

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 پالهنګ

۱۳۹۵

چهار



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.

دل‌تنگی‌های آدمی را باد ترانه‌ای می‌خواند

رویاهایش را آسمان پر ستاره نادیده می‌گیرد

و هر دانه‌ی برفی به اشکی نریخته می‌ماند.

سکوت سرشار از سخنان ناگفته است؛

از حرکات ناکرده،

اعتراف به عشق‌های نهان،

و شگفتی‌های به زبان نیامده،

در این سکوت حقیقت ما نهفته است؛

حقیقت تو و من.

برای تو و خویش

چشمانی آرزو می‌کنم،

که چراغ‌ها و نشانه‌ها را در ظلمات مان ببیند.

گوشی،

که صداها و شناسه‌ها 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 بندها کشیده است، سخن بگوییم.

پنجه درافکنده‌ایم با دست‌هایمان

به جای رها شدن

سنگین سنگین بر دوش می‌کشیم

بار دیگران را

به جای همراهی کردن‌شان!

عشق ما نیازمند دهان، است نه تصاحب

در راه خویش ایثار باید نه انجام وظیفه...

بی‌اعتمادی دری است

خودستایی، چفت و بست غرور است

و تهی دستی، دیوار است و لولا است

زندانی را که در آن محبوس رأی خویش ایم

دلتنگی مان را برای آزادی و دلخواه دیگران بودن

از رخنه‌هایش تنفس می‌کنیم...

- مارگوت بیگل

فهرست مطالب

صفحه	عنوان
یازده	فهرست مطالب
دوازده	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	۲-۳-۱ الگوریتم پیشنهادی
۱۳	۴-۲ علت کارکرد انتگرال فازی چوکت در انتقال دانش
۱۳	۲-۴-۱ ویژگی های انتگرال فازی چوکت
۱۳	۲-۴-۲ علت عملکرد بهتر مدل انتگرال فازی نسبت به مدل رایج مجموع وزنی؟
۱۳	مراجع
۱۶	چکیده انگلیسی

فهرست تصاویر

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^۱ و انتگرال فازی داریم که با هدف جمع آوری اطلاعات^۲ 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]:

^۱Fuzzy measures

^۲Aggregate 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- λ سوگنو^۱ 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 λ 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- λ سوگنو باید توجه کرد این است که در صورتی که $\lambda > 0$ باشد آنگاه خروجی تابع همیشه بزرگتر از مجموع ارزش تک تک منابع اطلاعاتی ورودی خواهد بود. این ویژگی باعث می شود که این تابع در بعضی از کاربردها (از جمله کاربرد مورد نظر در این پژوهش) کارایی نداشته باشد. انتگرال فازی در واقع یک تعمیمی به روش میانگین وزنی^۲ 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^۳ و چوکت^۴ 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

¹ Sugeno λ -Measure

² Weighted Arithmetic Mean

³ Sugeno

⁴ Choquet

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

$$\int_s h \circ g = S_g(h) = \bigvee_{i=1}^n h(x_{\pi_i}) \wedge g(A_i) \quad (5-1)$$

در انتگرال سوگنو لازم است که مقادیر منابع اطلاعاتی را مرتب کنیم که π عملگر جایگشت می‌باشد.

$$h \xrightarrow{\pi} h(\pi_1) \geq h(\pi_2) \cdots \geq h(\pi_n) \quad (6-1)$$

$$A_i = \{x_{\pi_1}, x_{\pi_2}, \dots, x_{\pi_i}\} \quad (7-1)$$

و انتگرال فازی چوکت به صورت زیر تعریف می‌شود [۶]:

$$\int_c h \circ g = C_g(f) = \sum_{i=1}^n (f(x_{\pi(i)}) - f(x_{\pi(i-1)})) \cdot g(A_i) \quad (8-1)$$

در رابطه‌ی بالا $f: X \rightarrow \mathbb{R}$ می‌باشد که یکی از وجه تمایز آن با انتگرال سوگنو می‌باشد و همچنین A_i و π همانند آنچه که در ۶-۱ و ۷-۱ آمده است می‌باشد با این تفاوت که بجای تابع h تابع f استفاده شده است. 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^۲ مناسب باشد؛ بدین معنی که تجمیع اعداد کاردینال^۳ (که اعداد دارای مفاهیم واقعی هستند) را انتگرال چوکت بهتر مدل می‌کند در حالی انتگرال سوگنو برای اعداد ترتیبی^۴ مناسب عمل می‌کند [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 می‌کند.

¹Support

²Positive Linear Transformation

³Cardinal Aggregation

⁴Ordinal 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^۱ را ارائه می‌دهد.

^۱ 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^۱ 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.

¹Interested

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» (یا به اختصار REFMAT^۱) 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 (یا به اختصار RCMAT^۲) 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)

^۱Reference Matrix

^۲Recall 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:        $r = c$ ;
5:     else
6:        $c = r$ ;

```

الگوریتم ۳ تابع FCI_Combiner معرفی شده در الگوریتم ۱

```

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

```

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 ۲-۱ 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^۲ 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.

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^۳ 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 ۲-۱ باشد تا خروجی انتگرال فازی همجنس مقادیر \vec{x} باشد.

$$f(w) = aw + b \quad (۱-۲)$$

متغیرهای a و b در ۲-۱ می‌تواند به عنوان پارامترهای سازگار^۴ در میزان کیفیت جدول Q مشارکتی خروجی 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

¹Factors

²Normalize

³Scalar

⁴Adaptive Parameters

نظر گرفته شده‌اند که یعنی از تابع همانی به عنوان تابع $f(\cdot)$ استفاده شده است.

۲-۴ علت کارکرد انتگرال فازی چوکت در انتقال دانش

۲-۴-۱ ویژگی‌های انتگرال فازی چوکت

هیچ وقت از حدود تجاوز نمی‌کند

مناسب برای اندازه‌گیری‌های غیرافزایشی

۲-۴-۲ علت عملکرد بهتر مدل انتگرال فازی نسبت به مدل رایج مجموع وزنی؟

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.

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

