



بسم الله الرحمن الرحيم



دانشگاه صنعتی اصفهان

دانشکده مهندسی برق و کامپیوتر

بهبود کیفیت و سرعت یادگیری در سیستم‌های چندعامله با استفاده از

ماتریس ارجاع و انتگرال فازی

پایان‌نامه کارشناسی ارشد مهندسی کامپیوتر – هوش مصنوعی و رباتیک

داریوش حسن‌پورآده

استاد راهنما

دکتر مازیار پالهنک

۱۳۹۵



دانشگاه صنعتی اصفهان

دانشکده مهندسی برق و کامپیوتر

پایان نامه کارشناسی ارشد رشته مهندسی کامپیوتر – هوش مصنوعی و رباتیک آقای

داریوش حسن پور آده

تحت عنوان

بهبود کیفیت و سرعت یادگیری در سیستم‌های چندعامله با استفاده از

ماتریس ارجاع و انتگرال فازی

در تاریخ ... توسط کمیته تخصصی زیر مورد بررسی و تصویب نهایی قرار گرفت:

دکتر مازیار پالهننگ

۱- استاد راهنمای پایان نامه

دکتر ...

۳- استاد داور (اختیاری)

دکتر ...

۴- استاد داور (اختیاری)

دکتر محمد رضا تابان

سرپرست تحصیلات تکمیلی دانشکده

**تشکر و قدردانی**

پروردگار منّان را سپاسگزارم .....

کلیه حقوق مادی مترتب بر نتایج مطالعات،  
ابتکارات و نوآوری‌های ناشی از تحقیق  
موضوع این پایان‌نامه متعلق به دانشگاه  
صنعتی اصفهان است.

دلتنگی های آدمی را باد ترانه ای می خواند  
رویا هایش را آسمان پر ستاره نادیده می گیرد  
و هر دانه ی برفی به اشکی نریخته می ماند.  
سکوت سرشار از سخنان ناگفته است؛  
از حرکات ناکرده،  
اعتراف به عشق های نهان،  
و شگفتی های به زبان نیامده،  
در این سکوت حقیقت ما نهفته است؛  
حقیقت تو و من.

برای تو و خویش  
چشمانی آرزو می کنم،  
که چراغ ها و نشانه ها را در ظلمات مان ببیند.  
گوشی،  
که صداها و شناسه ها را در بیهوشی مان بشنود.  
برای تو و خویش،  
روحي،  
که این همه را در خود گیرد و بپذیرد.  
و زبانی  
که در صداقت خود ما را از خاموشی خویش بیرون کشد،  
و بگذارد از آن چیزها که در بندهمان کشیده است، سخن بگوییم.

پنجه درافکنده ایم با دست هایمان  
به جای رها شدن  
سنگین سنگین بر دوش می کشیم  
بار دیگران را  
به جای همراهی کردن شان!  
عشق ما نیازمند رهایی است نه تصاحب  
در راه خویش ایثار باید نه انجام وظیفه...

بی اعتمادی دری است  
خودستایی، چفت و بست غرور است  
و تهی دستی، دیوار است و لولا است  
زندانی را که در آن محبوس رای خویش ایم  
دلتنگی مان را برای آزادی و دلخواه دیگران بودن  
از رخنه هایش تنفس می کنیم...

# فهرست مطالب

صفحه	عنوان
هشت	فهرست مطالب
نه	فهرست تصاویر
۱	چکیده
۲	فصل اول: مفاهیم علمی پیش نیاز پایان نامه
۲	۱-۱ اندازه گیری و انتگرال فازی
۵	فصل دوم: روش پیشنهادی
۵	۱-۲ مقدمه
۶	۲-۲ معیار خبرگی - ماتریس ارجاع و خاطره
۹	۳-۲ یادگیری مشارکتی $Q$ با استفاده از ماتریس ارجاع و انتگرال فازی
۱۰	۱-۳-۲ الگوریتم پیشنهادی
۱۲	۲-۳-۲ تعیین توابع $f(\cdot)$ و $g(\cdot)$ در انتگرال فازی چوکت
۱۴	۴-۲ علت کارکرد انتگرال فازی چوکت در انتقال دانش
۱۶	فصل سوم: نتایج عملی
۱۶	۱-۳ مقدمه
۱۷	۲-۳ رفتار الگوریتم های معرفی شده برای $g(\cdot)$
۱۸	۱-۲-۳ تعابیر مختلف انتگرال فازی چوکت از داده ها بر مبنای $g(\cdot)$
۱۹	۳-۳ مقایسه ی روش پیشنهادی با روش کوتاه ترین مسیر تجربه شده
۲۱	۱-۳-۳ مقایسه در محیط پلکان مارپیچ
۲۹	مراجع
۳۲	چکیده انگلیسی



## فهرست تصاویر

- ۱-۳ دو توزیع فرضی بجهت نمایش نحوه رفتار الگوریتم‌های ۴ تا ۷ بروی آن‌ها. . . . . ۱۷
- ۲-۳ نمایش توزیع‌های جدید بدست آمده بعد از اعمال الگوریتم‌های ۴ تا ۷ بروی دو توزیع فرضی شکل ۱-۳. . . . . ۱۸
- ۳-۳ نمایش رفتار انتگرال فازی بروی منابع اطلاعاتی  $y = 1$  و  $y = 2$  و  $y = 3$  به ازای توابع  $g(\cdot)$ های مختلف. . . . . ۱۹
- ۴-۳ مقایسه در سرعت و کیفیت یادگیری در محیط پلکان مارپیچ با تابع بولتزمن با تابع بولتزمن در محیط پلکان مارپیچ ۲۲
- ۵-۳ مقایسه در پیچیدگی زمانی روش‌ها به ازای تعداد تلاش‌های متفاوت برحسب میلی ثانیه با تابع بولتزمن در محیط پلکان مارپیچ . . . . . ۲۳
- ۶-۳ نمودار باروری الگوریتم‌ها مختلف با تابع بولتزمن در محیط پلکان مارپیچ . . . . . ۲۴
- ۷-۳ مقایسه تاثیر تعداد عامل‌ها میزان کیفیت و سرعت یادگیری با تابع بولتزمن در محیط پلکان مارپیچ . . . . . ۲۶
- ۸-۳ مقایسه در سرعت و کیفیت یادگیری در محیط پلکان مارپیچ با تابع حریصانه در محیط پلکان مارپیچ . . . . . ۲۷
- ۹-۳ نمودار باروری الگوریتم‌ها مختلف با تابع حریصانه در محیط پلکان مارپیچ . . . . . ۲۸
- ۱۰-۳ مقایسه تاثیر تعداد عامل‌ها میزان کیفیت و سرعت یادگیری با تابع حریصانه در محیط پلکان مارپیچ . . . . . ۲۹

## چکیده

واژه‌های کلیدی: ۱- سیستم‌های چندعامله، ۲- یادگیری مشارکتی، ۳- یادگیری تقویتی، ۴- دانش غیرافزایشی، ۵- انتگرال فازی.

## فصل اول

### مفاهیم علمی پیش نیاز پایان نامه

#### ۱-۱ اندازه گیری و انتگرال فازی

برای درک روش پیشنهادی نیاز به داشتن اطلاعات پایه در مورد اندازه گیری های فازی<sup>۱</sup> و انتگرال فازی داریم که با هدف جمع آوری اطلاعات<sup>۲</sup> ارائه شده اند. اندازه گیری های فازی پیش زمینه ای بر انتگرال های فازی هستند که قبل از آنکه آشنایی با انتگرال های فازی نیاز به معرفی اندازه گیری های فازی داریم. اگر فرض کنیم که تعداد منبع اطلاعاتی  $X = \{x_1, x_2, \dots, x_n\}$  که این منابع اطلاعاتی اطلاعات دریافتی از سنسورها، پاسخ های داده شده به یک پرسشنامه و غیره باشند. اندازه گیری فازی میزان ارزش اطلاعاتی این منابع را در اختیار ما می گذارد. معمولاً اندازه گیری فازی توسط تابع  $g : 2^{|X|} \rightarrow [0, 1]$  تعریف می شود که ورودی آن یک زیر مجموعه ای از منابع اطلاعاتی می باشد و خروجی آن یک مقدار مابین صفر و یک که میزان ارزش اطلاعاتی که آن زیر مجموعه از منابع اطلاعاتی ورودی تابع را مشخص می کند.

این تابع باید دارای شرایط مرزی تعریف شده و یکنوختی باشد که در ادامه به معرفی شرایط می پردازیم [۱]:

---

<sup>1</sup>Fuzzy measures

<sup>2</sup>Aggregate Information

۱. شرایط مرزی: اگر اطلاعاتی در دست نداریم ارزش صفر را دارد و کلیه اطلاعاتی حداکثر ارزش ۱ را دارد.

$$g(\emptyset) = 0, \quad g(X) = 1 \quad (1-1)$$

۲. یکنواختی - غیر کاهشی: اگر اطلاعات بیشتری به دست آمد ارزش کلیه اطلاعات که شامل اطلاعات جدید می باشد حداقل به اندازه زمانی است که آن اطلاعات جدید بدست نیامده است.

$$A \subseteq B \subseteq X \Rightarrow g(A) \leq g(B) \leq 1 \quad (2-1)$$

مقادیر تابع  $g$  یا توسط کارشناس ارائه می شود یا توسط یک تابعی مدل می شود، یکی از توابع معروف برای تخمین مقادیر تابع  $g$  تابع اندازه گیری- $\lambda$  سوگنو<sup>۱</sup> می باشد که به صورت زیر تعریف می شود [۲].

$$g(\{x_1, \dots, x_l\}) = \frac{1}{\lambda} \left[ \prod_{i=1}^l (1 + \lambda g_i) - 1 \right] \quad (3-1)$$

که در معادله ۱-۳ مقدار  $g_i$  ها مقادیر ارزش هریک از منابع اطلاعاتی است و  $\lambda$  بگونه ای تعیین می گردد که  $g_\lambda(X) = 1$  شود که این مقدار برابر با جواب معادله ی زیر باشد.

$$\lambda + 1 = \prod_{i=1}^n (1 + \lambda g_i), \quad \lambda \in (-1, \infty) \quad (4-1)$$

نکته ای که در رابطه با تابع اندازه گیری- $\lambda$  سوگنو باید توجه کرد این است که به ازای مقادیر  $n$  مختلف باید ریشه یابی بروی متغیر  $\lambda$  صورت گیرد؛ این ویژگی باعث می شود که این تابع در بعضی از کاربردها کارایی نداشته باشد.

انتگرال فازی در واقع یک تعمیمی به روش میانگین وزنی<sup>۲</sup> می باشد بطوری که نه تنها مشخصه های مهم تک تک ویژگی ها را در نظر می گیرد بلکه اطلاعات تعاملات بین ویژگی ها را نیز در نظر می گیرد [۳]. از میان انتگرال های فازی دو انتگرال سوگنو<sup>۳</sup> و چوکت<sup>۴</sup> از الگوریتم هایی هستند که می توانند بروی هر اندازه گیری فازی مورد استفاده واقع شود [۴]. فرض کنیم که تابعی چون  $h: X \rightarrow [0, 1]$  وجود دارد که مقادیر منابع اطلاعاتی را

<sup>1</sup>Sugeno  $\lambda$ -Measure

<sup>2</sup>Weighted Arithmetic Mean

<sup>3</sup>Sugeno

<sup>4</sup>Choquet

به بازه‌ی  $[1, 0]$  نگاشت می‌کند. در واقع  $h$  تابع پشتیبان<sup>۱</sup> منابع اطلاعاتی می‌باشد. انتگرال فازی سوگنو به صورت زیر تعریف می‌شود [۵]:

$$\int_s h \circ g = S_g(h) = \bigvee_{i=1}^n h(x_{\pi_i^s}) \wedge g(A_i^s) \quad (۵-۱)$$

$$h \xrightarrow{\pi^s} h(\pi_1^s) \leq h(\pi_2^s) \leq \dots \leq h(\pi_n^s) \quad (۶-۱)$$

$$A_i^s = \{x_{\pi_i^s}, x_{\pi_2^s}, \dots, x_{\pi_n^s}\} \quad (۷-۱)$$

در انتگرال سوگنو لازم است که مقادیر منابع اطلاعاتی را مرتب کنیم که  $\pi^s$  عملگر جایگشت انتگرال فازی سوگنو می‌باشد. نمادهای  $\vee$  و  $\wedge$  به ترتیب عملگرهای  $\max$  و  $\min$  می‌باشد. انتگرال فازی چوکت به صورت زیر تعریف می‌شود [۶]:

$$\int_c h \circ g = C_g(f) = \sum_{i=1}^n \left( f(x_{\pi_i^c}) - f(x_{\pi_{(i-1)}^c}) \right) \cdot g(A_i^c) \quad (۸-۱)$$

$$f \xrightarrow{\pi^c} f(\pi_1^c) \leq f(\pi_2^c) \leq \dots \leq f(\pi_n^c) \quad (۹-۱)$$

$$A_i^c = \{x_{\pi_i^c}, x_{\pi_2^c}, \dots, x_{\pi_n^c}\} \quad (۱۰-۱)$$

$$\pi_0^c = 0, \quad x_{\pi_0^c} = 0 \quad (۱۱-۱)$$

در رابطه‌ی بالا  $f : X \rightarrow \mathbb{R}$  می‌باشد که از وجه تمایز انتگرال فازی چوکت با سوگنو می‌باشد و  $\pi^c$  عملگر جایگشت انتگرال فازی چوکت می‌باشد. [۶].

انتگرال‌های فازی سوگنو و چوکت در حالت کلی دارای تفاوت‌هایی هستند که از جمله‌ی مهم‌ترین این ویژگی‌ها تفاوت تعریف توابع  $h$  و  $f$  در این انتگرال‌ها می‌باشد که باعث می‌شود انتگرال چوکت برای تبدیل‌های مثبت خطی<sup>۲</sup> مناسب باشد؛ بدین معنی که تجمیع اعداد کاردینال<sup>۳</sup> (که اعداد دارای مفاهیم واقعی هستند) را انتگرال چوکت بهتر مدل می‌کند در حالی انتگرال سوگنو برای اعداد ترتیبی<sup>۴</sup> مناسب است [۷]. به همین علت در این پژوهش انتگرال فازی چوکت مورد استفاده قرار گرفته است زیرا که ورودی انتگرال اعداد کاملاً معنی‌دار می‌باشد و اعمال تابع  $h$  بروی مقادیر منابع اطلاعاتی، معانی آن‌ها را تغییر داده و اطلاعات بدرد نخوری را تولید خواهد کرد.

<sup>۱</sup>Support

<sup>۲</sup>Positive Linear Transformation

<sup>۳</sup>Cardinal Aggregation

<sup>۴</sup>Ordinal Numbers

## فصل دوم

### روش پیشنهادی

#### ۱-۲ مقدمه

در این فصل جزییات روش پیشنهادی به طور مفصل معرفی خواهد شد، روش ارائه شده در حالت کلی از دو قسمت تشکیل شده است؛ اولین و مهم‌ترین قسمت ارائه یک معیار خبرگی جدید به نام معیار خبرگی «ارجاع» که برای هر عامل در هر چرخه یادگیری محاسبه و در یک «ماتریس ارجاع» نگه‌داری می‌شود. دومین قسمت مربوط به ترکیب دانش‌های عامل‌ها هستند که با استفاده از یک مدل انتگرال فازی، صورت می‌گیرد. همانطور که در فصل بعدی نیز نشان داده خواهد شد استفاده از مدل انتگرال فازی به دلیل خواصی مهمی که این مدل دارد باعث می‌شود سرعت و کیفیت یادگیری به طرز چشم‌گیری افزایش یابد. در این فصل ابتدا به معرفی معیار «ارجاع» و دلیل استفاده از آن می‌پردازیم سپس یادگیری مشارکتی چندعامله با استفاده از ماتریس ارجاع و انتگرال فازی معرفی خواهد شد و در نهایت نشان داده خواهد شد که چرا استفاده از انتگرال فازی نتایج بهتری را نسبت به مدل‌های سنتی چون مدل مجموع وزنی<sup>۱</sup> را ارائه می‌دهد.

---

<sup>۱</sup> Weighted Sum

## ۲-۲ معیار خبرگی - ماتریس ارجاع و خاطره

در دنیای واقعی «خبرگی» تعاریف متعددی به خود گرفته است، در روانشناسی خبرگی به معنی عملکرد برتر عامل تلقی می‌شود. در جامعه شناسی خبره به فردی گفتی برچسب خبرگی توسط یک گروهی به فرد زده شده است و آن گروه به توانایی که آن فرد در اختیار دارد علاقه‌مند<sup>۱</sup> است. در فلسفه خبره به فردی گفته می‌شود که دانشی که فرد تازه‌کار در اختیار ندارد را دارا می‌باشد [۸]. اگر تعاریف مختلف «خبرگی» را بررسی کنیم می‌بینیم که همه‌ی تعاریف در واقع تعبیری از میزان کیفیت عملکرد عامل نسبت به دیگر عامل‌ها می‌باشد. این تعبیر کلی از «خبرگی» انگیزه‌ای شد که درصدد معرفی معیاری برآیم که در حالت کلی بتوان به کلیه‌ی تعاریف «خبرگی» قابل تعمیم باشد.

**تئوری ۱-۲.** فرض می‌کنیم عامل  $A$  در محیط  $E$  در پی رسیدن به یک مجموعه اهداف  $G \subseteq \{g_1, g_2, \dots, g_n\}$  می‌باشد. میزان خبرگی عامل رابطه‌ی معکوسی با میزان تلاش عامل برای رسیدن به اهداف تعریف شده خود دارد.

طبق آنچه که در تئوری بالا آورده شده است از بین چند عاملی که در یک محیط و یک مجموعه از اهداف فعالیت می‌کنند، عاملی خبره‌تر است که تلاش کمتری برای رسیدن به آن مجموعه اهداف می‌کند. شاید این مساله در نگاه اول نامتعارف به ذهن برسد ولی در فعالیت‌های روزمره ما انسان‌ها نیز به کرات شاهد این امر می‌باشیم. به عنوان مثال رانندگی دو فرد مبتدی و حرفه‌ای را در نظر بگیریم؛ فرد مبتدی هنگام رانندگی تمام حواس خود را معطوف به رانندگی می‌کند تلاش بسیار زیادی برای کنترل نسبت میزان کلاچ و گاز می‌کند و هنگام رانندگی به طور طبیعی رانندگی نمی‌کند و ... ولی فرد خبره کلیه موارد ذکر شده را بطور خودکار و طبیعی انجام می‌دهد بطوری که انگار رانندگی مانند دیگر رفتارهای طبیعی وی چون نفس کشیدن می‌باشد، که بصورت خودکار صورت می‌پذیرد. از این گونه مثال‌ها از کاربرد تئوری ۱-۲ در زندگی روزمره ما زیاد می‌توان یافت.

توجه شود که در تئوری ۱-۲ عبارت «میزان تلاش» عامل می‌تواند در کاربردهای مختلف تعبیر مختلفی به خود بگیرد، مثلاً در مثال راننده‌ی مبتدی و خبره میزان نسبت مسافت طی شده بر زمان رانندگی را می‌توان به عنوان «میزان تلاش» عامل در نظر گرفت که در شرایط یکسان راننده‌ی خبره‌تر به طور نسبی در زمان کوتاه‌تری یک مسافت مشخصی را طی خواهد کرد (در رد کردن پیچ و خم‌های ترافیک و مدت زمان ترمز و ... زمان کمتری را تلف می‌کند). یا به عنوان مثال دیگر، دانشجوی قوی و دانشجوی ضعیف را مورد بررسی قرار دهیم، دانشجویی خبره هست که زمان کمتری را صرف حل صحیح یک مساله خاص کند (با فرض اینکه دانشجویها حتماً باید مساله را حل کنند). همانطور که دیدیم کمیت «میزان تلاش» عامل برای مسائل مختلف معیار متفاوتی را دربر می‌گیرد ولی همگی از همان اصل معرفی شده در تئوری ۱-۲ تبعیت می‌کنند.

<sup>1</sup>Interested

در یادگیری مشارکتی با استفاده از تئوری ۱-۲ می‌توان با تعریف ۱-۲ یک معیار خبرگی جدید را معرفی کرد که مبنی و پایه‌ی دستاوردهای این پژوهش می‌باشد.

**تعریف ۱-۲** (معیار خبرگی «میزان ارجاع»). فرض می‌کنیم مجموعه‌ای از عامل‌ها  $\mathbb{A} = \{A_1, A_2, \dots, A_m\}$  در محیط  $\mathcal{E}$  در پی رسیدن به یک مجموعه اهداف  $\mathcal{G} \subseteq \{g_1, g_2, \dots, g_n\}$  می‌باشند. اگر ما به طور مجازی و دلخواه محیط  $\mathcal{E}$  را به  $k$  ناحیه مانند  $e_i$  افراز کنیم بطوری که  $\mathcal{E} = \{\cup_{i=1}^k e_i \mid \forall i, j \in \{1, 2, \dots, k\} \wedge i \neq j : e_i \cap e_j = \emptyset\}$  طبق تئوری ۱-۲ در هر ناحیه  $i$ ام عاملی خبره‌تر است که میزان حضور آن عامل در آن ناحیه کمتر از دیگران است.

در تشریح آنچه که در تعریف ۱-۲ آمده است می‌توان گفت که در سیستم‌های چندعاملی که همگی عوامل در یک محیط به صورت مستقل در حال فعالیت هستند؛ محیط را به چند ناحیه دلخواه افراز می‌کنیم که اجتماع نواحی باهم کل محیط  $\mathcal{E}$  را تشکیل دهند و هیچ دو ناحیه‌ای اشتراکی باهم نداشته باشند [۹]. در این چنین افرازی از محیط، در هر ناحیه عاملی که نسبت به بقیه خبره‌تر است، نسبت به بقیه عوامل در همان ناحیه میزان تمایل حضور کمتری را از خود نشان می‌دهند. به عبارت دیگر عاملی که خبره‌تر است تمایل دارد کوتاه‌ترین مسیر رسیدن به اهداف خود را طی کند که نهایتاً منجر خواهد شد که میزان حضور عامل در هریک از نواحی محیط کمینه شود.

آنچه که در تئوری ۱-۲ در مورد «میزان تلاش» عامل آمده است در تعریف ۱-۲ در به صورت «میزان حضور عامل در هر ناحیه» تعریف شده است. بطوری که طبق تئوری مطرح شده میزان خبرگی عامل در هر ناحیه رابطه‌ی معکوسی با میزان حضور عامل در همان ناحیه را دارد. زیرا اگر عامل نسبت به محیط خود شناخت کامل‌تری داشته در هنگام تلاش برای رسیدن به اهداف خود به علت شناخت خوبی که از محیط دارد کمتر در محیط پرسه می‌زند (کمتر تلاش می‌کند) و با تعداد گام کمتری به سمت اهداف خود حرکت می‌کند - در واقع مسیر بهتری/کوتاه‌تری برای رسیدن به هدف را می‌شناسد. این موضوع در نهایت منجر می‌شود که عاملی که در هر ناحیه خبره‌تر است در همان ناحیه میزان پرسه زدن (حضور/تلاش) کمتری نسبت به دیگر عامل‌ها که از خبرگی نسبی کمتری برخوردار است را داشته باشد.

تا به اینجا گفته شد که عاملی که از خبرگی بیشتری برخوردار است لزوماً کمتر در محیط پرسه می‌زند و با طی کردن مسیر کوتاه‌تر به سمت اهداف خود، تلاش کمتری می‌کند ولی چند سوال در اینجا مطرح می‌شود که برای حل مساله نیازمند پاسخ به آن‌ها هستیم.

۱. میزان حضور عامل را در نواحی مختلف، که محیط از  $d$ -بعد تشکیل شده است چگونه مدل شود؟
۲. اگر عاملی که در هر چرخه یادگیری به یکی از نواحی کلا وارد نشد و میزان پرسه زدن عامل در آن ناحیه صفر شود؛ آیا این مقدار کمینه پرسه زدن، نشان دهنده‌ی خبرگی عامل در آن ناحیه است؟



۳. چگونه در معیار خبرگی ارائه شده باید مساله عدم حضور عامل در یکی از نواحی را مدل کرد، بگونه‌ای که اثر سوئی بر تجربه‌ی دیگر عامل‌ها در آن نواحی، در هنگام ترکیب دانش عامل‌ها نداشته باشد؟

پاسخ به این سوالات برای حل مساله با استفاده از معیار خبرگی پیشنهادی (تعریف ۲-۱) ضروری است. ما به ازای کلیه‌ی نواحی یک ماتریسی به نام «ماتریس ارجاع»<sup>۱</sup> (یا به اختصار REFMAT<sup>۱</sup>) در نظر میگیریم که در ابتدا صفر مقداردهی شده‌اند و هر دفعه که عامل از موقعیتی به موقعیت دیگر می‌رود مقدار آن ناحیه‌ای که موقعیت جدید در آن واقع است را یک واحد افزایش می‌دهیم بدین وسیله میزان حضور عامل در نواحی مختلف را می‌شماریم. همانطور که در قسمت آزمایشات این پایان‌نامه نشان داده شده است در صورتی که از تابع انتخاب عمل بولترمن استفاده کنیم میزان کوچک یا درشت بودن این نواحی در کیفیت نتیجه تاثیرگذار نیست. یعنی عملاً چه ما در حالت کلی، کل محیط را به عنوان یک ناحیه در نظر بگیریم و میزان حضور عامل در این ناحیه را بشماریم (که معادل می‌شود با تعداد گام‌های عامل در طی رسیدن به هدف) یا در حالت جزئی به ازای هر موقعیت موجود را یک ناحیه در نظر بگیریم (که معادل می‌شود با تعداد ملاقات هر یکی از موقعیت‌ها توسط عامل) به یک نتیجه می‌رسیم.

به همین دلیل در پاسخ به سوال دوم، اگر تعداد نواحی زیاد باشد (مثلاً هر موقعیت یک ناحیه باشد - حداکثر تعداد نواحی) ممکن است عامل در طی رسیدن به هدف برخی از نواحی را کلاً ملاقات نکند و مقدار ارجاع به آن نواحی صفر شود و از طرفی طبق تعریف ۲-۱ عاملی که تعداد حضور کمتری در نواحی مختلف داشته باشد از خبرگی بیشتری در آن نواحی برخوردار است و در این شرایط که مقدار ارجاع عامل به ناحیه‌ای صفر باشد را نمی‌توان به خبرگی عامل در آن ناحیه نسبت داد زیرا که آن عامل در کل، آن ناحیه را ملاقات نکرده است که بخواهد تجربه‌ای را در تعامل با آن ناحیه کسب کند تا بتواند خبرگی خود را در آن ناحیه افزایش دهد. برای حل این مشکل و پاسخ به سوال سوم، ماتریسی جدیدی به نام ماتریس خاطره<sup>۲</sup> (یا به اختصار RCMAT<sup>۲</sup>) را معرفی می‌کنیم. این ماتریس وظیفه‌ی نگهداری آخرین ارجاعات غیر صفر عامل را به هر کدام از نواحی تعریف شده را دارد و در زمان‌هایی که مقدار یک ناحیه در ماتریس REFMAT صفر باشد مقدار آن ناحیه از ماتریس RCMAT بروز رسانی می‌شود که میزان پرسه زدن عامل در آن ناحیه در آخرین باری عامل آن ناحیه را ملاقات کرده است را می‌دهد؛ در صورتی که مقدار پرسه زدن یک ناحیه در ماتریس REFMAT مقداری غیر صفر باشد مقدار ماتریس RCMAT با مقدار کنونی REFMAT آن ناحیه بروز رسانی می‌شود.

دلیل استفاده از ماتریس RCMAT این است که در یادگیری تقویتی عامل زمانی می‌توان دانش (سیاست/خبرگی)

<sup>1</sup>Reference Matrix

<sup>2</sup>Recall Matrix

خود را نسبت به نحوه‌ی عمل در یک موقعیت بهبود ببخشد که آن موقعیت را ملاقات کند. حال اگر عامل موقعیتی را ملاقات نکند دانش وی در آن موقعیت ثابت خواهد ماند به همین دلیل اگر عامل ناحیه‌ای را ملاقات نکند و مقدار REFMAT آن ناحیه صفر باشد می‌دانیم که دانش (خبرگی) عامل در آن ناحیه در این چرخه‌ی یادگیری ثابت مانده است و در صورتی که دوباره در آن ناحیه قرار می‌گرفت، حدودا به همان میزان آخرین ملاقات در آن محیط پرسه خواهد زد. به عبارت دیگر در یک چرخه یادگیری اگر هر ناحیه ملاقات نشده، مورد ملاقات واقع می‌شد، تقریباً به میزان آخرین تعداد ارجاع شده برای آن نواحی، مورد ارجاع واقع می‌شد.

## ۲-۳ یادگیری مشارکتی $Q$ با استفاده از ماتریس ارجاع و انتگرال فازی

آنچه که تا به اکنون در مورد روش پیشنهادی این پژوهش آورده شده، معرفی یک معیار خبرگی که در برعکس بسیاری از معیارهای خبرگی که تا به کنون معرفی شده است [۱۰-۱۲] در تمامی موقعیت‌های دنیای واقعی به وفور مشاهده می‌شود و آن ارائه این تئوری است عامل خبره‌تر برای رسیدن به یک مجموعه از اهداف تلاش نسبی کمتری نسبت به دیگر عامل‌ها با خبرگی کمتر در شرایط یکسان می‌کند. حال که معیاری برای میزان خبرگی عامل‌ها در اختیار داریم چالش بعدی برای بهبود کیفیت و سرعت یادگیری مشارکتی ارائه‌ی روشی برای ترکیب دانش‌های عامل‌ها از محیط (جداول  $Q$  آن‌ها) با استفاده از معیار ارائه شده می‌باشد. روش ترکیب باید بگونه‌ای باشد که کیفیت و سرعت یادگیری مشارکتی عامل‌ها را در طی زمان نسبت زمانی که عامل‌ها بدون مشارکت یاد می‌گیرند بهتر کند. همچنین کیفیت و سرعت یادگیری همبستگی مستقیمی داشته باشند با تعداد عامل‌هایی که درحال اشتراک گذاری هستند؛ به عبارت دیگر در صورت افزایش تعداد عامل‌هایی که دانش‌های خود را به اشتراک می‌گذارند مدل ترکیب کننده‌ی دانش‌های آن عامل‌ها باید بتواند دانش بهتری تولید کند که نهایتاً منجر به بهتر شدن کیفیت و سرعت کلی یادگیری عامل‌ها شود.

در این پژوهش ما انتگرال فازی را به عنوان مدل ترکیب کننده‌ی دانش‌های عامل‌ها پیشنهاد می‌دهیم. دلیل انتخاب این مدل ویژگی‌های منحصر به فردی است که این مدل کننده در اختیار دارد که مدل را کاملاً مناسب برای ترکیب دانش عامل‌ها می‌کند؛ که در بخش‌های آتی فصل این ویژگی‌ها و دلایل مناسب بودن آن‌ها برای ترکیب دانش عامل‌ها آورده شده است. لازم به یادآوری است که همانطور که در قسمت ۱-۱ این پایان‌نامه آورده شده است ما از به دلایل فنی از انتگرال فازی چوکت استفاده می‌کنیم که در بخش‌های بعدی این دلایل نیز بطور مفصل شرح داده می‌شود.

1: **procedure** REFMAT-COOPERATIVE-LEARNING( $m$ )

**Require:**  $m > 1$

▷ The number of agents.

**Ensure:** Initialize the  $Q$  matrix

**Ensure:** Initialize the RCMAT  $\leftarrow 0$

```

2:   while not End Of Learning do
3:     REFMAT  $\leftarrow 0$ 
4:     if In individual learning mode then
5:       Visit the state  $s$ ;
6:       Select an action  $a$  based on an action selection policy;
7:       Carry out the  $a$  and observe a reward  $r$  at the new state  $s'$ ;
8:        $Q[s, a] \leftarrow Q[s, a] + \alpha(r + \lambda \max_{a'} (Q[s', a']) - Q[s, a])$ ;
9:       Increment REFMAT( $\phi(s')$ ) by one;
10:       $s \leftarrow s'$ ;
11:    else if In cooperative learning mode then
12:      REFMAT, RCMAT  $\leftarrow$  Swap(REFMAT, RCMAT);
13:      CoQFCI  $\leftarrow$  FCI_Combiner(All agents'  $Q$  and REFMAT matrices);
14:      for each agent  $i \leftarrow 1, m$  do
15:         $Q_i \leftarrow$  CoQFCI;
```

## ۲-۳-۱ الگوریتم پیشنهادی

در این قسمت به معرفی الگوریتم پیشنهادی می‌پردازیم. آنچه که در الگوریتم ۱ آمده است به دو قسمت تشکیل شده است، یک قسمت که مربوط یادگیری مستقل (خطوط ۵ تا ۱۰) و قسمت دیگری مربوط به یادگیری مشارکتی (خطوط ۱۲ تا ۱۵) می‌باشد. ورودی الگوریتم تعداد عامل‌ها می‌باشد و در ابتدا ماتریس‌های  $Q$  و REFMAT و RCMAT مقداردهی می‌شود. سپس تا زمانی که یادگیری پایان نیافته است ابتدا عامل‌ها در قسمت یادگیری مستقل به صورت جدا گانه در محیط فعالیت می‌کنند که رویه‌های آورده شده در خطوط ۵ تا ۸ و همچنین خط ۱۰ همان الگوریتم یادگیری  $Q$  متعارف می‌باشد [۱۳]. در قسمت یادگیری مستقل تنها خط ۹ می‌باشد که در روش پیشنهادی به شبه‌کد اضافه شده است و این تنها یک وظیفه‌ی بسیار ساده را انجام می‌دهد و آن شمارش میزان حضور عامل در هر کدام از نواحی از پیش تعیین شده است؛  $\phi(\cdot)$  یک تابع نگاشت از یک موقعیت به یک ناحیه از محیط می‌باشد.

بعد از طی یادگیری مستقل عامل‌ها به قسمت اشتراک گذاری دانش‌های خود (جداول  $Q$ ) می‌رسند (خطوط ۱۲ تا ۱۵). در قسمت یادگیری مشترک ابتدا طبق آنچه که در در قسمت آورده شده است جداول REFMAT و RCMAT به صورت مشترک بروزرسانی می‌شود و سپس جداول  $Q$  و REFMAT تمامی عامل‌ها به مدل ترکیب کننده فازی معرفی شده در این پژوهش فرستاده می‌شود و مدل ترکیب کننده فازی وظیفه‌ی استخراج یک دانش جدید با در نظر گرفتن ورودی‌های آن برای جایگزینی دانش قبلی عامل‌ها می‌باشد.

---

**الگوریتم ۲** تابع Swap معرفی شده در الگوریتم ۱
 

---

```

1: procedure Swap(REFMAT, RCMAT)
Require: size(REFMAT) = size(RCMAT)
2:   for each element  $r$  in REFMAT and its corresponding element  $c$  in RCMAT do
3:     if  $r = 0$  then
4:       Update  $r = c$ ;
5:     else
6:       Update  $c = r$ ;
7:   return REFMAT, RCMAT

```

---



---

**الگوریتم ۳** تابع FCI\_Combiner معرفی شده در الگوریتم ۱
 

---

```

1: procedure FCI_Combiner( $\vec{K}, \vec{R}$ )
Require:  $\text{length}(\vec{K}) = \text{length}(\vec{R}) = m$ 
Ensure: Initialize  $\text{CoQ}_{\text{FCI}}$ 
2:   for each state  $s$  do
3:      $\vec{f} \leftarrow \{\}$ ;  $\triangleright$  Contains the normalized valued of REFMATs' value for state  $s$  for all agents
4:     for each REFMAT in  $\vec{R}$  do
5:        $\vec{f}.\text{add}(\text{REFMAT}(\phi(s)))$ ;
6:      $\vec{A} \leftarrow 1 - \text{normalize}(\vec{f})$ ;
7:     for each possible action  $a$  in state  $s$  do
8:        $\vec{x} \leftarrow \{\}$ ;  $\triangleright$  Contains the  $Q$  values of action  $a$  in state  $s$  for all agents
9:       for each  $Q$  in  $\vec{K}$  do
10:         $\vec{x}.\text{add}(Q[s, a])$ ;
11:         $\text{CoQ}_{\text{FCI}}[s, a] \leftarrow \sum_{i=1}^m (f(x_{\pi(i)}) - f(x_{\pi(i-1)})) \cdot g(\vec{A}_i)$   $\triangleright$  The Choquet Integral
12:   return  $\text{CoQ}_{\text{FCI}}$ ;

```

---

الگوریتم تابع بسیار ساده می‌باشد و مقادیر غیر صفر ماتریس ارجاع را در ماتریس خاطره کپی می‌کند و مقادیر صفر ماتریس ارجاع را از ماتریس خاطره جایگزین می‌کند. این تابع در الگوریتم ۲ آمده است. در این پژوهش در دو قسمت نوآوری صورت گرفته است، قسمت اول ارائه‌ی معیاری جدید برای سنجش معیار خبرگی که طبق تعریف ۲-۱ این معیار در خط ۹ الگوریتم ۱ پیاده‌سازی شده است؛ نوآوری دوم نحوه‌ی ترکیب اطلاعات دانش عامل‌ها با استفاده از انتگرال فازی که در خط ۱۳ الگوریتم ۱ و شرح جزئیات پیاده‌سازی آن در الگوریتم ۳ آمده است.

ورودی‌های الگوریتم ۳ به ترتیب مجموعه‌ای از جداول  $Q$  و ماتریس‌های ارجاع (REFMAT) تمامی عامل‌ها می‌باشد بطوری که در ازای هر جدول  $Q$  یک ماتریس REFMAT متناظر وجود دارد. خروجی این الگوریتم یک جدول  $Q$  می‌باشد که از ترکیب جداول  $Q$  ورودی با در نظر گرفتن میزان خبرگی هر کدام از عامل‌ها که توسط ماتریس‌های REFMAT آن‌ها تعیین می‌شود. الگوریتم ۳ به ازای کلیه‌ی موقعیت‌ها ( $s$ ها در خط ۲) ابتدا مقادیر REFMAT کلیه‌ی عامل‌ها در ناحیه‌ای که آن موقعیت در آن واقع است (که توسط تابع نگاشت  $\phi(\cdot)$  بدست

می‌آید) را استخراج می‌کند و در برداری بنام  $f^1$  ذخیره می‌کند (خطوط ۴ و ۵) که در واقع میزان ارجاعات هر کدام از عامل‌ها در ناحیه‌ی  $\phi(s)$  می‌باشد. بردار  $f^1$  معیاری برای سنجش میزان خبرگی کلی عامل‌ها در موقعیت  $s$  است، طبق آنچه که در تعریف ۱-۲ آمده است در هر ناحیه عاملی خبره‌تر است که مقدار REFMAT مربوط به آن ناحیه از دیگر عامل‌ها کمتر باشد. در نتیجه در خط ۶ بعد از عادی‌سازی<sup>۲</sup> مقادیر REFMAT عامل‌ها در ناحیه‌ی  $\phi(s)$  یک مکمل‌گیری صورت می‌گیرد تا عاملی که مقدار REFMAT کمتری دارد دارای بیشترین مقدار بعد از عادی‌سازی شود. در خط ۷ به ازای کلیه‌ی عمل‌های ممکن در موقعیت  $s$  ابتدا مقادیر  $Q$  تک‌تک عامل‌ها را در موقعیت  $s$  و عمل  $a$  در خطوط ۹ و ۱۰ در بردار  $\vec{x}$  ذخیره می‌کنیم و در نهایت در خط ۱۱ با استفاده از انتگرال فازی چوکت معرفی شده در ۱-۸ مقدار  $Q$  مشارکتی حاصل از میزان خبرگی بردار  $\vec{A}$  و مقادیر  $Q$ ‌های تک‌تک عامل‌ها در بردار  $\vec{x}$  در موقعیت  $s$  و عمل  $a$  بدست محاسبه می‌شود.

## ۲-۳-۲ تعیین توابع $f(\cdot)$ و $g(\cdot)$ در انتگرال فازی چوکت

بطور خلاصه در الگوریتم ۳ دو بخش عمده دارد بخش اول مربوط استخراج میزان خبرگی عامل‌ها بگونه‌ای که عاملی که خبره‌تر از دارای مقدار خبرگی بیشتری باشد که این بخش در خطوط ۴ تا ۶ صورت می‌گیرد؛ بخش دیگر محاسبه‌ی مقادیر  $Q$  مشارکتی کلیه‌ی عمل‌های ممکن در یک موقعیت با در نظر گرفتن میزان خبرگی عامل‌ها و مقادیر  $Q$  آن‌ها با استفاده از انتگرال فازی چوکت که در خطوط ۷ تا ۱۱ صورت می‌پذیرد.

آنچه که در خط ۱۱ الگوریتم ۳ مورد توجه واقع شود این است که توابع  $f(\cdot)$  و  $g(\cdot)$  چگونه تعریف باید تعریف شوند؟ برای تعیین تابع  $f(\cdot)$  منطقی که در این پژوهش استفاده کردیم بدین صورت است که از آنجایی که خروجی تابع  $g(\cdot)$  یک مقدار عددی<sup>۳</sup> بدون واحد می‌باشد و همچنین برای اینکه خروجی انتگرال فازی خط ۱۱ را بتوان به عنوان مقادیر جدول  $Q$  مشارکتی جدید در نظر گرفت تا بتوانیم در خطوط ۱۵ الگوریتم ۱ به عنوان جدول  $Q$  تک‌تک عامل‌ها جایگذاری کنیم باید خروجی انتگرال فازی خط ۱۱ الگوریتم ۳ از جنس جدول‌های  $Q$  عامل‌ها باشد در نتیجه تابع  $f(\cdot)$  باید یک تابع خطی بصورت ۱-۲ باشد تا خروجی انتگرال فازی همجنس مقادیر  $\vec{x}$  باشد.

$$f(\omega) = a\omega + b \quad (1-2)$$

متغیرهای  $a$  و  $b$  در ۱-۲ می‌تواند به عنوان پارامترهای سازگار<sup>۴</sup> در میزان کیفیت جدول  $Q$  مشارکتی خروجی

<sup>1</sup> Factors

<sup>2</sup> Normalize

<sup>3</sup> Scalar

<sup>4</sup> Addaptive Parameters

---

```

1: procedure Const-One( $\vec{A}_i$ )
2:   if length( $\vec{A}_i$ )  $\geq m$  then
3:     return 1;
4:   else if length( $\vec{A}_i$ ) = 0 then
5:     return 0;
6:   else
7:     return 1;

```

---

الگوریتم ۳ موثر واقع شود ولی با این حال در این پژوهش مقادیر  $a$  و  $b$  هر دو به ترتیب مقادیر ثابت ۱ و صفر در نظر گرفته شده‌اند که یعنی از تابع همانی به عنوان تابع  $f(\cdot)$  استفاده شده است.

تابع  $g(\cdot)$  یک ورودی مرتب شده طبق آنچه که در ۱-۱۰ آمده است می‌گیرد و در الگوریتم ۳ تعیین این تابع تاثیر زیادی بروی کیفیت خروجی الگوریتم خواهد داشت ولی چالش‌هایی برای تعیین این تابع داریم؛ تابع  $g(\cdot)$  باید دارای ویژگی‌های زیر باشد:

۱. پویا<sup>۱</sup> باشد: از آنجایی که تابع  $g(\cdot)$  میزان اندازه‌گیری غیرافزایشی<sup>۲</sup> منابع اطلاعاتی را در اختیار می‌گذارد [۱۴]، نیاز داریم تعیین کنیم که کدام منابع اطلاعاتی (در اینجا خبرگی عامل‌ها) در کنار هم چه ارزش افزوده‌ای دارد؛ ولی از آنجایی که در حین یادگیری مشترک روشی برای تعیین این ارزش افزوده نداریم بنابراین باید تابع  $g(\cdot)$  بصورت پویا بتواند مقادیر این ارزش افزوده را تخمین بزند.

۲. قابل گسترش<sup>۳</sup> باشد: زیرا که تعداد عامل‌ها در محیط متغیر است لذا باید تابع  $g(\cdot)$  بگونه‌ای باشد به ازای تغییر تعداد عامل‌ها (که تغییر در تعداد اعضای بردار  $\vec{A}$  را در پی دارد) قابل گسترش باشد.

یکی از روش تخمین  $g(\cdot)$  که دو ویژگی بالا را داشته باشد، تابع اندازه‌گیری- $\lambda$  سوگنو می‌باشد ولی این تابع نیاز به ریشه‌یابی روی متغیر  $\lambda$  دارد که طبق آنچه که در ۱-۴ آمده است به ازای تعداد عامل‌های مختلف نیاز به ریشه‌یابی معادلات غیرخطی دارد که بدلیل پیچدگی محاسباتی این ریشه‌یابی و همچنین طبق نتایج حاصل از دستاوردهای این پژوهش که در فصل نتیجه‌گیری آورده شده است، در آزمایشات صورت گرفته در این پژوهش از تابع اندازه‌گیری- $\lambda$  سوگنو به عنوان تابع  $g(\cdot)$  استفاده نشده است. یک سری توابع در این پژوهش بجهت استفاده، آزمایش و نتیجه‌گیری به عنوان  $g(\cdot)$  معرفی شده است که این توابع در الگوریتم‌های ۴ تا ۷ آمده‌اند.

در الگوریتم ۴ به ازای هر ورودی دلخواه مقدار ثابت ۱ به عنوان خروجی برگشت داده می‌شود، این بدین معنی است که ارزش افزوده‌ی هر نوع ترکیب اطلاعاتی (خبرگی) برای ما دارای حداکثر ارزش می‌باشد و این مساله

---

<sup>1</sup>Dynamic

<sup>2</sup>Non-additive

<sup>3</sup>Expandable

---

**الگوریتم ۵** Max برای تخمین تابع  $g(\cdot)$  در الگوریتم ۳
 

---

```

1: procedure Max( $\vec{A}_i$ )
2:   if length( $\vec{A}_i$ )  $\geq m$  then
3:     return 1;
4:   else if length( $\vec{A}_i$ ) = 0 then
5:     return 0;
6:   else
7:     return max;
        $\vec{A}_i$ 

```

---



---

**الگوریتم ۶** Mean برای تخمین تابع  $g(\cdot)$  در الگوریتم ۳
 

---

```

1: procedure Mean( $\vec{A}_i$ )
2:   if length( $\vec{A}_i$ )  $\geq m$  then
3:     return 1;
4:   else if length( $\vec{A}_i$ ) = 0 then
5:     return 0;
6:   else
7:     return  $\frac{\sum_{j=i}^m \vec{A}_i(j)}{\text{length}(\vec{A}_i)}$ ;

```

---

باعث می‌شود که نتیجه‌ی انتگرال فازی خط ۱۱ الگوریتم ۳ مقداری معادل با مقدار خبره‌ترین عامل (عاملی که کمترین پرسه را در محیط مربوطه داشته) را به عنوان مقدار جدید جدول  $Q$  مشارکتی تولید کند.

در الگوریتم ۵ میزان خبرگی خبره‌ترین عامل به عنوان خروجی تابع  $g(\cdot)$  برگشت داده می‌شود. در الگوریتم ۶ خروجی، میانگین خبرگی عامل‌ها در نظر گرفته شده است و در الگوریتم ۷ طبق رابطه‌ی نوشته شده میانگین  $k$ ام میزان خبرگی‌ها به عنوان خروجی برمی‌گردد به طوری که بزرگترین خبرگی در عدد  $k$  و کوچکترین خبرگی در عدد ۱ و هر آنچه که مابین این دو خبرگی وجود دارد در اندیس ترتیب مرتب شده آن‌ها ضرب می‌شود و میانگین این مجموع محاسبه می‌شود و برگشت داده می‌شود؛ توجه شود که ورودی‌های الگوریتم‌های ۴ تا ۷ طبق آنچه که در ۱-۱۰ آمده است یک مجموعه‌ی مرتب می‌باشد.

## ۲-۴ علت کارکرد انتگرال فازی چوکت در انتقال دانش

در این قسمت به بررسی شهودی اینکه چرا انتگرال فازی چوکت برای انتقال (ترکیب) دانش‌های عامل‌ها می‌تواند موثر واقع باشد می‌پردازیم. این شهود بعدها در آزمایش‌ها نشان داده خواهد شد که صحت دارد. انتگرال فازی چوکت یک سری ویژگی‌ها دارد که برای انتقال دانش مدل می‌کند. از مهم‌ترین ویژگی‌ها را می‌توان به موارد زیر اشاره کرد [۵].

۱. **محدود است:** اگر شرایط مرزی و یکنوایی تابع  $g(\cdot)$  برقرار باشد انتگرال فازی هیچ‌گاه بیشتر از حداکثر

---

```

1: procedure K-Mean( $\vec{A}_i$ )
2:   if length( $\vec{A}_i$ )  $\geq m$  then
3:     return 1;
4:   else if length( $\vec{A}_i$ ) = 0 then
5:     return 0;
6:   else
7:     return  $\frac{\sum_{j=i, k=1}^{m, \text{length}(\vec{A}_i)} k \cdot \vec{A}_i(j)}{(\sum_{j=1}^{\text{length}(\vec{A}_i)} j) - 1}$ ;

```

---

مقدار  $f(x_{\pi_i})$  ها و کمتر از حداقل مقدار آن‌ها خروجی نمی‌دهد [۶]. یعنی دانش تولیدی خارج از محدوده‌ی دانش فعلی عامل‌ها نمی‌باشد فقط ترکیب بهینه‌ای از این دانش‌ها به عنوان خروجی برگشت داده می‌شود که این در کاربرد یادگیری تقویتی به این معنی است که هیچ‌گاه مقادیر جدول  $Q$  بیشتر یا کمتر از آنچه که تجربه شده نمی‌شود و این باعث می‌شود که ضمانت همگرایی یادگیری تقویتی  $Q$  با اعمال انتگرال فازی چوکت نقض نشود و الگوریتم حتما همگرا شود؛ ولی در صورتی که روشی خارج از دانش کنونی عامل‌ها خروجی دهد ضمانتی برای همگرایی عامل‌ها وجود نخواهد داشت.

۲. می‌تواند اندازه‌گیری‌های غیرافزایشی مدل کند: معمولا روش‌هایی که تا به‌کنون در این زمینه ارائه شده است از میانگین وزنی خبرگی عامل‌ها برای بدست آوردن جدول  $Q$  مشترک استفاده کرده‌اند [۱۰-۱۲]. این در حال هست که میانگین وزن دار قسمتی از مدل اندازه‌گیری‌های غیرافزایشی می‌باشد. بنابراین با در نظر گرفتن مدل‌های غیرافزایشی که در ماهیت مساله هست قدرت و انعطاف بیشتری در اختیار داریم نسبت به روش‌هایی که فقط از میانگین وزنی استفاده کرده‌اند.

**تعریف ۲-۲** (اندازه‌گیری‌های غیرافزایشی). اگر فرض کنیم  $(X, A)$  فضای قابل اندازه‌گیری باشد که  $X$  مجموعه‌ی مرجع<sup>۱</sup> و  $A \subseteq X$ ، آنگاه تابع مجموعه‌ای مانند  $\mu$  که  $\mu : A \rightarrow [0, 1]$  اندازه‌گیر غیرافزایشی می‌گویند هرگاه شرایط زیر را ارضا کند [۱۴].

- $\mu(\emptyset) = 0, \quad \mu(X) = 1$

- $A \subseteq B \Rightarrow \mu(A) \leq \mu(B)$

تورا و همکاران [۱۴] یک مجموعه جامعی در مورد اندازه‌گیری‌های غیرافزایشی ایجاد کرده‌اند که جزئیات این مطلب خارج از حوصله‌ی این نوشتار است و در صورت تمایل به کسب اطلاعات بیشتر در مورد اندازه‌گیری‌های غیرافزایشی و انتگرال‌های فازی می‌توانید به آن مراجعه نمایید.

---

<sup>1</sup>Reference Set

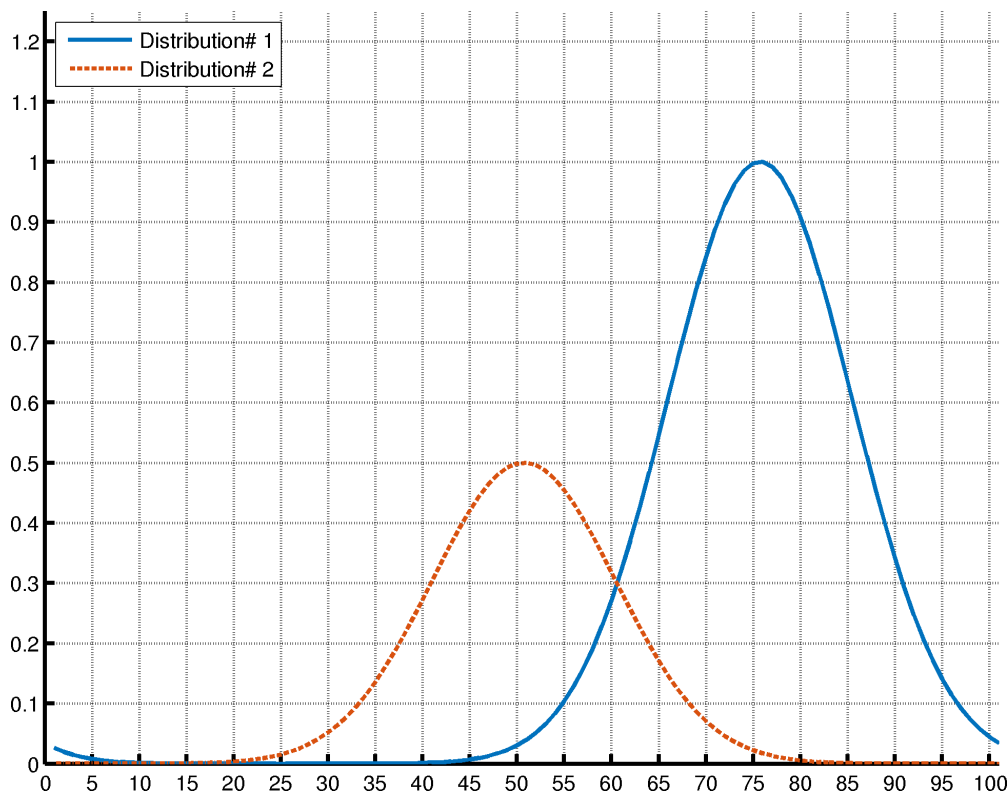


## فصل سوم

### نتایج عملی

#### ۱-۳ مقدمه

در این فصل به ارائه‌ی آزمایش‌های صورت گرفته بروی روش پیشنهادی می‌پردازیم و در طی این آزمایش‌ها روش پیشنهادی را با روش کوتاه‌ترین مسیر تجربه شده (یا به اختصار SEP) مقایسه می‌کنیم که آخرین و مدرن‌ترین روش ارائه شده در جهت بهبود یادگیری مشارکتی می‌باشد [۱۲]. آزمایش‌ها بروی دو محیط «پلکان مارپیچ» و «صید و صیاد» صورت گرفته است. آزمایش‌ها به دو دسته تقسیم بندی شده است؛ دسته اول آزمایش‌هایی که روش پیشنهادی را در مقابل روش SEP قرار می‌دهد و عملکرد روش پیشنهادی را مورد سنجش قرار می‌دهد. دسته دوم آزمایش‌ها مربوط به آزمون رفتار روش پیشنهادی در صورت تغییر در پارامترهای متخلف آن می‌باشد. همچنین اثر استفاده از سیاست‌های انتخاب عمل مختلف در الگوریتم ۱ نیز بررسی شده است. در روش‌های مرتبط مدرن قبلی [۱۱، ۱۲] که این پژوهش ادامه‌ی کار آن‌ها می‌باشد فقط از سیاست انتخاب عمل Boltzmann استفاده کرده‌اند؛ در این پژوهش علاوه بر Boltzmann تاثیر استفاده از روش  $\varepsilon$ -greedy بروی هردو روش پیشنهادی و SEP نیز مورد بررسی واقع گردیده است.



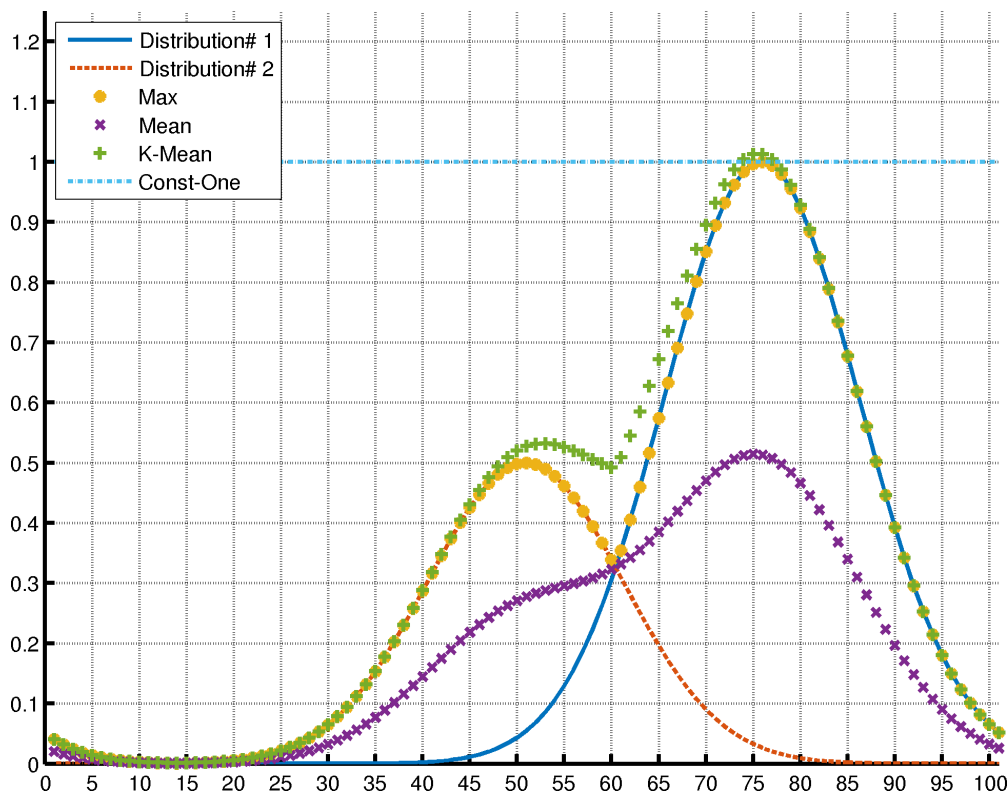
شکل ۳-۱: دو توزیع فرضی بجهت نمایش نحوه رفتار الگوریتم‌های ۴ تا ۷ بروی آن‌ها.

### ۳-۲ رفتار الگوریتم‌های معرفی شده برای $g(\cdot)$

در این قسمت به بررسی رفتار الگوریتم‌های ۴ تا ۷ معرفی شده برای  $g(\cdot)$  بروی دو توزیع فرضی خواهیم پرداخت، زیرا که در طی اجرای آزمایش‌های مختلف نتایج تاثیر این توابع بر اجرای الگوریتم پیشنهادی ۱ آورده شده است، لذا بجهت درک علت تاثیرات مختلف هرکدام از این توابع بروی نتیجه‌ی الگوریتم پیشنهادی در آزمایش‌ها، درک نحوه‌ی رفتار الگوریتم‌های ۴ تا ۷ ضروری است.

برای نمایش نحوه‌ی رفتار هرکدام از الگوریتم‌ها دو توزیع فرضی شکل ۳-۱ فرض شده است. در صورت اعمال الگوریتم‌های ۴ تا ۷ بروی دو توزیع آورده شده در شکل ۳-۱ توزیع‌های جدیدی بصورت آنچه که در شکل ۳-۲ آمده است بدست می‌آیند. همانطور که در شکل ۳-۲ می‌بینیم اعمال الگوریتم Const-One بروی دو توزیع مقدار ثابت ۱ را برمی‌گرداند. اعمال الگوریتم Max در هر نقطه حداکثر مقدار هر دو توزیع را برمی‌گرداند. الگوریتم Mean میانگین دو توزیع را در هر نقطه حساب می‌کند و در نهایت الگوریتم K-Mean میانگین  $k$ ام هردو توزیع را محاسبه میکند که همانطور که می‌بینیم میانگین  $k$ ام به سبب ماهیت الگوریتم به سمت بیشترین مقدار پیش‌قدر<sup>۱</sup> می‌باشد.

<sup>۱</sup>Bias



شکل ۳-۲: نمایش توزیع‌های جدید بدست آمده بعد از اعمال الگوریتم‌های ۴ تا ۷ بروی دو توزیع فرضی شکل ۳-۱

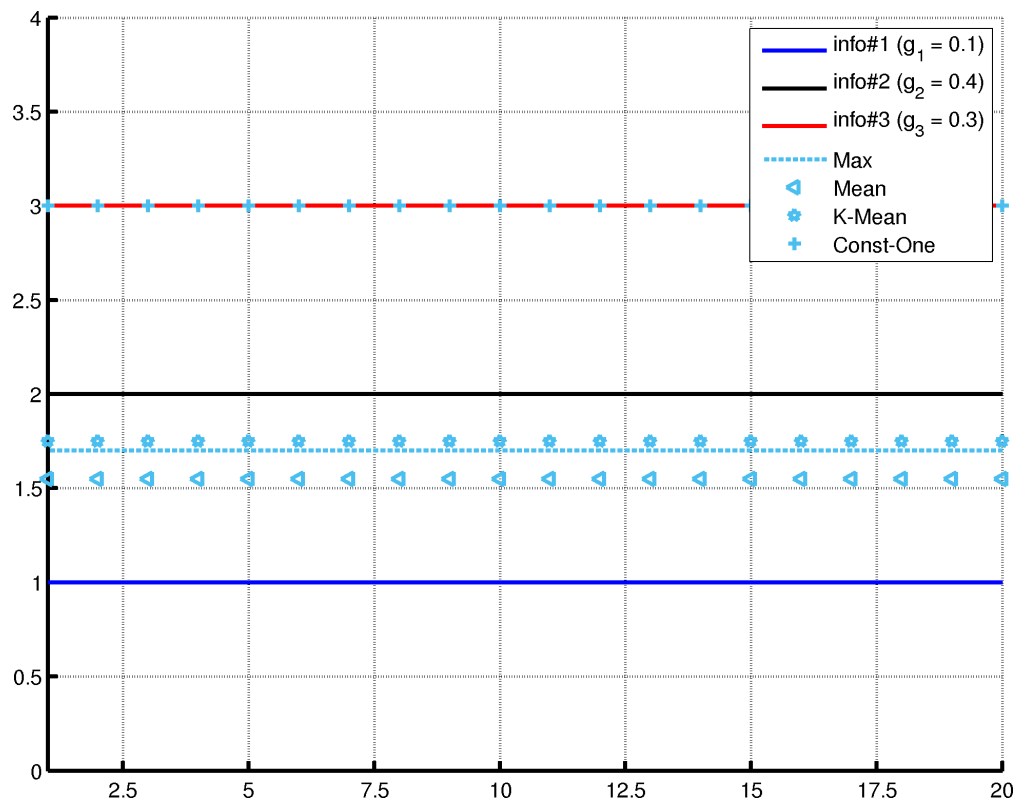
### ۳-۲-۱ تعابیر مختلف انتگرال فازی چوکت از داده‌ها بر مبنای $g(\cdot)$

الگوریتم‌های ۴ تا ۷ به تنهایی فقط در نقش یک عملگر بازی می‌کند ولی در هنگام ترکیب دانش با انتگرال فازی چوکت به دانش خروجی الگوریتم از دیدگاه‌های متفاوتی نگاه می‌کنند. از آنجایی که در فصل‌های قبلی نیز آورده شد انتگرال فازی در واقع یک تعمیم الگوریتم دهنده‌ی میانگین وزنی می‌باشد که علاوه بر ویژگی‌هایی که روش میانگین وزنی ارائه می‌دهد می‌تواند اندازه‌گیری‌های غیرافزایشی را نیز مدل کند. لذا با تغییر تابع  $g(\cdot)$  می‌توان باعث شد که انتگرال فازی چوکت تعابیر مختلفی از داده‌های ورودی خود ارائه دهد. از بین الگوریتم‌ها فقط الگوریتم Const-One دارای تعبیر صریح ریاضی می‌باشد که در ۳-۱ آمده است، بقیه‌ی الگوریتم‌ها دارای تعابیر صریح نیستند و فقط می‌توانیم بر اساسی نمایشی که در شکل ۳-۲ آمده است شهودی از نحوه‌ی تغییر رفتار انتگرال فازی به ازای هریک از الگوریتم‌ها ارائه داد.

$$g = \text{Const-One}(\cdot) \equiv \begin{cases} g(X) & = 1 \\ g(\emptyset) & = 0 \\ g_{A \subseteq X}(A) & = 1 \end{cases} \Rightarrow C_g(f) \equiv \max\{f(x_{\pi_{(1)}^c}), \dots, f(x_{\pi_{(n)}^c})\} \quad (3-1)$$

برای نمایش شهودی نحوه‌ی تغییر رفتار انتگرال فازی چوکت در شکل ۳-۳ سه منبع اطلاعاتی با مقادیر

$$g = [0.1 \quad 0.4 \quad 0.3]^T \text{ و } y = 3 \text{ و } y = 2 \text{ و } y = 1 \text{ در نظر گرفته شده است و مقدار ارزش هر کدام از این‌ها به ترتیب } g = [0.1 \quad 0.4 \quad 0.3]^T$$



شکل ۳-۳: نمایش رفتار انتگرال فازی بروی منابع اطلاعاتی  $y = 1$  و  $y = 2$  و  $y = 3$  به ازای توابع  $g(\cdot)$  های مختلف.

در نظر گرفته شده است. سپس انتگرال فازی چوکت را با در نظر گرفتن تابع همانی به عنوان تابع  $f(\cdot)$  بروی این ۳ منبع اطلاعاتی اعمال کردیم و همانطور که می بینیم مقداری که انتگرال فازی چوکت به ازای  $g = \text{Const-One}(\cdot)$  تولید می کند برابر با حداکثر مقدار منابع اطلاعاتی دریافتی می باشد. در حالت کلی هرچقدر میانگین تابع  $g_{A \subseteq X}(A)$  به سمت مقدار ۱ متمایل باشد خروجی انتگرال فازی چوکت به سمت بیشینه مقدار منابع اطلاعاتی پیش قدر می شود و در صورتی که این میانگین به سمت صفر متمایل باشد خروجی به کمینه مقدار پیش قدر می شود.

### ۳-۳ مقایسه‌ی روش پیشنهادی با روش کوتاه‌ترین مسیر تجربه شده

در این قسمت به مقایسه‌ی روش پیشنهادی با روش «کوتاه‌ترین مسیر تجربه شده» که از بروزترین تکنیک ارائه شده در این شاخه از یادگیری مشارکتی می باشد می پردازیم [۱۲]. کلیه‌ی این آزمایش‌ها در دو محیط «پلکان مارپیچ» و «صید و صیاد» صورت گرفته است. نتیجه‌ی هر آزمایش حاصل میانگین ۲۰ اجرای مستقل تمامی الگوریتم‌ها می باشد. همچنین به غیر از مواردی که صراحتاً قید شده است تعداد عوامل‌ها ۳ عدد می باشد - البته بدیهی است که یادگیری مستقل تک عامله (یا به اختصار IL<sup>۱</sup>) شامل این قاعده نمی باشد. همچنین در کلیه‌ی

<sup>۱</sup>Individual Learning

جدول ۳-۱: لیست اختصارهای استفاده شده در این فصل

معنی	اختصار
روش پیشنهادی	REFMAT
یادگیری مستقل تک عامله	IL
روش کوتاه‌ترین مسیر تجربه شده	SEP
میانگین وزنی	wsum
الگوریتم Max به عنوان مدل کننده‌ی تابع $g(\cdot)$	fci-max
الگوریتم Mean به عنوان مدل کننده‌ی تابع $g(\cdot)$	fci-mean
الگوریتم K-Mean به عنوان مدل کننده‌ی تابع $g(\cdot)$	fci-k-mean
الگوریتم Const-One به عنوان مدل کننده‌ی تابع $g(\cdot)$	fci-const-one
جستجوی کاملاً مکاشفانه محیط	Rand-Walk

آزمایش‌ها عامل‌ها از ۲۰۰ چرخه یادگیری بهره می‌برند و در هر چرخه عامل ۵ بار تلاش می‌کند که در مجموع ۱۰۰۰ تلاش صورت می‌گیرد. کلیه‌ی پارامترهای مربوط قسمت یادگیری مستقل الگوریتم ۱ اعمال شده در آزمایشات این فصل منطبق بر پارامترهای تعریف شده در [۱۲] می‌باشد که نتایج قایل قیاس باشند. در ضمن در این فصل اختصارهای جدول ۳-۱ را نیز داریم.

در این فصل در حالت کلی ما در دو بخش سیاست انتخاب عمل «بولتزمن» و « $\epsilon$ -حریصانه» (که از این به بعد، به اختصار «تابع بولتزمن» و «تابع حریصانه» خطاب خواهیم کرد.) به مقایسه‌ی نتایج می‌پردازیم. طبق آنچه که در ادامه مشاهده خواهیم کرد چه در صورت استفاده از تابع بولتزمن و چه تابع حریصانه روش پیشنهادی چه در سرعت یادگیری و چه در کیفیت یادگیری بهتر از روش SEP می‌باشد.

برای اینکه نشان دهیم که استفاده از انتگرال فازی در بهبود نتیجه تاثیر بسزایی دارد از تابع میانگین وزنی (یا به اختصار wsum<sup>۱</sup>) نیز استفاده کرده‌ایم. بدین صورت که بجای اینکه بعد از استخراج میزان خبرگی هر عامل جداول  $Q$  آن‌ها را به نسبت خبرگی‌ای که دارند باهم جمع می‌کنیم تا جدول  $Q$  مشارکتی تولید شود. تابع میانگین وزنی روشی است که در پژوهش‌های اخیر به کرات از آن استفاده کرده‌اند [۱۰-۱۲]. که یکی از اهداف ما در این پژوهش نمایش قدرت انتگرال‌های فازی در کاربردهای مختلف می‌باشد به‌طوری که اگر در پژوهش‌های قبلی به درستی از انتگرال فازی بهره برده می‌شد می‌توان به قطع گفت که می‌توانستند نتایج بهتری

<sup>۱</sup> Weighted Sum

را بدست بیاورند.

### ۳-۳-۱ مقایسه در محیط پلکان مارپیچ

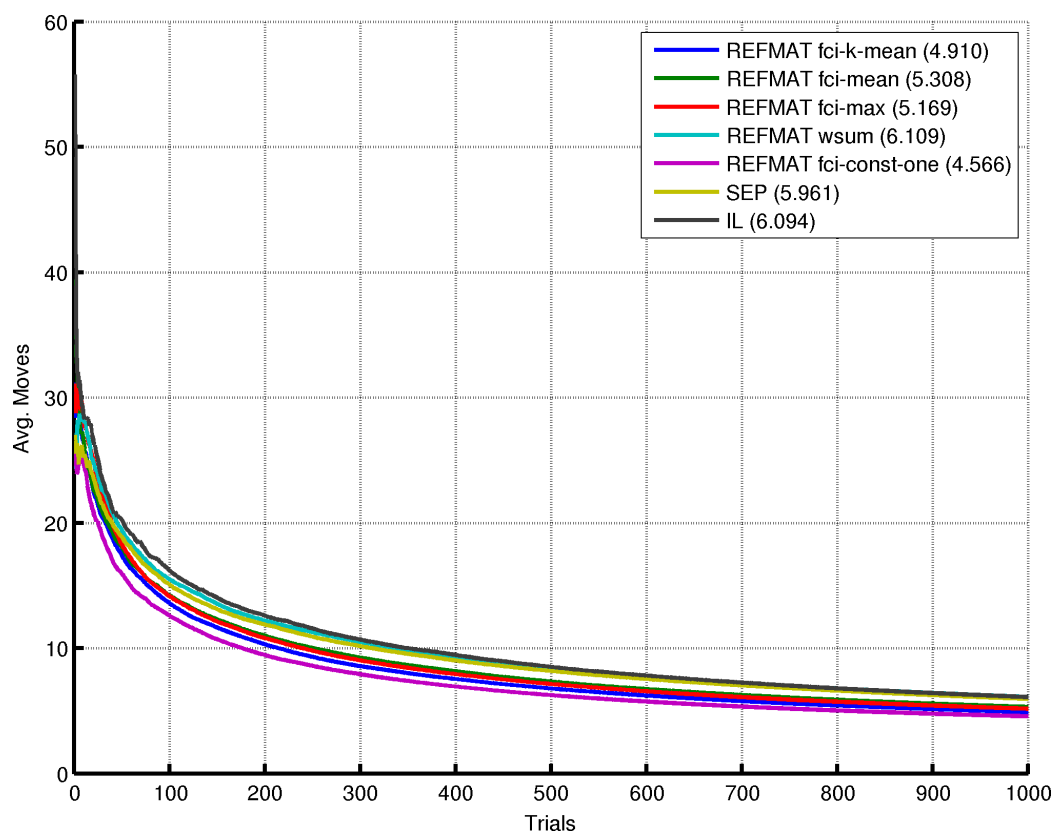
آزمایش‌های مربوط به این قسمت در ۴ بخش صورت گرفته است؛ ۱. مقایسه در سرعت و کیفیت یادگیری، ۲. مقایسه در پیچیدگی زمانی، ۳. مقایسه در میزان باروری، ۴. مقایسه تاثیر تعداد عامل‌ها میزان کیفیت و سرعت یادگیری می‌باشد.

#### سیاست انتخاب عمل «بولتزمن»

مقایسه در سرعت و کیفیت یادگیری: نتایج حاصل از اجرای الگوریتم‌ها در محیط پلکان مارپیچ در شکل ۳-۴ آمده است. در این شکل محور افقی تعداد تلاش‌های یادگیری عامل را نشان می‌دهد که در تلاش اول عامل بدون دانش اولیه شروع به تعامل با محیط می‌کند و در تلاش ۱۰۰۰ام عامل به اجرای خود پایان می‌دهد. محور عمودی نمودار میانگین تجمعی تعداد قدم‌های عامل را نشان می‌دهد. اعداد کناری برچسب‌ها (گوشه بالا سمت راست) متوسط تعداد قدم در آخرین تلاش عامل می‌باشد که انتظار می‌رود عامل آگاهی نسبی کاملی از محیط دارد را نشان می‌دهد که این عدد هرچقدر کمتر باشد نشان می‌دهد که عامل در طی رسیدن به هدف تعداد گام کمتری برداشته است و در نتیجه دانش و شناخت بهتری از محیط دارد.

همانطور که مشاهده می‌شود روش SEP دارای ۲٪ بهبود نسبت به IL می‌باشد در حالی که روش پیشنهادی در زمانی که از انتگرال فازی استفاده می‌کند در بدترین حالت دارای ۱۸٪ بهبود و در بهترین حالات دارای ۳۳٪ بهبود می‌باشد که نسبت به روش SEP تقریباً ۹ الی ۱۶ برابر نتیجه را بهبود داده است. در صورتی که از میانگین وزنی بجای انتگرال فازی استفاده شود نتایج با اختلاف اندکی (کمتر از ۱٪) بدتر از یادگیری IL بوده است که نشان می‌دهد که استفاده از انتگرال فازی چقدر می‌تواند نسبت به روش‌های سنتی و معمولی چون میانگین وزنی موثر واقع شود. نتایج این قسمت را می‌توان در جدول ۳-۲ خلاصه کرد.

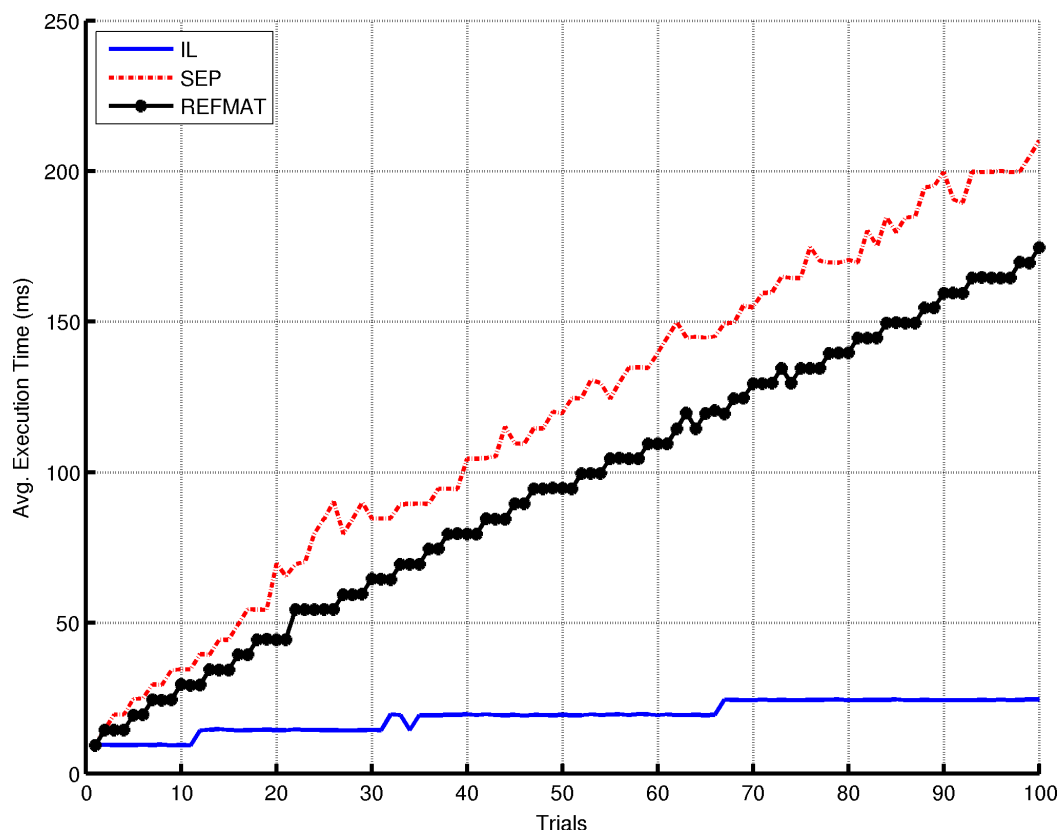
مقایسه در پیچیدگی زمانی: در این قسمت به مقایسه‌ی پیچیدگی زمانی روش پیشنهادی با روش SEP مورد بررسی قرار می‌گیرد، برای محاسبه‌ی پیچیدگی زمانی به روش ریاضی کار بسیار دشوار و پرخطایی می‌باشد؛ در اینجا ما بجای محاسبه‌ی پیچیدگی زمانی ریاضی دو الگوریتم از مدت زمانی که طول می‌کشد برنامه در سیستم اجرا و خاتمه یابد استفاده می‌کنیم. در شکل ۳-۵ میانگین زمانی ۲۰ اجرای مستقل برحسب میلی‌ثانیه به ازای هریک از تعداد تلاش‌ها آورده شده است. همان‌طور که در این شکل مشاهده می‌شود الگوریتم IL دارای حداکثر سرعت اجرا می‌باشد زیرا که هیچ سربار محاسباتی یادگیری مشترک را ندارد؛ هدف یادگیری اشتراکی این است



شکل ۳-۴: مقایسه در سرعت و کیفیت یادگیری در محیط پلکان مارپیچ با تابع بولتزمن در محیط پلکان مارپیچ

جدول ۳-۲: مقایسه در میزان بهبود کیفیت یادگیری در محیط پلکان مارپیچ با تابع بولتزمن

			REFMAT				
	IL	SEP	wsum	fci-mean	fci-max	fci-k-mean	fci-const-one
IL	%0.0						
SEP	%2.2	%0.0					
wsum	%-0.2	%-2.3	%0.0				
fci-mean	%14.9	%12.5	%15.1	%0.0			
fci-max	%18.0	%15.5	%18.2	%2.7	%0.0		
fci-k-mean	%24.0	%21.4	%24.2	%7.9	%5.1	%0.0	
fci-const-one	%33.6	%30.7	%33.8	%16.2	%13.2	%7.7	%0.0



شکل ۳-۵: مقایسه در پیچیدگی زمانی روش‌ها به ازای تعداد تلاش‌های متفاوت برحسب میلی‌ثانیه با تابع بولتزمن در محیط پلکان مارپیچ

که می‌خواهد در ازای یک سری سربار محاسباتی کیفیت و سرعت «یادگیری» عامل‌ها را افزایش دهد. با در نظر داشتن این موضوع همانطور که قبلاً دیدیم روش پیشنهادی سرعت و کیفیت یادگیری را بیشتر از روش SEP افزایش می‌دهد و در اینجا نیز می‌بینیم که دارای پیچیدگی زمانی کمتری نسبت به روش SEP می‌باشد که نشان از بهینه‌گی روش پیشنهادی نسبت به روش SEP می‌دهد.

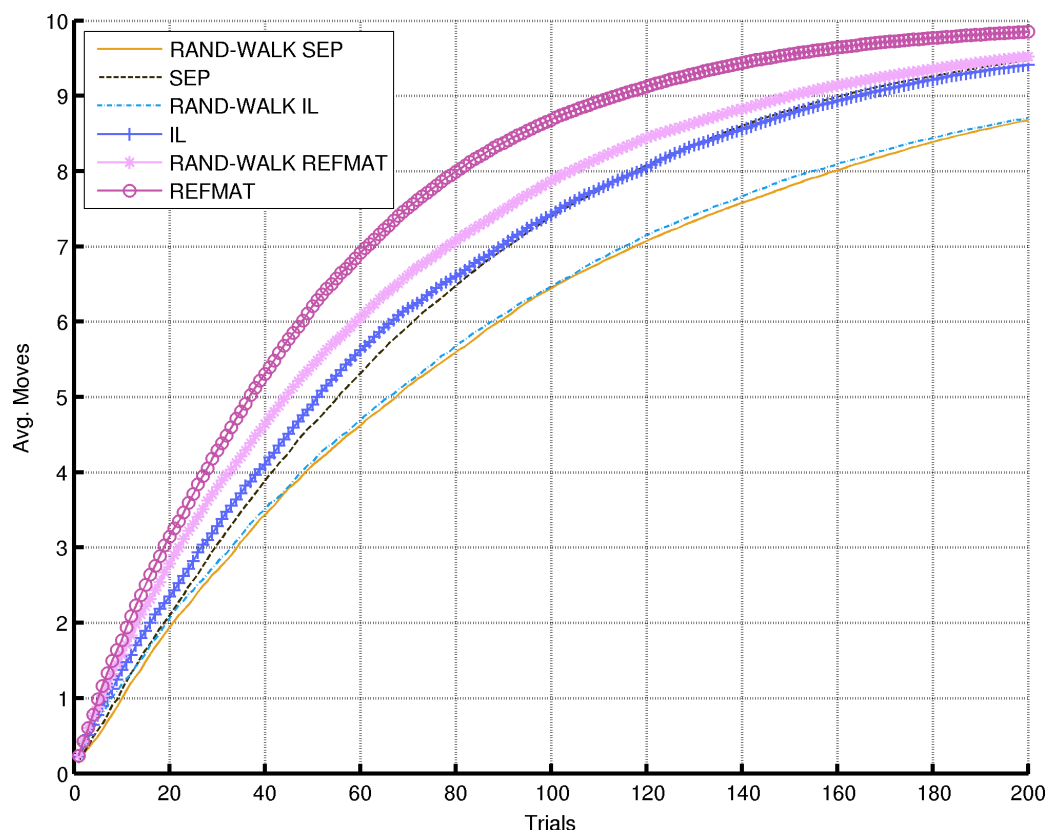
#### مقایسه در میزان باروری:

**تعریف ۳-۱** (سرعت باروری). اگر فرض کنیم الگوریتم یادگیری تقویتی  $\psi_Q(\mathcal{E})$  وجود دارد که در محیط  $\mathcal{E}$  فعالیت می‌کند و دانش خود را در جدولی مانند  $Q$  ذخیره می‌کند، سرعت باروری الگوریتم  $\psi_Q(\mathcal{E})$  را سرعت همگرایی حداکثر مقدار جدول  $Q$  به سمت حداکثر پاداش محیط قابل دریافت تعریف می‌کنیم.

**تعریف ۳-۲** (میزان باروری). انتگرال سرعت باروری را میزان باروری الگوریتم  $\psi_Q(\mathcal{E})$  که در محیط  $\mathcal{E}$  فعالیت می‌کند و دانش خود را در جدولی مانند  $Q$  ذخیره می‌کند، تعریف می‌کنیم.

**تئوری ۳-۱** (معیاری جدید برای سرعت یادگیری). طبق تعاریف ۳-۱ و ۳-۲ الگوریتمی میزان باروری بیشتری دارد که سریع‌تر مقادیر جدول  $Q$  خود را به سمت بیشینه مقداری که می‌توانند داشته باشند (یعنی بیشینه پاداشی که از محیط می‌توانند کسب کنند) سوق دهد. معمولاً این در الگوریتم‌های یادگیری تقویتی  $Q$  این کار با تنظیم مقدار سرعت یادگیری  $\alpha$  صورت می‌گیرد که باعث می‌شود





شکل ۳-۶: نمودار باروری الگوریتم‌ها مختلف با تابع بولتزمن در محیط پلکان مارپیچ

الگوریتم‌ها با سرعت بیشتری به یادگیری نحوه تعامل با محیط بپردازند. لذا در شرایط یکسان می‌توان گفت الگوریتمی بهتر عمل می‌کند که نحوه تعامل با محیط را سریع‌تر نسبت به دیگر الگوریتم‌ها یاد می‌گیرد و میزان باروری بیشتری داشته باشد.

در شکل ۳-۶ آورده شده است حداکثر میزان جدول  $Q$  روش‌ها در هر تلاش آورده شده است. همانطور که قبلاً در تعریف محیط پلکان مارپیچ آورده شده است حداکثر مقدار پاداش این محیط مقدار ۱۰ می‌باشد لذا همان‌طور که مشاهده می‌شود الگوریتم‌ها با شیب‌های متفاوتی حداکثر مقدار جدول خود را به سمت حداکثر مقدار پاداش قابل دریافت از محیط سوق می‌دهند. در این شکل سرعت باروری شیب نمودار در هر تلاش می‌باشد و میزان باروری مساحت زیر نمودار می‌باشد.

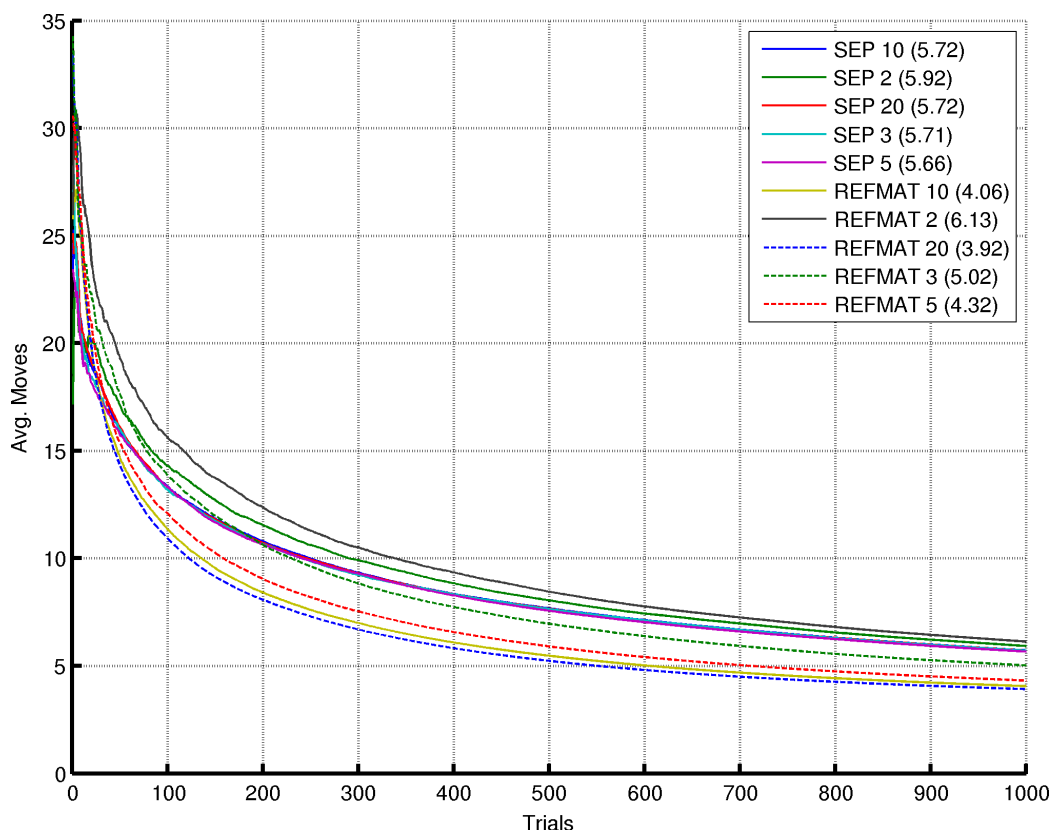
در شکل ۳-۶ منظور از RAND-WALK حرکت کاملاً تصادفی می‌باشد، به این صورت که عامل بعد از هر حرکت جدول  $Q$  خود را بروز رسانی می‌کند ولی هنگام انتخاب عمل در تابع بولتزمن مقدار  $\tau \rightarrow +\infty$  در نظر گرفته می‌شود تا میزان احتمال تمامی حرکت‌ها یکسان شود و در نتیجه حرکتی به صورت تصادفی انتخاب شود. همان‌طور که در قسمت‌های قبل دیدیم روش پیشنهادی هم در کیفیت و هم در سرعت یادگیری بهبود چشم‌گیری دارد و از طرفی هم در نمودار ۳-۶ دارای بیشترین میزان باروری (مساحت زیر نمودار) حداکثر مقدار جدول  $Q$  می‌باشد که این مساله تایید کننده تئوری ۳-۱ می‌باشد.

دلیل وجود نتایج آزمایش اجرای RAND-WALK در این قسمت این است که بررسی کنیم در صورتی که اگر عامل بصورت کورکورانه حرکت کند روش معرفی شده و SEP چقدر در میزان بارور شدن جدول  $Q$  عامل‌ها موثرند؟ به عبارت دیگر، در صورتی که استراتژی خاصی جهت انتخاب عمل وجود نداشته باشد، روش‌ها چقدر قدرت باروری دارند؟ همانطور که در شکل ۳-۶ مشاهده می‌کنیم روش معرفی شده در زمانی که به صورت تصادفی اقدام به انتخاب عمل می‌کند بیشتر از زمانی که IL با استفاده از تابع بولتزمن اقدام به انتخاب عمل می‌کند جدول  $Q$  را بارور می‌کند که از قدرت روش ارائه شده خبر می‌دهد. همچنین در مورد روش SEP می‌بینیم که در زمانی که بصورت تصادفی اقدام به عمل می‌کند باروری کمتری نسبت به روش پیشنهادی و IL دارد؛ یعنی میزان باروری روش SEP وابستگی زیادی به سیاست انتخاب عمل دارد و در صورت نداشتن سیاست انتخاب عمل خاصی بشدت عملکردش کاسته می‌شود ولی در روش پیشنهادی میزان این وابستگی از شدت کمتری برخوردار است که از دیگر امتیازات مثبت روش پیشنهادی می‌باشد.

**مقایسه تاثیر تعداد عامل‌ها میزان کیفیت و سرعت یادگیری:** در این مقایسه سعی شده است که تاثیر یک فاکتور بنیادی سیستم‌های چندعامله مشارکتی را مورد بررسی قرار دهیم، و آن میزان تاثیر پذیری روش‌های مورد مقایسه با افزایش تعداد عامل‌ها می‌باشد. در تئوری سیستم‌های چندعامله مشارکتی دیدگاه معقول براین است که اثر تعداد عامل‌ها در کیفیت و سرعت یادگیری مشارکتی باید مثبت باشد. در غیر این صورت سیستم‌های چندعامله‌ای که تعداد عامل‌ها تاثیری در خروجی سیستم نداشته باشد، دیگر ماهیت سیستم‌های چندعامله را ندارد.

همان‌طور که در شکل ۳-۷ آمده است، روش پیشنهادی و روش SEP به ازای تعداد عامل‌های ۲، ۳، ۵، ۱۰ و ۲۰ عدد به تعداد ۲۰ بار اجرا درآمده و میانگین اجراها به نمودار کشیده شده است. همانطور که می‌بینیم روش SEP در زمانی ۲۰ عامل در حال یادگیری و اشتراک گذاری دانش‌های خود هستند نسبت به زمانی که فقط ۲ عامل در حال تعامل مشارکتی با محیط هستند فقط ۳٪ در خروجی الگوریتم تاثیر مثبت داشته است. این در حالی است که در همین شرایط میزان بهبود نتیجه‌ی روش پیشنهادی ۵۶٪ می‌باشد. که نشان می‌دهد روش SEP نسبت به افزایش تعداد عامل‌ها رفتاری تقریباً خنثی از خود نشان می‌دهد درحالی که روش پیشنهادی در ازای افزایش تعداد عامل‌ها به دلیل اینکه دانش جمعی نیز افزایش می‌یابد کیفیت خروجی آن نیز بهتر می‌شود.

**نتیجه‌گیری:** نتیجه‌ای که از مقایسه‌ی روش پیشنهادی در هر چهار مقایسه‌ی بالا می‌توان گرفت این است که روش پیشنهادی بهبود چشم‌گیری به روش SEP در محیط پلکان مارپیچ و سیاست انتخاب عمل بولتزمن داده است.

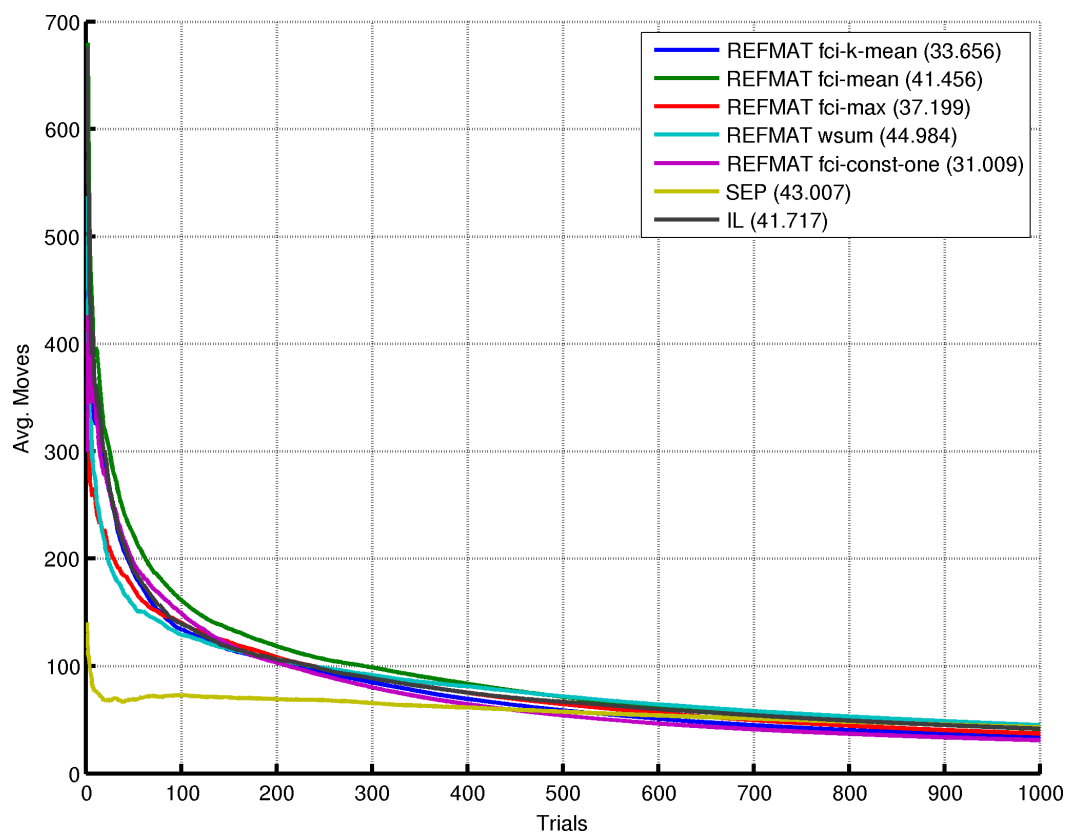


شکل ۳-۷: مقایسه تاثیر تعداد عامل‌ها میزان کیفیت و سرعت یادگیری با تابع بولتزمن در محیط پلکان مارپیچ

#### سیاست انتخاب عمل « $\epsilon$ -حریصانه»

مقایسه در سرعت و کیفیت یادگیری: نتایج حاصل از اجرای الگوریتم‌ها در محیط پلکان مارپیچ در شکل ۳-۸ آمده است. شرایط این آزمایش به مشابه شرایط آزمایش با تابع بولتزمن می‌باشد.

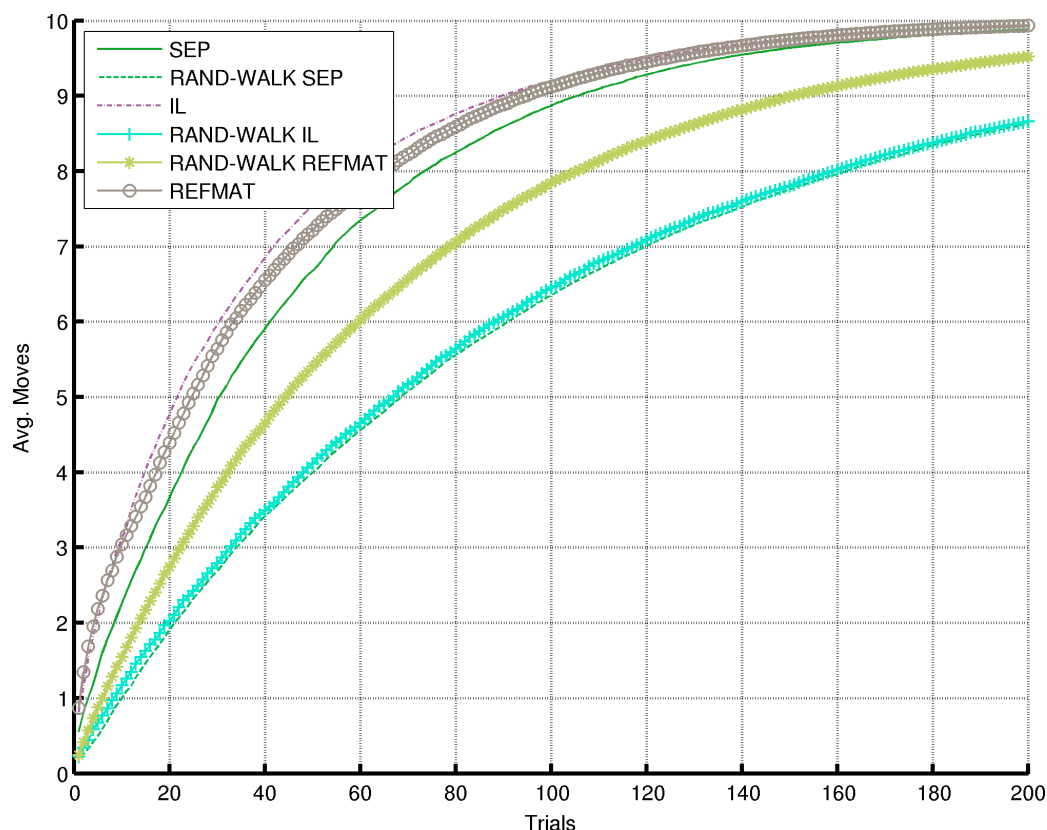
همانطور که مشاهده می‌شود روش SEP دارای ۳-٪ بهبود نسبت به IL می‌باشد در حالی که روش پیشنهادی در زمانی که از انتگرال فازی استفاده می‌کند در بدترین حالت دارای ۶/۰٪ بهبود و در بهترین حالات دارای ۳۴٪ بهبود می‌باشد که نسبت به روش SEP تقریباً ۴ الی ۳۸ برابر نتیجه را بهبود داده است. در صورتی که از میانگین وزنی بجای انتگرال فازی استفاده شود نتایج با اختلافی حدود ۷-٪ بدتر از یادگیری IL بوده است که نشان می‌دهد که استفاده از انتگرال فازی چقدر می‌تواند نسبت به روش‌های سنتی و معمولی چون میانگین وزنی موثر واقع شود. البته در شکل ۳-۸ باید توجه کرد که روش SEP در همان ابتدای کار خود به شدت میانگین حرکت عامل‌ها را کاهش داده ولی به دلیل ماهیت الگوریتم SEP اشباع جداول الگوریتم توانایی ادامه‌ی سرشکن کردن بیشتر میانگین حرکت عامل‌ها را ندارد. میانگین نتایج این قسمت را می‌توان در جدول ۳-۳ خلاصه کرد.



شکل ۳-۸: مقایسه در سرعت و کیفیت یادگیری در محیط پلکان مارپیچ با تابع حریصانه در محیط پلکان مارپیچ

جدول ۳-۳: مقایسه در میزان بهبود کیفیت یادگیری در محیط پلکان مارپیچ با تابع حریصانه

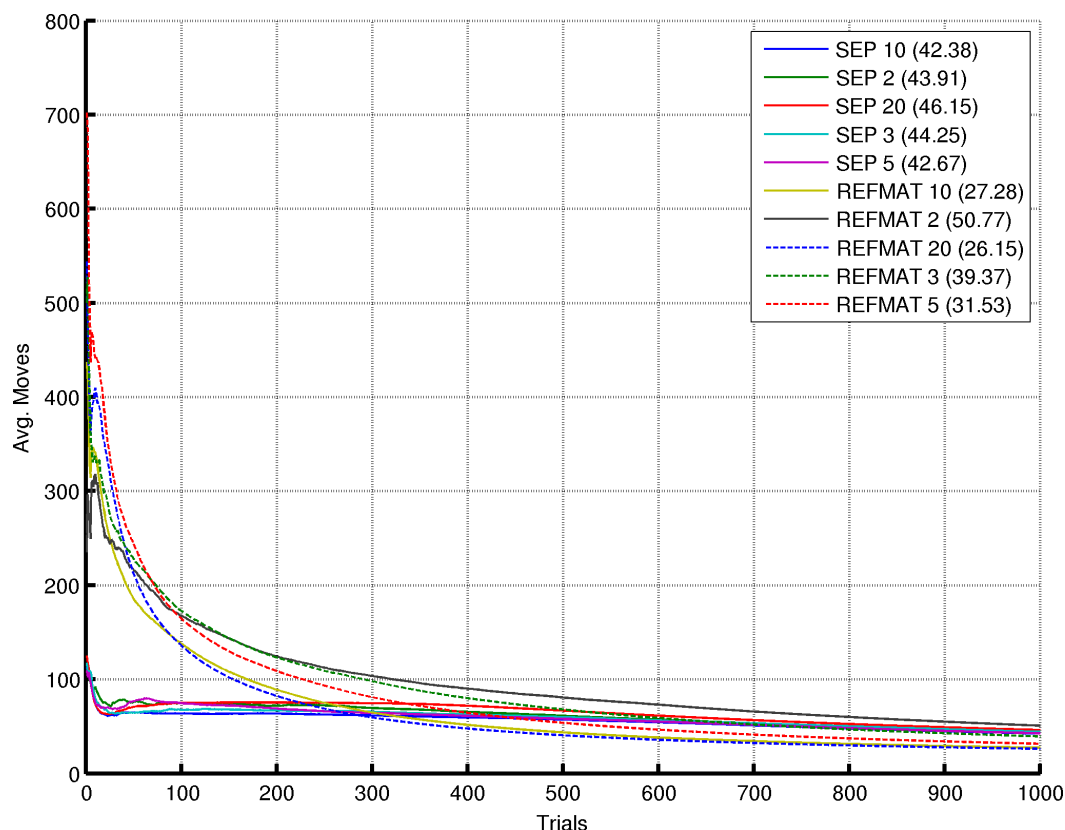
			REFMAT				
	IL	SEP	wsum	fci-mean	fci-max	fci-k-mean	fci-const-one
IL	%0.0						
SEP	%-3.0	%0.0					
wsum	%-7.3	%-4.4	%0.0				
fci-mean	%0.6	%3.7	%8.5	%0.0			
fci-max	%12.2	%15.6	%20.9	%11.5	%0.0		
fci-k-mean	%24.0	%27.8	%33.7	%23.2	%10.5	%0.0	
fci-const-one	%34.5	%38.7	%45.1	%33.7	%20.0	%8.5	%0.0



شکل ۳-۹: نمودار باروری الگوریتم‌ها مختلف با تابع حریصانه در محیط پلکان ماریچ

مقایسه در پیچیدگی زمانی: در شکل ۳-۹؟ میانگین زمانی ۲۰ اجرای مستقل برحسب میلی‌ثانیه به ازای هریک از تعداد تلاش‌ها آورده شده است. همان‌طور که در این شکل مشاهده می‌شود الگوریتم IL دارای حداکثر سرعت اجرا می‌باشد زیرا که هیچ سربار محاسباتی یادگیری مشترک را ندارد؛ هدف یادگیری اشتراکی این است که می‌خواهد در ازای یک سری سربار محاسباتی کیفیت و سرعت «یادگیری» عامل‌ها را افزایش دهد. با در نظر داشتن این موضوع همان‌طور که قبلاً دیدیم روش پیشنهادی سرعت و کیفیت یادگیری را بیشتر از روش SEP افزایش می‌دهد و در اینجا نیز می‌بینیم که دارای پیچیدگی زمانی کمتری نسبت به روش SEP می‌باشد که نشان از بهینه‌گی روش پیشنهادی نسبت به روش SEP می‌دهد.

مقایسه در میزان باروری: در شکل ۳-۹ میزان باروری IL از کلیه روش‌ها بهتر بوده (با اندک اختلاف نسبت روش پیشنهادی) ولی همچنان باروری روش پیشنهادی از روش SEP بیشتر بوده است و همچون آزمایش مشابه با تابع بولتزمن در اینجا نیز نشان داده شده است که روش SEP کاملاً وابسته است به این‌که در هنگام انتخاب عمل بر اساس دانش عامل عمل شود و اگر عامل بدون در نظر گرفتن دانش عامل حرکتی اتخاذ کند میزان باروری عامل بشدت تحت تاثیر قرار می‌گیرد در حالی که در روش پیشنهادی در شرایط یکسان از کلیه الگوریتم‌ها میزان باروری بیشتری دارد.



شکل ۳-۱۰: مقایسه تاثیر تعداد عامل‌ها میزان کیفیت و سرعت یادگیری با تابع حریصانه در محیط پلکان مارپیچ

مقایسه تاثیر تعداد عامل‌ها میزان کیفیت و سرعت یادگیری: همان‌طور که در شکل ۳-۱۰ آمده است، روش پیشنهادی و روش SEP به ازای تعداد عامل‌های ۲، ۳، ۵، ۱۰ و ۲۰ عدد به تعداد ۲۰ بار اجرا درآمده و میانگین اجراها به نمودار کشیده شده است. همان‌طور که می‌بینیم روش SEP در زمانی ۲۰ عامل در حال یادگیری و اشتراک گذاری دانش‌های خود هستند نسبت به زمانی که فقط ۲ عامل در حال تعامل مشارکتی با محیط هستند ۸-٪ در خروجی الگوریتم تاثیر منفی داشته است؛ بدین معنی که در زمانی که از تابع حریصانه استفاده شود روش SEP به افزایش تعداد عامل فقط منجر به بدتر شدن عملکرد عامل‌ها در یادگیری مشارکتی می‌شود. این در حالی است که در همین شرایط میزان بهبود نتیجه‌ی روش پیشنهادی ۹۲٪ می‌باشد. که نشان می‌دهد روش پیشنهادی در ازای افزایش تعداد عامل‌ها به دلیل اینکه دانش جمعی نیز افزایش می‌یابد کیفیت خروجی آن نیز بطور چشم‌گیری بهتر می‌شود. در حالی که در روش SEP اگر کار نتایج بدتر نشود بهتر نمی‌شود که از ضعف بزرگ روش SEP خبر می‌دهد.

نتیجه‌گیری: نتیجه‌ای که از مقایسه‌ی روش پیشنهادی در هر چهار مقایسه‌ی بالا می‌توان گرفت همچون نتیجه‌ای که از نتایج تابع بولتزمن، روش پیشنهادی بهبود چشم‌گیری به روش SEP در محیط پلکان مارپیچ و سیاست انتخاب عمل حریصانه داده است.

## مراجع

- [1] V. Torra and Y. Narukawa, "The interpretation of fuzzy integrals and their application to fuzzy systems," *International journal of approximate reasoning*, vol. 41, no. 1, pp. 43–58, 2006.
- [2] K. Leszczyński, P. Penczek, and W. Grochulski, "Sugeno's fuzzy measure and fuzzy clustering," *Fuzzy Sets and Systems*, vol. 15, no. 2, pp. 147–158, 1985.
- [3] A. F. Tehrani, W. Cheng, and E. Hullermeier, "Preference learning using the choquet integral: The case of multipartite ranking," *IEEE Transactions on Fuzzy Systems*, vol. 20, no. 6, pp. 1102–1113, 2012.
- [4] L. M. De Campos and M. Jorge, "Characterization and comparison of sugeno and choquet integrals," *Fuzzy Sets and Systems*, vol. 52, no. 1, pp. 61–67, 1992.
- [5] M. Grabisch, "Fuzzy integral in multicriteria decision making," *Fuzzy sets and Systems*, vol. 69, no. 3, pp. 279–298, 1995.
- [6] T. Murofushi, M. Sugeno, and M. Machida, "Non-monotonic fuzzy measures and the choquet integral," *Fuzzy sets and Systems*, vol. 64, no. 1, pp. 73–86, 1994.
- [7] M. Grabisch, "The application of fuzzy integrals in multicriteria decision making," *European journal of operational research*, vol. 89, no. 3, pp. 445–456, 1996.
- [8] "Expert - wikipedia." <https://en.wikipedia.org/wiki/Expert>. (Accessed on 11/12/2016).
- [9] E. Schechter, *Handbook of Analysis and its Foundations*, ch. 1, p. 16. Academic Press, 1996.
- [10] M. N. Ahmadabadi, M. Asadpur, S. H. Khodanbakhsh, and E. Nakano, "Expertness measuring in cooperative learning," in *Intelligent Robots and Systems, 2000.(IROS 2000). Proceedings. 2000 IEEE/RSJ International Conference on*, vol. 3, pp. 2261–2267, IEEE, 2000.
- [11] E. Pakizeh, M. Palhang, and M. M. Pedram, "Multi-criteria expertness based cooperative q-learning," *Applied intelligence*, vol. 39, no. 1, pp. 28–40, 2013.

- [12] M. ali mirzaei badizi, “Speed-up cooperative learning in multi-agent systems using shortest experimented path,” Master’s thesis, Department of Electrical and Computer Engineering, Isfahan University of Technology, Isfahan University of Technology, Isfahan 84156-83111, Iran, 3 2015.
- [13] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*, vol. 1. MIT press Cambridge, 1998.
- [14] V. Torra, Y. Narukawa, and M. Sugeno, *Non-Additive Measures*, pp. 3–7. Springer, 2014.



# Improvements in speed and quality of learning in multi-agent systems using the reference matrix and fuzzy integral

Dariush Hasanpour Adeh

d.hasanpoor@ec.iut.ac.ir

[DATE]

Department of Electrical and Computer Engineering  
Isfahan University of Technology, Isfahan 84156-83111, Iran

Degree: M.Sc.

Language: Farsi

**Supervisor: Assoc. Prof. Maziar Palhang (palhang@cc.iut.ac.ir)**

## **Abstract**

### **Key Words:**

Multi-agent Systems, Cooperative Learning, Reinforcement Learning, Non-additive Knowledges, Fuzzy Integral



**Isfahan University of Technology**

Department of Electrical and Computer Engineering

# Improvements in speed and quality of learning in multi-agent systems using the reference matrix and fuzzy integral

A Thesis

Submitted in partial fulfillment of the requirements  
for the degree of Master of Science

**by**

**Dariush Hasanpour Adeh**

Evaluated and Approved by the Thesis Committee, on ...

1. Maziar Palhang, Assoc. Prof. (Supervisor)
2. ..., Prof. (Examiner)
3. ..., Prof. (Examiner)

Mohamad Reza Taban, Department Graduate Coordinator

