کنترل عاطفی تفاوت زمانی[۱] سیستم‌های چند متغیره[۲]

جواد عبدی
فارغ التحصیل کارشناسی ارشد کنترل- دانشکده فنی- دانشگاه تهران
j.abdi@ece.ut.ac.ir
کارولوکس
استاد گروه مهندسی برق و کامپیوتر- دانشکده فنی- دانشگاه تهران
Lucas@ipm.ir
علی خاکی صدیق
استاد گروه مهندسی برق و کامپیوتر- دانشکده فنی- دانشگاه خواجه نصیرالدین طوسی
E-mail: Sedigh@eetd.kntu.ac.ir
مهرداد فتوره چی
دانشجوی دکتری دانشکده پردازش تصویر - دانشگاه بریتیش کلمبیا، ونکوور - کانادا
mehrdadf@ece.ubc.ca
(تاریخ دریافت ۸۱/۹/۲۷، تاریخ تصویب ۸۲/۱۱/۴)

چکیده

در این مقاله رویکردی عاملگرا برای کنترل سیستم‌های با اهداف چندگانه ارائه شده است. اصول این روش مبتنی یادگیری عاطفی و یادگیری تفاوت زمانی بوده و دارای ساختار فازی - عصبی[۳] می‌باشد. روش پیشنهادی می‌تواند با توجه به موقعیت فعلی، عملکرد سیستم در زمان‌های گذشته و اهداف کنترلی موجود، سیستم را به گونه‌ای کنترل نماید که این اهداف در حداقل زمان و به نحو بسیار مطلوبی برآورده شوند.

واژه‌های کلیدی: یادگیری عاطفی، یادگیری تفاوت زمانی[۴]، تخصیص اعتبار[۵]، نقاد فازی[۶]، عامل[۷]، سیستم‌های چند متغیره، ساختار فازی عصبی

مقدمه

از حدود دهه ۱۹۵۰، مهندسان کنترل با تحلیل سیتم‌هایی روبرو شدند که همزمان دارای چند ورودی و چند خروجی بوده در نتیجه روش‌های تحلیل پایداری سیستم‌های تک ورودی - چند خروجی قابل اعمال به آنها نیستند. مساله مشکلتر، طراحی یک سیستم کنترل برای سیستم‌های چند متغیره است که بتواند مشخصه‌های مطلوب سیستم حلقه بسته همانند ردیابی ورودی‌های مرجع، حذف اغتشاش و مقاوم بودن نسبت به تغییر پارامترهای سیستم را برآورده سازد.

یکی از کارهایی که هر موجود هوشمندی دائما در حال انجام آن است فرآیند استدلال می‌باشد. در عین حال از مهمترین مراحل هر استدلالی مرحله نگاشت است که در طی آن الگوی ورودی به الگوهای موجود در حافظه دائم عامل هوشمند تطبیق داده می‌شود تا شبیه‌ترین الگوها به عنوان پاسخ برگزیده شوند. لازمه این کار آن است که تمامی تجارب پیشین عامل هوشمند در حافظه فعال او وجود داشته باشند. علاوه بر این باید عامل هوشمند در یک زمان بسیار محدود قادر به مقایسه الگوی ورودی با تمامی تجارب پیشین باشد. از طرفی یک عامل هوشمند واقعی معمولا در یک محدوده زمانی مشخص، دارای چندین هدف فعال می باشد که گاها بعضی از این اهداف ممکن است در تضاد با هم باشند.

آنچه در این میان ضروری به نظر می‌رسد این است که عامل باید از بین این اهداف جاری زیر مجموعه‌ای را انتخاب نماید که با توجه به موقعیت فعلی مناسب‌ترین گزینه بوده و در حداقل زمان تمامی این اهداف به نحو مطلوبی برآورده شوند. برای رسیدن به این منظور، رفتار عامل هوشمند باید مبتنی داده‌های گذشته

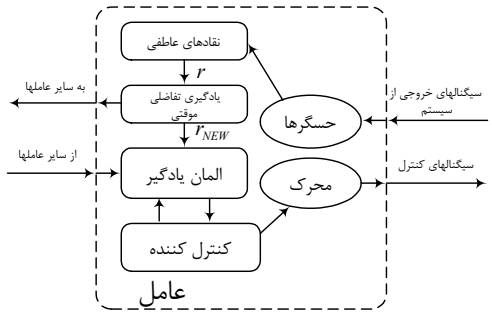
باشد تا بتواند قابلیت پیش‌بینی عامل را از محیط (به منظور برآورده ساختن اهداف خود) افزایش دهد. یادگیر تفاوت زمانی باعث می‌شود که عامل هوشمند بتواند با توجه به موقعیت فعلی محیط، عملکرد سیستم در زمان گذشته، مجموعه اهدافی که باید برآورده شوند و میزان برآورده شدن این اهداف تا کنون، برای هر یک از اهداف به صورت محلی اولویت‌هایی را در نظر بگیرد تا اینکه در حداقل زمان ممکن عملکرد سیستم بهینه شود. این اولویت‌بندی معمولاً به صورت وزن دادن به مجموعه اهدافی است که باید برآورده شوند. منظور از محلی بودن اولویت ها این است که وزن‌هایی که به اهداف تخصیص داده می‌شود بر اساس شرایط فعلی سیستم و میزان برآورده شدن هر یک از این اهداف می‌باشد و در هر لحظه زمانی این وزنها بسته به شرایط تغییر می‌کنند تا اینکه تمامی این اهداف در حداقل زمان به نحو مطلوبی برآورده شوند[۴-۱].

ساختار کلی این مقاله بدین صورت می‌باشد که ابتدا در بخش سوم کنترل مبتنی بر عامل، مورد بررسی قرار می‌گیرد. بخش چهارم به موضوع یادگیری عاطفی به عنوان یک روش کنترلی هوشمند پرداخته است. در بخش پنجم کنترل عاطفی تفاوت زمانی معرفی و مورد بحث قرار می‌گیرد. در بخش ششم ساختار جدید نقاد همراه با یادگیری تفاوتی زمانی ارائه شده و سپس معادلات توربین گازی و ستون تقطیر ساده شده که بدلیل تداخل زیاد از جمله مسائل مهم کنترلی می‌باشند معرفی شده و سپس در نهایت به بررسی عملکرد ساختار کنترلی ارائه شده در این مقاله پرداخته و به مقایسه نتایج عملی بدست آمده خواهیم پرداخت.

کنترل مبتنی بر عامل

شکل (۱) ساختار عامل کنترلی مورد استفاده در مقاله را نشان می‌دهد. عامل از طریق حسگرهایش، سیگنال‌های خروجی سیستم را حس نموده، تصویری از وضعیت کنترلی محول شده به خود به دست می‌آورد و از طریق محرک، سیگنال کنترلی لازم را به سیستم اعمال می‌نماید کنترل کننده نیز وظیفه نگاشت این سیگنال‌های ورودی به سیگنال‌های خروجی ذکر شده را بر عهده دارد.

عنصر یادگیری، دانش ذخیره شده در کنترل کننده را با توجه به یک شاخص عملکرد خارجی، به روز کند، به طوری که تلاش آن در جهت بهینه سازی عملکرد عامل صورت خواهد پذیرفت. مبنای عملکرد عنصر یادگیری، سیگنالی است که به وسیله واحد نقاد در اختیار آن قرار می‌گیرد. نقاد، وظیفه نقد و بررسی سیستم را بر عهده داشته، با توجه به نحوه عملکرد عامل محلی مربوطه و داده های قبلی که در اختیار یادگیر تفاوت زمانی قرار می دهد، سیگنال مناسب را که حاکی از عملکرد خوب یا بد سیستم کنترل است، فراهم می‌آورد. این سیگنال، مبنای کار عنصر یادگیری در به روزآوری ساختار کنترل کننده خواهد بود. سایر عاملها نیز می‌توانند از این سیگنال در به روزآوری دانش ذخیره شده در خود استفاده نمایند، به دلیل طبیعت غیر قطعی محیط، ساختارکنترل کننده به صورت فازی - عصبی انتخاب شده است. عنصر یادگیر نیز دارای قابلیت یادگیری عاطفی می‌باشد[۵،۶] .

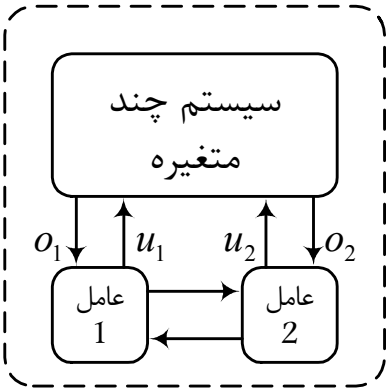


شکل ۱ : ساختار عامل کنترلی.

دیاگرام بلوکی ساختار کنترلی پیشنهادی برای سیستم‌های چند متغیره و با استفاده از مفهوم عامل، در شکل (۲) آمده است. در این ساختار برای هر متغیر ورودی یک عامل کنترلی در نظر گرفته می‌شود که وظیفه این عامل، فراهم کردن سیگنال کنترلی مناسب برای ورودی متناظرش است. ضمن اینکه هر عامل با عامل‌های دیگر، تبادل اطلاعات نموده، آنها را در امر کنترل یاری می‌دهد[۶].

یادگیری عاطفی

در مبحث یادگیری عاطفی در سیستم‌های کنترل چنین فرض می شود که معادلات سیستم بـــــرای کنترل



شکل ۲ : سیستم کنترل چند متغیره مبتنی بر عامل .

کننده شناخته شده نیستند وتنها اطلاعاتی که در دسترس می‌باشند عبارتند از حالت‌های سیستم و پسخوری از میزان عملکرد کنترل کننده به صورت یک سیگنال پیروزی یا شکست[۷]. علیرغم این کمبود اطلاعات سیستم کنترل باید یاد بگیرد که چگونه این سیستم ناشناخته را از حالت فعلی به حالت مطلوب که انتظار می‌رود بهتر باشد برساند[۸]. کنترل کننده با سعی وخطا در فضای جستجو و دریافت سیگنال پاداش (یا تنبیه) برای عملی که انتخاب کرده است باید به گونه ای عمل نماید که سیگنال پاداش بیشینه گردد. این در حالیست که با توجه به این که به سیستم گفته نمی‌شود که چه عملی را برگزیند خود آن باید با امتحان کردن اعمال گوناگون عملی را که بیشترین میزان پاداش را در پی خواهد داشت انتخاب نماید. اگر عمل انتخابی پاسخ مناسبی به دست دهد پاداشی به آن تعلق می‌گیرد تا احتمال اینکه آن عمل تکرار شود و در نتیجه سیستم پاداش بیشتری دریافت نماید افزایش یابد از طرف دیگر اگر سیستم به یک حالت نامطلوب برسد یک جریمه برای آن در نظر گرفته می شود. بنابر این در صورتی که سیستم به یک حالت مطلوب برسد تمایل به تولید مجدد آن حالت افزایش خواهد یافت [۹].

کنترل کننده‌های عاطفی تفاوت زمانی

شکل (۳) ساختار کلی کنترل کننده‌های عاطفی تفاوت زمانی را نشان می‌دهد همانطور که مشاهده می‌شود ساختار پیشنهادی شامل بخش‌های زیر می‌باشد:

نقاد: که سیگنال عاطفی را بر حسب وضعیت کنترلی تدارک می‌بیند.

کنترل کننده فازی-عصبی: که سیگنال کنترلی را برای دستگاه فراهم می‌آورد.

کنترل کننده تفاوت زمانی: که با توجه به سیگنال پاداش جاری و پاداش‌های قبلی به هر یک از سیگنال‌های کنترلی وزن‌هایی را اختصاص می‌دهد.

اکنون به بررسی اجمالی ساختار هر یک از بخش های فوق می پردازیم.

ساختار نقاد همراه با یادگیری تفاوت زمانی

ساختار نقاد خروجی

تعریف نقاد و طراحی ساختار آن، بستگی مستقیم به آن بخش از سیستم کنترل دارد که نقاد، وظیفه نقد آن را بر عهده گرفته است. در این مقاله نقاد طراحی شده وظیفه ارزیابی سیگنال خطا به همراه سیگنال تلاش کنترلی را به عهده دارد و از این رو آنرا نقاد چند کاره[۸] می‌نامیم. این نقاد پس از ارزیابی خروجی، سیگنال عاطفی r را تولید می‌کند که پیوسته بوده و هر مقداری را بین ۱- و ۱+ اخذ می‌کند، به طوری که r=+1 (یا r=-1) نشان دهنده شکست کامل کنترل کننده بوده و هر چه سیگنال عاطفی به صفر نزدیکتر باشد، موید آن است که تلاش کنترلی موفقیت آمیز بوده است. در اینجا سیستم کنترل به منظور یادگیری مجدد، منتظر یک شکست کامل نمی‌ماند، بلکه در همان زمانی که سیگنال کنترلی را اعمال می‌کند، به فرآیند یادگیری خود نیز ادامه می‌دهد. چون ارزیابی پیوسته وضعیت فعلی بر حسب امکان پیروزی یا شکست کامل، دیگر نوع ساده پیوند شرطی سازی نمی‌باشد، بلکه به تعریف تغییر حالت شناختی و کنترل تطبیقی نزدیک است، از آن به عنوان یادگیری عاطفی نام برده می‌شود.

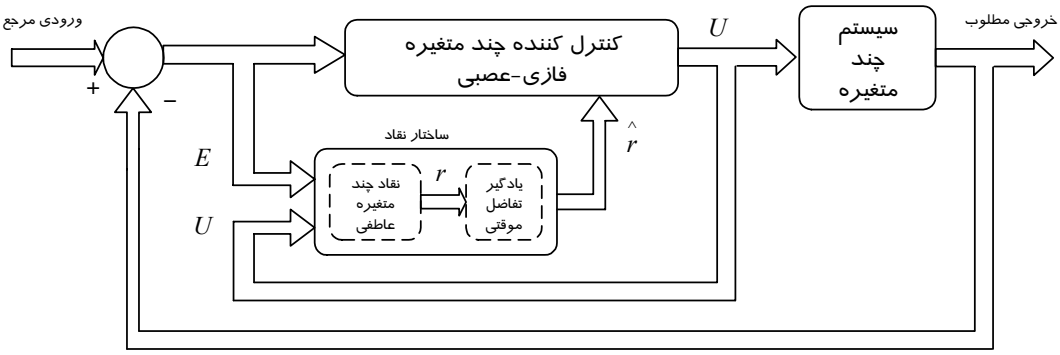
ساختار این نقاد، مشابه یک کنترل کننده فازی PD با پنج برچسب برای هر ورودی{ PL(مثبت بزرگ)، PS (مثبت کوچک)، ZE(صفر)، NS(منفی کوچک) و NL (منفی بزرگ)} و هفت برچسب برای خروجی {PL(مثبت بزرگ)، PM (مثبت متوسط)، PS(مثبت کوچک)، ZE (صفر)، NS(منفی کوچک)،NM (منفی متوسط) و NL (منفی بزرگ)} انتخاب شده است. در این مقاله مجموعه نقاد ارائه شده دارای چهار نقاد بوده که مجموع خروجیها

با توجه به وزنهای اختصاص داده شده به آنها در ساختن سیگنال عاطفی اولیه نقش دارند[۲].

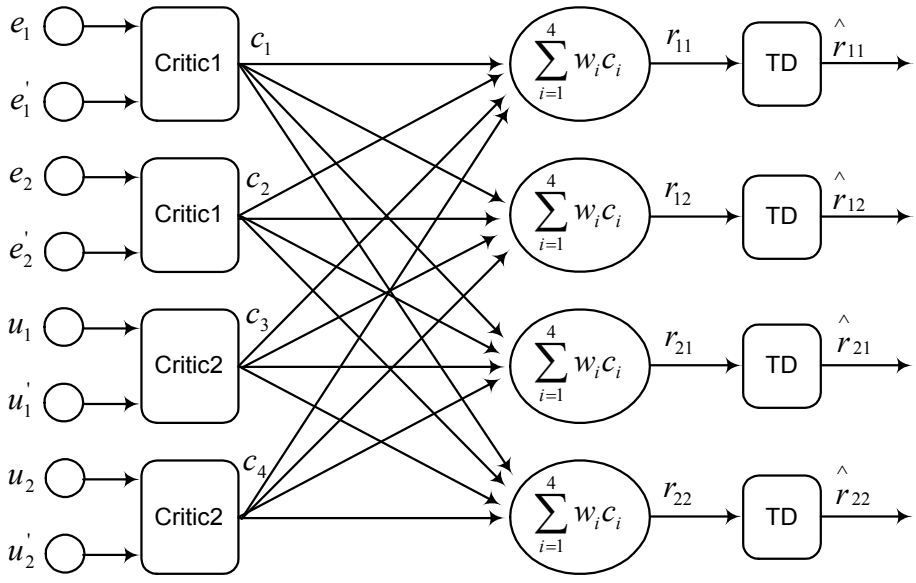
نمای کلی این نقاد در شکل (۴) نشان داده شده است. دو نقاد وظیفه نقد خطای اول و دوم را به عهده دارند که ورودی‌های آنها به ترتیب عبارتند از خطای خروجی اول و مشتق آن و خطای خروجی دوم و مشتق آن و خروجی‌های متناظر آنها به ترتیب عبارتند از C_1 و C_2، و دو نقاد دیگر وظیفه نقد سیگنال تلاش اول و دوم را به عهده دارند که ورودی‌های آنها به ترتیب عبارتند از سیگنال تلاش اول و مشتق آن و سیگنال تلاش دوم و مشتق آن و خروجیهای متناظر آنها به ترتیب عبارتند از C_3 و C_4، و مجموع وزندار این سیگنال‌ها، سیگنال‌های عاطفی r_{11}، r_{12}، r_{21} و r_{22} را برای یادگیر تفاوت زمانی مهیا می‌کنند. جدول قواعد این نقاد در جدول (۱) آمده است. عمل استنتاج به وسیله قاعده ماکزیمم - ضرب صورت می‌پذیرد و برای فازی زدایی، از قاعده مرکز ثقل استفاده شده است [۱۰].

ساختار یادگیر تفاوت زمانی

الگوریتم‌ها الزاما به یادگیری توابع مقدار معین از حالت‌ها یا زوجی از حالت-عمل متکی هستند که کاربرد آنها را نسبت به پاداش‌های نزولی آینده برآورد می‌نماید. سیاست عامل یا مطلقاً از این ارزیابی‌ها استخراج می‌شود و یا از طریق یک تابع ارزیابی ارائه می‌گردد.



شکل ۳ : ساختار کلی کنترل کننده های عاطفی تفاوت زمانی.



شکل ٤ : ساختار نقاد طراحی شده .

جدول ۱ : قواعد نقاد فازی.

الگوریتم‌های یادگیری تقویتی[۹] (RL) [۱۳-۱۱] بر مبنای روش‌های تفاوت زمانی(TD) [۱۴] در مورد تعامل اصلی بین عامل یاد گیرنده و محیط پیرامونی آن بکار می رود. در هر گام زمانی t، عامل حالت فعلی محیط، x_t، را مشاهده و عمل a_t را انجام می‌دهد. سپس یک ارزش یا پاداش تقویتی r_t دریافت می‌دارد، و حالت محیط به x_{t+1} تغییر می‌یابد. منظور از یادگیری شناخت بدون هیچ دانش قبلی از محیط است، که با یک سیاست تصمیم‌گیری (به عنوان مثال، یک ترسیم از حالت‌ها به عمل‌ها) منجر به افزایش مقدار ضریب تقویت، که عامل در طول دورهٔ زندگی‌اش دریافت می‌نماید، می‌گردد.

$$E\left[\sum_{t=0}^{\infty}\gamma^{t} r_t\right]$$

(۱)

که در آن $\gamma \in [0,1]$ یک ضریب کاهش می‌باشد که اهمیت نسبی پاداش‌های بلند مدت را نسبت به پاداش‌های کوتاه مدت تنظیم می‌کند.

گسترده‌ترین مطالعات الگوریتم‌های یادگیری تقویتی بر اساس روش تفاوت زمانی توسط ساتن[۱۰] [۱۴]، انجام شده است که معروف به $TD(\lambda)$ است. این قاعده عمومی یادگیری تقویتی بر مبنای یادگیری تفاوتی زمانی به صورت زیر نوشته شود:

$$update^{\beta}\left(U, x_t, r_t + \gamma U_t(x_{t+1}) - U_t(x_t)\right)$$

(۲)

که در آن، U تابع ارزیابی حالت[۱۱] است. این تابع به هر حالت x یک مقدار تخمینی از مجموع پاداش‌های تقویتی داده شده از ابتدای حالت تا سیاست جاری ارائه می‌نماید. در معادله (۲)، $U(x_t)$ برای پیش‌بینی مجموع تنزیلی پاداش‌های آینده می‌باشد:

$$z_t = \sum_{k=0}^{\infty}\gamma^{k} r_{t+k}$$

(۳)

که z_t پاسخ برگشتی TD برای زمان t می‌باشد. علائم بروز[۱۲] رسانی که در معادله (۲) بکار رفته است، به این معنی است که مقدار تابع U برای بحث x_t می‌بایست با استفاده از مقدار خطا $\Delta = r_t + \gamma U_t(x_{t+1}) - U_t(x_t)$ تنظیم گردد.

یعنی، تنظیمات $U_t(x_t) + \Delta$، به توجه به نرخ یادگیری β کنترل می‌شود.

$$update^{\beta}\begin{pmatrix} Q, x_t, a_t, r_t + \\ \gamma \max_a Q_t(x_{t+1}, a) - Q_t(x_t, a_t) \end{pmatrix}$$

(۴)

که در آن مقدار Q که به هر زوج حالت-عمل (x, a) نسبت داده شده، برای پیش‌بینی مقدار تقویت دریافت شده بعد از اجرای عمل a بر روی حالت x و با تعقیب یک سیاست حریصانه[۱۳] برای رسیدن به مقادیر Q فعلی. وقتی که مقادیر Q بهینه یاد گرفته شود، یک سیاست حریصانه در ارتباط با آنها، یک سیاست بهینه است.

در واقع، قانون عمومی ارائه شده در معادله (۲) با ساده‌ترین شکل روش‌های تفاوت زمانی TD(0) مطابقت دارد. برای مقادیر کلی از ضریب λ جدید قانون به روز آوری TD(λ)، برای هر حالت x در گام زمانی t توسط فرمول زیر اعمال می‌شود:

$$update^{\beta}\left(U, x, \left(r_t, \gamma U_t(x_{t+1}) - U_t(x_t)\right) e_x(t)\right)$$

(۵)

که در آن $e_x(t)=\sum_{k=0}^{t}(\gamma\lambda)^{t-k}\chi_x(k)$ مقدار اثر شایستگی[۱۴] [7][17] برای هر حالت x در زمان t، و مقدار $\chi_x(t)$ برای $x=x_t$ مساوی ۱ و در بقیه موارد مساوی صفر می باشد. آثار شایستگی برای تمام حالات در هر گام زمانی بر طبق قاعده به روزآوری زیر، بروز می‌گردند:

$$e_x(t):=\gamma\lambda e_x(t-1)+\chi_x(t)$$

(۶)

که در آن به صورت قراردادی $e_x(0)=\chi_x(t)$ می‌باشد.

برای λ=0 فقط یک حالت در گام زمانی t مقدار غیر صفر دارد و معادله شماره (۵) به معادله (۲) خلاصه می‌شود. برای λ مثبت، فرد می‌بایست در هر گام زمانی پیش‌بینی‌ها و آثار شایستگی را برای تمام حالات بروز در آورد. به همین دلیل است که این پیاده سازی با استفاده از λ>0 از نظر محاسباتی بسیار گرانتر از وقتی که λ=0 مورد استفاده قرار می‌گیرد می‌باشد، به ویژه برای کارهایی با فضای حالت بزرگتر. در هر صورت، استفاده از λ مثبت اغلب به یادگیری منجر می‌گردد که به صورت قابل ملاحظه‌ای سریعتر است. در مقالات [۱۱،۱۴،۱۵،۱۹] تکنیک ساده‌تری بنام TTD[۱۵] ارائه شده است که (تفاوت زمانی تخلیص شده)[۱۶] که استفاده از λ عمومی با هزینه‌های محاسباتی پایین را میسر می‌نماید. در شبیه‌سازی حاضر برای سادگی محاسبات از تکنیک TD(0) برای اعتبار بخشیدن به پاداش‌های گذشته استفاده شده است.

نتایج شبیه سازی

یکی از مثال‌های معروف در مبحث کنترل چند متغیره که دارای تداخل بالایی می‌باشد، کنترل توربین گازی است. ورودی‌های این سیستم به ترتیب، تحریک پمپ بنزین[۱۶] بر حسب mAmps و تحریک محرک نازل[۱۷] بر حسب ولت می‌باشند. خروجی‌های سیستم عبارتند از سرعت ژنراتور گازی[۱۸] و حرارت داخلی توربین[۱۹]. با اعمال تبدیل لاپلاس به مدل خطی شده معادلات دیفرانسیل غیر خطی سیستم، حول نقطه کار 80% سرعت ژنراتور گازی، ماتریس تابع تبدیل سیستم بصورت زیر در خواهد آمد [۱۸]:

$$G(s)=\begin{bmatrix}\dfrac{0.806s+0.264}{s^2+1.15s+0.202} & \dfrac{-(15s+1.42)}{s^3+12.8s^2+13.6s+2.36}\\ \dfrac{1.95s^2+2.12s+0.49}{s^3+9.15s^2+9.39s+1.62} & \dfrac{7.14s^2+25.8s+9.35}{s^4+20.8s^3+116.4s^2+111.6s+18.8}\end{bmatrix}$$

(۷)

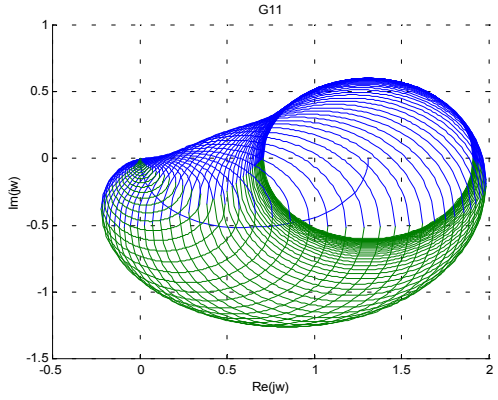
این سیستم ناپایدار و دارای تداخل بسیار بالایی می‌باشد. هدف بدست آوردن پاسخ پله در هر دو خروجی و با زمان برخاست کمتر از ۰/۳ ثانیه در خروجی اول وکمتر از ۱/۵ ثانیه در خروجی دوم است، از این رو فیلترهای ورودی بصورت زیر انتخاب شدند:

$$H_1(s)=\frac{306.25}{s^2+35s+306.25}$$

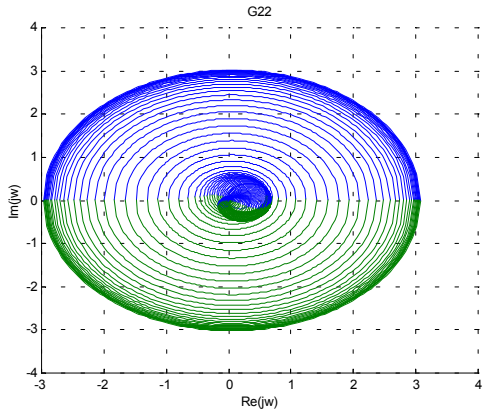
$$H_2(s)=\frac{7}{s^2+5s+7}$$

(۸)

برای نشان دادن کوپلینگ توربین گازی از تعریف باندهای گرشگورین استفاده می‌کنیم. دوایر گرشگورین توابع تبدیل G11 و G22 به ترتیب در شکل‌های (۵) و (۶) نشان داده شده است.



شکل ۵ : باندهای گرشگورین G11.



شکل ٦ : باندهای گرشگورین G22.

ساختار پیاده‌سازی برای شبیه‌سازی‌ها که توسط نرم افزار Matlab 5.3.1 صورت گرفته است در شکل (۷) نشان داده شده است.

کنترل توربین گازی

نتایج شبیه سازی برای پاسخ پله 1rpm برای سرعت سیستم و ورودی 1 درجه کلوین برای حرارت در شکل (۸) نشان داده شده‌اند که نشان از دست‌یابی به پاسخ پله بسیار خوب و تضعیف عالی تداخل دارند که بسیار برتر از نتایجی است که در مراجعی نظیر [۱۸] آمده‌اند. همچنین سیگنال‌های تلاش کنترلی، عاطفی و خطای سیستم نیز به ترتیب در شکل‌های (۹)، (۱۰) و (۱۱) ترسیم شده‌اند. با پیاده سازی روش آرائه معکوس[۲۰] نشان دهنده این است که سیگنال کنترل در روش عاطفی به مراتب کوچکتر از این روش کلاسیک می‌باشد.

کنترل مدل ساده شده یک ستون تقطیر

در این مثال، به کنترل مدل ساده شدهٔ یک ستون تقطیر خواهیم پرداخت. ماتریس تابع تبدیل این سیستم به‌صورت زیر می‌باشد:

$$Y(s) = G(s).U(s),$$

(۹)

$$G(s) = \begin{bmatrix} \dfrac{5}{9s^2+6.1s+1} & \dfrac{-1}{0.04s^2+0.4s+1} \\ \dfrac{-15}{23.52s^2+9.8s+1} & \dfrac{5}{0.16s^2+0.8s+1} \end{bmatrix}$$

(۱۰)

که در اینجا y_1, u_1, u_2 و y_2 متغیرهای ذیل را نمایش می‌دهند:

نرخ باز شارش	u_1
ورودی گرما	u_2
غلظت محصول	y_1
نرخ جریان	y_2

این سیستم دارای قطب و صفر انتقال پایدار است و با ترسیم باندهای گرشگورین آن مشخص می‌گردد که میزان تداخل بین عناصر g_{11} و g_{12} بسیار بالاست، ولی تداخل بین عناصر g_{21} و g_{22} در حد متوسط می‌باشد.

دوایر گرشگورین توابع تبدیل G11 و G22 ستون تقطیر به ترتیب در شکل‌های (۱۲) و (۱۳) نشان داده شده است.

در این مثال خاص، ژاکوبین عناصر روی قطر فرعی برابر با (۱-) انتخاب شد تا نتایج قابل قبولی بدست آید. هدف از طراحی دستیابی به پاسخ پله بدون فراجهش و زمان برخاست کمتر از یک ثانیه در هر دو خروجی است. مشکل بزرگی که در اینجا با آن روبرو می‌شویم، تنظیم ضرایب ورودی کنترل کننده عاطفی تفاوت زمانی به نحوی است که پاسخ مطلوب در خروجی ظاهر گردد. برای شبیه‌سازی این قسمت ابتدا ورودی پله را از یک فیلتر پائین گذر عبور می‌دهیم تا فرکانس‌های بالای آن حذف شوند و یک سیگنال پله هموار[۲۱] در ورودی داشته باشیم، TDMELIC[۲۲] بخوبی قادر به دنبال کردن این سیگنال خواهد بود. مشخصات این فیلتر از روی خواسته‌های مسأله (از قبیل میزان فراجهش، زمان نشست، زمان برخاست و...) در خروجی انتخاب می‌گردد. با توجه به موارد معین شده، مشخصات فیلترهای ورودی بطور تقریبی به‌صورت زیر تعیین می‌گردند:

$$H(s) = \frac{9}{s^2+6s+9}$$

(۱۱)

در اینجا ضرایب یادگیری هر دو کنترل کننده برابر با ۳۰ اختیار شدند. اکنون ضرایب تناسب کنترل کننده‌ها را در دو مرحله به‌صورت زیر تغییر می‌دهیم:

$$K_1 = 2, K_2 = 1$$

(۱۲)

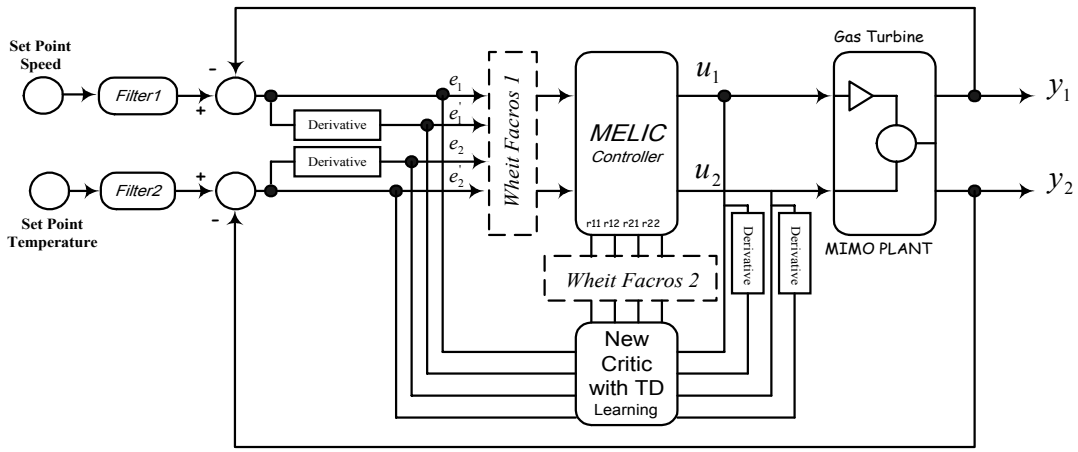
$$K_1 = 5, K_2 = 1$$

(۱۳)

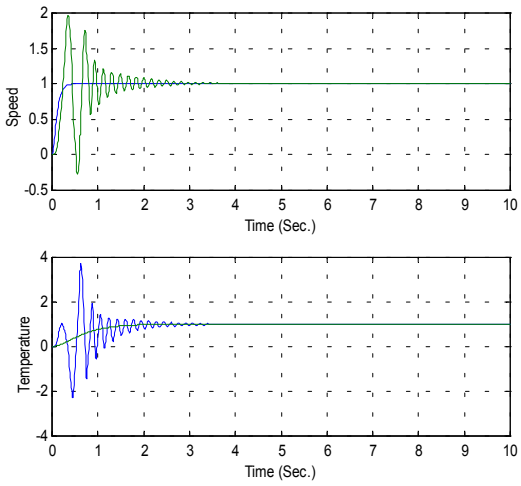
$$K_1 = 10, K_2 = 1$$

(۱۴)

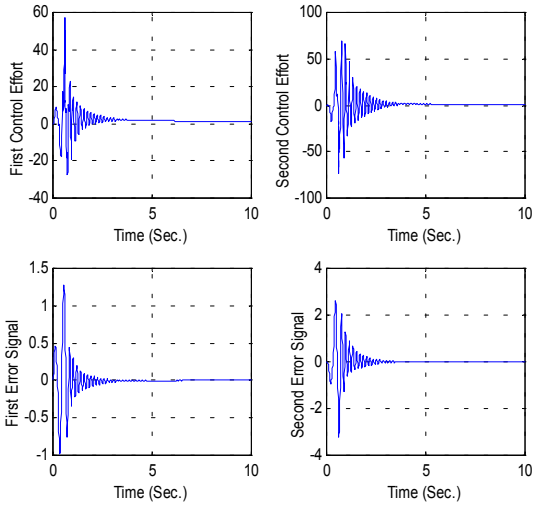
نتایج شبیه‌سازی برای این سه حالت در شکل‌های (۱۴) تا (۱۶) آمده‌اند. همان‌طور که ملاحظه می‌گردد با افزایش ضریب اهمیت کنترل کنندهٔ اول، میزان ردیابی آن درخروجی نیز بهبود می‌یابد. این نتایج بهتر از نتایج ارائه شده در [۲] می‌باشند.



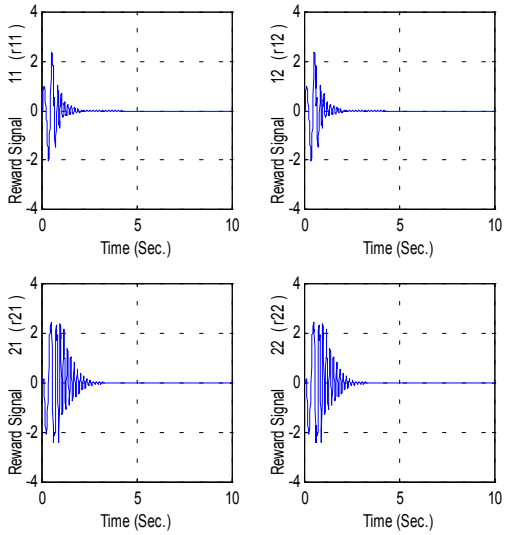
شکل ۷ : بلوک دیاگرام شبیه سازی.



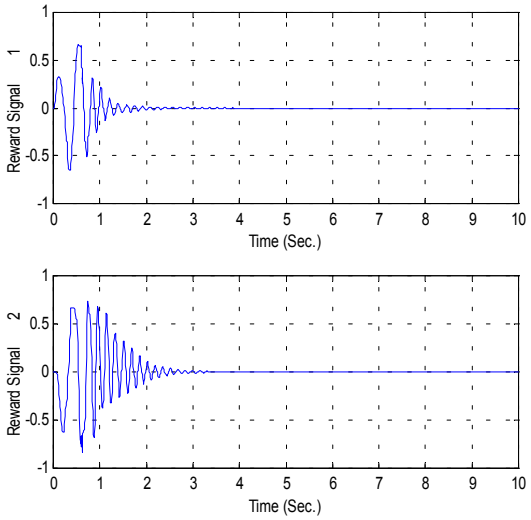
شکل ۸ : نتایج شبیه سازی برای توربین گازی.



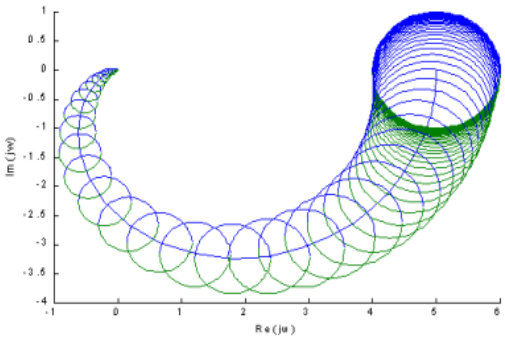
شکل ۹ : سیگنال‌های تلاش کنترلی و سیگنال‌های خطای خروجی.



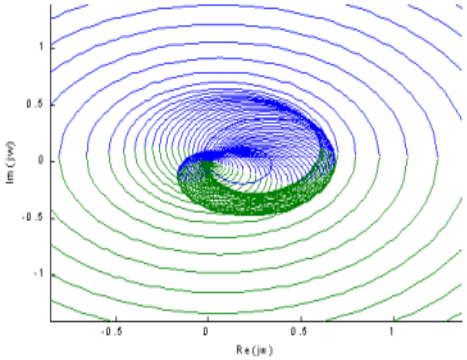
شکل ۱۰ : سیگنال‌های عاطفی r_{11}، r_{12}، r_{21} و r_{22}.



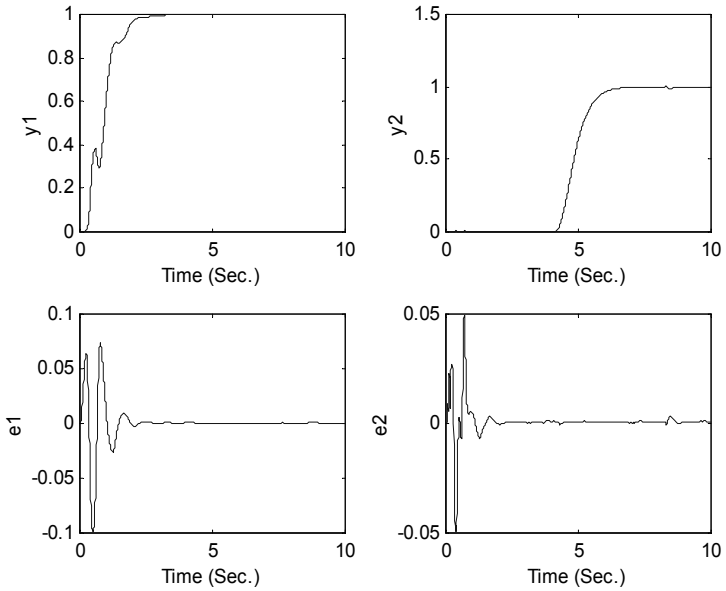
شکل ۱۱ : سیگنال‌های عاطفی r_1 و r_2.



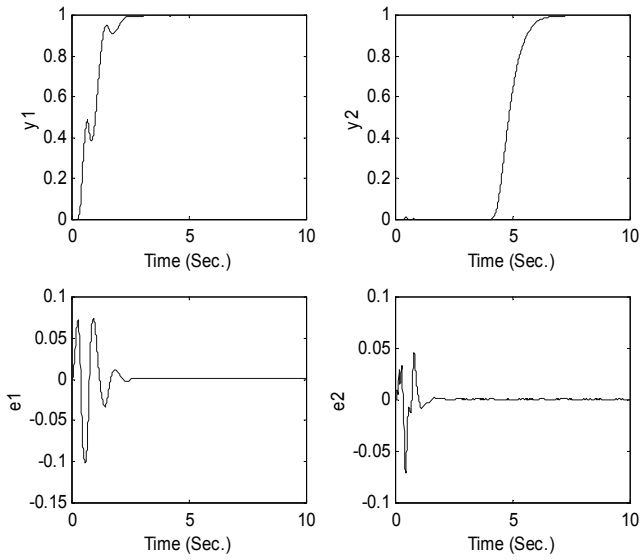
شکل ۱۲ : باندهای گرشگورین G11.



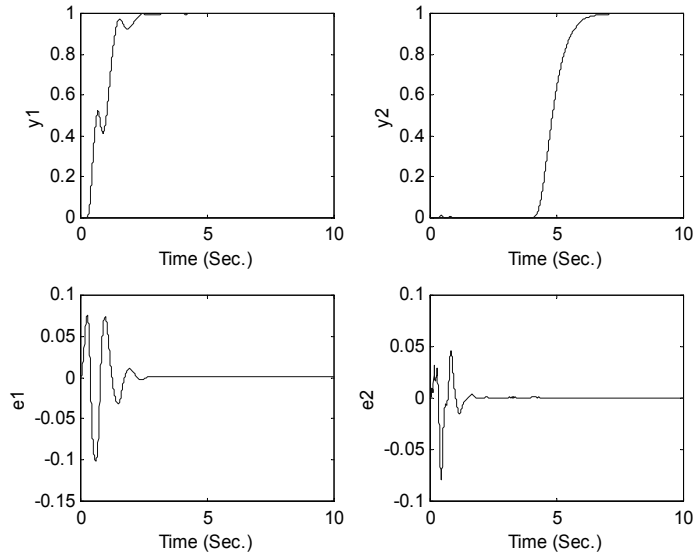
شکل ۱۳ : باندهای گرشگورین G22.



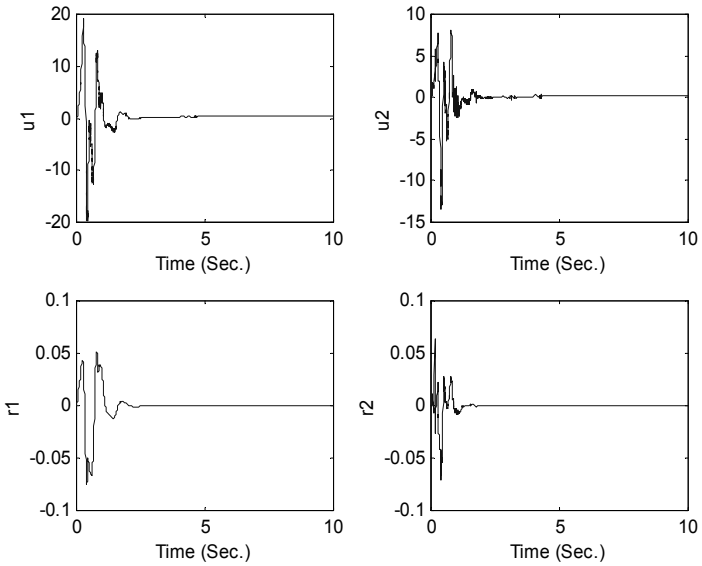
شکل ۱۴ : سیگنال‌های خروجی و خطا برای سیستم ستون تقطیر در حالت $(K_1 = 2, K_2 = 1)$.



شکل ۱۵ : سیگنال‌های خروجی و خطا برای سیستم ستون تقطیر در حالت $(K_1 = 5, K_2 = 1)$.



شکل ۱۶ : سیگنال‌های خروجی و خطا برای سیستم ستون تقطیر در حالت $(K_1 = 10, K_2 = 1)$.



شکل ۱۷ : سیگنال‌های کنترل و سیگنال‌های عاطفی در سیستم ستون تقطیر در حالت $(K_1 = 10, K_2 = 1)$.

همچنین در شکل (۱۷) سیگنال‌های مهم دیگر این سیستم کنترلی، نظیر سیگنال‌های کنترل و سیگنال‌های عاطفی سیستم ترسیم شده‌اند.

نتیجه گیری

با توجه به موفقیت‌های بدست آمده در کنترل عاطفی، هدف این مقاله بسط این روش برای پاسخگویی به مسائل پیچیده‌تر و اجرای اهداف مشکلتر بود. برای این منظور، قابلیت مدول یادگیری در کنترل کنندهٔ عاطفی با استفاده از یادگیری تفاوت زمانی برای تخصیص اعتبار به‌صورت پویاتر افزایش یافت. علاوه بر آن اهداف کنترلی چندگانه با استفاده از تعریف مناسب سیگنال عاطفی از نقاد به‌صورت توأم تجمیع شد تا کنترل کننده اعمال شده بتواند به تمامی مشخصات مطلوب جامه عمل بپوشاند. برای نشان دادن کارآیی، روش پیشنهادی آن را برای بسط کنترل سیستم‌های چندمتغیره به کار گرفتیم. توربین

گازی و ستون تقطیر، دو سیستم چندمتغیره با تزویج قوی می‌باشند که طراحی کنترل کنندهٔ خوب برای آنها حتی به‌صورت مبتنی بر مدل، امری پیچیده می‌باشد. نتایج اعمال کنترل کنندهٔ پیشنهادی حاکی از آن است که علاوه بر اینکه چندین پارامتر مثل خطا و سیگنال تلاش کنترلی مورد نقد قرار گرفتند خروجی مورد با سرعت مناسب و کمترین ماکزیمم فراجهشی ورودی را ردیابی نموده است.

مراجع

۱ - جذبی، ع. "توسعه روشهای یادگیری تقویتی در کنترل هوشمند و کاربردهای صنعتی و آزمایشگاهی آن." پایان نامه کارشناسی ارشد، دانشگاه تهران، (۱۳۷۷).

۲ - فتوره چی، م. "توسعه روش یادگیری عاطفی برای سیستمهای چند متغیره و سیستمهای با اهداف چند گانه." پایان نامه کارشناسی ارشد، دانشگاه تهران، (۱۳۸۰).

۳ - عبدی، ج. و لوکس، ک. "کاربرد یادگیری تفاوتی زمانی و یادگیری تقویتی در سیستمهای کنترلی." سمینار کارشناسی ارشد، دانشگاه تهران، (۱۳۸۱).

۴ - عبدی، ج.، لوکس، ک.، و صدیق، ع. خ. "کاربرد روش یادگیری تفاوت زمانی در مهندسی کنترل." پایان‌نامه کارشناسی ارشد، دانشگاه تهران، (۱۳۸۱)

5 - Lucas, C., Jazbi, S. A., Fatourechi, M. and Farshad, M. (2000). "Cognitive action selection with neurocontrollers." *Third Iran-Armenia Workshop on Neural Networks,* Yerevan, America

6 - Fatourechi, M., Lucas, C. and Sedigh A. K. (2001). "An agent-based approach to multivariable control." *Fourth Iran-Armenia Workshop on Neural Networks*.

7 - Barto, A. G., Sutton, R. S. and Anderson, C. W. (1983). "Neuron-like adaptive elements that can solve difficult learning problems." *IEEE Transactions on Systems, Man and Cybernetics*, Vol. SMC-13, No. 5, PP. 834-846.

8 - Berenji, H. R. and Khedkar, P. (1992). "Learning and tuning fuzzy logic controller through reinforcements." *IEEE Transactions on Neural Networks*, Vol. 3, No. 5, PP. 724-740.

9 - Berenji, H. R. (1994). "Fuzzy Q-Learning: a new approach for fuzzy dynamic programming." *Proceedings of IEEE 3rd Int'1 Conf. On Fuzzy Systems*, PP.486-491.

10 - Lee, C. C. (1990). "Fuzzy logic in control systems: fuzzy logic controller parts 1 and 2." *IEEE Transactions on Systems, Man and Cybernetics,* Vol. 20, No. 2, PP.404-435.

11 - Long-Ji Lin. (1993). *Reinforcement learning for robots using neural network.* PHD Thesis, School of Computer Science, Carnegie-Mellon University.

12 - Gordon, G. J. (1995). "Stable function approximation in dynamic programming." *In Proceedings of the 12th International Conference on Machine Learning (ML-95),* Morgan Kaufmann.

13 - Sutton, R. S. (1984). *Temporal credit assignment in reinforcement learning.* PHD Thesis, Department of Computer and Information Science, University of Massachusetts.

14 - Sutton, R. S. (1988). "Learning to predict by the methods of temporal differences." *Machine Learning*, Vol. 3, PP. 9-44.

15 - Sutton, R. S. (1995). "Generalization in reinforcement learning: successful examples using sparse coarse coding." *To Appear in Advances in Neural Information Processing Systems 8*.

16 - Cichosz, P. (1995). "Truncating temporal differences: on the efficient implementation of $TD(\lambda)$ for reinforcement learning." *Journal of Artificial Intelligence Research,* Vol. 2, PP. 287-318.

17 - Moore, A. W. (1990). *Efficient memory-based learning for robot control.* PHD Thesis, University of Cambridge Computer Laboratory.

18 - Patel, R. V. and Munro, M. (1984). *Multivariable system theory and design.* Pergamon Press.

19 - Abdi, J., Lux, C., Sedigh, A. K. and Khalili, A. F. (2003). "Truncating temporal differences in control." *11th International Conference on Electrical Engineering,* ICEE'03, Shiraz, Iran.

واژه های انگلیسی به ترتیب استفاده در متن

1 - Temporal Difference Emotional Control
2 - Multivariable System
3 - Neuro-Fuzzy
4 - Temporal Difference Learning
5 - Credit Assignment
6 - Fuzzy Critic
7 - Agent
8 - Multi Task Critic
9 - Reinforcement Learning
10 - Sutton
11 - State Utility Function
12 - Update
13 - Greedy Policy
14 - Eligibility Traces
15 - Truncated Temporal Difference
16 - Fuel Pump Excitation
17 - Nozzle Actuator Excitation
18 - Gas Generator Speed
19 - Inter-Turbine Temperature
20 - Inverse Nyquist Array
21 - Smooth
22 - Temporal Difference Multivariable Emotional Learning Based Intelligent Controller