The 2nd National Conference on
Geospatial Information Technology (NCGIT)

K.N.Toosi University of Technology
Faculty of Geomatics Engineering

18 January 2017

دومین کنفرانس ملی مهندسی فناوری اطلاعات مکانی

دانشگاه صنعتی خواجه نصیرالدین طوسی
دانشکده مهندسی نقشه‌برداری
۲۹ دی ماه ۱۳۹۵

مروری جامع بر روش‌های یادگیری ماشین و انتخاب ویژگی با تأکید بر موضوع طبقه‌بندی در کاربردهای سنجش‌ازدوری

مینا حمیدی[۱*]، حمید عبادی[۲]، عباس کیانی[۳]

۱- دانشجوی کارشناسی ارشد سنجش‌ازدور، دانشکده‌ی مهندسی ژئودزی و ژئوماتیک، دانشگاه صنعتی خواجه‌نصیرالدین طوسی
۲- استاد گروه فتوگرامتری و سنجش‌ازدور، دانشکده‌ی مهندسی ژئودزی و ژئوماتیک، دانشگاه صنعتی خواجه‌نصیرالدین طوسی
۳- دانشجوی دکترای فتوگرامتری، دانشکده‌ی مهندسی ژئودزی و ژئوماتیک، دانشگاه صنعتی خواجه‌نصیرالدین طوسی

چکیده :

امروزه، با توجه به پتانسیل بالای تصاویر ماهواره‌ای و پیشرفت روزافزون تکنولوژی‌های اخذ اطلاعات از فضا، یکی از سریع‌ترین و باصرفه‌ترین راه‌های تهیه نقشه، استفاده از داده‌های سنجش‌ازدور می‌باشد. با توجه به اینکه داده‌های اخذشده توسط این سنسورها، حاوی سطح بالایی از جزئیات هستند و تفسیر دستی بسیار وقت‌گیر و پرهزینه است و در صورت بروز شرایط بحرانی نظیر زلزله، برای سازمان‌دهی اقدامات فوری، نقشه‌برداری سریع ضروری است، به‌منظور کاهش زمان و هزینه، اتوماسیون موردنیاز خواهد بود. تکنیک‌های یادگیری ماشین ثابت کرده‌اند که برای سیستم‌هایی که با حجم بالای داده‌ها سروکار دارند مناسب بوده و روش‌های انتخاب ویژگی به‌عنوان راهی برای کاهش زمان محاسبات، بهبود عملکرد پیش‌بینی و درک بهتر داده‌ها در کاربردهای یادگیری ماشین می‌توانند مثمر ثمر واقع شوند. بر این اساس در این مقاله مروری جامع بر روش‌های یادگیری ماشین و انتخاب ویژگی با هدف ارائه یک دیدگاه کلی بر قابلیت‌ها و موارد کاربرد این روش‌ها ارائه شده است. بدین منظور، با توجه به مزایا و محدودیت‌های روش‌های موردبررسی، درنهایت نتایج به‌صورت مقایسه تحلیلی و جدول مدون ارائه گردید. درنتیجه برای یک کاربرد، الگوریتم یادگیری ماشین و انتخاب ویژگی می‌تواند بر اساس مواردی چون دقت طبقه‌بندی و زمان محاسبات موردنیاز انتخاب گردد.

واژه‌های کلیدی : روش‌های یادگیری ماشین، روش‌های انتخاب ویژگی، طبقه‌بندی، سنجش‌ازدور

نویسنده مکاتبه کننده : مینا حمیدی
آدرس پستی : تهران، خیابان ولیعصر، تقاطع میرداماد، روبروی ساختمان اسکان، دانشکده مهندسی نقشه برداری دانشگاه صنعتی خواجه نصیر الدین طوسی
تلفن : ۰۹۱۹۶۱۰۸۲۴۰
آدرس پست الکترونیک : Hamidi.mina94@Gmail.com

۱- مقدمه

در سال‌های اخیر، با توسعه‌ی سریع تکنیک‌های تصویربرداری فضایی، سنجنده‌های سنجش‌ازدوری قادر به فراهم کردن داده‌های مشاهدات زمین باقدرت تفکیک بالا در حوزه‌ی طیفی و مکانی هستند. این نوع جدید تصاویر باقدرت تفکیک بالای سنجش‌ازدوری حاوی اطلاعات دقیق زمین در هر دو حوزه‌ی طیفی و مکانی است. افزایش قدرت تفکیک مکانی به بیش از صد برابر و افزایش تعداد نمونه‌برداری طیفی به چند صد باند، پردازش این حجم عظیم داده‌ها را با چالش روبه‌رو کرده است. از سوی دیگر، همگام با توسعه مفاهیم و تکنولوژی‌های نوین در سنجش‌ازدور، روش‌های یادگیری ماشین نیز به‌صورت موازی با آن‌ها توسعه پیداکرده و قابلیت استفاده گسترده به‌منظور پردازش و طبقه‌بندی داده‌های سنجش‌ازدور را دارا می‌باشند. داشتن درک روشنی از الگوریتم‌ها و روش‌های پایه در کاربردهای یادگیری ماشین، قبل از تأمل عمیق بر روی توسعه‌ی سیستم‌ها و حل مشکلات با استفاده از این تکنیک‌ها، اجتناب‌ناپذیر است. به همین دلیل، هدف از این بررسی، ارائه‌ی یک بینش نظری به تکنیک‌های یادگیری ماشین و درک کارایی این تکنیک‌ها است تا بتوان با استفاده از این ابزار به‌استخراج اطلاعات بهینه از تصاویر سنجش‌ازدوری پرداخت.

با استفاده از یک مجموعه ویژگی، هر الگوریتم یادگیری ماشین می‌تواند طبقه‌بندی انجام دهد [۱]. در تصاویر با ابعاد بالا معمولاً بسیاری از ویژگی‌ها برای عمل یادگیری، نامربوط، اضافی و نویزی هستند که عملکرد الگوریتم یادگیری را کاهش می‌دهند و پردازش چنین داده‌هایی نیاز به زمان و منابع محاسباتی حجیمی دارد [۲, ۳]. بر این اساس، انتخاب ویژگی نقش مهمی در بهبود عملکرد الگوریتم‌های یادگیری ماشین ازنظر کاهش زمان برای ساخت مدل یادگیری و افزایش دقت در فرایند یادگیری ایفا می‌کند. به همین دلیل محققان توجه بیشتری بر روی انتخاب ویژگی برای بهبود عملکرد الگوریتم‌های یادگیری ماشین داشته‌اند. شناسایی الگوریتم انتخاب ویژگی مناسب به‌منظور یادگیری ماشین با مجموعه داده‌ها با ابعاد بالا بسیار ضروری است. ازاین‌رو، مطالعه روش‌های مختلف انتخاب ویژگی برای جامعه محققین به‌ویژه برای توسعه روش‌های مناسب انتخاب ویژگی به‌منظور افزایش عملکرد وظایف یادگیری ماشین بر روی‌داده‌ها با ابعاد بالا، نیاز است [۴, ۵].

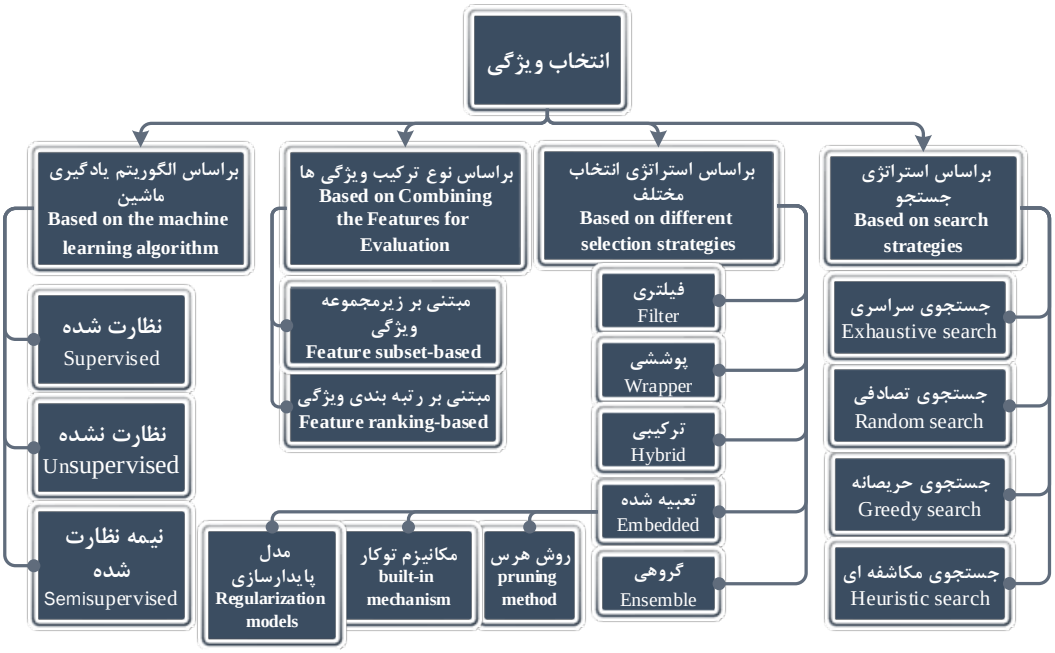
روش‌های متفاوتی به‌منظور انتخاب ویژگی در مطالعات مختلف پیشنهاد شده است [۱, ۵-۱۱]. هوانگ از الگوریتم ژنتیک به‌منظور انتخاب ویژگی و بهینه‌سازی پارامترها برای طبقه‌بندی ماشین‌های بردار پشتیبان استفاده نمود [۱۲]. بااین‌حال، انتخاب ویژگی آماری تولیدشده به‌وسیله الگوریتم ژنتیک ممکن است در مواجهه با مجموعه داده‌های سنجش‌ازدوری با ابعاد بالا، ناکارآمد و زمان‌بر باشد [۱۳]. اخیراً انتخاب ویژگی در رابطه با یک الگوریتم یادگیری ماشین بکار رفته است که الگوریتم یادگیری بخشی از فرایند اعتبارسنجی ویژگی را ایفا می‌کند. در این روش ویژگی‌های بهینه بر مبنای افزایش دقت طبقه‌بندی، انتخاب می‌گردند [۷]. برخی مطالعات که تمرکزشان بر روی استراتژی طراحی و انتخاب ویژگی‌ها به‌منظور طبقه‌بندی با استفاده از روش‌های یادگیری ماشین می‌باشد، تأثیر توجه به طراحی و انتخاب ویژگی‌ها را در عملکرد کارآمد روش‌های یادگیری در زمینه‌های مختلف مطالعاتی ازجمله طبقه‌بندی تصاویر با قدرت تفکیک مکانی بالای سنجش‌ازدوری و غیره، نشان داده‌اند [۱۴, ۱۵]. در زمینه‌ی ارزیابی عملکرد هم می‌توان گفت که اگرچه الگوریتم‌های زیادی توسعه‌یافته‌اند و با اعمال در طبقه‌بندی تصاویر ماهواره‌ای، بهبود دقت را گزارش کرده‌اند [۱۶]، اما پیدا کردن یک مقایسه‌ی سیستماتیک بر عملکرد الگوریتم‌های جدید، دشوار است. این امر به‌ویژه برای الگوریتم‌های یادگیری ماشین که بسیاری از آن‌ها کمتر از ۱۰ سال است که درزمینه‌ی سنجش‌ازدور معرفی‌شده‌اند یک واقعیت است. مقایسه‌ی عملکرد طبقه‌بندی‌کننده‌ها تنها به مقایسه‌ی یک الگوریتم جدید با یک طبقه‌بندی کننده مرسوم مانند طبقه‌بندی حداکثر احتمال یا مقایسه بین دو تا سه الگوریتم جدید محدود شده است [۱۷-۲۲].

به‌طورکلی، نتایج نهایی طبقه‌بندی تصویر به عواملی چون طرح طبقه‌بندی، داده‌های تصویری موجود، انتخاب نمونه‌های آموزشی، انتخاب و استخراج ویژگی‌ها و الگوریتم طبقه‌بندی وابسته است [۲۳]. با توجه به این عوامل تأثیرگذار در عملکرد طبقه‌بندی تصاویر سنجش‌ازدور، لازم است که هرکدام از این عوامل به‌طور مناسب تعیین گردند. در این مقاله، ابتدا به معرفی انتخاب ویژگی و انواع روش‌های آن پرداخته شده است؛ سپس با توجه به برخی از دسته‌بندی‌های ارائه‌شده برای تکنیک‌های یادگیری ماشین، به معرفی برخی از روش‌ها پرداخته شده است و درنهایت با توجه به نقاط قوت و ضعف روش‌های انتخاب ویژگی و یادگیری ماشین، عملکرد آن‌ها در طبقه‌بندی تصاویر سنجش‌ازدور مورد مقایسه قرار گرفته است.

۲- انتخاب ویژگی

انتخاب ویژگی به فرآیند انتخاب ویژگی‌های مناسب از مجموعه ویژگی ورودی با حذف ویژگی‌های نامناسب و یا زائد اطلاق می‌گردد [۶, ۲۴]. هدف اصلی یک روش انتخاب ویژگی، کاهش هزینه محاسباتی ویژگی و افزایش کارایی طبقه‌بندی است. در ابعاد بالاتر، شناسایی ویژگی‌هایی که ممکن است طبقه‌بندی کننده را در طول طبقه‌بندی گمراه کنند کار آسانی نیست. به همین دلیل انتخاب ویژگی‌های بهینه انگیزه بسیاری از محققان برای کار در این زمینه می‌باشد [۷]. از سویی دیگر، برای حذف یک ویژگی نامناسب، یک معیار انتخاب ویژگی موردنیاز است که ارتباط بین هر ویژگی و کلاس خروجی را اندازه‌گیری کند. از دیدگاه یادگیری ماشین اگر یک سیستم از ویژگی‌های نامناسب استفاده کند، برای داده جدید منجر به تعمیم ضعیف خواهد شد. هنگامی‌که یک معیار انتخاب ویژگی برگزیده شد، روشی باید توسعه داده شود که زیرمجموعه ویژگی‌های مناسب را بیابد [۱].

روش‌های مختلف انتخاب ویژگی با توجه به در دسترس بودن اطلاعات برچسب کلاس‌ها می‌توانند به الگوریتم‌های نظارت‌شده، نظارت‌نشده و نیمه نظارت‌شده دسته‌بندی گردد [۶]. بر اساس استراتژی‌های جستجوی مختلف الگوریتم‌های انتخاب ویژگی به روش‌های جستجوی سراسری، تصادفی، حریصانه و مکاشفه‌ای می‌توانند دسته‌بندی گردند [۵, ۷, ۹, ۱۱, ۲۵]. همچنین، بر اساس اینکه روش انتخاب ویژگی در تعامل با الگوریتم یادگیری باشد یا از خصوصیات کلی مجموعه داده‌ها برای حذف ویژگی‌های نامناسب صرف‌نظر از الگوریتم یادگیری استفاده کند، روش‌های انتخاب ویژگی می‌توانند به پنج گروه روش‌های فیلتری، پوششی، تعبیه‌شده، ترکیبی و گروهی دسته‌بندی گردند [۱۱]. قابل‌ذکر است که روش تعبیه‌شده خود به ۳ گروه روش هرس، مکانیزم توکار و مدل پایدارسازی می‌تواند دسته‌بندی گردد [۴]. علاوه بر این، الگوریتم‌های انتخاب ویژگی می‌توانند به‌صورت الگوریتم‌های انتخاب زیرمجموعه که یک زیرمجموعه از ویژگی‌ها را برمی‌گردانند یا الگوریتم‌های وزن ویژگی‌ها که وزن مربوط به هر ویژگی را برمی‌گردانند، دسته‌بندی شوند [۶]. با توجه به دسته‌بندی‌های فوق‌الذکر، یک دسته‌بندی کلی به‌صورت شکل (۱) ارائه شده است.



شکل ۱: دسته‌بندی روش‌های انتخاب ویژگی [۵, ۶, ۱۱]

۲-۱- انتخاب ویژگی بر اساس استراتژی جستجو

اگر تعداد کل ویژگی‌ها برابر n باشد، تعداد کل زیرمجموعه‌های ممکن 2^n حالت می‌شود. این تعداد برای n های متوسط هم خیلی زیاد است. بر اساس نحوه‌ی جستجو در میان این تعداد زیرمجموعه، روش‌های مختلف انتخاب ویژگی را می‌توان به چهار دسته جستجوی سراسری، جستجوی تصادفی، جستجوی حریصانه و جستجوی مکاشفه‌ای گروه‌بندی کرد [۵, ۷, ۹, ۱۱, ۲۵]. روش جستجوی سراسری تمام فضای جواب (ترکیبات ممکن) را برای یافتن جواب بهینه جستجو می‌کند؛ بنابراین این استراتژی جستجو ازنظر محاسباتی بسیار پرهزینه است و یک رویکرد عملی نیست. ازاین‌رو، استراتژی‌های جستجوی دیگری که ازنظر محاسباتی کارآمد باشند ارائه شدند. در جستجوی تصادفی، محدوده کمتری از فضای کل حالات ممکن جستجو می‌گردد که اندازه این محدوده به حداکثر تعداد تکرار الگوریتم بستگی دارد. در این روش‌ها پیدا شدن جواب بهینه به‌اندازه منابع موجود و زمان اجرای الگوریتم بستگی دارد. در جستجوی مکاشفه‌ای ازجمله الگوریتم ژنتیک، بهینه‌سازی ازدحام ذرات و غیره، تابع اکتشافی دانش قبلی را برای هدایت جستجو به‌منظور تولید زیرمجموعه ویژگی‌ها فراهم کرده و این زیرمجموعه‌ها توسط الگوریتم یادگیری ماشین ارزیابی می‌گردند. این استراتژی جستجو به‌منظور کاهش جستجو بدون از دست دادن شانس پیدا کردن جواب بهینه، طراحی شده است. روش‌های جستجوی حریصانه زمان اجرای کمتری نسبت به دیگر روش‌ها دارند. روش‌های SBS و SFS دو روش متداول از روش‌های با جستجوی حریصانه هستند. روش SBS با مجموعه تمام ویژگی‌ها شروع‌شده و به‌تدریج ویژگی‌های حاوی اطلاعات کمتر را حذف می‌کند. این روند زمانی که عملکرد الگوریتم یادگیری کمتر از یک حد آستانه گردد متوقف می‌شود. در مقابل SFS با یک مجموعه ویژگی تهی شروع‌شده و مکرراً ویژگی‌ها را اضافه کرده تا زمانی که دیگر دقت طبقه‌بندی افزایش پیدا نکند. در SFS یک ویژگی که اضافه شد دیگر حذف نمی‌گردد [۵].

۲-۲- انتخاب ویژگی بر اساس استراتژی انتخاب مختلف

روش‌های انتخاب ویژگی بر اساس استراتژی‌های انتخاب مختلف می‌توانند به پنج گروه روش‌های فیلتری، پوششی، تعبیه‌شده، ترکیبی و گروهی دسته‌بندی گردند [۸, ۱۱]. در روش‌های فیلتر، ویژگی‌ها بر اساس معیار آماری معین رتبه‌بندی می‌گردند و ویژگی‌ها با مقدار رتبه بالا انتخاب می‌شوند [۵, ۷, ۲۶]. ازاین‌رو، وابسته به الگوریتم‌های یادگیری

نیستند. از مزایای این رویکرد محاسبات سریع، جلوگیری از بیش برازش و کارکرد خوب در مجموعه داده‌های خاص می‌باشد. در این روش یافتن الگوریتم یادگیری مناسب ممکن است سخت باشد. همچنین، هیچ روش ایده آلی برای انتخاب ابعاد فضای ویژگی وجود ندارد. روش‌های انتخاب ویژگی پوششی از یک الگوریتم یادگیری برای ارزیابی کیفیت زیرمجموعه ویژگی‌ها در فضای جستجو استفاده می‌کنند. روش‌های پوششی باید یک الگوریتم جستجو برای جستجوی یک زیرمجموعه ویژگی بهینه بکار برند. این مدل با توجه به استفاده مکرر از الگوریتم یادگیری در فرایند جستجو، به زمان محاسباتی بالا، به‌ویژه برای مجموعه داده‌های با ابعاد بالا نیاز دارد [۱, ۹, ۲۵]. ازاین‌رو برای کاربردهای عملی با مجموعه داده‌های بزرگ روش فیلتر محبوب‌تر است. بااین‌حال، مدل پوششی به‌طور تجربی ثابت کرده است که ازنظر دقت طبقه‌بندی برتر از مدل فیلتر است [۲۷].

رویکردهای ترکیبی دارای مزیت هر دو روش فیلتر و پوششی هستند. ایده روش‌های ترکیبی این است که ویژگی‌ها را به روش فیلتر رتبه دهی کرده و جستجو را با الگوریتم‌های بهینه‌سازی مورداستفاده درروش پوششی هدایت کند [۱۰, ۲۸]. این رویکرد نسبت به روش پوششی، حساسیت کمتری به بیش برازش دارد اما مشابه روش‌های پوششی و تعبیه‌شده وابسته به یک طبقه‌بندی کننده است [۱۱]. در [۲۹] یک روش ترکیبی فیلتری-پوششی دنبال شده است که بیانگر نتایج دقیق‌تر با زیرمجموعه ویژگی کوچک‌تر در مقایسه با روش‌های جداگانه فیلتری و پوششی می‌باشد. در روش تعبیه‌شده، رویکرد اصلی این است که انتخاب ویژگی را به‌عنوان بخشی از فرایند آموزش قرار دهد [۱]. ازآنجایی‌که نگرانی اصلی این روش تعامل بین انتخاب ویژگی و طبقه‌بندی است عملکرد آن مشابه روش پوششی اما با سرعت پایین‌تر است [۹, ۱۰]. انتخاب ویژگی گروهی، یک روش نسبتاً جدید مورداستفاده برای به دست آوردن یک زیرمجموعه ویژگی پایدار است. الگوریتم انتخاب تک ویژگی بر روی زیرمجموعه‌های مختلف از نمونه‌ها اجراشده و نتایج برای به دست آوردن مجموعه ویژگی نهایی ادغام می‌شوند [۳۰, ۳۱]. این رویکرد برای داده با ابعاد بالا انعطاف‌پذیر و قوی است. هدف آن غلبه بر مشکل بی‌ثباتی سایر روش‌های انتخاب ویژگی است [۱۱].

۲-۳- انتخاب ویژگی بر اساس نوع ترکیب ویژگی‌ها

روش‌های انتخاب ویژگی مختلف را بر اساس چگونگی ترکیب آن‌ها برای ارزیابی به‌منظور انتخاب ویژگی‌های مهم از یک مجموعه داده به دو دسته روش‌های مبتنی بر زیرمجموعه ویژگی و روش‌های مبتنی بر رتبه‌بندی ویژگی‌ها می‌توان گروه‌بندی کرد [۴]. در روش مبتنی بر زیرمجموعه ویژگی، ویژگی‌ها به‌صورت ترکیبات ممکن از زیرمجموعه ویژگی‌ها با استفاده از یک استراتژی جستجو ترکیب می‌شوند. سپس، زیرمجموعه ویژگی با استفاده از روش‌های آماری یا الگوریتم‌های یادگیری با در نظر گرفتن اهمیت هر زیرمجموعه ویژگی ارزیابی‌شده و بهترین زیرمجموعه به‌عنوان زیرمجموعه ویژگی مناسب برای مجموعه داده انتخاب می‌گردد. در رویکرد مبتنی بر رتبه‌بندی ویژگی‌ها، هر یک از ویژگی‌های یک مجموعه داده بر اساس اندازه‌گیری‌های آماری یا اطلاعات تئوریک وزندهی شده و ویژگی‌ها بر اساس وزنشان رتبه‌بندی می‌گردند. سپس ویژگی‌های با رتبه بالاتر به‌عنوان ویژگی‌های مناسب با استفاده از یک حد آستانه از پیش تعریف‌شده برای تعیین تعداد ویژگی‌های انتخابی از یک مجموعه داده، انتخاب می‌گردند.

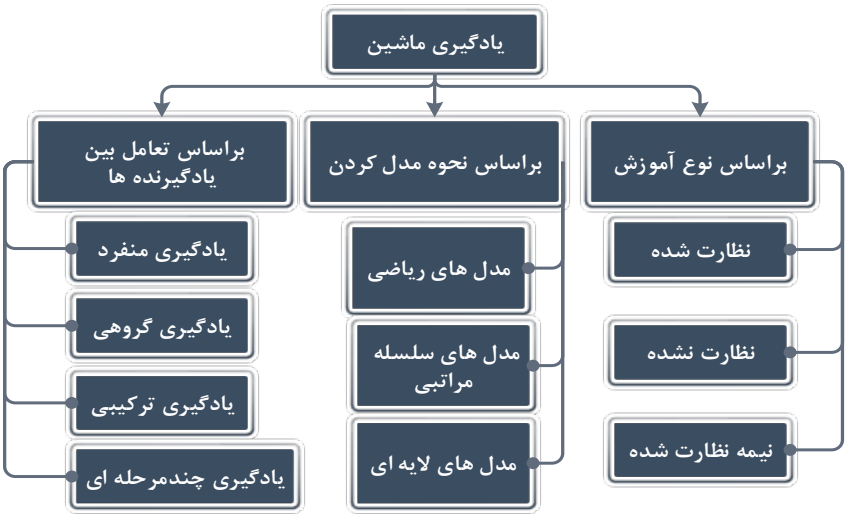
۲-۴- انتخاب ویژگی بر اساس الگوریتم‌های یادگیری ماشین

روش‌های انتخاب ویژگی بر اساس الگوریتم یادگیری ماشین، با توجه به در دسترس بودن اطلاعات برچسب کلاس‌ها می‌تواند به الگوریتم‌های نظارت‌شده، نظارت‌نشده و نیمه نظارت‌شده دسته‌بندی گردد. در الگوریتم‌های انتخاب ویژگی نظارت‌شده، ویژگی‌های مهم با برآورد ارتباط آن‌ها با برچسب کلاس‌ها یا عملکردشان در پیش‌بینی تعیین می‌شوند [۳۲, ۳۳]. الگوریتم‌های انتخاب ویژگی نظارت‌نشده، ویژگی‌ها را با به‌کارگیری واریانس یا توزیع داده‌ها انتخاب می‌کنند [۳۴, ۳۵]. تکنیک‌های خوشه‌بندی نمونه‌ای اولیه از یادگیری نظارت‌نشده هستند که سعی در کشف گروه‌بندی‌های طبیعی در مجموعه‌ای از اشیاء بدون آگاهی از برچسب کلاس‌ها دارند. انتخاب ویژگی

با استفاده از یادگیری نظارت‌نشده می‌تواند توصیف بهتر و قابل‌اطمینانی از داده‌ها نسبت به استفاده تنها از یادگیری نظارت‌نشده فراهم کند [۱]. در یک الگوریتم انتخاب ویژگی نیمه نظارت‌شده، تعداد کمی از داده‌های برچسب دار به‌عنوان اطلاعات اضافی برای بهبود عملکرد الگوریتم انتخاب ویژگی نظارت‌نشده بکار می‌رود [۳۶, ۳۷].

۳- روش‌های یادگیری ماشین

الگوریتم‌ها و مدل‌های بسیاری در تحقیقات یادگیری ماشین به‌منظور مسائل طبقه‌بندی مناسب در دسترس است. بااین‌حال، تعداد اندکی از آن‌ها برای طبقه‌بندی داده‌های با حجم بالا مناسب هستند. الگوریتم‌های یادگیری ماشین برحسب نوع آموزش به ۳ دسته‌ی یادگیری نظارت‌شده، نظارت‌نشده و نیمه نظارت‌شده تقسیم‌بندی می‌شوند. از طرفی مدل‌های یادگیری نظارت‌شده را می‌توان به ۳ دسته‌ی، مدل‌های ریاضی[1]، مدل‌های سلسله مراتبی[2] و مدل‌های لایه‌ای[3] دسته‌بندی کرد [۳۸]. از دیدگاه دیگر می‌توان به ۶ دسته‌ی، روش‌های نظارت‌شده، نظارت‌نشده، نیمه نظارت‌شده، هیبرید[4]، گروهی[5] و چندمرحله‌ای[6] تقسیم کرد [۳۹]. همچنین، می‌توان روش‌های یادگیری نظارت‌شده را به ۳ دسته‌ی روش‌های یادگیری منفرد و روش‌های یادگیری گروهی و روش‌های هیبرید گروه‌بندی کرد [۴۰]. با توجه به دسته‌بندی‌های فوق، در این مقاله به‌طورکلی دسته‌بندی به‌صورت شکل ۲ ارائه می‌شود.



شکل ۲: دسته‌بندی روش‌های یادگیری ماشین [۳۸-۴۰]

۳-۱- یادگیری ماشین بر اساس نوع آموزش

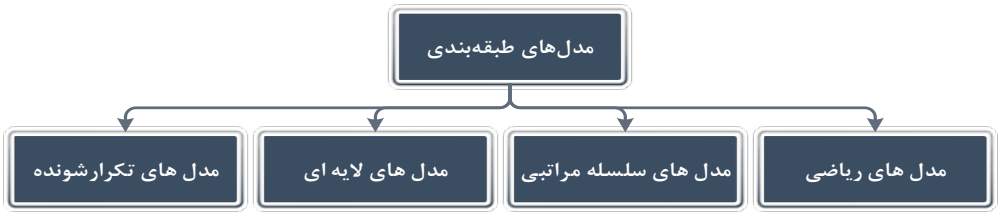
روش‌های یادگیری ماشین برحسب نوع آموزش به ۳ دسته‌ی یادگیری نظارت‌شده، یادگیری نظارت‌نشده و یادگیری نیمه نظارت‌شده تقسیم‌بندی می‌شوند؛ قسمت اصلی الگوریتم‌های یادگیری ماشین درزمینه‌ی سنجش‌ازدور، به طبقه‌بندی نظارت‌شده که اغلب در طبقه‌بندی‌های پوشش اراضی به کار می‌رود تمرکز دارد و طرح کلی یک طبقه‌بندی ساده را فراهم می‌کند تا بتواند به‌وسیله یادگیری از نمونه‌های آموزشی، اشیا را از صحنه‌های پیچیده به‌طور درست و سریع استخراج نماید. روش‌های طبقه‌بندی که فقط متکی بر اطلاعات تصویر بوده و کاربر در کل فرآیند آن دخالتی ندارد را روش‌های طبقه‌بندی نظارت‌نشده می‌گویند. طبقه‌بندی‌های نظارت‌نشده با استفاده از روش‌های خوشه‌بندی انجام می‌پذیرد؛ که در مطالعات گذشته، برای پردازش تصاویر چند طیفی و فرا طیفی با

[1] Mathematical Models
[2] Hierarchical Models
[3] Layered Models
[4] Hybrid
[5] Ensemble
[6] Multistage

کاهش ابعاد طیفی بکار رفته‌اند [۳۹]. روش‌های یادگیری نیمه نظارت‌شده از اطلاعات نمونه‌های بدون برچسب به همراه نمونه‌های برچسب دار در تعریف قوانین طبقه‌بندی بهره می‌برند [۴۱]. این ایده اصلی روش‌های یادگیری نیمه نظارت‌شده است [۴۲]. این رویکرد می‌تواند به پنج دسته‌ی خود آموزش[1]، آموزش تعاونی[2]، مدل احتمال مولد[3]، ماشین بردار پشتیبان نیمه نظارت‌شده[4] و روش‌های گراف مبنا[5] گروه‌بندی شود [۴۳].

۳-۲- یادگیری ماشین بر اساس نحوه مدل کردن

در یادگیری نظارت‌شده، با استفاده از داده‌های آموزشی یک تابع ایجاد می‌شود که در آن هر یک از داده‌های آموزشی شامل یک جفت‌بردار ورودی و خروجی (برچسب کلاس) هستند. وظیفه یادگیری این است که به تعیین رابطه‌ی تقریبی بین نمونه‌های ورودی-خروجی برای ایجاد یک طبقه‌بندی کننده (مدل) بپردازد. هنگامی‌که مدل ایجاد شد، طبقه‌بندی کننده می‌تواند نمونه‌های بدون برچسب را داخل کلاس‌های آموزش‌دیده برچسب دهی کند [۴۰]. مدل‌های طبقه‌بندی را می‌توان به ۳ دسته‌ی مدل‌های ریاضی، مدل‌های سلسله مراتبی و مدل‌های لایه‌ای تقسیم‌بندی کرد. علاوه بر این، با توجه به ماهیت تکرارشونده برخی از مدل‌های طبقه‌بندی ازجمله بگینگ، بوستینگ، آدابوست و غیره می‌توان یک دسته مدل تکرارشونده هم در نظر گرفت (شکل (۳)). در مدل‌های ریاضی، تمام نقاط داده در دامنه‌ی داده‌ها باهم برای طبقه‌بندی بکار گرفته می‌شوند. دو مدل معروف از این دسته، ماشین‌های بردار پشتیبان و k-نزدیک‌ترین همسایگی می‌باشد. مدل‌های سلسله مراتبی به طبقه‌بندی زوج گروه‌های جداشده از نقاط توسط اتصال آن‌ها به کلاس والدینشان، با استفاده از ساختارهای درخت مانند، کمک می‌کنند. این مدل‌ها همانند مدل‌های ریاضی، تمام نقاط داده در دامنه‌ی داده‌ها را باهم برای طبقه‌بندی بکار می‌گیرند. ازجمله مدل‌های این دسته می‌توان درختان تصمیم‌گیری و جنگل‌های تصادفی را نام برد. مدل‌های لایه‌ای به‌طور قابل‌توجهی از دو مدل قبلی متفاوت هستند. به این دلیل که تمام نقاط داده در دامنه داده‌ها را باهم بکار نمی‌گیرند، بلکه، در هرلحظه یک نقطه داده (نورون) توسط دنباله‌ای از لایه‌ها برای آموزش بکار گرفته می‌شود [۳۸]. شبکه‌های عصبی مصنوعی از این دسته مدل‌ها هستند.



شکل ۳: دسته‌بندی مدل‌های طبقه‌بندی [۳۸]

۳-۳- یادگیری ماشین بر اساس تعامل بین یادگیرنده‌ها

بسیاری از روش‌های طبقه‌بندی به‌صورت منفرد و تنها آموزش می‌بینند. ازجمله این روش‌ها می‌توان روش‌های k نزدیک‌ترین همسایگی، ماشین‌های بردار پشتیبان، شبکه‌های عصبی مصنوعی و درختان تصمیم‌گیری را نام برد. در مقابل، روش‌های طبقه‌بندی گروهی، مجموعه‌ی بزرگی از طبقه‌بندی کننده‌ها را از داده‌های آموزشی با استفاده از یک الگوریتم یادگیری پایه (مثل درخت طبقه‌بندی) ایجاد می‌کنند. سپس، مجموعه طبقه‌بندی کننده‌های متعدد با استفاده از یک رویکرد مبتنی بر قواعد (مثل رأی اکثریت) که پیکسل را به

[1] self-training
[2] cooperative training
[3] Generative probability model
[4] Semi-Supervised Support Vector Machine (S3VM)
[5] Graph based methods

کلاس با بیشترین تعداد رأی از طبقه‌بندی کننده‌های مجزا اختصاص می‌دهد، ترکیب می‌شوند [۲۰]. هدف اصلی متدلوژی گروهی، تلاش به بهبود عملکرد طبقه‌بندی کننده‌های منفرد، با استفاده از چندین طبقه‌بندی‌کننده و ترکیب آن‌ها برای رسیدن به یک طبقه‌بندی کننده جدید می‌باشد. این ایده از رفتار طبیعی انسان پیروی می‌کند که قبل از هر تصمیم‌گیری مهمی تمایل به دنبال کردن نظرات و عقاید مختلف است [۴۴]. از میان تمام روش‌های مختلف موجود، روش‌های جنگل‌های تصادفی، بگینگ، بوستینگ و آدابوست متداول‌ترین روش‌های یادگیری گروهی هستند.

روش‌های ترکیبی از بیش از یک یادگیرنده مانند ترکیب دو روش یادگیرنده نظارت‌شده و نظارت‌نشده یا ترکیب روش‌های نظارت‌شده پارامتریک و غیر پارامتریک برای طبقه‌بندی استفاده می‌کنند. تلفیق تصمیم‌گیری و طبقه‌بندی کننده نظارت‌شده فازی FCM را می‌توان دو نمونه خوب از روش‌های طبقه‌بندی ترکیبی معرفی کرد [۴۵, ۴۶].

روش‌های چندمرحله‌ای نه‌تنها به یک سیستم طبقه‌بندی خاص اشاره دارند بلکه مکانیزم پیش‌پردازش مربوطه را برای تحقق عملکرد بهتر سیستم ازنظر دقت طبقه‌بندی، استحکام و ثبات، پیچیدگی محاسباتی و غیره توصیف می‌کنند [۳۹]. شبکه‌های عصبی با طرح کاهش داده‌ها و جداسازی ویژگی‌ها از روش‌های مؤثر در روش‌های چندمرحله‌ای است [۴۷, ۴۸].

۴- بحث و بررسی

در این بخش ابتدا به‌طور خلاصه و مدون در جدول (۱)، به جمع‌بندی و بیان ویژگی‌های هر یک از روش‌های عنوان‌شده پرداخته می‌شود و در ادامه برخی از روش‌ها به‌صورت دودویی با یکدیگر مقایسه و مورد تحلیل قرار خواهند گرفت.

جدول ۱: مزایا و محدودیت‌های روش‌های یادگیری ماشین

روش‌های یادگیری ماشین	مزایا	محدودیت‌ها
ماشین‌های بردار پشتیبان (SVM)	✓ بدون نیاز به دانستن توزیع آماری داده‌ها (غیر پارامتریک) ✓ مرزهای کلاس غیرخطی را مدل می‌کند ✓ پیچیدگی محاسباتی کمتر نسبت به ANN ✓ در مقایسه با ANN، اخذ ویژگی‌های ضروری داده‌ها ✓ بیش برازش اتفاق نمی‌افتد ✓ کنترل پیچیدگی قانون تصمیم‌گیری و فرکانس خطا، آسان است ✓ قابلیت تعمیم خوب حتی با نمونه‌های آموزشی کم ✓ فراهم کردن دقت بالا در طبقه‌بندی ✓ بسیار مناسب برای کاربردهای داده‌های بزرگ	• مرحله آموزش در مقایسه با روش بیز و طبقه‌بندی درختی کندتر است • تعیین پارامترهای بهینه زمانی که داده‌ها به‌صورت غیرخطی تفکیک‌پذیر باشند، سخت است • نیاز به چندین پارامتر کلیدی برای رسیدن به بهترین نتیجه طبقه‌بندی • حساسیت به تنظیم پارامترها • درک ساختار الگوریتم سخت است • حساس به نویز • تبدیل به فضای با ابعاد بالاتر منجر به افزایش زمان محاسبات می‌شود • پیچیده‌تر برای طبقه‌بندی • تفسیر دشوار برای حل مدل پارامتری
K نزدیک‌ترین همسایگی (KNN)	✓ ساده، کارآمد، پیاده‌سازی آسان و غیر پارامتریک ✓ نرخ خطای پایین در فرایند آموزش فراهم می‌کند	• روند کند • زمان طبقه‌بندی طولانی • یافتن مقدار بهینه دشوار است

شبکه‌های عصبی مصنوعی (ANN)	✓ بدون نیاز به دانستن توزیع آماری داده‌ها (غیر پارامتریک) ✓ می‌توان برای طبقه‌بندی یا رگرسیون استفاده کرد ✓ توانایی ارائه توابع بولین (AND, OR, NOT) ✓ قادر به انجام طبقه‌بندی چندکلاسی ✓ تولید مدل برای طبقه‌بندی داده‌ها ✓ قابلیت تلفیق با روش‌های دیگر مثل الگوریتم ژنتیک ✓ ساختار موازی برای محاسبات آنی ✓ قابل‌آموزش با نمونه‌های محدود ✓ در حوزه‌های پیچیده، نتیجه‌های خوب فراهم می‌کند ✓ فرایند تست سریع است	• درک ساختار الگوریتم سخت است (جعبه سیاه هستند) • نیاز به تعریف پارامترهای اولیه توسط کاربر • توصیفات خیلی زیاد می‌تواند منجر به بیش برازش شود • ساختار بهینه شبکه را فقط باتجربه می‌توان تعیین کرد • حساسیت بالا به طراحی پارامترها • نیاز به داده‌های آموزشی بدون نویز • روند کند در آموزش

درخت تصمیم C4.5	✓ مناسب برای مشکلات دنیای واقعی ✓ اداره کردن مقادیر ازدست‌رفته ✓ تقسیم بسیار دقیق داده‌ها	✓ دارای قابلیت انتخاب ویژگی	• قوانین بسیار ساده • نیاز به نمونه‌های آموزشی بالا • نامطلوب در کاربرد عملی	• هرس درختان به علت وجود نویز و داده‌های پرت در نمونه‌های آموزشی

جنگل‌های تصادفی	✓ غیر حساس به بیش برازش ✓ مناسب برای طبقه‌بندی داده با ابعاد بالا ✓ قابلیت تعمیم دهی خوب ✓ در مقایسه با SVM تنها به تنظیم دو پارامتر نیاز است ✓ ماهیت تکراری منجر به نتایج بهتر ✓ آموزش سریع ✓ پیاده‌سازی بسیار ساده ✓ به هیچ اعتبارسنجی متقابل و یا مجموعه آزمایشی جداگانه، نیاز ندارد ✓ غیر حساس به تنوع داده‌های آموزشی ✓ حساسیت کم به داده‌های نویزی ✓ تعیین اهمیت ویژگی‌ها ✓ محاسبه احتمال تعلق کلاس‌ها به‌صورت جداگانه	• تحت تأثیر قرار گرفتن عملکرد طبقه‌بندی به علت، همبستگی بین دو درخت و استحکام یک درخت مجزا در جنگل • حساسیت به طرح نمونه‌برداری

آدابوست	✓ روش طبقه‌بندی قوی ✓ سریع، ساده و انعطاف‌پذیر ✓ تعداد دور بوستینگ در روند آموزش در دسترس است ✓ غیر حساس به بیش برازش ✓ علاوه بر واریانس، بایاس را هم کاهش می‌دهد ✓ نیازی به دانستن مرزهای خطا مبتنی بر طبقه‌بندی کننده‌های ضعیف ندارد	• بسیار پیچیده در داده‌های نویزی و پرت • در پیش‌بینی خطا قوی نیست • عدم موفقیت در اداره کردن یادگیرندگان ضعیف، اگر نرخ بیشتر از ½ باشد

بگینگ	✓ به نویز داده‌های کالیبراسیون، پایدارتر و قوی‌تر از درختان تصمیم‌گیری ✓ غیر حساس به تنوع داده‌های آموزشی ✓ به‌طور ذاتی مناسب برای محاسبات موازی ✓ قابلیت تعمیم دهی خوب ✓ کاهش چشمگیر واریانس ✓ با استفاده از یادگیرنده‌های پایه مستقل، خطای OOB به‌طور نمایی کاهش می‌یابد	• در مقایسه با جنگل تصادفی انعطاف‌پذیری کمتری دارد • به علت انتخاب تمام ویژگی‌ها، در مقایسه با جنگل تصادفی مرحله آموزش مؤثری ندارد
بوستینگ	✓ نسبت به نویز داده‌ها قوی‌تر ✓ عملکرد خوب در اندازه داده‌های آموزشی محدود ✓ پتانسیل کاهش خطاهای طبقه‌بندی، به علت ماهیت تکراری ✓ غیر حساس به تنوع داده‌های آموزشی	• به دلیل ماهیت تکرارشونده، دارای پروسه محاسباتی طولانی

با توجه به جدول (۱) و مطالعات صورت گرفته مشاهده شد که روش‌های بگینگ و بوستینگ به نویز داده‌های کالیبراسیون، پایدارتر و قوی‌تر از درختان تصمیم‌گیری هستند؛ همچنین دقت طبقه‌بندی بوستینگ کمی بالاتر از بگینگ برای طبقه‌بندی پوشش زمین می‌باشد [۴۹]. آدابوست برخلاف الگوریتم بوستینگ، نیازی به دانستن مرزهای خطا مبتنی بر طبقه‌بندی کننده‌های ضعیف ندارد؛ و همچنین نیازی به دانستن تعداد طبقه‌بندی‌کننده‌ها نیست. آدابوست علاوه بر واریانس، بایاس را هم کاهش می‌دهد و همانند ماشین‌های بردار پشتیبان، حاشیه را تقویت می‌کند [۴۴]. ماهیت جنگل‌های تصادفی اجازه می‌دهد تا بتوان ویژگی‌ها را برحسب اهمیت به‌عنوان معیاری برای انتخاب ویژگی، به‌منظور طبقه‌بندی داده چند یا ابر طیفی، زمانی که تعداد باندها زیاد است، رتبه‌بندی کند. الگوریتم‌های یادگیری گروهی می‌توانند در سناریوهای پیچیده نقشه‌برداری مفید واقع شوند، زیرا ماهیت تکرارشونده‌ی این الگوریتم‌ها، طبقه‌بندی‌های متعدد ایجاد می‌کند که رویکرد همگرا به برچسب‌دار شدن پیکسل را تضمین می‌کنند و به‌طور قابل‌توجهی دقت طبقه‌بندی را بهبود می‌بخشند، چون خطای یک طبقه‌بندی کننده بیشتر از خطای مجموعه طبقه‌بندی کننده‌های متعدد است [۲۰].

با مقایسه روش‌های منفرد مشاهده شد که برخلاف شبکه‌های عصبی، درختان تصمیم‌گیری دارای قوانین قابل تفسیر و محاسبات سریع هستند و همچنین دارای قابلیت انتخاب ویژگی می‌باشند [۵۰]. ماشین‌های بردار پشتیبان به‌طور مؤثری قادر به غلبه بر مشکلاتی مانند بیش برازش و مسائلی چون بهینه محلی که معمولاً در یادگیری شبکه عصبی رخ می‌دهد، می‌باشند. همچنین این روش، تنها به تعداد نقاط کمی برای تعریف ابر صفحه تفکیک‌کننده متکی است، بنابراین حساسیت چندانی به‌اندازه نمونه آموزشی ندارد و دقت بهتری در مقایسه با شبکه‌های عصبی مصنوعی با استفاده از اندازه نمونه آموزشی بسیار کمتر می‌تواند ایجاد کند [۵۱].

به‌طورکلی، روش‌های جنگل‌های تصادفی، شبکه‌های عصبی مصنوعی و ماشین بردار پشتیبان، نیازمند بررسی بیشتر روی مجموعه داده‌های متفاوت، به‌منظور بررسی عملکرد آن‌ها ازنظر استحکام، پیچیدگی و دقت هستند. چارچوب طبقه‌بندی گروهی، برای رسیدن به نتایج بهتر، نیاز است که بهینه شود. در بسیاری از موارد، جمع‌آوری نمونه‌های آموزشی کافی به‌منظور طبقه‌بندی به‌ویژه در مناطق ناهمگون، محدود به تجربه‌ی تحلیلگر، زمان و هزینه است؛ بنابراین، الگوریتم‌های مناسب برای تولید دقت قابل‌قبول با کمترین نمونه‌های آموزشی ممکن، نیاز است که بدین منظور از الگوریتم‌های یادگیری نیمه نظارت‌شده می‌توان بهره برد.

با توجه به مطالب بیان‌شده در بخش ۲، می‌توان مقایسه‌ای کلی بین روش‌های انتخاب ویژگی انجام داد. بر این اساس، با توجه به اینکه هر روش انتخاب ویژگی به یکسری موضوع مانند تعامل با الگوریتم یادگیری و اندازه‌گیری اهمیت ویژگی مربوط است. روش‌های پوششی، تعبیه‌شده، گروهی و ترکیبی در تعامل با الگوریتم یادگیری

هستند؛ درحالی‌که روش فیلتر از خصوصیات کلی مجموعه داده‌ها استفاده می‌کند؛ بنابراین ویژگی‌های منتخب توسط این روش ممکن است برای یک الگوریتم یادگیری خاص بهترین نباشند. روش ترکیبی سعی دارد نقطه قوت هر دو روش فیلتر و پوششی را اخذ کند. این رویکرد قابلیت مکان‌یابی راه‌حل مناسب را دارد اما رویکرد منفرد اغلب داخل یک راه‌حل نابالغ گیر می‌افتد. بااین‌حال، روش‌های ترکیبی فیلتر و پوششی ازنظر دقت رنج می‌برند زیرا ممکن است در تعاملات ویژگی‌ها یک ویژگی مناسب در انزوا، تفکیک‌پذیرتر از ویژگی نامناسب نباشد. رویکرد تعبیه‌شده، برخلاف روش پوششی حساسیت کمتری به بیش برازش دارد اما مشکل اصلی آن پیچیدگی محاسباتی در داده‌ها با ابعاد بالا می‌باشد و مشابه روش پوششی وابستگی بین ویژگی‌ها را در نظر می‌گیرد. از دیدگاه استراتژی جستجو، روش جستجوی سراسری برای مجموعه داده‌های بزرگ‌تر، بار محاسباتی سنگین دارد؛ بنابراین الگوریتم‌های ساده‌شده مانند جستجوی ترتیبی یا الگوریتم‌های تکاملی مانند الگوریتم ژنتیک یا بهینه‌سازی ازدحام ذرات که نتایج بهینه محلی حاصل می‌کنند، می‌توانند نتایج خوب و محاسبات کارآمد ارائه دهند.

۵- نتیجه‌گیری

یادگیری ماشین، روش‌های مفید زیادی را برای حل مشکلاتی که به نظر می‌رسید نیاز به راه‌حل دستی دارند، فراهم می‌کند. بااین‌حال، محققان یادگیری ماشین هنوز هم به‌طور کامل نتوانسته‌اند موفق به حل این مشکل شوند. مطالعات نشان داده است که انتخاب ویژگی نقش مهمی در بهبود عملکرد الگوریتم‌های یادگیری ماشین ایفا می‌کند. بر این اساس در این مقاله مروری جامع بر روش‌های یادگیری ماشین و انتخاب ویژگی با هدف ارائه یک دیدگاه کلی بر قابلیت‌ها و موارد کاربرد این روش‌ها ارائه گردید. از میان روش‌های موردبررسی، روش‌های یادگیری گروهی به دلیل به‌کارگیری چندین یادگیرنده در یک‌روند تکراری قادر به انتخاب ویژگی‌های بهینه در طول آموزش می‌باشند؛ که سبب بهبود عملکرد طبقه‌بندی می‌شوند. بااین‌حال این رویکردها به دلیل ماهیت تکرارشونده، بار محاسباتی سنگینی دارند. همچنین از میان روش‌های منفرد، روش ماشین بردار پشتیبان به دلیل استفاده از ویژگی‌های هندسی داده‌ها و تولید دقت‌های بالای طبقه‌بندی یک روش مناسب و متداول در سنجش‌ازدور است اما این روش از پیچیدگی محاسباتی رنج می‌برد. علاوه‌براین، الگوریتم‌های یادگیری نیمه نظارت‌شده را به دلیل دارا بودن مزیت روش‌های یادگیری نظارت‌شده و نظارت‌نشده، می‌توان به‌عنوان یک روش مؤثر برای انتخاب ویژگی و طبقه‌بندی بکار برد. در حال حاضر با توجه به مشکلات محاسباتی بسیار بالای ناشی از ابعاد بسیار بزرگ و پیچیدگی تصاویر سنجش‌ازدوری و نیز محدودیت‌های روش‌های موجود، توسعه‌ی روش‌های انتخاب ویژگی و طبقه‌بندی بسیار دقیق و سریع برای انتخاب زیرمجموعه ویژگی بهینه و رسیدن به عملکرد بهتر طبقه‌بندی یک موضوع باز است. به‌طورکلی با توجه به بررسی‌های صورت گرفته و اینکه هر روش یادگیری ماشین و انتخاب ویژگی دارای مزایا و محدودیت‌هایی است؛ روش مناسب برای یک کاربرد خاص بر اساس اهداف پژوهش و معیارهای موردنظر ازجمله سادگی، ثبات، تعداد ویژگی‌های کاهش‌یافته، دقت طبقه‌بندی، حافظه و زمان محاسبات موردنیاز می‌تواند انتخاب گردد.

مراجع

[1] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," Computers & Electrical Engineering, vol. 40, pp. 16-28, 1// 2014.

[2] H. Liu and H. Motoda, Computational methods of feature selection: CRC Press, 2007.

[3] R. O. Duda, P. E. Hart, and D. G. Stork, Pattern classification: John Wiley & Sons, 2012.

[4] D. A. A. Gnana, S. Appavu, and E. J. Leavline, "Literature Review on Feature Selection Methods for High-Dimensional Data," methods, vol. 136, 2016.

[5] I. A. Gheyas and L. S. Smith, "Feature subset selection in large dimensionality domains," Pattern recognition, vol. 43, pp. 5-13, 2010.

[6] S. Kanakam and C. Murthy, "Bridging Feature Selection and Extraction: Compound Feature Generation," IEEE Transactions on Knowledge and Data Engineering, 2016.

[7] S. Das, P. K. Singh, S. Bhowmik, R. Sarkar, and M. Nasipuri, "A Harmony Search Based Wrapper Feature Selection Method for Holistic Bangla Word Recognition," Procedia Computer Science, vol. 89, pp. 395-403, 2016.

[8] N. Spolaôr, M. C. Monard, G. Tsoumakas, and H. D. Lee, "A systematic review of multi-label feature selection and a new method based on label construction," Neurocomputing, vol. 180, pp. 3-15, 2016.

[9] F. P. Shah and V. Patel, "A review on feature selection and feature extraction for text classification," in Wireless Communications, Signal Processing and Networking (WiSPNET), International Conference on, 2016, pp. 2264-2268.

[10] A. Nouri and A. Nazari, "Improving Image steganalysis performance using a graph-based feature selection method," 2016.

[11] S. Mahajan and S. Singh, "Review On Feature Selection Approaches Using Gene Expression Data," Imperial Journal of Interdisciplinary Research, vol. 2, 2016.

[12] C.-L. Huang and C.-J. Wang, "A GA-based feature selection and parameters optimizationfor support vector machines," Expert Systems with applications, vol. 31, pp. 231-240, 2006.

[13] D. G. Stavrakoudis, G. N. Galidaki, I. Z. Gitas, and J. B. Theocharis, "A genetic fuzzy-rule-based classifier for land cover classification from hyperspectral imagery," IEEE Transactions on Geoscience and Remote Sensing, vol. 50, pp. 130-148, 2012.

[14] P. Tokarczyk, J. D. Wegner, S. Walk, and K. Schindler, "Features, color spaces, and boosting: New insights on semantic classification of remote sensing images," IEEE Transactions on Geoscience and Remote Sensing, vol. 53, pp. 280-295, 2015.

[15] B. Agarwal and N. Mittal, "Prominent feature extraction for review analysis: an empirical study," Journal of Experimental & Theoretical Artificial Intelligence, vol. 28, pp. 485-498, 2016/05/03 2016.

[16] D. Lu and Q. Weng, "A survey of image classification methods and techniques for improving classification performance," International journal of Remote sensing, vol. 28, pp. 823-870, 2007.

[17] C. Li, J. Wang, L. Wang, L. Hu, and P. Gong, "Comparison of classification algorithms and training sample sizes in urban land classification with Landsat thematic mapper imagery," Remote Sensing, vol. 6, pp. 964-983, 2014.

[18] X. Na, S. Zhang, X. Li, H. Yu, and C. Liu, "Improved land cover mapping using random forests combined with landsat thematic mapper imagery and ancillary geographic data," Photogrammetric Engineering & Remote Sensing, vol. 76, pp. 833-840, 2010.

[19] V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez, "An assessment of the effectiveness of a random forest classifier for land-cover classification," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 67, pp. 93-104, 2012.

[20] B. Ghimire, J. Rogan, V. R. Galiano, P. Panday, and N. Neeti, "An evaluation of bagging, boosting, and random forests for land-cover classification in Cape Cod, Massachusetts, USA," GIScience & Remote Sensing, vol. 49, pp. 623-643, 2012.

[21] S. E. Sesnie, B. Finegan, P. E. Gessler, S. Thessler, Z. Ramos Bendana, and A. M. Smith, "The multispectral separability of Costa Rican rainforest types with support vector machines and Random Forest decision trees," International Journal of Remote Sensing, vol. 31, pp. 2885-2909, 2010.

[22] J. C.-W. Chan and D. Paelinckx, "Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery," Remote Sensing of Environment, vol. 112, pp. 2999-3011, 6/16/ 2008.

[23] P. Gong and P. HOWARTH, "An assessment of some factors influencing multispectral land-cover classification," Photogrammetric Engineering and Remote Sensing, vol. 56, pp. 597-603, 1990.

[24] A. Abbasi, H. Chen, and A. Salem, "Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums," ACM Transactions on Information Systems (TOIS), vol. 26, p. 12, 2008.

[25] H. H. Inbarani, A. T. Azar, and G. Jothi, "Supervised hybrid feature selection based on PSO and rough sets for medical diagnosis," Computer Methods and Programs in Biomedicine, vol. 113, pp. 175-185, 1// 2014.

[26] Q. Song, J. Ni, and G. Wang, "A fast clustering-based feature subset selection algorithm for high-dimensional data," IEEE transactions on knowledge and data engineering, vol. 25, pp. 1-14, 2013.

[27] H. Liu and L. Yu, "Toward integrating feature selection algorithms for classification and clustering," IEEE Transactions on knowledge and data engineering, vol. 17, pp. 491-502, 2005.

[28] F. Temitayo, O. Stephen, and A. Abimbola, "Hybrid GA-SVM for efficient feature selection in e-mail classification," Comput Eng Intell Syst, vol. 3, pp. 17-28, 2012.

[29] R. S. Sreepada, S. Vipsita, and P. Mohapatra, "An efficient approach for microarray data classification using filter wrapper hybrid approach," in Advance Computing Conference (IACC), 2015 IEEE International, 2015, pp. 263-267.

[30] A.-C. Haury, P. Gestraud, and J.-P. Vert, "The influence of feature selection methods on accuracy, stability and interpretability of molecular signatures," PloS one, vol. 6, p. e28210, 2011.

[31] T. Abeel, T. Helleputte, Y. Van de Peer, P. Dupont, and Y. Saeys, "Robust biomarker identification for cancer diagnosis with ensemble feature selection methods," Bioinformatics, vol. 26, pp. 392-398, 2010.

[32] Z. Zhao, L. Wang, H. Liu, and J. Ye, "On similarity preserving feature selection," IEEE Transactions on Knowledge and Data Engineering, vol. 25, pp. 619-632, 2013.

[33] D. Wang, F. Nie, and H. Huang, "Global Redundancy Minimization for Feature Ranking," IEEE Transactions on Knowledge & Data Engineering, pp. 1-1.

[34] A. K. Farahat, A. Ghodsi, and M. S. Kamel, "An efficient greedy method for unsupervised feature selection," in 2011 IEEE 11th International Conference on Data Mining, 2011, pp. 161-170.

[35] M. Qian and C. Zhai, "Robust Unsupervised Feature Selection," in IJCAI, 2013.

[36] Z. Xu, I. King, M. R.-T. Lyu, and R. Jin, "Discriminative semi-supervised feature selection via manifold regularization," IEEE Transactions on Neural networks, vol. 21, pp. 1033-1047, 2010.

[37] K. Benabdeslem and M. Hindawi, "Efficient semi-supervised feature selection: constraint, relevance, and redundancy," IEEE Transactions on Knowledge and Data Engineering, vol. 26, pp. 1131-1143, 2014.

[38] S. Suthaharan, Machine learning models and algorithms for big data classification: Springer, 2016.

[39] D. Chutia, D. Bhattacharyya, K. Sarma, R. Kalita, and S. Sudhakar, "Hyperspectral Remote Sensing Classifications: A Perspective Survey," Transactions in GIS, 2015.

[40] C.-F. Tsai, Y.-F. Hsu, C.-Y. Lin, and W.-Y. Lin, "Intrusion detection by machine learning: A review," Expert Systems with Applications, vol. 36, pp. 11994-12000, 12// 2009.

[41] L. Bruzzone and C. Persello, "Recent trends in classification of remote sensing data: active and semisupervised machine learning paradigms," in Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International, 2010, pp. 3720-3723.

[42] Q. Shi, L. Zhang, and B. Du, "Semisupervised Discriminative Locally Enhanced Alignment for Hyperspectral Image Classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 51, pp. 4800-4815, 2013.

[43] F. Gao, W. Lv, Y. Zhang, J. Sun, J. Wang, and E. Yang, "A novel semisupervised support vector machine classifier based on active learning and context information," Multidimensional Systems and Signal Processing, pp. 1-20, 2016.

[44] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, "A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 42, pp. 463-484, 2012.

[45] H. Yang, Q. Du, and B. Ma, "Decision fusion on supervised and unsupervised classifiers for hyperspectral imagery," IEEE Geoscience and Remote Sensing Letters, vol. 7, pp. 875-879, 2010.

[46] J. Oldeland, W. Dorigo, L. Lieckfeld, A. Lucieer, and N. Jürgens, "Combining vegetation indices, constrained ordination and fuzzy classification for mapping semi-natural vegetation units from hyperspectral imagery," Remote Sensing of Environment, vol. 114, pp. 1155-1166, 2010.

[47] B. C. Sahoo, T. Oommen, D. Misra, and G. Newby, "Using the one-dimensional S-transform as a discrimination tool in classification of hyperspectral images," Canadian Journal of Remote Sensing, vol. 33, pp. 551-560, 2007.

[48] W. Elmahboub, "A combined methodology to produce highly accurate classification for AVIRIS hyperspectral data," Canadian Journal of Remote Sensing, vol. 35, pp. 321-335, 2009.

[49] P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson, "Random forests for land cover classification," Pattern Recognition Letters, vol. 27, pp. 294-300, 2006.

[50] Y. Wang and J. Li, "Feature-selection ability of the decision-tree algorithm and the impact of feature-selection/extraction on decision-tree results based on hyperspectral data," International Journal of Remote Sensing, vol. 29, pp. 2993-3010, 2008.

[51] G. Mountrakis, J. Im, and C. Ogole, "Support vector machines in remote sensing: A review," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 66, pp. 247-259, 2011.