ایجاد پایگاه قواعد فازی به کمک بردارهای پشتیبان

سهیلا نجفی گوجانی[۱]، سید حمید محمودیان *[۲،۳]

۱- دانش‌آموخته کارشناسی‌ارشد، دانشکده مهندسی برق، واحد نجف‌آباد، دانشگاه آزاد اسلامی، نجف‌آباد، ایران

*۲- استادیار، دانشکده مهندسی برق، واحد نجف‌آباد، دانشگاه آزاد اسلامی، نجف‌آباد، ایران، h_mahmoodian@pel.iaun.ac.ir

۳- مرکز تحقیقات پردازش دیجیتال و بینایی ماشین، واحد نجف‌آباد، دانشگاه آزاد اسلامی، نجف‌آباد، ایران

تاریخ دریافت: ۱۳۹۹/۸/۱ تاریخ پذیرش: ۱۳۹۹/۱۲/۱۰

چکیده: در این مقاله یک سیستم استخراج قواعد فازی طراحی شده و ایدهٔ استفاده از این روش در طراحی پایگاه قواعد فازی مورد توجه قرار گرفته است. ابتدا ماشین بردار پشتیبان فازی مبتنی بر روش خوشه‌بندی طراحی شده است. دراین صورت ماشین بردار پشتیبان عملکرد بهتری در داده‌های نویزی یا داده‌های خارج از محدوده خواهد داشت و در نتیجه به نظر می‌رسد که قواعد استخراج شده از آن، دقت بالاتری داشته باشند. مدل ماشین بردار پشتیبان فازی شبیه‌سازی شده و سپس از بردارهای پشتیبان آن برای استخراج قواعد فازی استفاده می‌شود. قواعد استخراج شده، قواعد فازی هستند که به شکل "اگر-آنگاه" به‌دست می‌آیند. استخراج قواعد به کمک دو مدل ماشین بردارپشتیبان ساده و ماشین بردار پشتیبان فازی صورت گرفته که نشان داده خواهد شد که قواعد استخراج شده از ماشین بردار پشتیبان فازی در اکثر موارد دقت بهتری در طبقه‌بندی نسبت به ماشین بردار پشتیبان معمولی دارند و با توجه به این موضوع که تعداد قواعد استخراج شده پارامتر مهمی در صحت عملکرد این پایگاه محسوب می‌شود، تعداد قواعد استخراج شده در روش فازی در اکثر موارد کمتر است.

واژه‌های کلیدی: ماشین‌بردارپشتیبان، پایگاه‌قواعدفازی، بردارپشتیبان، خوشه‌بندی‌فازی، طبقه‌بندهای‌فازی

۱- مقدمه

سیستم‌های فازی[۱]، سیستم‌های مبتنی بر قواعد هستند که دارای توانایی‌هائی مانند طبقه‌بندی[۲] وخوشه‌بندی[۳] داده‌ها هستند و در دو دهه اخیر استفاده از این سیستم‌ها به عنوان طبقه‌بندی الگو مورد توجه محققان قرارگرفته است [۱-۳]. روش ماشین بردارپشتیبان[۴] (SVM)، یکی از روش‌های قدرتمند و شناخته شده در طبقه‌بندی است و استفاده از این روش در طراحی پایگاه قواعد فازی مورد توجه قرار گرفته است [۴-۶].

استخراج قوانین از مدل‌های طبقه‌بندی مانند ماشین بردار پشتیبان، بخشی مهم از مسائل یادگیری ماشین را به خود اختصاص می‌دهد. برای حل این مساله، ابتدا نیاز به طراحی ماشین بردار پشتیبان و یا نسخه‌ای مناسب‌تر از آن است. به همین دلیل ماشین بردار پشتیبان فازی[۵] با رویکرد خوشه‌بندی طراحی می‌گردد. پیش از استخراج قوانین باید مدل طراحی شده توسط معیارهایی نظیر دقت سنجیده شده و از عملکرد آن اطمینان حاصل شود. در مرحله بعد، بردارهای پشتیبان به‌دست آمده از مدل طراحی شده را استخراج نموده و از آن برای استخراج قوانین کمک گرفته می‌شود. استخراج قوانین به این صورت است که با توجه به ویژگی‌های موجود در هر مجموعه داده و طبقه‌بندی هر کدام از آن‌ها، به هر ویژگی متغیرهای زبانی خاص خود اختصاص داده می‌شود. این متغیرها بیانگر محدوده خاصی از هر ویژگی هستند و توابع عضویت[۶] آن‌ها به‌صورت مثلثی در نظر گرفته می‌شود. پس از تعیین توابع عضویت، برای هر ناحیه به‌دست آمده به کمک داده‌های آموزش برچسب و یا همان بخش دوم قوانین را به‌دست می‌آید. در نتیجه قوانین فازی برای هر کدام از مجموعه داده‌ها و به کمک بردارهای پشتیبان حاصل می‌گردد.

تحقیقات زیادی بر روی استخراج قواعد فازی[۷] از ماشین‌های بردار پشتیبان تاکنون انجام شده است [۷]. اکثر این تحقیقات به گونه‌ای بوده‌اند که در صدد ارتقای قانون استخراج شده با تغییر سیاست استخراج قانون هستند.

[DOR: 20.1001.1.23222468.1399.9.4.6.0]
Journal of Novel Researches on Electrical Power - Vol. 9-No. 4- winter 2021

در مقاله [۸] عملکرد سیستم هوشمند پشتیبانی تصمیم گیری سخت افزاری با استفاده از ماشین های بردار پشتیبانی بررسی می شود. که هدف این سیستم پیش بینی بهره وری آینده بر اساس داده های تهیه شده توسط کارشناسان میدانی و به دنبال آن عوامل تأثیر بهره وری است. این ویژگی با ترکیب منطق فازی و ماشین بردار پشتیبانی انجام میشود.

الگوریتم ماشین بردار پشتیبانی فازی مبتنی بر SVDD در مقاله [۹] ارائه شده که در آن نویزها و پرتگاه ها توسط یک ابر کره با حداقل حجم در حالی که حداکثر نمونه‌ها را دارند، مشخص می‌شوند. در مقایسه با تعریف سنتی عضویت فازی بر اساس رابطه بین یک نمونه و مرکز خوشه آن، این روش به‌طور موثری سر و صداها یا فاصله‌ها را از بردارهای پشتیبانی متمایز می‌کند و ضرایب وزن مناسب را به آنها اختصاص می‌دهد حتی اگر در مرز بین مثبت و منفی توزیع شده باشند.

در این مقاله سعی شده است که اثر تغییر و بهبود ماشین بردار پشتیبان بر قوانین استخراج شده مشاهده گردد. به عبارت دیگر سؤالات زیر بخش مهمی از مسائل این پژوهش را پوشش می‌دهند:

- آیا تغییر مدل ماشین بردار پشتیبان و بهبود آن باعث بهبود دقت قوانین استخراج شده می‌گردد؟
- آیا در صورت بهبود نتایج، پیچیدگی محاسباتی بالا خواهد رفت؟

ساختار مقاله به این شرح است: در قسمت دوم به ماشین‌های بردار پشتیبان اشاره می‌شود. در قسمت سوم نتایج و بحث براساس داده‌های موجود بیان می‌شود. در قسمت چهارم نیتجه‌گیری مقاله آمده است.

۲- ماشین‌های بردار پشتیبان

الگوریتم ماشین بردار پشتیبان بر اساس پیشرفت در نظریه یادگیری آماری پیشنهاد شد و در سالیان اخیر بسیاری از پژوهش‌ها به این موضوع اختصاص پیدا کرده و به ابزار پیشرفته‌ای برای حل مسائل طبقه‌بندی، رگرسیون و انتخاب ویژگی تبدیل شده است [۱۰،۱۱].

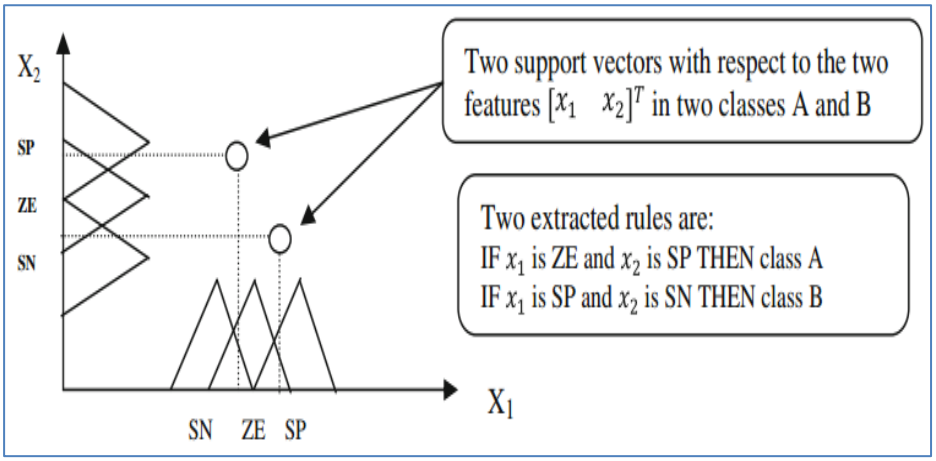
در مسائل طبقه‌بندی SVM یک ابرصفحه[۸] بهینه چنان تعیین می‌کند که داده‌های طبقات مختلف را از هم جدا کند. این ابر صفحه بنا به نظریه یادگیری آماری بیشترین قابلیت تعمیم‌دهی را برای داده‌هایی که هنگام آموزش از آن‌ها استفاده نشده (داده‌های تست) داراست و به همین دلیل از کلمه بهینه استفاده شده است. این ابر صفحه بهینه به‌وسیله حل یک مسئله بهینه‌سازی حاصل می‌شود. همچنین قادر است داده‌هایی با الگوهای پیچیده را که به‌صورت خطی جداپذیر نیستند، با اعمال یک تبدیل غیرخطی که به‌عنوان کرنل[۹] شناخته می‌شود، به فضای جدیدی برده و در آن به کمک ابرصفحه از یکدیگر جدا کند. این قابلیت SVMها را در بسیاری از مسائل واقعی همچون تشخیص چهره و تشخیص سرطان سینه[۱۰] دارای کاربرد کرده است.

روش‌های یادگیری که اغلب مورد استفاده قرار می‌گیرند بر اساس کمینه‌کردن خطا بر روی داده‌های آموزش هستند که به آن کمینه-سازی ریسک تجربی اطلاق می‌شود. روش‌هایی مانند شبکه‌های عصبی[۱۱] و درخت‌های تصمیم‌گیری[۱۲] از نمونه‌های این دسته هستند. از طرف دیگر، بردار پشتیبان بر اساس کمینه سازی ریسک ساختاری است که دسته دیگری از روش‌ها را تشکیل می‌دهند [۱۲،۱۳]. بردار پشتیبان قدرت تعمیم‌دهی بالاتری دارد و کمینه‌کردن ریسک ساختاری از طریق کمینه نمودن حد بالای خطای تعمیم‌دهی به‌دست می‌آید [۱۴،۱۵].

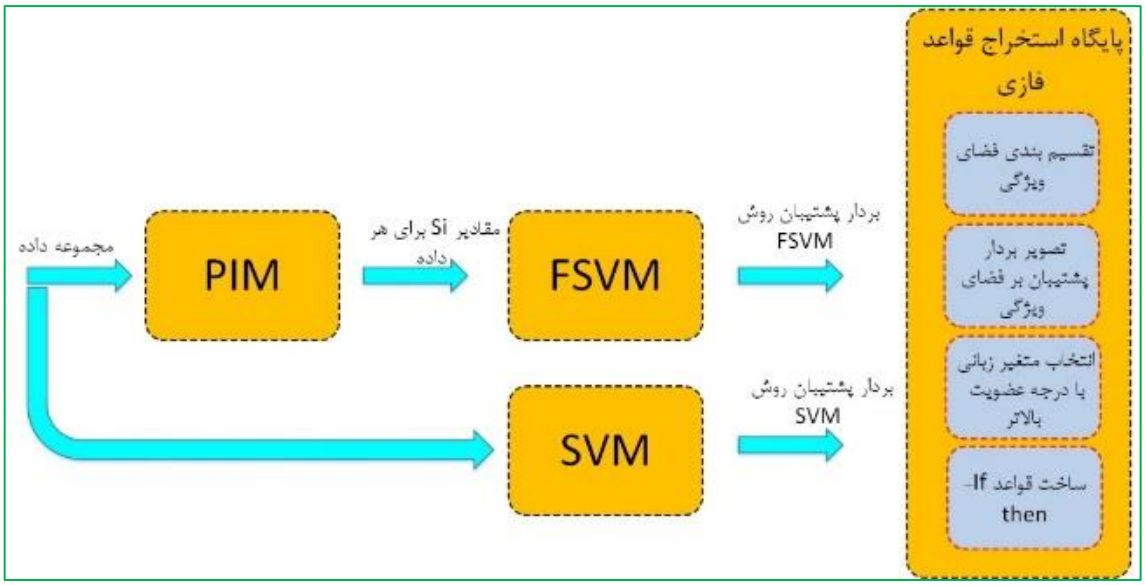
روش‌های متفاوتی در تعیین قواعد فازی استفاده می‌شوند که می‌توان به چهار روش زیر اشاره کرد [۱۶-۱۹]:

- ساخت قواعد فازی با استفاده از میانگین و واریانس[۱۳] ویژگی‌ها
- ساخت قواعد فازی با استفاده از هیستوگرام[۱۴] ویژگی‌ها
- ساخت قواعد فازی با استفاده از قطعیت ویژگی‌ها
- ساخت قواعد فازی با استفاده از تقسیم فضاهایی که هم‌پوشانی دارند

در روش استخراج قواعد با کمک بردارهای پشتیبان با توجه به شکل (۱) ابتدا فضای ویژگی به تعداد مشخصی بازه تقسیم می‌شود و برای هر بازه یک تابع عضویت مشخص قرار داده می‌شود [۲۰].



شکل (۱): استفاده از بردار پشتیبان برای استخراج قواعد



شکل (۲): پایگاه استخراج قواعد فازی

هر داده بردار پشتیبان از چند ویژگی تشکیل شده است. هرکدام از ویژگی‌های داده بردار پشتیبان در یکی از توابع عضویت مقدار بیشتری خواهند داشت. آن تابع عضویت به عنوان متغیر زبانی انتخاب می‌شود. در نتیجه به تعداد بردارهای پشتیبان قانون وجود دارد.

پس از تعیین قوانین برای هر بردار پشتیبان، داده تست را در تمام قوانین قرار داده می‌شود. مقدار سازگاری[۱۵] برای هر قانون از حاصل‌ضرب مقادیر عضویت برای هر ویژگی به‌دست می‌آید. قانونی که بیشترین مقدار سازگاری را داشته باشد، خروجی مورد نظر را می‌دهد و کلاس آن قانون که متناظر با کلاس بردار پشتیبان آن قانون است، به‌عنوان کلاس داده تست انتخاب می‌گردد.

درنهایت می‌توان خلاصه روش مورد نظر را مطابق شکل (۲) به این‌صورت بیان نمود: ابتدا دو مدل FSVM-PIM[۱۶] و SVM ساخته می‌شوند. سپس بردارهای پشتیبان استخراج شده و بر روی فضای ویژگی تصویر می‌شوند. در نهایت قوانین If-Then با توجه به متغیرهای زبانی انتخاب‌شده ساخته می‌شوند. پس از تعین قواعد به کمک داده‌های آزمایش ابتدا برای هر کلاس، طبقه‌بند FSVM با کمک درجه‌های عضویت فازی به‌دست آمده طراحی می‌گردد و سپس برای هر نمونه داده آزمون، نتیجه هر طبقه‌بند FSVM محاسبه می‌گردد و در نهایت به کمک روش رأی‌گیری حداکثری برچسب کلاس مشخص می‌گردد.

۳- نتایج و بحث

مجموعه داده های مورد بررسی، در دو دسته و در جدول‌های (۱) و (۲) آمده که به ترتیب برای بررسی و مقایسه دقت در روش SVM و FSVM و برای استخراج قواعد به کار گرفته شده است.

برنامه‌نویسی نیز در محیط متلب صورت گرفته است. در زیر به توضیح مجموعه داده‌های مورد استفاده می‌پردازیم:

جدول (۱): ویژگی‌های مجموعه داده‌های استفاده‌شده جهت استخراج قواعد

شماره	مجموعه داده	تعداد ویژگی	تعداد نمونه
۱	Ripley	۲	۱۲۵۰
۲	Iris	۴	۱۵۰
۳	Herberman	۳	۳۰۶
۴	Bank note	۴	۱۳۷۲
۵	Hayer Roth	۵	۱۳۲

جدول (۲): ویژگی‌های مجموعه داده‌های استفاده‌شده جهت شبیه سازی

شماره	مجموعه داده	تعداد ویژگی	تعداد نمونه
۱	Ripley	۲	۱۲۵۰
۲	PIMA	۸	۷۶۸
۳	Banana	۲	۵۳۰۰
۴	Monk	۷	۵۵۴
۵	Yeast	۸	۸۹۲

۱. مجموعه داده Monk: این مجموعه داده از ۵۵۶ نمونه تشکیل شده است که هر نمونه ۷ ویژگی دارد. تعداد ۳۸۰ نمونه به‌صورت تصادفی برای فرآیند آموزش و ۱۸۶ نمونه برای آزمایش استفاده شده‌اند. این مجموعه داده ۲ کلاس دارد.

۲. مجموعه داده Hayer roth: این مجموعه ۱۳۲ نمونه و ۶ ویژگی دارد که در ۶ کلاس قرار گرفته‌اند. تعداد ۷۰ نمونه به‌صورت تصادفی برای آموزش و ۶۲ نمونه برای آزمایش انتخاب شده است.

۳. مجموعه داده Iris: این مجموعه ۱۵۰ نمونه دارد و هر نمونه از ۴ ویژگی تشکیل شده است. نمونه‌ها در ۳ دسته جای می‌گیرند و تعداد ۱۰۰ نمونه به‌صورت تصادفی برای آموزش و ۵۰ نمونه برای آزمایش انتخاب شده‌اند.

۴. مجموعه داده Bank note: این مجموعه داده ۱۳۷۲ نمونه دارد که از ۵ ویژگی تشکیل شده‌اند. تعداد کلاس‌های این مجموعه ۲ است و

۶۰۰ نمونه به‌صورت تصادفی برای آموزش و مابقی برای آزمایش استفاده شده است.

۵. مجموعه داده Herberman: این مجموعه ۳۰۶ نمونه و ۴ ویژگی دارد. این مجموعه ۳ کلاس دارد و تعداد ۲۰۰ نمونه به‌صورت تصادفی برای آموزش انتخاب شده‌اند.

۶. مجموعه داده Ripley: این مجموعه داده ۱۲۵۰ نمونه دارد که از ۳ ویژگی تشکیل شده‌اند. تعداد کلاس‌های این مجموعه ۲ است و۷۲۰ نمونه به‌صورت تصادفی برای آموزش و مابقی برای آزمایش استفاده شده است.

۷. مجموعه داده Banana: این مجموعه داده ۵۳۰۰ نمونه دارد که از ۳ ویژگی تشکیل شده‌اند. تعداد کلاس‌های این مجموعه ۳ است و ۳۱۵۰ نمونه به‌صورت تصادفی برای آموزش و مابقی برای آزمایش استفاده شده است.

۸. مجموعه داده Pima: این مجموعه داده ۷۶۸ نمونه دارد که از ۹ ویژگی تشکیل شده‌اند. تعداد کلاس‌های این مجموعه ۲ است و ۴۲۰ نمونه به‌صورت تصادفی برای آموزش و مابقی برای آزمایش استفاده شده است.

۹. مجموعه داده Yeast: این مجموعه داده ۸۹۲ نمونه دارد که از ۹ ویژگی تشکیل شده‌اند. تعداد کلاس‌های این مجموعه ۲ است و ۵۲۰ نمونه به‌صورت تصادفی برای آموزش و مابقی برای آزمایش استفاده شده است.

در مساله طبقه‌بندی به روش ماشین بردار پشتیبان، عملکرد سیستم به داده‌هایی که بسیار دورتر از ناحیه کلاس خود هستند حساس است. این گونه داده‌ها تحت عنوان نویز یا داده‌های خارج از محدوده نیز شناخته می‌شوند. برای مقابله با این اثر نامطلوب رویکرد فازی به داده‌های مورد نظر پیشنهاد شده است به این ترتیب که برای هر داده یک میزان عضویت به کلاس خود فرض می‌شود. این مقدار برای داده‌های خارج از محدوده نزدیک به صفر و برای داده‌هایی که به خوبی قابل تمایز از سایر کلاس‌ها هستند نزدیک به یک است. با این روش داده‌های خارج از محدوده در تابع هزینه بهینه‌سازی ماشین بردار پشتیبان اهمیت کمتری داشته و داده‌های با عضویت بالاتر در این تابع نقش مهمتری را ایفا خواهند کرد.

یکی از مهمترین قسمت‌های ماشین بردار پشتیبان فازی تعیین میزان عضویت داده‌ها است. برای این کار روش‌های مختلف وجود دارد که با شناسایی مراکز کلاس‌های داده‌ها و تعیین فاصله داده‌ها از آن مراکز، مقدار عضویت هر داده را مشخص می‌کند.

در این بخش با کمک ۵ مجموعه داده معرفی شده، قواعد فازی از مدل ماشین بردار پشتیبان معمولی و همچنین مدل ماشین بردار پشتیبان فازی استخراج می‌گردند. یکی از معیارهای مهم برای قواعد استخراج شده، دقت قواعد بوده که در هر مورد گزارش شده است. داده‌های مورد استفاده در این قسمت، در جدول (۳) معرفی‌شده‌اند. یکی از نکاتی که باید مورد نظر باشد این است که تعداد ویژگی‌های مورد استفاده نباید زیاد باشد، زیرا این ویژگی‌ها در بخش مقدم قوانین حاضر می‌شوند و در نتیجه تعداد بالای آن‌ها باعث پیچیدگی قوانین و افزایش قابل ملاحظه تعداد قوانین می‌گردند. همان‌طور که در جدول (۴) نشان داده شده است، تعداد اعضای مجموعه داده‌ها متفاوت می‌بخشد تا ارزیابی روش مورد بررسی، بهتر صورت بگیرد.

ابتدا به بررسی تعداد قواعد استخراج شده توسط هر روش اشاره می‌شود که در جدول (۵) ۰نشان داده شده است. یکی از مزایای روش میانگین-واریانس این است که برای هر ویژگی یک قانون ایجاد می‌نماید. در نتیجه تعداد قوانین استخراج شده برای آن کمتر بوده و تفسیر قوانین راحت‌تر است. بیشترین تعداد قوانین مربوط به روش شبکه‌بندی است. این روش به دلیل تقسیم نمودن فضای ویژگی به-صورت شبکه‌ای، تعداد قوانین بالایی خواهد داشت که تعداد آن‌ها با تعداد ویژگی‌ها رابطه مستقیم دارد. دو روش FSVM و SVM از نقطه نظر تعداد قوانین نسبتاً به هم نزدیک بوده زیرا تفاوت بین مسئله بهینه‌سازی این دو روش معمولاً تغییرات اندکی در بردارهای ویژه ایجاد می‌نماید. نکته قابل توجه این است که در هر دوی این روش‌ها تعداد قوانین از تعداد قوانین روش شبکه‌بندی کمتر است. در نتیجه قوانین استخراج شده از ماشین بردار پشتیبان سرعت بالاتری در ارائه نتایج خواهند داشت.

نتایج دقت طبقه‌بندی در جدول (۵) نشان داده شده است. مشاهده می‌شود که برای چهار مورد از پنج مجموعه داده استفاده شده، دقت طبقه‌بندی توسط قوانین استخراج شده FSVM بالاتر است. این افزایش چشم‌گیر بوده به‌گونه‌ای که در مورد مجموعه داده Iris این دقت به مقدار ۱/۰۰۰ رسیده است که نسبت به دقت به‌دست آمده از قوانین SVM حدود ۰/۱۷ بیشتر است. تنها در مورد مجموعه داده Ripley اندکی کاهش وجود داشته است که مقدار آن قابل چشم‌پوشی است.

جدول (۳): مجموعه داده‌های مورد استفاده برای استخراج قواعد

تعداد نمونه	تعداد ویژگی	مجموع داده
۱۲۵۰	۲	Ripley
۱۵۰	۴	Iris
۳۰۶	۳	Herberman
۱۳۷۲	۴	Bank note
۱۳۲	۵	Hayer Roth

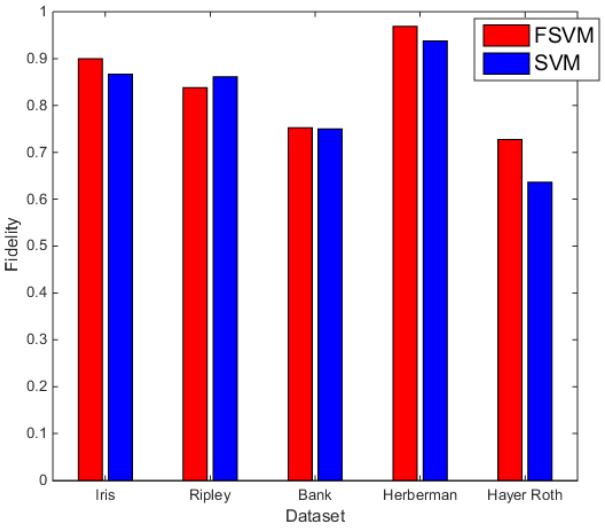
جدول (۴): تعداد قواعد استخراج شده

روش FSVM	روش SVM	روش شبکه‌بندی	روش میانگین واریانس	مجموع داده
۱۸۵	۲۱۲	۳۶	۲	Ripley
۱۴	۱۲	۱۲۹۶	۴	Iris
۱۷۵	۱۶۳	۲۱۶	۳	Herberman
۳۷۵	۳۷۵	۱۲۹۶	۴	Bank note
۲۶	۵۵	۷۷۷۶	۵	Hayer Roth

جدول (۵): دقت طبقه‌بندی توسط قواعد استخراج شده

مجموع داده	دقت به کمک میانگین واریانس	دقت به کمک شبکه‌بندی	دقت به کمک SVM	دقت به کمک FSVM
Ripley	۰/۸۹۴۰	۰/۹۰۸۰	۰/۷۸۶۰	۰/۷۸۵۰
Iris	۱/۰۰۰	۱/۰۰۰	۰/۸۳۳۳	۱/۰۰۰
Herberman	۰/۶۵۶۳	۰/۶۷۷۱	۰/۷۰۸۳	۰/۷۰۸۳
Bank note	۰/۸۵۸۶	۰/۹۷۰۹	۰/۸۱۰۷	-/۸۱۰۷
Hayer Roth	۰/۷۲۷۳	۰/۶۳۶۴	۰/۶۳۶۴	۰/۷۲۷۳

معیار دیگری که می‌تواند به بررسی قوانین استخراج شده کمک کند سازگاری نام دارد. این معیار نشان‌دهنده سازگار بودن خروجی قوانین استخراج شده و مدل مورد بررسی خواهد بود. به‌عبارت دیگر تعریف این معیار عبارت است از درصد تعداد نمونه‌هایی که خروجی آن‌ها از استخراج قوانین و مدل یکسان بوده است. دلیل استفاده از این معیار، سازگاری قواعد و مدل بیانگر درستی قواعد استخراج شده است، به عبارت دیگر تنها دقت بالای قوانین استخراج شده ملاک تفسیر مدل انتخابی نیست. زیرا ممکن است حتی قوانین استخراج شده دقت بالایی داشته باشند اما مدل مبنای مورد استفاده در تضاد با این قوانین باشد. به همین منظور این معیار در کنار دقت استفاده خواهد شد. معیار سازگاری برای مجموعه داده‌های مورد استفاده در شکل (۳) نشان داده شده است.



شکل (۳): معیار سازگاری برای قوانین استخراج‌شده

همان‌طور که مشاهده می‌شود، معیار سازگاری برای چهار مورد از پنج مجموعه داده مورد استفاده افزایش داشته است. روش پیشنهادی تنها در مجموعه داده Ripley با کاهش سازگاری مواجه شده است. همچنین با توجه به شکل تقریباً میزان سازگاری بالای ۷۰ درصد است و در مورد مجموعه داده Herberman این مقدار به حدود ۹۷ درصد رسیده است که نشان دهنده سازگاری بالای مدل ماشین بردار پشتیبان فازی با قوانین استخراج شده است. با توجه به شکل می‌توان نتیجه گرفت که قوانینی که از روش FSVM-PIM استخراج شده‌اند با مدل سازگاری بیشتری دارند و در نتیجه اعتبار آن‌ها بالاتر است.

۴- نتیجه‌گیری

در این مقاله روش استخراج قواعد فازی مبتنی بر ماشین بردار پشتیبان فازی برای مجموعه داده‌های نویی و دارای داده خارج از محدوده ارائه گردید. ابتدا برای مجموعه داده‌های مورد استفاده، ماشین بردار پشتیبان طراحی شد و سپس با کمک روش ماشین بردار پشتیبان فازی، دقت طبقه‌بندی برای این روش افزایش داده شد. در روش ماشین بردار پشتیبان فازی نیاز به تعیین مقدار عضویت هرکدام از نمونه‌ها به دسته‌ای خاص است که با کمک روش خوشه‌بندی PIM انجام شد. در این روش مجموعه داده به تعدادی دسته تقسیم می‌شود و سپس درجه عضویت هر مجموعه داده به کمک مقدار عضویت به هر دسته تعیین می‌گردند. در نهایت با توجه به تفاوت دو روش در خروجی و بردارهای پشتیبان، قواعد فازی برای این روش استخراج شد. همان‌طور که اشاره شد، روش‌های متفاوتی برای استخراج قواعد فازی با رویکرد طبقه‌بندی وجود دارد. دو روش استخراج به کمک میانگین و واریانس و استخراج قواعد فازی به کمک روش شبکه‌ای مبنای مقایسه قرار گرفته‌اند. نتایج عددی نشان می‌دهد که استخراج قواعد با کمک روش ماشین بردار پشتیبان فازی، در اکثر موارد از استخراج به کمک ماشین بردار پشتیبان معمولی بهتر عمل نموده است. دو معیار دقت و معیار سازگاری برای مقایسه قواعد فازی استخراج شده استفاده شد. معیار دقت بیانگر درستی قواعد فازی در حوزه دانشی مورد استفاده خواهد بود و هر چه میزان بالاتری را به خود اختصاص دهد نشان دهنده مطلوبیت بالای سیستم خواهد بود. معیار دقت در پنج مجموعه داده مورد استفاده اعداد بالایی را به خود اختصاص داده است. مقایسه معیار دقت نشان می‌دهد که قوانین استخراج شده از ماشین بردار پشتیبان فازی مبتنی بر خوشه‌بندی PIM دقت بالاتری از خود نشان می‌دهد. معیار دوم سازگاری است که نشان‌دهنده مطابقت قوانین استخراج شده با مدل ماشین بردار پشتیبان است. همان‌طور که اشاره شده است مطابقت بالا نشان‌دهنده تفسیر مناسب مدل توسط قوانین خواهد بود. نتایج عددی نشان می‌دهد که در اکثر موارد میزان سازگاری بالای ۷۰ درصد است و در بعضی موارد این عدد به بالای ۹۰ درصد می‌رسد. نتایج نشان می‌دهد که قواعد استخراج شده از ماشین بردار پشتیبان فازی در چهار مورد از پنج مجموعه داده مورد استفاده سازگاری بهتری نسبت به ماشین بردار پشتیبان معمولی دارند. همچنین تعداد قواعد استخراج شده برای روش‌ها با یکدیگر مقایسه شده‌اند. نتایج نشان می‌دهد که تعداد قواعد در روش فازی در اکثر موارد کمتر است.

مراجع

[۱] فقیه‌نیا الهه، کامل‌طباخ‌فریضنی سیدرضا، خیرآبادی مریم، **"سیستم تشخیص نفوذ بهبود یافته مبتنی بر الگوریتم ژنتیک خود تطبیق جزیره ای برای حل ماشین بردار پشتیبان به‌صورت یادگیری چندهسته‌ای با کد کننده‌های خودکار"**، روش‌های هوشمند در صنعت برق، سال ۱۲، شماره ۴۵، ص. ۹۳-۷۹، بهار ۱۴۰۰.

[2] E. Lotfi, "*A novel hybrid system based on fractal coding for soccer retrieval from video database*", Majlesi Journal of Electrical Engineering, vol. 6, no. 1, 2012.

[۳] حاجیان مهدی، اکبری‌فرود اصغر، نوروزیان حسین، **"بررسی پایداری استاتیکی ولتاژ با استفاده از ماشین بردار پشتیبان و شبکه عصبی"**، روش‌های هوشمند در صنعت برق، سال ۴، شماره ۱۳، ص.: ۱۴-۳، بهار ۱۳۹۲.

[۴] معلم سپهر، محمدعلی‌پوراهری رویا، شاهقلیان غضنفر، معظمی مجید، کاظمی سیدمحمد، **"پیش‌بینی بلندمدت تقاضا در "زنجیره تامین انرژی الکتریکی صنایع سنگ آهن اسپیدان" با استفاده از شبکه عصبی عمیق و ماشین یادگیری شدید"**، روش‌های هوشمند در صنعت برق، سال ۱۳، شماره ۴۹، ص.: ۱-۲۰، ۱۴۰۱.

[5] S. Rezvani, X. Wang, F. Pourpanah, "*Intuitionistic fuzzy twin support vector machines*", IEEE Trans. on Fuzzy Systems, vol. 27, no. 11, pp. 2140-2151, Nov. 2019.

[6] H. Mahmoodian, L. Ebrahimian, "*Using support vector regression in gene selection and fuzzy rule generation for relapse time prediction of breast cancer*", Biocybernetics and Biomedical Engineering, vol. 36, no. 3, pp. 466-472, 2016.

[7] C. Liu, W. Wang, M. Wang, F. Lv, M. Konan, "*An efficient instance selection algorithm to reconstruct training set for support vector machine*", Knowledge-Based Systems, vol. 116, pp. 58-73, 2017.

[8] G. Prabakaran, D. Vaithiyanathan, M. Ganesan, "*FPGA based effective agriculture productivity prediction system using fuzzy support vector machine*", Mathematics and Computers in Simulation, vol. 185, pp. 1-16, 2021.

[9] W. Liu, L. Ci, L. Liu, "*A new method of fuzzy support vector machine algorithm for intrusion detection*", Applied Sciences, vol. 10, no. 3, Article Number: 1065, 2020.

[10] Y. Jiang, X.G. Wang, Z.J. Zou, Z.L. Yang, "*Identification of coupled response models for ship steering and roll motion using support vector machines*", Applied Ocean Research, vol. 110, Article Number: 102607, May 2021.

[11] D. Martens, B. Baesens, T.V. Gestel, J. Vanthienen, "*Comprehensible credit scoring models using rule extraction from support vector machines*", European Journal of Operational Research, vol. 183, no. 3, pp. 1466-1476, Dec. 2007.

[12] X. Wang, S. Wang, Z. Huang, Y. Du, "*Condensing the solution of support vector machines via radius-margin bound*", Applied Soft Computing, vol. 101, Article Number: 107071, March 2021.

[13] N. Singh, P. Singh, D. Bhagat, "*A rule extraction approach from support vector machines for diagnosing hypertension among diabetics*", Expert Systems with Applications, vol. 130, pp. 188-205, Sept. 2019.

[14] N.H. Barakat, A.P. Bradley, "Rule extraction from support vector machines: A sequential covering approach", IEEE Trans. on Knowledge and Data Engineering, vol. 19, no. 6, pp. 729-741, June 2007.

[15] F. Borges et al., "*An unsupervised method based on support vector machines and higher-order statistics for mechanical faults detection*", IEEE Latin America Transactions, vol. 18, no. 06, pp. 1093-1101, June 2020.

[16] R. Jain, A. Abraham, "*A comparative study of fuzzy classification methods on breast cancer data*", Australasian Physics & Engineering Sciences in Medicine, vol. 27, pp. 213–218, 2004.

[17] G. Shahgholian, A. Hakimi, N. Behzadfar, "*Motor speed maximum control in the resonance ratio controller for two-mass system using self-organizing fuzzy controller*", International Journal of Research Studies in Electrical and Electronics Engineering, vol. 1, no. 6, pp. 1-8, 2020.

[۱۸] خسروی عادل، چترایی عباس، شاهقلیان غضنفر، کارگر سیدمحمد، **" مدل‌سازی کمپرسور** k-250 **با استفاده از روش سری موازی نارکس و فازی سلسله مراتبی"**، نشریه مهندسی برق و مهندسی کامپیوتر ایران، سال ۱۸، شماره ۳، ص.: ۱۹۱-۱۹۸، پاییز ۱۳۹۹.

[19] E. Aghadavoodi, G. Shahgholian, "*A new practical feed-forward cascade analyze for close loop identification of combustion control loop system through RANFIS and NARX*", Applied Thermal Engineering, Vol. 133, pp. 381-395, 2018.

[20] Y. Chen, T. Wang, B. Wang, *et al.* "*A survey of fuzzy decision tree classifier*", Fuzzy Information and Engineering, vol. 1, pp. 149–159, 2009.

زیرنویس‌ها

1-Fuzzy systems
2-Classification
3-Clustering
4-Support vector machine
5-Fuzzy support vector machine
6-Membership function
7-Fuzzy rules
8-Hyper plane
9-Kernel
10-Breast cancer
11-Neural networks
12-Decision tree
13-Variance
14-Histogram
15-Fidelity
16-Fuzzy SVM-partition index maximization

[DOR: 20.1001.1.23222468.1399.9.4.6.0]
Journal of Novel Researches on Electrical Power - Vol. 9-No. 4- winter 2021

Create a Database of Fuzzy Rules with the Help of Support Vectors

Soheila Najafi Gojani[1], Seyed Hamid Mahmoodian[2,3*]

[1]MSc student, Department of Electrical Engineering, Najafabad Branch, Islamic Azad University, Najafabad, Iran
[*2]Assistant Professor- Department of Electrical Engineering, Najafabad Branch, Islamic Azad University, Najafabad, Iran, h_mahmoodian@pel.iaun.ac.ir
[3]Digital Processing and Machine Vision Research Center, Najafabad Branch, Islamic Azad University, Najafabad, Iran

Abstract: In this paper, a fuzzy rule extraction system is designed and the idea of using this method in designing a fuzzy rule database is considered. First, the fuzzy backup vector machine is designed based on the clustering method. In this case, the backup vector machine will perform better in noise data or out-of-range data, and as a result, the rules extracted from it seem to be more accurate. The fuzzy backup vector machine model is simulated and then its backup vectors are used to derive fuzzy rules. The extracted rules are fuzzy rules that are obtained in the form of "if-then". The rules are extracted using two models of simple support vector machine and fuzzy support vector machine, which will show that the rules extracted from fuzzy backup vector machine in most cases have better accuracy in classification than ordinary backup vector machine. The number of extracted rules is an important parameter in the accuracy of the operation of this database, the number of rules extracted in the fuzzy method is less in most cases.

Keywords: Support vector machine, fuzzy rule base, fuzzy clustering, fuzzy classification.