amely adott környezet mellett ugyanazokat az $a_1, a_2, a_3, ..., a_N$ alternatívákat (például végkimeneteleket) eredményezi, mint egy adott KDSZ, feltéve, hogy az *1, 2, 3, ..., N* ágensek egy adott *S* játékelméleti megoldási elvnek (például domináns stratégiák, Nash-egyensúly [7]) megfelelően választják $m_1, m_2, m_3, ..., m_N$ üzeneteiket (avagy a mechanizmusban játszott stratégiáikat). Amennyiben az adott feltételek mellett létezik ilyen mechanizmus, úgy a KDSZ-t *S-implementálható*nak nevezzük.

A fenti megközelítésnek több előnye is van. Képes például szociális intézmények, külsődleges társadalmi ráhatások, ágensek közötti megállapodások modellezésére. Számos közgazdasági, politikai helyzet modellezésére alkalmas. Ismeretes például, hogy, ha az ágensek által követett *S* játékelméleti megoldási elv a domináns stratégiák (azaz, ha az ágensek mindig a domináns stratégiájukat választják, amely minden más stratégiájuknál jobb eredményt ad függetlenül attól, hogy a többi ágens milyen stratégiát választ), akkor kizárólag diktatórikus KDF-ek implementálhatók. A diktatórikus KDF mindig egy adott ágens – kimenetelek felett értelmezett – preferenciáinak kedvez, azaz olyan kimenetelt eredményez, ami az adott ágens hasznát maximálja. Ennek az igen „negatív” eredménynek az egyik legfőbb oka az, hogy nem minden játékban van a játékosoknak domináns stratégiája.

Az előnyök mellett természetesen a megközelítésnek több hátránya is van. Nem közgazdasági, vagy társadalmi helyzetekben, hanem például az informatikában, mesterséges intelligens rendszerek (szoftver ágensek, robotok stb.) tervezésekor a Tervezőnek *közvetlen* ráhatása van a rendszer belső felépítésére, működésére, programjára). Az *S* megoldási elv viszont csak egy *közvetett* feltételezés erre vonatkozólag. Nyilván ennek az okai az implementációs elmélet társadalomtudományi gyökereiben keresendők, ahol az ágensek (vállalatok, emberek stb.) nem megváltoztatható módon adottak. Felvetődhet tehát a kérdés, hogy miért is kellene az ágenseknek éppen egy adott *S* elvnek megfelelően működnie? Az ilyen, és ehhez hasonló kérdésekre az implementációs elmélet sajnos már nem ad magyarázatot. Hátránynak tekinthető továbbá, hogy az ágensek egy *központi* mechanizmuson keresztül kénytelenek cselekedni, ami ráadásul *globális hozzáférésel* bír az ágensek környezetéhez. Ez általában véve egy irreális feltevés, főként intelligens ágens-rendszerek tervezésekor, ahol az ágensek működése legtöbbször *decentralizált*, és a környezethez (például Internet, Mars felszíne) való hozzáférés többnyire csak lokális.

További hátrány, hogy ahhoz, hogy egy KDSZ implementálható legyen, általában igen sok *speciális feltételnek* kell eleget tennie (monotonitás, ordinalitás, indíték kompatíbilitás stb.), ami igencsak leszűkíti az implementálható KDSZ-ek körét. Végül, de nem utolsó sorban hátrány, hogy általános esetben csakis *approximatív implementáció* lehetséges, azaz tetszőleges – a speciális feltételeknek eleget tevő – KDSZ implementálható, de csak megközelítőleg, valamekkora hibával. Ezt nevezik *virtuális implementációnak* [8].

3. Közösségi döntések implementációjának új megközelítése

Az implementációs elmélet fentebb felsorolt hátrányainak kiküszöbölését tűzi ki célul a virtuális haszon alapú döntéshozás elve. A környezet legyen egy *játék*, melyben az ágensek legyenek a *játékosok*, terveik pedig a *játékosok stratégiái*, továbbá minden játékoshoz tartozzon egy *haszonfüggvény* is, amely minden lehetséges stratégia-kombináció esetén megadja az adott játékos környezetben vett valós hasznosságát. Ez lesz tehát az a hasznosság, amit az adott játékos valójában elér, ha mindenki a stratégia-kombinációban neki megfelelő stratégiát játssza. Minden egyes játékos rendelkezzen továbbá *a játék egy belső reprezentációjával,* azaz a játék egy modelljével (beleértve a többi ágenst, stratégiáikat, és haszonfüggvényeiket).

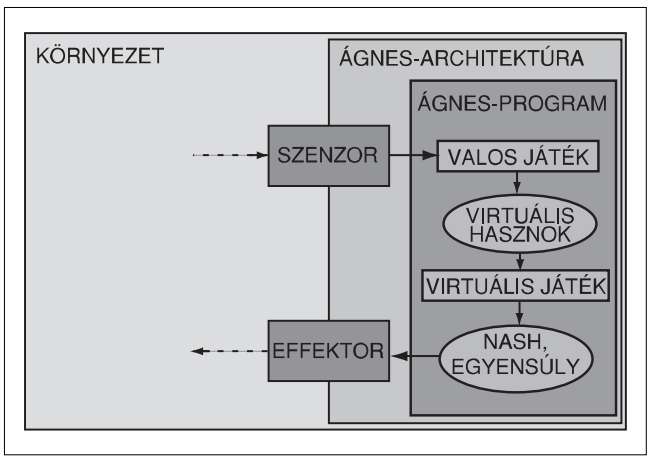
Az ágensek felépítése ekkor legyen a következő: legyen adott egy *architektúrájuk*, és egy *programjuk*, ahol az architektúra legyen felelős a program futtatásáért, a program pedig a játék belső reprezentációja alapján válassza ki az ágens által játszott stratégiát. Az ágensek programjának modellje ekkor legyen a követ-

kező: minden játékosnak legyen egy *virtuális haszonfüggvénye,* amely a játék belső reprezentációjában lehetséges összes stratégia-kombinációhoz egy-egy virtuális haszonértéket rendel. Ekkor a játékos programja felfogható úgy, mint ami éppen azt a stratégiát választja ki a játékos számára, amit a virtuális játék (melyben a játékosok stratégia-kombinációkhoz tartozó haszna az adott játékos által részükre feltételezett virtuális haszonfüggvényük által adott) valamely Nash-egyensúlya ír elő (az a stratégia-kombináció, amelytől egyik játékosnak se éri meg egyoldalúan eltérni). Ezt a döntési mechanizmust szemlélteti *3. ábra:*

Az előbbiek alapján jól látható, hogy az ágensek működését (programját) sikerült explicit módon modellezni. Ebből következően a Tervező immár decentralizált, lokális környezeti hozzáférésű mechanizmus formájában implementálhat egy-egy KDSZ-t azáltal, hogy megadja az ágensek ehhez szükséges architektúráját, és programját (azaz lényegében a virtuális haszonfüggvényeiket). Bizonyítást nyert, hogy teljes információs játékoknak megfelelő környezetekben tetszőleges KDF, megkötés nélkül, egzakt módon implementálható bináris virtuális haszonfüggvények felhasználásával [9]. A bizonyítás konstruktív, így tehát adott ágens architektúrák mellett megadja azokat a virtuális haszonfüggvényeket, melyek mellett a fentebb leírt ágens-működés közösségi szinten éppen egy tetszőleges választott KDF-et implementál.

Az új megközelítés hátrányának tekinthető, hogy viszonylag magas absztrakciós szinten modellezi az ágensek működését, s így még további kutatás szükségeltetik ahhoz, hogy egy-egy konkrétan adott ágens-architektúrának (pl. JADE) is megfeleltethető legyen. Előnye viszont, hogy közvetlenül, egzakt módon, megkötés nélkül tervezhetünk általa ágens-közösségeket. Ennek következtében bizonyítható módon válik implementálhatóvá az „optimális" közösségi működés (például Pareto optimális – azaz ha nincs a közösségnek olyan része, amely jobban jár akkor, ha eltér stratégiájától, miközben a többiek egyike se jár rosszabbul –, vagy korlátosan optimális). Érdekes, és fontos ágens-társadalmi jelenségek is modellezhetővé válnak továbbá.

3. ábra Ágensek működésének újfajta megközelítése



Például az ágensek közti kooperáció (ahol a kooperáló ágenseknek azonos a virtuális haszonfüggvénye), vagy éppen az áldozathozatal jelensége (ahol az „áldozatkész" ágens virtuális haszna ott magas, ahol valós haszna alacsony) stb. Ezen felül a játékelméletben már jól ismert típus-központú megközelítés [10] felhasználásával (ahol a játékosok típusa most az ágens architektúrájának, és programjának együtteseként értendő) kezelhetővé válik a nem teljes információs játékoknak megfelelő környezetek esete is.

4. Összefoglalás

A cikkben bemutatott virtuális haszon alapú döntéshozási elv lehetővé tette, hogy létrehozzuk az ágensek, és az ágensközösségek működésének egy olyan absztrakt, magas-szintű modelljét, amely segítségével az eddigieknél sokkalta hatékonyabban válik megoldhatóvá a közösségi döntések implementációjának problémája. A további kutatás az említett modell már meglévő, alacsonyabb szintű ágens-modellekkel való összekapcsolását; a nem teljes információs játékoknak megfelelő probléma-környezetek vizsgálatát; és az elképzelés – intelligens rendszerek egységes tervezésére, és elemzésére irányuló – átfogó rendszerspecifikációs elvbe történő integrációját tűzi ki célul.

Irodalom

[1] D. L. Kovács:
"Intelligens rendszerek egységes tervezése,"
Híradástechnika, 2004/10. pp.29–38.
[2] S. Russell, D. Subramanian:
"Provably bounded-optimal agents,"
Journal of AI Research, Nr.2,1995, pp.1–36.
[3] R. Serrano: "The Theory of Implementation of Social Choice Rules," SIAM Review, 46:377-414, 2004.
[4] S. Russell, P. Norvig: Artificial Intelligence:
A Modern Approach, Prentice Hall, 1995.
[5] J. von Neumann, O. Morgenstern:
Theory of games and economic behavior,
Princeton University Press, 1947.
[6] M. Bowling, R. Jensen, M. Veloso:
"A Formalization of Equilibria for Multiagent planning,"
in Proc. of IJCAI'03 Worskshop, August 2003.
[7] J. F. Nash: "Non-cooperative games,"
Annals of Mathematics, 54(2), 1951. pp.286–295.
[8] D. Abreu, A. Sen:
"Virtual Implementation in Nash Equilibrium,"
Econometrica, Nr.59, 1991. pp.997–1021.
[9] D. L. Kovács:
A general model to strategy selection in games,
Technical report, BUTE-DMIS, Hungary, May 2004.
[10] J. C. Harsányi: "Games with incomplete information played by Bayesian players I-II-III,"
Management Science,14. pp.159–182; 320–334;
486–502, 1967–1968.