مجله مهندسی مکانیک مدرس، دی 1394، دوره 15، شماره 10، ص‌ص 269-278

ماهنامه علمی پژوهشی

مهندسی مکانیک مدرس

mme.modares.ac.ir



کنترل تعادل دوچرخه بدون سرنشین با استفاده از یک سیستم طبقه‌بند بهبودیافته

سعید هاشم نیا[1]، مسعود شریعت پناهی[2*]

1- دانشجو دکترا، مهندسی مکانیک، دانشگاه تهران، تهران
2- دانشیار، مهندسی مکانیک، دانشگاه تهران، تهران
* تهران، 14399-55961، mshariatp@ut.ac.ir

اطلاعات مقاله

مقاله پژوهشی کامل
دریافت: 08 خرداد 1394
پذیرش: 31 مرداد 1394
ارائه در سایت: 28 شهریور 1394

کلید واژگان:
توابع عضویت فازی
دوچرخه بدون سرنشین
سیستم طبقه‌بند یادگیرنده
کنترل پایداری

چکیده

در این مقاله، استفاده از یک سیستم طبقه‌بند یادگیرنده بهبود یافته برای کنترل تعادل دوچرخه بدون سرنشین در حال حرکت پیشنهاد شده است. ویژگی بارز سیستم‌های مزبور، بی‌نیازی از مدل دینامیکی سیستم و توانایی یادگیری رفتار سیستم به‌کمک مجموعه‌ای از حرکت‌ها در محیط واقعی (نظیر فرایند یادگیری موجودات هوشمند) است. بر خلاف پژوهش‌های پیشین که در آن‌ها دامنه متغیرهای خروجی کنترل‌گر بصورت گسسته در نظر گرفته شده و به همین دلیل استفاده از آن‌ها در کاربردهای واقعی عملاً ناممکن شده است، در پژوهش حاضر کارایی سیستم طبقه‌بند یادگیرنده با استفاده از تعریف دامنه پیوسته برای خروجی‌های آن ارتقا پیدا کرده و سپس از آن برای کنترل تعادل دوچرخه بدون سرنشین استفاده شده است. برای این کار با استفاده از توابع عضویت فازی، امکان تعریف دامنه متغیرهای خروجی بصورت پیوسته فراهم شده است. سیستم پیشنهادی همچنین دارای یک مکانیزم تخصیص پاداش پویاست که برای مقابله با پاسخ تأخیری دوچرخه که ناشی از جرم آن است، ابداع شده است. با استفاده از این مکانیزم، امکان محاسبه سریع پاسخ فراهم شده و استفاده از کنترل‌گر در کاربردهای زمان- حقیقی مانند کنترل وسایل بدون سرنشین امکان پذیر می‌گردد. به‌منظور نشان دادن کارایی سیستم بهبودیافته پیشنهادی، یک مدل استاندارد از دوچرخه با دو درجه آزادی مورد استفاده قرار گرفته است. نتایج شبیه‌سازی نشان می‌دهد که سیستم پیشنهاد شده نسبت به سیستم‌های طبقه‌بند مرسوم و همچنین نسبت به برخی استراتژی‌های کنترلی مرسوم گزارش شده در پژوهش‌های پیشین بهتر عمل می‌کند.

Balance control of an unmanned bicycle using an improved classifier system

Saeed Hashemnia, Masoud Shariat Panahi*

Department of Mechanical Engineering, University of Tehran, Tehran, Iran.
* P.O.B. 14399-55961 Tehran, Iran, mshariatp@ut.ac.irIran, maligoodarz@srttu.edu

ARTICLE INFORMATION

Original Research Paper
Received 29 May 2015
Accepted 22 August 2015
Available Online 19 September 2015

Keywords:
Balance control
Fuzzy Membership Function
Learning Classifier System
Unmanned Bicycle

ABSTRACT

In the present article, an improved Learning Classifier System (LCS) is proposed to control the balance of a moving unmanned bicycle. A significant characteristic of learning classifier systems is that they can learn through a set of system actions in the real world (similar to intelligent creatures) while no dynamic model of the system is needed. Contrary to studies reported in the literature where action domain of the controller is discrete and accordingly such controller cannot be used in real world applications, in the present study efficacy of the classifier system is enhanced by definition of continuous domain for the outputs, and then is used to control the balance of unmanned bicycle. A scheme based upon fuzzy membership functions is proposed which makes it possible for the domain of actions to be continuous. The proposed LCS features a dynamic reward assignment mechanism which is invented to cope with the bicycle's delayed response due to its mass inertias. This allows the rapid calculation of the reward and hence enables the controller to be used in such real time applications as the balance control of unmanned vehicles. A standard 2 degree of freedom (2-DOF) bicycle model is employed to demonstrate the efficiency of the enhanced LCS. Simulation results show that the proposed classifier system outperforms traditional classifier system as well as some of the more common balance-control strategies reported in the literature.

1- مقدمه

سیستم‌های مجهز به مکانیزم یادگیری امروزه به‌طور گسترده برای حل مسائل مربوط به محیط‌های دینامیکی و نیمه‌مشاهده‌پذیر به‌کار می‌روند. سیستم‌های طبقه‌بند یادگیرنده[1] به‌عنوان یکی از این سیستم‌های یادگیرنده، از مجموعه‌ای از قوانین (کلاسیفایر) شرط-عملکرد[2] استفاده می‌کنند که چنانچه به میزان کافی تکامل پیدا کنند، عملکرد مناسب در پاسخ به اطلاعات

1- Classifier Systems
2- Action

برای ارجاع به این مقاله از عبارت ذیل استفاده نمایید:

Please cite this article using:

S. Hashemnia, M. Shariat Panahi, Balance control of an unmanned bicycle using an improved classifier system, *Modares Mechanical Engineering*, Vol. 15, No. 10, pp. 269-278, 2015 (In Persian)

[Downloaded from mme.modares.ac.ir on 2024-03-29]

[DOR: 20.1001.1.10275940.1394.15.10.25.9]

دریافتی از محیط را تجویز می‌کنند. در مقایسه با رهیافت‌هایی که صرفا مبتنی بر تکامل هستند، سیستم‌های طبقه‌بند با توجه به بازدهی تکنیک‌های یادگیری تقویتی سریعتر هستند [1].

گرچه سیستم‌های طبقه‌بند یادگیرنده در دامنه وسیع و متنوعی از مسائل مختلف به‌طور موفقیت‌آمیز مورد استفاده قرار گرفته‌اند، اما تلاش‌های ناچیزی برای آزمودن بازدهی آنها برای حل مسائل پیچیده کنترلی انجام شده است. از میان نمونه‌های کاربردهای کنترلی سیستم‌های طبقه‌بند می‌توان به ناوبری ربات متحرک [2] و کاربرد ساده جستجوی نور [3] اشاره کرد.

در مقاله حاضر سیستم طبقه‌بند دارای بخش عملکرد پیوسته همراه با مکانیزم تخصیص پاداش پویا[1] به‌منظور پایدار کردن حرکت دوچرخه بدون سرنشین پیشنهاد می‌شود. مشارکت و نوآوری اصلی این نوشتار پیشنهاد طرحی بر مبنای منطق فازی است که با استفاده از آن این امکان برای سیستم طبقه‌بند مقدار- حقیقی[2] فراهم می‌شود تا بتواند با عملکرد پیوسته کار کند. بنابراین با ایجاد این امکان، مشکل گسسته بودن عملکرد در سیستم طبقه‌بند مقدار- حقیقی حل شده و با حل این مشکل استفاده از آن در مسائل کنترلی مناسب خواهد بود. در عین حال روشی ویژه برای تشکیل تابع پاداش در مسائل کنترلی واقعی –که تقریبا همیشه در آنها جسمی دارای جرم وجود دارد- پیشنهاد شده است تا بتوان نقایص تخصیص پاداش استاتیکی را پوشش داد.

در ادامه مقاله و در بخش 2، استراتژی‌های کنترلی مختلفی که برای حفظ تعادل دوچرخه بدون سرنشین در مقالات پیشین مورد استفاده قرار گرفته است، مرور شده‌اند. در بخش 3، ابتدا مدل دوچرخه مورد استفاده در این پژوهش توضیح داده شده و سپس نحوه ایجاد سیستم طبقه‌بند دارای بخش عملکرد پیوسته همراه با مکانیزم تخصیص پاداش پویا به عنوان کنترل‌گر شرح داده شده است. نتایج شبیه‌سازی‌ها در بخش 4 و نهایتاً جمع‌بندی در بخش 5 ارائه شده است.

2- پژوهش‌های پیشین

تفاوت اصلی سیستم‌های دوچرخ (موتورسیکلت و دوچرخه) با دیگر سیستم‌ها، نیاز آنها به حفظ تعادل است، مسأله‌ای که در خودروهای چرخ‌دار مشابه [4] وجود ندارد. از آنجا که وسایل دوچرخ زیرفعال[3] هستند، کنترل تعادل آنها با چالش مواجه است. دوچرخه بدون سرنشین در محدوده‌ای مشخص از سرعت می‌تواند به‌طور خودکار تعادل خود را حفظ کند. دوچرخه خارج از محدوده ذکر شده ناپایدار است و باید با استفاده از کنترل‌گر مناسب پایدار گردد. پژوهش‌های مختلفی پایدارسازی دوچرخه را مورد مطالعه قرار داده و راه‌های گوناگونی برای این منظور پیشنهاد شده است.

شارپ [5] از اولین محققانی بود که پایداری وسایل دوچرخ را تحلیل کرد. شارپ و همکاران [6] بر مبنای اندازه‌گیری خطای ردیابی، از کنترل‌گر PID برای تولید گشتاور مناسب اعمالی به فرمان استفاده کردند. با وجود اینکه کنترل‌گرهای PD و PID [7] قبلا برای حفظ تعادل دوچرخه در موقعیت عمودی استفاده شده بود، شارپ و همکاران [6] بهره‌های متفاوتی را با توجه به سرعت پیشنهاد نمودند. سپس شارپ [8] از تئوری کنترل پیش نمایش خطی بهینه[4] برای کنترل مسیر موتورسیکلت استفاده نموده و کنترل راننده دوچرخه را مورد مطالعه قرار داد [9].

به‌طور کلی دو راهبرد در پژوهش‌های پیشین برای پایدارسازی دوچرخه مورد استفاده قرار گرفته است: یکی کنترل مرکز گرانش دوچرخه و دیگری کنترل زاویه فرمان. یوچی و همکاران [10] مدلی را ایجاد کردند که در آن بدنه دوچرخه و آونگ به عنوان یک سیستم آونگ معکوس دوگانه مورد ملاحظه قرار می‌گرفت. آنها به ارائه راهبردی برای ایجاد حرکت خودکار دوچرخه با استفاده از هر دو ابزار فرمان و مرکز گرانش پرداختند. شوآب و همکاران [11] دو نوع کنترل‌گر را بررسی کردند: طراحی کنترل بهینه با تنظیم‌کننده خطی درجه دوم و کنترل‌گر شهودی بر مبنای فرمان دادن به سمت زاویه مایل جانبی[5]. سرون و همکاران [12] کنترل‌گر بازخورد حالت پارامتر-متغیر خطی را برای حفظ تعادل دوچرخه در موقعیت عمودی پیشنهاد کردند.

علاوه بر کنترل‌گرهای خطی، رهیافت کنترل غیرخطی نظیر کنترل‌گر فازی نیز برای پایدارسازی وسایل دوچرخ مورد توجه بوده‌اند. چن و داو [13] از یک کنترل‌گر PID و یک کنترل‌گر فازی برای حفظ تعادل دوچرخه استفاده کردند. آنها از یک کنترل‌گر فازی دیگر نیز برای کنترل زاویه مایل جانبی به نحو دلخواه که منجر به حرکت دایره‌ای پایدار دوچرخه می‌گردد، بهره بردند. سپس آنها [14] کنترل‌گری فازی قابل تطبیق با تغییرات سرعت برای کنترل دوچرخه جهت ردیابی زاویه مایل جانبی ایجاد کردند. داو و چن [15] کنترل ردیابی زاویه مایل جانبی دوچرخه بدون سرنشین با استفاده از کنترل‌گر مود لغزشی را بررسی کردند. هوآنگ و همکاران [16] جدول قوانین فازی شامل 121 قانون را با استفاده از سطح لغزش و مشتق آن معرفی نمودند. ننر و همکاران [17] از یک مدل خطی و نظریه بازخورد کمّی برای ایجاد کنترل‌گر مرتبه پایین طبقاتی[6] که موتورسیکلت را در دامنه سرعت بین 2/5 تا 6/5 متر بر ثانیه پایدار کند، استفاده کردند. تینگ و چنگ [18] طرح کنترل مقاومی را بررسی کردند که از شبکه عصبی فازی و روش کنترل پسگام[7] استفاده می‌کند. هاشم‌نیا و همکاران [19] کنترل‌گر فازی با پایداری تضمین‌شده را برای پایدارسازی حرکت دوچرخه بدون سرنشین پیشنهاد کردند. آنها یک کنترل‌گر فازی ایجاد کردند که تعادل دوچرخه را حفظ کرده و به‌طور همزمان پایداری را بر اساس معیار پایداری لیاپانوف تضمین می‌کند و در عین حال خطای ردیابی زاویه مایل جانبی و مشتق آن را حداقل می‌کند. آنها نشان دادند که کنترل‌گر ایجاد شده می‌تواند با انواع شرایط اولیه تطبیق پیدا کند. در نهایت با استفاده از اعمال کنترل‌گر فازی بر روی طراحی‌های مختلفی از دوچرخه (مجموعه‌های متفاوتی از پارامترهای دوچرخه با مقادیر مختلف) آنها مقاوم بودن کنترل‌گر در برابر تغییر پارامترها را نشان دادند.

در پژوهش حاضر سیستم طبقه‌بند دارای بخش عملکرد پیوسته همراه با مکانیزم تخصیص پاداش پویا به عنوان کنترل‌گر برای پایدارسازی دوچرخه بدون سرنشین پیشنهاد شده است. از آنجا که برخی از پارامترهای سیستم به سرعت حرکت دوچرخه وابسته هستند و بنابراین با زمان تغییر می‌کنند، مسأله مورد نظر چالش برانگیز است. علاوه بر آن برخی از پارامترها نظیر گشتاورهای اینرسی وابسته به مورد هستند و تخمین دقیق آنها مشکل است. بر خلاف رهیافت‌های کنترل کلاسیک، مکانیزم یادگیری پیشنهاد شده در این پژوهش بدون نیاز به مدل ریاضی سیستم قابل پیاده‌سازی است. قوانین (کلاسیفایرهای) این سیستم بدون هیچ‌گونه دانش قبلی نسبت به سیستم

1- Continuous-Action Classifier System with Dynamic Reward Assignment
2- Real-Value Classifier System (CSR)
3- Under-Actuated
4- Optimal Linear Preview Control Theory
5- Lean
6- Cascaded
7- Backstepping Control Technique

[Downloaded from mme.modares.ac.ir on 2024-03-29]
[DOR: 20.1001.1.10275940.1394.15.10.25.9]

عمل کرده و به سمت مجموعه‌ای از قوانین که یک راهکار کنترلی مقاوم را تشکیل می‌دهد، تکامل می‌یابند. در واقع راهکار کنترلی در معرض خطاهای احتمالی ناشی از اندازه‌گیری مشخصات دوچرخه و یا تغییر در محیط نظیر شرایط زمینی که دوچرخه بر روی آن حرکت می‌کند نیست. همچنین راهکار پیشنهادی بر خلاف رهیافت کنترل فازی نیازی به استفاده از تجربیات فرد متخصص برای تولید قوانین و تنظیم آنها ندارد.

در پژوهش حاضر گشتاور اعمالی به فرمان تنها متغیر کنترلی (بخش عملکرد اگر-آنگاه) است و زاویه مایل جانبی و مشتق آن متغیرهای ورودی (بخش شرط اگر-آنگاه) را تشکیل می‌دهند.

3- نحوه ایجاد سیستم طبقه‌بند دارای بخش عملکرد پیوسته همراه با مکانیزم تخصیص پاداش پویا

در سیستم‌های طبقه‌بند اولیه، هنگامی که قوانین عملکردی را توصیه می‌کنند که پاداش خارجی را در پی دارد، این پاداش هم به قوانین کنونی و هم به قوانینی که در گذشته فعال[1] شده‌اند (کل زنجیره از گذشته تا قوانین کنونی)، اعمال می‌گردید. علاوه بر آن در چنین سیستم‌های اولیه‌ای مناسب بودن[2] یک قانون تنها سنجشی از پاداشی که ممکن است دریافت کند نبود، بلکه سنجشی از پاداشی که تاکنون دریافت کرده است نیز بود. به این ترتیب، سنجش یکسانی برای مکانیزم فعال کردن هر قانون و همچنین تعیین مقدار ارزش عمومی آن به کار می‌رفت. این روش باعث مشکلات متعددی می‌شد، به ویژه برای قوانینی که به ندرت فعال می‌شدند اما بسیار ضروری بودند. این قوانین قبل از آنکه صحیح بودن خود را نشان دهند، به تدریج به سبب رقابت تکاملی از مجموعه قوانین کنار گذاشته می‌شدند.

هدف اصلی از ایجاد سیستم طبقه‌بند توسعه‌یافته[3] [20] فائق آمدن بر این مشکلات است. سنجش مناسب بودن در سیستم طبقه‌بند توسعه‌یافته بر مبنای صحت پیش‌بینی (و نه قدرت) است. به همین ترتیب قوانین در سیستم طبقه‌بند توسعه‌یافته دارای یک مقدار مناسب بودن برای الگوریتم ژنتیک هستند که این مقدار بر مبنای دقت تخمین پاداش (و نه خود مقدار پاداش که برای انتخاب قوانین جهت فعال شدن به کار می‌رود) است. به این ترتیب مشکل قوانین صحیحی که به ندرت فعال می‌شوند حل می‌شود. در [21] نشان داده شده است که سیستم طبقه‌بند توسعه‌یافته، قوانین صحیح با حداکثر عمومیت ایجاد می‌کند. جهت آشنایی بیشتر با سیستم طبقه‌بند توسعه‌یافته و ویژگی‌های آن می‌توان به منبع [22] رجوع کرد. توضیح الگوریتمی روش نیز در منبع [23] وجود دارد. اصلاح بیشتری نیز بر روی سیستم طبقه‌بند توسعه‌یافته انجام شده است که در آن ویلسون [24] برای کار با ورودی‌های حقیقی بخش شرط قوانین را به شکل بازه‌ای حقیقی (و نه صرفا چند صفر و یک باینری) نمایش داده است. از این رو این سیستم امکان کار با مقادیر حقیقی ورودی به جای مقادیر باینری را دارد و به صورت سیستم طبقه‌بند توسعه‌یافته مقدار-حقیقی[4] نامگذاری شده است. این سیستم به طور ویژه برای کار با متغیرهای ورودی پیوسته مناسب است؛ اما در آن همچنان متغیر خروجی باید گسسته باشد و امکان کار با متغیر خروجی پیوسته فراهم نیست. از آنجا که در زمینه‌های متنوعی همچون کنترل ربات‌ها نیاز به استفاده از بخش عملکرد پیوسته است، هیچ کدام از سیستم‌های پیشنهاد شده در بالا با توجه به گسسته بودن خروجی نمی‌توانند به کارایی بالایی دست پیدا کنند.

پیوسته بودن عملکرد هدفی است که سیستم‌های طبقه‌بند به ندرت مورد خطاب آن قرار گرفته‌اند [25]. ویلسون [25] سه ساختار سیستم طبقه‌بند (1- بر مبنای درون یابی، 2- بر مبنای الگوی عمل‌کننده-منتقد[5]، 3- بر مبنای در نظر گرفتن عملکرد به عنوان یک متغیر پیوسته همگن با ورودی) را معرفی کرد که به سیستم‌ها اجازه می‌دهد عملکرد پیوسته (غیر گسسته) داشته باشند. یکی از راه حل‌های ارائه شده در [25] سیستم طبقه‌بند عمومی است که در آن ورودی x با عملکرد a در فرایند تطبیق نمونه‌های محیط مرتبط می‌گردد. مشکل سیستم طبقه‌بند عمومی این است که در آن باید بخش شرط به شکل $t(x,a)$ به نحوی تکامل پیدا کند که عملکرد پیوسته باشد. راه حل دیگری نیز وجود دارد که توسط ترن و همکاران [26] ارائه شده است و در آن عملکرد به صورت ترکیب خطی حالت ورودی و برداری از وزن‌های متناظر محاسبه می‌گردد. رهیافت دیگری [27] نیز پیشنهاد شده است که در آن عملکرد قطعه‌ای کدگذاری شده[6] (بیان درختی شبیه به درختی که در برنامه‌نویسی ژنتیک تولید می‌شود) جایگزین عملکرد گسسته می‌گردد. در این روش مقدار عملکرد با استفاده از مقادیر حالات ورودی به عنوان پایانه-های عملکرد قطعه‌ای کدگذاری شده تعیین می‌گردد. ما در پژوهش حاضر با استفاده از منطق فازی، سیستم طبقه‌بند توسعه‌یافته مقدار-حقیقی را به نحوی بهبود می‌بخشیم تا بتواند با عملکرد پیوسته کار کند و از این رو قابلیت استفاده در کاربردهای کنترلی (بخصوص کنترل تعادل دوچرخه بدون سرنشین که موضوع این مقاله است) را دارا باشد.

علاوه بر آن، طرحی برای حل چالشی پیشنهاد می‌شود که در به وجود آمدن فرایندهای استدلال داخلی[7] در سیستم‌های طبقه‌بند با آن مواجهیم: کشف و نگهداری[8] زنجیره‌های طولانی از قوانین. یکی از راه‌حل‌های مرسوم، یادگیری از پاداش‌های تأخیری است که در منبع [28] پیشنهاد شده است. راه حل دیگر سیستم طبقه‌بند با عملکرد تأخیری [29] است که به طور ویژه برای یادگیری رفتار پیش‌گویانه ابداع شده است. این روش از طریق تأخیر انداختن عملکرد عمل می‌کند و از آنجا که تأخیر در قوانین بر روی ژنوم آنها رمزگذاری شده است، الگوریتم ژنتیک قادر خواهد بود به طور همزمان فضاهای عملکرد و تأخیر را اکتشاف کند [29]. در مسائل کنترلی و دینامیکی اغلب جسمی دارای جرم و خاصیت اینرسی مشاهده می‌شود. با توجه به پاسخ تأخیری ناشی از خاصیت اینرسی در چنین مسائلی، تعیین سریع پاداش صحیح بعد از ارسال عملکرد به محیط مشکل است. بنابراین در پژوهش حاضر رهیافت نوینی معرفی می‌شود که تخصیص پاداش پویا برای چنین سیستم‌هایی را ممکن می‌سازد. در این تحقیق سیستم طبقه‌بند دارای بخش عملکرد پیوسته همراه با مکانیزم تخصیص پاداش پویا برای پایدارسازی حرکت دوچرخه بدون سرنشین –که یک سیستم پیچیده و غیرخطی دارای چندین جرم و چندین گشتاور اینرسی جرمی است- به کار برده شده است.

3-1- مدل دوچرخه

سیستم‌های طبقه‌بند یادگیرنده برای کنترل سیستم مورد نظر به مدل دینامیکی آن نیاز ندارند و قادرند همانند موجودات هوشمند، رفتار سیستم را به‌کمک مجموعه‌ای از حرکت‌ها در محیط واقعی فرا بگیرند. با این وجود، از آنجا که در پژوهش حاضر امکان استفاده از سیستم واقعی برای تربیت

1- Fire
2- Fitness
3- eXtended Classifier System (XCS)
4- eXtended Classifier System for Real-Valued Inputs (XCSR)

5- Actor-Critic Paradigm
6- Code Fragment Action
7- Internal Reasoning Processes
8- Discovery Aad Maintenance

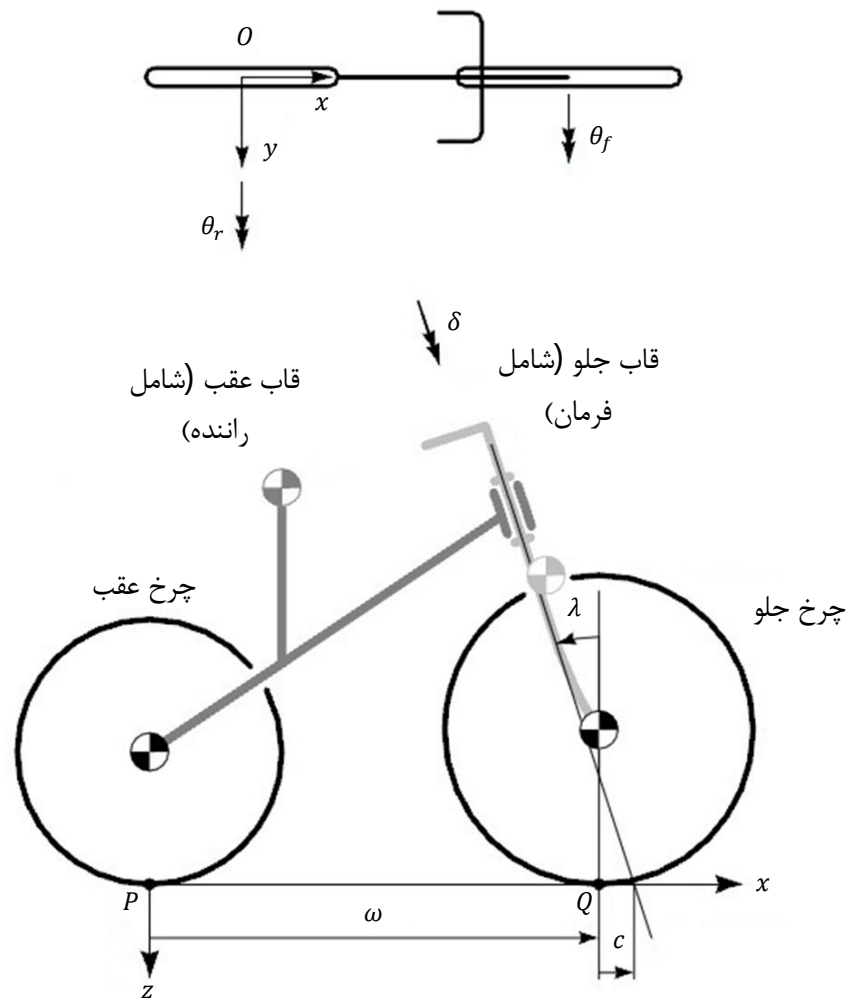
[Downloaded from mme.modares.ac.ir on 2024-03-29]
[DOR: 20.1001.1.10275940.1394.15.10.25.9]

کنترل‌گر هوشمند فراهم نبوده است، برای انجام این کار موقتاً از مدل دینامیکی سیستم استفاده شده است.

مدل دوچرخه مورد استفاده در این پژوهش قبلاً به عنوان مدل معیار[1] در [30،31] ارائه شده است و درستی آن بصورت تجربی نشان داده شده است [32]. در منبع [33] بیان شده است که این مدل برای کنترل خودکار دوچرخه کاملاً مناسب است. همان‌طور که در شکل 1 نشان داده شده است، این مدل دارای دو درجه آزادی شامل زاویه مایل جانبی $\varphi(t)$ و زاویه فرمان $\delta(t)$ است. سرعت v یک پارامتر ثابت است. مدل از چهار بخش صلب شامل قاب عقب، بخش چنگال مانند جلو و فرمان، چرخ عقب و چرخ جلو تشکیل شده است. این بخش‌ها از طریق مفاصل چرخشی متصل شده و متقارن جانبی هستند (چپ و راست با هم متقارنند). تماس چرخ‌ها با زمین به طور تک‌نقطه‌ای و غلتش بدون لغزش درنظر گرفته شده است. حالت تماس بوسیله قید هولونوم در جهت عمودی و قید غیر هولونوم در جهات طولی و جانبی مدل شده است. معادلات حرکت خطی شده شامل دو معادله دیفرانسیل معمولی کوپله مرتبه دوم است که به شکل ماتریسی (1) قابل نمایش است:

$$M\ddot{q} + (C_1 v)\dot{q} + (K_0 + K_2 v^2)q = f(t)$$
$$q = \begin{cases}\varphi(t)\\ \delta(t)\end{cases} \quad , \quad f(t) = \begin{cases}T_{\varphi(t)}\\ T_{\delta(t)}\end{cases} \qquad (1)$$

که در آن $T_{\varphi(t)}$ اغتشاش گشتاور رول خارجی و $T_{\delta(t)}$ گشتاور اعمالی بوسیله عملگر بر فرمان است. M ماتریس متقارن جرمی، $C_1 v$ ماتریس میرایی و $K_0 + K_2 v^2$ ماتریس سختی به شکل حاصل‌جمع بخش متقارن و ثابت K_0 و بخش وابسته به سرعت $K_2 v^2$ است. مقادیر درایه‌های این ماتریس‌ها وابسته به مشخصات هندسی دوچرخه است. لازم به ذکر است که $T_{\varphi(t)}$ گشتاور کنش-واکنش بین فضای ثابت و قاب عقب مربوط به زاویه مایل جانبی است.



شکل 1 مدل دوچرخه، سیستم مختصات، درجات آزادی و پارامترهای مربوطه (اقتباس شده از [30،31])

در عمل چنین گشتاوری ممکن است بوسیله باد جانبی، چرخ‌های اضافی مخصوص آموزش کودکان، و یا والدینی که به کودک خود آموزش می‌دهند، اعمال شود. در مورد یک دوچرخه عادی، این گشتاور صفر درنظر گرفته می‌شود.

با انتخاب زاویه مایل جانبی، زاویه فرمان، و مشتقات آنها به عنوان متغیرهای حالت، معادلات دیفرانسیل معمولی خطی شده (1) را می‌توان به شکل فضای حالت نمایش داد. در این حالت ورودی کنترلی $u(t)$ برابر با گشتاور $T_{\delta(t)}$ اعمال شده توسط عملگر بر روی فرمان است. فرض می‌شود که همه متغیرهای حالت با استفاده از خروجی $y(t)$ اندازه‌گیری می‌شوند. بنابراین معادلات فضای حالت به شکل (2) خواهد بود [30،31]:

$$\dot{x}(t) = A(v)x(t) + Bu(t)$$
$$y(t) = Cx(t) + Du(t) \qquad (2)$$

همان‌طور که توضیح داده شد در آن روابط (3) و (4) برقرارند:

$$x(t) = [\varphi(t)\ \ \delta(t)\ \ \dot{\varphi}(t)\ \ \dot{\delta}(t)] \qquad (3)$$
$$u(t) = T_{\delta(t)} \qquad (4)$$

مقادیر درایه‌های ماتریس‌های $A(v)$، B، C و D وابسته به پارامترهای فیزیکی و هندسه دوچرخه مورد نظر است:

$$A(v) = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ a_1 & a_{21} + a_{22}v^2 & a_3 v & a_4 v \\ d_1 & d_{21} + d_{22}v^2 & d_3 v & d_4 v \end{bmatrix} \qquad (5)$$

$$B = \begin{bmatrix} 0 \\ 0 \\ b_1 \\ b_2 \end{bmatrix} \qquad (6)$$

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad (7)$$

$$D = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \qquad (8)$$

در ادامه نحوه ایجاد سیستم طبقه‌بند دارای بخش عملکرد پیوسته همراه با مکانیزم تخصیص پاداش پویا به عنوان کنترل‌گر جهت حفظ تعادل دوچرخه بدون سرنشین توضیح داده می‌شود. الگوریتم مربوطه دارای چندین بخش به شرح زیر است:

3-2- نحوه نمایش قوانین

در پژوهش حاضر بخش شرط قوانین مشابه با بخش شرط در سیستم طبقه‌بند توسعه‌یافته مقدار-حقیقی در نظر گرفته می‌شود؛ اما برای کار با مقادیر پیوسته (به جای گسسته)، مقدار عملکرد به صورت یک عدد حقیقی در نظر گرفته می‌شود. مسأله نحوه برخورد با یک عدد حقیقی به جای باینری (مشابه سیستم طبقه‌بند توسعه یافته [20] و سیستم طبقه‌بند توسعه‌یافته مقدار-حقیقی [24]) در زیربخش‌های بعدی توضیح داده خواهد شد. ساختار قانون پیشنهادی به شکل زیر است:

$$(l_1, u_1), \dots, (l_i, u_i), \dots, (l_m, u_m): \qquad AV.$$

که در آن AV مقدار عملکرد مربوطه است. بخش شرط هر قانون مشابه با بخش شرط سیستم طبقه‌بند توسعه‌یافته مقدار-حقیقی [24] تشکیل شده است: هر شرط بوسیله حدود بالایی و پایینی مشخص شده است. به عبارت دیگر بخش شرط هر قانون از چند جفت تشکیل شده که هر جفت شامل دو عدد حقیقی (l_i, u_i) است. اندیس i نشان دهنده شماره ورودی در بخش

1- Benchmark

[Downloaded from mme.modares.ac.ir on 2024-03-29]
[DOR: 20.1001.1.10275940.1394.15.10.25.9]

شرط است و l نماینده حد پایینی و u نماینده حد بالایی است. این اعداد در ابتدا به صورت تصادفی با توزیع یکنواخت بر روی بازه [1 0] انتخاب می‌گردند. از آنجا که این اعداد متناظر با محدوده هر ورودی هستند، پس از انتخاب ضریب مقیاس مناسب نشان‌دهنده محدوده واقعی ورودی مورد نظر خواهند بود. برخلاف سیستم طبقه‌بند توسعه‌یافته [20] و سیستم طبقه‌بند توسعه‌یافته مقدار-حقیقی [24]، مقدار عملکرد یک عدد تصادفی با توزیع یکنواخت بر روی بازه [1 0] است که پس از انتخاب ضریب مقیاس مناسب نشان‌دهنده مقدار واقعی عملکرد خواهد بود.

از آنجا که مدل دوچرخه مورد استفاده در این پژوهش فاقد آونگ معکوس است، گشتاور اعمالی به فرمان به عنوان تنها ورودی کنترلی درنظر گرفته می‌شود. دو متغیر قابل اندازه‌گیری زاویه مایل جانبی و مشتق آن به عنوان متغیرهای ورودی (بخش شرط قانون‌ها) انتخاب شده‌اند. بنابراین گشتاور اعمالی به فرمان بخش عملکرد قانون‌ها بوده و ساختار قانون‌ها به شکل زیر خواهد بود:

$$(l_{\mathrm{R}}, u_{\mathrm{R}}), (l_{\dot{\mathrm{R}}}, u_{\dot{\mathrm{R}}}): \quad ST$$

که در آن R زاویه مایل جانبی، $\dot{\mathrm{R}}$ مشتق این زاویه، و ST گشتاور اعمالی به فرمان است. با انتخاب ضریب مقیاس‌های مناسب، بخش شرط قانون‌ها نشان‌دهنده زاویه مایل جانبی و مشتق آن بر حسب درجه و درجه بر ثانیه، و بخش عملکرد آنها نماینده گشتاور اعمالی بر فرمان برحسب نیوتن-متر خواهد بود.

3-3- نحوه ایجاد مجموعه مطابقت یافته[1]

مشابه با سیستم طبقه‌بند توسعه‌یافته مقدار-حقیقی [24]، هنگام مواجهه با یک مثال جدید، ابتدا الگوریتم تطابق آن با بخش شرط قانون‌ها را بررسی می‌کند. چنانچه المان مربوطه از مثال درون بازه متناظر از قانون باشد، تطابق برای (فقط) آن المان صحیح است. لازم به ذکر است که تطابق کامل برای یک قانون فقط هنگامی صورت می‌پذیرد که تمامی المان‌های آن تطابق داشته باشند. قانون‌هایی که بخش شرط آنها با مثال داده شده تطابق داشته باشد، مجموعه مطابقت یافته [M] را تشکیل می‌دهند. چنانچه چنین قانونی وجود نداشته باشد و یا تعداد چنین قانون‌هایی کمتر از تعدادی از پیش تعیین شده -که با $AS_{\min}$ نمایش داده می‌شود- باشد، پوشش اتفاق می‌افتد تا قوانینی جدید که بخش شرط آنها با مثال تطابق دارد، تولید گردد. در این حالت تعداد قانون‌های درون مجموعه [M] برابر با $AS_{\min}$ خواهد بود. قانون‌های جدید باید به کل جمعیت نیز افزوده گردند. چنانچه به سبب پوشش تعداد کل جمعیت از مقدار از پیش تعیین شده N فراتر رود، تعدادی دیگر از قوانین بر حسب مناسب بودن و تجربه آنها، از جمعیت حذف خواهند شد.

3-4- نحوه ایجاد مجموعه عملکرد پیوسته

لازم به ذکر است که چون عملکرد در سیستم طبقه‌بند توسعه‌یافته مقدار-حقیقی گسسته است، قانون‌هایی که مجموعه تطابق‌یافته را تشکیل داده‌اند، بوسیله مقدار عملکرد مشترک بین آنها گروه‌بندی شده و یک زیرمجموعه تشکیل می‌دهند. الگوریتم سیستم طبقه‌بند توسعه‌یافته مقدار-حقیقی برای هر کدام از این زیرمجموعه‌ها میانگینی از پیش‌بینی[2] با ضریب وزنی مناسب بودن محاسبه می‌کند. سپس الگوریتم سیستم طبقه‌بند توسعه‌یافته مقدار-حقیقی عملکردی را از مجموعه [M] بر اساس میانگین محاسبه شده در مرحله قبل انتخاب کرده و آن را جهت محاسبه پاداش به محیط ارسال می کند. به این صورت مجموعه عملکرد [A] به صورت زیرمجموعه‌ای از [M] که همه قانون‌های موجود در آن دارای عملکرد مشترک ارسال شده به محیط هستند، تشکیل می‌گردد [24]. مراحل اشاره شده در بالا برای سیستم طبقه‌بند دارای بخش عملکرد پیوسته همراه با مکانیزم تخصیص پاداش پویا با توجه به گسسته نبودن بخش عملکرد در آن کاملا متفاوت است: با توجه به مطالب ارائه شده در بخش 3-2، محدوده عملکرد قبلا بر روی بازه [1 0] نگاشت شده است. این محدوده باید به منظور کار با مقادیر پیوسته به n بازه (زیرمجموعه) مساوی به طول i/n افراز شود. بنابراین محدوده i-امین بازه $[i-1/n \quad i/n]$ خواهد بود. هریک از این بازه‌ها ممکن است تعدادی از مقادیر عملکرد را دربرگیرد و همچنین ممکن است یکی یا برخی از این بازه‌ها دربرگیرنده هیچ مقدار عملکردی نباشد. برای هرکدام از بازه‌هایی که دارای یک یا تعدادی بیشتر مقدار عملکرد هستند، میانگین پیش‌بینی با ضریب وزنی مناسب بودن قوانین مربوطه محاسبه شده و به هر بازه احتمالی متناسب با این میانگین اختصاص می‌یابد. احتمال اختصاص داده شده به بازه‌هایی که دربرگیرنده هیچ مقدار عملکردی نیستند، صفر خواهد بود. یکی از بازه‌ها با استفاده از گردونه شانس و بر اساس احتمال تخصیص یافته انتخاب شده و مجموعه عملکرد [A] نامگذاری می‌شود. بازه انتخاب شده دارای یک یا تعداد بیشتری مقدار عملکرد خواهد بود. در صورتیکه بازه مذکور دارای فقط یک مقدار عملکرد باشد، این مقدار برای دریافت پاداش به محیط فرستاده شده و پارامترهای قانون متناظر نیز بسته به میزان پاداش/تنبیه به‌روزآوری خواهد شد.

مشکل هنگامی بروز می‌کند که چندین عملکرد مختلف –که دارای مقادیر پیوسته نیز هستند- در بازه مزبور وجود داشته باشد. در این حالت باید به دو سوال پاسخ داد: کدام قانون باید برای ارسال به محیط انتخاب شود و پاداش دریافتی از محیط چگونه باید بین قانون‌هایی که دارای مقادیر عملکرد متفاوتی هستند، توزیع گردد؟

در روش پیشنهادی ما، عملکرد مربوط به قانون با بزرگترین مناسب بودن برای ارسال به محیط انتخاب می‌گردد. آزمایشات گسترده ما نشان داده است که مناسب بودن یک قانون (و نه پیش‌بینی و یا پیش‌بینی با ضریب وزنی مناسب بودن) بهترین نشان دهنده شایستگی یک قانون برای جمع‌آوری پاداش است. برای پاسخ به دومین سوال نیز در پژوهش حاضر از رهیافت مبتنی بر منطق فازی برای تخصیص پاداش/مجازات جداگانه برای هریک از اعضای بازه استفاده شده است که منجر به افزایش نرخ یادگیری سیستم طبقه‌بند می‌گردد. مکانیزم تخصیص پاداش پیشنهادی مطابق دستورالعمل زیر کار می‌کند:

الف. عملکرد انتخاب شده –که متناظر با قانون دارای بیشترین مقدار مناسب بودن است- فعال گردیده، به محیط ارسال شده و مقدار پاداش/مجازات دریافتی از محیط (R_{HF}) محاسبه می‌گردد.

ب. پاداشی که باید به هر قانون اختصاص یابد بوسیله رابطه (9) که تابعی از مقدار عملکرد قانون مربوطه است، با استفاده از تابع تعلق گوسی که بر روی بازه انتخاب شده تعریف گردیده است، محاسبه می‌گردد (شکل 2):

$$\mu = \mu_{\mathrm{reward}}(x) = R_{\mathrm{HF}} e^{\left(-\frac{(x-c)^2}{2st^2}\right)} \qquad (9)$$

که در آن c موقعیت مربوط به نقطه حداکثر تابع تعلق گوسی،

1- Match Set
2- Prediction

[Downloaded from mme.modares.ac.ir on 2024-03-29]
[DOR: 20.1001.1.10275940.1394.15.10.25.9]

st انحراف معیار و R_{HF} پاداش دریافتی به‌وسیله عملکرد فعال شده است. موقعیت مربوط به نقطه حداکثر تابع تعلق منطبق با مقدار عملکرد فعال شده در نظر گرفته شده است. به این ترتیب قانونی که عملکرد آن انتخاب شده بود، بیشترین پاداش را دریافت خواهد نمود. دیگر قانون‌های عضو مجموعه عملکرد نیز پاداشی متناسب با عکس فاصله بین عملکرد خود با عملکرد انتخاب شده دریافت خواهند نمود. با تعیین پاداش مربوط به هر قانون، پارامترهای آن با استفاده از دستورالعمل بخش 3-7 به روزآوری خواهد شد.

3-5- نحوه تعریف تابع پاداش

عبارت "پاداش" عموماً به شایستگی قابل اندازه‌گیری عملکرد فعال شده اطلاق می‌گردد. هدف تابع پاداش اندازه‌گیری مؤثر بودن عملکرد یک قانون در پایدارسازی دوچرخه است. مشکل اینجاست که در باره دوچرخه پاداش نمی‌تواند بلافاصله محاسبه شود، چراکه این محاسبه نیازمند اطلاع از پاسخ سیستم نسبت به عملکرد است که این پاسخ با تأخیر اتفاق می‌افتد. برای مثال حالتی فرض می‌شود که ورودی به شکل $[\varphi_0, \dot{\varphi}_0]$ باشد. چنانچه زاویه مایل جانبی و مشتق آن هر دو مثبت باشند ($\varphi_0 > 0$ و $\dot{\varphi}_0 > 0$)، با نگاه به دوچرخه از نمای پشت (شکل 3)، مشاهده می‌شود که مرکز جرم دوچرخه در قسمت منفی محور z قرار گرفته و به سمت قسمت مثبت محور y حرکت می‌کند (چرخش ساعتگرد). در این حالت به سادگی نتیجه می‌شود که گشتاوری در جهت u_1 برای بازگرداندن دوچرخه به موقعیت تعادل لازم است. بنابراین پاداش باید به گشتاوری در جهت u_1 (و نه u_2) اعطا شود. هنگامی که گشتاوری در جهت u_1 اعمال شود، مدت زمانی طول می‌کشد تا سیستم دینامیکی واکنش نشان دهد و چرخش ساعتگرد آن به چرخش پادساعتگرد تبدیل شود. به طور دقیق‌تر، گشتاور اعمالی در جهت u_1 منجر به شتاب در جهت u_1 می‌شود که بزرگی این شتاب متناسب با گشتاور مربوطه است. نتیجه چنین شتابی، کاهش سرعت زاویه‌ای ساعتگرد است. چنانچه گشتاور اعمالی در جهت u_1 در طول مدت زمان به اندازه کافی طولانی Δt اعمال گردد، سرعت زاویه‌ای به صفر نزدیک شده و سپس به سرعت زاویه‌ای در جهت پادساعتگرد تبدیل خواهد شد. حالتی را در نظر بگیرید که گشتاوری با اندازه کوچک در جهت u_1 در طول مدت زمان کوتاه Δt اعمال شود. در این حالت ممکن است به سبب اینرسی دوچرخه، مرکز جرم آن حتی پس از گذشت مدت زمان Δt باز هم در جهت ساعتگرد حرکت نماید. بنابراین پس از گذشت مدت زمان کوتاه Δt، زاویه مایل جانبی دوچرخه برابر با φ_1 خواهد بود به نحوی که $\varphi_1 > \varphi_0 > 0$. همان‌طور که قبلاً اشاره شد، هدف از پیش تعیین شده حفظ تعادل دوچرخه در موقعیت عمودی بود. حال چنانچه صرفاً بر اساس موقعیت زاویه‌ای نهایی φ_1 و مقایسه آن با موقعیت زاویه‌ای اولیه φ_0 قضاوت شود، گشتاور در جهت u_1 –که گشتاوری در جهت صحیح است- ناصحیح خواهد بود، چرا که در ظاهر باعث بدتر شدن موقعیت دوچرخه نسبت به حالت عمودی شده است. برای حل این تناقض که ناشی از پاسخ تأخیری سیستم‌های دینامیکی است، نباید سیستم را صرفاً در حالت کنترل شده بررسی کرد، بلکه باید آن را درحالتی که سیگنال کنترلی وجود ندارد نیز بررسی نمود و سپس به مقایسه این دو حالت پرداخت. مقایسه بین این دو حالت منجر به تعریف صحیحی برای خوب بودن سیگنال کنترلی خواهد شد. بنابراین دو حالت زیر باید برای محاسبه پاداش مد نظر قرار گیرند:

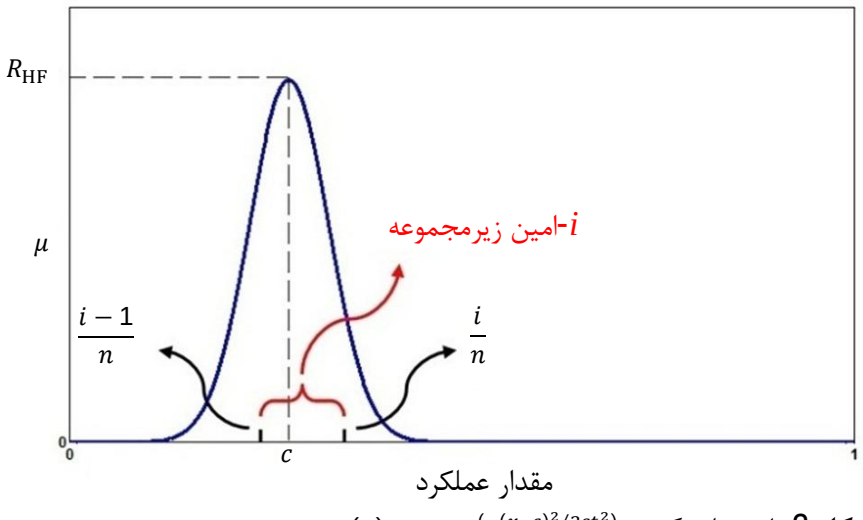
حالت 1- دوچرخه بدون سرنشین و بدون کنترل‌گر: با این فرض که کنترل‌گری بر دوچرخه اعمال نمی‌شود، زاویه مایل جانبی آن پس از $\Delta t_{\mathrm{reward}}$ ثانیه محاسبه می‌شود (φ_{unc}). لازم به ذکر است که چون در مقاله حاضر از سیستم *طبقه‌بند* برای کاربرد کنترلی استفاده شده است، بنابراین فرض بر این است که مدل سیستم در اختیار نیست. از این رو از معادلات حاکم بر دوچرخه برای این منظور استفاده شده است. بنابراین زاویه مایل جانبی با استفاده از معادله شتاب ثابت (10) تخمین زده می‌شود:

$$\varphi_{\mathrm{unc}} = \frac{1}{2}\ddot{\varphi}_0 \Delta t_{\mathrm{reward}}^2 + \dot{\varphi}_0 \Delta t_{\mathrm{reward}} \qquad (10)$$

که در آن شتاب در بازه زمانی کوچک $\Delta t_{\mathrm{reward}}$ ثابت در نظر گرفته شده است. لازم به ذکر است که سیگنال کنترلی در این حالت صفر است. علاوه بر آن، گشتاور ناشی از وزن که به عنوان گشتاور خارجی بر سیستم اعمال می‌شود نیز با توجه به تغییر ناچیز در بازوی گشتاورگیری در طول زمان کوتاه $\Delta t_{\mathrm{reward}}$ تقریبا عملا ثابت است. این دلیل موجه بودن فرض شتاب ثابت در طول زمان کوتاه $\Delta t_{\mathrm{reward}}$ است.

حالت 2- دوچرخه به‌وسیله سیگنال کنترلی اعمال شده کنترل می‌شود: در این حالت باید سیگنال کنترلی بر دوچرخه واقعی اعمال شده و زاویه مایل جانبی پس از مدت زمان $\Delta t_{\mathrm{reward}}$ اندازه‌گیری شود. با حل معادلات حاکم بر سیستم (3-6) با استفاده از روش رانجی-کوتا مرتبه چهار، مقدار زاویه مایل جانبی در پایان مدت زمان $\Delta t_{\mathrm{reward}}$ محاسبه می‌شود[1]. چنانچه این زاویه φ_{cont} نامیده شود، پاداش با استفاده از معادله (11) قابل محاسبه است:

$$\begin{cases} |\varphi_{\mathrm{unc}}| - |\varphi_{\mathrm{cont}}| > 0 \rightarrow R = |\varphi_{\mathrm{unc}}| - |\varphi_{\mathrm{cont}}| \\ |\varphi_{\mathrm{unc}}| - |\varphi_{\mathrm{cont}}| < 0 \rightarrow R = 0 \end{cases} \qquad (11)$$



شکل 2 تابع تعلق گوسی $\mu = \mu_{\mathrm{reward}}(x) = R_{\mathrm{HF}} e^{(-(x-c)^2/2st^2)}$ تعریف شده بر روی i-امین بازه. حد پایینی بازه i-1/n و حد بالایی آن i/n است.

شکل 3 نمای دوچرخه از پشت برای نمایش جهت‌های مثبت φ_0 و $\dot{\varphi}_0$.

1- با توجه به اینکه مجموعه تست آزمایشگاهی مربوطه هنوز به‌طور کامل راه‌اندازی نشده است، در اینجا از معادلات حاکم بر سیستم به جای دوچرخه واقعی استفاده شده است.

[Downloaded from mme.modares.ac.ir on 2024-03-29]
[DOR: 20.1001.1.10275940.1394.15.10.25.9]

با توجه به معادله (11)، چنانچه اندازه زاویه مایل جانبی نهایی در حالت وجود کنترل‌گر کمتر از حالت بدون کنترل‌گر باشد (یعنی دوچرخه به حالت تعادل عمودی نزدیک‌تر شده است)، عملکرد مربوطه مناسب تشخیص داده شده و پاداش اختصاص می‌یابد؛ در غیر اینصورت پاداش صفر خواهد بود.

3-6- اختصاص پاداش پویا

همان‌طور که پیش‌تر نیز اشاره شد، پاسخ تأخیری سیستم دینامیکی مورد نظر مانعی برای محاسبه آنی پاداش مربوط به عملکرد فعال شده است. در اینجا طرحی پیشنهاد می‌شود تا تابع پاداش پیشنهاد شده در بخش 3-5، بوسیله تنظیم پارامتر مدت زمان Δt_{reward}، به صورت پویا محاسبه شود:

$$\Delta t_{\text{reward}} = t_0|\varphi_0| + t_1 \quad (12)$$

که در آن t_0 و t_1 دو عدد ثابت برای تنظیم Δt_{reward} بر حسب اندازه زاویه مایل جانبی اولیه (φ_0) هستند. بر طبق معادله (12) برای مقادیر کوچک $|\varphi_0|$ (یعنی هنگامی که دوچرخه نزدیک موقعیت تعادل عمودی است)، مقدار کوچک Δt_{reward} پیشنهاد می‌شود؛ در حالی که برای مقادیر بزرگ $|\varphi_0|$ (یعنی هنگامی که دوچرخه دور از موقعیت تعادل عمودی است)، مقدار بزرگتر Δt_{reward} ارجحیت دارد تا سیستم زمان کافی برای رسیدن به حالت تعادل داشته باشد.

انتخاب Δt_{reward} به صورت وابسته به φ_0 با این واقعیت قابل توجیه است که وقتی Δt_{reward} کوچک است، سیستم زمان کافی برای پاسخ به سیگنال کنترلی ندارد؛ در حالی که Δt_{reward} بزرگ ممکن است باعث شود مرکز جرم دوچرخه به حالت عمودی رسیده و از آن عبور نماید و در ادامه به زاویه مایل جانبی با اندازه‌ای بزرگتر از مقدار اولیه برسد. در این حالت قانونی که منجر به چنین حرکتی شده است، با وجود اینکه سیگنال کنترلی صحیحی ارائه کرده و دوچرخه را به سمت موقعیت تعادل عمودی هدایت کرده است، غیر مؤثر انگاشته می‌شود و پاداشی دریافت نمی‌کند. از این رو Δt_{reward} باید تابعی از زاویه مایل جانبی اولیه باشد.

3-7- به‌روزآوری پارامترها

پس از آنکه پاداش برای عملکرد مربوط به قانون مشخص شد، پارامترهای آن قانون شامل تجربه، پیش‌بینی، خطا و مناسب بودن به‌روزآوری می‌شود. دستورالعمل به‌روز رسانی در روش حاضر مشابه روش‌های سیستم طبقه‌بند توسعه‌یافته [20] و سیستم طبقه‌بند توسعه یافته مقدار-حقیقی [24] است. برای به‌روزآوری پارامترهای هر قانون متعلق به مجموعه عملکرد [A]، مراحل زیر انجام می‌شود:

به‌روزآوری تجربه:

$$exp_i = exp_i + 1 \quad (13)$$

به‌روزآوری پیش‌بینی (P) و خطا (ε):

اگر: $exp_i < \frac{1}{\beta} \rightarrow$

$$P_i = P_i + \frac{R_i - P_i}{exp_i} \qquad \varepsilon_i = \varepsilon_i + \frac{(|R_i - P_i| - \varepsilon_i)}{exp_i}$$

اگر: $exp_i \geq \frac{1}{\beta} \rightarrow$

$$P_i = P_i + \beta(R_i - P_i) \qquad \varepsilon_i = \varepsilon_i + \beta(|R_i - P_i| - \varepsilon_i) \quad (14)$$

که در آن β ضریب یادگیری و R_i پاداش دریافتی از محیط است. به‌منظور بروزآوری مناسب بودن، ابتدا پارامتر صحت (k_i) متناظر با هر قانون موجود در [A] باید بروزآوری شود:

اگر: $\varepsilon_i < \varepsilon_0 \rightarrow k_i = 1$

اگر: $\varepsilon_i \geq \varepsilon_0 \rightarrow k_i = \beta\left(\frac{\varepsilon_i}{\varepsilon_0}\right)^{-\gamma}$ (15)

که در آن ε_0 حداقل خطای پیش‌بینی و γ توان صحت است. سپس مناسب بودن (F) هر قانون در [A] با توجه به صحت نسبی ($k_i/\sum k_i$) آن قانون به‌روزآوری می‌شود:

$$F_i = F_i + \beta\left[\left(\frac{k_i}{\sum k_i}\right) - F_i\right] \quad (16)$$

با به‌روزآوری‌های انجام شده در بالا یک دوره[1] آموزش کامل می‌شود.

3-8- تشکیل نسل بعد

پس از اینکه تعداد از پیش تعیین شده (که حد آستانه الگوریتم ژنتیک نامیده می‌شود) از دوره‌ها تکمیل شود، الگوریتم ژنتیک برای تکامل بهتر قانون‌ها اجرا می‌شود. الگوریتم ژنتیک مشابه حالات عادی دارای سه مرحله است: انتخاب، تقاطع و جهش. در مرحله انتخاب، دو قانون (والدین) با استفاده از گردونه شانس که احتمال بیشتر را به قانون‌هایی که دارای مقدار مناسب بودن بالاتر هستند می‌دهد، انتخاب می‌شوند. عملگر تقاطع بر روی والدین با نرخی از پیش تعیین شده (نرخ تقاطع) اعمال می‌شود. سپس با نرخ از پیش تعیین شده دیگری (احتمال جهش)، در هر یک از حدود (بالایی یا پایینی) فرزندان تولید شده جهش اتفاق می‌افتد. فرزندان حاصل به جمعیت افزوده می‌شوند و دو قانون دیگر از جمعیت حذف می‌شوند تا اندازه جمعیت ثابت باقی بماند. قانون‌های حذف شده دارای مقدار مناسب بودن پایینی هستند و باید قبلاً در تعداد حداقلی مشخصی آزمایش شرکت کرده باشند؛ یعنی باید قبلاً زمان کافی برای تخمین صحیح پارامترهای آنها در اختیارشان قرار داده شده باشد.

4- نتایج شبیه‌سازی

به‌منظور نشان دادن کارایی راهبرد کنترلی پیشنهادی، این راهبرد برای کنترل دوچرخه‌ای بدون سرنشین با پارامترهای ذکر شده در [30،31] استفاده شده است. مقادیر درایه‌های ماتریس‌های $A(v)$ و B که وابسته به مشخصات فیزیکی و هندسی دوچرخه هستند، عبارتند از:

$$A(v) = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 9.4702 & -0.589 - 0.887v^2 & -0.104v & -0.328v \\ 12.399 & 31.559 - 2.042v^2 & 3.618v & -3.139v \end{bmatrix} \quad (17)$$

$$B = \begin{bmatrix} 0 \\ 0 \\ -0.1226 \\ 4.265 \end{bmatrix} \quad (18)$$

مقدار Δt_{reward} بوسیله رابطه زیر تعریف شده است:

$$\Delta t_{\text{reward}} = \frac{\text{round}(|10\varphi_0|)}{100} + 0.03 \quad (19)$$

شبیه‌سازی با استفاده از روش رانجی-کوتا مرتبه چهارم با گام زمانی 0/01 ثانیه انجام شده است. در رابطه (19)، عملگر گرد کردن (round) برای این منظور استفاده شده است که Δt_{reward} حتما ضریبی از گام زمانی شبیه‌سازی باشد. اندازه جمعیت برابر با 10000 انتخاب شده و 10000 مثال برای آموزش سیستم طبقه‌بند استفاده شده است. دیگر پارامترهای مورد نیاز برای آموزش سیستم طبقه‌بند در جدول 1 فهرست شده است.

1- Epoch

[Downloaded from mme.modares.ac.ir on 2024-03-29]
[DOR: 20.1001.1.10275940.1394.15.10.25.9]

جدول 1 پارامترهای مورد نیاز برای آموزش سیستم طبقه‌بند

پارامتر	نماد	مقدار
حداقل تعداد قانون‌های مطابقت یافته	$AS_{\min}$	2
ضریب یادگیری	β	0/2
توان صحت	γ	5
حداقل خطای پیش‌بینی	ε_0	0/01
مناسب بودن اولیه	F_{initial}	0/02
پیش‌بینی اولیه	P_{initial}	0/02
خطای اولیه	$\varepsilon_{\text{initial}}$	2
حد آستانه الگوریتم ژنتیک	$GA_{\text{threshold}}$	100
احتمال جهش	P_{mut}	0/01
نرخ تقاطع	R_{cross}	0/05
تعداد زیرمجموعه‌ها (بازه‌ها)	N_{sub}	10
انحراف معیار تابع گوسی	st	0/2

نمودار زاویه مایل جانبی $\varphi(t)$ دوچرخه در سرعت دو متر بر ثانیه با شرایط اولیه الف) $\varphi_0 = 3\text{deg}$ و $\delta_0 = 3\text{deg}$ و ب) $\varphi_0 = -3\text{deg}$ و $\delta_0 = 3\text{deg}$ با صفر درنظر گرفتن دیگر شرایط اولیه در شکل 4 ترسیم شده است. به‌منظور نشان دادن قابلیت تطبیق کنترل‌گر طراحی شده با شرایط اولیه مختلف، شرایط اولیه در شکل 5 به صورت الف) $\varphi_0 = 3\text{deg}$ و $\dot{\varphi}_0 = -3\,\text{deg/sec}$ و ب) $\varphi_0 = -3\text{deg}$ و $\dot{\varphi}_0 = -3\,\text{deg/sec}$ انتخاب شده است (دیگر شرایط اولیه صفر درنظر گرفته شده‌اند).

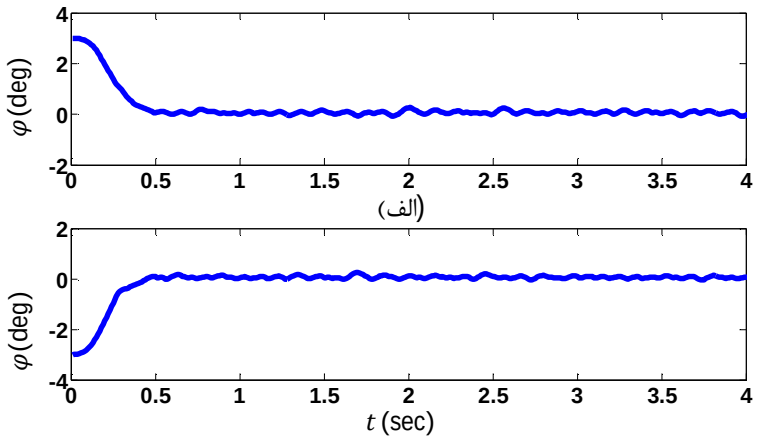
علاوه بر این، نمودار زاویه مایل جانبی $\varphi(t)$ دوچرخه در سرعت دو متر بر ثانیه با شرایط اولیه الف) $\varphi_0 = 6\text{deg}$ و $\delta_0 = -6\text{deg}$ و ب) $\varphi_0 = -6\text{deg}$ و $\delta_0 = -6\text{deg}$ با صفر درنظر گرفتن دیگر شرایط اولیه در شکل 6 ترسیم شده است.

ترسیمه زاویه مایل جانبی $\varphi(t)$ دوچرخه در سرعت دو متر بر ثانیه با انتخاب شرایط اولیه به صورت الف) $\varphi_0 = 6\text{deg}$ و $\dot{\varphi}_0 = 6\,\text{deg/sec}$ و ب) $\varphi_0 = -6\text{deg}$ و $\dot{\varphi}_0 = 6\,\text{deg/sec}$ در شکل 7 به تصویر کشیده شده است (دیگر شرایط اولیه صفر درنظر گرفته شده‌اند).

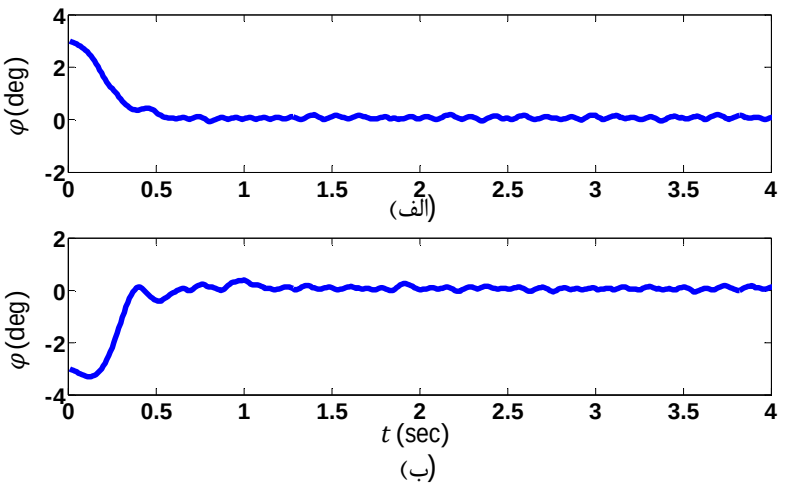
شکل‌های ترسیم شده نشان‌دهنده این واقعیت هستند که راهبرد کنترلی پیشنهاد شده قادر به حفظ تعادل دوچرخه در موقعیت عمودی با شروع از شرایط اولیه مختلف است.

مقایسه بین سیستم طبقه‌بند پیشنهادی به عنوان کنترل‌گر در این پژوهش با سیستم طبقه‌بند توسعه‌یافته مقدار-حقیقی مرسوم برای حفظ تعادل دوچرخه در حالت عمودی در سرعت دو متر بر ثانیه و با استفاده از پارامترهای مشابه برای هر دو سیستم طبقه‌بند در شکل 8 نشان داده شده است. با توجه به شکل قابل مشاهده است که با شروع از شرایط اولیه الف) $\varphi_0 = 5\text{deg}$ و $\delta_0 = 5\text{deg}$ هر دو سیستم طبقه‌بند قادر به حفظ تعادل دوچرخه در موقعیت عمودی هستند و در این حالت با توجه به جدول 2، سیستم طبقه‌بند پیشنهادی در این مقاله عملکرد بهتری (بر مبنای زمان نشست و فراجهش) نسبت به سیستم طبقه‌بند توسعه‌یافته مقدار-حقیقی مرسوم دارد؛ در حالی که با شروع از شرایط اولیه ب) $\varphi_0 = -5\text{deg}$ و $\delta_0 = 5\text{deg}$ فقط راهبرد پیشنهادی در این پژوهش قادر به حفظ تعادل دوچرخه و بازگرداندن آن به موقعیت عمودی است و سیستم طبقه‌بند توسعه‌یافته مقدار-حقیقی مرسوم توانایی پایدار نمودن دوچرخه در این حالت را ندارد.

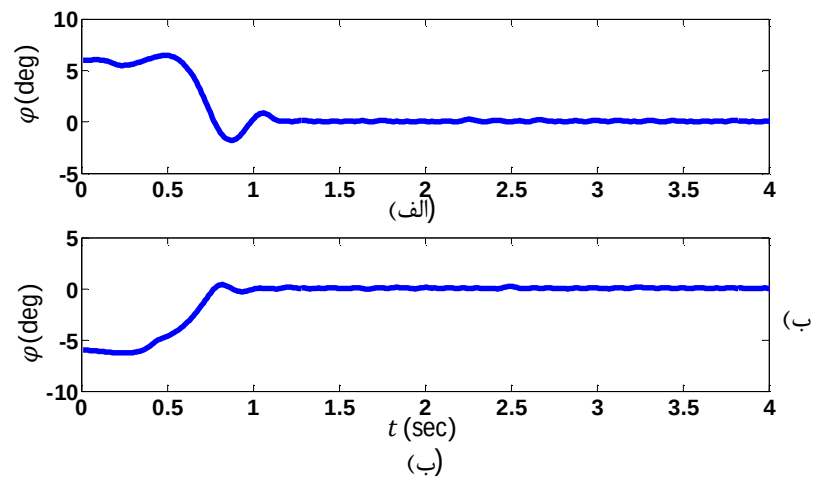
در نهایت راهبرد پیشنهادی در این پژوهش با دو راهبرد کنترلی که در



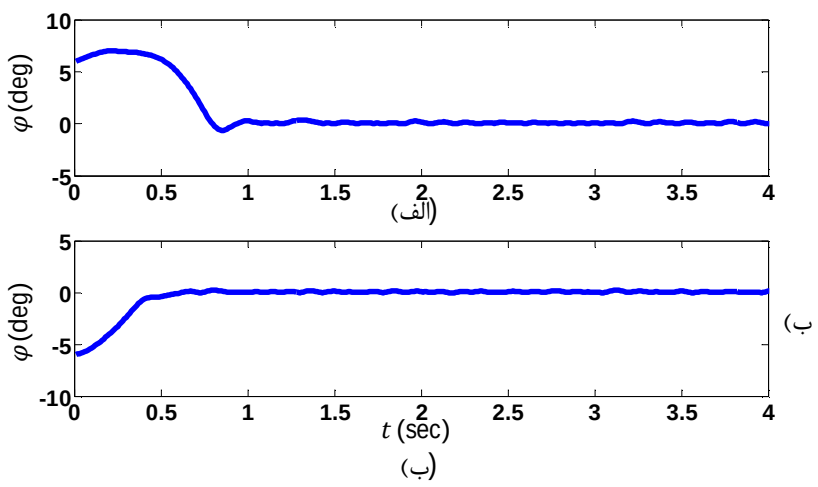
شکل 4 زاویه مایل جانبی $\varphi(t)$ دوچرخه در سرعت دو متر بر ثانیه با شرایط اولیه الف) $\varphi_0 = 3\text{deg}$ و $\delta_0 = 3\text{deg}$ و ب) $\varphi_0 = -3\text{deg}$ و $\delta_0 = 3\text{deg}$



شکل 5 زاویه مایل جانبی $\varphi(t)$ دوچرخه در سرعت دو متر بر ثانیه با شرایط اولیه الف) $\varphi_0 = 3\text{deg}$ و $\dot{\varphi}_0 = -3\frac{\text{deg}}{\text{sec}}$ و ب) $\varphi_0 = -3\text{deg}$ و $\dot{\varphi}_0 = -3\frac{\text{deg}}{\text{sec}}$



شکل 6 زاویه مایل جانبی $\varphi(t)$ دوچرخه در سرعت دو متر بر ثانیه با شرایط اولیه الف) $\varphi_0 = 6\text{deg}$ و $\delta_0 = -6\text{deg}$ و ب) $\varphi_0 = -6\text{deg}$ و $\delta_0 = -6\text{deg}$



شکل 7 زاویه مایل جانبی $\varphi(t)$ دوچرخه با شرایط اولیه الف) $\varphی_0 = 6\text{deg}$ و $\dot{\varphi}_0 = 6\frac{\text{deg}}{\text{sec}}$ و ب) $\varphi_0 = -6\text{deg}$ و $\dot{\varphi}_0 = 6\frac{\text{deg}}{\text{sec}}$

[Downloaded from mme.modares.ac.ir on 2024-03-29]
[DOR: 20.1001.1.10275940.1394.15.10.25.9]

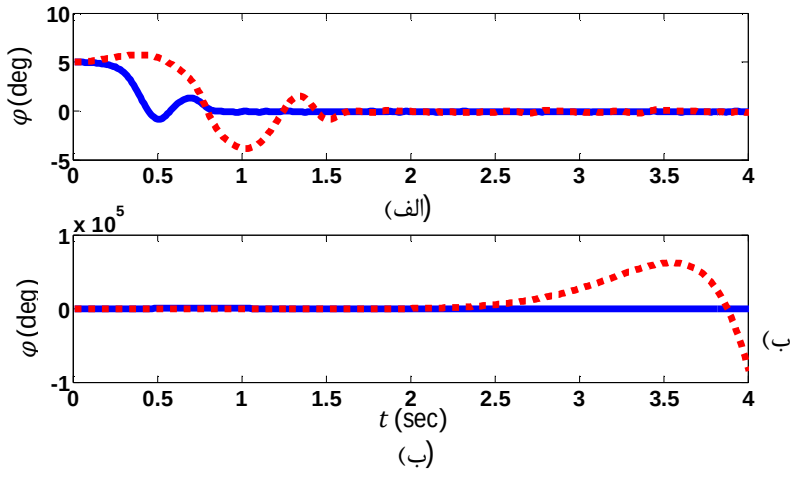
پژوهش‌های پیشین برای حفظ تعادل دوچرخه بدون سرنشین ارائه شده‌اند، مقایسه گردیده است. در ابتدا در شکل 9 راهبرد پیشنهادی در این مقاله با کنترل‌گر شهودی معرفی شده در [11] مقایسه شده است. هر دو کنترل‌گر قادر به حفظ تعادل دوچرخه هستند، اما با توجه به جدول 3، کنترل‌گر پیشنهادی در این پژوهش با در نظر گرفتن زمان نشست و فراجهش موثرتر است، یعنی دارای زمان نشست کوتاهتر و فراجهش کمتر است.

مقایسه بین عملکرد کنترل‌گر پیشنهادی در این پژوهش با کنترل‌گر بازخورد حالت پارامتر-متغیر خطی پیشنهاد شده در [12] نیز در شکل 10 به تصویر کشیده شده است. لازم به ذکر است که در این مقایسه هر دو کنترل‌گر به دوچرخه‌ای با ماتریس‌های $A(v)$ و B با مقادیر زیر اعمال شده‌اند:

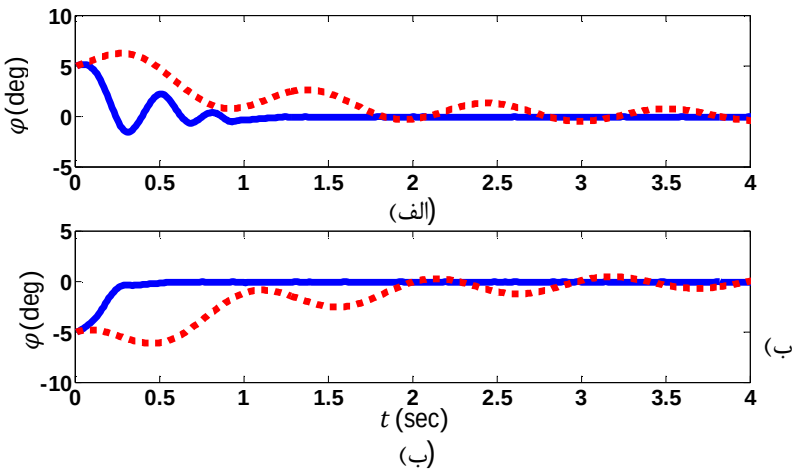
$$A(v) = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 13.67 & 0.225-1.319v^2 & -0.164v & -0.552v \\ 4.857 & 10.81-1.125v^2 & 3.621v & -2.388v \end{bmatrix} \quad (20)$$

$$B = \begin{bmatrix} 0 \\ 0 \\ -0.339 \\ 7.457 \end{bmatrix} \quad (21)$$

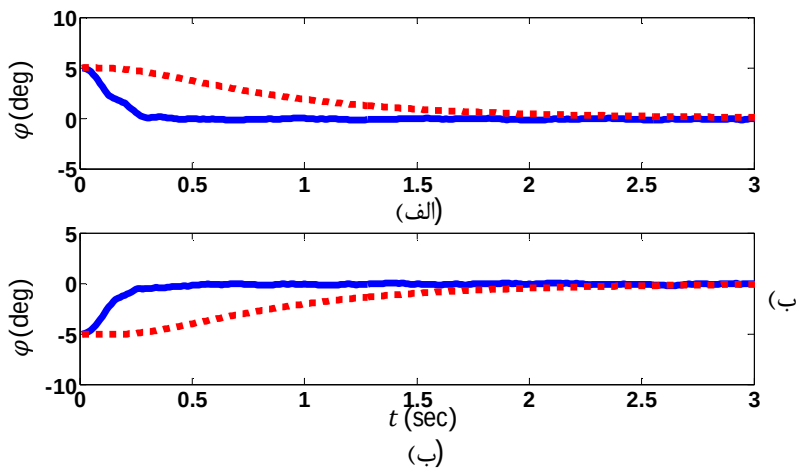
با توجه به جدول 4، مقایسه انجام شده بیانگر این حقیقت است که راهبرد پیشنهادی در این پژوهش نسبت به کنترل‌گر بازخورد حالت پارامتر-متغیر خطی پیشنهاد شده در [12] دارای زمان نشست کوتاهتری است و بنابراین عملکرد بهتری دارد.



شکل 8 مقایسه بین زاویه مایل جانبی $\varphi(t)$ دوچرخه در دو حالت استفاده از راهبرد پیشنهادی این پژوهش (خط توپر) و استفاده از سیستم طبقه‌بند توسعه‌یافته مقدار-حقیقی مرسوم ارائه شده در مرجع [24] (خط چین) با شرایط اولیه الف) $\varphi_0 = 5\text{deg}$ و $\delta_0 = 5\text{deg}$ و ب) $\varphi_0 = -5\text{deg}$ و $\delta_0 = 5\text{deg}$



شکل 9 زاویه مایل جانبی $\varphi(t)$ دوچرخه در دو حالت استفاده از راهبرد پیشنهادی این پژوهش (خط توپر) و کنترل‌گر شهودی مرجع [11] (خط چین) با شرایط اولیه الف) $\varphi_0 = 5\text{deg}$ و $\dot{\varphi}_0 = 5\text{deg}$ و ب) $\varphi_0 = -5\text{deg}$ و $\dot{\varphi}_0 = 5\text{deg}$



شکل 10 زاویه مایل جانبی $\varphi(t)$ دوچرخه با استفاده از راهبرد پیشنهادی این پژوهش (خط توپر) و کنترل‌گر پیشنهادی مرجع [12] (خط چین) با شرایط اولیه الف) $\varphi_0 = 5\text{deg}$ و $\delta_0 = 5\text{deg}$ و ب) $\varphi_0 = -5\text{deg}$ و $\delta_0 = 5\text{deg}$

جدول 2 مقایسه بین روش کنترلی پیشنهادی این پژوهش و روش پیشنهادی مرجع [24] با شرایط اولیه $\varphi_0 = 5\text{deg}$، $\delta_0 = 5\text{deg}$، $\dot{\varphi}_0 = 0$ و $\dot{\delta}_0 = 0$

روش	زمان نشست 5% برحسب ثانیه	درصد فراجهش
روش پیشنهادی پژوهش حاضر	0/8023	0
روش سیستم طبقه‌بند توسعه‌یافته مقدار-حقیقی مرسوم ارائه شده در مرجع [24]	1/5966	14/77

جدول 3 مقایسه بین روش کنترلی پیشنهادی این پژوهش و روش پیشنهادی مرجع [11] با شرایط اولیه $\varphi_0 = 5\text{deg}$، و $\dot{\varphi}_0 = 5\text{deg}$ ، $\delta_0 = 0$ و $\dot{\delta}_0 = 0$

روش	زمان نشست 5% برحسب ثانیه	درصد فراجهش
روش پیشنهادی پژوهش حاضر	1/1000	0
روش کنترل شهودی ارائه شده در مرجع [11]	5/7202	24/51

جدول 4 مقایسه بین روش کنترلی پیشنهادی این پژوهش و روش پیشنهادی مرجع [12] با شرایط اولیه $\varphi_0 = 5\text{deg}$، $\delta_0 = 5\text{deg}$، $\dot{\varphi}_0 = 0$ و $\dot{\delta}_0 = 0$

روش	زمان نشست 5% برحسب ثانیه	درصد فراجهش
روش پیشنهادی پژوهش حاضر	0/2711	0
روش کنترل بازخورد حالت پارامتر-متغیر خطی ارائه شده در مرجع [12]	2/3850	0

5- جمع‌بندی

در این پژوهش رهیافت نوینی برای کنترل تعادل دوچرخه بدون سرنشین بدون نیاز به مدل دینامیکی سیستم (که روش‌های کنترل کلاسیک به آن نیاز دارند) و یا شخص خبره (که روش‌های کنترل فازی به آن نیاز دارد) پیشنهاد گردید. راهبرد پیشنهادی از قابلیت یادگیری سیستم‌های طبقه‌بند مقدار-حقیقی استفاده کرده و این قابلیت را به نحوی گسترش می‌دهد که مقادیر عملکرد پیوسته را که در ذات مسائل کنترلی واقعی است، در بر بگیرد. برای پیاده‌سازی این رهیافت از دو متغیر زاویه مایل جانبی و مشتق آن به عنوان متغیرهای ورودی (بخش شرط قوانین) و از گشتاور اعمالی به فرمان به عنوان متغیر خروجی (بخش عملکرد قوانین) استفاده گردید. روش پیشنهادی

[Downloaded from mme.modares.ac.ir on 2024-03-29]
[DOR: 20.1001.1.10275940.1394.15.10.25.9]

همچنین مکانیزم تخصیص پاداش پویا بر مبنای اصول منطق فازی را جهت بهبود تطبیق‌پذیری سیستم با طبیعت دینامیکی محیط به کنترل‌گر طراحی شده اضافه می‌کند. این مکانیزم سیستم را قادر می‌سازد تا بر دشواری‌های ناشی از پاسخ تأخیری سیستم‌های الکترومکانیک که به سبب اینرسی جرم موجود در آنها پدید می‌آید، فائق آید. بنابراین عملکرد سیستم با توجه به این واقعیت که با مکانیزم توصیف شده در بالا پاداش به سرعت پس از فعال شدن یک قانون قابل محاسبه است، بهبود می‌یابد. این ویژگی کنترل‌گر را قادر می‌سازد تا در چنین کاربردهای زمان-حقیقی همانند کنترل تعادل وسایل بدون سرنشین مورد استفاده قرار گیرد. پیشرفت‌ها و اصلاحات انجام شده در مقایسه با سیستم‌های طبقه‌بند مقدار-حقیقی کلاسیک همچنین شامل اضافه شدن طرح توزیع پاداش است که به نحوی انجام می‌گیرد که تمامی اعضای مجموعه برنده امکان می‌یابند تا متناسب با نزدیکی به قانون فعال شده پاداش دریافت کنند. به این صورت و با به‌روزآوری پارامترهای چندین قانون در هر دوره، نرخ همگرایی روش پیشنهادی نیز بهبود می‌یابد.

به منظور نشان دادن کاربرد و کارایی رهیافت پیشنهادی -که "سیستم طبقه‌بند دارای بخش عملکرد پیوسته همراه با مکانیزم تخصیص پاداش پویا" نامیده شد- آن به مسأله کنترل تعادل دوچرخه بدون سرنشین برای چندین حالت مختلف شرایط اولیه اعمال گردید. نتایج شبیه‌سازی نشان داد که رهیافت پیشنهادی در تمامی حالات مورد مطالعه نسبت به سیستم طبقه‌بند توسعه‌یافته مقدار-حقیقی مرسوم و همچنین در مقایسه با دیگر راهبردهای کنترل تعادل گزارش شده در پژوهش‌های پیشین عملکرد مطلوب‌تری داشت. به‌منظور تأیید تجربی نتایج به دست آمده، نویسندگان در حال ایجاد دوچرخه‌ای بدون سرنشین در آزمایشگاه هستند تا با استفاده از کار آزمایشگاهی و به‌طور عملی نیز صحت نتایج را تأیید نمایند.

6- مراجع

[1] C. Wang, P. Wiggers, K. Hindriks, C. M. Jonker, Learning Classifier System on a humanoid NAO robot in dynamic environments, *12th International Conference on Control Automation Robotics & Vision (ICARCV),* Guangzhou. pp. 94-99, 2012.

[2] M. Shariat Panahi, A. Karkhaneh Yousefi, M. Khorshidi, Combining accuracy and success-rate to improve the performance of eXtended Classifier System (XCS) for data-mining and control applications, *Engineering Applications of Artificial Intelligence,* Vol. 26, No. 8, pp. 1930-1935, 2013.

[3] J. Hurst, L. Bull, C. Melhuish, TCS Learning Classifier System Controller on a Real Robot, *Lecture Notes in Computer Science,* Vol. 2439, pp. 588-597, 2002.

[4] M. R. Bidgoli, A. K. Khalaji, S. A. A. Moosavian, Trajectory tracking control of a wheeled mobile robot by a non-model-based control algorithm using PD-action filtered errors, *Modares Mechanical Engineering*, Vol. 14, No. 12, pp. 171-178, 2014. (In Persian)

[5] R. S. Sharp, The stability and control of motorcycles, *J. Mech. Eng. Sci.,* Vol. 13, No. 5, pp. 316–329, 1971.

[6] R. S. Sharp, S. Evangelou, D. J. N. Limebeer, Advances in the modeling of motorcycle dynamics, *Multibody Syst. Dyn.,* Vol. 12, pp. 251–283, 2004.

[7] S. Miyagishi, I. Kageyama, K. Takama, M. Baba, H. Uchiyama, Study on construction of a rider robot for two-wheeled vehicle, *JSAE Rev.,* Vol. 24, pp. 321–326, 2003.

[8] R. S. Sharp, Optimal preview speed-tracking control for motorcycles, *Multibody Syst. Dyn.,* Vol. 18, pp. 397–411, 2007.

[9] R. S. Sharp, On the Stability and Control of the Bicycle, *App. Mech. Rev.,* Vol. 61, pp. 1-24, 2008.

[10] K. Iuchi, H. Niki, T. Murakami, Attitude control of bicycle motion by steering angle and variable COG control, *Proc. 31st Annu. Conf. IEEE IECON,* pp. 2065–2070, 2005.

[11] A. L. Schwab, J. D. G. Kooijman, J. P. Meijaard, Some recent developments in bicycle dynamics and control, *Proceedings of the Fourth European Conference on Structural Control (4ECSC),* St. Petersburg, Russia, 2008.

[12] V. Cerone, D. Andreo, M. Larsson, D. Regruto, Stabilization of a Riderless Bicycle A Linear-Parameter-Varying Approach, *IEEE Contr. Syst. Mag.*, Vol. 30, pp. 23-32, 2010.

[13] C. K. Chen, T. S. Dao, Fuzzy Control for Equilibrium and Role Angle Tracking of an Unmanned Bicycle, *Multibody Syst. Dyn.,* Vol. 15, pp. 325-350, 2006.

[14] C. K. Chen, T. K. Dao, Speed-adaptive roll-angle-tracking control of an unmanned bicycle using fuzzy logic, *Veh. Syst. Dyn*., Vol. 48, pp. 133–147, 2010.

[15] T. K. Dao, C. K. Chen, Sliding-mode control for the roll-angle tracking of an unmanned bicycle, *Veh. Sys. Dyn*., Vol. 49, pp. 915–930, 2011.

[16] C. L. Hwang, H. M. Wu, C. L. Shih, Fuzzy Sliding-Mode Underactuated Control for Autonomous Dynamic Balance of an Electrical Bicycle, *IEEE T. Contr. Syst. T.,* Vol. 17, pp. 658-670, 2009.

[17] U. Nenner, R. Linker, P.O. Gutman, Robust feedback stabilization of an unmanned motorcycle, *Control Eng. Pract*., Vol. 18, pp. 970–978, 2010.

[18] C.S. Ting, Y.N. Chang, A Robust Fuzzy Neural Control Approach for Vehicle Lateral Dynamics, *Procedia Eng*., Vol. 29, pp. 479-483, 2012.

[19] S. Hashemnia, M. Shariat Panahi, M. J. Mahjoob, Unmanned Bicycle Balancing via Lyapunov Rule-Based Fuzzy Control, *Multi-Body System Dynamics,* Vol. 31, pp. 147-168, 2014.

[20] S. W. Wilson, Classifier Systems Based on Accuracy, *Evolutionary Computation,* Vol. 3, No. 2, pp. 149-175, 1995.

[21] T. Kovacs, *XCS classifier system reliably evolves accurate, complete, and minimal representations for Boolean functions,* Roy, Chawdhry, and Pant (Eds.), *Soft Computing in Engineering Design and Manufacturing,* Springer-Verlag, London, pp. 59-68, 1997.

[22] T. Kovacs, *Strength or Accuracy: Credit Assignment in Learning Classifier Systems, Distinguished Dissertations,* Springer, 2004.

[23] M. V. Butz, S. W. Wilson, An algorithmic description of XCS, *Journal of Soft Computing,* Vol. 6, No. 3-4, pp. 144-153, 2002.

[24] S. W. Wilson, *Get Real! XCS with Continuous-Valued Inputs, Learning Classifier Systems, From Foundations to Applications,* edited by Pier Luca Lanzi and Wolfgang Stolzmann and Stewart W. Wilson. Springer-Verlag, Berlin, pp. 209–219, 2000.

[25] S. W. Wilson, Three Architectures for Continuous Action, *Learning Classifier Systems, Lecture Notes in Computer Science,* Vol. 4399, pp. 239-257, 2007.

[26] T. H. Tran, C. Sanza, Y. Duthen, D. T. Nguyen, XCSF with Computed Continuous Action, *Proceedings of the Genetic and Evolutionary Computation Conference,* pp. 1861–1869, 2007.

[27] M. Iqbal, W. N. Browne, M. Zhang, XCSR with Computed Continuous Action, *AI 2012: Advances in Artificial Intelligence, Lecture Notes in Computer Science,* Vol. 7691, pp. 350-361, 2012.

[28] C. J. C. H Watkins, *Learning from Delayed Rewards,* PhD Thesis, King's College, 1989.

[29] C. Carse, Learning anticipatory behaviour using a delayed action classifier system, *Evolutionary Computing, Lecture Notes in Computer Science,* Vol. 865, pp. 210-223, 1994.

[30] A. Schwab, J. Meijaard, J. Papadopoulos, Benchmark results on the linearized equations of motion of an uncontrolled bicycle, *Int. J. Mech. Sci. Technol.,* Vol. 19, pp. 292–304, 2005.

[31] J. Meijaard, J. Papadopoulos, A. Ruina, A. Schwab, Linearized dynamics equations for the balance and steer of a bicycle: A benchmark and review, *Proc. R. Soc*., Vol. 463, pp. 1955–1982, 2007.

[32] J. Kooijman, A. Schwab, J. Meijaard, Experimental validation of a model of an uncontrolled bicycle, *Multibody Syst. Dyn*., Vol. 19, pp. 115–132, 2008.

[33] K. J. Astrom, R. E. Klein, A. Lennartsson, Bicycles dynamics and control, *IEEE Control Syst. Mag*., Vol. 25, No. 4, pp. 26–47, 2005.

[Downloaded from mme.modares.ac.ir on 2024-03-29]
[DOR: 20.1001.1.10275940.1394.15.10.25.9]