دوفصل‌نامه علمی منادی امنیت فضای تولید و تبادل اطلاعات (افتا)
سال ۱۴۰۰، جلد ۱۹، شماره ۱ (۲۰-۲۶)
https://monadi.isc.org.ir

ارائه‌شده در شانزدهمین کنفرانس بین المللی انجمن رمز ایران در دانشگاه فردوسی مشهد، ۶ و ۷ شهریور ۱۳۹۸

شناسایی ربات‌های وب با استفاده از ترکیب رویکردهای مبتنی بر ماشین‌های بردار پشتیبان فازی★

مصطفی سبزه‌کار*[۱،۲]، مجتبی سبزه‌کار[۲]، سید ابوالفضل اسلامی[۱] و علی مهری خانیکی[۲]

[۱]گروه مهندسی کامپیوتر، دانشگاه صنعتی بیرجند، بیرجند، ایران
[۲]گروه مهندسی کامپیوتر، دانشگاه آزاد اسلامی واحد بیرجند، بیرجند، ایران

اطلاعات مقاله

کلمات کلیدی:
تشخیص ربات‌های وب
ماشین بردار پشتیبان
ماشین بردار پشتیبان فازی
امنیت
طبقه‌بندی

dor: 10.0000/000000000

نوع مقاله: پژوهشی

چکیده

ربات‌های وب، برنامه‌هایی هستند که به طور خودکار و به صورت بازگشتی اطلاعات و محتوای وب‌سایت‌ها را بازبینی می‌نمایند. برخی از این ربات‌ها نقش مخربی دارند. بنابراین، تشخیص ربات‌های وب از جمله مهم‌ترین چالش‌ها در زمینه امنیت وب است. در این مقاله به ارائه یک روش جدید ترکیبی بر مبنای ماشین‌های بردار پشتیبان فازی با هدف افزایش کارایی در تشخیص ربات‌های وب پرداخته شده است. به این منظور از سه روش ماشین بردار پشتیبان فازی استفاده شده و توابع تعلق آنها به منظور افزایش دقت پیش‌بینی، با یکدیگر ترکیب شده است. نتایج روش پیشنهادی با ماشین بردار پشتیبان استاندارد و همچنین هر یک از سه ماشین بردار پشتیبان فازی مورد استفاده، مقایسه شده است. معیارهای مورد مقایسه، دقت، ویژگی و حساسیت می‌باشد. نتایج حاکی از برتری روش پیشنهادی در معیارهای مورد بررسی نسبت به الگوریتم‌های مورد مقایسه در این مقاله است.

© ۱۴۰۰ انجمن رمز ایران

۱ مقدمه

ربات‌های وب با رفتارهای متفاوت اعم از مخرب و غیرمخرب، در کنار کاربران انسانی، جزء بازدیدکنندگان وب به حساب می‌آیند. بعضی از ربات‌های وب و یا خزنده‌ها برای جاسوسی و دیگر اهداف شخصی طراحی شده‌اند. به عنوان مثال، دسته‌ای از ربات‌های وب به این منظور طراحی می‌شوند که به یک سایت به شکل ناشناس وارد شده و اطلاعات تجاری آن را هوشمندانه جمع‌آوری کنند. اغلب ربات‌ها برای استخراج دانش از سایت‌های تجاری، آثار مخربی را در آن به جای می‌گذارند. بنابراین مسئله‌ی تشخیص خودکار دسترسی‌های انسان از ربات یکی از مهم‌ترین مسائل در وب‌کاوی و امنیت وب است.

روش‌های متعددی در زمینه تشخیص ربات‌های وب مورد مطالعه قرار گرفته‌اند که تلاش آنها برای تشخیص و دسته‌بندی بازدیدکنندگان وب بوده است. متداول‌ترین راه برای تشخیص خودکار ربات‌های وب، استفاده از فایل‌های نشست، استخراج ویژگی‌های مرتبط (که دسترسی‌های انسان را از ربات بهتر تفکیک می‌نماید) و در نهایت استفاده از یک روش طبقه‌بندی برای حل مسئله است. به طور کلی می‌توان تلاش‌هایی که محققان برای حل این مسئله انجام داده‌اند، را در دو دسته مطالعه نمود. دسته‌ی اول مربوط به روش‌هایی است که روی انتخاب ویژگی‌های مناسب‌تر برای جدا کردن دسترسی‌های انسان از ربات تمرکز کرده‌اند (مانند روش‌های ارائه شده در [۱، ۲]) و دسته‌ی دیگر روش‌هایی که بر روی بهبود الگوریتم طبقه‌بندی تمرکز کرده‌اند (مانند روش‌های ارائه شده در [۳] با استفاده از زنجیره‌های مارکوف و [۴] با استفاده از تست تورینگ). در این مقاله، هدف افزایش دقت از طریق ارائه روشی جدید برای طبقه‌بندی است.

★از کمیته علمی شانزدهمین کنفرانس بین‌المللی انجمن رمز ایران برای داوری این مقاله تشکر می‌شود.
*نویسنده مسئول
آدرس‌های رایانامه: sabzekar@Birjandut.ac.ir (مصطفی سبزه‌کار)، sabzekar.Mojtaba@Gmail.com (مجتبی سبزه‌کار)، abolfazl.Eslami@Birjand.ac.ir (سید ابوالفضل اسلامی)، alimehrimail@gmail.com (علی مهری خانیکی)
© ۱۴۰۰ تمامی حقوق متعلق به انجمن رمز ایران است.

یکی از اولین روش‌های ارائه شده برای حل مساله توسط تان و کومار [1] ارائه شد. آن‌ها دسته‌بندی بازدیدکنندگان وب را در دو دسته کاربران انسانی و ربات‌های وب با استفاده از درخت تصمیم C4.5 روی یک مجموعه داده با ۲۵ ویژگی انجام دادند [۲، ۵]. در پژوهشی دیگر [۴] با استفاده از تست تورینگ که شامل تشخیص بر اساس تست کپچا [2] و تشخیص با رفتار مروری انسان می‌باشد، روشی برای دسته‌بندی بازدیدکنندگان ارائه شد. همچنین نویسندگان در [۳] با استفاده از زنجیره‌های مارکوف روشی جدید برای حل مسئله ارائه نمودند. از دیگر روش‌هایی که برای حل مسئله ارائه شده است می‌توان به روش بیزین [۶] و شبکه‌های عصبی [۷] اشاره نمود. هم‌چنین در پژوهشی دیگر [۸] و با این استدلال که روش‌های طبقه‌بندی دو کلاسه به ساختار ربات‌ها وابسته هستند و با تغییر ساختار آنها کارآیی خود را از دست می‌دهند، به ارائه روشی تک کلاسه مبتنی بر ماشین‌های بردار پشتیبان تک کلاسه (SVDD) پرداخته شده است. در این پژوهش، با تغییر در ساختار SVDD، روشی آگاه به نویز برای حل مسئله ارائه شده است. از پژوهش‌های اخیر انجام شده در این حوزه نیز می‌توان به استفاده از نظریه مجموعه سخت [3] [۹]، ترکیب روش‌های برخط (با استفاده از قواعد) و روش‌های برون خط (با استفاده از روش‌های یادگیری ماشین) [۱۰] و یادگیری عمیق [۱۱] اشاره نمود.

با وجود انجام پژوهش‌های متعدد در این حیطه، نیاز به ارائه روش‌های دقیق‌تر با توجه به اهمیت تشخیص دسترسی‌های ربات احساس می‌شود. اگر چه در صورتی که یک دسترسی انسان توسط ماشین به عنوان ربات در نظر گرفته شود، فقط سبب نارضایتی کاربر می‌شود اما اگر دسترسی ربات مخرب به درستی تشخیص داده نشود، آسیب‌های جبران ناپذیری به سیستم وارد خواهد شد. رویکرد ترکیب طبقه‌بندها [4]، یکی از تدابیر موفق در جهت افزایش دقت طبقه‌بندی است. این رویکرد سبب می‌شود بتوانیم از مزایای طبقه‌بندهای مختلف همزمان بهره ببریم. در این مقاله به ارائه روشی ترکیبی بر مبنای ماشین‌های بردار پشتیبان فازی با هدف افزایش کارایی در تشخیص ربات‌های وب پرداخته‌ایم. ماشین بردار پشتیبان فازی [5]، توسعه‌ای از ماشین بردار پشتیبان است به گونه‌ای که نمونه‌های آموزشی دارای درجه اهمیت متفاوتی هستند. باید برای هر نمونه علاوه بر ویژگی‌های آن یک تعلق فازی در نظر گرفت که به منظور وزن‌دهی به نمونه‌های آموزشی به‌کار می‌روند. بدین ترتیب، نمونه‌های با مقادیر تعلق بالاتر، نقش بیشتری در فرایند آموزش خواهند داشت. همچنین از توابع تعلق برای شناسایی داده‌های نویزی و پرت نیز می‌توان کمک گرفت [۱۲]. روش های متعددی برای تعیین تابع تعلق فازی FSVM توسعه داده شده‌اند [۱۳، ۱۴] که در هریک از این روش‌ها بر اساس معیاری خاص میزان اهمیت هر داده‌ی آموزش تعیین شده است.

در این مقاله هدف این است که با ترکیب این روال‌های امتیازدهی از مزایای تعدادی از آنها بهره بگیریم. با ترکیب چند روش به همراه تخصیص وزن به هر تابع تعلق می‌توان تابع تعلقی تعریف کرد که این رویکرد ترکیبی

[1] Tan and Kumar [2] captcha tests [3] rough set theory [4] classier combination
[5] Fuzzy Support Vector Machines (FSVM)

شکل ۱. مراحل آماده سازی مجموعه داده‌ها

دقت FSVM را در حل این مسئله بهبود خواهد بخشید. از آنجا که میزان اهمیت و کارآیی این روش‌ها با هم مقایسه نمی‌شوند در نتیجه وزن‌ها به صورت ثابت در نظر گرفته می‌شود. سپس توابع تعلق به‌صورت خطی وزن‌دار با هم ترکیب می‌شود. روش پیشنهادی در این مقاله، شامل چهار مرحله است: پیش‌پردازش، انتخاب ویژگی، برچسب گذاری نشست‌ها و طبقه‌بندی توسط ماشین بردار پشتیبان فازی. نتایج آزمایشات حاکی از برتری روش پیشنهاد شده دارد.

۲ روش پیشنهادی

مقاله حاضر به ارائه روشی جدید بر مبنای ترکیب توابع تعلق فازی مختلف ارائه شده در پژوهش‌های گذشته برای حل مسئله تشخیص ربات‌های وب می‌پردازد. یکی از مشکلات روش FSVM، انتخاب تابع مناسب تعلق برای نمونه‌های یادگیری است که زمینه‌ی پژوهشی جدیدی را پیش روی محققان باز نموده است. از آنجا که این تعلق‌ها در تابع هدف مسئله‌ی بهینه‌سازی FSVM نیز قرار می‌گیرد، انتخاب تابع نادرست کارایی این روش را تحت تاثیر خود قرار می‌دهد. بنابراین، در این مقاله، توابع مختلف تعلق با یکدیگر ترکیب می‌شوند تا از مزایای جمعی آنها استفاده شود. حل مسئله در چهار فاز انجام می‌شود: پیش‌پردازش، انتخاب ویژگی، برچسب‌گذاری نشست‌ها و طبقه‌بندی توسط ماشین بردار پشتیبان فازی. شکل ۱ جزئیات مراحل حل مسئله را از مرحله ثبت و شناسایی نشست‌ها تا بررسی معیارها را نمایش می‌دهد.

مسئله تشخیص ربات وب شامل مراحل زیر است:

۱.۲ پیش پردازش

در مسئله تشخیص ربات‌ها، گام پیش‌پردازش شامل تعیین نشست‌های موجود در فایل ثبت وقایع است، که با تقسیم و دسته‌بندی نشست‌ها براساس آدرس IP و اطلاعات عامل کاربر صورت می‌گیرد. هر نشست

جدید براساس یک بازه‌ی زمانی مشخص ایجاد می‌شود [۱۵–۱۹].

۲.۲ استخراج ویژگی

عملیات دسته بندی بازدیدگان وب توسط مدل پیشنهادی با در نظر گرفتن چندین ویژگی بررسی می‌شود.برخی ویژگی‌هایی که در منابع از آنها استفاده شده که به ۱۸مورد آنها اشاره شده است [۱۵، ۱۹، ۲۰]. این ویژگی‌ها در غالب مجموعه داده در نمونه کارهای دانشگاهی کشور انجام شده است.مشخصات برخی از ویژگی های مجموعه داده شامل: مدت نشست، درخواست فایل robot.txt، درخواست‌های Head، درخواست‌های GET، درخواست فایل‌های دیگر، درصد خطاهای 4xx، مجموع صفحات درخواستی، طول نشست، درصد تکرار درخواست، پهنای باند، وب و سند، اسکریپت، تصویر، چندرسانه‌ای، دانلود، Hit و تعداد صفحات یکتا می‌باشد.

۳.۲ برچسب‌گذاری نشست‌ها

پس از تعیین خصیصه‌ها برای هر نشست، نوبت به برچسب‌گذاری نشست‌ها می‌رسد. در مرحله اول تمام نشست‌هایی که خصیصه robot.txt آنها مقدار یکتاست، به عنوان ربات در نظر گرفته می‌شوند. در ادامه رشته عامل کاربر کلیه نشست‌های باقیمانده با لیست رشته عامل کاربر ربات‌های شناخته شده وب مقایسه می‌شود. در صورتی‌که رشته عامل کاربر با لیست بروز شده عامل‌های کاربر ربات‌های وب مطابقت داشت، آن نشست نیز به عنوان ربات وب شناخته می‌شود. اگر رشته عامل کاربر با لیست بروزشده مرورگرهای وب مطابقت داشته باشد و فایل robot.txt در این نشست خوانده نشده باشد، نشست به عنوان انسان برچسب‌گذاری می‌شود.

۴.۲ طبقه‌بندی توسط روش پیشنهادی

در این مقاله هدف بهره جستن از مزایای روش FSVM در طبقه‌بندی داده‌های دو کلاسی است. رابطه (۱) مدل ماشین بردار پشتیبان فازی را نمایش می‌دهد.

$$\begin{aligned} &\text{Minimize} && \frac{1}{2}\|w\|^2 + c\sum_{1}^{n} s_i\xi_i \\ &\text{subject to} && y_i(wz_i+b) \geq 1-\xi_i, \quad \xi_i \geq 0, \\ & && i = 1,\ldots,l \end{aligned} \tag{1}$$

هر کدام از کاربردهای ماشین بردار پشتیبان فازی برای بدست آوردن تابع تعلق از روش خاص استفاده نموده‌اند به عنوان نمونه در [۱۶] طریقه محاسبه s_i به صورت آنچه در رابطه (۲) آمده، در نظر گرفته شده است. در این مرجع از یک استراتژی ساده با عنوان نزدیک‌ترین همسایه (KNN) به منظور تخمین نزدیک‌ترین نمونه آموزش استفاده شده است. نمونه‌ای که تصور می‌شود مهم‌تر است، مقدار عضویت فازی بالاتر به آن تخصیص داده می‌شود و به مرکز کلاس خود نزدیک‌تر است و نمونه‌ای که کم اهمیت‌تر است از مرکز کلاس خودش فاصله می‌گیرد. این روش ممکن است برخی از نمونه‌های نویزی را به عنوان نمونه نرمال در نظر بگیرد، به دلیل اینکه نمی‌توانیم آنها را عضو نزدیک‌ترین کلاس نمونه در نظر بگیریم. همان‌طور که در این رابطه مشخص است، برای هر نمونه، فاصله تا مرکز دسته محاسبه شده و بر ماکزیمم فاصله داده‌های آن دسته تا مرکز تقسیم می‌شود. بدیهی است هر چه این کسر بزرگتر باشد نشان دهنده داده با اهمیت کمتر است. لذا این مقدار بایستی از یک کم شود.

$$S_i = \begin{cases} 1 - \frac{d_i^{cen^+}}{\max_j(d_i^{cen^+})+\delta} & \text{if } x_i \in O^+ \\ i = 1,2,\ldots,p & \\ 1 - \frac{d_i^{cen^-}}{\max_j(d_i^{cen^-})+\delta} & \text{if } x_i \in O^- \\ i = p+1,p+2,\ldots,l & \end{cases} \tag{2}$$

در [۱۸] که در مورد بازیابی و جستجوی تصاویر مبتنی بر محتوا بحث شده است با انگیزه استفاده از مفاهیم منطق فازی به خصوص ماشین بردار پشتیبان فازی، به منظور توسعه یک تابع عضویت فازی به عنوان یک ارزش نرم و ملایم با توجه به ارتباط متناسب با برچسب کلاس‌ها مطرح شده است. از آنجا که خوشه‌بندی به شکل کلاس مثبت و منفی به طور جداگانه انجام شده است و خوشه‌های متعدد برای هر کلاس (تصاویر مربوط یا نامربوط) فراهم می‌گردد، خوشه‌های به دست‌آمده در هر کلاس می تواند جهت تولید مقدار عضویت بکار گرفته شود. تابع تعلق در این مرجع به دو سناریو تقسیم شده است که برای این تحقیق رابطه (۳) را در نظر گرفته‌ایم. همان‌گونه که در [۱۸] توضیح داده شده، برای هر کلاس از داده‌ها یک تابع توزیع سیگموئیدی در نظر گرفته می‌شود که a_2 پارامتر مقیاس‌پذیری و y فاصله هر نمونه تا SVM آموزش داده شده است:

$$\begin{aligned} S2_i &= w_2(X_p) \\ &= \begin{cases} \frac{1}{1+\exp(-a_2 y)} & \text{pseudo-label is positive} \\ \frac{1}{1+\exp(a_2 y)} & \text{otherwise} \end{cases} \end{aligned} \tag{3}$$

در مرجع [۱۴]، در مورد SVM و چالش آن که عدم دقت بالا در داده‌های نویزی است، بحث شده و برای حل این مشکل ایده فازی را به SVM اضافه کرده است تا اثرات ناشی از داده های نامتوازن و نویزی از بین برود. تابع تعلق فازی به صورت رابطه (۴) می‌باشد. در این تابع نیز معیار فاصله تا مرکز دسته برای هر نمونه به عنوان معیاری برای اهمیت آن تعیین شده است:

$$S3_i = f_{lin}^{cen}(x_i) = 1 - \frac{d_i^{cen}}{\max(d_i^{cen}) + \Delta} \tag{4}$$

در این مقاله با هدف بهره‌مندی از مزایای برخی از این توابع تعلق ((۲)-(۴)) و افزایش کارایی روش ماشین بردار پشتیبان فازی در زمینه تشخیص ربات‌های وب، آنها را با یکدیگر ترکیب خطی وزن‌دار می‌نماییم. در رابطه (۵) مقدار w وزن هر کدام از توابع تعلق است که به صورت برابر در نظر گرفتیم و از این رابطه استفاده کردیم. برای هر تابع تعلق s_i، یک وزن

w_i در نظر گرفته شده است. تعلق نهایی اختصاص داده شده برای جمع خطی تعلق‌هاست.

$$s'_i = \sum_{i=1}^{n} w_i s_i \quad w_i = \frac{1}{n} \quad i = 1, \ldots, n \quad n = 3 \qquad (5)$$

می‌توان برای هر روش وزن متفاوتی در نظر گرفت، ولی در این مقاله باتوجه به اینکه دید دقیقی از قدرت هیچ کدام از روش‌های ذکر شده نیست وزن را مساوی در نظر گرفتیم.

۳ ارزیابی

مجموعه داده‌ای که برای این تحقیق در نظر گرفته شده است یک فایل ورود به سیستم دانشگاه برای یک ماه است شامل ۱۸۷۴۲ نشست است که ۶۳۹۵ نمونه مربوط به خزنده‌های وب و بقیه دسترسی‌های انسان هستند. در روش مورد استفاده در این مقاله، طبقه‌بندی را توسط روش جدید ماشین بردار پشتیبان فازی که تابع تعلق آن ترکیبی از توابع تعلق مراجع [۱۶-۱۸] است، انجام می دهیم. تمامی پیاده‌سازی‌ها با استفاده از نرم‌افزار MATLAB 2011 انجام شده است.

ما در این مقاله برای ارزیابی از معیارهای دقت[1]، ویژگی[2]، حساسیت[3] و امتیاز F1[4] استفاده می‌کنیم که از جمله مهم‌ترین معیارهای مورد استفاده در مسائل طبقه‌بندی[5] هستند.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \qquad (6)$$

$$\text{Specificity} = \frac{TP}{TP + FN} \qquad (7)$$

$$\text{Sensitivity} = \frac{TN}{FP + TN} \qquad (8)$$

$$\text{F1-Score} = \frac{2 * TP}{(2 * TP) + FP + FN} \qquad (9)$$

برای رسیدن به معیارهای ارزیابی الگوریتم‌های طبقه‌بندی، آنها را برای یک مسئله با دو دسته ربات و غیر ربات در نظر می‌گیریم و دسته بندی را بر اساس معیارهای ارزیابی زیر انجام داده‌ایم. طبق تعریف داریم:

- TP[6] برابر تعداد نمونه‌هایی است که متعلق به کلاس غیرربات هستند و الگوریتم نیز برای آن‌ها کلاس غیرربات را پیش‌بینی نموده است.
- FP[7] برابر تعداد نمونه‌هایی است که متعلق به کلاس غیرربات هستند و الگوریتم برای آن‌ها کلاس ربات را پیش‌بینی نموده است.
- TN[8] برابر تعداد نمونه‌هایی است که متعلق به کلاس ربات هستند و الگوریتم نیز برای آن‌ها کلاس ربات را پیش‌بینی نموده است.
- FN[9] برابر تعداد نمونه‌هایی است که متعلق به کلاس ربات هستند و الگوریتم برای آن‌ها کلاس غیرربات را پیش‌بینی نموده است.

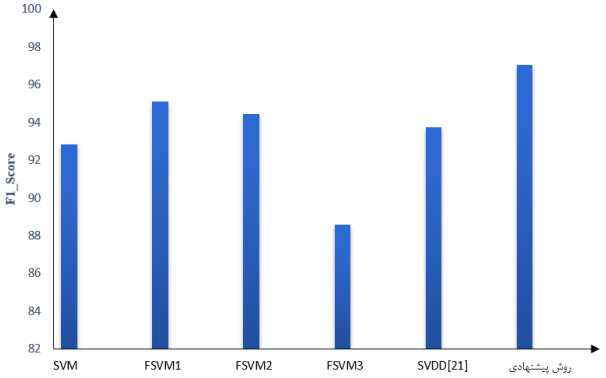
[1]Accuracy [2]Specificity [3]Sensitivity [4]F1-Score [5]Classification [6]True Positive [7]False Positive [8]True Negative [9]False Negative

جدول ۱. نتایج به دست آمده مقادیر TP، FP، TN و FN حاصل از اعمال روش پیشنهادی روی مجموعه داده مورد استفاده

	TP	TN	FP	FN
FSVM1	۳۹	۲۷	۲	۲
FSVM2	۳۴	۳۲	۱	۳
FSVM3	۳۵	۳۱	۲	۲
SVDD [۲۱]	۳۰	۳۶	۲	۲
روش پیشنهادی	۳۳	۳۵	۱	۱

جدول ۲. مقایسه دقت و ویژگی و حساسیت روش پیشنهادی (بر حسب درصد)

	Accuracy	Specificity	Sensitivity	F1-score
FSVM1	۹۴٫۲۸	۹۵٫۱۲	۹۳٫۱۰	۹۵٫۱۲
FSVM2	۹۴٫۲۸	۹۱٫۸۹	۹۶٫۹۷	۹۴٫۴۴
FSVM3	۹۴٫۲۸	۹۴٫۵۹	۹۳٫۹۳	۸۸٫۶۰
SVDD [۲۱]	۹۴٫۲۸	۹۳٫۷۵	۹۴٫۷۴	۹۳٫۷۵
روش پیشنهادی	۹۷٫۱۴	۹۷٫۰۶	۹۷٫۲۲	۹۷٫۰۵



شکل ۲. نمودار مقایسه F1-score الگوریتم‌ها

نتایج به دست آمده برای مقادیر TP، FP، TN و FN درجدول ۱ نمایش داده شده اند.

مقادیر معیارهای ارزیابی با استفاده از توابع تعلق FSVM1 [۱۳] و FSVM2 [۱۵] و FSVM3 [۱۴] و روش پیشنهادی به دست آمده است.

معیارهای FP، TN و FN در روش پیشنهادی بهتر عمل کرده‌اند، اما در مورد TP نتیجه دلخواه گرفته نشده است. در ادامه به بررسی مقادیر نمونه‌ها با استفاده از روابط ((۶)-(۹)) می‌پردازیم.

همان طور که در جدول ۲ مشاهده می شود نتایج روش پیشنهادی برای معیارهای دقت، حساسیت، ویژگی و امتیاز F1 بهتر بوده است.

در شکل ۲ به بررسی نموداری الگوریتم‌ها و روش پیشنهادی براساس معیار F1-score پرداخته شده است.

۴ نتیجه

یکی از مهم‌ترین چالش‌ها در حوزه‌ی وب‌کاوی مسئله تشخیص ربات‌های مخرب وب است. در این مقاله، با توجه به توانایی ماشین‌های بردار پشتیبان فازی در حل مسائل طبقه‌بندی، با ارائه یک روش ترکیبی از مزایای توابع تعلق مختلف ارائه شده بهره‌برداری گردید. با توجه به ارزیابی نتایج مشخص شد که رویکرد پیشنهاد شده عملکرد بهتری نسبت به سایر روش‌ها داشت. با توجه به نتایج این پژوهش، مهم‌ترین پیشنهاداتی که می‌تواند به‌عنوان چارچوبی برای تحقیقات آتی مدنظر قرار گیرند عبارتند از:

- استفاده از دیگر توابع تعلق ارائه شده برای ترکیب و ایجاد ماشین بردار پشتیبان فازی جدید و مقایسه نتیجه با نتایج به‌دست آمده در این تحقیق.
- در فرمول پیشنهادی برای ترکیب توابع تعلق وزن‌ها را برابر در نظر گرفتیم که پیشنهاد می‌شود با استفاده از الگوریتم‌های بهینه‌سازی مثل ژنتیک برای هر تابع تعلق یک وزن مجزا و متفاوت با توجه به درصد اهمیت آن محاسبه نماییم.
- استفاده از الگوریتم‌های ترکیبی تکاملی جهت ترکیب با ماشین بردار پشتیبان فازی و استفاده در تشخیص ربات‌های وب

مراجع

[1] Chang, Yin-Wen and Lin, Chih-Jen. Feature ranking using linear svm. in *Causation and prediction challenge*, pp. 53–64. PMLR, 2008.

[2] Tan, Pang-Ning and Kumar, Vipin. Discovery of web robot sessions based on their navigational patterns. in *Intelligent Technologies for Information Analysis*, pp. 193–222. Springer, 2004.

[3] Lu, Wei-Zhou and Yu, Shun-Zheng. Web robot detection based on hidden markov model. in *2006 International Conference on Communications, Circuits and Systems*, vol. 3, pp. 1806–1810. IEEE, 2006.

[4] Ahn, Luis von, Blum, Manuel, Hopper, Nicholas J, and Langford, John. Captcha: Using hard ai problems for security. in *International conference on the theory and applications of cryptographic techniques*, pp. 294–311. Springer, 2003.

[5] Tan, Pang-Ning and Kumar, Vipin. Modeling of web robot navigational patterns. in *Workshop on Web Mining for E-Commerce, Challenges and Opportunities Working Notes*, pp. 111–117. (KDD2000), Boston, MA, 2000.

[6] Suchacka, Grażyna and Sobkow, Mariusz. Detection of internet robots using a bayesian approach. in *2015 IEEE 2nd International Conference on Cybernetics (CYBCONF)*, pp. 365–370. IEEE, 2015.

[7] Stevanovic, Dusan, Vlajic, Natalija, and An, Aijun. Detection of malicious and non-malicious website visitors using unsupervised neural network learning. *Applied Soft Computing*, 13(1):698–708, 2013.

[۸] سبزه کار، م.، هاشمی، س.ر.، رزم دیده، پ.، و کمال پور، ا. تشخیص ربات‌های وب با استفاده از الگوریتم طبقه بند تک کلاسه با قیود فازی. در بیست و چهارمین کنفرانس انجمن کامپیوتر ایران، تهران، ایران، ۱۳۹۷. دانشگاه صنعتی شریف.

[9] Hamidzadeh, Javad, Zabihimayvan, Mahdieh, and Sadeghi, Reza. Detection of web site visitors based on fuzzy rough sets. *Soft Computing*, 22(7):2175–2188, 2018.

[10] Zhu, Weiping, Gao, Hang, He, Zongjian, Qin, Jiangbo, and Han, Bo. A hybrid approach for recognizing web crawlers. in *International Conference on Wireless Algorithms, Systems, and Applications*, pp. 507–519. Springer, 2019.

[11] Cabri, Alberto, Suchacka, Grażyna, Rovetta, Stefano, and Masulli, Francesco. Online web bot detection using a sequential classification approach. in *2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, pp. 1536–1540. IEEE, 2018.

[12] Lin, C-fu, Wang, S-de, et al. Fuzzy support vector machines with automatic membership setting. in *Support vector machines: Theory and applications*, pp. 233–254. Springer, 2005.

[13] Tang, Hao and Qu, Liang-sheng. Fuzzy support vector machine with a new fuzzy membership function for pattern classification. in *2008 International Conference on Machine Learning and Cybernetics*, vol. 2, pp. 768–773. IEEE, 2008.

[14] Lin, Chun-Fu and Wang, Sheng-De. Fuzzy support vector machines. *IEEE transactions on neural networks*, 13(2):464–471, 2002.

[15] Stevanovic, Dusan, An, Aijun, and Vlajic, Natalija. Feature evaluation for web crawler detection with data mining techniques. *Expert Systems with Applications*,

39(10):8707–8717, 2012.

[16] Ju, Zhe, Cao, Jun-Zhe, and Gu, Hong. Predicting lysine phosphoglycerylation with fuzzy svm by incorporating k-spaced amino acid pairs into chou′ s general pseaac. *Journal of Theoretical Biology*, 397:145–150, 2016.

[17] Dai, Hong-Liang. Class imbalance learning via a fuzzy total margin based support vector machine. *Applied Soft Computing*, 31:172–184, 2015.

[18] Wu, Kui and Yap, Kim-Hui. Fuzzy svm for content-based image retrieval: a pseudo-label support vector machine framework. *IEEE Computational Intelligence Magazine*, 1(2):10–16, 2006.

[19] Stassopoulou, Athena and Dikaiakos, Marios D. Web robot detection: A probabilistic reasoning approach. *Computer Networks*, 53(3):265–278, 2009.

[20] Doran, Derek and Gokhale, Swapna S. Web robot detection techniques: overview and limitations. *Data Mining and Knowledge Discovery*, 22(1):183–210, 2011.

[21] Tax, David MJ and Duin, Robert PW. Support vector data description. *Machine learning*, 54(1):45–66, 2004.