# ارائه یک مدل هم‌تکاملی رقابتی جدید برای یادگیری بازی‌ها

احمد نیک‌آبادی*، رضا صفابخش†

## چکیده

**یادگیری بازی‌ها از جمله مسائل مورد توجه و پرکاربرد در هوش مصنوعی است. تا کنون روش‌های مختلفی برای یادگیری بسیاری از بازی‌ها ارائه شده است. این روش‌ها بر اساس ویژگی‌های مختلفی مورد بررسی قرار می‌گیرند. یکی از این ویژگی‌ها میزان استفاده از دانش فرد خبره در زمینه بازی مورد نظر است. یکی از روش‌های موفق یادگیری بازی که از هیچ دانش بشری در زمینه بازی مورد نظر استفاده نمی‌کند روش هم‌تکاملی است. در این روش بازیکنان در ضمن تکامل برای یادگیری بازی از هیچ تابع ارزیابی شایستگی خارجی استفاده نمی‌کنند و صرفاً نتایج بازی‌های انجام شده در مقابل یکدیگر برای ارزیابی بازیکنان استفاده می‌شود. این روش تا کنون در یادگیری بازی‌های بسیاری همچون چکرز ، اتلو و امثال آن با موفقیت مورد استفاده قرار گرفته است. الگوریتم هم‌تکاملی مورد استفاده در یادگیری این بازی‌ها دارای معایبی است که باعث می‌شود این الگوریتم قادر به یادگیری بازی ساده‌ای همانند بازی OX نباشد. این مشکلات در این مقاله به طور دقیق مورد بررسی قرار گرفته و چارچوب هم‌تکاملی رقابتی جدیدی برای رفع آنها ارائه می‌شود. در چارچوب پیشنهادی برای یادگیری نقش دو بازیکن متفاوت دو جمعیت مجزا در نظر گرفته شده و برای ارزیابی هر جمعیت، جمعیت ارزیابی متشکل از تعدادی از افراد جمعیت مقابل تشکیل می‌شود. نتایج تجربی بدست آمده نشان دهنده آنست که الگوریتم جدید به خوبی قادر به رفع مشکلات برشمرده شده می‌باشد.**

## کلمات کلیدی

**یادگیری بازی، هم‌تکاملی رقابتی، شبکه عصبی، بازی OX، استراتژی تکامل.**

# A New Competitive Co-evolutionary Model for the Learning of Games

Ahmad Nickabadi, Reza Safabakhsh
Amirkabir University of Technology

## Abstract

Game learning is one of the most popular problems in artificial intelligence studies. So far, a lot of techniques have been proposed for game learning goal. These techniques are examined according to different criteria. One of these criteria is the amount of expert knowledge used in the learning process of the game. Co-evolutionary method is a successful technique that doesn't use any initial expert knowledge. In this method during the evolving of players to learn the game, no external fitness evaluation function is used and just the results of direct competitions between the players is used in the evaluation. Until now, this method has been employed to learn many games such as Checkers and Go. This co-evolutionary algorithm has some drawbacks that make it impossible for this algorithm to learn simple games such as OX. These drawbacks have been studied in this paper and a new competitive co-evolutionary algorithm has been proposed to solve them. In this algorithm, two distinct populations are developed for learning two different roles involved in the game. For evaluation of each population, an evaluator population consisting of some individuals from the other population are selected. The experimental results show that the new algorithm can properly resolve the mentioned problems.

## Keywords

Game Learning, Competitive Co-evolutionary, Neural Networks, OX Game, Evolutionary Strategy.

* دانشجوی کارشناسی ارشد هوش ماشین و رباتیک، دانشگاه صنعتی امیرکبیر، دانشکده مهندسی کامپیوتر و فناوری اطلاعات،nikabadi@ce.aut.ac.ir.
† استاد تمام و عضو هیأت علمی دانشگاه، دانشگاه صنعتی امیرکبیر، دانشکده مهندسی کامپیوتر و فناوری اطلاعات،safa@ce.aut.ac.ir.

## ۱- مقدمه

مسأله یادگیری بازی‌ها به عنوان یکی از مسائل پرکاربرد در هوش مصنوعی در سالیان گذشته توجه بسیاری از محققان را به خود جلب کرده و حجم بزرگی از تحقیقات انجام شده در زمینه هوش مصنوعی را به خود اختصاص داده است. تا کنون روش‌ها و تکنیک‌های مختلفی برای این منظور مورد استفاده قرار گرفته و بازیکنان هوشمند بسیاری برای بازی‌های مختلف آموزش داده شده است که برخی از آنان به سطوح بسیار بالایی از مهارت در بازی مورد نظر دست یافته‌اند. در یک دسته‌بندی کلی روش‌های یادگیری بازی‌ها را می‌توان به دو دسته عمده تقسیم کرد. دسته‌ای که از دانش افراد خبره در زمینه بازی استفاده می‌کنند و دسته دیگری که از هیچ‌گونه دانش اولیه‌ای در زمینه نحوه بازی استفاده نمی‌کنند. البته مرزبندی دقیقی بین روش‌های ارائه شده برای یادگیری بازی‌ها نمی‌توان انجام داد. به عنوان نمونه‌ای از روش‌های یادگیری بازی که تا حد زیادی از دانش بشری استفاده می‌کند می‌توان به Logistello که یکی از قویترین بازیکنان بازی اتلو است اشاره کرد[1,2]. یکی از مشکلات عمده این دسته از روش‌ها اینست که دانش بازیکن هوشمند به مهارت افراد خبره‌ای که از بازی‌های آنان برای آموزش استفاده شده است محدود می‌شود و دانش جدیدی تولید نمی‌شود.

دسته دوم روش‌هایی هستند که در آنها از هیچ گونه دانش فرد خبره‌ای استفاده نمی‌شود. در این حالت معمولاً بازیکنان هوشمند با بازی مقابل یکدیگر به سطوح بالای مهارت در یک بازی دست پیدا می‌کنند. به عنوان نمونه کاملاً موفقی از این نوع روش‌های یادگیری می‌توان به TD-Gammon اشاره کرد[3].

الگوریتم‌های تکاملی در بسیاری از مسائل بهینه‌سازی با موفقیت مورد استفاده قرار گرفتند. آقای فوگل در روش پیشنهادی خود برای یادگیری بازی‌ها بدون استفاده از دانش بشری از یک الگوریتم تکاملی استفاده کرد[4]. یادگیری بازی با استفاده از الگوریتم‌های تکاملی و استفاده از نتایج بازی بازیکنان در مقابل یکدیگر به عنوان معیار ارزیابی شایستگی را "هم‌تکاملی" می‌نامند. روش‌های هم‌تکاملی خود به دو دسته هم‌تکاملی رقابتی و هم‌تکاملی همکاری‌کننده تقسیم می‌شوند. در هم‌تکاملی همکاری‌کننده هر یک از دو بازیکن (موجود مورد ارزیابی در الگوریتم تکاملی) سعی در افزایش مجموع امتیاز کسب شده توسط دو بازیکن دارد در حالی که در هم‌تکاملی رقابتی امتیازی که به طرفین داده می‌شود رابطه عکس با یکدیگر دارد و لذا هر فرد سعی در افزایش امتیاز خود و کاهش امتیاز حریف دارد.

دراین مقاله نشان داده می‌شود که علیرغم موفقیت روش هم‌تکاملی ارائه شده توسط آقای فوگل در یادگیری بازی‌های پیچیده‌ای همانند چکرز [4]و اتلو[5]، این روش قادر به یادگیری بازی ساده‌ای همانند بازی OX نیست. این امر به دلیل برخی ویژگی‌های خاص این بازی است که در کارهای انجام گرفته تا کنون کمتر مورد توجه قرار گرفته است.

در ادامه این مقاله ابتدا به معرفی بازی OX می‌پردازیم. در بخش ۳ روش هم‌تکاملی متداول بررسی شده و مشکلات این روش در یادگیری بازی OX در بخش ۴ مطرح می‌شود. سپس در بخش ۵ چارچوب هم‌تکاملی جدید ارائه شده و در نهایت نتایج تجربی حاصل از این روش در بخش ۶ مورد بررسی قرار می‌گیرد.

## ۲- معرفی بازی OX

بازی دوز سه در سه که با نام OX نیز شناخته می‌شود، مثال ساده‌ای از یک بازی مجموع صفر است که اطلاعات کامل بازی در اختیار بازیکنان قرار دارد. در این بازی دو بازیکن و یک صفحه سه در سه وجود دارد. در ابتدای کار صفحه خالی است. هر بازیکن در هر حرکت علامتی در یکی از خانه‌های خالی قرار می‌دهد. مرسوم است که ابتدا بازیکنی که از علامت X استفاده می‌کند، بازی را شروع می‌کند. هدف بازی قرار دادن سه علامت یکسان در یک سطر، ستون و یا قطر است. این امر منجر به برد برای بازیکن صاحب علامت و باخت برای بازیکن دیگر می‌شود.

## ۳- استراتژی‌های تکامل

استراتژی‌های تکامل دسته خاصی از الگوریتم‌های تکاملی هستند. این استراتژی‌ها اولین بار در دهه ۱۹۷۰ ارائه شدند[6]. از این دسته الگوریتم‌های تکاملی معمولاً در کارهای بهینه‌سازی عددی استفاده می‌شود. از ویژگی‌های این دسته از الگوریتم‌های تکاملی می‌توان به سرعت بالای آنان، بهینه‌سازی مناسب برای مقادیر حقیقی و زمینه تئوری بیشتر نسبت به سایر الگوریتم‌های تکاملی اشاره کرد.

استراتژی‌های تکامل نیز همانند سایر الگوریتم‌های تکاملی از یک چرخه تکامل تشکیل می‌شوند. جمعیتی از جواب‌های مورد نظر به صورت تصادفی ایجاد می‌شود. با استفاده از یک تابع ارزیابی شایستگی به هر یک از افراد جمعیت مقداری عددی که بیانگر شایستگی آن فرد است نسبت داده می‌شود. تعدادی از افراد با استفاده از یکی از روش‌های انتخاب والدین برای تولید نسل انتخاب می‌شوند. با استفاده از عملگرهای ژنتیکی همانند جهش و بازترکیبی تعدادی فرزند ایجاد می‌شود. فرزندان نیز با استفاده از تابع ارزیابی شایستگی ارزیابی شده و از میان این فرزندان و جمعیت قبلی، جمعیت تشکیل دهنده نسل بعد انتخاب می‌شوند.

معمولاً هر فرد در استراتژی‌های تکامل (یک کروموزوم) با بردار زیر ارائه می‌شود:

$< X_1, X_2, ..., X_n, \sigma_1, \sigma_2, ..., \sigma_{n\sigma} >$ که در آن $< X_1, X_2, ..., X_n >$ بردار اعداد حقیقی است که استراتژی تکاملی سعی در یافتن مقدار بهینه آن دارد. مقادیر $\sigma$ و X را با استفاده از روابط زیر جهش می‌دهیم:

$$\sigma_i' = \sigma_i.\exp(\tau'.N(0,1) + \tau.N_i(0,1))$$
$$X_i' = X_i + \sigma_i'.N_i(0,1) \qquad (۱)$$

در این روش از دو نرخ یادگیری $\tau$ و $\tau'$ استفاده می‌شود. $\tau$ نرخ یادگیری سراسری و مقدار آن متناسب با $1/\sqrt{2n}$ و $\tau'$ نرخ یادگیری مربوط به هر یک از مختصه‌های X و مقدار آن متناسب با $1/\sqrt{2\sqrt{n}}$ است. N(0,1) نیز بیانگر یک عدد تصادفی حاصل از توزیع نرمال با میانگین 0 و واریانس 1 است.

## ۴- مشکلات چارچوب تکامل فعلی

در چارچوب هم‌تکاملی که در بخش قبل به آن اشاره شد و در بسیاری از کارهای مهم انجام شده در زمینه یادگیری بازی‌ها همانند [4] مورد استفاده قرار گرفته است، برای ارزیابی میزان شایستگی (قدرت) هر یک از اعضاء جمعیت (بازیکنان تکامل یافته)، تورنمنتی از مسابقات میان آن بازیکن و تعدادی از بازیکنان دیگر هم نسل انجام می‌شود. از امتیازات بدست آمده توسط هر یک از افراد به عنوان میزان شایستگی آن در قیاس با سایر بازیکنان استفاده می‌شود. این روش علیرغم مزایایی که دارد، دارای معایبی نیز هست بطوریکه حتی در برخی بازی‌ها استفاده از آن برای یادگیری سطح قابل قبولی از بازی ممکن نیست. به عنوان مثالی از این روش، دقیقاً با همان ساختاری که دربخش پیش معرفی شد، برای یادگیری بازی OX استفاده شد. نتایج حاصل در بخش ۶ نشان داده شده است. عملاً هیچ گونه یادگیری در این حالت مشاهده نمی‌شود. بررسی دقیقتر رفتار این الگوریتم تکاملی در مورد بازی OX نشان داد که علت عدم توانایی یادگیری این الگوریتم مشکلات این روش خصوصاً در مواجهه با بازی‌هایی نظیر OX است. در ادامه به برخی از این مشکلات اشاره می‌کنیم.

### ۴-۱- تداخل مکانی در شبکه عصبی

یادگیری بازی OX را می‌توان به یادگیری دو استراتژی عمده این بازی تقسیم کرد. استراتژی بازی در نقش بازیکن اول و استراتژی بازی در نقش بازیکن دوم. در واقع بازیکن اول با امکان برد بسیار بیشتر نسبت به بازیکن دوم نوعی بازی تهاجمی و بازیکن دوم در مقابل این بازی نوعی بازی تدافعی ارائه می‌کنند. هنگامی که تنها از یک شبکه عصبی برای یادگیری بازی OX استفاده می‌شود، این شبکه باید به طور همزمان این دو کار را یاد بگیرد. این پدیده باعث ایجاد تداخل مکانی در شبکه عصبی شده و سرعت یادگیری بازی را کاهش می‌دهد.

### ۴-۲- یادگیری بیش از حد سریع یکی از نقش‌ها

در الگوریتم‌های هم‌تکاملی رقابتی برای تعیین میزان شایستگی هر یک از افراد جمعیت، تعدادی از افراد جمعیت مقابل با استفاده از مکانیسمی (مثلاً به صورت تصادفی) انتخاب و فرد مورد نظر در مقابل هر یک از این افراد بازی کرده و امتیاز بدست آمده توسط هر شخص به عنوان معیاری برای ارزیابی آن شخص در نظر گرفته می‌شود. زمانی که یکی از نقش‌های حاضر در تکامل بسیار سریع رشد می‌کند، عملاً امکان پیشرفت جمعیت مقابل را از بین می‌برد. در این حالت به دلیل ضعیف بودن حریف، خود جمعیت اول نیز پیشرفت چندانی نخواهد داشت.

### ۴-۳- عدم کارآیی جمعیت ارزیاب

مشکل دیگری که در الگوریتم هم‌تکاملی موجود وجود دارد عدم توجه به استفاده از یک جمعیت ارزیاب کارآست. فرد کارآ در جمعیت ارزیاب کسی است که در مقابل تعدادی از افراد جمعیت برنده و در مقابل تعداد دیگری از آنان بازنده شود. در صورتی که فردی همه افراد جمعیت را شکست دهد یا از همه آنها شکست بخورد فرد مؤثری در ارزیابی نیست.

### ۴-۴- ارزیابی تنها در مقابل بازیکنان قوی

در الگوریتم‌های تکاملی پس از چند تولید نسل شایستگی همه افراد جمعیت افزایش می‌یابد. در الگوریتم‌های هم‌تکاملی این عامل باعث می‌شود که در تولید نسل‌های بعدی افراد جمعیت تنها در برابر افراد قوی ارزیابی شوند و ممکن است توانایی بازی در مقابل افراد ضعیف را نداشته باشند.

### ۴-۵- افراد یکسان در جمعیت ارزیاب

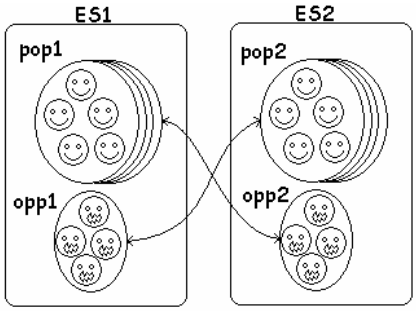
وجود افراد مشابه در جمعیت ارزیاب نیز سبب می‌شود که بازیکنان تنها در مقابل یک نوع بازی ارزیابی شوند.

در بخش بعد روش پیشنهادی این مقاله برای حل کلیه مشکلات فوق بیان می‌شود.

## ۵- چارچوب هم‌تکاملی رقابتی پیشنهادی

در این مقاله از یک شبکه عصبی MLP برای یادگیری بازی استفاده می‌شود. شبکه مورد نظر دارای ۹ نورون در لایه ورودی ۸ نورون در لایه مخفی و ۹ نورون در لایه خروجی است. هر یک از نورون‌های لایه ورودی و خروجی متناظر یکی از خانه‌های صفحه بازی OX است. تعداد پارامترهای این شبکه ۱۶۱ است که با استفاده از الگوریتم تکاملی یاد گرفته می‌شود.

شمای کلی الگوریتم هم‌تکاملی رقابتی پیشنهادی این مقاله در شکل (۴) نشان داده شده است. در چارچوب پیشنهادی دو جمعیت مستقل pop1 و pop2 به ترتیب برای یادگیری بازی در نقش بازیکن اول و دوم توسعه داده می‌شوند. هر یک از این دو جمعیت به صورت مجزا تکامل می‌یابد. برای ارزیابی بازیکنان نقش اول به بازیکنی که در نقش بازیکن دوم بازی کند نیاز داریم و در مورد جمعیت بازیکنان نقش دوم برعکس. برای این کار هر یک از جمعیت‌ها می‌توانند از بازیکنان جمعیت دیگر به عنوان رقیب استفاده کنند.

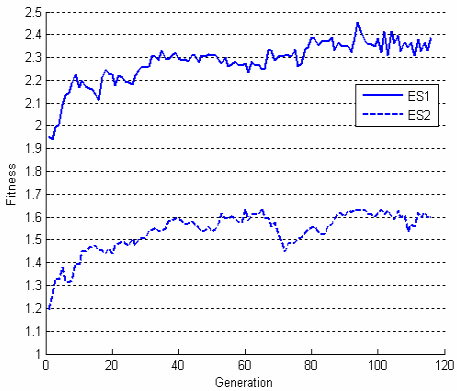



شکل (۱) چارچوب تکاملی پیشنهادی برای یادگیری بازی OX. در این چارچوب دو استراتژی تکامل ES1 و ES2 به صورت موازی اجرا می‌شوند. هر استراتژی تکامل دارای جمعیت تکامل یابنده(pop) و جمعیت ارزیاب (opp) مخصوص است. هر جمعیت ارزیاب برای محاسبه میزان شایستگی جمعیت تکامل یابنده مقابل استفاده و متناسب با آن تکامل می‌یابد.

در این روش به جای آنکه اساس کار ارزیابی (جمعیت ارزیاب) بر مبنای جمعیت رقیب فعلی باشد، جمعیت ارزیاب به صورت مؤثر و پویا ایجاد می‌شود. مؤثر از آن جهت که عناصری که تأثیری در ایجاد تمایز در بازیکنان جمعیت اصلی ندارند از آن حذف شده، و پویا از آن جهت که متناسب با سطح بازی جمعیت اصلی، جمعیت ارزیاب نیز تغییر می‌یابد. برای این منظور به جای اینکه از هر یک از جمعیت‌های تکامل یابنده به عنوان جمعیت ارزیاب جمعیت دیگر استفاده شود در کنار هر جمعیت، جمعیت ارزیابی نیز تشکیل می‌شود. این جمعیت‌ها در شکل (۴) با عنوان opp1 و opp2 نشان داده شده‌اند. اعضاء هر یک از این جمعیت‌های ارزیاب از جمعیت‌های تکامل یابنده نظیر آنها تأمین می‌شود. در مرحله مقداردهی اولیه جمعیت‌ها، تعدادی از افراد جمعیت اصلی به تصادف انتخاب و در جمعیت ارزیاب قرار می‌گیرند. سپس از این جمعیت ارزیاب برای ارزیابی جمعیت اصلی مقابل استفاده می‌شود. در هر تولید نسل امتیازاتی که هر یک از افراد جمعیت ارزیاب بدست آورده است مورد بررسی قرار می‌گیرد. در صورتی که فردی از این جمعیت قادر به کسب هیچ امتیازی نشده باشد و یا امتیاز کامل را کسب کرده باشد، به دلیلی که قبلاً گفته شد فرد مناسبی برای ارزیابی نیست. در این حالت فرد مذکور از جمعیت ارزیابی حذف شده و فرد دیگری جای آن را می‌گیرد. به جای بازیکنان قوی حذف شده بازیکنان ضعیف و به جای بازیکنان ضعیف حذف شده بازیکنان قوی جایگزین می‌شوند. اگر دو بازیکن در جمعیت ارزیاب در چند تولید نسل متوالی امتیاز یکسانی در مقابل جمعیت تکامل یابنده کسب کنند، به عنوان بازیکنان ارائه دهنده بازی یکسان شناسایی و از جمعیت ارزیاب حذف و بازیکن دیگری به تصادف از مخزن بازیکن جایگزین آن می‌شود.

## ۶- نتایج عملی

نتایج استفاده از مدل هم‌تکاملی ارائه شده برای یادگیری بازی OX در شکل (۲) نشان داده شده است. همگرایی الگوریتم و افزایش سطح مهارت بازیکنان در این حالت نسبت به الگوریتم هم‌تکاملی معمول که اصولاً هیچ همگرایی ندارد کاملاً قابل توجه است.




شکل (۲) یادگیری بازی OX با استفاده از چارچوب پیشنهادی.

## مراجع

[1] Buro, M., “Statistical feature combination for the Evaluation of game Positions,” jurnal of Artificial Intelligence Research, vol. 3, pp. 373-382, 1995.
[2] Buro, M., “Improving heuristic mini-max search by supervised learning,” Artificial intelligence, vol 134, pp. 85-99, 2002.
[3] Tesauro, G., “Practical issues in temporal difference learning,” Machine Learning, vol. 8, pp. 257-277, 1992.
[4] Chellapilla, K., and Fogel, D.B., "Evolving Neural Networks to Play Checkers without Expert Knowledge," IEEE Transactions On Neural Networks, Vol. 10, pp. 1382-1391, 1999.
[5] Chong, S. Y., Tan, M. K., and White, J. D., “Observing the Evolution of Neural Networks Learning to Play the Game of Othello,” IEEE Transactions On Evolutionary Computiation, vol. 9, no. 3, 2005.
[6] Back, T., Hammel, U., and Schwefel, H.P., “Evolutionary Computation: Comments on the history and current state,” IEEE Transaction on Evolutionary computing, vol. 1, no. 1, pp. 3-17, Apr, 1996.

## زیرنویس‌ها